Eliciting and Utilizing Willingness-to-Pay: Evidence from Field Trials in Northern Ghana

James Berry, Greg Fischer and Raymond Guiteras*

January 2019

Abstract

Using the Becker-DeGroot-Marschak (BDM) mechanism, we estimate the willingness-to-pay (WTP) for and impact of clean water technology through a field experiment in Ghana. Although WTP is low relative to the cost, demand is relatively inelastic at low prices. In the short-run, treatment effects are positive – the incidence of children's diarrhea falls by one third – and consistent throughout the WTP distribution. After a year, use has fallen, particularly for those with relatively low valuations. Strikingly, the long-run average treatment effect is negative for those with valuations below the median. Combining estimated treatment effects with individual willingness-to pay measures implies households' valuations of health benefits are much smaller than those typically used by policymakers. Finally, we explore differences between BDM and take-it-or-leave-it valuations and make recommendations for effectively implementing BDM in the field.

JEL Classifications: C26, C93, D12, L11, L31, O12, Q51

Keywords: price mechanism, heterogeneous treatment effects, health behavior, Becker-DeGroot-Marschak, field experiments

---

*Berry: University of Delaware, jimberry@udel.edu; Fischer: London School of Economics, g.fischer@lse.ac.uk; Guiteras: North Carolina State University, rpguiter@ncsu.edu. We thank Alex Nisichenko and Abdul-Salaam Sayuti for superb work in coordinating the field activities in Ghana, Kelly Bidwell and Jessica Kiessel at Innovations for Poverty Action, and Mary-Kay Jackson and Susan Murcott at Pure Home Water. We are grateful for helpful comments from the editor, four anonymous referees, Douglas Almond, Nava Ashraf, Abhijit Banerjee, Gharad Bryan, Sylvain Chassang, Vivian Hoffman, B. Kelsey Jack, Leigh Linden, Dean Karlan, Sheila Olmstead, Gerard Padró i Miquel, Giordano Palloni, Jesse Shapiro, Erik Snowberg, Chris Udry, Sergio Urzúa, Roger von Haefen, Ed Vytlacil and seminar participants at Colgate University, Columbia University, Duke University, HECER, the Inter-American Development Bank, the London School of Economics, Queens University Belfast, Resources for the Future, Stanford University, Triangle Resource and Environmental Economics Seminar, Universitat Pompeu Fabra, the University of California Santa Barbara, the University of Virginia, and Wesleyan University. S. Aqil Esmail, Kate Loose, Diego Rojas and Ben Zou provided excellent research assistance. This research was supported by the Improving Institutions for Pro-poor Growth research consortium, the International Growth Centre and the Maryland Population Research Center.

1 Introduction

Unsafe drinking water is a significant threat to health and welfare in the developing world. Approximately 30 percent of the world's population lacks access to safe water, and diarrheal disease kills nearly 1.4 million people per year, including over 500,000 children under age five. The problem is especially acute in sub-Saharan Africa, where diarrheal disease causes nearly 10 percent of deaths of children under age five, and 41 percent of the rural population drinks water from unimproved sources (WHO 2016; WHO and UNICEF 2017). Rural infrastructure improvements, such as bore wells or spring protection, suffer from poor governance, frequent outages, and recontamination of water between collection and consumption (Wright et al. 2004; Kremer et al. 2011), leading to interest in household water treatment as a potentially attractive alternative. Simple, relatively inexpensive technologies are known to be micro-biologically effective and have reduced diarrhea in controlled field trials (Clasen et al. 2015).

Despite these potential benefits, demand for household water treatment is typically low (Ahuja et al. 2010). This is an example of a general puzzle in development economics: households appear to underinvest in seemingly beneficial technologies across many domains (Foster and Rosenzweig 2010; Jack 2011; Dupas and Miguel 2017). When demand is low, measuring willingness-to-pay (WTP) provides a key input for pricing policy, guiding the magnitude and targeting of subsidies. Furthermore, understanding the relationship between WTP and a product's benefits is critical for distinguishing when the price mechanism allocates goods where their benefits are greatest and when it simply reduces access. In addition, combining measures of WTP with estimated treatment effects can yield insights into how households value health.

We study the demand for and impact of a household water filter in a field experiment with 1,265 households in rural northern Ghana. The filter requires effort to use, but if used properly produces safe drinking water for the household. After normal marketing efforts, we made sales offers to households and distributed the filters to those who purchased it.

We conducted follow-up surveys one month and one year after the sale to measure filter use and health outcomes related to water quality.

In our study we used the Becker-DeGroot-Marschak mechanism (BDM, Becker et al. 1964) to elicit precise measures of WTP. In BDM, an individual states her bid for an item. Then a random price is drawn. If the random price is greater than her bid, she cannot purchase the product. If the random price is less than or equal to her bid, she purchases the product, but pays the random price draw rather than her stated bid. Because the subject's stated WTP affects only whether or not she purchases the item, not the price she pays, BDM is *incentive-compatible*: the subject's dominant strategy is to bid her true maximum WTP.[1] In contrast to take-it-or-leave-it (TIOLI) offers, which yield only a bound on WTP, BDM produces an exact measure. In addition, BDM induces random variation in both treatment status and price paid, conditional on WTP. This allows researchers to separately identify screening and sunk cost effects.[2] Embedded in a field experiment, BDM can extract richer information than is typically available, but with the potential cost of added complexity. To assess the performance of BDM in a field setting, we randomly allocated half the households to a BDM sales treatment and half to a more traditional sales treatment using a TIOLI offer at a random price.

This study makes five contributions. First, we measure demand for clean water technology in a population facing a stark decision: how much of their scarce resources should they allocate to improving poor water quality? Demand estimates can provide important information on welfare and policy priorities, but measuring demand in developing countries is difficult because revealed-preference tools such as hedonic valuation or com-

[1]Deviations from expected-utility maximization may lead a subject's optimal bid to deviate from her true maximum WTP (Horowitz 2006), which we discuss in Section 6 below.

[2]Screening and sunk-cost effects typically cannot be separately identified, either in observational data or through TIOLI offers. Karlan and Zinman (2009) and Ashraf et al. (2010), among others, use a second-stage randomized discount to identify the causal effect of price paid. However, for the first-stage offer price to be incentive-compatible, subjects cannot anticipate the possibility of a second-stage discount. This is not feasible in many contexts, including ours, since information spread quickly within villages. By contrast, BDM allows a researcher to identify screening and sunk-cost effects in a single stage, without a surprise discount. In this paper, we focus on screening effects because we find no evidence of sunk-cost effects (see Appendix E).

pensating differentials rely on strong assumptions of complete markets (Greenstone and Jack 2015). This paper adds to a small but growing literature measuring demand for health goods directly through sales to households.[3] Similar to previous research for other preventative health products, demand is low. Median WTP is only 10 to 15 percent of the manufacturing cost, and demand is close to zero at a break-even price. However, almost all households have positive WTP, and demand is relatively inelastic at low prices.

Second, we use exogenous variation in filter allocation provided by our sales exercise to estimate the causal effect of receiving the filter on child health. In the short run, the filter reduces the probability that a child aged five or under has a case of diarrhea in the previous two weeks by about 7 percentage points, relative to the baseline rate of 21 percent. However, these benefits do not persist. The average treatment effect of the filter at our one-year follow-up visit is negative: diarrhea *increased*.

Third, we shed light on this surprising finding by estimating the distribution of treatment effects with respect to WTP. The importance of estimating distributions of treatment effects to policy analysis and uncovering structural parameters has been emphasized in the marginal treatment effects (MTEs) literature (Heckman and Vytlacil 2007), but estimating MTEs typically requires strong structural assumptions or multiple or multi-valued instruments. In contrast, by jointly eliciting WTP and generating exogenous variation in treatment conditional on WTP, BDM allows us to estimate the distribution of MTEs with respect to WTP in a simple and transparent way.[4] We find that after one year, the benefit of the filter is increasing in WTP, and the negative effect occurs in households with below-median WTP. The pattern of filter use resembles the pattern of treatment effects: households with low WTP were less likely to be using the filter after one year, suggesting that household effort, in particular proper maintenance and use of the filter, is an

[3]See, for example, Ashraf et al. (2010), Cohen and Dupas (2010) and Guiteras et al. (2015). Ito and Zhang (2016) provide an alternative approach using observational data, carefully isolating the price premium for goods with varying environmental benefits.

[4]The ability of BDM to improve information extraction from randomized control trials is emphasized by Chassang et al. (2012), who describe BDM as a type of a "selective trial."

important mediator of benefits. These findings have two important implications. First, in this sample, charging a positive price would allocate the filter to households where it is beneficial. Second, it underscores the importance of household behavior. Even technologically sound health products may not achieve their potential without appropriate household inputs (Brown and Clasen 2012; Hanna et al. 2016).

Fourth, because we have precise revealed-preference WTP data as well as WTP-specific impacts, we can estimate the distribution of demand for health. This contributes to the limited set of revealed-preference estimates for the value of health in low-income countries (Greenstone and Jack 2015). Using our short-run estimates, median WTP to avert one episode of children's diarrhea is USD 1.12. With additional assumptions, this implies a median WTP of USD 3,604 to avoid one statistical child death or USD 40 to avoid the loss of one disability-adjusted life year, well below standard cost-effectiveness thresholds.

Fifth, by randomizing households to either BDM or TIOLI we can compare the two mechanisms. Although BDM has the potential to enhance the information gained from field experiments, little is known about its performance in the field. BDM has been extensively used in laboratory settings, but anomalous behavior among subjects has been observed, such as sensitivity to the distribution of draws (Bohm et al. 1997; Mazar et al. 2014) or misunderstanding of the dominant strategy (Cason and Plott 2014). It is therefore an open question whether BDM's potential advantages outweigh its potential drawbacks. We present what is, to our knowledge, the first direct comparison of BDM and TIOLI in a developing-country field setting.[5] Results from both methods of demand elicitation follow a similar pattern and imply similar price elasticities. Furthermore, the cross-validated, predictive power of BDM estimates for TIOLI behavior is comparable to that of TIOLI itself. However, TIOLI acceptance rates are above the BDM demand curve. We explore a number of potential explanations and find that risk aversion accounts for much of the gap.

[5]Section 6 summarizes the theoretical and experimental literature studying behavior under BDM.

The paper proceeds as follows. Section 2 describes the setting and data. Section 3 describes demand for the filter. Section 4 presents the health impacts of the filter and heterogeneous treatment effects by WTP. Section 5 discusses policy counterfactuals and WTP for children's health. Section 6 compares the BDM and TIOLI mechanisms and discusses implications for future research using BDM. The final section concludes.

2 Experimental Setting and Design

We study the *Kosim* water filter (Figure A1), marketed in northern Ghana by Pure Home Water, an NGO. The filter consists of a clay pot treated with colloidal silver and a plastic storage container with a tap. The filter is micro-biologically effective, removing more than 99 percent of *E. coli* in field trials (Johnson 2007). This effectiveness is sustained with proper use: field tests one to three years after purchase found that well-maintained filters remove more than 95 percent of *E. coli* (Clopeck 2009). At the time of the study, the cost of production and delivery to a rural household in a village-level distribution was about GHS 21 (USD 15). We offered the filter to 1,265 respondents in 15 villages in Northern Ghana between October 2009 and June 2010. To select our sample, we identified villages that had limited access to clean drinking water and had not previously been exposed to the *Kosim* filter. Our subjects were women who were primary caregivers of children.[6] Figure A2 provides an illustrative timeline.

2.1 Data Collection and Experimental Design

2.1.1 Preliminary Activities & Household Survey

MARKETING MEETING. In each village, we held an initial village meeting. The NGO conducted its usual demonstration and marketing, and our field staff demonstrated the

[6] These were primarily mothers, but occasionally were others caring for children whose parents had migrated or were permanently absent for other reasons. We also included pregnant women and women who might become pregnant (married and of childbearing age).

sales mechanisms. During these demonstrations, field staff performed mock versions of BDM and TIOLI for a token item. The staff also practiced the sales mechanisms with volunteer attendees, again for a token item. We informed villagers that a filter would be installed at the village health worker's home and encouraged them see it in use, taste the water, and ask questions. We announced that we would visit households in two weeks to sell the filter and encouraged them to discuss with their families what they were willing to pay. The two-week interim period was to allow families time to try the filter, determine their WTP, and obtain necessary funds. On the same day as the marketing meeting, we conducted a village census to identify subjects.

REMINDER VISIT AND WATER QUALITY TESTING. One week later, we visited each household to remind them of the upcoming sale. In all households, we collected a 100 ml sample of drinking water. Budget constraints prevented testing all samples, so we tested levels of *E. coli* and turbidity in a randomly-selected half of the samples.

HOUSEHOLD SURVEY. One week after the reminder visit, we conducted the survey and sales visit. The survey included demographics, asset ownership, water collection and treatment practices, basic health knowledge, and recent episodes of diarrhea among household members. Subjects were compensated with GHS 1 cash, given in small coins so respondents could submit fine-scale bids in the practice rounds described below. There were always at least 30 minutes between the gift and the sale.

2.1.2 Filter Sale

At the end of the survey, we conducted the sale. Respondents were randomized in roughly equal proportions to either a BDM or TIOLI sales treatment.[7] Treatments were randomized at the compound level, stratified by number of respondents in the com-

[7] Within each broad category, we included three sub-treatments, described in Appendix J, to examine mechanisms underlying differences between BDM and TIOLI responses. However, demand was statistically indistinguishable by sub-treatment, and we group sub-treatments together for the primary analysis.

pound.[8] Each sale began with a practice round in which we offered the respondent the opportunity to purchase a bar of soap with retail value of GHS 1 using her assigned sales mechanism. After the practice round, we offered the *Kosim* filter using the same mechanism. If a sale resulted, the subject paid for the filter and received a receipt that could be redeemed for the filter at a central location in the village, typically the health liaison's home. To maintain realism – households routinely make small loans to each other for purchases – we permitted households to gather the money by the end of the day. If the respondent initially agreed to the purchase but was ultimately unable to obtain the funds, we code her as not purchasing. Our scripts are provided in Appendix A.

BDM TREATMENT. First, the surveyor read a brief description of the BDM procedure. We emphasized that the respondent would have only one chance to obtain the filter, could not change her bid after the draw, and must be able to pay that day. The surveyor then played a practice round for the bar of soap. The respondent was asked to bid her maximum WTP for the soap. The surveyor then asked the respondent if she would want to purchase the soap if she drew slightly more than her bid. The respondent was then allowed to adjust her bid. This process repeated until she was no longer willing to adjust her bid. Next, the surveyor reminded her that if she drew a price equal to her bid she must be willing and able to make this payment. At several points during the process, the surveyor reviewed various hypothetical outcomes to test the respondent's understanding. Once the final bid was established, the price was drawn and the subject either purchased or did not purchase the soap. The procedure for the filter was similar.[9]

We did not require respondents to present cash in the amount of their bid before the draw. However, before the draw, we asked multiple times whether the respondent would

[8]Most subjects live in extended patrilineal family compounds, small clusters of individual huts, usually enclosed by a wall. Many resources are shared within the compound, although in most cases each mother is responsible for providing water for her husband and children.

[9]Prices were written on wooden beads and placed in an opaque cup. The subject drew the price herself. For soap, the prices were distributed uniformly from 0 to 100 in increments of 10 pesewas (GHS 0.10). For the filter, the distribution of prices was 0, 0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4, 5, 6, 7, 8, 9, 10, 11, 12 in equal proportions. In neither case did we inform respondents of the distribution.

have access to the necessary funds. Of the 272 respondents who drew a price less than or equal to their bid, 269 (98.9 percent) completed the purchase. For the three respondents who did not, their failure to purchase appears to have been due to an unexpected inability to gather funds, for example because a family member was unavailable.

TIOLI TREATMENT. The standard TIOLI treatment was a simple sales offer at a randomized price. We emphasized that there would be no bargaining. We first conducted a practice round for a bar of soap. We then offered the filter at one of three prices: GHS 2, 4, and 6, the approximate 25th, 50th and 75th percentiles of BDM bids in piloting.

2.1.3 Follow-up Surveys

We conducted follow-up surveys one month and one year after the sale.[10] BWe obtained caretaker reports on diarrhea over the previous two weeks among children aged five and under. Among households that purchased the filter, surveyors recorded objective indicators of its condition and use. In the one-year survey, we also measured risk aversion, ambiguity aversion, digit span, and other preferences and beliefs we hypothesized could be related to behavior under the two sales mechanisms. Appendix B provides details.

The one-month survey was conducted in all 15 villages. Due to funding constraints, we randomly selected eight villages for the one-year survey. We re-surveyed 87.1 percent of targeted households in the one-month follow-up and 90.5 percent in the one-year follow-up. Attrition is largely balanced along observable dimensions. Most importantly, attrition is not related to the BDM draw or to the TIOLI price. See Appendix C for details.

[10]With good maintenance practices, in particular regular cleaning of the ceramic element, the filter's useful life is expected to be two years. We chose the one-year horizon as half this expected life. In practice, 40-50 percent of filters were found to be undamaged and in use after one year.

2.2 Sample Characteristics and Balance

Table 1 displays summary statistics, with the full sample in Column 1.[11] Only 9 percent of respondents had ever attended school, and the average number of children aged 0 to 5 was 1.1 per respondent. On average, households had 0.24 episodes of diarrhea among children aged 0 to 5 in the previous two weeks. Only 19 percent of households had access to an improved water source year round. Only 11.5 percent of households regularly treat their water using an effective method such as boiling (10.9 percent) or a ceramic filter (0.6 percent), reflecting the lack of affordable water treatment options.[12]

Columns 2 and 3 display sample means by treatment (BDM or TIOLI), and Column 4 tests differences between the two. There are a few marginally significant differences: 0.13 fewer children aged 0 to 5 per household in the BDM treatment ($p < 0.1$), 0.17 more children aged 6 to 17 ($p < 0.1$), 0.07 fewer children aged 0 to 5 with diarrhea in the past two weeks ($p < 0.1$), and 0.55 fewer respondents in the compound ($p < 0.1$).

In Column 5, we check balance of the BDM draw by regressing the BDM draw on the same set of characteristics, as well as the BDM bid. Of the 13 variables in the regression, one is significant at the 0.1 level: a higher number of respondents in the compound is associated with a higher draw ($p < 0.01$). Column 6 regresses the TIOLI price on these characteristics. Higher prices were associated with more children aged 6 to 17 with diarrhea in the past two weeks ($p < 0.1$) and higher turbidity in stored water ($p < 0.01$).

[11] Due to budget constraints, water quality (*E. coli* and turbidity) was measured for only half of the sample. Since households were randomly selected for water quality testing, this explanatory variable data is, by design, missing completely at random (MCAR).

[12] At the time of the study, household chlorination products were not widely available in Northern Ghana, and even if they had been the highly turbid source water would have limited their effectiveness. In a subsequent survey of 12 similar villages in Northern Ghana, Lu (2012) also found no use of chlorine and low levels of use of other effective treatment methods.

3 Demand for Filters

This section describes the demand for water filters measured through sales to households. Here, our focus is on the pattern of demand estimated through either the BDM or TIOLI mechanisms. Section 6 compares the two mechanisms in detail.

Figure 1a shows the inverse demand curve generated across all 15 villages using data from all 608 BDM and 658 TIOLI subjects. For the BDM observations, we plot for each price p the share of subjects whose bid was greater than or equal to p. For the TIOLI subjects, we show the share who purchased at each of the three randomly-assigned price points, $P = 2, 4, 6$.

There are several features of this inverse demand curve worth noting. WTP is almost universally positive: across the full sample, 95 percent of respondents were willing to pay at least GHS 1.[13] However, WTP is low relative to the filter's cost: the median BDM bid of GHS 2.5 corresponds to approximately 10 to 15 percent of the cost of manufacturing and delivery. This result is consistent with the relatively low WTP for water treatment and other health goods found in previous work (Ahuja et al. 2010). Figure 1b displays the price elasticity of demand at prices from 0 to 10 GHS as calculated from the BDM-elicited WTP data, and, for TIOLI subjects, the arc price elasticity of demand from 0 to 2, 2 to 4, and 4 to 6. In both groups, demand at low prices is relatively inelastic. In fact, demand is price inelastic up to roughly the median of the WTP distribution. While the lack of a steep drop in demand above a price of zero is largely consistent with existing estimates of demand for health products, we observe less price sensitivity than what has been found in much of the prior literature (Dupas and Miguel 2017).

[13]"House money" effects could provide an explanation for high demand at small positive prices; individuals may be be less price sensitive when spending funds given to them as a participation fee. The sale of soap before the filter bid allows us to test for such effects. We find no relationship between participation fees remaining after soap purchase (computed as 1 minus the draw for soap among those who purchased soap) and the filter bid, conditional on WTP for soap.

4 Health Impacts and Heterogeneous Treatment Effects

This section presents estimates of the filter's impact on children's diarrhea. In Section 4.1, we present standard IV estimates using the random offer price as an instrument for TIOLI subjects and the random price draw as an instrument for BDM subjects. In Section 4.2, we introduce heterogeneous treatment effects (HTEs) and how we can use BDM to estimate HTEs, in particular the relationship between effects and WTP. In Section 4.3, we apply this method and uncover important heterogeneity: benefits and WTP are positively related in our one-year follow-up data. Section 4.4 shows a similar positive relationship between use and WTP, and Section 4.5 further investigates mechanisms.

4.1 Average Effects on Child Health

We begin with the basic treatment effects equation

$$y_{jic} = \beta_0 + \beta_1 T_{ic} + \varepsilon_{jic}, \tag{1}$$

where y_{jic} indicates whether child j of subject i in compound c has had one or more cases of diarrhea in the previous two weeks, T_{ic} is dummy variable indicating whether subject i purchased the filter, and ε_{jic} captures unobservable determinants of y. The coefficient of interest is β_1, the effect of purchasing a filter on children's diarrhea.

To instrument for the treatment variable, we estimate the first-stage equation

$$T_{ic} = \gamma_0 + \gamma_1 P_{ic} + v_{ic}, \tag{2}$$

where P_{ic} is the TIOLI offer price for TIOLI subjects and the BDM draw for BDM subjects. Since P_{ic} is random, it is uncorrelated with ε_{jic} and therefore is a valid instrument for treatment. Table 2 presents the linear probability model estimates of the first stage. Price strongly predicts treatment, with a 1 GHS reduction in price leading to a 9.3 to 18.4

percentage point increase in the probability of treatment.

Panel A of Table 2 presents linear 2SLS estimates from our one-month data for the pooled, TIOLI, and BDM samples. Using the pooled data, the likelihood of diarrhea in the two weeks before the survey is reduced by about one-third, comparable to other trials (Ahuja et al. 2010). The estimates for TIOLI and BDM subjects are similar – TIOLI point estimates are slightly higher, but not statistically different. In Panel B of Table 2 we examine our long-term data, collected in a random sub-sample of half our villages. After one year, there is no evidence of benefits. The point estimates are positive, *i.e.*, the filter appears to have increased the likelihood of diarrhea. The effect is only statistically significant with controls, but the point estimates are consistently positive across specifications.[14]

4.2 Heterogeneous Treatment Effects: Theory

The standard IV approach of the previous subsection estimates a single average treatment effect. As discussed by Heckman and Urzúa (2010), this may not be the parameter of interest. In our setting, understanding the relationship between benefits and WTP is critical for pricing policy. It may be that those most likely to benefit are aware of this and have the resources to pay, in which case charging for the product improves targeting. Alternatively, those likely to benefit may be unaware of the extent to which they will benefit or simply too poor or credit constrained to purchase, in which case higher prices will restrict access without improved targeting (Cohen and Dupas 2010). Because BDM both elicits respondents' WTP and randomizes treatment conditional on WTP, it provides a simple way to estimate the relationship between benefits and WTP.

Consider the following econometric model,[15] adapted from Heckman et al. (2006),

[14] The above results assume a linear demand schedule in the first stage. As a robustness check, we estimated models with a more flexible demand specification: for TIOLI subjects, we use dummies for each of the three randomized prices (GHS 2, 4, 6); for BDM subjects, we use a quadratic in the random price draw. The results are similar, see Table A1.

[15] We provide a more complete treatment in Appendix D.

which generalizes (1) to allow β_1 to vary by WTP :

$$y = \beta_0 + \beta_1 (w) T + \varepsilon. \tag{3}$$

$\beta_1 (w)$ is the treatment effect for those with WTP $= w$, and WTP has distribution $F_{WTP}(w)$. Let $\overline{\beta}_1 = E_{F_{WTP}} [\beta_1 (w)]$ be the average effect in the population, and let $\tilde{\beta}_1 (w) = \beta_1 (w) - \overline{\beta}_1$ be the difference between $\beta_1 (w)$ and this average.

Now, consider the usual case where WTP is unobserved. The estimable model is

$$y = \beta_0 + \overline{\beta}_1 T + u, \tag{4}$$

with compound error term $u = \tilde{\beta}_1 (w) T + \varepsilon$. OLS estimation of (4) is biased if

$$E [Tu] = E \left[T \left(\tilde{\beta}_1 (w) T + \varepsilon\right)\right] \neq 0. \tag{5}$$

There are two potential sources of bias. The first is selection on levels, $E [T\varepsilon] \neq 0$, when treatment is correlated with unobservable determinants of y in the absence of treatment. The second is selection on gains: if WTP and benefits are correlated, then $E \left[T\tilde{\beta}_1 (w)\right] \neq 0$.

Selection on levels is traditionally addressed by an instrument: a source of variation in treatment uncorrelated with unobservables. One natural candidate is a randomized price, $Z \in \{P_L, P_H\}$, which for simplicity takes on two values, $P_L < P_H$. If demand is downward-sloping, then $\Pr (T|P_L) > \Pr (T|P_H)$, so the instrument is relevant. The instrument is valid if

$$E [Zu] = E [Z\varepsilon] + E \left[Z\tilde{\beta}_1 (w) T\right] = 0. \tag{6}$$

Since Z is random, $E [Z\varepsilon] = 0$, which solves the problem of selection on levels. However, the problem of selection on gains remains. Since $T = 1 \{\text{WTP} > Z\}$, if there is a relationship between WTP and benefits then $E \left[Z\tilde{\beta}_1 (w) T\right] \neq 0$.

Therefore, when there is selection on gains, IV using the offer price Z will not produce a consistent estimate of $\bar{\beta}_1$. By the LATE theorem of Imbens and Angrist (1994), IV does estimate the average effect on compliers: those whose treatment status is changed by the instrument. Here, this is the group with $P_L \leq \text{WTP} \leq P_H$, who would buy a filter at P_L but not at P_H. Formally, IV using Z estimates

$$\beta_1^{IV}\left(P_L \leq \text{WTP} \leq P_H\right) = \int_{P_L}^{P_H} \beta_1\left(w\right) dF_{WTP}\left(w\right),$$

the average $\beta_1\left(w\right)$ between P_L and P_H weighted by $F_{WTP}\left(\cdot\right)$. As argued in Heckman and Urzúa (2010), this may not be a useful parameter, since it only tells us the effect of changing price from P_H to P_L in a population with WTP distributed $F_{WTP}\left(\cdot\right)$.

BDM provides a simple method to estimate $\beta_1\left(w\right)$.[16] Intuitively, BDM provides a measure of WTP, then the BDM draw randomizes treatment conditional on this measure.[17] With a large enough sample, we could estimate the function $\beta_1\left(w\right)$ nonparametrically by comparing outcomes of winners and losers at each WTP. Our sample is not large enough to condition on exact WTP, so we compute kernel-weighted linear 2SLS estimates on a WTP grid.

4.3 Heterogeneous Treatment Effects: Application

The kernel IV approach reveals substantial heterogeneity with respect to WTP. The outcome variable, as above, is an indicator for whether the child has had one or more cases

[16] Remarkably, the *local instrumental variables* (LIV) method of Heckman et al. (2006) can estimate $\beta_1\left(w\right)$ without observing WTP. LIV estimates a propensity score in a first step, then regresses the outcome on the propensity score. The BDM approach has the advantage of observing WTP directly, rather than inferring it through a first-step selection model. This increases power – in our application, confidence intervals are 40% narrower on average. (See comparison in Appendix D.2.) LIV also allows non-price instruments, although continuous, many-valued, or multiple instruments are typically required to estimate the propensity score flexibly. Furthermore, interpretation is more subtle with non-price instruments since heterogeneity is estimated with respect to unobservables. See Appendix D.1.

[17] Using the BDM draw as an instrument requires an exclusion restriction: the draw cannot directly affect the outcome. This is violated if there are wealth effects, since the draw determines the price paid. This is a common problem in IV estimation (Jones 2015), and applies equally to random TIOLI prices. Similarly, a causal effect of price paid on use may violate the exclusion restriction. We do not observe a causal effect of price paid on use (Appendix E).

of diarrhea in the previous two weeks. We estimate kernel-weighted treatment effects $\hat{\beta}_1(w)$ for each GHS 0.1 step from GHS 1 to GHS 6, which correspond approximately to the 0.1 and 0.9 quantiles of WTP in the BDM sample.[18]

Figure 2 presents the kernel first stage: the point estimate on the draw Z_i (left axis) and the associated F-statistic (right axis), at each grid point $w = \{1, 1.1, \ldots, 6\}$. Results from the one-month follow-up are in the top panel (a), with results from the one-year follow-up in the bottom panel (b). The random price draw is negatively associated with treatment, as expected, and is a statistically strong instrument. These regressions are unconditional; results with controls are similar.

Kernel IV estimates of the outcome equation are presented in Figure 3. We reverse the sign of $\hat{\beta}(w)$, so benefits are positive. In the top panel (Fig. 3a), we consider the effect at one month. The point estimates are positive, although not statistically significant at any level of WTP, and there is little heterogeneity. In the bottom panel (Fig. 3b), we use the one-year data, and observe important heterogeneity: the perverse negative effect occurs among those with below-median WTP. The estimated benefit increases with WTP, becoming positive at roughly GHS 3 and peaking at roughly GHS 4.5. Above GHS 4.5, point estimates decrease, although confidence intervals are wide. We discuss this finding at length in Sections 4.5 and 5, but here we emphasize that this implies price is an effective screening mechanism for this product in this context. In fact, charging 3 GHS would not just improve targeting, but would actually prevent harm in the medium term.

While the flexibility of the kernel IV and the sample size limit the precision of our estimates, we can reject that the one-year treatment effects at WTP $= 4$ and WTP $= 2$ are equal (estimated difference $\hat{\beta}(4) - \hat{\beta}(2) = 0.450$, std. err. 0.141, $p = 0.001$). If we assume that the treatment effect is linear in WTP, the slope term is statistically significant (point estimate 0.170, std. err. 0.076, $p = 0.024$).

[18] Non-parametric estimators are prone to bias at boundaries. Restricting to the 0.1 and 0.9 quantiles of WTP reduces this risk. Furthermore, our estimator is analogous to a local linear regression rather than a local constant regression, and local linear regressions are less subject to boundary bias (Li and Racine 2007).

4.4 WTP and Use

In this section we analyze use of the filter in our one-month and one-year data. The potential health gains of the filter may not be achieved if it is not used properly or cleaned regularly. Variation in use over time and across individuals with different levels of WTP could produce the patterns of impacts observed in the previous section.

We collected three objective indicators of use from all subjects who purchased the filter: (i) whether the filter was found in the compound and was undamaged; (ii) whether water was in the plastic storage reservoir above the level of the tap (an indicator of whether filtered water was immediately available to drink); and (iii) whether water was in the clay filter pot. To aggregate the three measures in an agnostic way, we create an index by normalizing each measure to have mean 0 and standard deviation 1 and taking their average (Kling, Liebman and Katz 2007).

For comparability with our analysis of heterogeneous treatment effects in Section 4.3, we restrict the sample to winning households with children aged 0 to 5 and model the relationship between WTP and use nonparametrically using kernel regression. Figure 4 displays these results for both the aggregate use index and, for ease of interpretation, the indicator of whether filtered water was immediately available to drink.[19]

In the short term, use is generally high. The filter is present and operational in nearly 90 percent of households that purchased, and filtered water is available to drink in more than 75 percent. As shown in Figure 4, there is little heterogeneity with respect to WTP. In contrast, use has fallen substantially in the one-year follow-up. Filtered water is immediately available in fewer than half of households. Although the confidence intervals are wide, the kernel estimates now reveal substantial variation in use with respect to WTP. The conditional mean of filtered water being available ranges from 35 percent in households with a WTP of GHS 2 to 59 percent in households with a WTP of GHS 4 ($p = 0.036$). The usage index follows a similar pattern, with a difference of 0.29 standard deviations

[19] See Table A2 for linear regressions and additional outcomes, with similar results.

between those with a valuation of GHS 2 and those with a valuation of GHS 4 ($p = 0.096$).

The similarity between the patterns of use and of benefits is consistent with effort as an important mediator of treatment effects. In the short-term, effort is uniformly fairly high and there is evidence of benefits for most of the population. In the longer-term, effort and benefits have both fallen overall, and benefits are greatest in the population that is exerting the most effort.

4.5 Understanding the Pattern of Treatment Effects

In this section, we explore mechanisms behind the detrimental long-run impacts of the filter observed in the lower half of the WTP distribution, with a formal model and additional discussion in Appendix H. The possibility that compensatory responses to health interventions could offset the intended effects has been studied extensively in other contexts (e.g., Peltzman 1975; Lakdawalla et al. 2006). In settings closer to ours, Bennett (2012) finds that the introduction of piped water in the Philippines led to decreased private investment in sanitation, reversing the gains from cleaner water, and Gross et al. (2017) show that improved water sources in Benin led to decreases in point-of-use water quality, likely through changes in water handling practices. If households perceive the filter and other health behaviors as substitutes, receiving the filter will reduce other health investment, assuming that households maximize overall utility, not child health.

While standard models of compensatory behavior could explain a muted benefit, they would not generate negative effects. However, in our context, certain aspects of the utility or production function may have combined with compensatory behavior to generate detrimental effects. In Appendix H, we consider three particular mechanisms that may have been operating. First, households may have failed to adjust other health behaviors in response to a decrease in the filter's effectiveness over time. Second, due to limited supply, adults may have restricted children's access to filtered water. Third, non-convexities in health production technology may have led to increases in utility but decreases in health.

With the caveat that use is endogenous, the pattern of outcomes and use over time is consistent with compensatory behavior playing a role in the negative treatment effects. Diarrhea rates are higher for children in households that purchased the filter but are no longer using it after a year than in those that never purchased (0.34 vs. 0.24; $p = 0.069$). This difference is driven by households that were using the filter after one month, and hence might rationally engage in compensatory behavior. Among this group, the incidence of children's diarrhea increases to 0.37 relative to 0.24 for those who never purchased ($p = 0.036$). Those who purchased but were not using the filter after one month – and hence were unlikely to engage in compensatory behavior – report outcomes similar to those who never purchased (0.22). Figure A7 displays these results.

These results add to a nascent literature in economics exploring the role of subjects' behavior as a moderator of treatment effects (e.g., Chassang et al. 2012; Hanna et al. 2016). Our analysis also highlights the challenges in studying these mechanisms, which are less amenable to experimental variation than assignment of a program or product, such as the filter.

5 Policy Counterfactuals and Valuing Health

In this section, we explore policy implications of the treatment effects estimated above. First, we analyze the effects of different counterfactual prices to inform optimal pricing policy. In the short run, prices merely reduce access. In the longer term, prices screen out those with the lowest benefits and improve allocative efficiency. Second, we estimate households' valuation of the filter's health benefits by combining our treatment effects estimates with our WTP data. We find low valuations compared with those typically assumed by policy makers.

5.1 Policy Counterfactuals

In this section, we show how the distribution of treatment effects estimated above can be used to simulate impacts of different pricing policies. We consider a social planner who values disability-adjusted life years (DALYs) at B. The planner's choice variable is the sales price P. The social planner places equal weight on subsidy and private expenditure: P is of interest only for its effect on allocation, not for revenue.

Under these assumptions, the social planner will lower the price P as long as the marginal cost per DALY is less than B. If the benefits of the filter are constant at all prices, the marginal cost per DALY will be constant. The filter will be fully subsidized if the marginal cost per DALY is below B, or not distributed at all if the marginal cost per DALY is above B. On the other hand, if the benefits of the filter are increasing in price, the social planner will set the price such that the marginal cost per DALY equals B. At this point, decreasing the price will include households whose benefits cost more than B, and increasing the price will screen out households for whom the benefits cost are less than B.

We consider two scenarios. First, we assume that the health gains from the one-month survey persist for a full year. While in practice the average treatment effects diminished over time, this provides a bound on the health gains if use patterns could be maintained over the life of the filter. Since there is little evidence of heterogeneity in the short term, we assume these effects are constant with respect to WTP. As we describe in more detail in Appendix F, the constant treatment effects imply that the marginal costs per DALY gained are constant and equal to USD 369. A social planner valuing DALYs above USD 369 would maximize gains by distributing the filter for free. This value falls below cost-effectiveness thresholds typically used by policy makers. Although precise thresholds are subject to debate, the 1993 World Development Report presents interventions costing less than USD 150 per DALY as cost-effective (World Bank 1993), and this figure has been cited in a number of subsequent cost-effectiveness analyses (Shillcutt et al. 2009).[20]

[20] An alternative—and considerably higher—threshold, used by the WHO-CHOICE project, is one to

In our second scenario, we assume health effects initially equal our short-term estimates and then evolve smoothly over 12 months to the long-term estimates. We again assume the short-term effects are constant with respect to WTP and impose a linear functional form on the one-year effects. Because the benefits are now increasing in price, the marginal and costs per DALY gained are decreasing in price. As we show in Appendix F, a policymaker with a value per DALY of at least USD 361 would optimally sell the filter at a price of GHS 4. A lower price would reduce total benefits, and a higher price would reduce coverage among those whose benefits cost less than USD 361 per DALY.

5.2 Valuing Children's Health

By combining our WTP data with our estimates of the impact of the filter on child health, we can directly estimate households' valuation of children's health. There are few well-identified revealed-preference estimates of this parameter, or of WTP for health or environmental quality more generally, in spite of its importance for optimal policy (Greenstone and Jack 2015). A notable exception is Kremer et al. (2011), in which the authors randomize water quality improvements at springs in Western Kenya and observe how much additional time households travel to collect better quality water. They then use wage data to convert this implicit valuation in terms of time to monetary valuation. Using this travel cost model, estimated mean WTP to avoid a case of children's diarrhea equals USD 0.89, which, with additional assumptions, translates to a value of a DALY of USD 23.7 and a value of a statistical life (VSL) of USD 754. A key advantage of our approach is that we observe WTP directly, rather than inferring it through travel time and an assumed value of time. We can simply calculate the household's observed WTP to avoid a case of diarrhea as the household's WTP for the filter divided by the number of cases avoided over the anticipated life of the filter.

three times annual per capita PPP GDP, or USD 2,997 to 8,991 for Ghana at the time of our study (Hutubessy et al. 2003).

While this quantity is simple to calculate in our setting, interpreting it as the household's underlying value of child health requires several assumptions. First, households know the effect of the filter on children's health. Second, households only value the filter's effect on children's health. That is, the household's WTP does not reflect other potential benefits of the filter, such as improved taste or prestige. Third, households only value reductions in diarrhea for children aged five and below. This assumption is made because diarrhea has more severe health consequences for young children, but it is also made due to data limitations: our pilot surveys indicated respondents were unable to accurately report diarrhea cases among older children or adults. Fourth, households are not liquidity constrained. Fifth, using the filter entails no change in convenience or time costs relative to current practices. We return to these assumptions at the end of this section.

We estimate households' WTP to avoid a case of diarrhea under two scenarios, making the same assumptions on treatment effects as in Section 5.1 above.[21] In the first scenario, we use the estimated impact from the one-month follow-up survey to project benefits over a year. This corresponds to the household believing that its own short-run use and maintenance practices as well as the filter's impact will persist over the first year. Again, we restrict the treatment effect to be constant with respect to WTP since there is little evidence of heterogeneous treatment effects in the short run. Figure 5a plots the distribution of WTP to avoid a case of children's diarrhea. The resulting median WTP is GHS 1.58, or USD 1.12. If we assume deaths from diarrhea are proportional to incidence and that households value only the reduction in mortality risk, not the reduced morbidity, we can compute the value of a statistical life using a ratio of mortality to incidence of one death per 3,216 cases of diarrhea in children under five, estimated for Ghana in 2010 (Global Burden of Disease Collaborative Network, 2017). The resulting median VSL is GHS 5,081 (USD 3,604). Again assuming that the reduction in DALYs is proportional to the reduction in incidence, we can apply a ratio of one DALY for each 35.3 cases of children's diarrhea

[21] See Appendix G for details on these calculations.

(Global Burden of Disease Collaborative Network, 2017) to calculate a median value of a DALY of GHS 55.77 (USD 39.56). Similar to the findings of Kremer et al. (2011), this is well below the typical cost effectiveness thresholds described in the previous subsection.

In the second scenario, we use both the short-term and one-year effects and compute the total effect of the filter over the first year as if the effect changed smoothly over the course of the year. We again assume the short-term effects are constant and impose a linear functional form with respect to the WTP on the one-year effects. Figure 5b plots the distribution of these estimates. The most striking feature of the graph is the large share of households with negative WTP to avoid children's diarrhea: the median WTP is GHS -0.20 (USD -0.14). Mechanically, this occurs because the average of the one-month and one-year treatment effects are negative for just over half of the population even though they exhibit positive WTP.

It is unlikely that households have a negative WTP for children's health. We posit two key explanations for this result related to Section 4.5's discussion of compensatory behavior. First, households may have misperceived the benefits of maintaining the filter or using it regularly. Improper use or a failure to re-optimize compensatory behaviors over time could produce negative long-run treatment effects. If a household failed to foresee these actions, it might pay a positive amount for these negative treatment effects even if it valued health, and we would estimate a negative value for health. Second, as in Kremer et al. (2011), the calculations above are based on the assumption that the filter produces a single good: children's health. In fact, the filter produces multiple goods, for example, adults' health and better tasting water, that may also be valued by the household. A household's total WTP for the filter is the sum of its value for all of these goods. As discussed further in Appendix H, this bundling can explain why a household might rationally be willing to pay for the filter despite a negative impact on children's health.

While our empirical setting does not allow us to precisely identify the individual components of a household's valuation for the filter, by simply comparing valuations from households with and without children under age five we estimate that the valuation of the other goods produced by the filter could represent as much as 83 percent of total WTP.[22] Incorporating this information in our estimates of the WTP to avoid a case of children's diarrhea would eliminate many of the negative valuations implied by the longer-term impacts.[23] These households may be willing to accept a reduction in children's health in exchange for the bundle of goods the filter provides. This highlights both the challenge and importance of constructing accurate WTP measures for health and environmental goods in developing countries.

6 Comparing Mechanisms

In addition to using BDM to conduct analyses of demand for the filter and its benefits, we designed our study to compare demand elicited under BDM and TIOLI. While BDM produces more precise information than TIOLI offers at randomized prices, this benefit may be mitigated by its complexity. Furthermore, although bidding one's true maximum WTP is the dominant BDM strategy for expected utility maximizers, this does not necessarily hold for non-expected utility maximizers (Karni and Safra 1987; Horowitz 2006).

There is an extensive literature in experimental economics studying the behavior of BDM among subjects in laboratory settings. It raises several issues. Several papers find that BDM-elicited valuations can be sensitive to the distribution of prices (Bohm et al. 1997; Mazar et al. 2014). Cason and Plott (2014) show that subjects' misunderstanding of the best response can also influence the WTP elicited by BDM. In addition, several studies explicitly compare BDM with other incentive-compatible elicitation mechanisms and find differences in elicited WTP (Rutstram 1998; Shogren et al. 2001; Noussair et al. 2004).

[22] Average WTP for households with no children under 5 is GHS 2.67, while average WTP for households with children under 5 is 3.22. Other goods produced by the filter could include adult health, taste of the water, or prestige of owning the filter. We lack the data to examine these components directly.

[23] Assigning a value to other goods produced by the filter would also reduce the mean and median estimates based on our short-term treatment effects.

In spite of the large laboratory literature on BDM, little is known about its performance in field settings. We therefore designed our study to allow direct comparison of the demand estimates from BDM and TIOLI and to investigate the causes of any differences. Although both mechanisms are research tools and may not map directly to typical market interactions, TIOLI offers at randomized prices are common in applied research. They provide a useful benchmark for the signal contained in BDM offers. We present what is, to our knowledge, the first direct comparison of BDM and TIOLI in a developing-country field setting with the aim of better understanding the suitability of BDM for extracting additional information from field experiments.[24]

We organize the analysis comparing BDM and TIOLI as follows. Section 6.1 compares the demand estimates and out-of-sample predictive accuracy of both mechanisms. The BDM-based demand model has similar accuracy in predicting out-of-sample TIOLI decisions as the TIOLI model itself, indicating that the BDM bids contain substantial signal. As is common in the consumer behavior literature, there is substantial unobserved heterogeneity in demand estimates using either mechanism, which underscores the utility of measuring demand directly. Section 6.2 tests several potential explanations for the BDM-TIOLI demand gap. Our main finding is that the gap is largest among the most risk-averse subjects and negligible for the most risk tolerant.

6.1 Comparing Demand Estimates and Predictive Accuracy

This section compares the correlates of demand obtained using each mechanism as well as the accuracy of each mechanism for predicting out-of-sample purchase behavior. In addition to providing a point of comparison between mechanisms, understanding the relationship between household characteristics and WTP can be directly useful by informing how pricing policies target particular types of households. Previous studies have found

[24]Subsequent to our study, Cole et al. (2016) study demand for weather insurance and an agricultural information service in India using BDM and TIOLI. They find that BDM-measured demand is similar to that of TIOLI on average, although the relationship depends on the product offered.

limited evidence that WTP for health goods in low-income countries is related to health characteristics or wealth (Ashraf, Berry and Shapiro 2010; Cohen and Dupas 2010), reflecting a common finding in the consumer behavior literature: choice is often only weakly correlated with standard consumer attributes (Nevo 2011). This makes predicting individual purchase behavior difficult and underscores the usefulness of direct measurement of WTP.

We model the relationship between WTP and characteristics as

$$\text{WTP}_{ic} = \alpha_0 + X_{ic}^{'}\beta + \varepsilon_{ic}, \tag{7}$$

where X_{ic} is a vector of characteristics for subject i in compound c, and ε_{ic} is an error term.

In our BDM sample, we observe WTP directly and can estimate Equation (7) via ordinary-least-squares. Columns 1 and 2 of Table 3 present these results. The BDM bid is positively related to the number of children aged five and under with diarrhea, a result significant at the 10 percent level. One additional child with diarrhea in the household (conditional on the total number of children), is associated with an increase of GHS 0.55 in the BDM bid. The BDM bid is also positively related to durables ownership and education, although the latter is not significant. These relationships are consistent with hypotheses from the pricing literature. However, we note that, also consistent with that literature, the estimates are generally imprecise. Household characteristics explain very little of the variation in WTP. Moreover, as shown in Column 2, the best predictor of WTP for the filter is a household's WTP for soap, a related health product. When we control for a household's bid for soap in the BDM practice rounds, the share of variation explained by the model increases from 0.053 to 0.214.

For TIOLI subjects, WTP is an unobserved latent variable, so we estimate (7) indirectly using a discrete choice model:

$$\text{buy}_{i,p} = 1\left\{\text{WTP}_i \geq p_i\right\} = 1\left\{\text{WTP}_i - p_i \geq 0\right\} = 1\left\{\alpha_0 + X_{ic}^{'}\beta + \varepsilon_{ic} - p_i \geq 0\right\} \tag{8}$$

where $\text{buy}_{i,p}$ is an indicator equal to 1 if respondent i agreed to buy when assigned price p_i. We estimate (8) on TIOLI subjects by probit. In the estimation, we normalize the coefficient on price (in GHS) to -1, so the estimated coefficients β are interpreted in terms of GHS and are comparable to those obtained by estimating Equation (7) directly with BDM subjects. Columns 3 and 4 of Table 3 presents these results.[25]

When we compare the correlates of demand using each mechanism (Column 5 of Table 3), there are a few significant differences between the estimates for BDM and TIOLI. In several key cases, the BDM coefficient conforms more closely to hypothesized mechanisms from the literature and to our prior beliefs. For example, respondents that are more educated tend to express a higher WTP under BDM but are significantly less likely to accept a TIOLI offer at a given price. That said, and consistent with the aforementioned consumer behavior literature, much of the heterogeneity across subjects remains unexplained. For both mechanisms, a household's purchase decision for soap are more predictive of filter demand than the set of all other household characteristics combined. Because both the filter and soap are health products, this result suggests that a household's unobserved demand for health is a strong determinant of WTP for both goods. Appendix I.1 describes the results of applying LASSO regression to determine the most relevant attributes to predict filter demand. Here too the WTP for soap in the practice round is the dominant feature predicting filter demand.

An alternative method of evaluating BDM is to analyze the extent to which it can predict non-BDM purchase behavior. We therefore compare both mechanisms on their ability to predict out-of-sample TIOLI decisions. Appendix I.2 details the procedure and provides additional results. In summary, we split each of the BDM and TIOLI samples into 10 roughly equally-sized parts or folds. For each fold k in the TIOLI sample, we use the remaining $k-1$ folds in each of the BDM and TIOLI samples to predict purchase behavior in the k^{th}, holdout, fold. We then calculate prediction error for each model and

[25]Bivarate regressions of BDM bid or TIOLI purchase on each variable separately yield broadly similar results (not shown).

combine the estimates of the 10 folds. BDM and TIOLI correctly predict TIOLI behavior in the holdout samples correctly in 76.0 percent and 73.9 percent of observations, respectively, relative to a base rate of 56.2 percent. While additional work is required to link behavior under either mechanism to actual market purchase behavior, in this setting the predictive ability of BDM for TIOLI behavior is comparable to that of TIOLI itself.

6.2 Mechanism Effects

As shown in Figure 1a, demand is lower under BDM than TIOLI at each of the three TIOLI price points. This gap is 18.2 percentage points at a price of 2 GHS ($p = 0.000$), 16.3 percentage points at 4 GHS ($p = 0.002$), and 10.0 percentage points at 6 GHS ($p = 0.012$).[26] The adjustment to BDM bids that minimizes the differences in demand at the three TIOLI price points is approximately GHS 1. Under the assumption that TIOLI reflects true WTP, this implies a BDM "mechanism effect" of GHS 1. In this sub-section, we investigate potential explanations for this gap.

First, we examine the relationship between the BDM-TIOLI gap and risk aversion. Theory predicts no gap in elicited WTP between BDM and TIOLI when agents are expected utility (EU) maximizers. In our setting, there are multiple likely sources for deviations from EU maximization including loss aversion, ambiguity aversion, and non-standard beliefs about probability. Based on survey responses to questions on hypothetical gambles, 30.4% of our subjects exhibit loss aversion, 41.6% exhibit some degree of ambiguity aversion, and 64.6% at least one of these two. The theoretical literature on the BDM mechanism finds that, among non-EU maximizers, the optimal BDM bid can differ from the TIOLI reservation price, and this difference is likely to be increasing in risk aversion (Safra et al. 1990; Keller et al. 1993).

To test this hypothesis, in the one-year followup villages we collected standard survey measures of risk aversion using stated-preference responses to hypothetical gambles. (See

[26]See Appendix J.1 for full presentation of these results.

Appendix B for detail.) We then divide the sample into terciles by risk aversion and estimate the gap separately for each tercile. As predicted by theory, risk aversion appears to be an important determinant of the mechanism effect: the gap is largest BDM-TIOLI gap is largest among the most risk-averse subjects (mean BDM effect -0.200, $p = 0.000$) and has largely closed among the least risk-averse subjects (mean BDM effect -0.051, $p = 0.425$). See Appendix J.2 for details on these tests and robustness checks.

Second, we examine how BDM-TIOLI gap differs with respect to other household observables, with the caveat that this is ex post hypothesizing rather than guided by theory. Here, we highlight the most interesting findings; we present the methods and full set of results in Appendix J.3. The mean BDM-TIOLI gap is 13.8 percentage points narrower for subjects with a child age 0 to 5 than for subjects without ($p = 0.002$). Furthermore, within the set of subjects with children age 0 to 5, the gap is 14.2 percentage points narrower if the subject reported a case of diarrhea among her young children in the previous two weeks ($p = 0.015$). In fact, among this latter group, the BDM-TIOLI gap is negligible (point estimate -0.009 , standard error of estimate 0.052, $p = 0.865$). This suggests that respondents with more at stake may have taken the exercise more seriously.[27] These estimates are from single comparisons but are similar when testing multiple possible determinants of the BDM-TIOLI gap jointly (see Tables A9 and A10, with discussion in Appendix J.3).

Third, based on our piloting, we tested two hypotheses for reasons underlying a potential BDM-TIOLI gap: (a) that the TIOLI price offer could serve as an anchor; and (b) that subjects might be generally uncomfortable with the randomness involved in BDM. We included several variations of our basic BDM and TIOLI procedures as experimental sub-treatments designed to test these hypotheses. We found little evidence in support of our hypotheses from these sub-treatments. We provide details on the sub-treatments and analysis in Appendix J.4.

[27]In the language of Harrison (1992), these subjects may perceive their payoff functions to be steeper below their optimum bid, and so face a greater possible penalty for a bid that does not equal their true maximum WTP.

Fourth, evidence is not consistent with the gap being driven by lack of familiarity with the filter or by uncertainty about its benefits. As shown in Appendix J.5, we observe a BDM-TIOLI gap in demand for soap, a familiar product, during the practice rounds.

Finally, ex post regret – BDM subjects regretting their bid after the draw was realized – could be responsible for the BDM-TIOLI gap. This could arise from either misunderstanding the mechanism or non-EU preferences in which the resolution of uncertainty increases one's reference point. Immediately after the BDM price draw, we asked losing respondents if they wished they had bid more. A substantial share, 19.2 percent, said that they did, and Appendix J.6 explores this as a potential explanation of the differences between BDM and TIOLI. We note, however, that a comparable share of TIOLI subjects, 17.0 percent, attempted to bargain with surveyors even though the script emphasized there would be no bargaining.

7 Conclusion

This paper has demonstrated the use of the BDM mechanism to elicit willingness to pay for and estimate impacts of point-of-use water technology in rural Northern Ghana. We find that WTP for the filter is low, corresponding to less than 15 percent of the cost of production. Under the standard set of neoclassical assumptions, including full information, complete markets, and an efficient household, this low WTP implies that the effect of the filter on household welfare is low as well. The presence of selection on gains, i.e., the observed positive relationship between WTP and benefits, provides some support for the view that WTP reflects welfare. On the other hand, market failures may provide a rationale for subsidies, as is often assumed for health products in these contexts.

Although average WTP is low, our estimates imply that a small positive price would not dramatically reduce coverage. In fact, it would improve outcomes by screening out those for whom long-term treatment effects were negative. Combining WTP and treatment effects yields a low implied valuation for children's health: less than USD 40 per DALY and a VSL on the order of USD 3,600. Consistent with Kremer et al. (2011) in Kenya, the implied valuation is far below those typically used by public health planners or estimated in higher income countries (Viscusi and Aldy 2003).

We also show that behavior matters: the filter's benefits decrease over time and become negative for households exerting low effort. Even a technically sound product can have its effects blunted by slippage in consistency or quality of use, and policymakers should not underestimate the importance of costly effort. One direction to pursue is to invest in understanding user behavior and sustaining behavioral change. A second is to develop products that are less dependent on correct use or impose lower effort costs.

As we demonstrate, embedding BDM in field experiments can also provide insights into how use and treatment effects vary with WTP. There are numerous potential applications. In sectors where heterogeneity in returns is particularly important, such as microfinance (Meager 2018) or agriculture (Jack 2011), incorporating BDM into field experiments could enhance our understanding of such heterogeneity. In other contexts, researchers have already used incentive-compatible WTP elicitation within field experiments, and future experiments could combine BDM and treatment effect estimation to provide additional information for policy. Examples include other health products (Meridith et al. 2013), sanitation (BenYishay et al. 2017), electrification (Lee et al. 2016), and insurance (Cole et al. 2014).[28] For researchers interested in using BDM in the field, Appendix K discusses some of the practical tradeoffs between BDM and TIOLI.

However, the added information provided by BDM comes with the cost of added complexity. Most experimental mechanisms to recover valuations differ from normal market interactions, but BDM can seem particularly unusual. While the predictive power of BDM estimates for TIOLI behavior is comparable to that of TIOLI itself, demand under BDM is systematically lower than TIOLI at each of the TIOLI price points, particularly among the

[28]Within the larger set of studies eliciting WTP within field experiments, several have used BDM (e.g., Hoffmann 2009; Cole et al. 2014; Guiteras et al. 2016; Grimm et al. 2017; BenYishay et al. 2017).

most risk-averse households.

Further research on understanding the performance of BDM in field settings would be highly valuable. Our results suggest at least two useful directions. The first follows from the finding that the BDM-TIOLI gap was close to zero among subjects with lowest risk aversion. This suggests exploring ways to frame BDM to reduce the salience of randomness and further emphasize the dominance of bidding one's true maximum WTP (Cason and Plott 2014). The additional confirmation steps we added were an attempt to move in this direction, creating explicit choices similar to a multiple price list exercise (Andersen et al. 2006) in the neighborhood of subjects' initial BDM bids. Further work aimed at getting subjects to focus less on the randomization and more on how they value a product relative to a fixed sum of money would be valuable.

Second, in our exploratory analysis we found that the BDM-TIOLI gap was smaller for subjects with children aged five or under, and smaller still for those who reported that a child aged five or under had a case of diarrhea in the previous two weeks. We speculate that these subjects may have perceived that they had more at stake and taken the BDM task more seriously, thinking more carefully about their true maximum WTP. This suggests further investigation of how carefully subjects consider the BDM exercise and how best to frame BDM to increase subjects' engagement. Of course, these factors are likely to be context- and product-specific, so there may not be general answers. We expect that iteration between the field and the lab will be useful in understanding in understanding how subjects form their bids and how different aspects of the BDM protocol may influence behavior.

References

Ahuja, A., M. Kremer, and A. P. Zwane, "Providing Safe Water: Evidence from Randomized Evaluations," *Annual Review of Resource Economics*, 2010, *2*, 237–256.

Andersen, S., G. W. Harrison, M. I. Lau, and E. E. Rutström, "Elicitation Using Multiple Price List Formats," *Experimental Economics*, 2006, *9* (4), 383–405.

Ashraf, N., J. Berry, and J. M. Shapiro, "Can Higher Prices Stimulate Product Use? Evidence from a Field Experiment in Zambia," *American Economic Review*, 2010, *100* (5), 2383–2413.

Becker, G. M., M. H. DeGroot, and J. Marschak, "Measuring utility by a single-response sequential method," *Behavioral Science*, 1964, *9*, 226–236.

Bennett, D., "Does Clean Water Make You Dirty? Water Supply and Sanitation in the Philippines," *Journal of Human Resources*, 2012, *47* (1), 146–173.

BenYishay, A., A. Fraker, R. P. Guiteras, G. Palloni, N. Shah, S. Shirrell, and P. Wang, "Microcredit and willingness to pay for environmental quality: Evidence from a randomized-controlled trial of finance for sanitation in rural Cambodia," *Journal of Environmental Economics and Management*, 2017, *86*, 121–140.

Bohm, P., J. Linden, and J. Sonnegård, "Eliciting Reservation Prices: Becker-DeGroot-Marschak Mechanisms vs. Markets," *Economic Journal*, 1997, *107* (443), 1079–1089.

Brown, J. and T. Clasen, "High Adherence Is Necessary to Realize Health Gains from Water Quality Interventions," *PLoS ONE*, 2012, *7* (5), e36735.

Cason, T. N. and C. R. Plott, "Misconceptions and Game Form Recognition: Challenges to Theories of Revealed Preference and Framing," *Journal of Political Economy*, 2014, *122* (6), 1235–1270.

Chassang, S., G. Padró i Miquel, and E. Snowberg, "Selective Trials: A Principal-Agent Approach to Randomized Controlled Experiments," *American Economic Review*, 2012, *102*, 1279–1309.

Clasen, T. F., K. T. Alexander, D. Sinclair, S. Boisson, R. Peletz, H. H. Chang, F. Majorin, and S. Cairncross, "Interventions to Improve Water Quality for Preventing Diarrhoea," in "Cochrane Database of Systematic Reviews," Vol. 10 October 2015.

Clopeck, K. L., "Monitoring and Evaluation of Household Water Treatment and Safe Storage Technologies: The Sustained Use of the KOSIM Ceramic Water Filter in Northern Region Ghana." MA Thesis, M.I.T., 2009.

Cohen, J. and P. Dupas, "Free Distribution or Cost-Sharing? Evidence from a Randomized Malaria Prevention Experiment," *Quarterly Journal of Economics*, 2010, *125* (1), 1–45.

Cole, S. A., A. N. Fernando, D. Stein, and J. Tobacman, "Field Comparisons of Incentive-Compatible Preference Elicitation Techniques," *Mimeo, University of Delaware*, 2016.

_ , D. Stein, and J. Tobacman, "Dynamics of Demand for Index Insurance: Evidence from a Long-Run Field Experiment," *American Economic Review Papers and Proceedings*, 2014, *104* (5), 284–290.

Dupas, P. and E. Miguel, "Impacts and Determinants of Health Levels in Low-Income Countries," in A. V. Banerjee and E. Duflo, eds., *Handbook of Economic Field Experiments*, Vol. 2 2017, pp. 3–93.

Foster, A. D. and M. R. Rosenzweig, "Microeconomics of Technology Adoption," *Annual Review of Economics*, 2010, *2*, 395–424.

Global Burden of Disease Collaborative Network, "Global Burden of Disease Study 2016 Results," 2017.

Greenstone, M. and B. K. Jack, "Envirodevonomics: A Research Agenda for an Emerging Field," *Journal of Economic Literature*, 2015, *53* (1), 5–42.

Grimm, M., L. Lenz, J. Peters, and M. Sievert, "Demand for off-grid solar electricity – Experimental evidence from Rwanda," February 2017. ZEF Discussion Paper 232.

Gross, E., I. Guenther, and Y. Schipper, "Water Supply and Water Handling: Perceived Substitutes but Revealed Supplements for Safe Water," 2017.

Guiteras, R. P., D. I. Levine, S. P. Luby, T. H. Polley, K. K. e Jannat, and L. Unicomb, "Disgust, Shame and Soapy Water: Tests of Novel Interventions to Promote Safe Water and Hygiene," *Journal of the Association of Environmental and Resource Economists*, 2016, *3* (2), 321–359.

_ , J. Levinsohn, and A. M. Mobarak, "Encouraging Sanitation Investment in the Developing World: A Cluster-Randomized Trial," *Science*, 2015, *348* (6237), 903–906.

Hanna, R., E. Duflo, and M. Greenstone, "Up in Smoke: The Influence of Household Behavior on the Long-Run Impact of Improved Cooking Stoves," *American Economic Journal: Economic Policy*, 2016, *8* (1), 80–114.

Harrison, G. W., "Theory and Misbehavior of First-Price Auctions: Reply," *American Economic Review*, 1992, *82* (5), 1426–1443.

Heckman, J. J. and E. J. Vytlacil, "Econometric Evaluation of Social Programs, Part II: Using the Marginal Treatment Effect to Organize Alternative Econometric Estimators to Evaluate Social Programs, and to Forecast their Effects in New Environments," in J. Heckman and E. Leamer, eds., *Handbook of Econometrics*, Vol. 6B 2007, pp. 4875–5143.

_ and S. Urzúa, "Comparing IV with Structural Models: What Simple IV Can and Cannot Identify," *Journal of Econometrics*, 2010, *156* (1), 27–37.

_ , _ , and E. J. Vytlacil, "Understanding Instrumental Variables in Models with Essential Heterogeneity," *Review of Economics and Statistics*, 2006, *88* (3), 389–432.

Hoffmann, V., "Intrahousehold Allocation of Free and Purchased Mosquito Nets," *American Economic Review*, 2009, *99* (2), 236–41.

Horowitz, J. K., "The Becker-DeGroot-Marschak Mechanism is Not Necessarily Incentive Compatible, Even for Non-random Goods," *Economics Letters*, 2006, *93* (1), 6–11.

Hutubessy, R., D. Chisholm, T. T.-T. Edejer, and WHO-CHOICE, "Generalized cost-effectiveness analysis for national-level priority-setting in the health sector," *Cost Effectiveness and Resource Allocation*, 2003, *1* (8).

Imbens, G. W. and J. D. Angrist, “Identification and Estimation of Local Average Treatment Effects,” *Econometrica*, 1994, *62* (2), 467–475.

Ito, K. and S. Zhang, “Willingness to Pay for Clean Air: Evidence from Air Purifier Markets in China,” Working Paper 22367, National Bureau of Economic Research June 2016.

Jack, B. K., “Constraints on the Adoption of Agricultural Technologies in Developing Countries,” White paper, Agricultural Technology Adoption Initiative, Boston: J-PAL (MIT) and Berkeley: CEGA (UC Berkeley) 2011.

Johnson, S. M., “Health and Water Quality Monitoring of Pure Home Water’s Ceramic Filter Dissemination in the Northern Region of Ghana.” MA Thesis, M.I.T., 2007.

Jones, D., “The Economics of Exclusion Restrictions in IV Models,” NBER Working Paper 21391 July 2015.

Karlan, D. and J. Zinman, “Observing Unobservables: Identifying Information Asymmetries With a Consumer Credit Field Experiment,” *Econometrica*, 2009, *77* (6), 1993–2008.

Karni, E. and Z. Safra, “Preference Reversal and the Observability of Preferences by Experimental Methods,” *Econometrica*, 1987, *55* (3), 675–685.

Keller, L. R., U. Segal, and T. Wang, “The Becker-DeGroot-Marschak Mechanism and Generalized Utility Theories: Theoretical Predictions and Empirical Observations,” *Theory and Decision*, 1993, *34* (2), 83–97.

Kling, J. R., J. B. Liebman, and L. F. Katz, “Experimental Analysis of Neighborhood Effects,” *Econometrica*, 2007, *75 (1)*, 83–119.

Kremer, M., J. Leino, E. Miguel, and A. P. Zwane, “Spring Cleaning: Rural Water Impacts, Valuation, and Property Rights Institutions,” *Quarterly Journal of Economics*, 2011, *126*, 145–205.

Lakdawalla, D., N. Sood, and D. Goldman, “HIV breakthroughs and risky sexual behavior,” *Quarterly Journal of Economics*, 2006, *121* (3), 1063–1102.

Lee, K., E. Miguel, and C. Wolfram, “Experimental Evidence on the Demand for and Costs of Rural Electrification,” May 2016. NBER Working Paper 22292.

Li, Q. and J. S. Racine, *Nonparametric Econometrics*, Princeton University Press, 2007.

Lu, C. C., “Monitoring and Evaluation of a Ceramic Water Filter and Hand-washing Intervention in Northern Ghana,” MA Thesis, M.I.T., 2012.

Mazar, N., B. Koszegi, and D. Ariely, “True Context-dependent Preferences? The Causes of Market-Dependent Valuations,” *Journal of Behavioral Decision Making*, 2014, *27* (3), 200–208.

Meager, R., “Aggregating Distributional Treatment Effects: A Bayesian Hierarchical Analysis of the Microcredit Literature,” July 2018. Manuscript, London School of Economics.

Meridith, J., J. Robinston, S. Walker, and B. Wydick, "Keeping the doctor away: Experimental evidence on investment in preventative health products," *Journal of Development Economics*, 2013, *105*, 196–210.

Nevo, A., "Empirical Models of Consumer Behavior," *Annual Review of Economics*, 2011, *3* (1), 51–75.

Noussair, C., S. Robin, and B. Ruffieux, "Revealing Consumers' Willingness-to-Pay: A Comparison of the BDM Mechanism and the Vickrey Auction," *Journal of Economic Psychology*, 2004, *25* (6), 725–741.

Peltzman, S., "The Effects of Automobile Safety Regulation," *Journal of Political Economy*, 1975, *83* (4), 677–725.

Rutstram, E., "Home-grown Values and Incentive Compatible Auction Design," *International Journal of Game Theory*, 1998, *27* (3), 427–441.

Safra, Z., U. Segal, and A. Spivak, "The Becker-DeGroot-Marschak Mechanism and Non-expected Utility: A Testable Approach," *Journal of Risk and Uncertainty*, 1990, *3* (2), 177–190.

Shillcutt, S. D., D. G. Walker, C. A. Goodman, and A. J. Mills, "Cost Effectiveness in Low- and Middle-Income Countries: A Review of the Debates Surrounding Decision Rules," *Pharmacoeconomics*, 2009, *27* (11), 903–917.

Shogren, J. F., S. Cho, C. Koo, J. List, C. Park, P. Polo, and R. Wilhelmi, "Auction mechanisms and the measurement of WTP and WTA," *Resource and Energy Economics*, 2001, *23* (2), 97–109.

Viscusi, W. K. and J. E. Aldy, "The Value of a Statistical Life: A Critical Review of Market Estimates Throughout the World," *Journal of Risk and Uncertainty*, 2003, *27* (1), 6–76.

WHO, "Global Health Estimates 2015: Deaths by Cause, Age, Sex, by Country and by Region, 2000-2015," Technical Report, World Health Organization, Geneva 2016.

WHO and UNICEF, "Progress on Drinking Water, Sanitation and Hygiene: 2017 Update and SDG Baselines," Technical Report, Joint Monitoring Programme of World Health Organization (WHO) and the United Nations Children's Fund (UNICEF), Geneva 2017.

World Bank, *World Development Report 1993: Investing in Health*, New York: Oxford University Press, 1993.

Wright, J., S. Gundry, and R. Conroy, "Household Drinking Water in Developing Countries: A Systematic Review of Microbiological Contamination between Source and Point-of-Use," *Tropical Medicine & International Health*, 2004, *9* (1), 106–117.

Table 1: Sample Composition and Descriptive Statistics

	Mean			Diff.	Regressions	
	Full Sample	BDM	TIOLI	BDM-TIOLI	BDM Draw	TIOLI Price
	(1)	(2)	(3)	(4)	(5)	(6)
Number of respondents in compound (census)	3.593	3.305	3.859	-0.554*	0.236***	-0.051
	[2.323]	[1.816]	[2.683]	(0.323)	(0.079)	(0.045)
Husband lives in compound	0.794	0.792	0.796	-0.004	0.453	-0.243
	[0.404]	[0.406]	[0.403]	(0.022)	(0.367)	(0.168)
Number of children age 0-5 in household	1.135	1.069	1.196	-0.127*	0.195	0.028
	[0.978]	[0.941]	[1.008]	(0.073)	(0.159)	(0.078)
Number of children age 6-17 in household	1.303	1.389	1.224	0.165**	0.028	-0.013
	[1.282]	[1.304]	[1.258]	(0.084)	(0.129)	(0.047)
Number of children age 0-5 with diarrhea in past two weeks	0.243	0.208	0.277	-0.069*	-0.372	0.075
	[0.525]	[0.487]	[0.557]	(0.035)	(0.376)	(0.128)
Number of children age 6-17 with diarrhea in past two weeks	0.049	0.050	0.048	0.002	-0.499	0.463*
	[0.272]	[0.302]	[0.241]	(0.016)	(0.417)	(0.267)
Respondent has ever attended school	0.090	0.079	0.100	-0.021	-0.025	-0.077
	[0.286]	[0.270]	[0.301]	(0.016)	(0.515)	(0.195)
First principal component of durables ownership	0.132	0.059	0.198	-0.139	-0.046	0.005
	[1.555]	[1.512]	[1.592]	(0.126)	(0.091)	(0.056)
All-year access to improved water source	0.187	0.196	0.179	0.017	-0.126	0.119
	[0.390]	[0.397]	[0.384]	(0.038)	(0.376)	(0.252)
Currently treats water	0.115	0.109	0.120	-0.011	0.567	0.048
	[0.319]	[0.312]	[0.325]	(0.024)	(0.468)	(0.257)
E. coli count, standardized	-0.052	-0.026	-0.076	0.050	-0.102	0.038
	[0.949]	[1.012]	[0.887]	(0.089)	(0.162)	(0.120)
Turbidity, standardized	-0.065	-0.099	-0.032	-0.068	-0.008	0.224***
	[0.997]	[0.922]	[1.063]	(0.096)	(0.178)	(0.081)
BDM Filter Bid (GHS)					-0.093	
					(0.062)	
Number of households	1265	607	658		607	658
Number of compounds	558	275	283		275	283

Notes: Columns 1, 2 and 3 display sample means in the full sample, BDM treatment and TIOLI treatment, respectively. Column 4 displays the differences in means betweenthe BDM and TIOLI treatments. Column 5 displays the results of a regression of BDM draw on the listed characteristics. Column 6 displays the results of a regression of TIOLI price on the listed characteristics. Missing values of independent variables in Columns 5 and 6 are set to 0, and dummy variables are included to indicate missing values. Standard deviations in brackets. "Currently treats water"refers to boiling or use of a microbiologically effective filter. Standard errors clustered at the compound (extended family) level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2: Constant-Effects Instrumental Variables Estimates

	Combined all subjects		TIOLI subjects		BDM subjects	
	(1)	(2)	(3)	(4)	(5)	(6)
A. One-month followup						
A.1 Structural Equation – Dependent variable: Child age 0 to 5 has had diarrhea over previous two weeks						
Bought Filter	-0.065*	-0.072**	-0.100*	-0.098*	-0.049	-0.058
	(0.037)	(0.035)	(0.054)	(0.051)	(0.050)	(0.043)
Mean dependent variable	0.145	0.145	0.149	0.149	0.142	0.142
A.2 First Stage – Dependent variable: Household Purchased Filter						
Randomized offer price (GHS)	-0.109***	-0.107***	-0.178***	-0.172***	-0.093***	-0.093***
	(0.004)	(0.004)	(0.013)	(0.013)	(0.004)	(0.005)
F-stat	655.6	591.4	202.4	185.4	571.4	424.1
Number of compounds	472	472	244	244	229	229
Number of households	786	786	418	418	368	368
Number of children	1244	1244	665	665	579	579
B. One-year followup						
B.1 Structural Equation – Dependent variable: Child age 0 to 5 has had diarrhea over previous two weeks						
Bought Filter	0.093	0.121*	0.148	0.220**	0.090	0.108
	(0.070)	(0.071)	(0.099)	(0.100)	(0.089)	(0.090)
Mean dependent variable	0.241	0.241	0.215	0.215	0.262	0.262
B.2 First Stage – Dependent variable: Household Purchased Filter						
Randomized offer price (GHS)	-0.109***	-0.105***	-0.184***	-0.174***	-0.094***	-0.091***
	(0.006)	(0.007)	(0.017)	(0.020)	(0.006)	(0.007)
F-stat	305.9	244.0	116.2	72.3	252.6	195.0
Number of compounds	247	247	121	121	126	126
Number of subjects	387	387	197	197	190	190
Number of children	539	539	266	266	273	273
Controls	No	Yes	No	Yes	No	Yes
Village FEs	No	Yes	No	Yes	No	Yes

Notes: Each column in A.1 and B.1 displays the results of a linear two-stage least squares regression of child diarrhea status at the child level on filter purchase, where filter purchase is instrumented by random BDM draw for BDM subjects and by randomly assigned TIOLI price for TIOLI subjects. Each column in A.2 and B.2 displays the results of a linear probability model first-stage regression, where the dependent variable is an indicator for whether the household purchased a filter and the independent variable of interest is a randomized price, and the instruments are as in A.1 and B.1. Controls include all variables (other than BDM bid) listed in Table 1. Standard errors clustered at the compound (extended family) level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

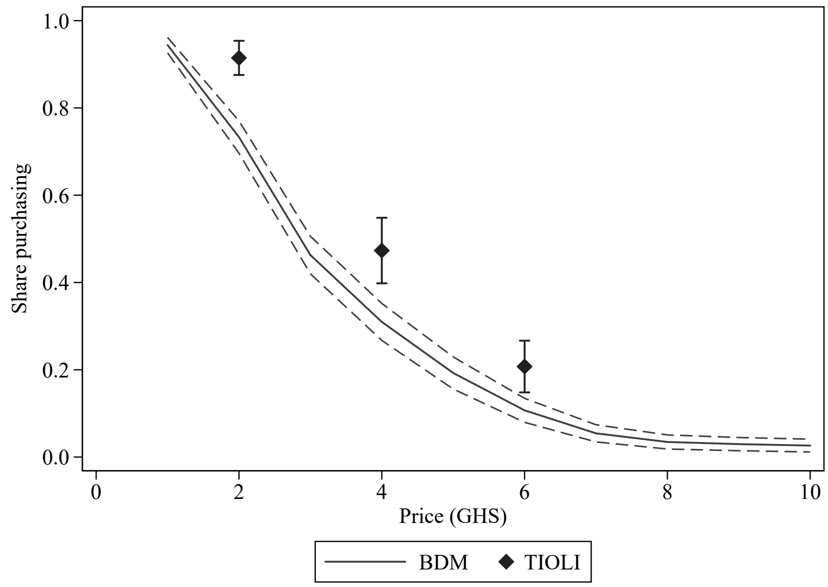
Table 3: Correlates of Willingness to Pay

	BDM		TIOLI		
	OLS		Probit		Diff. (2)-(4)
	(1)	(2)	(3)	(4)	(5)
Number of respondents in compound	0.053	0.085	-0.089^{***}	-0.117^{***}	0.203^{***}
	(0.061)	(0.059)	(0.034)	(0.035)	(0.068)
Husband lives in compound	-0.005	0.157	-0.463^{*}	-0.471^{**}	0.629^{**}
	(0.249)	(0.220)	(0.244)	(0.233)	(0.318)
Number of children age 0-5 in household	0.067	0.098	-0.066	-0.053	0.151
	(0.114)	(0.098)	(0.092)	(0.093)	(0.134)
Number of children age 6-17 in household	0.018	-0.013	0.197^{**}	0.172^{**}	-0.185^{*}
	(0.068)	(0.064)	(0.080)	(0.080)	(0.102)
Number of children age 0-5 with diarrhea in past two weeks	0.550^{*}	0.387	-0.260	-0.284	0.671^{**}
	(0.290)	(0.266)	(0.170)	(0.175)	(0.315)
Number of children age 6-17 with diarrhea in past two weeks	-0.187	-0.210	-0.663^{*}	-0.592^{*}	0.382
	(0.223)	(0.228)	(0.355)	(0.343)	(0.409)
Respondent has ever attended school	0.604	0.556	-0.535^{**}	-0.542^{**}	1.098^{**}
	(0.418)	(0.410)	(0.236)	(0.239)	(0.470)
First principal component of durables ownership	0.128^{*}	0.011	0.099	0.102	-0.092
	(0.075)	(0.066)	(0.072)	(0.068)	(0.094)
All-year access to improved water source	-0.307	-0.074	-0.259	-0.220	0.146
	(0.253)	(0.231)	(0.265)	(0.257)	(0.344)
Currently treats water	0.560	0.526	0.246	0.076	0.451
	(0.378)	(0.344)	(0.270)	(0.274)	(0.435)
E. coli count, standardized	-0.123	-0.180^{*}	0.134	0.088	-0.269
	(0.111)	(0.103)	(0.161)	(0.166)	(0.194)
Turbidity, standardized	-0.190^{**}	-0.217^{**}	0.076	0.042	-0.259^{*}
	(0.087)	(0.089)	(0.123)	(0.117)	(0.146)
BDM Soap Bid (GHS)		3.527^{***}			
		(0.579)			
Purchased soap				1.195^{***}	
				(0.261)	
R-squared	0.053	0.214			
Log-likelihood			-347.1	-321.2	
Number of households	607	607	657	656	
Number of compounds	275	275	283	282	

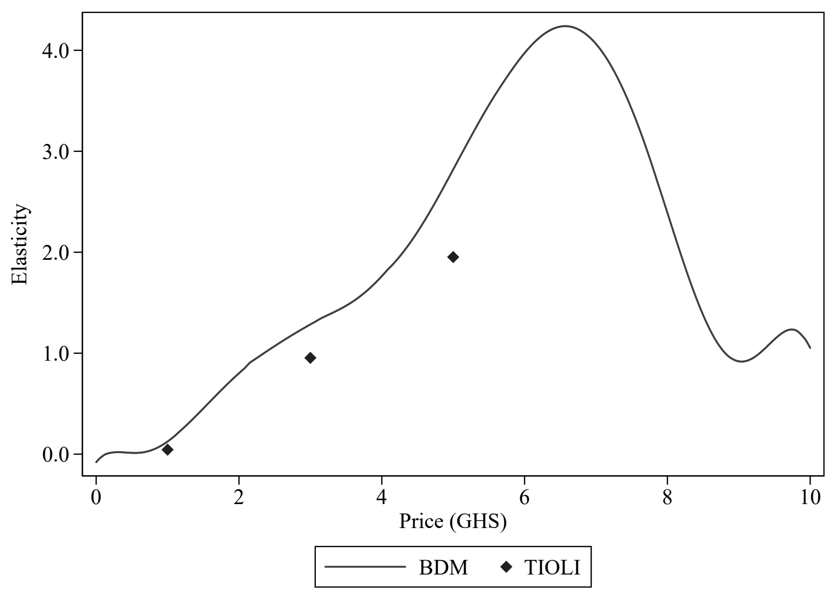
Notes: Columns (1) and (2) display coefficients from a linear regression of directly reported willingness to pay (the BDM bid) on baseline characteristics. Columns (3) and (4) report coefficients from probit models, where the dependent variable is the TIOLI purchase decision. As discussed in the text, by restricting the coefficient on price to equal -1 in the probit estimation, the estimated coefficients can be interpreted in terms of willingness to pay and are comparable to the OLS estimates from the BDM subjects. Missing values of the independent variables are set to 0, and dummy variables are included to indicate missing values. Column (5) reports differences in the estimated coefficients between BDM (Column (2)) and TIOLI (Column (4)), with standard errors calculated via SUR. Standard errors clustered at the compound (extended family) level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure 1: Demand and Price Elasticity

(a) Inverse Demand Curve



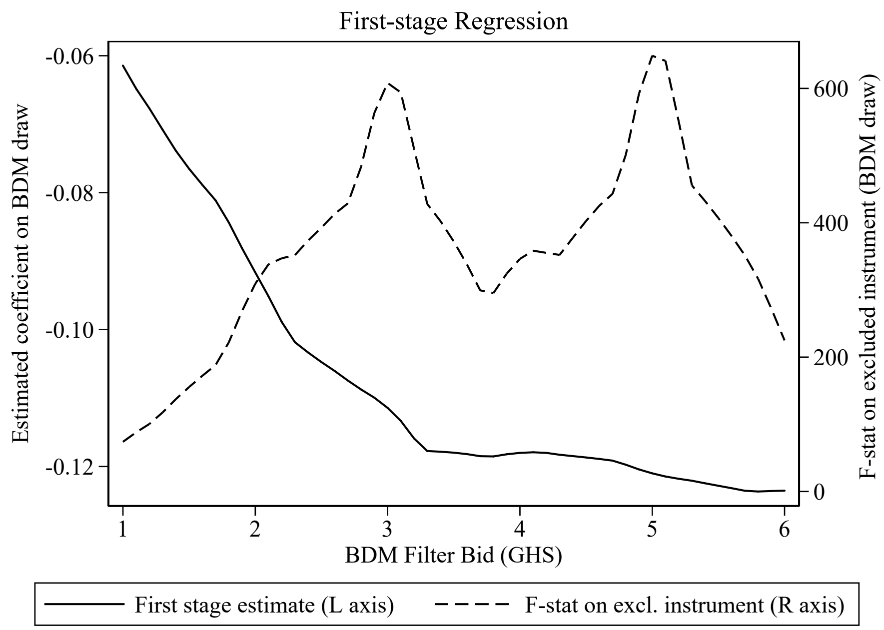
(b) Price Elasticity of Demand



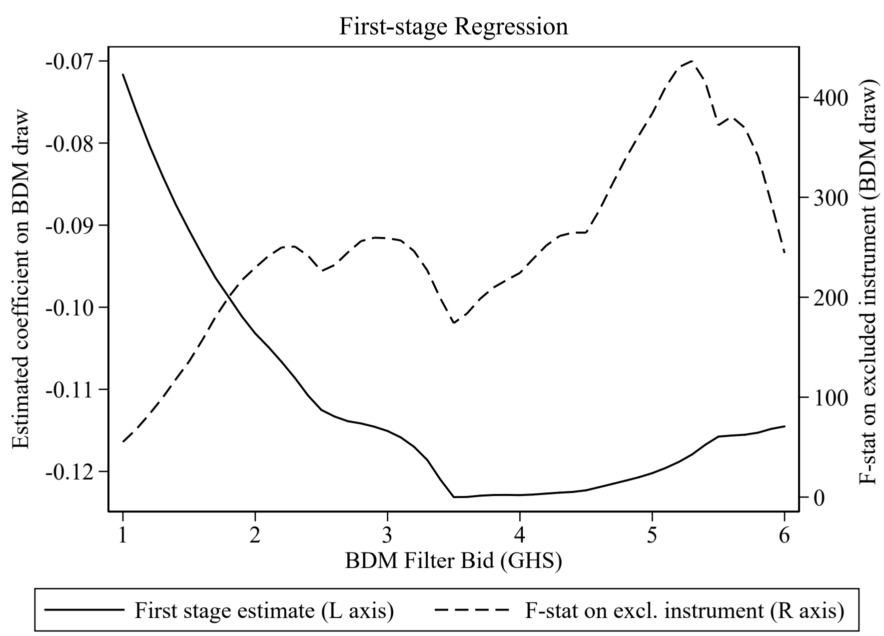
Notes: The top panel plots the BDM demand curve, with a 90% confidence band, and take-it-or-leave-it (TIOLI) demand at three price points (2, 4 and 6 GHS), with 90% confidence intervals. The BDM demand curve indicates the share of respondents with a BDM filter bid greater than or equal to the indicated price. The TIOLI markers indicate the share of respondents who purchased the filter at the corresponding (random) price. Point-wise inference from logit regressions (at prices GHS 1, 2, . . . , 10 for BDM, 2, 4, 6 for TIOLI). Standard errors clustered at the compound (extended family) level. 607 BDM observations. 658 TIOLI observations, of which 246 at a price of 2, 224 at a price of 4, and 188 at a price of 6. The bottom panel plots demand elasticities among BDM and TIOLI respondents. The BDM elasticity is calculated by a local polynomial regression, using an oversmoothed Epanechnikov kernel. The TIOLI elasticity is an arc elasticity calculated between GHS 0-2, 2-4 and 4-6 and plotted at the midpoint of each segment (GHS 1, 3 and 5, respectively). For both BDM and TIOLI, demand at a price of zero is is assumed to be 1.

Figure 2: Kernel IV Estimates of Treatment Effects
First Stage Regression

(a) Short-term: One-Month Follow-Up



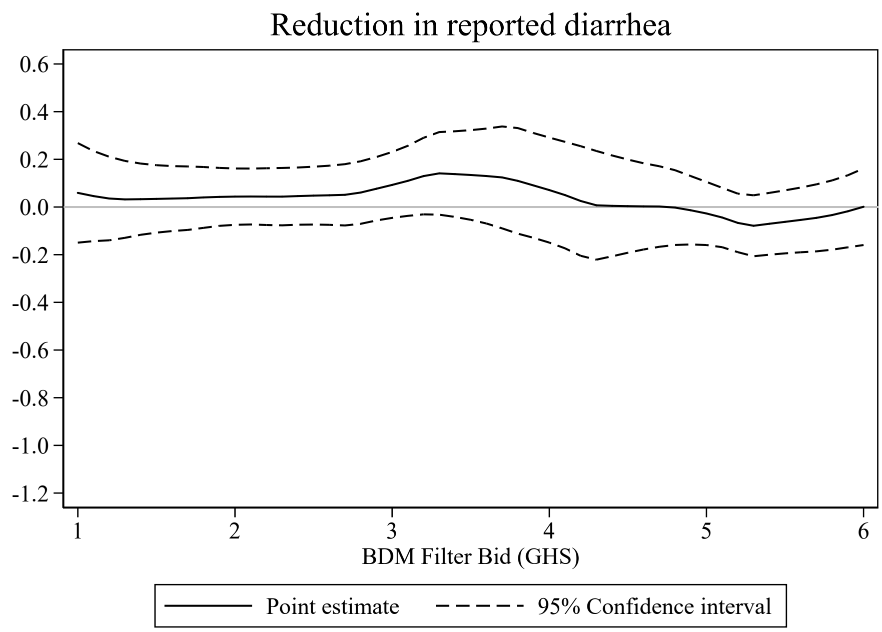
(b) Long-term: One-Year Follow-Up



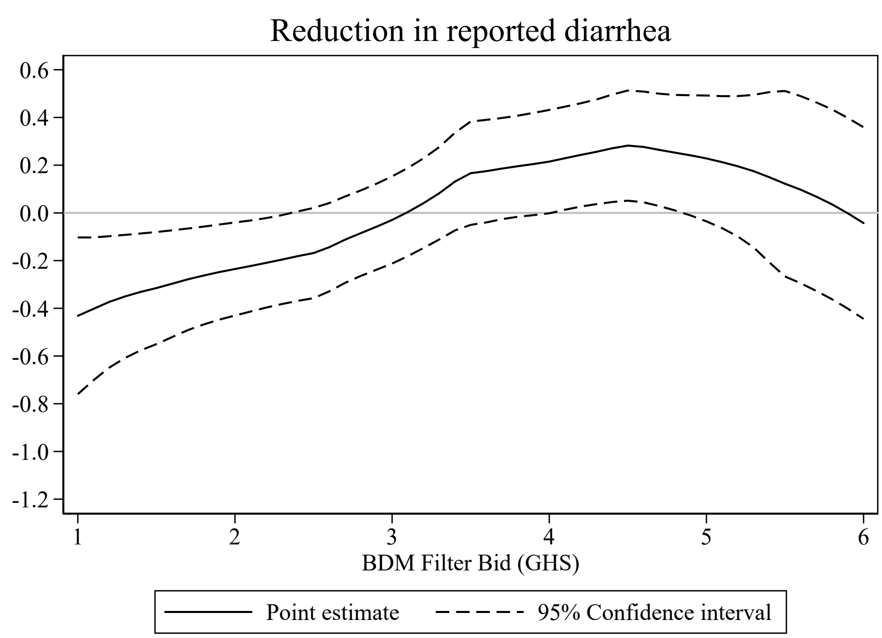
Notes: The solid line (left axis) plots the estimated coefficient on the price draw from the first-stage regression at each evaluation point WTP $= 1.0, 1.1, \ldots, 6.0$ (GHS). The dashed line (right axis) plots the F-statistic on the excluded instrument (the BDM price draw) in the first-stage regression at each evaluation point.

Figure 3: Kernel IV Estimates of Treatment Effects

(a) Short-term: One-Month Follow-Up



(b) Long-term: One-Year Follow-Up

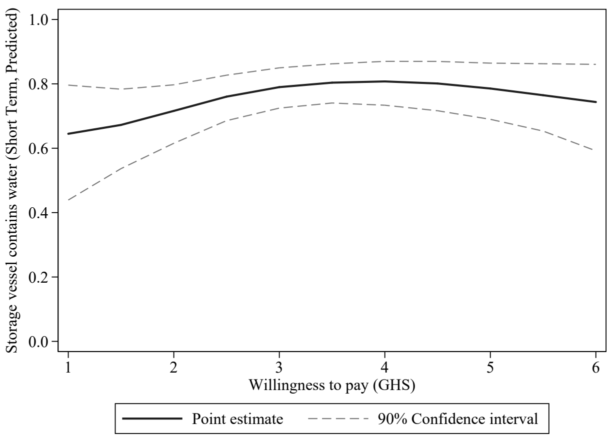


Notes: These graphs present estimated treatment effects (reduction in diarrhea among children age 0 to 5) as a function of willingness-to-pay (WTP). Estimates are by linear two-stage least squares at WTP $= 1.0, 1.1, \ldots, 6.0$, weighting observations by their distance from the evaluation point (Epanechnikov kernel, bandwidth by Silverman's rule of thumb). The endogenous treatment variable is an indicator for whether the household purchased a filter, and the exogenous instrument is the household's BDM draw. Standard errors are clustered at the compound (extended family) level. See Section 4.3 for details, and Figures 2 and A4 for first-stage results and for ancillary statistics.

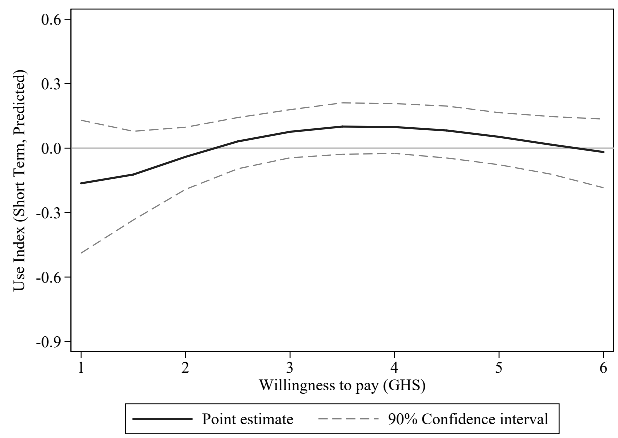
Figure 4: Relationship between Use and Willingness to Pay BDM Purchasers with Children 0 to 5

(a) One-Month Follow Up

Storage Vessel Contains Water

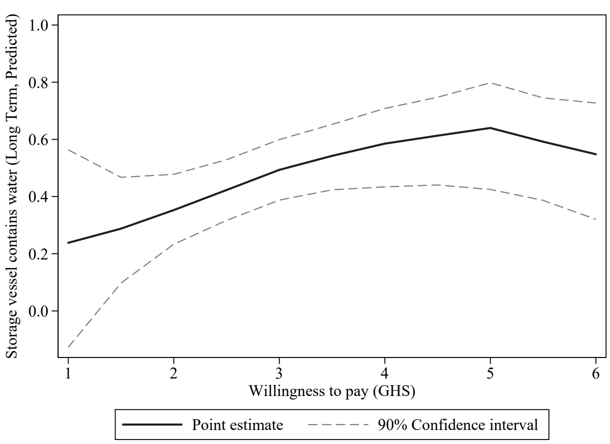


Use Index

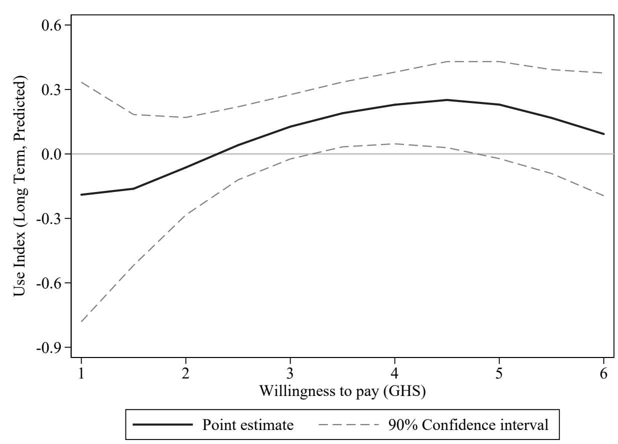


(b) One-Year Follow Up

Storage Vessel Contains Water



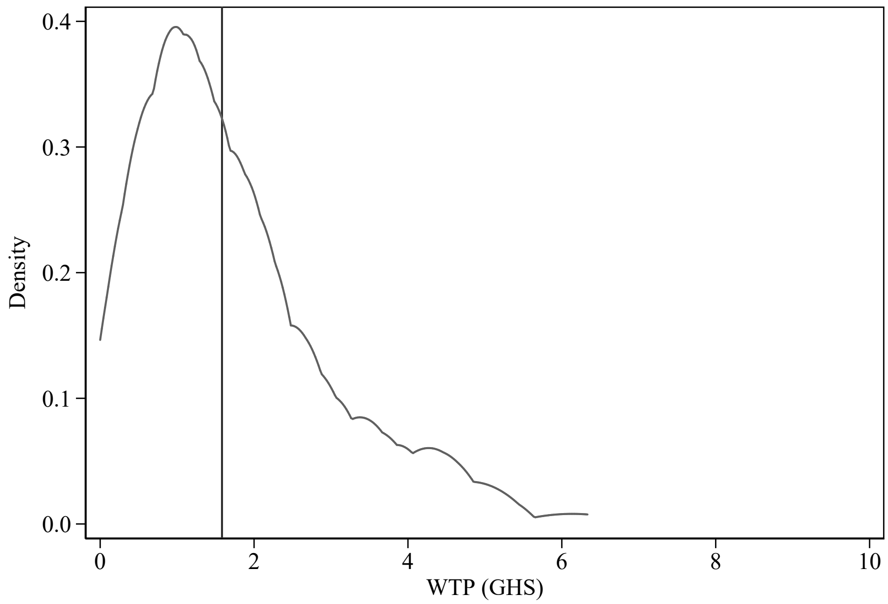
Use Index



Notes: These figures show predicted values from a kernel regression (local polynomial of degree 1) for measures of use on the household's willingness-to-pay (WTP), as stated in the BDM sale. The left figures display an indicator for whether the safe storage container contained water at or above the level of the spigot. The right figures display an index of use measures comprising indicators for whether the filter was observed in the compound, whether the ceramic pot contained water, and whether the safe storage container contained water at or above the level of the spigot. These measures are standardized and combined following Kling et al. (2007). The sample consists of households that won a filter in the BDM sale and have one or more children age 0 to 5. Confidence intervals robust to clustering at the compound (extended family) level are computed by bootstrapping, resampling compounds with replacement (1,000 repetitions).

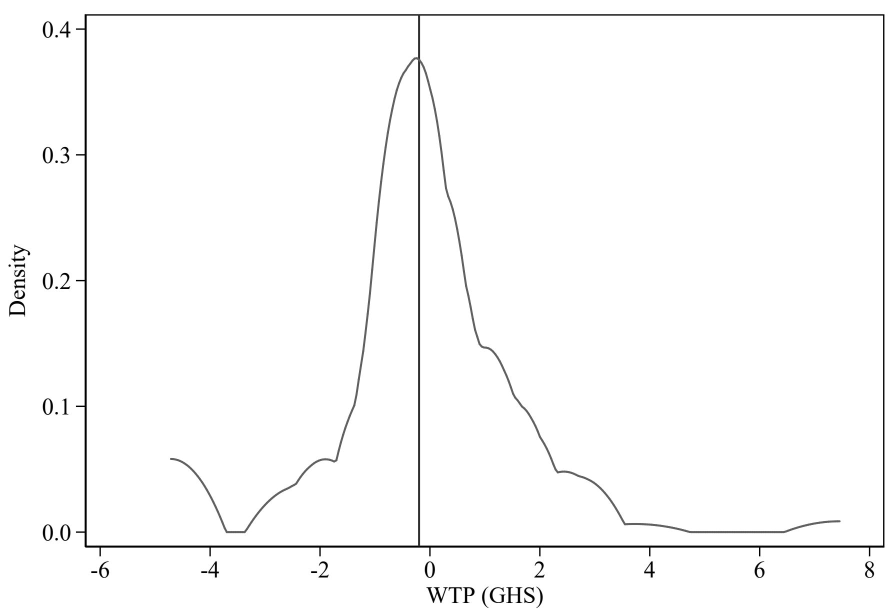
Figure 5: WTP to Avoid a Case of Children's Diarrhea

(a) One-Month Treatment Effect



Median WTP of GHS 1.58 indicated by vertical line.

(b) Average Treatment Effect over One Year



Median WTP of GHS -0.20 indicated by vertical line.

Notes: These figures present distributions of the WTP to avoid a case of diarrhea based on BDM bids and the treatment effects estimated in Section 4. In the top panel, short-term impacts on diarrhea are assumed to be constant and last for one year. The bottom panel assumes the average of short- and long-term impacts last for one year. In the bottom panel, the short-term impacts are constant and the long-term impacts are linear in willingness-to-pay.

Eliciting and Utilizing Willingness-to-Pay: Evidence from Field Trials in Northern Ghana

Online Appendices

James Berry, Greg Fischer and Raymond Guiteras*

January 2019

*Berry: University of Delaware, jimberry@udel.edu; Fischer: London School of Economics, g.fischer@lse.ac.uk; Guiteras: North Carolina State University, rpguiter@ncsu.edu.

A BDM Script

Section numbers refer to survey instrument. For full text of all sales treatments, see the Supplemental Materials.

J. REGULAR_BDM

READ EXACTLY FROM SCRIPT. DO NOT SAY ANYTHING THAT IS NOT IN SCRIPT.

READ:

- We would like to sell you a filter but the price is not yet fixed. It will be determined by chance in a game we are about to play.

- You will not have to spend any more for the filter than you really want to.

- You may even be able to buy it for less.

Here is how the promotion works:

- I will ask you to tell me the maximum price (***dan kuli***) you are willing to pay (***ka a ni sagi dali***) for the *Kosim* filter (koterigu di mali lokorigu).

- In this cup, I have many different balls with different numbers on them.

- The numbers represent prices for the filter.

- Then I will ask you to pick a ball from the cup, and we will look at the price together.

- If the number you pick is less than or equal to your bid, you will buy (***ani too dali***) the filter and you will pay the price you pick from the cup.

- If the number you pick is greater than your bid, then you cannot buy the filter.

- You will only have one chance to play for the filter.

- You cannot change your bid after you draw from the cup.

- You must state a price that you are actually able to pay now.

- We will practice in one moment, but for now, do you have any questions?

Answer any questions respondent has.

J.1 REGULAR_BDM PRACTICE

REMEMBER: Get respondent to state HIGHEST price they are WILLING AND ABLE to pay right now.

NOTE: Refer to p.23 for correct Dagbani translation of Cedi amounts.

- Before we play for the filter, let's practice the game. We'll play the same game, but instead of playing for the filter, we will play for this bar of soap. ***Show respondent soap.***

1) What is the maximum amount (dan kuli) that you are willing to pay for this soap? **[*Respondent states price X*]**

2) Now, if you pick a number that is less than or equal to X, you will buy the soap at the price you pick. If you pick a number greater than X, you will not be able to purchase the soap, even if you are willing to pay the greater number. You cannot change your bid after you pick a price. Do you understand?

3) Please, tell me - if you pick [X+5 peswas] now, what happens? ***If respondent does not give correct answer, explain the rules again and then ask question again.***

4) And if you pick [X-5 peswas] now, what happens? ***If respondent does not give correct answer, explain the rules again and then ask question again.***

5) If you draw [X+5], will you want to purchase the soap for [X+5]?
IF YES: → **5)**
IF NO: → **6)**

6) Do you want to change your bid to [X+5]?
IF YES: OK, your new bid is [X+5]. → **2) [*use* X+5 *as new* X]**
IF NO: → **6)**

7) So, is X truly the most you would want to pay?
IF YES: → **7)**
IF NO: → **1)**

8) If you pick X, you must be able to pay X. Are you able to pay X now?
IF YES: → **J.1.1**
IF NO: What is the maximum price you are willing **and able** to pay now? → **2) [*use new X*]**

→ ***Record respondent's Final Bid*** (J.1.1, page 29)

9) Could you please fetch the amount you have stated you are willing to pay and show it to me?
Wait for respondent to fetch money and check to see she has enough funds for Final Bid.

10) Now you will pick a price from the cup. If you pick X or less, you will buy the soap at the price you pick. If you pick more than X, you will not be able to buy the soap. Are you ready to pick a ball?

Mix balls in cup, hold cup above eye level of respondent and have her pick a ball without looking.

11) Now you can draw a ball from the cup. ***Let respondent draw ball. Together, look at the ball and read the price picked.*** **[*Drawn price is Y*]**

→ ***Record Drawn Price*** (J.1.2, page 29)

12) Let us look at the ball together.

→ ***Record if Drawn Price is lower/equal to or higher than Final Bid Survey*** (J.1.3, page 29)

a. **[*If Y <= X*]:** The price is Y which is [less than/equal to] the amount you said you would be willing and able to pay for this soap. You can now buy the item at this price.

→ ***Exchange payment for soap.***

b. **[*If Y > X*]:** The price is Y, which is greater than the amount you said you would be willing to spend. You cannot purchase the soap.

13) Do you have any questions about the game?

Address any questions or concerns respondent has. Make sure she understands rules of game.

J.2 REGULAR_BDM FILTER SALE

REMEMBER: Get respondent to state HIGHEST price they are WILLING AND ABLE to pay right now.

NOTE: Refer to p.23 for correct Dagbani translation of Cedi amounts.

Read:

- Now you will play to buy the filter
- Recall the community meeting on [day of community meeting]
- Have you thought about how much you are willing to pay for the filter?
- Do you have the funds available now?

Let's begin:

1) What is the maximum amount (dan kuli) that you are willing to pay for this filter? **[*Respondent states price X*]**

2) Now, if you pick a number that is less than or equal to X, you will buy the soap at the price you pick. If you pick a number greater than X, you will not be able to purchase the soap, even if you are willing to pay the greater number. You cannot change your bid after you pick a price. Do you understand?

3) Please, tell me - if you pick [X+1 cedis] now, what happens? ***If respondent does not give correct answer, explain the rules again and then ask question again.***

4) And if you pick [X-1 cedis] now, what happens? ***If respondent does not give correct answer, explain the rules again and then ask question again.***

5) If you draw [X+1], will you want to purchase the filter for [X+1]?

IF YES: **→ 5)**
IF NO: **→ 6)**

6) Do you want to change your bid to [X+1]?

IF YES: OK, your new bid is [X+1]. **→ 2) [*use* X+1 *as new* X]**
IF NO: **→ 6)**

7) So, is X truly the most you would want to pay?

IF YES: **→ 7)**
IF NO: **→ 1)**

8) If you pick X, you must be able to pay X. Are you able to pay X now?

IF YES: **→ J.2.1**
IF NO: What is the maximum price you are willing **and able** to pay now?
→ 2) [*use new X*]

→ ***Record respondent's Final Bid*** (J.2.1, page 29)

9) Could you please fetch the amount you have stated you are willing to pay and show it to me?
Wait for respondent to fetch money and check to see she has enough funds for Final Bid.

10) Now you will pick a price from the cup. If you pick X or less, you will buy the filter at the price you pick. If you pick more than X, you will not be able to buy the filter. Are you ready to pick a ball?

Mix balls in cup, hold cup above eye level of respondent and have her pick a ball without looking.

11) Now you can draw a ball from the cup. ***Let respondent draw ball. Together, look at the ball and read the price picked.* [*Drawn price is Y*]**

→ ***Record Drawn Price*** (J.2.2, page 29)

12) Let us look at the ball together.

→ ***Record if Drawn Price is lower/equal to or higher than Final Bid*** (J.2.3, page 29)

a. **[*If Y <= X*]:** The price is Y which is [less than/equal to] the amount you said you would be willing and able to pay for this filter. You can now buy the filter at this price.

→***Receive payment for filter. Record filter tracking code on survey*** (I.2.5, page 29). ***Record filter tracking code on receipt and give it to respondent. Inform her of where and when she can pick up the filter.***

b. **[*If Y > X*]:** The price is Y, which is greater than the amount you said you would be willing to spend. You cannot purchase the filter.

→***Go to Household Survey question*** J.24, page 29

B Measurement of Risk and Ambiguity Aversion

This section provides additional detail on the hypothetical gambles used to measure risk and ambiguity aversion in the one-year follow-up surveys.

To measure of risk aversion, we presented subjects with a series of choices between (a) a 50-50 gamble for a gain of 8 GHS and (b) a certain gain of X. The certain gain X began at 0.5 GHS and increased by 0.5 GHS until the subject chose the certain sum over the risky gamble. We create an integer variable to indicate the switching point and reverse the scale to yield a measure increasing in risk aversion. For example, for a subject who chose the certain 0.5 GHS over the risky gamble–the most risk-averse choice–the variable takes on a value 11, while a switching point of GHS 1 corresponds to a value of 10. The median switching point was GHS 2, corresponding to an integer value of 8. We then repeated this exercise in the loss domain, in which we measured the minimum payment at which the subject would choose a 50-50 gamble for a loss of 8 GHS over a certain payment to the experimenter. Finally, we conducted the exercise in the gain-loss domain, in which we measured the minimum sum the subject would be willing to pay to avoid a 50-50 gamble for winning 4 GHS vs. losing 4 GHS, or, if the subject were risk-loving, how much the subject would need to be compensated to forgo such a gamble. In our analysis, we use the first principal component of these three measures, but the results in Section 6.2 are robust to other methods of combining them.

To measure ambiguity aversion, we presented subjects with a version of the game posed by Ellsberg (1961). Subjects were presented with one bag that contained 5 black balls and 5 white balls, and another bag that contained 10 black and white balls in unknown proportions. The subject would choose the winning color and draw from a bag. Subjects were asked to choose between the first bag with a payout of 4 GHS and a second bag with varying payouts. The payout of the second bag started at GHS 0.5 and increased by 0.5 GHS until the subject chose the second bag. We identify subjects as ambiguity averse if they required at least 4.5 GHS to choose the second bag. By this measure, 41.6% percent of subjects are classified as ambiguity averse. We also create an integer measure of ambiguity aversion that corresponds to point at which the subject chose the second bag.

C Attrition

In this section, we discuss attrition from the follow-up surveys. The overall attrition rate was 12.9 percent in the one-month survey and 9.5 percent in the one-year survey. Table A3 shows that attrition from the one-month survey was fairly well-balanced on assignment to BDM vs. TIOLI, the BDM bid, the BDM draw, the TIOLI price, and most observable characteristics. Households that attritted were somewhat more likely to have a young child than households that were captured (7.9 pp, $p < 0.05$). In the one-year follow-up, attrition was again largely balanced on observable variables. Attritted households had significantly more young children ($p < 0.05$) and reported more young children having diarrhea in the two weeks ($p < 0.1$) prior to the baseline survey. We also find that attritters in the BDM treatment had lower WTP for the filter than non-attritters (GHS 1.0, $p < 0.01$).

While attritters in the one-year survey had lower WTP, on average, than non-attritters, our heterogeneous treatment effects are estimated *across* the distribution of WTP. The most relevant test in this case thus whether treatment is correlated with attrition at different levels of WTP. To implement this test, we estimate the following equation at different levels of WTP:

$$y_{ic} = \beta_0 + \beta_1 T_{ic} + \varepsilon_{ic}, \tag{9}$$

In this equation, y_{ic} is an indicator for whether subject i in compound c attritted from the follow-up survey, T_{ic} is an indicator for treatment (subject i's BDM bid was greater than her draw). To condition on WTP, we estimate equation (9) using a kernel (local linear) regression. As in Section 5.2, we estimate at each GHS 0.1 step from GHS 1 to GHS 6, which correspond approximately to the 0.1 and 0.9 quantiles of WTP in the BDM sample. We use an Epanechnikov kernel and Silverman's rule of thumb to choose the bandwidth. Following our analysis of heterogeneous treatment effects, we restrict the sample to BDM subjects with one or more children age 0 to 5 in one-year follow-up villages.

The results are plotted in Figure A5. As shown in the figure, there is no significant difference in attrition between treated (BDM winners) and untreated (BDM losers) once we condition on WTP. While we cannot test whether attrition is balanced on unobservables, this null result may mitigate the potential concern regarding the correlation between WTP and attrition shown in Table A4.

D Heterogeneous Treatment Effects, Detail

D.1 Heterogeneous Treatment Effects: Theory, Detail

This section provides a more detailed treatment of the theory introduced in Section 4.2 and provides greater detail on the LIV estimator of Heckman et al. (2006).

We begin with the generalized treatment effects model of Equation (3) in the main text:

$$y = \beta_0 + \beta_1 (w) T + \varepsilon. \tag{10}$$

As in Section 4.2, suppose the product is offered at two random TIOLI prices, $Z \in \{P_L, P_H\}$. If there is differential take-up at the two prices, $\Pr(T|P_L) > \Pr(T|P_H)$, then Z is correlated with T, so the instrument is relevant. For the instrument to be valid, it is necessary that $E[Zu] = 0$. Expanding u as in Equation (5), we require

$$E\left[Z\left(\tilde{\beta}_1 (w) T + \varepsilon\right)\right] = 0. \tag{11}$$

As in Section 4.2, we consider levels and gains separately. By randomization,

$$E[Z(\varepsilon)] = 0, \tag{12}$$

so the instrument solves the problem of selection on levels. However, we must also consider the selection-on-gains term

$$E[Zu] = E\left[Z\tilde{\beta}_1 (w) T\right], \tag{13}$$

which need not be zero. Even though Z is *unconditionally* random, it may not be independent of $\tilde{\beta}_1 (w) T$: since $T = 1\{\text{WTP} > Z\}$, if there is a relationship between WTP and gains then (13) will be nonzero. As a simple example, suppose $\tilde{\beta}_1 (w)$ is positively related to w. Then when $Z = p_H$, the population selecting into treatment will have, on average, high values of $\tilde{\beta}_1 (w)$ relative to the population treated when $Z = p_L$. As discussed by Heckman et al. (2006), (13) is only zero if (a) there is no heterogeneity in gains ($\beta_1 (w) = \overline{\beta}_1$ for all w, or, equivalently,$\tilde{\beta}_1 (w) = 0$ for all w,) or (b) agents either have no information on $\tilde{\beta}_1 (w)$ or, if they do have such information, they cannot or do not act on it.

As described in Section 4.2, rather than estimating either $\bar{\beta}_1$ or $\beta_1(w)$, IV estimation using TIOLI estimates:

$$\beta_1^{IV} (P_L \leq \text{WTP} \leq P_H) = \int_{P_L}^{P_H} \beta_1 (w)\, dF_{\text{WTP}} (w).$$

In order to provide estimates of $\beta_1(w)$, one could add more randomized prices $P_1, \ldots, P_M$, and, using instrumental variables as above, estimate treatment effects piecewise:

$$\beta_1^{IV}\left(P_1 \leq \text{WTP} \leq P_2\right), \ldots, \beta_1^{IV}\left(P_{M-1} \leq \text{WTP} \leq P_M\right).$$

As in the case of estimating a full demand curve using randomized TIOLI prices, this will require a relatively large sample.

A second strategy to estimate $\beta_1(w)$ is provided in the marginal treatment effects literature (Heckman and Vytlacil 2007). Given an instrument Z, the marginal treatment effect, $\Delta^{\text{MTE}}(z)$, is defined as the treatment effect on those just on the margin of indifference between being treated or not when the instrument has value z. When Z is a randomized price, $\Delta^{\text{MTE}}(w)$ is equivalent to $\beta_1(z)$, since by definition someone with $\text{WTP} = z$ is indifferent between purchasing and not purchasing at a price of w. Heckman et al. (2006) show that $\Delta^{\text{MTE}}(w)$ can be estimated even though WTP is typically not observed. Heckman et al. (2006) show that the marginal treatment effect is equal to the *local instrumental variables* parameter

$$\Delta^{\text{MTE}}(w) = \Delta^{\text{LIV}}(w) = \left.\frac{\partial E\left[y \mid \Pr(z) = \Pr\right]}{\partial \Pr}\right|_{w=z},$$

where $\Pr(Z)$ is the propensity score with respect to the instrument, representing the probability of treatment among those facing (random) price Z. The marginal treatment effect at z, then, is the change in the outcome of interest on those brought into treatment by small changes in Z around z, $\partial E\left[y \mid \Pr(z)\right] / \partial \Pr(z)$.[29] Heckman et al. (2006) provide a local instrumental variables estimator, which estimates the propensity score $\Pr(z)$ in a first step and then regresses the outcome of interest on the propensity score. As with the first strategy, this will require a large sample with a broad range of prices, since the MTE is only identified on the support of $\Pr(z)$, and the precision of the estimate depends on the precision of the estimated propensity score.[30]

BDM can estimate $\beta_1(w)$ with greater precision than these two alternatives. In the case of piecewise randomized prices, the reason is straightforward – as in the case of estimating demand curves, each BDM observation provides much more information on

[29]For intuition, note that this is a differential analogue of the traditional Wald estimator $\left(E\left[Y|Z=1\right] - E\left[Y|Z=0\right]\right) / \left(\Pr\left[T|Z=1\right] - \Pr\left[T|Z=0\right]\right)$ in the case of a binary instrument.

[30]We focus on price as an instrument for comparability with our application. However, the method of Heckman et al. (2006) applies more broadly. For example, in their empirical example, they estimate the effect of high school graduation on wages using mother's graduation status and number of siblings as instruments. Note that continuous, many-valued, or multiple instruments will be required to estimate $\Pr(z)$ flexibly. Furthermore, the interpretation of the MTE is more subtle with non-price instruments: what is estimated is $\Delta^{\text{MTE}}(u_D)$, the effect on those with unobservables $u_D \in [0,1]$ such that they are indifferent between treatment and non-treatment when the value of the instrument Z is z such that $\Pr(z) = u_D$. See Brinch et al. (2017) for progress on estimating MTEs with a discrete instrument and Kowalski (2016) for the interpretation of MTEs as a function of unobservables.

WTP than TIOLI. In the case of marginal treatment effects / local instrumental variables, BDM allows us to observe this dimension of heterogeneity directly, rather than obtaining it indirectly through the first-step propensity score estimation.

D.2 Comparison with Local Instrumental Variables

In this section, we compare estimated treatment effects using the BDM-IV method with the Local Instrumental Variables (LIV) methods of Heckman, Urzúa and Vytlacil (2006, herafter HUV). We compare estimates on the primary outcome of interest in the main text: long-term (one-year followup) cases of diarrhea among children age five and younger, based on caretaker recall over the previous two weeks.

In the first LIV step, we estimate the propensity score $\Pr(z) = \Pr[T = 1|Z = z]$, where z is the BDM draw. Following HUV, we estimate $\Pr(z)$ using locally linear regression. The estimated propensity score $\widehat{\Pr}(z)$, with a 95 percent confidence band, is plotted in Figure A6a.

In the second step, we estimate $\partial E[y \mid \Pr(z)] / \partial \Pr$ by regressing the outcome y on the estimated propensity score $\widehat{\Pr}(z)$. For comparability with our BDM-IV estimates, again we use local linear regression. The results are plotted in Figure A6b. As in the main text, we have flipped the sign of the dependent variable so benefits (reductions in diarrhea) correspond to positive point estimates.

Third, by Equation (16) in HUV (pg. 397), this derivative is the treatment effect for those at the margin of indifference when $\widehat{\Pr}(z) = \mathrm{pr}$. That is, $\partial E[y \mid \Pr(z)] / \partial \Pr = E[y_1 - y_0 \mid \Pr(z) = \mathrm{pr}]$, which, in turn, is equal to $\Delta^{\mathrm{MTE}}(\mathrm{pr})$.

Fourth, in an "inversion step," we use the fact that $\Pr(z)$ is strictly monotonic (decreasing) in z to translate $\Delta^{\mathrm{MTE}}(\mathrm{pr})$, effects plotted as a function of pr as in Figure A6b, into $\Delta^{\mathrm{MTE}}(z) = \Delta^{\mathrm{MTE}}(z : \Pr(z) = \mathrm{pr})$, effects as a function of the price draw z. This marginal treatment effect, $\Delta^{\mathrm{MTE}}(z)$, is plotted in Figure A6c. Note that, perhaps counterintuitively, relatively *low* values of the propensity score in Figure A6b correspond to relatively *high* values of the draw in Figure A6c, since the probability of treatment is low when the draw is high.

Fifth, since z is a price, $\Delta^{\mathrm{MTE}}(z)$ represents the effect those on the margin of indifference at a price of z, and this is exactly $\beta(w)$, the effect on those with WTP $= w$. That is, we can simply re-label the x-axis of Figure A6c as WTP rather than the price draw Z. Comparing Figure A6c with Figure 2.b, we observe that the pattern of estimated treatment effects is similar, in that they are increasing with respect to WTP.

Finally, in Figure A6d we compare the precision of the estimates by plotting the width of the 95 percent confidence intervals. In the case of LIV, since the regressors in the second step are estimates from the first step, we bootstrap the entire process, resampling by compound with replacement. The confidence intervals for BDM-IV are narrower over most of the range of WTP (GHS 1 to 5).

E Sunk Cost Effects

BDM embeds a double randomization that allows researchers to separately identify two factors that may be important for understanding the relationship between prices and use: the causal effect of price paid conditional on WTP (a *sunk-cost* effect), and the correlation between WTP and use (a *screening* effect). In Section 4.4, we analyze screening effects, showing that there is evidence for a positive association between WTP and use in the long-term follow-up survey.

Because the price draw is random, we can test for causal effects of price paid by comparing measures of use for subjects with the same WTP but who paid different prices. For example, BDM generates the following experiment: consider three subjects, each willing to pay GHS 6 for a filter; one doesn't receive the filter; another pays GHS 6; and the other pays GHS 2. Thus, at every level of WTP above the minimum price, there is variation in both allocation and the price paid conditional on allocation.

Following the analysis of WTP and use in Section 4.4, we use three indicators of use: presence of an undamaged filter, presence of water in the storage reservoir, and presence of water in the clay pot. We estimate the impact of price paid on each measure separately and on an index following Kling, Liebman and Katz (2007).

Specifically, we estimate

$$\text{use}_{ic} = \alpha_0 + \alpha_1 D_{ic} + \alpha_2 f(WTP_{ic}) + \varepsilon_{ic}, \tag{14}$$

where use_{ic} represents the use measure, D_{ic} is the respondent's draw, and $f(WTP_{ic})$ is a cubic polynomial of bid. It is important to control adequately for WTP since, although the price draw was unconditionally random, conditional on receiving the filter it is positively correlated with WTP.

Table A5 presents results from OLS estimation of Equation (14). Panel A shows that there is little evidence for an effect of the price paid on use in the one-month follow-up. Panel B shows a similar null result in the one-year follow-up data. Taken together, this suggests there are no significant sunk-cost effects.

F Policy Counterfactuals, Detail

This section provides further detail on the policy counterfactuals described in Section 5.1. As outlined in that section, we consider a social planner who values DALYs at B. The filter costs C_F, inclusive of production, marketing and delivery. For simplicity, we treat these costs as variable, although in reality there is likely to be a substantial fixed cost at the village level. We also abstract from time costs of use. The planner chooses the sales price P. Given a price P, we find $Q^D(P)$, the share of households purchasing the filter from our analysis in Sections 3. The total cost of filters is $C_F \cdot Q^D(P)$, the cost of the filter times demand.

We compute the reduction in cases of diarrhea per household, given by $\Delta_H(w)$, under two assumptions on treatment effects. In the first scenario, we use the treatment effect $\hat{\beta}_1^{1M}(w)$ from the one-month follow-up survey. Since there is little evidence of heterogeneous treatment effects in the short run, we restrict $\hat{\beta}_1^{1M}(w)$ to be constant with respect to WTP. Formally, this is given by:

$$\Delta_H^{1M}(w) = \hat{\beta}_1^{1M} \cdot 26 \cdot n_k,$$

where $\hat{\beta}_1^{1M}$ is the average reduction in children's diarrhea in each 2-week period, which we estimated in Section 5.2 to be 0.049, and n_k is the number of children in the household.

In the second scenario, we take the average of the short-term and one-year effects. That is, we compute the total effect of the filter over the first year as if the effect changed smoothly over the course of the year. We again assume the short-term effects are constant with respect to WTP, and impose a linear functional form on the one-year effects:

$$\Delta_H^{1Y}(w) = \left(\left(\hat{\beta}_1^{1M} + \hat{\beta}_1^{1Y}(w)\right)/2\right) \cdot 26 \cdot n_k.$$

Finally, let $F_{\text{WTP}}(w)$ be the CDF of WTP in the population. Since households with WTP $\geq P$ purchase when the price is P, the reduction in cases of diarrhea when the price is P is given by

$$\mathrm{H}(P) = \int_{w \geq P} \Delta_H(w)\, dF_{\text{WTP}}(w),$$

where $\Delta_H(w)$ is either $\Delta_H^{1M}(w)$ or $\Delta_H^{1Y}(w)$ depending on the scenario.

Following Kremer et al. (2011), we assume that the gain in DALYs is proportional to the reduction in cases of diarrhea:

$$\text{DALY}(P) = 0.028 \cdot \mathrm{H}(P),$$

where 0.028 is the ratio of DALYs to diarrhea incidence for Ghana in 2010 from Global Burden of Disease Collaborative Network (2017).

We then compute the average costs per DALY gained as the ratio of the total cost of filters divided by the total gain in DALYs:

$$\mathrm{AC}\,(P) = \frac{C_F \cdot Q^D\,(P)}{\mathrm{DALY}\,(P)}.$$

where C_F equals USD 15, as described in Section 2.

In order to avoid parameterization of the cost function, we compute the marginal cost per DALY in terms of a discrete price change from $P + 0.5.$ to $P - 0.5$. This reflects the cost per DALY of reducing the price from $P + 0.5$ to $P - 0.5$:

$$MC(P) = \frac{C_F \cdot Q^D\,(P - 0.5) - C_F \cdot Q^D\,(P + 0.5)}{DALY(P - 0.5) - DALY(P + 0.5)}.$$

This function equals the increase in costs resulting from increased demand at a lower price, divided by the increased DALYs from including the additional households purchasing at the lower price.

Table A6 displays diarrhea cases averted, DALYs averted, and average and marginal costs per DALY averted under two different assumptions about treatment effects. In Panel A, we assume constant treatment effects using data from our one-month follow-up survey. As the price increases (across columns), coverage decreases. Since we have assumed a constant treatment effect, cases reduced conditional on purchase are constant, and total cases reduced per household in the population decrease proportionally with demand. The same holds for DALYs gained conditional on purchase and total DALYs gained per household in the population. Because the treatment effect is constant, both average and marginal costs per DALY are also constant at USD 361.

In Panel B, we assume treatment effects are an average of the effects estimated from the one-month and one-year surveys. The one-month effect is assumed to be constant, while the one-year effect is assumed to be linear in WTP. Now, as price increases, negative-gains purchasers – those with low WTP – no longer purchase the filter, and diarrhea cases reduced conditional on purchase increase. For small positive prices, total gains in the population increase as well. Above a price of GHS 4, the decrease in coverage outweighs the increasing gain per household and total gains decline. We see a similar pattern in DALYs gained, both conditional on purchase (monotonically increasing with price) and total DALYs gained in the population (increasing, then decreasing, with a maximum at GHS 4). Because the treatment effect is increasing in WTP, higher prices screen out those with lower treatment effects, and average and marginal costs per DALY decrease with price.

G Valuing Health, Detail

We calculate WTP_H, the household's observed WTP to avoid a case of diarrhea, as w, the household's WTP for the filter, divided by Δd, the number of cases avoided over the anticipated life of the filter:

$$\text{WTP}_H = \frac{w}{\Delta_H(w)}. \tag{15}$$

We obtain the numerator of Equation 15, w, directly from our WTP data. As per the discussion in Appendix F, we compute the denominator, $\Delta_H(w)$, under two scenarios about the filter's impact on child health. In the first scenario, we assume the one-month treatment effects $\hat{\beta}_1^{1M}(w)$ are constant. That is, the household correctly anticipates the average benefit, but not necessarily its own benefit.

$$\text{WTP}_H = \frac{w}{\hat{\beta}_1^{1M} \cdot 26 \cdot n_k},$$

where $\hat{\beta}_1^{1M}$ defined as in Appendix F.

In the second scenario, we assume the treatment effects evolve smoothly between the short-term and one-year effects. We again assume the short-term effects are constant and that the one-year effects are linear in WTP. A linear functional form for $\hat{\beta}_1^{1Y}(w)$ implies households' beliefs and valuations are, on average, consistent: households with WTP $=$ w believe that they will receive a health benefit given by the best linear approximation of $\beta_1^{1Y}(w)$ and, on average, households are correct in this belief. The household's WTP to avoid a case of children's diarrhea, then, is

$$\text{WTP}_H = \frac{w}{\left(\left(\hat{\beta}_1^{1M} + \hat{\beta}_1^{1Y}(w)\right)/2\right) \cdot 26 \cdot n_k}.$$

H Framework for Compensating Behavior

This section provides additional detail for the discussion of possible mechanisms for the detrimental long-run impacts of the filter described in Section 4.5. We begin with a simple model of household health production to frame the issue. Consider a world where households maximize additively separable utility over children's health (h) and all other consumption (x) subject to a budget y. Children's health is a function of both general health behavior s, which we label "sanitation" to fix ideas,and the consumption of clean water (w): $h(s,\, w)$, where both inputs are continuous and non-negative with unit costs p_s and p_w. We make the usual assumptions: $h_s > 0$, $h_{ss} < 0$, $h_w > 0$, and $h_{ww} < 0$. The household's maximization problem is then to pick a vector of inputs, (s, w, x), that maximize utility, $h(s,\, w) + x$, subject to the budget constraint, $p_s s + p_w w + p_x x \leq y$.The filter reduces the per unit cost of clean water, p_w. If clean water and sanitation are substitutes in the health production function, $h_{sw} < 0$, a reduction in the price of clean water will reduce sanitation and other investments in children's health. The substitution between s and w alone could explain a muted or even zero impact from the filter; however, it could not generate the perverse effects that we observe.

In Section 4.5, we outline the three factors that we believe are mostly likely to have combined with compensatory behavior to generate detrimental effects in our context: sporadic reoptimization in response to gradual declines in use of the filter or the filter's effectiveness; intrahousehold allocation decisions that limited children's access to filtered water; and non-convexities in the alternative health technologies. Combined with compensatory behavior, each can produce negative treatment effects.

First, upon receipt of the filter – a large shock to their health production function – households may have reoptimized, engaging in compensatory behavior. Then, in response to a gradual decrease in use or the filter's effectiveness over time, they may have failed to reoptimize again, either due to rational inattention (Tobin 1982; Reis 2006; Da et al. 2014) or simple mistakes: households may have misperceived the benefits of maintaining or using the filter. If households that value the filter more also tend to be more attentive, we would expect more failures to reoptimize among those with low WTP.

Second, even in households that purchased the filter, some children may not have had access to the filtered water. The filter produces a limited supply of drinking water, but this water comprises multiple goods. Most importantly, we consider children's health and better tasting water for adults. Before receiving the filter, households made health investments (such as traveling to cleaner water sources or boiling their water) that jointly produced both goods. The filter can decouple this production.. With the filter, adults can obtain better tasting water with less effort devoted to activities that improve water quality for the entire household. In particular, field reports indicated that some children were not allowed to drink filtered water because of concerns that they might damage the filter or that there would be insufficient "sweet tasting water" for the male head of the household.[31] The pattern of treatment effects we observe is consistent with this mechanism. Households with a low value for children's health would be less likely to provide filtered

[31] In response to the field reports, we added survey questions regarding children's access to filtered water, but subjects' answers proved unreliable.

water for their children and, all else equal, tend to have a lower WTP for the bundle of goods produced by the filter.

Finally, compensatory behavior can worsen the targeted outcome while improving utility if the alternative health production technology is non-convex. Many health behaviors have a fixed cost component. For example, suppose a household can either obtain its water at low cost from a dirty source or at ha higher cost from a cleaner source. Without the filter, the household chooses to incur the higher cost and drink relatively clean water. The filter improves the quality of the dirty water sufficiently that, if the household has the filter, it optimally chooses not to incur the cost of obtaining clean water. If filtered water with low other investment produces less health than unfiltered water with high other investment, purchasing the filter can increase utility but reduce health.

Consider a setting in which the individual can (1) either choose to obtain water from a clean source (cw) or a dirty source (dw) and (2) either use the filter or not. Note that although we describe this as clean vs. dirty water, this could be any of a set of sanitation practices. Without the filter, the individual has utility as follows:

$$U^{NF} = max\left\{h^{cw} - c^{cw},\ h^{dw} - c^{dw}\right\},$$

and with the filter, the individual has utility:

$$U^{F} = max\left\{h^{cw} + F^{cw} - c^{cw},\ h^{dw} + F^{dw} - c^{dw}\right\},$$

where h^s represents the utility of health when water is obtained from source $s \in \{cw, dw\}$, F^s is the effect of the filter on the utility of water obtained from s, and c^s is the cost of obtaining water from s.

Assumption 1. *The clean water is better for your health:* $h^{cw} > h^{dw}$,

and

Assumption 2. *The filter works:* $F^{cw} > 0;\ F^{dw} > 0.$

Consider the case where without the filter, individuals use the clean water source:

$$h^{cw} - c^{cw} > h^{dw} - c^{dw} \tag{16}$$

Based on the assumption that the filter works, receiving the filter could lead to worse health if with the filter the individual switches to the dirty water source:

$$h^{dw} - c^{dw} + F^{dw} > h^{cw} - c^{cw} + F^{cw}. \tag{17}$$

If health worsens after receiving the filter, this implies:

$$h^{cw} > h^{dw} + F^{dw}. \tag{18}$$

For notational compactness, we define $\Delta h = h^{cw} - h^{dw}$ and the analogous variables similarly. Combining the preceding equations implies:

$$\Delta c - \Delta F > \Delta h > F^{dw}$$

$$\Delta c > F^{cw}.$$

That is: (1) the health benefit of clean vs. dirty water needs to be greater than the health benefit of using the filter with dirty water (thus health goes down if households with the filter switch to dirty water) and (2) the cost of obtaining clean vs. dirty water needs to be greater than the health benefit of using the filter with clean water (thus households improve utility by switching to dirty water once they have the filter even though this reduces health). Note this also implies that $F^{cw} < F^{dw}$, that is, the filter has less of a health benefit with clean water than with dirty. Further, the willingness to pay for the filter equals $U^F - U^{NF}$, so those who have perverse effects – and hence all else equal get less net benefit from the filter – will also have relatively low willingness to pay. As in Peltzman (1973), the non-convexity of the health production technology generates the possibility of perverse effects and differentiates this setting from benchmark models of health investment such as described in Greenstone and Jack (2015).

We also considered alternative mechanisms, such as improper use of the filter or sanitation externalities (as discussed in Bennett 2012). While either could, in principle, produce detrimental impacts, both are unlikely in our setting. We find no evidence of improper use causing detrimental effects. In fact, conditional on use, nearly all filters were in good condition and well maintained. As for externalities, our heterogeneity analysis finds that those with a low WTP who receive the filter have worse one-year outcomes than households in the same village with the same WTP who were not randomly assigned – via the BDM price draw or randomized price – to receive the filter. Since the treatment status of one's neighbors, who could be generating the negative externalities, is independent of one's own treatment status, sanitation externalities are unlikely to explain the observed pattern of effects.

I Correlates of WTP, Detail

I.1 Feature Selection

As described in Section 6, we find that a model of demand using an a priori list of covariates such as wealth, education, and health status has limited predictive power for both TIOLI purchase decisions and WTP elicited directly through BDM. This reflects a common pattern for studies of health goods in low-income countries (Ashraf et al. 2010; Cohen and Dupas 2010) and the consumer behavior literature more generally (Browning and Carro 2007; Nevo 2011). In this appendix, we describe the use of LASSO for covariate selection.

The LASSO (Tibshirani 1996), common in the machine learning literature, is a penalized regression approach to variable subset (model) selection in which the data determine the set of covariates. It solves a similar minimization problem to ordinary least squares, but with a penalty for model complexity. This produces something similar to a linear regression in which only a small number of predictors have non-zero coefficients. The parameter estimates are given by

$$argmin\, L(\beta|x) + \lambda \sum_{j=1}^{p} \left|\beta_j\right| ,$$

where $L(\beta|x)$ is the loss function, usually a quantity proportional to the negative log likelihood. In our setting, we use the residual sum of squared errors for the BDM data, where WTP is directly observed, and the negative log likelihood of the logistic function for the TIOLI data, where WTP is a latent variable. The term $\lambda \sum_{j=1}^{p} \left|\beta_j\right|$ penalizes the inclusion of additional regressors, and λ is a tuning parameter that determines the extent of this penalty. For small λ, the penalty is minor and LASSO recovers the OLS regression coefficients. When λ is sufficiently large, some of the coefficients will be set to zero and the LASSO performs variable selection.

In order to perform the LASSO, we standardize all predictors to have variance one. We then randomly allocate half of the sample to a training set, which will be used to select the tuning parameter with the best out-of-sample prediction properties, and half to a hold-out set, on which we will estimate the model using the selected tuning parameter. We choose a grid of λ values for which we compute the out-of-sample prediction error using 10-fold cross validation.

To construct the cross-validation error, we divide our training sample, for example, the BDM observations allocated to the training set, into 10 groups, or folds, of approximately equal size. We reserve the observations in the first fold as a validation set and fit the model, for each value of λ in the grid, on the observations in the other nine folds. We then calculate the prediction error for the observations in the first (validation) fold. This is calculated as the mean squared error in the BDM sample and the classification error in the TIOLI sample. We repeat this procedure ten times, using each fold as a validation set. This produces ten estimates of the cross-validation error for each value of λ. We select then refit the model using all available observations and the largest value of λ (the most parsimonious model) that produced the lowest cross-validation error. This determines the set of covariates. We then estimate the model using the hold-out sample.

We include the following set of baseline features in the models for both BDM and TIOLI purchase decisions: number of adult females in the compound; number of adult males in the compound; number of children in the compound; number of children aged 5 or less; marital status; whether the respondent is the primary caregiver; indicators and counts of household assets (bicycle, bucket, chair, sewing machine, cooking pot, cutlass, lantern, light bulb, mattress, mobile phone, motorcycle, radio, refrigerator, sewing machine, television, torches, video player); the first principle component of all assets; educational attainment; acres farmed; acres owned; shared land farmed; total loans outstanding; primary occupation; pregnant; beliefs about actions that prevent diarrhea (boiling water, clean clothes, clean dishes, eating clean food, cooking food, drink clean water, filter water, good hygiene, using a latrine, medication, prayer or God, treating water, washing hands, nothing, does not know); respondent has primary responsibility for water collection in the compound; water source in the dry season (well; dugout; dam; borehole; rainwater; private standpipe; public standpipe; public dug well; river, stream or pond); water source in the rainy season (same categories as dry); status of water source in dry season; protection status of water source in rainy season; water treatment activities (boiling, ceramic filter, chemicals, cloth filter, pipe filter, letting settle); and village fixed effects.

We also consider measures of preferences and numeracy that were elicited for a subset of households at the one-year follow-up: risk aversion in the gains domain, risk aversion in the loss domain, and risk aversion in the gain/loss domain, forward digit span, backwards digit span, total digit span, ambiguity aversion categories, and whether the respondent indicated that she felt lucky in games of chance.

For the BDM subjects, we include the expressed willingness to pay for soap in the BDM practice round. For TIOLI subjects, we include each subject's purchase decision for soap in the TIOLI practice round. We do not include the TIOLI price for soap, which was randomly assigned and orthogonal to both individual characteristics and the TIOLI price for the filter.

Table A7 reports the selected features and estimated coefficients in the hold-out samples for both the BDM and TIOLI groups. In both samples, the dominant feature relates to the respondent's purchase decision for soap in the practice round. Those with a higher valuation for soap were more likely to purchase the filter or value it highly. This predictive power holds even when including all other household characteristics for which we would imagine soap purchase behavior might serve as a proxy, such as education, wealth, income, health beliefs, practices and status. Only village fixed effects, in the case of TIOLI, have comparable predictive power. In contrast, other household characteristics explain little of purchasing behavior. Characteristics related to education and asset ownership, which are often considered predictive of demand for health (Ashraf, Berry and Shapiro 2010; Cohen and Dupas 2010), appear in the regularized model for BDM demand but carry relatively limited explanatory power; they do not appear in the model for TIOLI purchase decisions.

We then expand the set of features to include preferences and numeracy and estimate the model on the smaller sample for which this data is available. Digit span (0.130), risk aversion in the gains domain (-0.074), and the indicator for whether the respondent indicated that she felt lucky in games of chance (0.046) appear in the regularized model for BDM subjects. None of these features appears in the TIOLI model.

I.2 Cross Validation

In this appendix we describe the k-fold cross validation procedure used to assess the performance of the BDM and TIOLI mechanisms in predicting TIOLI responses. We begin by randomly dividing the TIOLI sample into 10 groups of approximately equal size. The first group is treated as a validation set, and we fit the latent demand model in Equation (8) on the remaining nine groups via probit. We repeat this procedure ten times, treating each group as a validation set in turn. We denote by $\hat{r}^{(-k)}(X_{ic}; p)$ the predicted probability of purchasing at price p for an individual with characteristics X_{ic}, computed with the k^{th} part of the data removed. We then form a predicted binary purchase decision, $\hat{b}_i = 1\left(\hat{r}^{(-k)}(X_{ic}; P_{ic}) \geq 0.5\right)$ for each observation in the validation set, where P_{ic} is the randomized TIOLI price actually faced by household i in compound c. We repeat this procedure for all ten folds. We then estimate the accuracy of TIOLI for out-of-sample prediction of behavior under the TIOLI mechanism itself based on the share of correct predictions in the full TIOLI sample. The resulting accuracy rate is 76.0 percent.

To calculate the analogous accuracy rate of prediction based on the BDM mechanism, we randomly divide the BDM sample into 10 groups of approximately equal size. Since the validation set is drawn from the TIOLI sample, this procedure serves to replicate the sampling variability and sample size effects of the cross-validation procedure within the TIOLI sample. We estimate Equation (7) from the main text for the test set via ordinary-least-squares and then estimate $w\hat{t}p^{(-k)}(X_{ic})$ for each observation in the corresponding validation set from the TIOLI sample. Based on this estimation, we form a predicted binary purchase decision $\hat{b}_i = 1\left(w\hat{t}p^{(-k)}(X_{ic}) \geq P_{ic}\right)$, for each observation in the validation set, where again P_{ic} is the randomized TIOLI price actually faced by household i in compound c. We repeat this procedure for all ten folds and estimate the accuracy of BDM for out-of-sample prediction of behavior under the TIOLI mechanism. The share of correct predictions in the full TIOLI sample is 73.9 percent. These accuracy rates compare to a base rate—the accuracy of trivially predicting the most-frequent decision within each validation set—of 56.2 percent. Consistent with the pattern of demand estimated by two mechanisms, TIOLI more accurately predicts affirmative purchase decisions while BDM performs better when predicting refusals.

To explore the relative performance of the two mechanisms in greater depth, we construct ROC curves for both mechanisms and compare model accuracy via their respective areas under the curve (AUCs). The ROC curve plots the sensitivity of the predictive model (the rate of true positives) on the y-axis against the specificity (the rate of true negatives) on the x-axis as we vary the cutoff for predicting a purchase. The simple comparison above is equivalent to setting the cutoff at a 50 percent probability of purchase. The AUC is a commonly used measure to summarize the performance of a classifier over all possible thresholds. Figure A8 displays the AUCs for the BDM and TIOLI models.

The diagonal represents the performance of a model that randomly classified each observation. For TIOLI and BDM, the AUCs are 83 percent and 79 percent respectively. While TIOLI outperforms BDM in predicting TIOLI behavior, their performance is remarkably close. We consider this encouraging evidence that, at least in this setting, the noise generated by the BDM mechanism is outweighed by the additional information it provides.

J Mechanism Effects, Detail

In this section we extend the discussion in Section 6.2 by providing further analysis of the magnitude and potential sources of the differences between BDM and TIOLI-elicited WTP.

J.1 Comparing Demand Under BDM and TIOLI

This subsection presents regression estimates of the differences in demand between BDM and TIOLI at the three TIOLI price points, as displayed in Figure 1a of the main text.

In order to perform the comparison, we run a similar regression to that presented in Section 6.2. We estimate

$$\text{buy}_{icp} = \alpha_p + \beta_p \text{BDM}_{ic} + x_{ic}'\gamma + \varepsilon_{icp}, \tag{19}$$

where buy_{icp} indicates whether subject i in compound c purchased at price p (under the TIOLI mechanism), or would have purchased at price p given her bid (under the BDM mechanism), and BDM_{ic} is an indicator for whether subject i was assigned to the BDM mechanism. For each price p, α_p represents the share purchasing under TIOLI and β_p represents the difference in shares between BDM and TIOLI at price p.[32]

The regression results are presented in Table A8. As shown in Columns 1, 3, and 5, the difference between the two mechanisms is significant at the 5 percent level or greater for each of the three prices. The test of joint significance of all three differences yields a p-value of less than 0.001. While the absolute (percentage point) differences are declining with each price, we cannot reject that all three differences are equal ($p = 0.239$ without controls; $p = 0.354$ with controls), and there is no such pattern in relative (percentage) differences. As shown in Columns 2, 4, and 6, the results are virtually unchanged with the inclusion of controls.

J.2 Correlation Between BDM-TIOLI Gap and Risk Aversion

This sub-section presents details on the comparison of the BDM-TIOLI gap across terciles of risk aversion, discussed in Section 6.2 of the main text. In order to implement the comparison between BDM and TIOLI, we collapse the more precise individual WTP information from BDM to the binary purchase indicators generated by TIOLI. Our outcome variable is $\text{buy}_{i,p}$, which represents subject i's purchase decision when facing a price $p \in \{2, 4, 6\}$. For TIOLI subjects, this is just whether they agreed to purchase at the offer price. For BDM subjects, $\text{buy}_{i,p} = 1\{\text{WTP}_i \geq p\}$, where WTP_i is subject i's BDM

[32] Since each BDM subject's bid can be used to simulate purchase behavior at all three prices, each regression contains about three times as many BDM observations as TIOLI observations. We estimate the system (one equation for each of $p = 2, 4, 6$) via seemingly unrelated regression to account for correlation of errors across equations and to conduct cross-equation tests.

bid. We create the variables RA_i^1, RA_i^2, RA_i^3 to indicate that subject i is in the first (most risk-averse), second, or third (least risk-averse) tercile, respectively. We then estimate

$$\mathrm{buy}_{icp} = \sum_{t=1}^{3} \alpha_p^t \mathrm{RA}_i^t + \sum_{t=1}^{3} \beta_p^t \left(\mathrm{RA}_i^t \times \mathrm{BDM}_i\right) + x_{ic}'\gamma + \varepsilon_{icp}, \tag{20}$$

where BDM_i is an indicator for whether subject i was assigned to the BDM mechanism. For each price p, α_p^t represents the purchase probability for TIOLI subjects in the t-th tercile, while β_p^t represents the "BDM effect" in the t-th tercile. The differences without controls are presented in Figure A9. The top panel plots the estimated coefficients $\hat{\beta}_2^1$, $\hat{\beta}_4^1$, $\hat{\beta}_6^1$, with 90 percent confidence intervals, for tercile 1 of risk aversion (the most risk-averse subjects), while the middle and bottom panels plot the same set of coefficients for terciles 2 and 3 (the least risk-averse subjects), respectively. As Figure A9 makes clear, the BDM-TIOLI gap is largest among the most risk-averse subjects (mean BDM effect -0.200, $p = 0.000$), and has largely closed among the least risk-averse subjects (mean BDM effect -0.051, $p = 0.425$). These results are unconditional, but they are robust to controlling for a large set of household controls (see Figure A10) and when testing multiple possible determinants of the BDM-TIOLI gap jointly (see Table A9).

J.3 Correlation Between BDM-TIOLI Gap and Observables

To supplement the analysis of risk aversion presented in Section 6.2, this section presents an exploratory analysis of the correlation between the BDM-TIOLI gap and other relevant observables. For binary observables, we compare the BDM-TIOLI gap between the two levels of the variable; for continuous observables, we break the sample into terciles and compare the top to the bottom tercile. Similar to Equation (20) of the main text, we estimate

$$\mathrm{buy}_{icp} = \alpha_{0p} + \alpha_{1p} D_i + \beta_{0p}\mathrm{BDM}_i + \beta_{1p}\left(D_i \times \mathrm{BDM}_i\right) + \varepsilon_{icp}, \tag{21}$$

where D_i is an indicator for the subgroup of interest and the other variables are as in Equation (20). For each price p, β_{0p} is the BDM-TIOLI gap (the difference in purchase probabilities) for subjects with $D_1 = 0$, $\beta_{0p} + \beta_{1p}$ is the BDM-TIOLI gap for subjects with $D_1 = 1$, and β_{1p} is the difference between the two sub-groups. We then average the coefficients over the three TIOLI prices to obtain $\overline{\beta}_0$, $\overline{\beta}_1$ and $\overline{\beta}_0 + \overline{\beta}_1$.

Figure A11 presents the results, with levels ($\overline{\beta}_0$ and $\overline{\beta}_0 + \overline{\beta}_1$) in the top panel (Figure A11a) and differences ($\overline{\beta}_1$) in the bottom panel (Figure A11b). First, household wealth is associated with a smaller BDM-TIOLI gap, but not to the same extent as risk aversion (see Figure A9 and discussion above). Second, the gap among subjects who have attended school is approximately zero, although this is imprecisely estimated since only nine percent of subjects have ever attended school. On the other hand, the gap is wider among subjects who scored in the top tercile of the digit span test. Third, the gap is narrower among subjects who have a child age 0 to 5, and narrower still (with a point estimate close to zero) if one or more children has had a case of diarrhea in the previous two weeks. This may suggest that respondents with more at stake took the exercise more seriously. On the other hand, the gap is significantly wider among respondents whose water samples were in the top tercile in terms of *E. coli* (those with the poorest water quality). This is somewhat surprising, especially since the gap is largely unaffected by turbidity, which, unlike *E. coli*, is visible.

The results in Figure A11 test one covariate at a time, but the results are generally similar when we test several covariates jointly, as shown in Table A9. Note that water quality is not included in this comparison because of sample size limitations – we collected the risk aversion measure only in the one-year sample of villages (8 of 15 villages), and tested water only in a 50 percent subsample of households. Table A10 repeats this exercise for the full sample with just the variables available for all households household survey, and shows that the coefficients and statistical significance of these variables are similar in single and joint tests.

J.4 BDM and TIOLI Experimental Sub-Treatments

This section describes the experimental sub-treatments designed to test mechanisms behind the BDM-TIOLI gap. We test the "standard" presentation of the BDM and TIOLI mechanisms against four sub-treatments that incorporated modifications to the sales scripts. Descriptive graphs demand across treatments are provided in Figures A12 and A13, with formal statistical tests reported in Tables A11 and A12.[33]

The first two sub-treatments were designed to test the hypothesis that the stated prices in the TIOLI treatment could cause respondents to anchor their valuations to those prices. In the "anchoring" treatments for both BDM and TIOLI, we informed subjects that the price of the filter in the Tamale town market (the nearest market town) was GHS 20. Based on our pilot results, we believed this information would dominate any conveyed in the TIOLI price, placing both mechanisms on equal footing and allowing us to estimate any anchoring or signaling effects from the offer price. However, these anchoring treatments did not produce any consistent effect on BDM bids or TIOLI purchase behavior. There was a significant effect on TIOLI demand at GHS 4 (-0.233, $p < 0.05$), but there was no effect at the other TIOLI prices or in BDM bids.

We also included a "random TIOLI" sub-treatment, in which the TIOLI offer price was drawn by the respondent from a cup of numbered wooden beads, the same mechanism used to determine the BDM price. The aim was to make salient the arbitrariness of the TIOLI prices and reduce the likelihood that they were serving as signals of quality. Based

[33]Table A11 presents the results of two tests that compare the distributions of the BDM sub-treatments using both the Wilcoxon-Mann-Whitney rank-sum and Kolmogorov-Smirnov tests. Cluster-robust significance levels for the distributional tests are constructed via a bootstrap percentile method. We pool data from the two treatments being compared, draw block-bootstrap samples, where the compound is the block, and then randomly assign placebo treatments by compound and run the distributional test in question. Since the placebo treatments are randomly generated, the null hypothesis of equality of distribution is true by construction. By sampling compounds and assigning placebo treatments by compound, we preserve the clustering structure in the data. We repeat this for 500 bootstrap repetitions, and then obtain a p-value for our test by finding where the original test statistic falls in the distribution of bootstrap test statistics.

on our pilot results and the evidence that in some settings BDM bids are sensitive to the underlying price distribution (Bohm et al. 1997; Urbancic 2011; Mazar et al. 2014), we hypothesized that the randomness in the price draw may contribute to the BDM-TIOLI gap, through a failure to reduce compound lotteries, subjects' general discomfort with randomness and ambiguity, or other departures from expected utility maximization. However, demand under the random TIOLI treatment was statistically indistinguishable from standard TIOLI, indicating that our efforts to equate the perceived randomness in the two mechanisms had no effect on subjects' purchase behavior. We note, however, that our modifications were designed to increase the perceived randomness of the TIOLI mechanism, and as we speculate in the text, reducing the perceived randomness of the BDM mechanism may narrow the gap.

Unrelated to explaining potential differences between BDM and TIOLI, we also conducted a "market study BDM" treatment in which we told respondents that we were using the information from the study to help decide on the future price of the filter in similar villages. If strategic bidding was important, then this sub-treatment could lead to enhanced strategic bidding and decrease BDM bids. However, we found that the market study treatment increased valuations, with marginal statistical significance.

J.5 Comparing Demand for Soap

As we argue in Section 6.2, the gap in elicited WTP between BDM and TIOLI also does not appear to be driven by lack of familiarity with the filter and uncertainty of its benefits. Although the sale of soap was primarily intended to be a practice round for the elicitation mechanism, the data provides suggestive evidence of the BDM-TIOLI gap for a more familiar product. Using these data, we find a similar difference in elicited WTP between the mechanisms: as shown in Table A13, BDM predicts between 8 and 45 percentage points lower purchase at the TIOLI price points for the soap.

J.6 Ex Post Regret

As discussed in Section 6.2, 19.2 percent of BDM respondents stated that they wished they had bid more. As shown in table A14, the proportion expressing regret is highest for those who narrowly missed winning in BDM: roughly 40 percent of those who missed by GHS 1 or less wished that they had bid more, with this percentage declining to approximately 12 percent among those who missed by GHS 5 to 10. To estimate the influence of regret on elicited WTP, we calculate what the adjustment to BDM bids would have been if all respondents who wished they had bid more had actually bid the value of the draw. Because those whose bid exceeded their draw cannot express regret, we apply an adjustment to this group that equals the average adjustment of BDM losers who have similar bids. Calculated in this manner, the average adjustment across all subjects equals GHS 0.6, or about 60 percent of the gap between BDM and TIOLI. Note that this likely represents an upper bound on underbidding due to regret because responses were not tied to an actual purchase decision. The share of respondents who actually offered to pay more than their final bid is substantially less than those who stated they wished they had more, at 5.4 percent. If we adjust the bids of those who offered to pay more up to the value of the draw (and apply an adjustment for those whose bid exceeded the draw following the procedure described above) this would result in an average increase in WTP of GHS 0.07, which would account for little of the BDM-TIOLI gap.

Although the upwards revision of bids after the price draw could result from respondents misunderstanding the BDM mechanism, it is also consistent with non-expected utility maximization in which a respondent revises her reference point upwards when the price is revealed. Further, a substantial share of TIOLI subjects, 17.0 percent, attempted to bargain with surveyors over the randomly drawn price. As noted in the text, we take this as evidence that both mechanisms may have seemed unusual to respondents who are unaccustomed to fixed prices.

K Using BDM in the Field

This section offers additional discussion of the practical tradeoffs between BDM and TIOLI for researchers considering using one or the other method. The key advantages of BDM are precision in measuring WTP, the ability to separately identify selection by WTP and the impacts of price paid, and the ability to estimate heterogeneous treatment effects with respect to WTP. The key disadvantage is complexity, which carries both fixed costs – time to tailor BDM to local context and train enumerators – and variable costs – time to explain BDM to subjects, conduct practice rounds, etc. Which method is preferable will depend on context and the questions the researcher is asking, but the relative advantages and disadvantages just mentioned offer some general guidelines.

First, the number of prices at which the researcher would like to measure demand affects the choice. The more prices of interest there are, the more advantageous BDM is likely to be, since more prices will require ever greater TIOLI sample sizes. Second, if the causal effect of price paid is of interest and a surprise randomized discount is not feasible, then BDM becomes attractive, since TIOLI cannot separately identify selection by WTP from the effect of price paid. Third, the extent to which it is plausible that treatment effects vary by WTP affects the choice. If there is strong prior evidence that treatment effects are constant, or constant with respect to WTP, then this tips the balance towards TIOLI. Fourth, developing a context-specific BDM protocol is a significant investment, and it is important to spend time explaining to and practicing with subjects. Based on our experience, multiple demonstration rounds with a different product or products, emphasis on the bid as the subject's optimal response, training the subjects to understand their bid as their maximum WTP, and the understanding check after respondents stated their bids are essential to successful implementation. These procedures have been emphasized elsewhere in the laboratory literature as important for eliciting accurate WTP through BDM (Plott and Zeiler 2005), although more research is needed on how each detail may contribute to subjects' understanding. Finally, the cost of each observation (including the cost of the item itself, the cost of collecting follow-up data on use or the outcome of interest, etc.) affects the tradeoff. If each observation is very cheap, then the burden of increased sample size from TIOLI is less of a concern. If each observation is relatively expensive, it becomes more important to obtain as much information as possible from each subject and the balance tilts towards BDM.

Table A1: Constant-Effects Instrumental Variables: Flexible Demand Curve
Dependent Variable: Child age 0 to 5 has had diarrhea over previous two weeks

	Combined all subjects		TIOLI subjects		BDM subjects	
	(1)	(2)	(3)	(4)	(5)	(6)
A. One-month followup						
Bought Filter	-0.057*	-0.066**	-0.083	-0.085*	-0.036	-0.048
	(0.033)	(0.032)	(0.051)	(0.049)	(0.047)	(0.040)
Mean dependent variable	0.145	0.145	0.149	0.149	0.142	0.142
First-stage F-statistic	441.0	228.4	111.2	96.1	504.0	338.8
Number of compounds	472	472	244	244	229	229
Number of subjects	786	786	418	418	368	368
Number of children	1244	1244	665	665	579	579
B. One-year followup						
Bought Filter	0.116*	0.138**	0.142	0.211**	0.115	0.127
	(0.067)	(0.067)	(0.100)	(0.100)	(0.089)	(0.085)
Mean dependent variable	0.241	0.241	0.215	0.215	0.262	0.262
First-stage F-statistic	132.3	80.2	58.8	36.0	179.3	170.0
Number of compounds	247	247	121	121	126	126
Number of subjects	387	387	197	197	190	190
Number of children	539	539	266	266	273	273
Controls	No	Yes	No	Yes	No	Yes
Village FEs	No	Yes	No	Yes	No	Yes

Notes: Each column displays the results of a linear two-stage least squares regression of child diarrhea status at the child level on filter purchase. For TIOLI subjects, filter purchase is instrumented dummies for each level of the randomly assigned TIOLI price (GHS 2, 4, 6). For BDM subjects, filter purchase is instrumented by a quadratic in the random BDM price draw. Controls include all variables (other than BDM bid) listed in Table 1. Missing values of control variables are set to 0, and dummy variables are included to indicate missing values. Standard errors clustered at the compound (extended family) level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A2: Relationship between Use and Willingness to Pay

	Filter present and unbroken (1)	Storage vessel contains water (2)	Clay pot contains water (3)	Usage index (4)
A. Short-term effects				
Bid (GHS)	-0.010	-0.008	-0.009	-0.022
	(0.010)	(0.012)	(0.013)	(0.021)
Mean dep. var.	0.877	0.753	0.728	-0.003
Adj. R-sqd.	0.002	-0.002	-0.002	0.002
Num. Obs.	235	235	235	235
B. One-year effects				
Bid (GHS)	0.013	0.027*	-0.013	0.018
	(0.014)	(0.014)	(0.012)	(0.021)
Mean dep. var.	0.641	0.486	0.380	0.066
Adj. R-sqd.	-0.002	0.016	-0.002	-0.003
Num. Obs.	142	142	142	142

Notes: The sample includes those subjects in the BDM treatment who purchased the filter, i.e., drew a price less than or equal to their bid. Each column presents the results of a separate regression of the depend variable, listed in the column heading, on the willingness to pay, i.e, the subject's bid in BDM. Usage index is the average of the normalized values of the three individual usage measures. Usage measures are observed by the enumerator at indicated follow-up survey. Standard errors clustered at the compound (extended family) level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A3: Attrition -- 1-month survey

	Baseline (1)	Surveyed (2)	Not Surveyed (3)	Difference (4)
Share of households		0.871	0.129	
Assigned to BDM treatment	0.480	0.479	0.485	0.006
	[0.500]	[0.500]	[0.501]	(0.043)
Number of respondents in compound	3.593	3.580	3.681	0.101
	[2.323]	[2.378]	[1.914]	(0.207)
Respondent's husband lives in compound	0.794	0.804	0.730	-0.074**
	[0.404]	[0.397]	[0.445]	(0.037)
One or more children age 0-5 in household	0.723	0.713	0.791	0.078**
	[0.448]	[0.452]	[0.408]	(0.035)
Number of children age 0-5 in household	1.569	1.561	1.620	0.059
(conditional on positive)	[0.801]	[0.760]	[1.017]	(0.111)
Num. children age 0-5 w. diarrhea in prev. 2 wks.	0.337	0.328	0.388	0.059
(among households with children age 0-5)	[0.592]	[0.597]	[0.563]	(0.055)
Num. children age 6-17 w. diarrhea in prev. 2 wks.	0.075	0.081	0.033	-0.047**
(among households with children age 0-5)	[0.335]	[0.349]	[0.181]	(0.023)
Respondent ever attended school	0.090	0.088	0.104	0.016
	[0.286]	[0.283]	[0.307]	(0.025)
Respondent's spouse ever attended school	0.233	0.225	0.300	0.075
	[0.423]	[0.418]	[0.464]	(0.078)
Wealth index	0.132	0.115	0.245	0.130
	[1.555]	[1.556]	[1.549]	(0.131)
Improved water source, all year	0.187	0.185	0.202	0.017
	[0.390]	[0.389]	[0.403]	(0.036)
Treats water with an effective method	0.115	0.107	0.166	0.059*
	[0.319]	[0.309]	[0.373]	(0.035)
Water quality: E. coli (MPN, z-score)	-0.052	-0.050	-0.067	-0.018
	[0.949]	[0.958]	[0.877]	(0.108)
Water quality: Turbidity (index, z-score)	-0.065	-0.080	0.072	0.151
	[0.997]	[0.985]	[1.100]	(0.141)
Bid for filter (GHS)	3.051	3.022	3.243	0.221
(among BDM respondents)	[2.268]	[2.247]	[2.414]	(0.276)
Filter draw (GHS)	4.650	4.621	4.842	0.221
(among BDM respondents)	[3.663]	[3.669]	[3.641]	(0.434)
Filter offer price (GHS)	3.824	3.864	3.548	-0.316
(among TIOLI respondents)	[1.616]	[1.604]	[1.682]	(0.199)

Notes: Standard deviations in brackets. Standard errors clustered at the compound (extended family) level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A4: Attrition -- 1-year survey

	Baseline (1)	Surveyed (2)	Not Surveyed (3)	Difference (4)
Share of households		0.904	0.096	
Assigned to BDM treatment	0.479	0.488	0.400	-0.088
	[0.500]	[0.500]	[0.494]	(0.084)
Number of respondents in compound	3.676	3.620	4.215	0.596
	[2.552]	[2.523]	[2.781]	(0.378)
Respondent's husband lives in compound	0.800	0.816	0.646	-0.170**
	[0.400]	[0.388]	[0.482]	(0.070)
One or more children age 0-5 in household	0.722	0.725	0.692	-0.033
	[0.448]	[0.447]	[0.465]	(0.062)
Number of children age 0-5 in household	1.556	1.578	1.333	-0.245**
(conditional on positive)	[0.763]	[0.771]	[0.640]	(0.114)
Num. children age 0-5 w. diarrhea in prev. 2 wks.	0.367	0.379	0.244	-0.134*
(among households with children age 0-5)	[0.606]	[0.620]	[0.435]	(0.072)
Num. children age 6-17 w. diarrhea in prev. 2 wks.	0.080	0.079	0.088	0.009
(among households with children age 0-5)	[0.288]	[0.288]	[0.288]	(0.052)
Respondent ever attended school	0.099	0.093	0.154	0.061
	[0.298]	[0.290]	[0.364]	(0.047)
Respondent's spouse ever attended school	0.303	0.311	0.200	-0.111
	[0.461]	[0.464]	[0.414]	(0.103)
Wealth index	0.030	0.070	-0.341	-0.411
	[1.574]	[1.553]	[1.726]	(0.269)
Improved water source, all year	0.216	0.220	0.185	-0.035
	[0.412]	[0.414]	[0.391]	(0.069)
Treats water with an effective method	0.107	0.106	0.123	0.017
	[0.310]	[0.308]	[0.331]	(0.043)
Water quality: E. coli (MPN, z-score)	-0.132	-0.113	-0.313	-0.200
	[0.928]	[0.951]	[0.664]	(0.131)
Water quality: Turbidity (index, z-score)	-0.348	-0.353	-0.298	0.055
	[0.532]	[0.536]	[0.495]	(0.088)
Bid for filter (GHS)	3.068	3.150	2.115	-1.035***
(among BDM respondents)	[2.383]	[2.428]	[1.519]	(0.369)
Filter draw (GHS)	4.606	4.632	4.308	-0.324
(among BDM respondents)	[3.585]	[3.614]	[3.290]	(0.631)
Filter offer price (GHS)	3.768	3.778	3.692	-0.085
(among TIOLI respondents)	[1.636]	[1.648]	[1.559]	(0.298)

Notes: Standard deviations in brackets. Standard errors clustered at the compound (extended family) level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A5: Casual Effect of Prices

	Filter Present and Undamaged (1)	Storage Vessel Contains Water (2)	Clay Pot Contains Water (3)	Usage Index (4)
A. Short-term effects				
Draw	0.017	0.037*	-0.003	0.043
	(0.018)	(0.022)	(0.024)	(0.038)
Mean Dependent Variable	0.877	0.753	0.728	-0.003
R-squared	0.020	0.033	0.010	0.025
Observations	235	235	235	235
B. One-year effects				
Draw	-0.013	0.021	0.019	0.018
	(0.034)	(0.033)	(0.033)	(0.051)
Mean Dependent Variable	0.641	0.486	0.380	0.066
R-squared	0.029	0.033	0.010	0.015
Observations	142	142	142	142

Notes: The sample includes those subjects in the BDM treatment who purchased the filter, i.e., drew a price less than or equal to their bid. Each column presents the results of a separate regression of the dependent variable, listed in the column heading, on BDM draw and BDM bid. See Section 5 for discussion of data. Usage index is the average of the normalized values of the three individual usage measures. Usage measures are observed by enumerator at indicated follow-up survey. Standard errors clustered at the compound (extended family) level in parentheses.

Table A6: Estimated Impacts of Pricing Policy

	Price (GHS)						
	0	1	2	3	4	5	6
Share Purchasing	1.00	0.94	0.73	0.46	0.31	0.19	0.11
A. Constant one-month effects							
Diarrhea cases averted per household (conditional on purchase)	1.43	1.43	1.43	1.43	1.43	1.43	1.43
Diarrhea cases averted per household (unconditional)	1.43	1.35	1.05	0.66	0.44	0.28	0.15
DALYs averted per household (conditional on purchase)	0.041	0.041	0.041	0.041	0.041	0.041	0.041
DALYs averted per household (unconditional)	0.041	0.038	0.030	0.019	0.013	0.008	0.004
Average social cost per DALY (USD)	369	369	369	369	369	369	369
Marginal cost per DALY (USD)		369	369	369	369	369	369
B. Average of one-month effects and one-year effects							
Diarrhea cases averted per household (conditional on purchase)	-1.09	-0.72	0.62	2.73	4.29	5.73	6.81
Diarrhea cases averted per household (unconditional)	-1.09	-0.68	0.46	1.26	1.33	1.10	0.73
DALYs averted per household (conditional on purchase)	-0.031	-0.021	0.018	0.077	0.121	0.162	0.193
DALYs averted per household (unconditional)	-0.031	-0.019	0.013	0.036	0.038	0.031	0.021
Average social cost per DALY (USD)	–	–	849	194	123	92	78
Marginal cost per DALY (USD)		–	–	–	361	128	79

Notes: In Panel A, short-term impacts on diarrhea are assumed to be constant and last for one year. Panel B assumes the average of short- and long-term impacts last for one year. In Panel B, the short-term impacts are constant and the long-term impacts are linear in willingness-to-pay. Diarrhea incidence is converted to DALYs at the rate of 35.3 cases per year to one DALY, using data from the Global Burden of Disease Collaborative Network (2017). The average social cost does not account for revenue generated from sales. The marginal cost per DALY is computed as the difference in costs between price $P - 0.5$ and price $P + 0.5$ divided by the difference in DALYs averted between price $P - 0.5$ and price $P + 0.5$. Missing entries in the average and marginal cost rows indicate that costs cannot not be computed because treatment effects are negative for average or marginal households at the prices indicated.

Table A7: Correlates of Willingness to Pay, LASSO Regularization

	Regularized Coefficient	
Variable	BDM	TIOLI
BDM soap bid	0.322	-
Purchased TIOLI soap	-	0.232
Village fixed effect, V08	-0.007	-0.164
Village fixed effect, V09	-	-0.050
Village fixed effect, V12	-	0.230
Highest education attained, kindergarten	0.075	-
Highest education attained, other	0.001	-
Spouse occupation, animal tending	0.023	-
Primary occupation, non-ag wage labor	-	-0.010
Primary occupation, household enterprise	-	0.001
Household has mobile phone	0.019	-
Number of phones in household	0.016	-
Number of cutlasses in household	-0.002	-
Household has chair	-0.010	-
Primary water source, dry season: dugout	-0.005	-
Primary water source, dry season: dam	-	0.009
Household treats water by boiling	0.004	-
Believes hygiene prevents diarrhea	-	0.062

Notes: Regularized coefficients reported for all features with non-zero coefficients in test sample using regularization parameter (λ) with minimum out-of-sample error rate in training sample. See Appendix I.1 for details.

Table A8: Demand Comparison by Purchase Mechanism
Take-it-or-leave-it Price Points

	Dependent Variable: WTP $\geq$ Price (GHS)					
	Price = 2		Price = 4		Price = 6	
	(1)	(2)	(3)	(4)	(5)	(6)
Difference (BDM - TIOLI)	-0.182***	-0.171***	-0.163***	-0.152***	-0.100**	-0.103***
	(0.033)	(0.033)	(0.052)	(0.052)	(0.040)	(0.039)
Mean TIOLI Purchase	0.915	0.915	0.473	0.473	0.207	0.207
Controls:	No	Yes	No	Yes	No	Yes
Number of BDM Respondents	607	607	607	607	607	607
Number of TIOLI Respondents	246	246	224	224	188	188
Number of clusters	395	395	390	390	376	376

Notes: BDM acceptance rate calculated based on share of respondents bidding greater than or equal to the evaluated price. TIOLI acceptance rate equal to share of respondents offered the evaluated price who agreed to purchase. Controls include all variables listed in Table 1 (except BDM bid). Missing values of the control variables are set to 0, and dummy variables are included to indicate missing values. Standard errors clustered at the compound (extended family) level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Joint p-value testing significance of BDM across all three equations (GHS 2, 4, 6): 0.000 without controls; 0.000 with controls. P-value testing equality of BDM across equations: 0.239 without controls; 0.354 with controls.

Table A9: Relationship between Household Observables and BDM-TIOLI Gap Risk Aversion Sample (One-Year Follow-Up Villages)

	Pairwise (1)	Joint (2)
Top vs. bottom tercile of risk aversion	−0.138*	−0.150*
	(0.081)	(0.083)
First principal component of durables ownership	0.030	0.030
	(0.028)	(0.030)
Respondent has ever attended school	0.141	0.180*
	(0.113)	(0.107)
Has child age 0-5	−0.002	0.004
	(0.095)	(0.100)
Husband lives in compound	0.066	0.037
	(0.094)	(0.098)
All-year access to improved water source	−0.150	−0.104
	(0.106)	(0.097)
Currently treats water	−0.077	0.011
	(0.129)	(0.155)
Ambiguity aversion category (more is more AA)	0.003	0.003
	(0.009)	(0.010)
Total digit span score forward + backward	−0.028	−0.035*
	(0.018)	(0.019)
Number of compounds	233	233
Number of households	399	399

Notes: This table presents estimates of the interaction between the mean BDM-TIOLI gap (the probability of purchase at 2, 4 or 6 GHS) and the household observable indicated. Column (1) shows the results of pairwise comparisons: an indicator for whether the household would agree to purchase the filter at the given price on an indicator for the BDM treatment, a level term for the indicated covariate, and the interaction between the two. Column (2) shows the estimated interaction terms in a joint regression. Coefficients are estimated for offer prices of 2, 4 and 6 and then averaged across the three prices, with standard errors calculated by SUR. The sample consists of households surveyed in the one-year followup (conducted in 8 of the 15 study villages) in the top or bottom tercile of risk aversion. Standard errors clustered at the compound (extended family) level in parentheses.

Table A10: Relationship between Household Observables and BDM-TIOLI Gap
All Villages; Household Survey Measures Only

	Pairwise (1)	Joint (2)
First principal component of durables ownership	0.007	0.001
	(0.016)	(0.016)
Respondent has ever attended school	0.137**	0.128**
	(0.061)	(0.059)
Has child age 0-5	0.138***	0.129***
	(0.045)	(0.045)
Husband lives in compound	0.050	0.036
	(0.053)	(0.053)
All-year access to improved water source	0.056	0.072
	(0.060)	(0.059)
Currently treats water	0.022	0.028
	(0.075)	(0.071)
Number of compounds	556	556
Number of households	1265	1265

Notes: This table presents estimates of the interaction between the mean BDM-TIOLI gap (the probability of purchase at 2, 4 or 6 GHS) and the household observable indicated. Column (1) shows the results of pairwise comparisons: an indicator for whether the household would agree to purchase the filter at the given price on an indicator for the BDM treatment, a level term for the indicated covariate, and the interaction between the two. Column (2) shows the estimated interaction terms in a joint regression. Coefficients are estimated for offer prices of 2, 4 and 6 and then averaged across the three prices, with standard errors calculated by SUR. Standard errors clustered at the compound (extended family) level in parentheses.

**Table A11: Equality of Bid Distributions
Comparison with Standard BDM**

	Market (1)	Anchor (2)
A. Wilcoxon		
Z-statistic	2.754	−0.900
P-value	0.022	0.748
Num. Obs.	411	408
B. Kolmogorov-Smirnov		
D-statistic	0.144	0.058
P-value	0.050	0.777
Num. Obs.	411	408

Notes: This table reports results of nonparametric tests for equality of bid distributions across BDM treatments. The anchoring and marketing BDM treatments (describe in the text) are separately compared to the standard BDM treatment. P-values robust to clustering at the compound level are calculated via randomization inference.

Table A12: Differences Between TIOLI sub-treatments

	Price=2	Price=4	Price=6
	(1)	(2)	(3)
Random TIOLI	0.018	-0.133	0.022
	(0.063)	(0.104)	(0.084)
Anchoring TIOLI	0.066	-0.233**	-0.087
	(0.059)	(0.114)	(0.075)
Constant	0.890	0.600	0.232
	(0.053)	(0.083)	(0.056)
Mean Dependent Variable	0.915	0.473	0.207
R-squared	0.009	0.036	0.014
Observations	246	224	188

Notes: This table reports results of a linear probability model for purchase of the filter at the TIOLI price indicated in the column header. The ommitted category is standard TIOLI. The p-values for joint tests across equations are calculated from SUR estimation. P-value for joint test that coefficient on Random TIOLI=0 in all three equations: 0.587. P-value for joint test that coefficient on Anchoring TIOLI=0 in all three equations: 0.077. Standard errors clustered at the compound (extended family) level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A13: Soap Demand Comparison by Purchase Mechanism
Soap Take-it-or-leave-it Price Points

	Dependent Variable: WTP $\geq$ Price (GHS)					
	Price = 0.3		Price = 0.5		Price = 0.7	
	(1)	(2)	(3)	(4)	(5)	(6)
Difference (BDM - TIOLI)	-0.077***	-0.067**	-0.169***	-0.171***	-0.454***	-0.449***
	(0.028)	(0.029)	(0.053)	(0.054)	(0.058)	(0.055)
Mean TIOLI Purchase	0.947	0.947	0.757	0.757	0.663	0.663
Controls:	No	Yes	No	Yes	No	Yes
Number of BDM Respondents	607	607	607	607	607	607
Number of TIOLI Respondents	189	189	148	148	172	172
Number of clusters	364	364	359	359	356	356

Notes: BDM acceptance rate calculated based on share of respondents bidding greater than or equal to the evaluated price. TIOLI acceptance rate equal to share of respondents offered the evaluated price who agreed to purchase. Controls include all variables listed in Table 1. Missing values of the control variables are set to 0, and dummy variables are included to indicate missing values. Standard errors clustered at the compound (extended family) level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Joint p-value testing significance of BDM across all three equations (GHS 0.3, 0.5, 0.7): 0.000 without controls; 0.000 with controls. P-value testing equality of BDM across equations: 0.000 without controls; 0.000 with controls.

Table A14: BDM: Respondents Interested in Changing Their Bid Ex Post

Difference between draw and bid	Number whose draw exceeded bid (1)	Frac. who wished they bid more (2)	Frac. who tried to pay more (3)
Difference<1	20	0.45	0.30
1≤Difference<2	45	0.33	0.13
2≤Difference<3	44	0.25	0.07
3≤Difference<4	32	0.19	0.06
4≤Difference<5	33	0.12	0.03
Difference>5	159	0.12	0.00
Total	333	0.19	0.05

Notes: Column 1 presents the number of subjects in the BDM treatment whose draw exceeded their bid. Column 2 presents the fraction of those subjects who answered "Yes" to the question "Do you wish you had bid higher?" Column 3 presents the fraction of those subjects who attempted to pay more than their bid after the draw was realized.

Figure A1: The *Kosim* filter

Figure A2: Experimental Timeline for a Typical Village

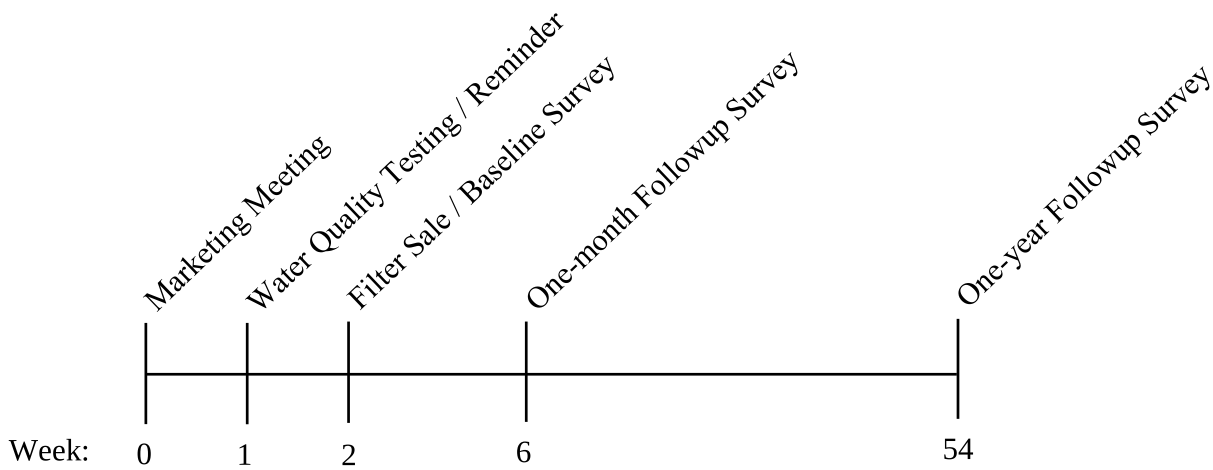


Figure A3: Participant Flow Diagram

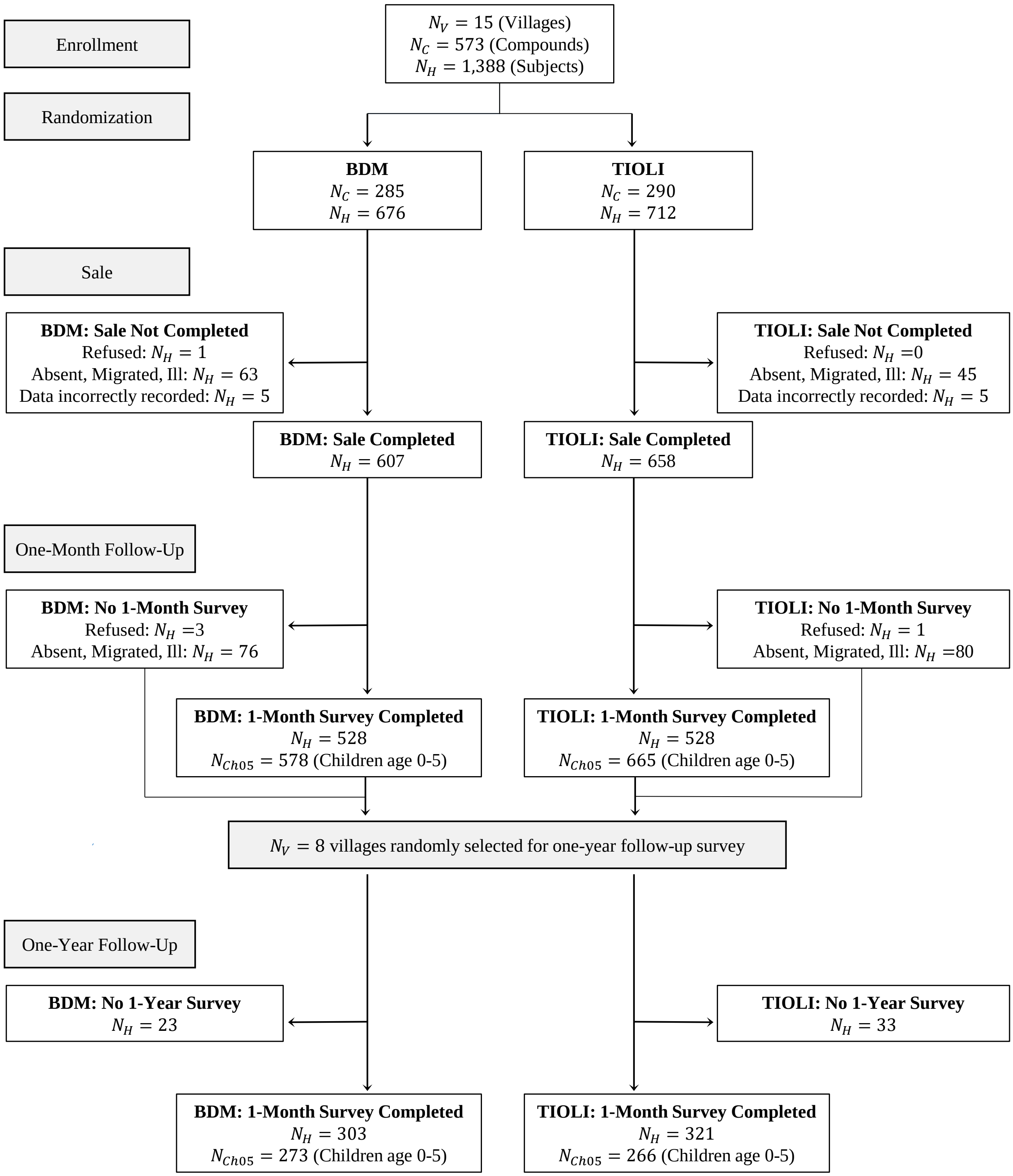
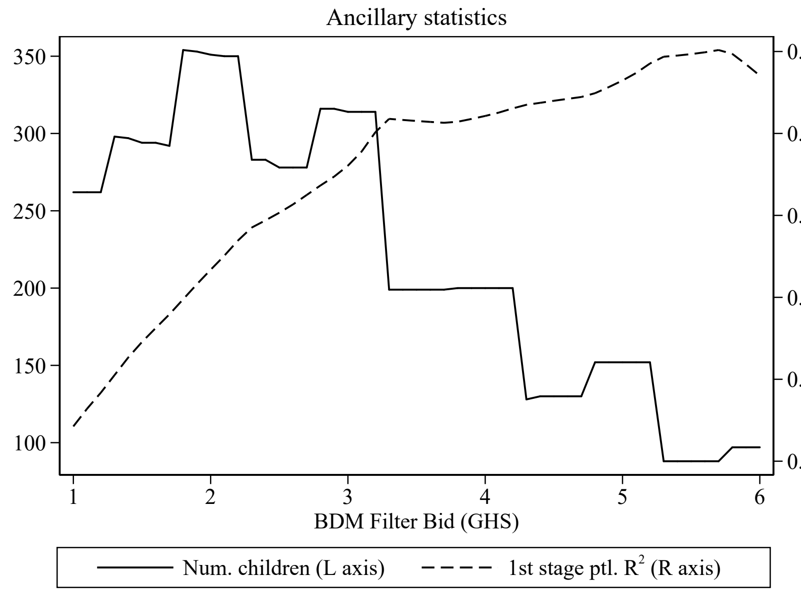
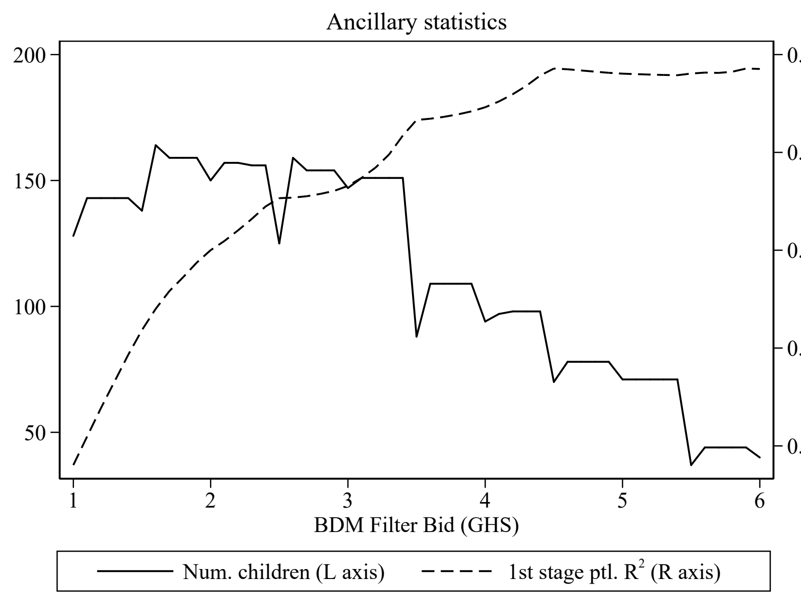


Figure A4: Kernel IV Estimates of Treatment Effects
Ancillary Statistics

(a) One-Month Follow-Up

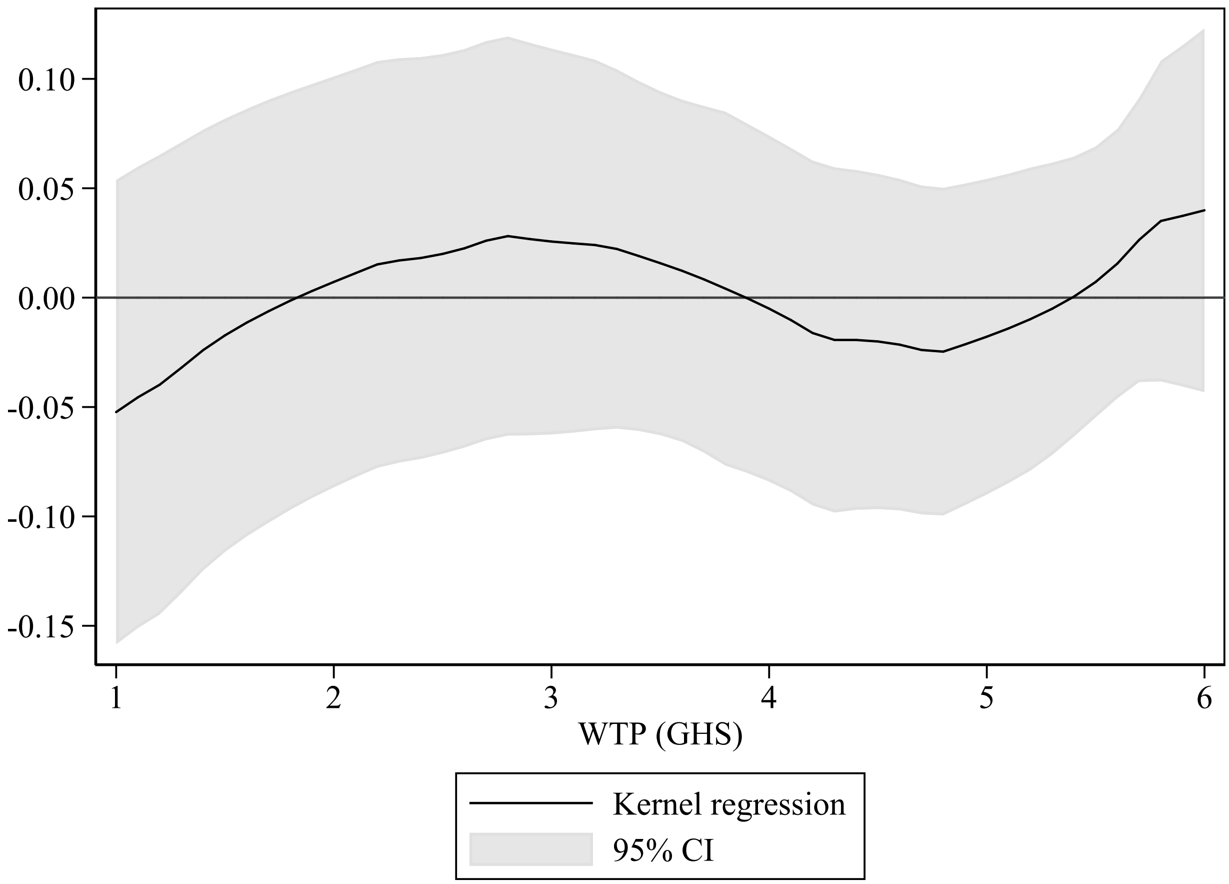


(b) One-Year Follow-Up



Notes: The solid line (left axis) plots the sample size, i.e., the number of children receiving positive weight in the kernel regression, at each evaluation point WTP $= 1.0, 1.1, \ldots, 6.0$ (GHS). The dashed line (right axis) plots Shea's partial R-squared for the excluded instrument (the BDM price draw) in the first-stage regression at each evaluation point.

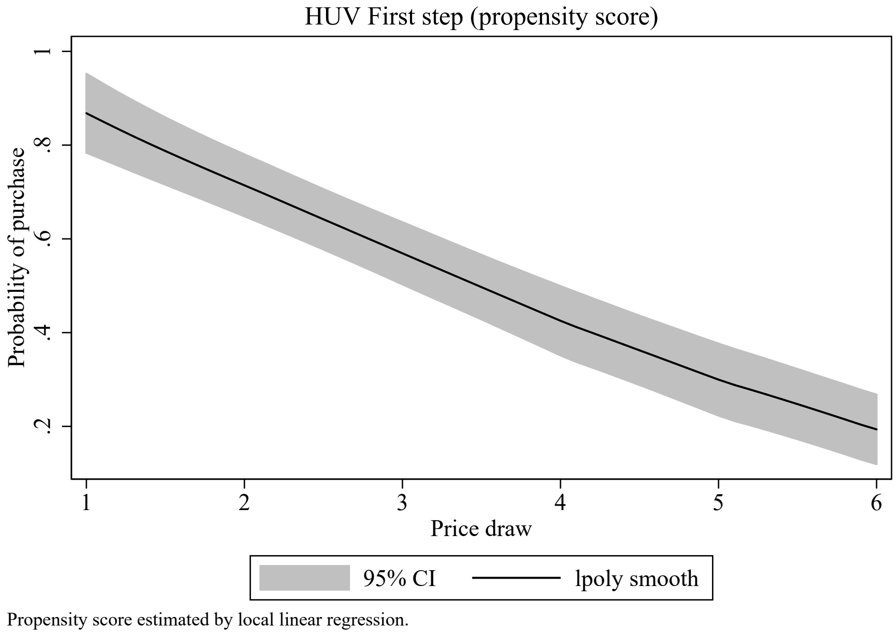
Figure A5: Difference in attrition rates: BDM Winners vs. Losers 1-year follow-up survey; BDM participants with children 0 to 5 years old



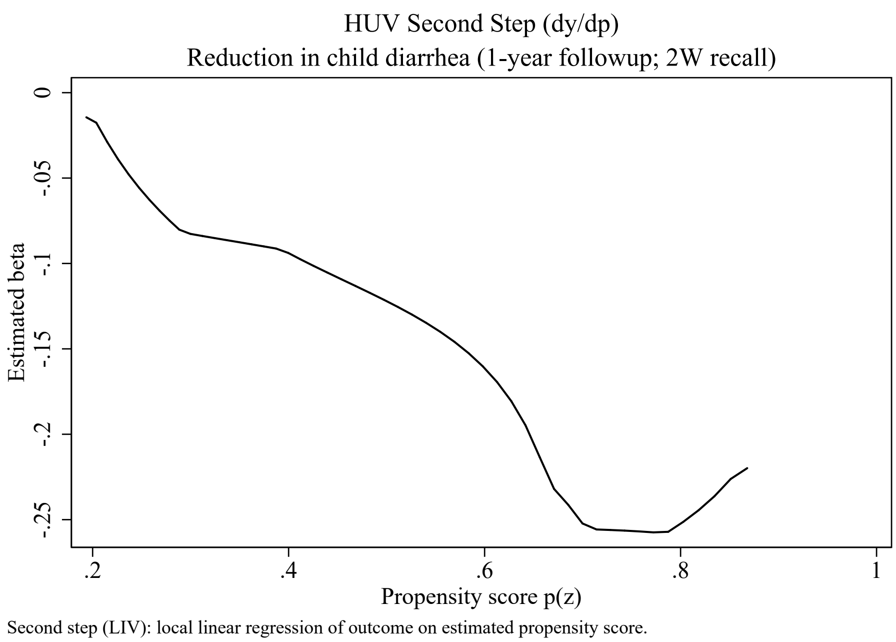
Notes: This figure plots estimated differences, with 95 percent confidence bands, in the rate of attrition from the one-year follow-up survey between BDM subjects who won the filter and subjects who did not win. The line plots estimates from kernel regressions of attrition on winning the filter, using Epanechnikov kernel with Silverman's rule-of-thumb bandwidth. Standard errors are robust to clustering at the compound (extended family) level.

Figure A6: Local Instrumental Variables and Marginal Treatment Effects

(a) Step 1: Propensity Score



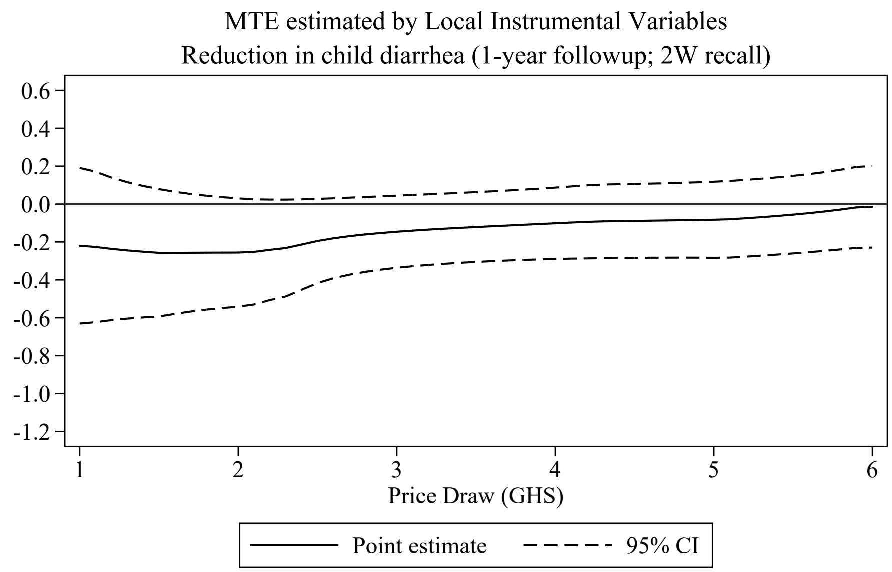
(b) Step 2: Local Instrumental Variables



(Continued next page)

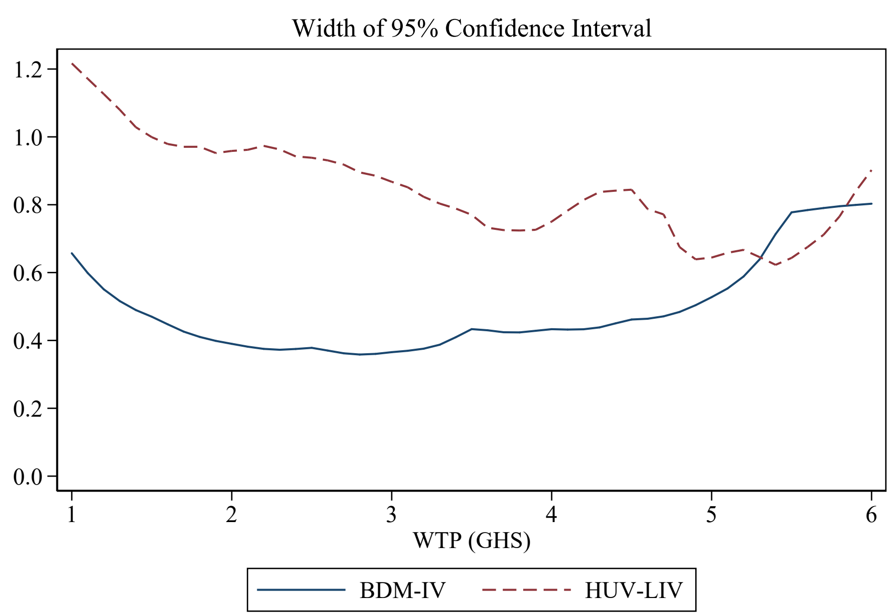
Figure A6: Local Instrumental Variables (Continued)

(c) Marginal Treatment Effect



First step (propensity score): local linear regression of purchase (T) on draw (Z).
Second step (LIV): local linear regression of outcome on estimated propensity score.
Clustered standard errors in second stage (no bootstrapping).

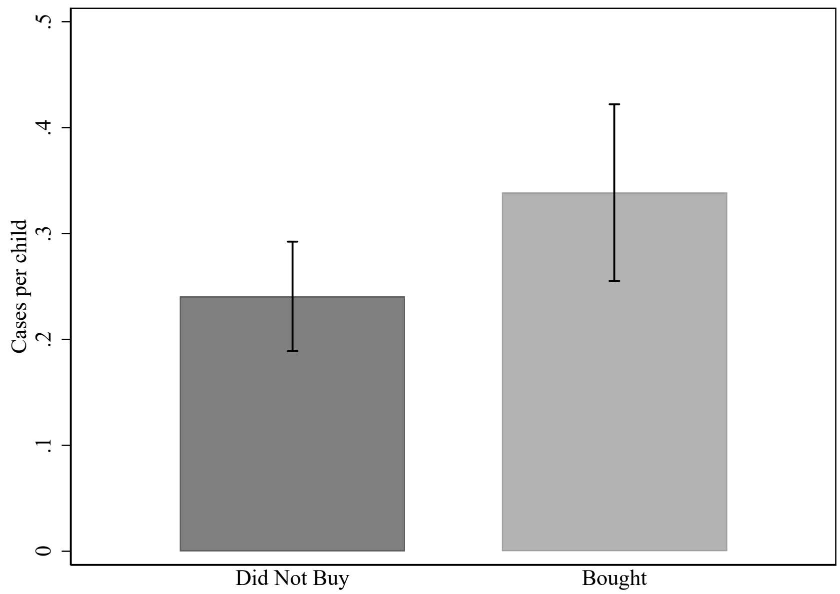
(d) Comparison of BDM-IV and LIV



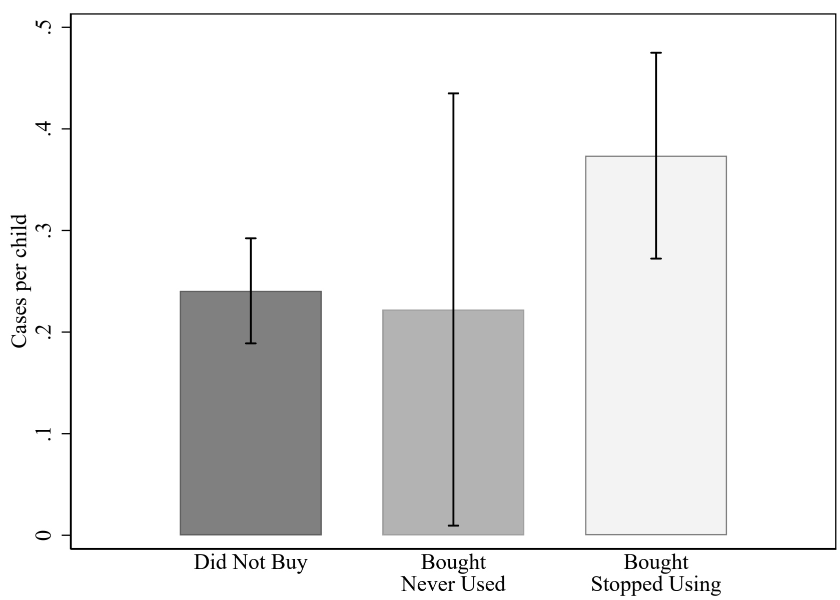
HUV: Confidence interval from cluster bootstrap.

Figure A7: Health Outcomes for Long-Term Non-Users
All Subjects with Children 0 to 5

(a) Reported Cases (1 Year)

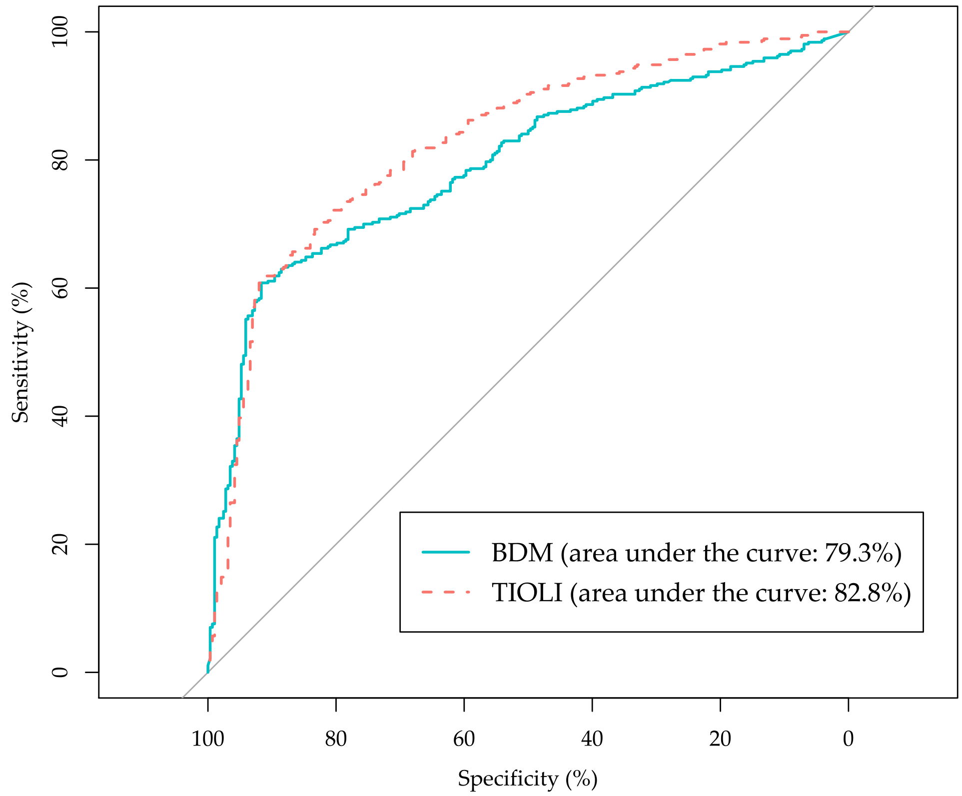


(b) Reported Cases (1 Year), By use History



Notes: This figure displays incidence of diarrhea (lower is better) in the prior two weeks for children aged five or under among all households that are not using the filter at the one-year follow up. The second figure separates purchasers into those who were not using the filter after one month and those who were using the filter but stopped using at the one-year follow up. For all households, use is defined as an indicator for the filter being present, operational, and either containing water in the clay pot or storage vessel. Whiskers represent 95 percent confidence intervals.

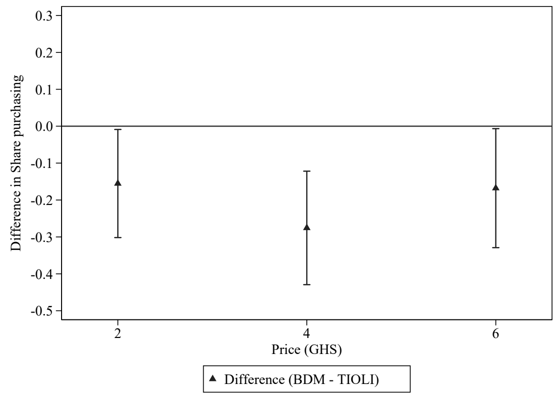
Figure A8: ROC Comparison of BDM vs. TIOLI for Predicting TIOLI Purchase Behavior



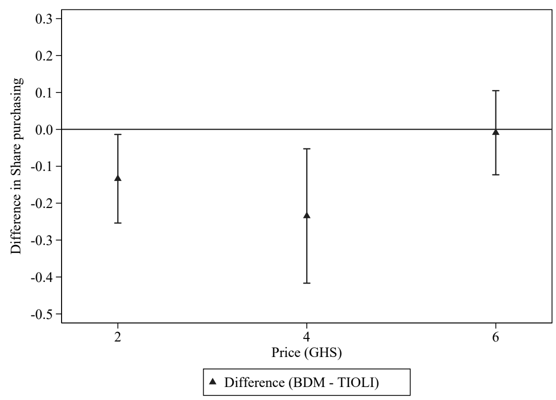
Notes: The target outcome is the TIOLI purchase decision (yes/no) in cross-validation sample. The ROC curves plot the sensitivity of each predictive model (the rate of true positives) vs. the specificity (the rate of true negatives) as we vary the threshold for predicting purchase. The 45-degree line represents the performance a model that randomly classified each observation.

Figure A9: BDM–TIOLI gap by tercile of risk aversion

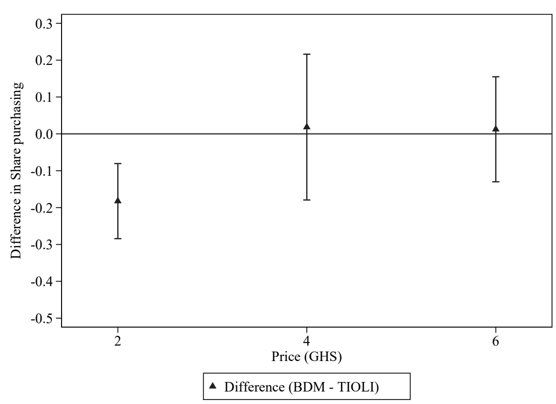
(a) Tercile 1 (most risk-averse)



(b) Tercile 2



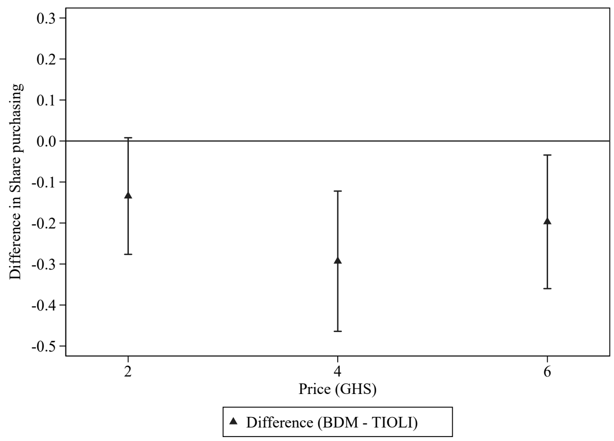
(c) Tercile 3 (least risk-averse)



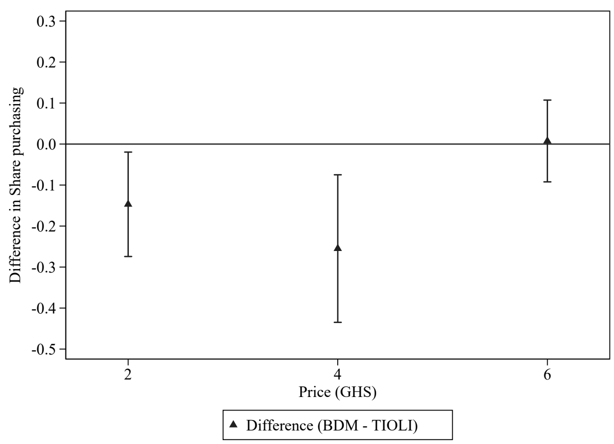
Notes: These figures plot estimated differences, with with 90 percent confidence intervals, between the share of BDM subjects and the share of TIOLI subjects agreeing to purchase at each TIOLI price (GHS 2, 4, 6), separately by tercile of risk aversion. The results here are unconditional, see Figure A10 for robustness checks with additional controls.

Figure A10: BDM–TIOLI gap by tercile of risk aversion
Robustness check: with controls, including ambiguity aversion

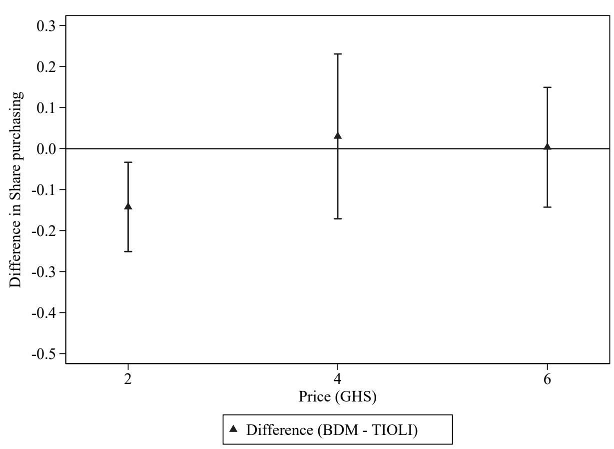
(a) Tercile 1 (most risk-averse)



(b) Tercile 2



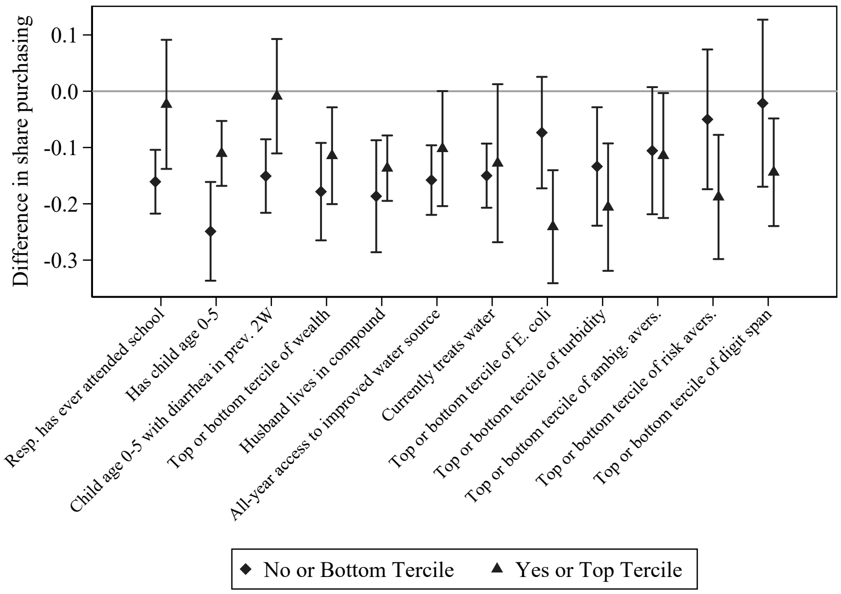
(c) Tercile 3 (least risk-averse)



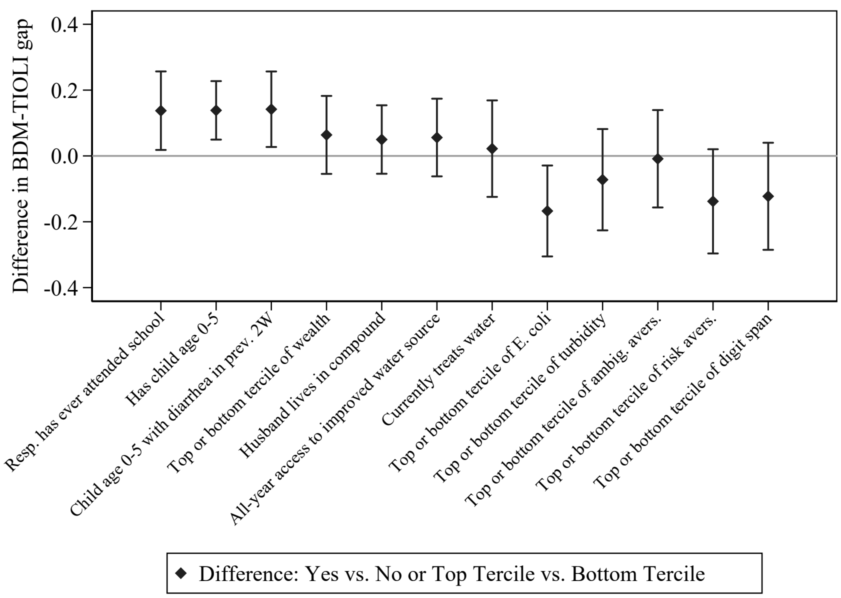
Notes: These figures plot the difference between the share of BDM subjects and the share of TIOLI subjects agreeing to purchase at each TIOLI price (GHS 2, 4, 6), separately by tercile of risk aversion. The regression includes the standard set of household controls and our measure of ambiguity aversion, described in Appendix B.

Figure A11: Heterogeneity in BDM–TIOLI gap across relevant observables

(a) Levels

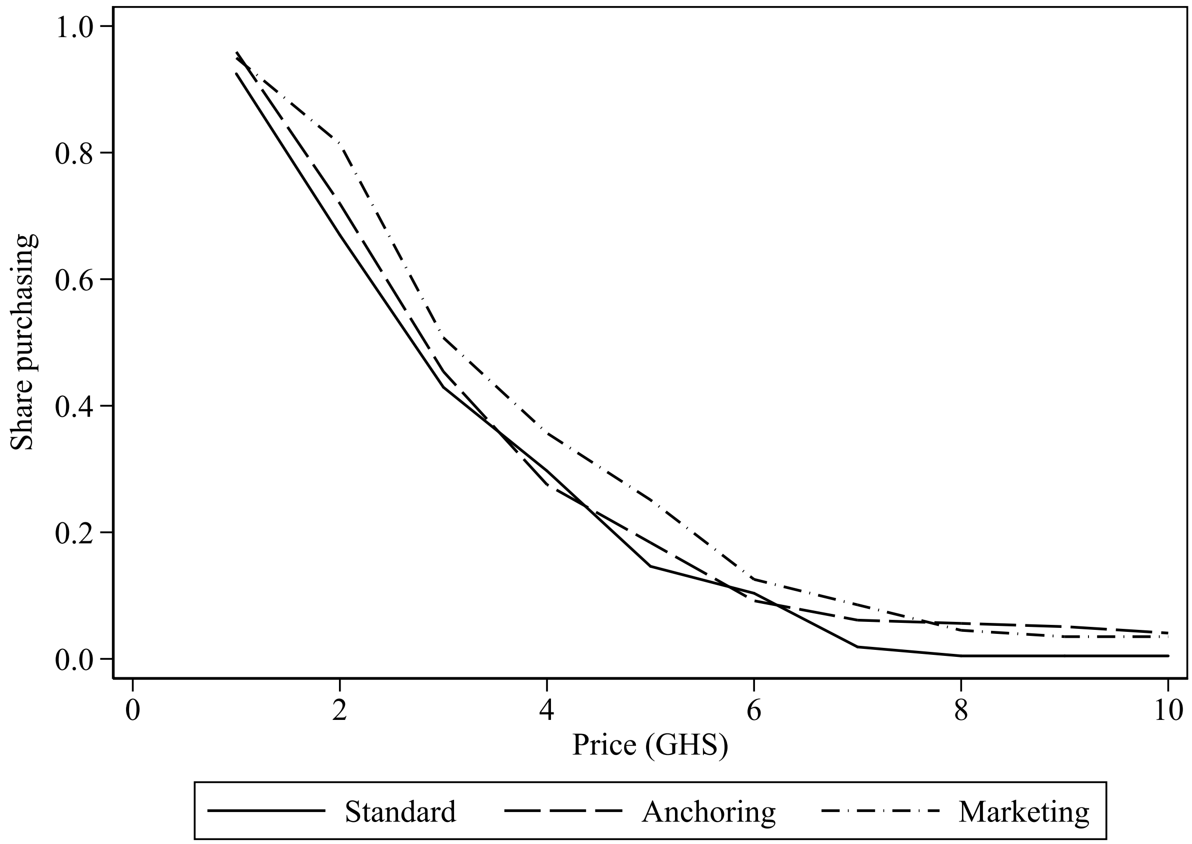


(b) Differences



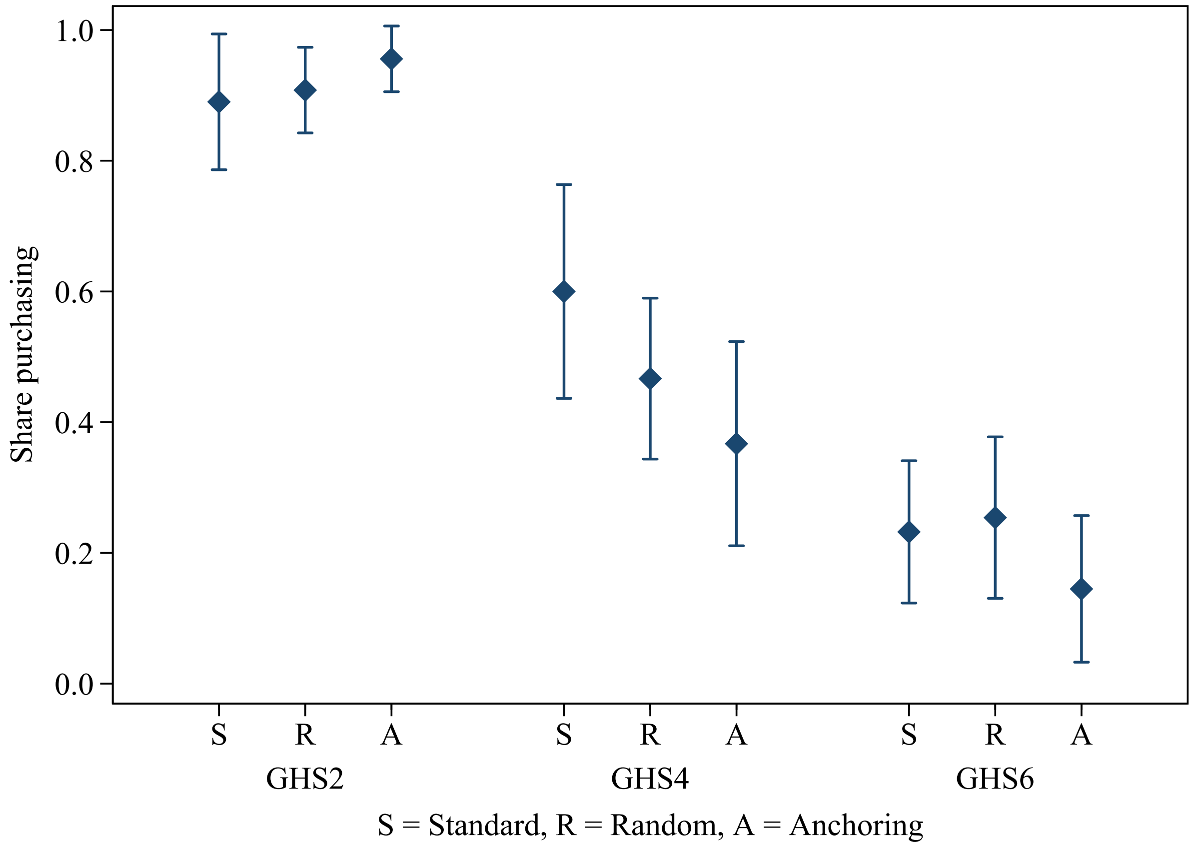
Notes: These figures compare the average BDM-TIOLI gap (percentage point difference in shares purchasing, averaged over the three TIOLI prices) in different sub-groups. For binary observables, we compare the two levels of the variable. For continuous observables, we divide the sample into terciles and compare the top and bottom terciles. The top panel shows the level of the gap for the two sub-groups; a more negative number indicates a larger BDM gap. The bottom panel shows the difference in the gap between the two groups; a positive number that the BDM-TIOLI gap is narrower among the "Yes" or "Top Tercile" subgroup than among the "No" or "Bottom Tercile" subgroup. "Child age 0-5 with diarrhea in prev. 2W" is limited to respondents with one or more children age 0-5. For "Top or bottom tercile of E. coli" and "Top or bottom tercile of turbidity" the top tercile category refers to the highest levels of E. coli and highest levels of turbidity, respectively, i.e., the poorest quality water.

Figure A12: Comparison of BDM Sub-treatments



Notes: The standard, anchoring and marketing treatments are described in detail in the text. 607 observations total, of which 212 are standard BDM, 199 are marketing BDM and 196 are anchoring BDM. All treatments were randomized at the compound (extended family) level.

Figure A13: Comparison of TIOLI Sub-treatments



Notes: This graph plots demand for the filter at each take-it-or-leave-it price, for each TIOLI sub-treatment. The random, anchoring and standard sub-treatments are described in detail in the text. Each treatment was randomized at the compound level. For the standard and anchoring TIOLI treatments, the price was also randomized at the compound level. For the random TIOLI treatment, the price was drawn by individual respondents. 658 observations, of which: standard 217 (GHS2 91, GHS4 70, GHS6 56); random 225 (GHS2 87, GHS4 75, GHS6 63); anchoring 216 (GHS2 68, GHS4 79, GHS6 69).