Online Supplement to “Water Quality Awareness and Breastfeeding: Evidence of Health Behavior Change in Bangladesh” Pinar Keskin, Gauri Kartini Shastry & Helen Willis

1. Provision of information and health behavior change

In this appendix, we discuss two puzzles related to this evidence of behavior change. One explanation for why mothers choose to breastfeed longer instead of switching to an arsenic-safe well (which would protect more members of the family) is because breastfeeding is within their sphere of influence while the choice of water source is not. We discuss ethnographic evidence on intra-household decisions from Bangladesh and provide evidence of heterogeneous responses among women in our sample that is consistent with this explanation. Next, we discuss features of the information campaign that may explain why the Bangladesh information campaign was able to elicit behavior change, while many other interventions have had little success.

1.A. Spheres of influence and heterogeneous effects.

Ethnographic research conducted by Farhana Sultana (2009, 2011) describes a household structure where men and women are both responsible for providing safe drinking water, but in distinct ways. Sultana studied 18 villages between 2003 and 2005 in four districts that were acutely affected by arsenic. She describes the practice of well-sharing in detail and how geographic variation in access to safe water affected power dynamics between households. Within the household, even though women do most of the water collection, male heads own the tubewells and assume responsibility for providing a tubewell with arsenic-safe water. Men expressed anxiety about not being able to purchase a deep tubewell to provide safe water. When unable to provide safe water within the household, men still exert authority over where the women (usually the youngest) collect water. Traditional social norms frown on women travelling into public spaces, and for some men (and also older women), these norms outweighed the need for safe water. Women reported social and emotional stress when they had to fetch water from sources that did not belong to them, often requiring them to navigate social hierarchies and power relations. Many women reported both verbal and physical conflicts that were exacerbated when safe water sources were overcrowded. Sultana also notes that pre-existing social networks play an important role in well-sharing. Owners of safe wells restrict use of their wells even among relatives, although they do feel some obligation to share water, especially for uses such as drinking or cooking and especially to mothers. Corroborating this ethnographic evidence, Inauen et al. (2013) survey households with access to different safe water options (piped water, deep wells, well-sharing, etc.) and find that well-sharing has one of the lowest satisfaction levels on dimensions such as taste, time to collect water and whether users felt others disapproved.

In such a setting, it is not surprising that women may resort to other means to avoid arsenic-contaminated water when available, such as breastfeeding for young children. There is a large literature suggesting that women choose to allocate more resources to child health and nutrition, relative to their husbands (Duflo 2003, Qian 2008, Miller and Mobarak 2013). We present heterogeneity among women in the two responses to the campaign that is consistent with the idea that women respond by breastfeeding more because switching to an arsenic-free well is outside their control (table A5). Specifically, we demonstrate differential impacts of the campaign by measures of maternal social capital, education and autonomy.

The top panel of table A5 repeats the exercise in table 2 – predicting which households continue to use arsenic-contaminated water or drink surface water using measures of arsenic levels and access to clean wells – but with triple interactions between arsenic, access to clean wells and maternal characteristics. Column 1 looks at whether mothers who participate in voluntary organizations such as Grameen Bank or mother's clubs organized by the Bangladesh Rural Advancement Committee are differentially responsive to the campaign on these two dimensions (using arsenic-safe water and breastfeeding longer). Panel A shows that women who participate in such clubs (25% of the mothers in the sample) are more likely to be using an arsenic-safe water source (the coefficient is negative because they are less likely to be using arsenic contaminated or surface water). Women who participate in such clubs are likely to have stronger social networks, either because they are better at maintaining relationships or because their households allow them to socialize outside the home. Either way, this is consistent with Sultana's documentation of the social challenges in accessing arsenic-safe water. Panel B of table A5 replicates our main result from table 4 – estimating how much longer mothers breastfeed their children after the campaign in contaminated areas using the DID strategy – but with triple interaction terms between "post," contamination and maternal characteristics. Participation in clubs does not affect the breastfeeding response to the campaign.

Column 2 of table A5 demonstrates heterogeneity in both responses with respect to maternal education, although the coefficients are statistically significant only at the 10% level. Panel A shows that mothers with any formal education (60% of mothers) are more likely to use clean water sources. This result could be due to greater decision-making authority on the part of the women, better social skills to facilitate well-sharing or the fact that more educated women are likely to have married into households with different preferences. It could also be that more educated women are better able to internalize the health related information provided by the campaign (Grossman 1972, Schultz 1975, Rosenzweig 1995). Dupas (2011b) discusses the complementarity between information provision and education, noting that behavior change will be difficult if people are unable to process the information. Much of the evidence for this health-cognition gradient is from developed countries (see Grossman 2000, De Walque 2004 about smoking and Rosenzweig and Schultz 1989 about contraception in the U.S.), but there is a small, growing literature from developing countries (for example, see De Walque 2007 about HIV/AIDS prevention in Uganda).

Similarly, one would expect that educated women are more likely to make the connection between arsenic and protection provided by breastfeeding on their own if, indeed, the campaign did not mention it. The fact that breastfeeding provides protection for infants in areas with poor sanitation is not new information in Bangladesh, but educated women may be more likely to have processed it. Panel B column 2 shows that more educated women were also more likely to breastfeed for longer periods after the information campaign.

At the same time, in order to change their breastfeeding practices to protect children from contaminated water, these mothers would have to have some decision-making power. Measuring decision-making authority within the household is notoriously difficult. Our measure is based on the question: “Who in your family usually has the final say on the following decisions: your own health care, making large household purchases, making household purchases for daily needs, visits to family, friends, or relatives and what food should be cooked each day.” Women can either respond: themselves, just their husband, themselves together with their husband, someone else, or themselves together with a third person. We add up the number of areas in which the respondent feels that she has some say and then divide the sample into women with above and below average decision-making power. While this could measure decision-making ability within the household, it could also measure decision-making ability in general: whether women feel they have control over things that happen to them and their families. Column 3 shows that women who self-report that they are involved in more decisions are more likely to increase breastfeeding duration in response to the campaign (panel B). This finding supports the hypothesis of multiple decision-makers and separate spheres: otherwise, our results would be consistent with an explanation where men make all household decisions including how much their wives breastfeed and the breastfeeding response is preferred (by men) to using a clean well only because of the taboo against women in public spaces. These women are not more likely to be using a clean well (panel A), but this is not entirely surprising given Sultana's description since the choice of well is outside the women's sphere of influence. Interestingly, this measure of autonomy is not strongly correlated with education.

Finally, in column 4, we include interactions with all three maternal characteristics to ensure that our results are not driven by correlations between the characteristics. Our results persist, but we acknowledge that we cannot claim definitely that these heterogeneous effects are not picking up differences in responses due to unobservable characteristics correlated with participation in clubs, formal education or decision-making ability.

1.B. Distinctive features of the National Arsenic Communication Strategy Campaign

Next, we describe some aspects of the water quality information campaign that might partly explain its success, in light of recent empirical evidence of slow changes in health behavior in other contexts. The campaign was surprisingly effective at generating awareness and appreciation of the severity of arsenicosis, particularly since most symptoms appear in the long run. Mobarak et al. (2012) point out that despite the well-publicized health risks of indoor air pollution, 76% of respondents claimed that smoke from traditional cook stoves is less harmful than arsenic-contaminated drinking water. They also document a relatively high willingness to pay for arsenic-free water access compared to other important goods and services, including an electricity connection, clean cook stoves or sanitary latrines. One explanation is that arsenic poisoning is more frightening than other health risks, for some unknown reason. Another explanation is that some feature of this campaign made it particularly effective, such as comprehensive information provision, a constant visual reminder, or a communal design.

1. Comprehensive Information: Unlike many health campaigns in developing countries, the campaign in Bangladesh gave villagers comprehensive information. For example, instead of focusing on one type of preventative behavior, the DPHE, UNICEF and local governments identified many different mitigation methods (Inauen et al. 2013), aided by the blanket testing and labelling of tubewells. Importantly, the different mitigation methods varied in cost and included some with relatively small financial costs, such as well-sharing, in addition to more costly methods (installing a deep tubewell). Providing comprehensive information has proven to be more effective in other contexts as well. Duflo et al. (2015) found that when an HIV prevention campaign focused only on abstinence, young women got married earlier. Dupas (2011a) concluded that informing girls about the relative riskiness of partners by different age was more effective in reducing early pregnancies.

2. Visual and Continual Reminders: The color-coding of the wells provided a visual and lasting reminder that men and women saw every day and throughout the village. Previous research has also found visual reminders to be effective in preventative health campaigns. In a randomized control trial in Kenya, Luoto (2009) shared information on fecal contamination near drinking water sources with households while providing free water treatment products. Some households were asked to make a public commitment to treat their drinking water and given a poster as a visual reminder. Luoto found that this intervention increased water treatment. Visual reminders are being increasingly used in the design of large-scale public health interventions in developed countries. Kessler and Zhang (2014) summarize recent studies in behavioral and health economics and discuss how visual cues can affect individual's health related decisions.

3. Communal Persuasion: The National Arsenic Communication Strategy Campaign disseminated information mostly through public forums, instead of individual household visits. For example, testing a well takes approximately 30 minutes and tubewell testers generally spent that time providing information about arsenic. Thus, users were aware of the messages that other households received. Combined with the public, visual color-coding of wells, it is possible that sourcing water from a red-painted well developed its own stigma. Kremer et al. (2009) offer similar evidence that a community meeting design might be more successful than household visits in increasing treatment of household drinking water with chlorine disinfectant.

2. BAMWSP Results

The British Geological Survey (BGS), from which we calculate measures of arsenic contamination and measures of access to clean wells, only tested 3,534 wells out of the approximately 8 million wells in the country in 1998-99. While the sampling was nationally representative, this small sample might lead to measurement error in our independent variables of interest. In this appendix, we compare our measures to alternative measures calculated from another dataset that includes arsenic levels for 4.5 million wells across contaminated regions of the country, gathered by the Bangladeshi government. This new dataset was collected as part of the Bangladesh Arsenic Mitigation Water Supply Program (BAMWSP). We do not use this data as our primary data for multiple reasons. Most importantly, the BAMWSP data does not include GPS coordinates for each well making it difficult to generate precise local measures of arsenic exposure and access to uncontaminated wells. The data does provide the village of the well and the mouza (an administrative unit just above village), which we will use to generate the new measures described in this appendix. Other reasons we do not focus on the BAMWSP data are i) we do not know when the wells were built, possibly introducing some reverse causality, and ii) the data does not cover the entire country.

After describing the data and the creation of the alternative contamination measures, we first demonstrate that these measures are closely related to our BGS-sourced measures. We then re-estimate our specifications using this new data. Our DID results are robust to using these alternative measures of contamination – in fact, the magnitudes are remarkably similar. While the triple difference results are robust to some combinations of measures, other combinations suffer from limited power due to the lack of variation in access to clean wells, conditional on contamination. We believe this is driven by the fact that we do not have the exact locations for the 4.5 million wells, limiting our use of very local variation; this causes very high correlations between arsenic contamination and access to clean wells, making it difficult to identify the separate effects in a regression. We discuss this issue further below.

2.A. BAMWSP data: measures of arsenic contamination

While the BAMWSP data does not contain GPS coordinates for each well, it does list the well's village and mouza location. We use GIS shapefiles for the mouzas (CEGIS 2002) to assign the wells to clusters (villages) from the Bangladesh Demographic and Health Survey (BDHS). Bangladesh is divided into 64 districts, which are further divided into around 500 upazilas, which are further divided into approximately 60,000 mouzas. Mouzas are 3 square kilometers in size on average and contain multiple villages; the best sources of information we have found suggest that most mouzas contain between 1 and 5 villages, but some contain many more (BAMWSP data, authors' calculations). One concern with the BAMWSP data is that it does not cover all mouzas in the vicinity of a BDHS cluster location and it is not clear whether all villages in a mouza were visited. While some areas with no data are arsenic-safe as indicated on a map of this data (the Welltracker map) produced by the Earth Institute at Columbia University (van Geen and Immel 2006) and confirmed with the nationally representative BGS information, there are many mouzas in contaminated regions which were not visited by BAMWSP as well.

For our first BAMWSP measure of arsenic contamination, we draw a 5 km circle around each cluster's GPS location and calculate the percent of wells in the mouzas that intersect this circle (some may lie only partly inside the circle) that are contaminated (arsenic levels at or above 50 µg/L). Note that some mouzas in this circle may not have any BAMWSP tested wells, only some of which are labeled arsenic-safe according to the Welltracker map (see figure A4). Thus, we then multiply this fraction with the fraction of the 5 km circle that is not considered arsenic-safe. The assumptions this makes are i) mouzas that are bisected by the 5 km circle have uniformly distributed contamination levels, ii) average contamination levels of mouzas with data (darkly shaded in figure A4) are representative of the contamination levels of those mouzas that have no BAMWSP data but are not labeled arsenic-safe (medium shading in figure A4) iii) no wells in mouzas labeled arsenic-safe are contaminated and iv) the well density of the mouzas with BAMWSP data is representative of all the mouzas, regardless of whether they are labeled arsenic-safe.

The second BAMWSP measure of arsenic contamination uses a GIS exact interpolation tool called Inverse Distance Weighting (IDW) that predicts the arsenic level at approximately 1 million GPS locations across Bangladesh using the BAMWSP data. First, we calculate the average arsenic level of all wells in a mouza and assign this value to the center of the mouza. The IDW process then predicts the arsenic level at each GPS location by calculating a weighted average of arsenic levels of mouza centers near each GPS location, where the weights are the distance between the mouza center and the GPS locations. The explicit assumption that objects closer to each other are more similar than objects that are far apart is the only one made in implementing IDW. Using this interpolation, we calculate the fraction of GPS locations (as a proxy for households) within 5 km of the cluster location that are contaminated.

In table A6, we regress the new BAMWSP measures on five BGS measures. The two sets of measures are strongly related to each other, with correlation coefficients ranging from 0.69 to 0.78. Table A7 replicates table 1 but using the BAMWSP measures of arsenic contamination.

Clearly, these measures of arsenic are also picking up something real: households in more contaminated areas are more likely to have heard of arsenic (Column 1), more likely to still be using contaminated wells (despite being more than half way into the campaign) and more likely to have a painted well and, specifically a red-painted well (Columns 4 and 5).

2.B. Difference in difference results with BAMWSP measures

Next, we re-estimate specification (1) using the BAMWSP measures of arsenic contamination in table A8. As in table 4, children born in clusters exposed to arsenic in 2002 or later are breastfed longer than children born before 2002, relative to children born in clusters less exposed to arsenic. Of note is the fact that the magnitude of the effect is remarkably similar. One standard deviation in the original BGS measures of contamination was approximately 10%, an increase of which would cause mothers in rural areas to breastfeed their children approximately 0.6 months longer (table 4, panel A, column 6). One standard deviation in the first BAMWSP measure is 28%, an increase of which would lead to 0.57 month increase in breastfeeding duration. A one standard deviation increase in the second BAMWSP measure (34% points) would lead to a 0.54 month increase in breastfeeding duration. While these magnitudes may seem small, a 10% point increase in the BGS measure of contamination increases the probability of a child being exclusively breastfed by 2.84% points on a basis of 12%. Using the two BAMWSP measures, a one standard deviation increase in exposure increases exclusive breastfeeding by 2.91% points and 2.61% points, respectively. This is a non-trivial effect given the low incidence of exclusive breastfeeding overall. As in table 4, the results are driven by rural areas with small (or negative) and insignificant coefficients estimated for urban areas. Estimating the event study and plotting the coefficients gives us graphs that look very similar to figure 3 (see figures A5 and A6). As before, we can strongly reject the hypothesis that the average of the coefficients in the pre period is equal to the average of the coefficients in the post period.

Table A9 replicates table 5 using the two new BAMWSP measures. While our overall results for whether children are exclusively breastfed are robust to the two new BAMWSP measures (table A8, panel C), we only see a statistically significant coefficient (at 10%) for whether children under the age of 6 months are exclusively breastfed using the second BAMWSP measure. Again, the magnitude of that coefficient suggests a similar sized effect to our original BGS measures. Our results for whether children between 6 and 14 months are given any plain water are robust to both BAMWSP measures and give us similar sized effects.

Table A10 re-estimates the effects on child health from table 7 with the two new BAMWSP measures. In the interest of space we replicate only the first three columns from table 7, mortality until age 24 months and health status until 36 months. Our health results are quite strongly robust to these measures: we see reductions in mortality and diarrheal incidence and an increase in weight for height for infants (those most likely to be breastfeeding exclusively). (The reduction in mortality is significant until age 24 months, but is driven by infant mortality: if we redefine the outcome variable to be mortality between 12 and 24 months, the interaction term is positive and significant at 10% for the second BAMWSP measure, consistent with the findings in Field, Glennerster and Hussam 2011.) Again, the magnitudes of the effects are remarkably similar: a one standard deviation increase in exposure results in a 1.1%, 1.4% or 1.7% point reduction in infant mortality and a 4.3%, 4.8% and 6% point reduction in diarrheal incidence using the BGS measure and the two BAMWSPmeasures, respectively.

2.C. BAMWSP measures of access to clean wells

As mentioned above, estimating our triple difference specification is made difficult by the lack of variation in access to clean wells, conditional on exposure to arsenic, in the BAMWSP data. This lack of variation is possibly due to the fact that we do not have the exact GPS locations for the wells, limiting our use of very local variation. It is not ex-ante obvious whether measures from the BGS data (small sample of wells, but with precise GPS location) or from the BAMWSP data (large sample of wells, but without GPS location) would better capture a household's access to uncontaminated wells. It is worth considering what the ideal measures would be. The best village-level measure of contamination would be the fraction of households for which the closest well (usually within the compound of the residence) is contaminated with arsenic. The fraction of BAMWSP sampled wells that are contaminated provides a good measure of this. The best village-level measure of access to clean wells, however, would be a measure of the distance a household member would need to travel in order to reach an uncontaminated well. It is not difficult to imagine that with enough local variation in arsenic contamination, there could be variation in access to clean wells even conditional on the contamination. However, it may not be possible to accurately estimate or simulate this measure at the local level without

GPS coordinates of each well. Here, we describe three approaches to estimating village-level access to clean wells. The measures of access to clean wells from the BAMWSP data are very highly correlated with the BAMWSP measures of contamination and not all combinations provide enough independent variation to separately identify both in a regression (which is not a problem with the BGS measures). In fact, the BGS measures do a somewhat better job of predicting household behavior (from the BDHS) than the BAMWSP measures.

The first measure we use was developed originally in Gelman et al. (2004) and further in van Geen et al. (2006). These papers use the BAMWSP data to estimate "safe-depth thresholds" for each village: that is, a well depth at which they estimate the probability of accessing arsenic-contaminated water to be very low. The search algorithm used data from spatial clusters of ~75 wells and identifies two possible safe-depth thresholds: between the deepest unsafe well in a cluster and the shallowest safe well that is strictly deeper than it, and the analogous next-deepest unsafe-safe well pair. The final safe-depth threshold is the depth of the safe well of the pair with the higher probability that a well deeper than it is safe based on a posterior probability model. The algorithm to calculate the safe-depth threshold was first developed in Gelman et al. (2004) but this paper focused on the small area in Araihazar district studied extensively by researchers at Columbia University. van Geen et al. (2006) extend this algorithm to all BAMWSP villages.[1]

To aggregate the safe depth measure from the BAMWSP village level to the BDHS cluster level, we first calculate the average safe depth for villages within a mouza. We then calculate the weighted average of the mouza mean safe depth across all mouzas that intersect the 5 km circle around each cluster's GPS location, where the weights are the fraction of the circle covered by the particular mouza. As with our first contamination measure described above, we make the assumptions that i) average access levels of mouzas with any data are representative of the access levels of those mouzas that have no BAMWSP data but are not labeled arsenic-safe in the Welltracker map, and ii) the safe depth threshold for mouzas labeled arsenic-safe is 0 feet. The main difference in how we aggregate from the mouza to the 5 km cluster circle level is that for the exposure measure, mouzas are weighted by the number of wells tested while for the safe depth measure, mouzas are weighted by the area of the 5 km cluster circle covered by the mouza.

[1] Personal communications with Alexander van Geen indicates that his research team is not very satisfied with the properties of this measure anymore, but since it does predict water source switching behavior in the 2004 BDHS data, we include it here for completeness.

The second measure uses the IDW results described above that predicted the arsenic level at 1 million GPS locations across Bangladesh. From this interpolation, we calculate the distance between each GPS location (as a proxy for a household) and the nearest GPS location that is not contaminated (as a proxy for a clean well). We then calculate the average distance across all GPS locations within 5 km of a cluster location.

The third measure we calculate randomly assigns contaminated and uncontaminated wells to GPS locations within each mouza (using the number of wells tested and the number found to be contaminated from the BAMWSP data) and then calculates the minimum distance between each GPS location (as a proxy for a household) and the nearest GPS location assigned a clean well. We repeat this simulation 20 times and take the average as a measure of distance to clean wells at the mouza level. We aggregate this measure from the mouza level to the 5 km cluster circle in the same manner as the safe depth measure. Mouzas that are labeled arsenic-safe in the Welltracker map are assigned a distance to the nearest clean well of 0 meters and the mouza-level mean distance is weighted by how much of the 5 km cluster circle is covered by this particular mouza.

The second measure (interpolation) and this third measure (simulation) have slightly different strengths: the interpolation measure uses the information across mouzas (average arsenic level of all tested wells), but not information within a mouza, such as how many wells were tested, how many were contaminated or the shape of the mouza. The simulation measure uses this latter within-mouza information, but assigns contaminated and uncontaminated wells to GPS locations uniformly, without using information from neighboring mouzas. A fourth measure combines the two approaches: first, we use the IDW tool to predict the probability that a well at each of the 1 million GPS locations is contaminated. Then we use these probabilities to assign the contaminated and uncontaminated wells to the GPS locations within a mouza. However, the data is not fine enough to predict large enough variation in the probability that a well is contaminated for this fourth measure to differ significantly from the third measure.

2.D. Relationship between BAMWSP and BGS measures of access to clean wells

In table A11, we regress the new BAMWSP measures on the four BGS measures of access to clean wells, separately. Each regression demonstrates a statistically significant relationship between the two measures (significant at the 1% level). The correlation coefficients between measures from the two sources are not as high as for the measures of exposure (they range from 0.1183 to 0.3428 in absolute value), however, kernel-weighted local polynomials of the relationship between the BAMWSP measures and the BGS measures demonstrate close relationships. For example, figure A7 plots the BGS measures, the probability of living within 1 mile of a clean well (weighted and unweighted), on the Y axis and the simulated minimum distance to a clean well from the BAMWSP data on the X axis, with bootstrapped confidence intervals. The two measures of access to clean wells are closely related. In the interest of space, we omit similar looking plots documenting the relationship between with the interpolated BAMWSP measure and between the other BGS measure (the average distance to the closest uncontaminated well) with any of the BAMWSP measures.[2] Foreshadowing the difficulty in identifying the effect of exposure and access separately, we also find that the BAMWSP measures of access are highly correlated with the BAMWSP measures of exposure (correlation coefficients ranging from 0.53 to 0.91).

We next demonstrate that the BAMWSP measures of access to clean wells are able to predict switching behavior in the 2004 BDHS (by replicating figure 2 and table 2), but we also show that the new measures are not any better at doing this than the original BGS measures. Thus, while having information on arsenic contamination for a much larger number of wells should improve the accuracy of our measures, the lack of GPS information on the location of the wells hinders our ability to estimate variation in access to clean wells, conditional on exposure to arsenic. Figures A9, A10 and A11 replicate figure 2 using the three BAMWSP measures of access to clean wells. The relationship between the safe depth threshold and a household's choice of water source seems reasonably linear after taking the natural log of the depth measure (figure A9). However, the relationship between the IDW measure of average distance and a household's choice of water sources does not appear to be monotonic or linear. Recall that the IDW measure uses more of the variation across mouzas than variation within mouzas, unlike the simulated measure; the fact that more than 97% of the mouzas covered by the BAMWSP data have at least one clean well suggests that the simulated measure should be able to capture local variation in access to clean wells better. Figure A11 demonstrates that the average distance to a

[2] Even though the BGS measures do strongly predict the BAMWSP safe depth measure in table A11, the relationship does not appear to be monotonic or linear (see figure A8). Recall that the authors of the safe depth measures are not entirely convinced by this measure anymore; we include it because it does predict switching behavior (see figure A9, which replicates figure 2 with the safe depth measure and the natural log of the safe depth threshold, after adding 1 to deal with the zeros).

clean well, simulated using the number of clean and contaminated wells in each mouza according to the BAMWSP data, monotonically predicts whether a household sources water from a contaminated well or a surface source.

Table A12 extends table 2 to include both BGS and BAMWSP measures of exposure and access to clean wells. Each cell is from a regression of whether the BDHS household uses arsenic-contaminated well water or surface water for drinking (in panel A) or just arsenic-contaminated well water (in panel B) on a measure of arsenic exposure, a measure of access to clean wells and the interaction of the two. The regressions also include district fixed effects and the standard errors are clustered by BDHS cluster. In the interest of space, we focus on the unweighted measures of access to clean water from the BGS data and the BAMWSP measures in levels, not logs (the results are very similar). The top row in panel A uses our preferred BGS measure of arsenic contamination: the first two columns reproduce the coefficients on the interaction term from columns (1) and (3) from table 2. The next two rows demonstrate that the choice of water source does respond to the BGS measures of access to clean water as we would expect even when we use measures of arsenic exposure from the BAMWSP data. The next three columns, however, suggest that the BAMWSP measures of access to clean water do not robustly predict whether the household uses a dirty water source. In Panel B, however, the interaction terms in Columns (3) to (5) are often statistically significant at the 5% level. Households are more likely to be using an arsenic-contaminated well the deeper they would have to dig or the farther they would have to walk to access clean water. The results in this table provide support for the BGS measures of access to clean water. The BGS measures are the most consistent in predicting a household’s water source and the R-squared from all the regressions are in the same ballpark, suggesting that the BAMWSP measures are no better at predicting switching behavior.

2.E. Triple difference results with BAMWSP measures of access to clean wells

Finally, we replicate the triple difference results in table 6 using the BAMWSP measures of access to clean wells in table A13. Panel A uses our preferred BGS measure of arsenic exposure, panel B uses the first BAMWSP measure of arsenic exposure (the fraction of contaminated wells) and panel C uses the second BAMWSP measure of arsenic exposure (the interpolation). Column (1) uses the safe depth measure, column (2) uses the interpolation measure and column (3) uses the simulated measure of access to clean wells. Using a log transformation of the access to clean wells measure does not change the results appreciably (results available upon request). Since the three BAMWSP measures of access to clean wells are decreasing in access, we would expect the coefficients on the triple interaction terms to be positive, which they are for the most part. In panel A, the coefficient on the main difference-in-difference effect is not statistically significant in any column: mothers in clusters with the best access to clean wells do not respond to the information campaign by increasing breastfeeding duration. The triple interaction terms are statistically significant at the 5% level for safe depth measure as well as the simulated measure: mothers in clusters with worse access to clean wells increase the duration of breastfeeding. Both of these measures use more of the within mouza information relative to the interpolation measure, which mainly uses across mouza information on average arsenic levels. Only one of the triple interaction terms in panels B and C is statistically significant and only at the 10% level. However, note that the correlation coefficient between the measure of arsenic contamination and access to clean wells for each regression is very high (shown at the bottom of each panel, ranging from 0.53 to 0.91): we suspect there is too little variation in the BAMWSP access to clean wells measures, conditional on contamination, for us to identify the differential effect. Note also that the main DID coefficient is usually statistically significant and the joint F-test of the difference-in-difference coefficient and the triple interaction term often rejects the null hypothesis that variation in arsenic contamination (both exposure and access to clean wells) does not affect a mother's breastfeeding choices at the 10% level. It is also possible that the (decreasing) measure of access is simply picking up actual exposure to arsenic, such as noise in our measure of arsenic contamination.

The fact that we are unable to replicate the triple difference results using just the BAMWSP data can be interpreted in two ways: 1) the BGS measures of access to clean water are better at allowing us to identify variation in access conditional on contamination or 2) the BGS triple difference results are spurious. Two results from this analysis support the first interpretation over the second: First, the BAMWSP measures of access to clean water are no better at predicting water sourcing behavior from the 2004 BDHS survey than are the BGS measures of access to clean water (table A12). Second, regressions that use the BAMWSP measures of access to clean wells with the original BGS measure of arsenic exposure, with lower correlation coefficients between the two measures, support our original conclusion that the breastfeeding response is driven by mothers in clusters with less access to clean well water (table A13, panel A).

3. References

Bangladesh Arsenic Mitigation Water Supply Project (BAMWSP). 2004. “The Distribution of Arsenic in Groundwater in Bangladesh.” www.bamwsp.org

British Geological Survey and Department of Public Health Engineering (BGS and DPHE). 2001. “Arsenic contamination of groundwater in Bangladesh.” Edited by D.G. Kinniburgh and P.L. Smedley. Keyworth, U.K.: British Geological Survey. http://www.bgs.ac.uk/arsenic/bangladesh/ (accessed August 1, 2013).

Center for Environmental and Geographic Information Services (CEGIS). *Mauza Database*. 2002. http://www.cegisbd.com/

de Walque, D. 2004. “Education, Information, and Smoking Decisions: Evidence from Smoking Histories, 1940-2000.” World Bank Policy Research Working Paper 3362.

de Walque, D. 2007. “How Does the Impact of an HIV/AIDS Information Campaign Vary with Educational Attainment? Evidence from Rural Uganda.” *Journal of Development Economics*, 84: 686-714.

Duflo, E. 2003. Grandmothers and Granddaughters: Old Age Pensions and Intra-household Allocation in South Africa. *The World Bank Economic Review*, *17*(1), 1-25.

Duflo, E., P. Dupas and M. Kremer. 2015. “Education, HIV and Early Fertility: Experimental Evidence from Kenya.” *American Economic Review*, 105 (9), 2757-97.

Dupas, P. 2011a. “Do Teenagers Respond to HIV Risk Information? Evidence from a Field Experiment in Kenya.” *American Economic Journal: Applied Economics,* 3 (1): 1–34.

Dupas, P. 2011b. “Health Behavior in Developing Countries.” *Annual Review of Economics*, 3 (1): 425-449.

Field, Erica, Rachel Glennerster and Reshmann Hussam. 2011. “Throwing the Baby out with the Drinking Water: Unintended Consequences of Arsenic Mitigation Efforts in Bangladesh.” http://scholar.harvard.edu/files/field/files/arsenic_infantmortality_feb10.pdf

Gelman, Andrew, Matilde Trevisani, Hao Lu, and Alexander van Geen. 2004. “Direct Data Manipulation for Local Decision Analysis as Applied to the Problem of Arsenic in Drinking Water from Tube Wells in Bangladesh.” *Risk Analysis* 24(6): 1597-1612.

Global Hazards Information Network (GHIN). National Geospatial-Intelligence Agency (vector digital data file bdg_pplp_nga). *Bangladesh: Populated Places.* (2007) http://www.pdc.org/geodata/unocha/bgd_populated_places.zip

Grossman, M. 1972. “On the Concept of Health Capital and the Demand for Health.” *Journal of Political Economy*, 80 (2): 223-255.

Grossman, M. 2000. “The Human Capital Model.” *Handbook of Health Economics*, 1, 347-408.

Inauen, J., M.M. Hossain, R.B. Johnston and H.J. Mosler. 2013. “Acceptance and Use of Eight Arsenic-Safe Drinking Water Options in Bangladesh.” *PLoS ONE*, 8 (1): e53640.

Kessler, J.B. and C. Yiwei Zhang. 2014. “Behavioral Economics and Health.” In preparation for the *Oxford Textbook of Public Health*. http://assets.wharton.upenn.edu/~czhan/KesslerZhang_BehavioralEconomicsHealth.pdf.

Kremer, M., E. Miguel, S. Mullainathan, C. Null, & A.P. Zwane. 2009. Making Water Safe: Price, Persuasion, Peers, Promoters, or Product Design. *Unpublished manuscript*.

Luoto, J. 2009. “Information and Persuasion: Achieving Safe Water Behavior in Kenya.” *Unpublished manuscript.*

Miller, G. and A. M. Mobarak. 2013. “Gender Differences in Preferences, Intra-Household Externalities, And Low Demand For Improved Cookstoves.” NBER Working Paper Series, Working Paper No. 18964.

Mobarak, A. M., P. Dwivedi, R. Bailis, L. Hildemann and G. Miller. 2012. “Low Demand for Nontraditional Cookstove Technologies.” *Proceedings of the National Academy of Sciences*, *109* (27), 10815-10820.

Qian, N. 2008. “Missing Women And The Price Of Tea In China: The Effect Of Sex-specific Earning On Sex Imbalance.” *The Quarterly Journal of Economics*, 123 (3): 1251-1285.

Rosenzweig, M. R. and T. P. Schultz. 1989. “Schooling, Information and Nonmarket Productivity: Contraceptive Use and Its Effectiveness.” *International Economic Review*, 30 (2): 457-477.

Rosenzweig, M. R. "Why are There Returns to Schooling?.” 1995. *American Economic Review*, 85(2), 153-158.

Schultz, T.W. 1975. “The Value of the Ability to Deal with Disequilibria.” *Journal of Economic Literature*, 13 (3): 827-846.

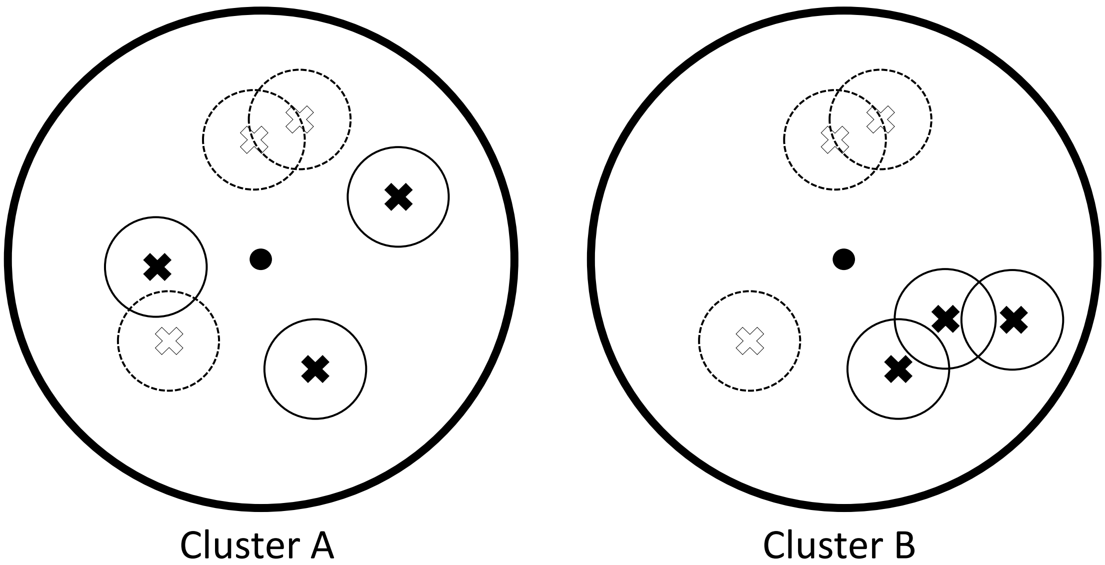
Sultana, F. 2009. “Fluid Lives: Subjectivities, Gender and Water in Rural Bangladesh.” *Gender, Place and Culture*, 16 (4): 427-444.

Sultana, F. 2011. “Suffering for Water, Suffering from Water: Emotional Geographies of Resource Access, Control and Conflict.” *Geoforum*, 42: 163-172.

van Geen, A., M. Trevisani, J. Immel, Md. Jakariya, N. Osman, Z. Cheng, A. Gelman, and K.M. Ahmed. 2006. “Targeting Low-arsenic Groundwater with Mobile-phone Technology in Araihazar, Bangladesh. *Journal of Health, Population, and Nutrition* 24 (3) : 282-297.

van Geen, Alexander and John Immel. 2006 "Arsenic Map of Bangladesh." *Columbia University Welltracker*. BAMWSP, Dhaka University, and Columbia University. http://www.ldeo.columbia.edu/res/pi/welltracker/map.php

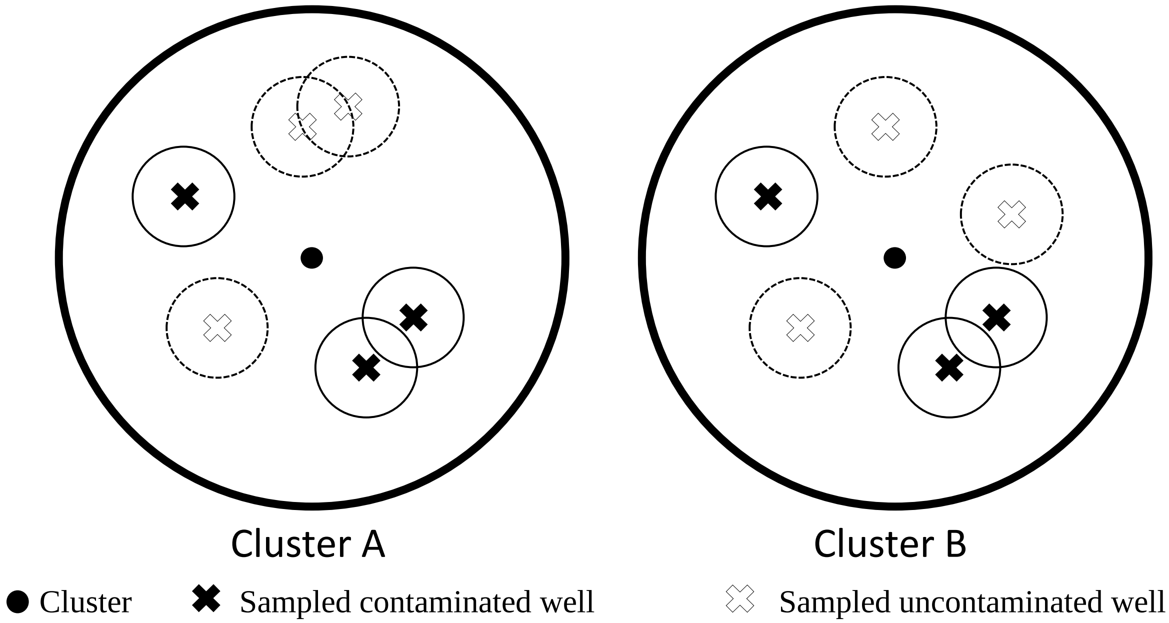
Appendix Figure A1: Calculation of Preferred Contamination Measure



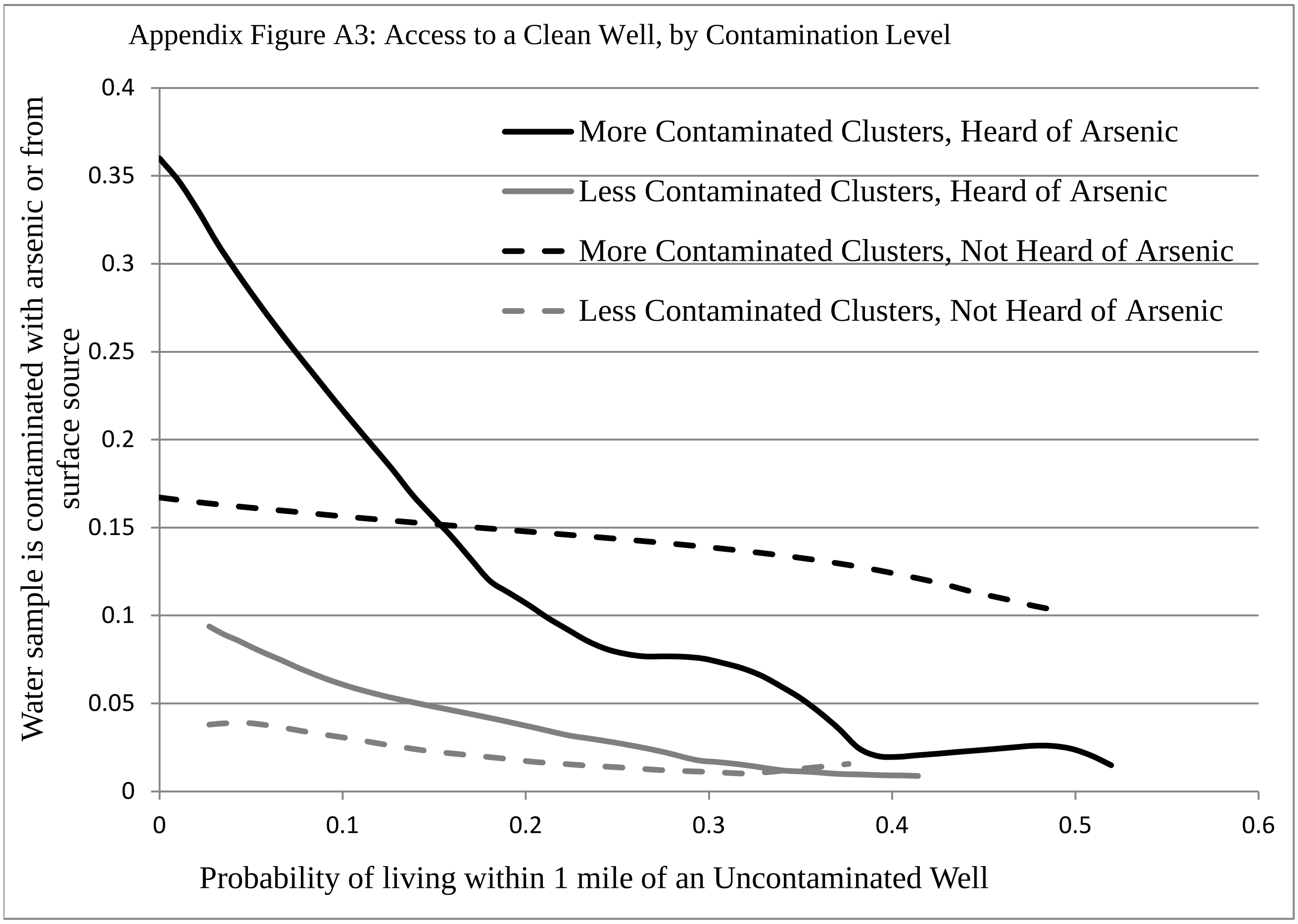
● Cluster ✖ Sampled contaminated well ✖ Sampled uncontaminated well

Note: This figure depicts the 5-mile circles around 2 fictitious cluster locations. Recall that the cluster's GPS location only approximates the location of the household because it is displaced and because it marks the closest enumeration point, not the household's residence. Suppose the BGS tested 6 wells within 5 miles of each cluster, 3 of which were contaminated (solid X's) and 3 were not (outline X's). Despite having the same number of contaminated and uncontaminated wells, a household in Cluster A is more likely to be exposed to arsenic contaminated drinking water than a household in Cluster B.

Appendix Figure A2: Variation in Access to Uncontaminated Wells, conditional on Contamination

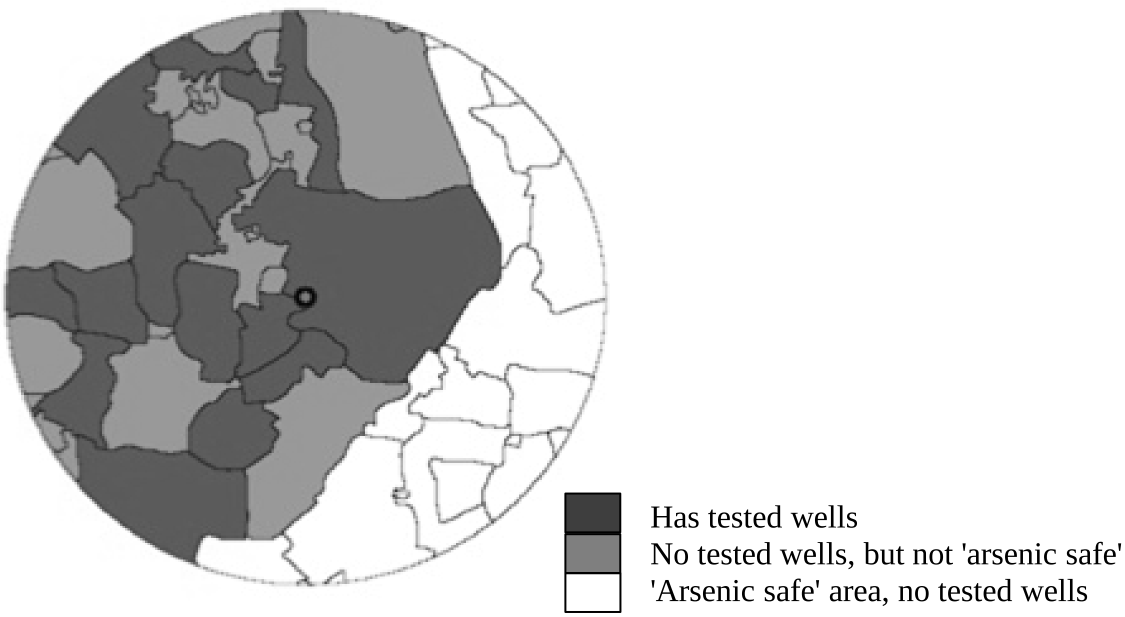


Note: This figure depicts the 5-mile circles around 2 fictitious cluster locations. Recall that the cluster's GPS location only approximates the location of the household because it is displaced and because it marks the closest enumeration point, not the household's residence. Suppose the BGS tested 6 wells within 5 miles of each cluster, 3 of which were contaminated (solid X's) and 3 were not (outline X's). Despite having the same number of contaminated and uncontaminated wells and the same placement of the contaminated wells, a household in Cluster A is less likely to have access to a clean well than a household in Cluster B.



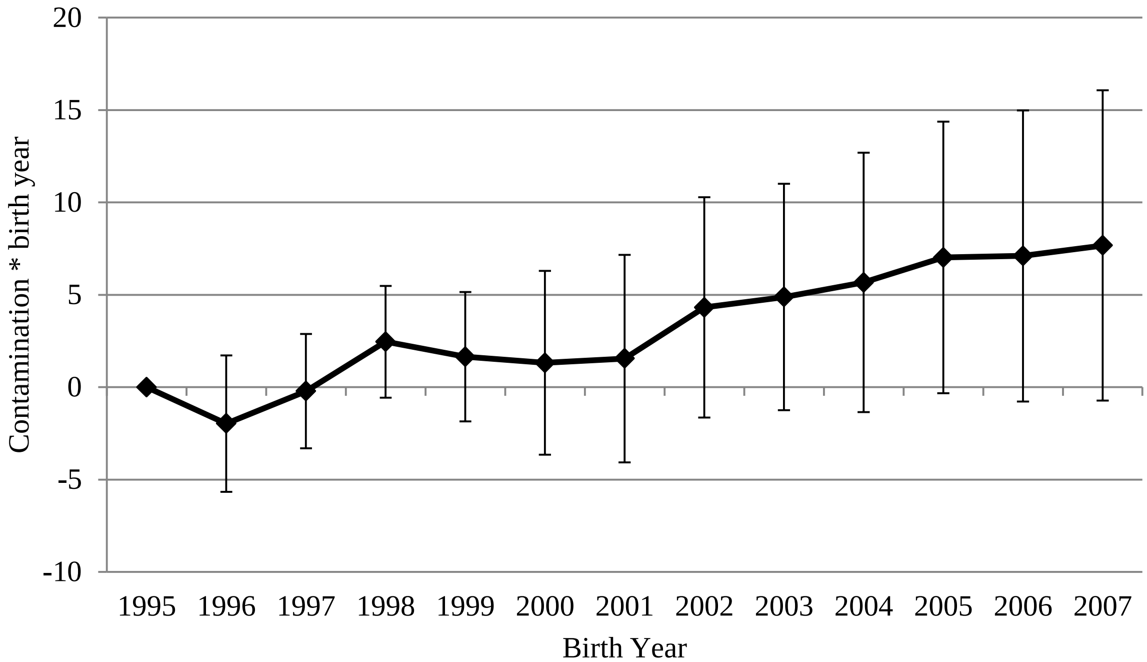
Note: This figure plots a Kernel-weighted local polynomial of the relationship between a household's access to a clean well and whether a household gets water from a contaminated well or a surface source. We exclude clusters with zero tested wells. The plots use an epanechnikov kernel.

Appendix Figure A4: BAMWSP Measure of Arsenic Contamination
Mouzas in a 5km circle centered at the DHS cluster GPS location



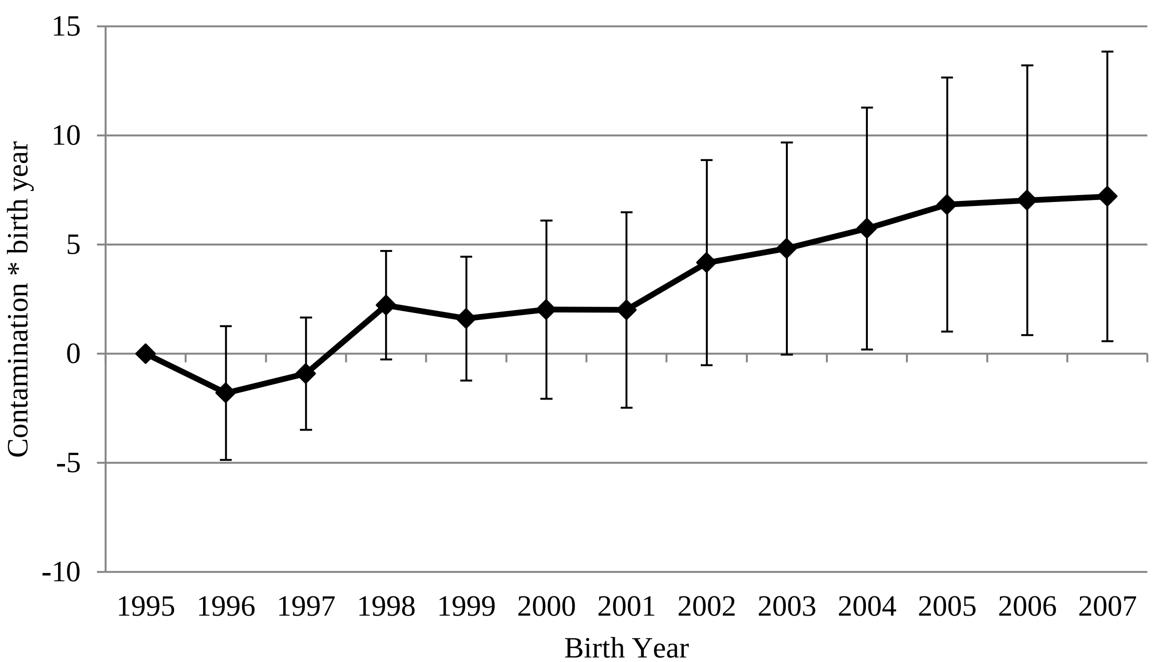
Note: The BAMWSP data does not cover all of the mouzas in the area around each DHS cluster GPS location. Some of these unvisited mouzas (the white area) have been determined to be arsenic safe (as seen from the BGS data and the Welltracker map), but there is no information available about the other unvisited mouzas (the light gray areas). The average arsenic contamination of mouzas that have tested wells (the dark gray areas) ranges above and below the safe level of 50 µg/L.

Appendix Figure A5: Months breastfed and first BAMWSP measure



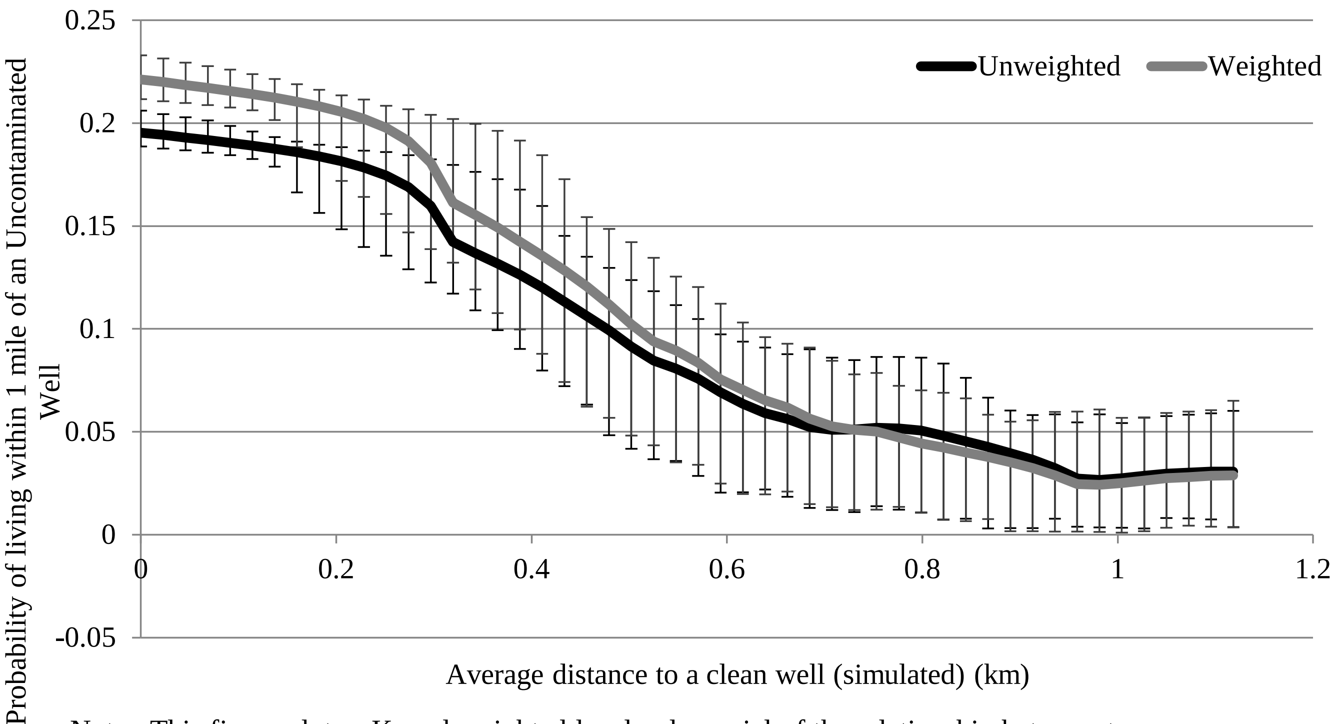
Note: This figure plots the coefficients on the interactions between birth year dummies and arsenic exposure in a regression similar to those in table A7. Only rural areas are included in this figure. The error bars are 95% confidence intervals, after clustering on BDHS cluster.

Appendix Figure A6: Months breastfed and second BAMWSP measure



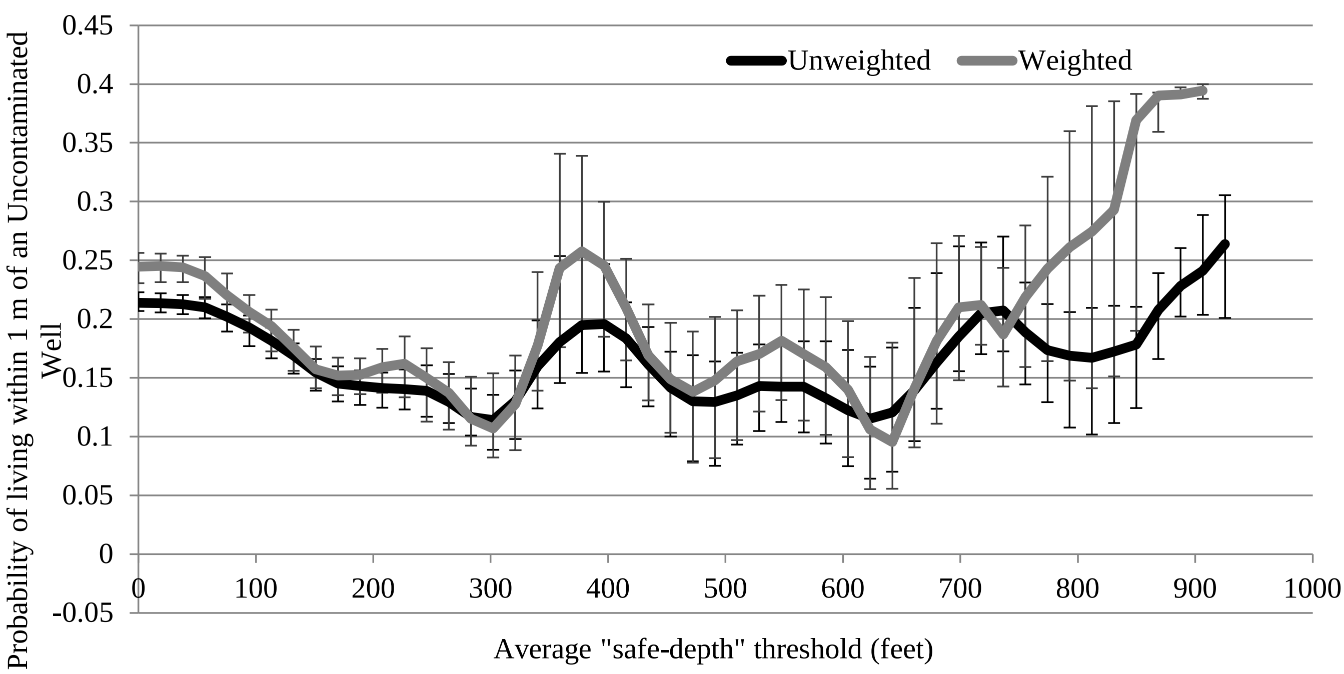
Note: This figure plots the coefficients on the interactions between birth year dummies and arsenic exposure in a regression similar to those in table A7. Only rural areas are included in this figure. The error bars are 95% confidence intervals, after clustering on BDHS cluster.

Appendix Figure A7: Measures of Access



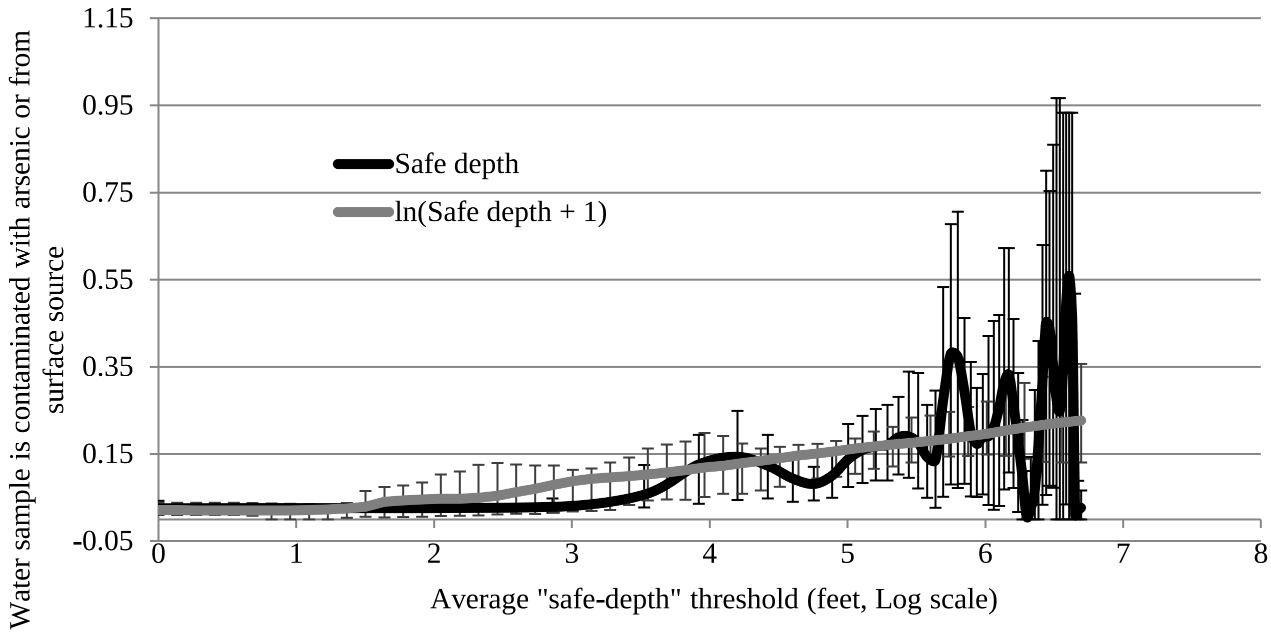
Note: This figure plots a Kernel-weighted local polynomial of the relationship between two measures of access to a clean well, with bootstrapped 95% confidence intervals and an epanechnikov kernel.

Appendix Figure A8: Measures of Access



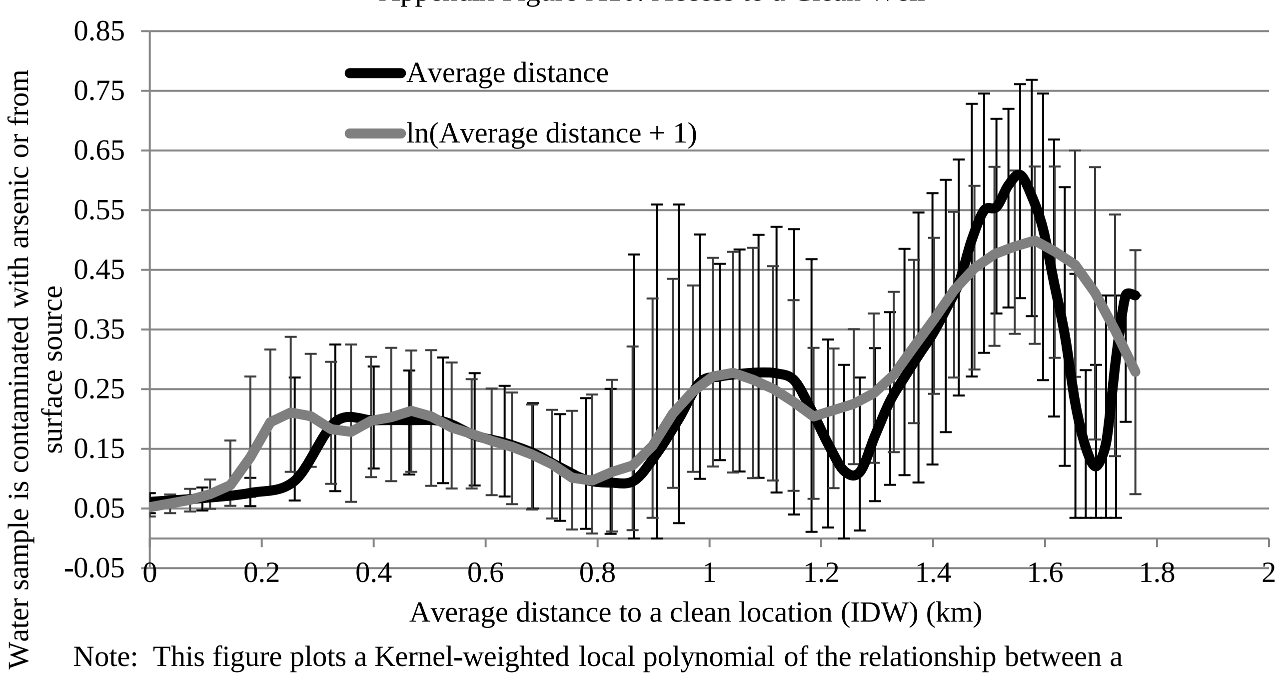
Note: This figure plots a Kernel-weighted local polynomial of the relationship between two measures of access to a clean well, with bootstrapped 95% confidence intervals and an epanechnikov kernel.

Appendix Figure A9: Access to a Clean Well



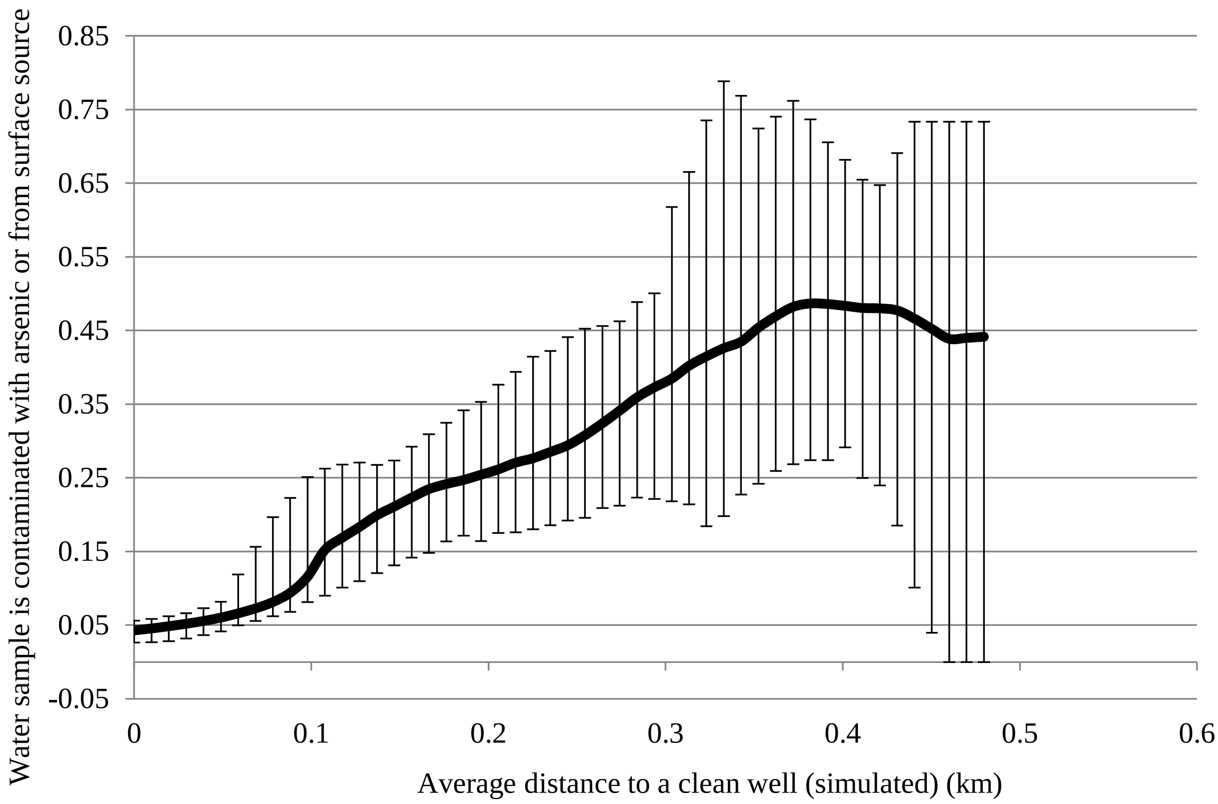
Note: This figure plots a Kernel-weighted local polynomial of the relationship between a household's access to a clean well and whether a household gets water from a contaminated well or surface sources, with cluster-bootstrapped 95% confidence intervals. The plots uses an epanechnikov kernel.

Appendix Figure A10: Access to a Clean Well



Note: This figure plots a Kernel-weighted local polynomial of the relationship between a household's access to a clean well and whether a household gets water from a contaminated well or surface sources, with cluster-bootstrapped 95% confidence intervals. The plots uses an epanechnikov kernel.

Appendix Figure A11: Access to a Clean Well



Note: This figure plots a Kernel-weighted local polynomial of the relationship between a household's access to a clean well and whether a household gets water from a contaminated well or surface sources, with cluster-bootstrapped 95% confidence intervals. We drop the three highest values which appear to be outliers. The plots uses an epanechnikov kernel.

Appendix table A1. Testing robustness to different measures of arsenic exposure

	(1)	(2)	(3)	(4)	(5)	(6)
Areas included	All	Urban	Rural	All	Urban	Rural
Panel A: Months Breastfed						
Post*probability contaminated	7.205***	3.569	8.853***	6.778***	2.334	7.147***
(unweighted)	(2.432)	(5.509)	(2.682)	(2.260)	(5.152)	(2.422)
Post*fraction contaminated	1.894**	0.110	2.691***	1.473**	0.121	1.788**
	(0.770)	(1.762)	(0.844)	(0.705)	(1.657)	(0.731)
Post*number contaminated	0.230**	0.0595	0.302***	0.219***	0.0164	0.247***
	(0.0890)	(0.202)	(0.0983)	(0.0815)	(0.188)	(0.0869)
Post*average As level (g/L)	7.375**	1.368	9.844***	4.593	-0.310	5.833*
	(3.141)	(6.603)	(3.613)	(2.991)	(6.410)	(3.281)
Panel B: Exclusive Breastfeeding						
Post*probability contaminated	0.264**	0.0768	0.321**	0.177	-0.0755	0.269**
(unweighted)	(0.123)	(0.234)	(0.141)	(0.121)	(0.246)	(0.132)
Post*fraction contaminated	0.102**	0.0497	0.111**	0.0499	-0.0201	0.0657
	(0.0408)	(0.0857)	(0.0472)	(0.0374)	(0.0815)	(0.0421)
Post*number contaminated	0.00894**	0.00157	0.0113**	0.00591	-0.00342	0.00960**
	(0.00410)	(0.00767)	(0.00481)	(0.00421)	(0.00851)	(0.00458)
Post*average As level (g/L)	0.247	-0.0464	0.311	0.0179	-0.317	0.0810
	(0.197)	(0.397)	(0.214)	(0.167)	(0.302)	(0.179)

Note: This table tests the robustness of the results in table 4 to different measures of arsenic exposure. Each number is from a separate regression, determined by the sample (columns) and measure of arsenic exposure (rows). The dependent variable is the number of months the child was breastfed in panel A and a dummy variable for whether the child is currently exclusively breastfed in panel C. The independent variable of interest is the interaction between various measures of arsenic contamination and an indicator for being born in 2002 or later. We include fixed effects for BDHS cluster and the child's year of birth, as well as the child's current age (or age at death) in months and a dummy for whether the child had died in panel A. Standard errors, clustered by BDHS cluster, are shown in parentheses. Significant at *10%, **5%, ***1%.

Appendix table A2. Testing robustness with respect to right-censored dependent variable

	(1)	(2)	(3)
Panel A: Right-censoring due to still breastfeeding			
Post*contamination	19.05***	7.83*	
	(6.46)	(4.13)	
Number of observations	5266	13609	
R-squared	0.30	0.59	
Panel B: Right-censoring due to child death while breastfeeding			
Post*contamination		5.02*	5.59**
		(2.94)	(2.37)
Number of observations		13610	13610
R-squared		0.44	0.54

Note: This table tests the robustness of the results in table 4 panel A to different assumptions regarding the right-censoring of the dependent variable, months breastfed. In column (1), we include only children who have already stopped breastfeeding. In column (2), we replace months breastfed with the maximum in the data for those who are still breastfeeding, in panel A, and for those who died, in panel B. In column (3), we replace months breastfed with the age the child would have been at the time of the survey (the greatest number of months breastfed that would have been possible had the child not died, given the timing of the survey). As in table 4, the independent variable of interest is the interaction between arsenic contamination and an indicator for being born in 2002 or later. We include fixed effects for BDHS cluster and the child's year of birth, as well as the child's current age (or age at death) in months (in panel A), the child's potential age at the time of the survey had she or he not died (in panel B), a dummy for whether the child had died, and district-specific linear trends. Only childen living in rural areas are included. Standard errors, clustered by BDHS cluster, are shown in parentheses. Significant at *10%, **5%, ***1%.

Appendix table A3. Testing robustness to specific breastfeeding durations

	(1)	(2)	(3)	(4)	(5)
Dependent Variable	Breastfeeding for longer than … months				
	6	12	18	24	36
Post*contamination	0.12***	0.11**	0.23***	0.29**	0.45**
	(0.04)	(0.05)	(0.09)	(0.13)	(0.19)
Number of observations	11498	10241	8788	7648	5061
R-squared	0.06	0.06	0.07	0.09	0.17
Mean dependent variable	0.97	0.95	0.89	0.78	0.36
Mean contamination	0.07	0.07	0.07	0.07	0.07

Note: This table shows the relationship between breastfeeding patterns and exposure to arsenic contaminated wells after the information campaign. The dependent variable is a dummy variable for whether the child was breastfed for at least 6, 12, 18, 24 or 36 months; we include in the sample only children who have attained the respective age by the time of the survey. As in table 4, the independent variable of interest is the interaction between arsenic contamination and an indicator for being born in 2002 or later. We include fixed effects for the nearest 2004 BDHS cluster, the child's year of birth and the survey year, as well as the child's current age (or age at death) in months, the arsenic exposure main effect and district-specific linear trends. The sample size drops considerably for longer lengths because survival status until age 36 months, for example, will not be known for children born within 3 years of the survey date. Only children living in rural areas are included. Standard errors, clustered by BDHS cluster, are shown in parentheses. Significant at *10%, **5%, ***1%.

Appendix table A4. Robustness checks regarding village trends

	(1)	(2)	(3)
	With cluster trends		
Areas included	All	Urban	Rural
Panel A: Dependent variable: Months breastfed			
Post*contamination	3.907	-3.766	7.083**
	(3.453)	(8.587)	(3.587)
Number of observations	19420	5811	13609
R-squared	0.661	0.616	0.679
Mean months breastfed	19.42	18.95	19.63
Mean contamination	0.0713	0.0698	0.0720

	(1)	(2)	(3)	(4)	(5)	(6)
				With cluster X birth year F.E.		
		Contamination level			Contamination level	
Areas included	All	Low	High	All	Low	High
Panel B: Dependent variable: Months breastfed						
Post*heard of arsenic	0.0643	-0.601	4.616***	0.375	0.0850	4.862**
	(0.575)	(0.601)	(1.407)	(0.865)	(0.900)	(2.399)
Number of observations	4756	3249	1507	4756	3249	1507
R-squared	0.681	0.686	0.680	0.755	0.761	0.741
Mean months breastfed	19.59	19.98	18.75	19.59	19.98	18.75
Mean contamination	0.0739	0.0110	0.210	0.0739	0.0110	0.210

Note: This table provides robustness checks for table 4, providing support for the identifying assumption that cluster-level trends are not driving the results. The regressions in panel A are similar to those in panel A of table 4, except that we include cluster-specific linear trends instead of district-specific linear trends. The regressions in panel B are similar to those in panel A of table 4 as well, except that the measure of arsenic contamination is whether or not the household responded that they had heard of arsenic. Since many households have heard of arsenic, even in relatively uncontaminated areas, we break up the sample into less and more contaminated regions in columns (2), (3), (5) and (6). Columns (4) to (6) also include cluster X birth year fixed effects, focusing entirely on intra-village variation in arsenic knowledge. The sample in panel B is restricted to rural households. Standard errors, clustered by BDHS cluster, are shown in parentheses. Significant at *10%, **5%, ***1%.

Appendix table A5. Heterogeneous responses to the information campaign

	(1)	(2)	(3)	(4)
Panel A: Dependent variable: Water sample is contaminated with arsenic or from surface source				
Contamination * access to clean well	-0.206***			-0.231***
* participation in any clubs	(0.0781)			(0.0761)
Contamination * access to clean well		-0.132*		-0.157**
* any education		(0.0701)		(0.0696)
Contamination * access to clean well			0.0444	0.0765
* decision-making			(0.0523)	(0.0518)
Number of observations	4776	4776	4771	4771
R-squared	0.565	0.563	0.563	0.567
Panel B: Dependent variable: Months breastfed				
Post*contamination	6.913**	0.920	0.411	-2.083
	(2.675)	(3.374)	(3.220)	(3.891)
Post*contamination	-5.520			-5.903
* participation in any clubs	(4.354)			(4.415)
Post*contamination		6.801*		5.995*
* any education		(3.501)		(3.445)
Post*contamination			7.966**	7.932**
* decision-making			(3.814)	(3.821)
Number of observations	4772	4772	4767	4767
R-squared	0.681	0.681	0.681	0.682

Note: This table explores heterogeneity among women in the two responses to arsenic contamination and the information campaign: using arsenic-safe wells for drinking water and breastfeeding longer. Panel A presents regressions similar to those in table 2 (column 4), except for the inclusion of triple interactions between arsenic contamination, access to clean wells and maternal characteristics. Please see the notes from table 2 for more details. The only other differences are that panel A also includes cluster fixed effects, which table 2 does not include, and the level of the observation is the child, not the household as in table 2. Panel B presents regressions similar to those in table 4 (panel A column 6), except for the inclusion of triple interactions between arsenic contamination, being born in 2002 or later and maternal characteristics. The only other difference is that we focus on data from 2004 in order to compare the results with panel A. Please see the text for more details on how the maternal characteristics are measured. Standard errors, clustered by BDHS cluster, are in parentheses. Significant at *10%, **5%, ***1%.

Appendix table A6. Relationship between BAMWSP and BGS contamination measures

BAMWSP measures of contamination:	(1) Percent of BAMWSP wells contaminated in 5km circle	(2) Percent of GPS locations in 5km circle contaminated (IDW)
BGS measures of contamination		
Probability of living within 1 mile of a contaminated well (weighted)	1.734***	2.125***
	(0.0795)	(0.0888)
Probability of living within 1 mile of a contaminated well (unweighted)	2.385***	2.937***
	(0.0805)	(0.0886)
Fraction of wells contaminated within 5 mi	0.648***	0.800***
	(0.0264)	(0.0278)
Number of wells contaminated within 5 mi	0.0832***	0.103***
	(0.00313)	(0.00349)
Average arsenic level of wells within 5 mi (g/L)	2.417***	2.982***
	(0.112)	(0.133)

Note: This table shows the relationship between the new BAMWSP measures of arsenic contamination and the original measures calculated from the BGS-tested wells. An observation is a cluster and each cell is from a separate regression of the BAMWSP measure on the BGS measure. Robust standard errors are in parentheses. Significant at *10%, **5%, ***1%.

Appendix table A7. Relationship between BAMWSP measures of arsenic contamination and BDHS household level variables

	(1)	(2)	(3)	(4)	(5)
	Heard of Arsenic	Level of arsenic in HH water source (µg/L)	HH water source contaminated	HH well painted red	HH well painted
Percent of BAMWSP wells in 5km circle that are contaminated	0.306***	99.47***	0.416***	0.566***	0.515***
	(0.0269)	(22.83)	(0.0617)	(0.0822)	(0.0587)
Percent of GPS locations in 5km circle that are contaminated (IDW)	0.243***	77.71***	0.332***	0.412***	0.367***
	(0.0189)	(18.16)	(0.0495)	(0.0634)	(0.0451)
Mean dependent variable	0.843	15.17	0.0813	0.170	0.315

Note: This table shows the relationship between the new BAMWSP measures of arsenic contamination and the household-level measures of arsenic contamination and information dissemination available in the 2004 wave of the BDHS. Each cell is from a separate regression of the household-level variable on the cluster-level measure of arsenic exposure. Standard errors, clustered by BDHS cluster, are in parentheses. Significant at *10%, **5%, ***1%.

Appendix table A8. The information campaign's effect on breastfeeding patterns, BAMWSP measures

	(1)	(2)	(3)	(4)	(5)	(6)
Measure of contamination	Percent of BAMWSP wells in 5km			Percent of GPS locations in 5km		
Areas included	All	Urban	Rural	All	Urban	Rural
Panel A: Dependent variable: months breastfed						
Post*contamination	1.717**	-0.823	2.039**	1.120*	-1.585	1.603**
	(0.819)	(1.905)	(0.897)	(0.653)	(1.501)	(0.701)
Number of observations	18823	5541	13282	19734	5851	13883
R-squared	0.619	0.573	0.641	0.619	0.571	0.641
Mean dependent variable	19.45	18.99	19.64	19.41	18.95	19.60
Mean contamination	0.184	0.129	0.207	0.198	0.146	0.220
Panel B: Dep Var: Breastfed for 12 or more months, among children 12 months and older						
Post*contamination	0.0153	-0.0887	0.0665***	0.0139	-0.0780	0.0504***
	(0.0228)	(0.0707)	(0.0230)	(0.0181)	(0.0556)	(0.0180)
Number of observations	14232	4246	9986	14924	4480	10444
R-squared	0.0562	0.0908	0.0620	0.0555	0.0890	0.0609
Mean dependent variable	0.939	0.913	0.950	0.939	0.912	0.951
Mean contamination	0.184	0.128	0.208	0.199	0.146	0.221
Panel C: Dependent variable: exclusively breastfeeding						
Post*contamination	0.0733*	0.00127	0.104**	0.0481	-0.0212	0.0767**
	(0.0425)	(0.0832)	(0.0478)	(0.0351)	(0.0678)	(0.0384)
Number of observations	9650	2742	6908	10078	2889	7189
R-squared	0.364	0.362	0.375	0.368	0.364	0.378
Mean dependent variable	0.111	0.0930	0.118	0.113	0.0935	0.120
Mean contamination	0.179	0.131	0.198	0.192	0.146	0.210

Note: This table shows the relationship between breastfeeding patterns and exposure to arsenic contaminated wells after the information campaign. The dependent variable is the number of months the child was breastfed in Panel A, an indicator variable for whether a child 12 months old or older was breastfed for at least 12 months in Panel B, and an indicator variable for whether the child is currently exclusively breastfed in Panel C. The independent variable of interest is the interaction between "post," defined as being born in 2002 or later, and the BAMWSP measure of arsenic contamination indicated by the column headers. We also include fixed effects for BDHS cluster in Panels A and C, for the nearest 2004 DHS cluster in Panel B, and the child's year of birth in all panels. Additional control variables are the child's current age (or age at death) in months and district-specific linear trends; a dummy for whether the child had died in Panel A; and survey year and the arsenic measure's main effect in Panel B. Columns (1) and (4) include all children, while the other columns divide the sample by urban or rural location. Standard errors, clustered by BDHS cluster, are shown in parentheses. Significant at *10%, **5%, ***1%.

Appendix table A9. Heterogeneous effects on exclusive breastfeeding by age, BAMSWP measures

	(1)	(2)	(3)	(4)	(5)	(6)
Measure of contamination	Percent of BAMWSP wells in 5km circle that are contaminated			Percent of GPS locations in 5km circle that are contaminated (IDW)		
Ages included	< 6 m	6 - 14 m	> 12 m	< 6 m	6 - 14 m	> 12 m
Panel A: Dependent variable: Breastfeeding exclusively						
Post*contamination	0.157	0.0173	0.0189	0.267*	0.0881	0.00413
	(0.174)	(0.0566)	(0.0247)	(0.144)	(0.0578)	(0.0196)
Number of observations	1320	1795	4245	1382	1875	4405
R-squared	0.392	0.249	0.109	0.386	0.260	0.105
Mean dependent variable	0.502	0.0496	0.0177	0.507	0.0528	0.0179
Mean contamination	0.225	0.202	0.191	0.239	0.216	0.202
Panel B: Dependent Variable: Had plain water in past 24 hours						
Post*contamination	-0.128	-0.191**	0.0391	-0.245*	-0.174**	0.0349
	(0.147)	(0.0793)	(0.0250)	(0.135)	(0.0718)	(0.0228)
Number of observations	1432	1850	7674	1501	1931	7988
R-squared	0.430	0.269	0.0771	0.428	0.273	0.0787
Mean dependent variable	0.411	0.925	0.962	0.409	0.921	0.962
Mean contamination	0.220	0.201	0.205	0.232	0.215	0.218

Note: This table shows the relationship between exclusive breastfeeding patterns or being given plain water at different ages and exposure to arsenic contaminated wells after the information campaign. The independent variable of interest is the interaction between "post," defined as being born in 2002 or later, and the BAMWSP measure of arsenic contamination indicated by the column headers. We include fixed effects for the nearest 2004 BDHS cluster in both panels. We also include fixed effects for the child's year of birth and survey year, as well as district-specific linear trends. Additional control variables are the child's current age in months and the arsenic exposure main effect. Panel B also includes a control variable for whether the child had died. The dependent variable in Panel A is a dummy for whether the child is exclusively breastfed and in Panel B is a dummy for whether the child has consumed plain water in the last 24 hours, but the sample is restricted to children under 6 months, between 6 and 14 months and more than 12 months, respectively. Only childen living in rural areas are included. Standard errors, clustered by BDHS cluster, are shown in parentheses. Significant at *10%, **5%, ***1%.

Appendix table A10. Health effects by age, BAMSWP measures

	(1)	(2)	(3)	(4)	(5)	(6)
Measure of contamination	Percent of BAMWSP wells in 5km circle that are contaminated			Percent of GPS locations in 5km circle that are contaminated (IDW)		
Panel A: Child died before the age of …						
Age in months	6	12	24	6	12	24
Post*contamination	-0.0567***	-0.0665***	-0.0549*	-0.0550***	-0.0623***	-0.0611***
	(0.0216)	(0.0243)	(0.0294)	(0.0176)	(0.0196)	(0.0226)
Number of observations	11942	10737	8117	12481	11222	8469
R-squared	0.0417	0.0465	0.0581	0.0394	0.0443	0.0551
Mean dependent variable	0.0564	0.0652	0.0748	0.0558	0.0647	0.0745
Mean contamination	0.205	0.206	0.208	0.218	0.219	0.219
Panel B: Health status of children						
Ages included	0 - 12 m	12 - 24 m	24 - 36 m	0 - 12 m	12 - 24 m	24 - 36 m
Incidence of diarrhea in previous two weeks						
Post*contamination	-0.171**	0.146	-0.0416	-0.179***	0.0723	-0.0540
	(0.0740)	(0.0980)	(0.0662)	(0.0631)	(0.0813)	(0.0504)
Weight for height Z-Score						
Post*contamination	0.895***	0.411	0.199	0.911***	0.437*	0.116
	(0.262)	(0.360)	(0.310)	(0.224)	(0.231)	(0.248)
Height for age Z-Score						
Post*contamination	0.562*	0.416	-0.0757	0.331	0.264	-0.0481
	(0.308)	(0.422)	(0.349)	(0.280)	(0.365)	(0.253)
Number of observations	2707	2494	2527	2829	2620	2620
Mean incidence of diarrhea	0.0842	0.121	0.0768	0.0848	0.122	0.0752
Mean weight-for-height Z-score	-0.542	-1.379	-1.119	-0.541	-1.379	-1.124
Mean height-for-age Z-score	-1.063	-1.990	-1.880	-1.065	-1.990	-1.889
Mean contamination	0.209	0.212	0.212	0.223	0.231	0.223

Note: This table shows the relationship between various health outcomes and exposure to arsenic contaminated wells after the information campaign. The dependent variable in Panel A is a dummy variable for whether the child died before the age of 6, 12, 24, or 36 months; we include in the sample only children who have attained the respective ages by the date of the survey. The dependent variables in Panel B are the incidence of diarrhea in the previous two weeks, weight for height or height for age. The independent variable of interest is the interaction between "post," defined as being born in 2002 or later, and the BAMWSP measure of arsenic contamination indicated by the column headers. We include fixed effects for the nearest 2004 BDHS cluster in both panels. We also include the arsenic exposure main effect, fixed effects for the child's year of birth and survey year, as well as district-specific linear trends. In Panel A, we also include the child's current age or potential age, had the child died. Additional control variables in Panel B are the child's current age or age at death in months and a dummy for whether the child had died. Only children living in rural areas are included. Standard errors, clustered by BDHS cluster, are shown in parentheses. Significant at *10%, **5%, ***1%.

Appendix table A11. Relationship between BAMWSP and BGS measures of access to clean wells

	(1)	(2)	(3)
BAMWSP measures of contamination:	Average "safe-depth" threshold (feet)	Average distance to a clean location (IDW) (km)	Average distance to a clean well (simulated) (km)
BGS measures of contamination			
Probability of living within 1 mile of an uncontaminated well	-347.3***	-3.459***	-0.358***
	(60.13)	(0.349)	(0.0448)
Probability of living within 1 mile of an uncontaminated well (weighted)	-204.9***	-2.264***	-0.246***
	(48.71)	(0.257)	(0.0336)
Average distance to closest uncontaminated well (miles)	13.28***	0.130***	0.0199***
	(4.603)	(0.0311)	(0.00444)
Average distance to closest uncontaminated well (weighted) (miles)	12.88***	0.128***	0.0194***
	(4.359)	(0.0298)	(0.00415)

Note: This table shows the relationship between the new BAMWSP measures of access to clean wells and the original measures calculated from the BGS-tested wells. An observation is a cluster and each cell is from a separate regression of the BAMWSP measure on the BGS measure. Robust standard errors are in parentheses. Significant at *10%, **5%, ***1%.

Appendix table A12: Predicting which households use arsenic-contaminated water or drink surface water, BGS and BAMSWP measures

	(1)	(2)	(3)	(4)	(5)
Measure of Access to Clean Well	Probability of living within 1 mile of an uncontaminated well	Average distance to closest uncontaminated well (miles)	Average "safe-depth" threshold (feet)	Average distance to a clean location (IDW) (km)	Average distance to a clean well (simulated) (km)
Source of Measure	BGS	BSG	BAMWSP	BAMWSP	BAMWSP
Panel A: Dep Var: Water sample is contaminated with arsenic or from surface source					
Probability of living within 1 mile of a contaminated well (weighted) (BGS)	-3.148***	0.408***	0.000109	0.193	0.413
*access to clean well	(1.159)	(0.114)	(0.000759)	(0.145)	(1.391)
R-squared	0.259	0.269	0.249	0.249	0.280
Percent of BAMWSP wells in 5km circle that are contaminated	-1.652***	0.133***	0.000388	0.188***	-0.206
*access to clean well	(0.455)	(0.0398)	(0.000353)	(0.0683)	(0.423)
R-squared	0.277	0.281	0.263	0.265	0.274
Percent of GPS locations in 5km circle that are contaminated (IDW)	-0.957**	0.0984***	0.0000268	0.297***	-0.128
*access to clean well	(0.387)	(0.0337)	(0.000314)	(0.0973)	(0.357)
R-squared	0.255	0.260	0.249	0.253	0.273
Panel B: Dep Var: Water sample is contaminated with arsenic					
Probability of living within 1 mile of a contaminated well (weighted) (BGS)	-3.890***	0.372***	0.000146	0.194	3.180**
*access to clean well	(0.999)	(0.111)	(0.000746)	(0.143)	(1.239)
R-squared	0.244	0.251	0.229	0.238	0.247
Percent of BAMWSP wells in 5km circle that are contaminated	-1.828***	0.106***	0.000793***	0.202***	0.713**
*access to clean well	(0.376)	(0.0408)	(0.000293)	(0.0640)	(0.334)
R-squared	0.255	0.251	0.249	0.247	0.238
Percent of GPS locations in 5km circle that are contaminated (IDW)	-1.115***	0.0737**	0.000317	0.302***	0.708**
*access to clean well	(0.341)	(0.0353)	(0.000282)	(0.0885)	(0.279)
R-squared	0.232	0.230	0.228	0.234	0.233

Note: This table shows the relationship between whether a household drinks arsenic-contaminated water or surface water from the 2004 BDHS (the dependent variable) and the cluster-level measures of arsenic contamination, access to clean wells and the interaction of these two variables. Only the interaction term is shown. All regressions include district fixed effects. Standard errors, clustered by BDHS cluster, are in parentheses. Significant at *10%, **5%, ***1%.

Appendix table A13. Differential responses with respect to access to clean wells, BAMSWP measures

Measure of Access to Clean Well	(1) Average "safe-depth" threshold (feet)	(2) Average distance to a clean location (IDW) (km)	(3) Average distance to a clean well (simulated) (km)
Measure of arsenic contamination			
Panel A: Probability of living within 1 mile of a contaminated well (weighted)			
Post*contamination	-0.728	3.998	2.141
	(4.329)	(3.259)	(3.077)
Post*contamination	0.0234**	0.00331	36.92**
*measure of access	(0.0112)	(0.00212)	(15.47)
Number of observations	13070	13609	13179
R-squared	0.641	0.641	0.641
Correlation coefficient	0.570	0.566	0.301
p-value of joint test	0.00560	0.0217	0.000158
Panel B: Percent of BAMWSP wells in 5km circle that are contaminated			
Post*contamination	0.901	3.075*	2.501**
	(1.862)	(1.611)	(1.205)
Post*contamination	0.00411	-0.000308	0.509
*measure of access	(0.00441)	(0.000869)	(4.218)
Number of observations	13173	13282	13282
R-squared	0.641	0.641	0.641
Correlation coefficient	0.768	0.841	0.612
p-value of joint test	0.183	0.126	0.0751
Panel C: Percent of GPS locations in 5km circle that are contaminated (IDW)			
Post*contamination	0.798	4.379***	1.714*
	(1.604)	(1.385)	(0.883)
Post*contamination	0.00178	0.00289*	0.807
*measure of access	(0.00396)	(0.00151)	(3.665)
Number of observations	13204	13883	13313
R-squared	0.641	0.641	0.641
Correlation coefficient	0.744	0.908	0.534
p-value of joint test	0.412	0.00150	0.0998

Note: This table shows the relationship between breastfeeding patterns and exposure to arsenic contaminated wells after the information campaign differentially with respect to different measures of the household's use of and access to uncontaminated wells. The specification is identical to the one in Table AA3, except for the triple interaction and all relevant main effects and two-way interactions. The dependent variable is the number of months the child was breastfed. The independent variables of interest are the interaction between "post," defined as being born in 2002 or later, and the BAMWSP measure of arsenic contamination indicated by the panel titles, and the triple interaction between this variable and the BAMSWP measures of the distance to an uncontaminated well. We also include fixed effects for BDHS cluster, the child's year of birth, as well as district-specific linear trends. Additional control variables are the measure of the distance to an uncontaminated well, its interactions with post and with arsenic exposure, the child's current age (or age at death) in months, and a dummy for whether the child had died. Only children living in rural areas are included. Standard errors, clustered by BDHS cluster, are shown in parentheses. Significant at *10%, **5%, ***1%.