

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

James Berry, Greg Fischer and Raymond Guiteras*

March 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 Vytlačil 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-

¹Deviations 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.

²Screening 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).

compensating 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

³See, 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.

⁴The 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.

⁵Section 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 (Appendix 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 Appendix 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

⁶These 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-

⁷Within 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

⁸Most 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.

⁹Prices 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 We 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.

¹⁰With 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 12 percent of households regularly treat their water using an effective method such as boiling (11 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 ($p < 0.01$).

¹¹Due 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).

¹²At 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 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

$$y_{jic} = \beta_0 + \beta_1 T_{ic} + \varepsilon_{jic}, \quad (1)$$

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

$$T_{ic} = \gamma_0 + \gamma_1 P_{ic} + v_{ic}, \quad (2)$$

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),

¹⁴The 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 Appendix Table A1.

¹⁵We provide a more complete treatment in Appendix D.

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

$$y = \beta_0 + \beta_1(w) T + \varepsilon. \quad (3)$$

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

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

$$y = \beta_0 + \bar{\beta}_1 T + u, \quad (4)$$

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

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

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 $E[T\tilde{\beta}_1(w)] \neq 0$.

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

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

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 $E[Z\tilde{\beta}_1(w) T] \neq 0$.

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 $P_L \leq WTP \leq P_H$, who would buy a filter at P_L but not at P_H . Formally, IV using Z estimates

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

the average $\beta_1(w)$ between P_L and P_H weighted by $F_{WTP}(\cdot)$. 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 $F_{WTP}(\cdot)$.

BDM provides a simple method to estimate $\beta_1(w)$.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 $\beta_1(w)$ 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

¹⁶Remarkably, the *local instrumental variables* (LIV) method of Heckman et al. (2006) can estimate $\beta_1(w)$ 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.

¹⁷Using 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 $w = \{1, 1.1, \dots, 6\}$. 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, $p = 0.024$).

¹⁸Non-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

¹⁹See 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 F. 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 F, 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). Appendix 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 G, 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

²⁰An 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

²¹See Appendix H 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 F, 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).

²²Average 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.

²³Assigning 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

²⁴Subsequent to our study, Cole et al. (2018) 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

$$\text{WTP}_{ic} = \alpha_0 + X'_{ic}\beta + \varepsilon_{ic}, \quad (7)$$

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:

$$\text{buy}_{i,p} = 1 \{ \text{WTP}_i \geq p_i \} = 1 \{ \text{WTP}_i - p_i \geq 0 \} = 1 \{ \alpha_0 + X'_{ic}\beta + \varepsilon_{ic} - p_i \geq 0 \} \quad (8)$$

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

²⁵Bivariate 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 76.1 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 percent of our subjects exhibit loss aversion, 41.6 percent exhibit some degree of ambiguity aversion, and 64.6 percent 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

²⁶See 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 Appendix 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

²⁷In 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.

analysis in Appendix J.4.

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

²⁸Within 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).

systematically lower than TIOLI at each of the TIOLI price points, particularly among the 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," 2018. Manuscript, University of Delaware.
- , 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. <http://ghdx.healthdata.org/gbd-results-tool>.
- 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. Ruhr Economic Papers 745.
- Gross, E., I. Guenther, and Y. Schipper, "Water Supply and Water Handling: Perceived Substitutes but Revealed Supplements for Safe Water," 2017. Manuscript, University of Bayreuth.
- 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. Vytlačil, "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. Vytlačil, "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," 2016. NBER Working Paper 22367.
- 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," 2015. NBER Working Paper 21391.
- 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. Manuscript, University of California, Berkeley.
- 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 (1)	BDM (2)	TIOLI (3)	BDM-TIOLI (4)	BDM Draw (5)	TIOLI Price (6)
Number of respondents in compound (census)	3.593 [2.323]	3.305 [1.816]	3.859 [2.683]	-0.554* (0.323)	0.236*** (0.079)	-0.051 (0.045)
Husband lives in compound	0.794 [0.404]	0.792 [0.406]	0.796 [0.403]	-0.004 (0.022)	0.453 (0.367)	-0.243 (0.168)
Number of children age 0-5 in household	1.135 [0.978]	1.069 [0.941]	1.196 [1.008]	-0.127* (0.073)	0.195 (0.159)	0.028 (0.078)
Number of children age 6-17 in household	1.303 [1.282]	1.389 [1.304]	1.224 [1.258]	0.165** (0.084)	0.028 (0.129)	-0.013 (0.047)
Number of children age 0-5 with diarrhea in past two weeks	0.243 [0.525]	0.208 [0.487]	0.277 [0.557]	-0.069* (0.035)	-0.372 (0.376)	0.075 (0.128)
Number of children age 6-17 with diarrhea in past two weeks	0.049 [0.272]	0.050 [0.302]	0.048 [0.241]	0.002 (0.016)	-0.499 (0.417)	0.463* (0.267)
Respondent has ever attended school	0.090 [0.286]	0.079 [0.270]	0.100 [0.301]	-0.021 (0.016)	-0.025 (0.515)	-0.077 (0.195)
First principal component of durables ownership	0.132 [1.555]	0.059 [1.512]	0.198 [1.592]	-0.139 (0.126)	-0.046 (0.091)	0.005 (0.056)
All-year access to improved water source	0.187 [0.390]	0.196 [0.397]	0.179 [0.384]	0.017 (0.038)	-0.126 (0.376)	0.119 (0.252)
Currently treats water	0.115 [0.319]	0.109 [0.312]	0.120 [0.325]	-0.011 (0.024)	0.567 (0.468)	0.048 (0.257)
E. coli count, standardized	-0.052 [0.949]	-0.026 [1.012]	-0.076 [0.887]	0.050 (0.089)	-0.102 (0.162)	0.038 (0.120)
Turbidity, standardized	-0.065 [0.997]	-0.099 [0.922]	-0.032 [1.063]	-0.068 (0.096)	-0.008 (0.178)	0.224*** (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 between the 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)
<i>A. One-month followup</i>						
<i>A.1 Structural Equation – Dependent variable: Child age 0 to 5 has had diarrhea over previous two weeks</i>						
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
<i>A.2 First Stage – Dependent variable: Household Purchased Filter</i>						
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
<i>B. One-year followup</i>						
<i>B.1 Structural Equation – Dependent variable: Child age 0 to 5 has had diarrhea over previous two weeks</i>						
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
<i>B.2 First Stage – Dependent variable: Household Purchased Filter</i>						
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. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

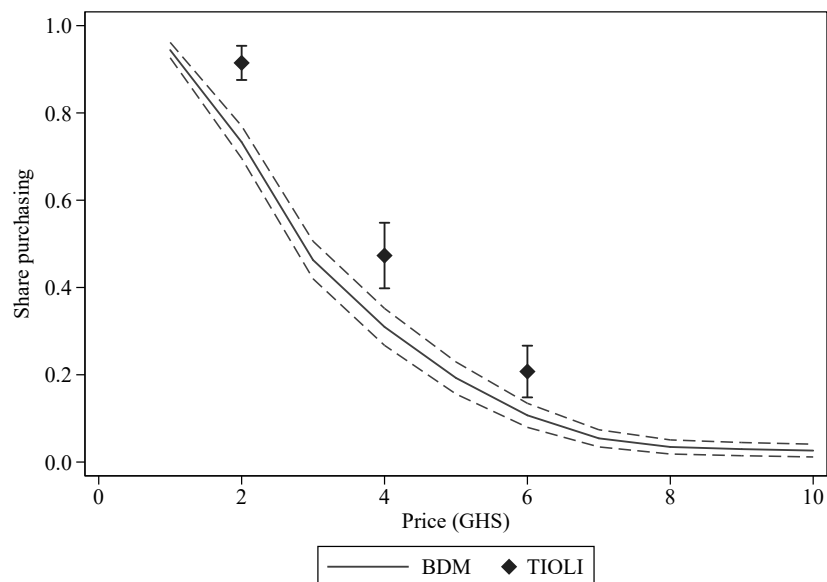
Table 3: Correlates of Willingness to Pay

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

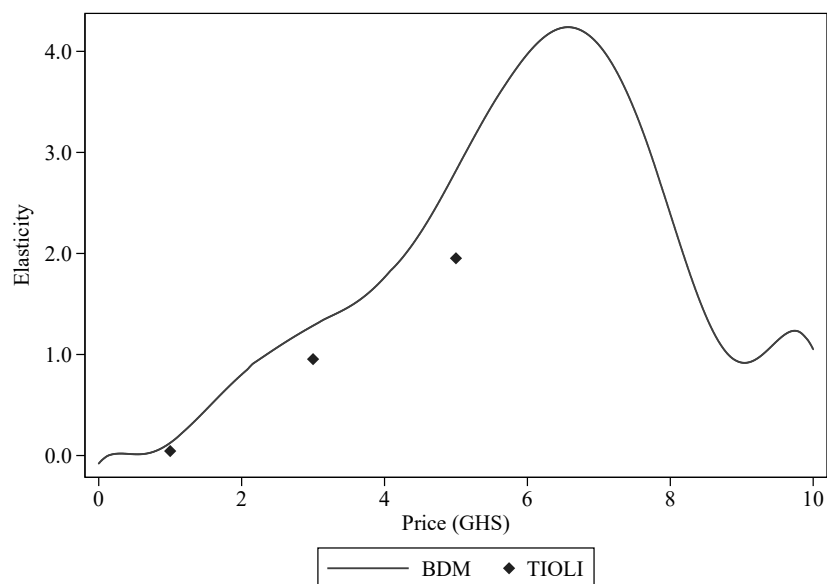
Notes: Columns (1) and (2) display coefficients from a linear regression of directly reported willingness to pay (the BDM bid) on baseline characteristics. Columns (3) and (4) report coefficients from probit models, where the dependent variable is the TIOLI purchase decision. As discussed in the text, 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



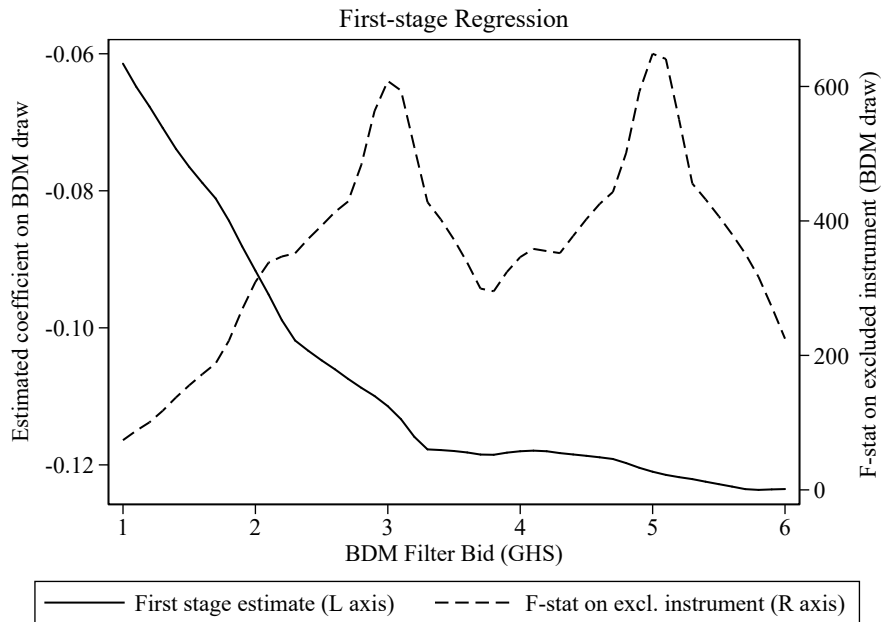
(b) Price Elasticity of Demand



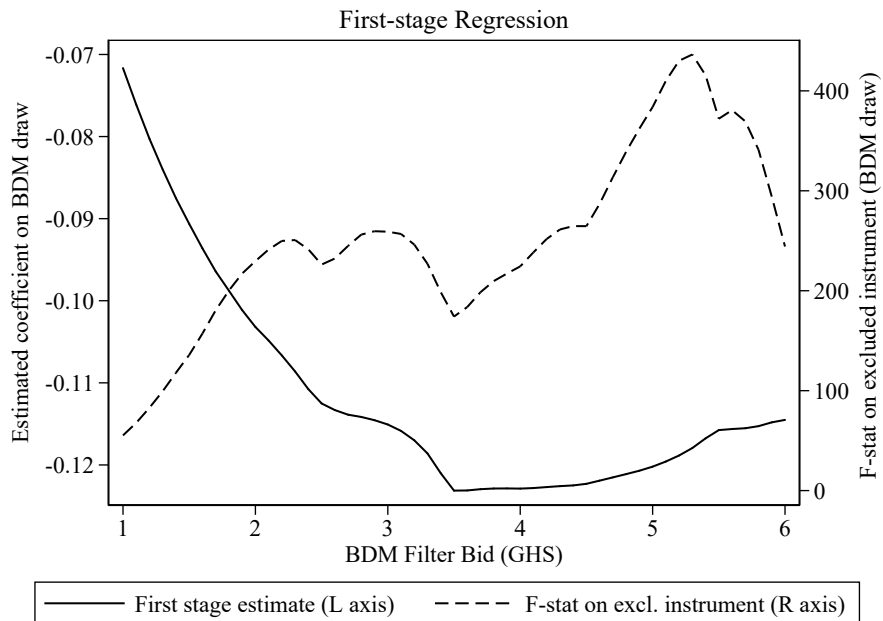
Notes: The top panel plots the BDM demand curve, with a 90% confidence band, and take-it-or-leave-it (TIOLI) demand at three price points (2, 4 and 6 GHS), with 90% confidence intervals. The BDM demand curve indicates the share of respondents with a BDM filter bid greater than or equal to the indicated 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 assumed to be 1.

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

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



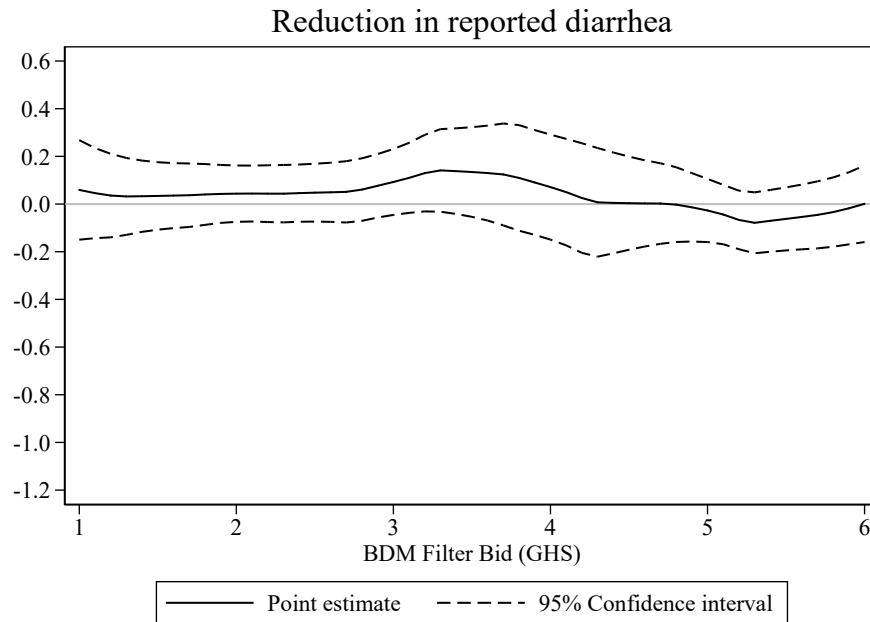
(b) Long-term: One-Year Follow-Up



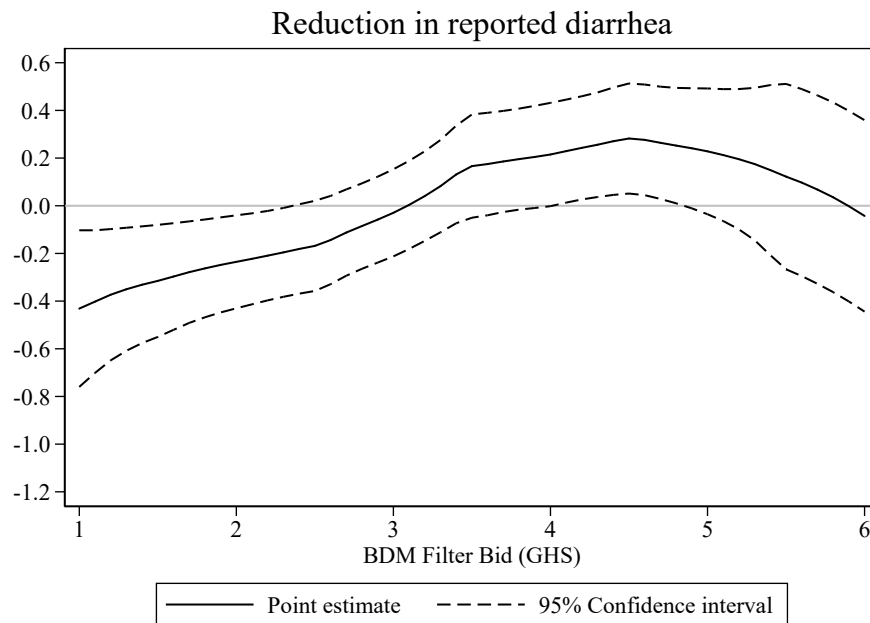
Notes: The solid line (left axis) plots the estimated coefficient on the price draw from the first-stage regression at each evaluation point $WTP = 1.0, 1.1, \dots, 6.0$ (GHS). The dashed line (right axis) plots the F-statistic on the excluded instrument (the BDM price draw) in the first-stage regression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

Figure 3: Kernel IV Estimates of Treatment Effects

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



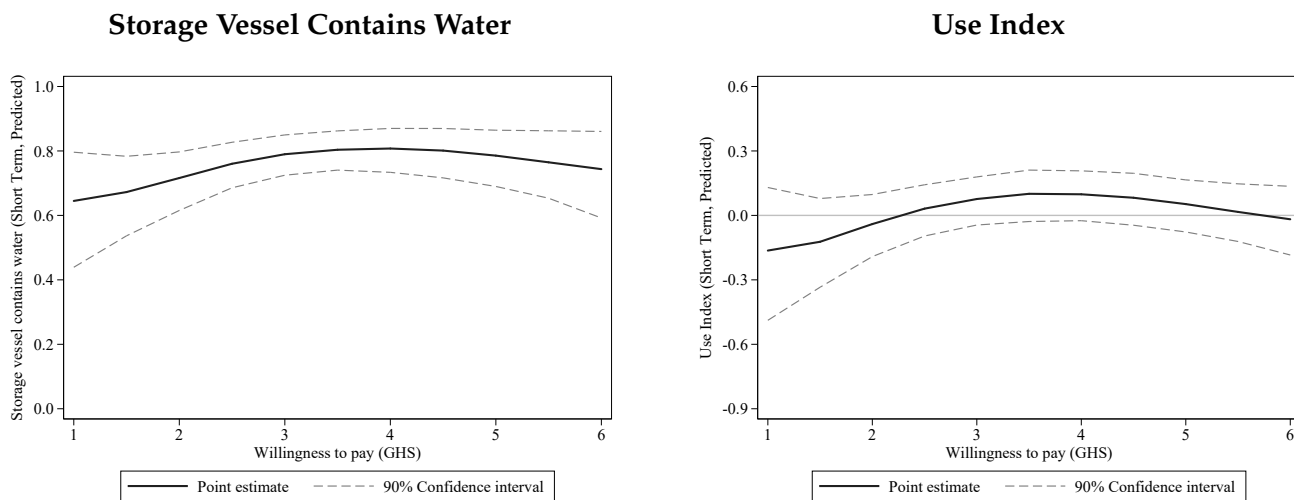
(b) Long-term: One-Year Follow-Up



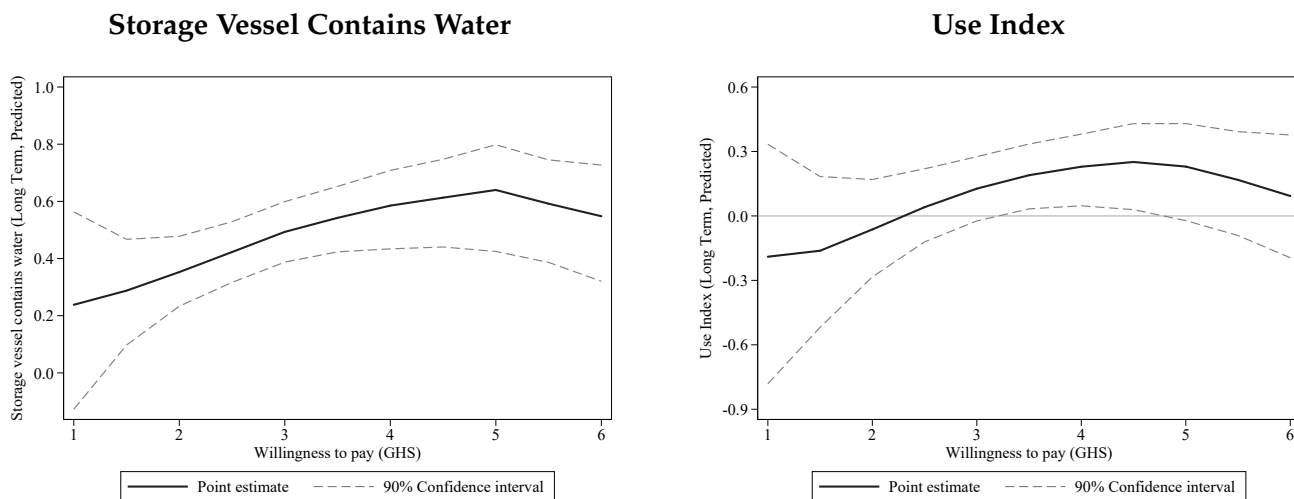
Notes: These graphs present estimated treatment effects (reduction in diarrhea among children age 0 to 5) as a function of willingness-to-pay (WTP). Estimates are by linear two-stage least squares at $WTP = 1.0, 1.1, \dots, 6.0$, weighting observations by their distance from the evaluation point (Epanechnikov 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 Figure 2 and Appendix Figure 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



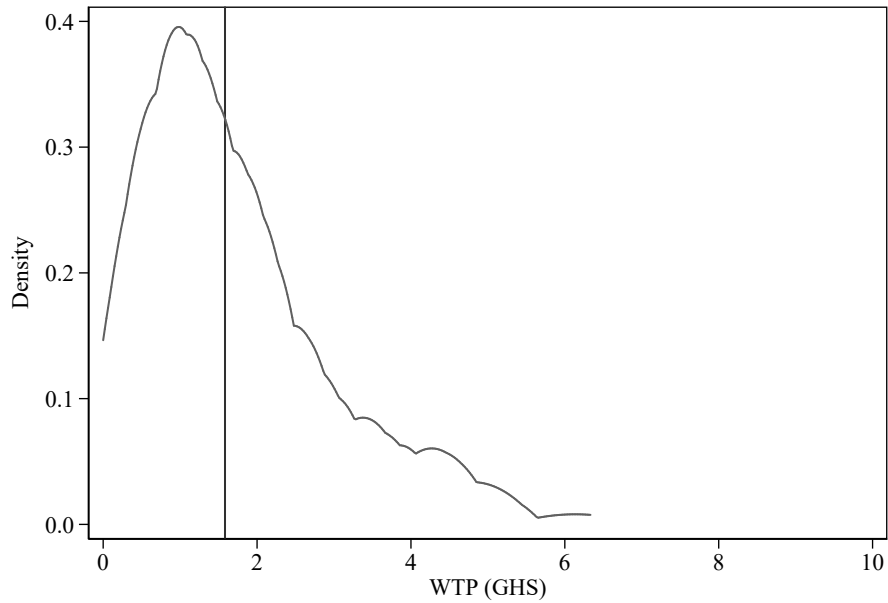
(b) One-Year Follow Up



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).

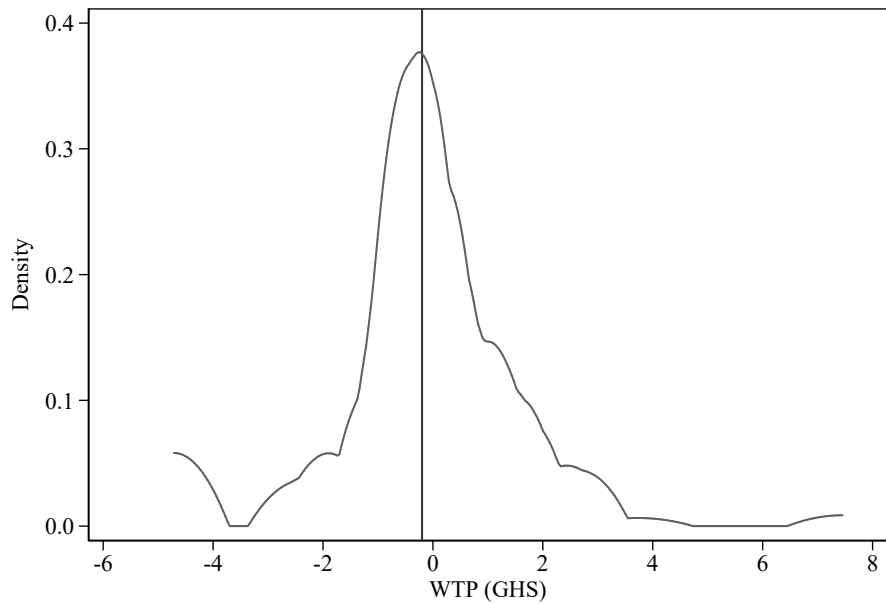
Figure 5: WTP to Avoid a Case of Children's Diarrhea

(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.