Improved Source, Improved Quality? Demand for Drinking Water Quality in Rural India

Katrina Jessoe*

March 1, 2010

Abstract

To reduce waterborne disease from unsafe drinking water supplies, the Government of India expanded protected drinking water sources throughout its rural areas. An increase in the supply of improved drinking water sources may reduce private expenditure on water quality enhancing behaviors, and subsequently offset some of the quality benefits from source protection. Using a combination of data from a 1998 survey in rural India and field work conducted in 2007, I detect if households reduce demand for in-home treatment with source protection, and then quantify in water quality and Rupees the effect of this offsetting behavior. I estimate demand for in-home treatment using a variety of discrete choice models. Hydrological data that exogenously measure the price and supply of improved sources identify a household's drinking water source. Source protection reduces the probability of in-home treatment between 27 and 39 percentage points, with households shifting from boiling, a costly and effective treatment technology, to no treatment. This reduction in treatment offsets coliform abatement from improved sources by 5 percent and reveals that annual willingness to pay for source protection is 1% of total expenditure. Lastly, I estimate the effect of source protection on childhood diarrhea. While exogenous improvements in water quality reduce child diarrhea, unobservables that are correlated with source protection negate this benefit.

*Assistant Professor, Dept. of Ag and Resource Economics, UC Davis. Email: kkjessoe@ucdavis.edu; telephone: (530)-752-6977. I am extremely grateful to Sheila Olmstead for her support and advice. I would also like to thank Michael Boozer, Robert Mendelsohn, Erin Mansur and Robert Repetto for their time and comments. Achyuta Adhvaryu, Reena Badiani, Rahul Deb, Joshua Lustig, Maher Said, Steven Wallander, Juliana Wang and participants at the Yale Environmental Economics Seminar provided many helpful suggestions. Kapil Narula, T.S. Panwar and members of the water division at The Energy and Resources Institute offered invaluable suggestions regarding the collection and analysis of field data in India. Vidur Chopra, Priyanaka Chatterjee and Amrita Singh provided excellent assistance in field data collection. I thank Mr. Swaroop for assistance in accessing and interpreting groundwater data. This has been funded in part by the United States Environmental Protection Agency (EPA) under the Science to Achieve Results (STAR) Graduate Fellowship Program (FP 91690401), the Heinz Fellowship for Environmental Scholars and the V. Kann Rasmussen Foundation grant. All remaining errors are my own.

1 Introduction

At least 1 million child deaths occur annually in India from waterborne disease. To combat preventable waterborne disease in rural India, UNICEF and the Government of India launched in 1969 what evolved into the largest rural water supply program in the world (Black and Talbot 2005). This initiative increased the availability of safe water sources in under-supplied communities by investing heavily in improved drinking water sources - defined as taps, tube wells and hand pumps - which enable households to access groundwater protected from surface contamination, thereby lowering exposure to fecal coliform and other waterborne pathogens.[1] However, in some developing country contexts improvements in source quality may have produced negligible reductions in diarrheal disease (Bennett 2008, Jalan and Ravallion 2003).

Limited health benefits may occur because improvements in source quality do not generate substantial water quality gains at the point of consumption. Increased source water quality may induce households to reduce private expenditure on in-home abatement, offsetting the quality improvements from source protection. This paper detects if households reduce private expenditure on drinking water quality in response to source protection and then quantifies in coliform abatement (counts per 100 ml) and Rupees the effects of this offsetting behavior. Lastly, I estimate the reduction in diarrheal disease from an exogenous improvement in drinking water quality and compare this to the reduction in diarrheal disease when households choose their primary drinking water source.

While the public health literature has established that improvements in source quality significantly reduce exposure to fecal coliform, it continues to debate whether source protection causes measurable reductions in diarrhea and other waterborne diseases. Recently, a meta-analysis assessing the impacts of water, sanitation, and hygiene on diarrhea in developing countries, echoed previous meta-analyses (Esrey and Habicht 1986, Esrey et al. 1991) in finding that safe drinking water supplies reduce diarrheal incidence (Fewtrell et al. 2005). Similarly, in the U.S. the proliferation of piped drinking water supplies in urban areas, in the late 1800s and early 1900s, caused a rapid reduction in mortality rates (Cutler and Miller 2005). By contrast, another strand of literature finds that a safe water supply results in minimal health improvements (Brick et al. 2004, Checkley et al. 2004). These studies conclude that health benefits from access to clean drinking water only occur when bundled with sanitation, hygiene, or education programs (Brick et al. 2004, Checkley et al. 2004).

To better understand the mechanisms underlying the relationship between source protection and childhood diarrhea, economists have begun exploring the role of information, education and income, as well as behavioral responses to source protection (Bennett 2008, Jalan and Ravallion 2003, Kremer et al. 2009). A study estimating the effects of randomized spring source protection in Kenya found a 62 percent increase in drinking water quality, increased usage of protected (as compared to

[1] The World Health Organization (WHO) describes a sustainable and safe drinking water source as an improved source. The Government of India defines improved sources as taps, tube wells and hand pumps (Planning Commission 2002).

unprotected) springs and a 33 percent reduction in childhood diarrhea after the addition of spring protection (Kremer et al. 2009). However, other studies find limited or no health benefits from source protection. In India, piped water supplies only benefit high income households while in the urban Philippines substitution between piped drinking water supplies and sanitation eliminates the health gains from piped water (Bennett 2008, Jalan and Ravallion 2003). In rural India, the links between source protection, drinking water quality and childhood diarrhea remain unclear. Using a combination of survey data from 1998 and field data collected in 2007, I investigate whether trade-offs between improvements in source water quality and in-home treatment limit the quality benefits from source protection, and as a result compromise some of the health benefits of source protection.[2]

I create a simple theoretical model of household demand for in-home treatment, my measure of private expenditure on drinking water quality. The model demonstrates that improvements in source quality induce households to shift to less expensive modes of in-home treatment or no treatment. Additionally, this reduction in in-home treatment provides a lower bound estimate of willingness to pay for source protection.

To empirically test the relationship between in-home treatment and source quality improvements, I estimate the effect of improved drinking water sources on demand for in-home treatment. Demand for in-home treatment is estimated using a variety of discrete choice models with logit and multinomial logit models serving as the primary models. These specifications will generate biased (attenuated or overestimated) coefficient estimates of a household's primary drinking water source. Unobservables captured in the error term such as health endowments, village amenities, or preferences for water quality will be correlated with a household's primary drinking water source. Spatial groundwater data collected during field work in India are used to identify a household's primary drinking water source. District hydrological characteristics such as rock type, discharge rate, and well depth provide information on the price of improved sources. As shown in Figure 1, geological data spatially characterize the feasible locations of dug wells, tube wells, and bore wells. I estimate demand for in-home treatment using two stage least squares, where in the first stage a household's primary drinking water source is predicted using groundwater characteristics.[3] To test the sensitivity of my results to the restrictions imposed by a mulitnomial logit model, I estimate demand for in-home treatment using nested logit, ordered probit and multinomial probit models.

Results from all models support the hypothesis that demand for in-home treatment declines with an improvement in source water quality. In the logit model, the probability of any mode of in-home treatment decreases by 5 to 7 percentage points for households with an improved source; this result is robust to a variety of fixed effects models and to the exclusion of in-home piped users.

[2] The data exploit household, village, tehsil, sub-district, district, sub-state and state variation. A tehsil represents a unit of government that consists of a collection of villages and cities, with the U.S. equivalent of a civil township. A sub-district describes a region within a district (e.g. Western Jhansi). A district is the U.S. equivalent of a county. A collection of districts compose a sub-state (e.g. Northeast Uttar Pradesh), and 4-6 sub-states make up a state. There are a total of 28 states according to the 2001 Census.

[3] Technically, since demand for in-home treatment is estimated using discrete choice models, I use a control function estimator to predict the effect of an improved source on demand for in-home treatment.

Once I control for the endogeneity of source, households with an improved source are 27 to 39 percent less likely to engage in in-home treatment.

I also consider demand for each mode of in-home treatment. Modes of in-home treatment vary in the water quality provided by each technology, the market cost to purchase the technology, and the time cost to filter drinking water with the technology. In the 2SLS, I find that source protection reduces demand for time intensive treatment technologies and induces a larger reduction in the more expensive technology. Demand for cloth filters reduces by 11 percentage points and in some cases is not impacted by source protection, whereas boiling treatment reduces by 22 to 27 percentage points. Boiling describes a relatively high cost technology when compared to cloth filters; boiling also eliminates coliforms from drinking water, whereas cloth filters reduce concentrations, but not to zero.

Data collected during field work in India measure, in coliform counts per 100 ml, source water quality and abatement provided by in-home treatment technologies. Source water quality data are not representative of rural India, but information on source quality and the effectiveness of in-home treatment technologies provide a starting point to estimate the quality gains from source protection. Results suggest that the trade-off between source protection and in-home treatment offsets the coliform abatement from improved sources by 5 percent. Nonetheless, improvements in source quality lead to a 53 percent reduction in bacteriological contaminants. Despite these gains, my results suggest that on average coliform counts in drinking water amount to 125 counts per 100 ml which exceeds the 10 counts per 100 ml standard imposed by the Government of India.

To measure household willingness to pay for improved sources, I monetize the change in demand for in-home treatment from source protection using three distinct measures of price - the village market price of fuel wood, the district agricultural wage rate and the market cost of risk averting technologies in peri-urban Delhi. To value environmental quality, revealed preference studies often interchange market price and the opportunity cost of time; in providing valuation estimates using these two price measures, I can evaluate the degree of substitutability between them. In my sample, households are willing to pay 1 percent of total expenditure for source protection, with wealthier and more educated households displaying a higher willingness to pay for source protection; these results are consistent with previous valuation studies in India (Jalan et al. 2009). I generate similar results using market prices and the opportunity cost of time.

Lastly, I use data from the REDS 1998 survey to estimate the reduction in diarrheal disease from source protection. In a village fixed effects model, I find that source protection only reduces the incidence of childhood diarrhea for households with an in-home improved source. By contrast in a 2SLS model where I instrument for a household's primary drinking water source using hydrological characteristics, I find that improved drinking water sources significantly reduce the probability of childhood diarrhea. Together these results suggest that while exogenous improvements in drinking water quality lower the incidence of diarrhea, household unobservables such as hygiene or in-home treatment that are negatively correlated with improved sources offset these quality benefits.

2 Related literature

My research relates and contributes to the literature on offsetting behavior from regulatory innovation, risk averting behavior and willingness to pay for drinking water quality; this section provides a brief overview of these literatures.

An extensive literature has examined how technological and regulatory innovations designed to improve welfare may be partially compromised by offsetting behavior. The compensating behavior model first introduced in the auto safety literature argues that the protection gained from mandatory seat belt laws lowered the marginal cost of an accident, inducing individuals to drive faster and limiting the reduction in deadly car accidents associated with stricter seat belt laws (Lave 1985, Peltzman 1975). In the transportation literature, fuel efficiency technologies designed to reduce emissions lower the per mile cost of travel, and drivers compensate by increasing annual mileage (Small and Van Dender 2005). The compensating behavior hypothesis has also been applied to explore the impact of sexual education programs on teen pregnancy rates; sexual education programs appear to decrease the costs of sexual activity (lowering pregnancy probability), thereby increasing the volume of sexual activity and ultimately neutralizing programs' impact on teenage pregnancy rates (Oettinger 1999). In my research, I explore the extent to which improvements in source water quality lead to reductions in private expenditure on drinking water quality.

While no previous studies have measured the effects of source protection on demand for in-home treatment in India, a sizable literature has explored demand for in-home treatment. A recent experimental study in urban India informed households if their drinking water source was clean or dirty; upon learning that their drinking water was contaminated, in-home treatment increased by 13 percentage points (Jalan and Somanathan 2008). The authors suggest that ignorance about the relationship between drinking water quality and health may deter demand for in-home treatment. Along with information and awareness, previous studies have shown that education, wealth and cost all influence demand for in-home treatment (Jalan et al. 2009, Jalan and Somanathan 2008, Mintz et al. 2001, Quick et al. 1999).

In developing countries, since drinking water most often describes a non-market good, economists have relied on revealed and stated preference techniques to elicit willingness to pay for water quality. The contingent valuation literature consistently demonstrates that households in developing countries actually pay or are willing to pay high prices for improved water quality and water service (Briscoe et al. 1990, Johnson and Baltodano 2004, McPhail 1994, Whittington et al. 1991). Revealed preference studies echo these findings, though estimates using revealed preferences techniques tend to report a lower willingness to pay. A recent study in urban India used expenditure on in-home treatment to reveal the effect of education and income on willingness to pay for drinking water quality, reporting a mean annual expenditure of 137 Rs per person per year or 0.3 percent of annual expenditure on in-home treatment (Jalan et al. 2009). However, this study did not estimate willingness to pay for source protection. With the exception of a travel cost study in India that generates exceptionally high willingness to pay estimates (half a day's wage), I present the first measure of willingness to pay for source protection in rural India (Asthana 1997).

3 Drinking water in rural India

The Government of India defines safe or improved sources to include taps, tube wells and hand pumps (Planning Commission 2002).[4] Taps, tube wells and hand pumps enable households to access groundwater that is protected from surface contamination, thus lowering exposure to fecal coliform and other water-borne pathogens.[5] By contrast, unimproved sources such as surface water or open dug wells are exposed to the surface and susceptible to pathogen contamination from free-flowing sewage (Black and Talbot 2005). The increased likelihood of disease transmission in unimproved sources occurs because fecal matter deposited on the open ground leaches into the water-ways and contaminates the surface water (Helmer 1999). The relative absence of sanitation, defined as latrines, bathrooms, waste disposal, and waste removal, in rural India increases the probability of pathogen contamination in surface and exposed drinking water sources (NSSO 1998).

The expansion of groundwater sources via the drilling of tube wells shifted primary drinking water sources from surface water to protected groundwater, and in general provided households with a safer (in terms of microbes) supply of drinking water. In addition to providing improved sources, the government also established microbial standards for drinking water. A source is considered safe for human consumption if there are less than 10 coliform counts per 100 ml of drinking water (Dept. Drinking Water Supply 2007).[6] Despite these standards, there is little monitoring and enforcement of regulations and both taps and tube wells, as well as unimproved sources have been shown to contain high levels of zooplankton, coliform and other bacteria (Islam et al. 2007, Planning Commission 2002). Some deep ground water sources expose households to the separate health risk of unsafe concentrations of arsenic (Chaterjee et al. 1995, Madajewicz et al. 2007). Although arsenic contamination poses a critical health concern, my study restricts its attention to the acute health outcomes associated with bacterial contamination. If India's groundwater expansion increased the risk of exposure to other contaminants, this, along with compensating behavior, will offset the water quality gains from source protection.

While an improved source presents a safer (as compared to an unimproved source) supply of drinking water, significant contamination occurs during the transport and storage of the water supply.[7] A recent meta-analysis of field studies measuring bacterial contamination at the source

[4]According to the World Health Organization (WHO) an improved source must adhere to the following criteria: (i) a significant increase in the probability that the water is safe; (ii) a more accessible source; and (iii) sufficient measures taken to protect the water source from contamination. In contrast to India, WHO extends the scope of improved sources to include protected dug wells, protected springs and rainwater collection (WHO 2000).

[5]The Census of India (2001) defines tube well water as sub-soil water that is removed via electricity or diesel pump. The term hand pump describes a device that is attached to a non-electric tube well and used to manually extract drinking water (Census of India 2001, WHO 2004). Tap characterizes water that is supplied to households through a pipe after suitable treatment, as required by a corporation, a municipality or another private group. If piped water does not undergo the required treatment, it is not considered tap water (NSS 1999).

[6]Coliform bacteria describe a broad class of bacteria that are common in the environment. Though coliform bacteria are generally not harmful, the presence of coliforms serves as an indicator for potentially harmful pathogens and bacteria. Water suppliers test for coliform because coliform testing offers a simple and relatively inexpensive way to examine and monitor drinking water quality. *E. coli* and fecal coliform are bacteria that may originate from human or animal wastes. Microbes found in *E. coli* can cause short-term health effects, such as diarrhea, cramps, nausea, headaches, or other symptoms.

[7]The vast majority of households in rural India travel outside the dwelling to obtain their supply of drinking water

and in the household finds that all water sources incur substantial bacterial contamination during transport and storage, with this finding exacerbated for high quality drinking sources (Wright et al. 2005). Storage duration and storage material also influence the cleanliness of drinking water (Checkley et al. 2004). In a study in urban South India, water storage of 1-9 days in duration increases contaminant loads by 69 percent. The same study also documents that *Escherichia coli* (*E. coli*) loads are linked to storage material, with earthenware containing the highest contaminant loads and brass containing the lowest contaminant loads (Brick et al. 2004).

Point-of-use drinking water treatment provides a final method to remove pathogens. Studies suggest that in-home treatment, as compared to source or storage improvements, provides the most effective method to ensure consumption of clean drinking water (Brick et al. 2004, Fewtrell et al. 2005). Chemical treatment and boiling treatment can eliminate coliform in drinking water, however they are costly and because of this infrequently practiced by households (Mintz 1995, Quick et al. 1999). It remains uncertain why households do not engage in less costly but less effective treatment methods such as plain cloth filters. In Bangladesh, old saris effectively removed all zooplankton, most phytoplankton and particulates $> 20\mu$m from the water supply, and reduced the incidence of cholera by 48 percent. Coarser treatment methods may not eliminate bacteria, but they can remove larger particulates that have been found in improved and unimproved sources (Colwell et al. 2003).

4 A Stylized Model

I construct a simple model of demand for in-home treatment to explore the effect of an improvement in source water quality on in-home treatment. I show that source protection causes households to shift to less expensive modes of in-home treatment or no treatment. Lastly, I relate the monetized reduction in demand for in-home treatment to willingness to pay for an improved source.

4.1 Decision to treat

Household i living in district j derives health benefits from drinking water quality $B_{ij}(P_i, X_i)$. Benefits are decreasing in water contamination (a bad) denoted by P_i and depend on household characteristics, X_i. The health benefits from drinking water quality are household specific since households vary in health endowments and susceptibility to illness from contamination; $F(B_i)$ characterizes the cumulative distribution function of the health benefits from water quality.

A household may reduce water pollution by choosing among $k+1$ treatment technologies, including no treatment, that differ in abatement and cost, $t_k \in \{0, 1.., k\}$. The cost of in-home treatment equals $w_j * L_k$; this is the product of the wage rate (w_j) for agricultural field labor in district j and the minutes (L_k) it takes to treat drinking water with technology k. The utility from using treatment technology k is $U_{ijk} = B(P(t_k), X_i) + Y_i - w_j L_k$ as compared to $U_{ij0} =$

(McKenzie and Ray 2004, NSS 1999). Of the households with an in-dwelling water source, most face a discontinuous supply; for example a tap may flow for only 4 hours each day (McKenzie and Ray 2004).

$B(P(t_0), X_i) + Y_i$ if the household chooses no treatment; Y_i denotes incomes. Household i will choose to engage in treatment k if and only if for other technology choices $m \neq k$.

$$B_{ijk}(P(t_k), X_i) - B_{ijm}(P(t_m), X_i) > w_j(L_k - L_m) \tag{1}$$

For a household to engage in any mode of in-home treatment, as opposed to no treatment, the additional benefit from in-home treatment must exceed the costs of treatment. The fraction of households that engage in in-home treatment is $F(w_j)$, those for which the benefits of no treatment are less than a benefit threshold w_j.

4.2 Impact of source protection

Drinking water quality also depends on whether a household drinks from an improved (s_c) or unimproved (s_u) source, $s_i \in \{u, c\}$. Water pollution in s_u is at least as high as pollution in s_c; that is for a given choice of treatment $P_c = P(t, s_c) \leq P(t, s_u) = P_u$. Because of this $F^{s_c} \leq F^{s_u}$. Source and treatment are partial substitutes, $P(t_k, s_c) - P(t_0, s_c) \leq P(t_k, s_u) - P(t_0, s_u)$.[8]

In equation 1 I assume that households drink from s_u; now suppose that s_u is replaced with an improved source, s_c, at zero cost. Households will now choose to treat water if and only if $B_{ijk}(P(t_k), s_c)) - B_{ijk}(P(t_0, s_c)) > L_k w_j$. Some households, those characterized by

$$U_{ij0}(P(t_0, s_c)) > U_{ijk}(P(t_k, s_u)) > U_{ij0}(Q(t_0, s_c)) \tag{2}$$

will choose to treat an unimproved source but will not treat an improved source. Similarly, some households will shift from more expensive to less expensive treatment technologies. The reduction in demand for in-home treatment, measured as $F^{s_c}(w_j) - F^{s_u}(w_j)$, from source protection occurs for two reasons. Since households with an improved source are less susceptible to illness from no treatment, $F^{s_c} \leq F^{s_u}$, the benefits of no treatment exceed the wage rate (a benefits threshold). There is also a price effect; since source and treatment are partial substitutes, the price to produce an additional unit of water quality has increased.

Linearizing the utility that consumer i in district j receives from the consumption of treatment technology k, demand for in-home treatment can be characterized as

$$U_{ijk} = \beta_k X_i + \alpha_k s_i + \eta_k Y_i - \psi_k w_j \tag{3}$$

As shown in equation 2, the utility from in-home treatment should decrease if a household drinks from an improved source; that is $\alpha_k < 0$.

[8] Water quality tests were used to validate assumptions about drinking water quality, primary source and in-home treatment.

4.3 Willingness to pay for improved sources

Household willingness to pay for an improvement, s_u to s_c, in source water quality is measured as the change in income Y_u to Y_c that leaves household utility constant; that is $V(s_c, Y_c) = V(s_u, Y_u)$, where V() denotes indirect utility. Since household income must adjust to changes in source water quality, income becomes an implicit function of source ($Y^*(s)$). To hold utility constant along indifference curve V_u,

$$V(Y_c, s_c) - V(Y_u, s_u) = 0 = V(Y_c, s_u) - V(Y_u, s_u)(Y(s_c) - Y(s_u)) + V(Y_u, s_c) - V(Y_u, s_u) \quad (4)$$

or

$$Y(s_c) - Y(s_u) = -(V(Y_u, s_c) - V(Y_u, s_u))/V(Y_c, s_u) - V(Y_u, s_u) \quad (5)$$

As a proxy for willingness to pay, I use the change in expenditure on in-home treatment that occurs with source protection. For a discussion on the relationship between willingness to pay and averting expenditure see appendix. Total expenditure on drinking water quality E is defined as $w_j L_k$. The observed change in averting expenditure from an improvement in source water quality is,

$$E_{s_c} - E_{s_u} = w_j(F^{s_c}(w_j) - F^{s_u}(w_j)); \quad (6)$$

the fraction of households that switch from treatment to no treatment with source protection multiplied by the cost of in-home treatment.

5 Estimation strategy

This section empirically tests whether households reduce demand for in-home treatment in response to source protection. Demand for in-home treatment is estimated using a variety of discrete choice models to test the sensitivity of the results to model specification. To allow drinking water source to exogenously determine a household;s choice of in-home treatment, I use hydrological characteristics as instruments for the probability of using an improved source.

5.1 Demand for in-home treatment

Household $i's$ demand for in-home treatment technology k, T_{ik}, is estimated using binary logit, multinomial logit, nested logit, ordered probit and multinomial probit models with sub-state (or district) fixed effects and standard errors clustered at the district (or sub-district) level, [9]

$$T_{ik} = \beta_k X_i + \alpha_k S_i + \eta_k Y_i + \psi_k D_j + \gamma_l + u_i. \quad (7)$$

Demand for treatment technology k depends upon observed household and village characteristics (X_i), the price of fuel wood (p_v) in village v which is also included in X_i, a household's

[9]Where possible, I estimate demand for in-home treatment using district fixed effects. Clustering will differ depending on the level of variation in the variable of interest.

primary drinking water source (S_i), household and village measures of income (Y_i), rural district characteristics (D_j), unobservable sub-state (district) characteristics (γ_l), and the district rural wage rate (w_j) which is also included in D_j. Measures of price include the village market price of fuel wood, which measures the price of boiling treatment, and the district wage rate.[10] Measures of income (Y_i) include a durable goods index, hectares of land privately owned, and mean per capita expenditure in a village. Household preferences (X_i) include social group, percentage of females, percentage of children under the age of 5, the education level of children in school and median education level in the village. Rural district variables including literacy rates, discriminated social groups, field wage rates, population density, employment rates, and the composition of the labor force capture preferences at the district level (D_j). Sub-state fixed effects (γ_l) (or district fixed effects γ_j) control for unobserved heterogeneity at the sub-state (district) level such as government programs or state (district) regulations.

The main prediction derived in the theoretical model suggests that improvements in source water quality should reduce demand for in-home treatment. If households reduce private expenditure on in-home treatment in response to source protection then $\alpha < 0$.

5.2 Identification

This estimating equation will generate biased coefficient estimates of source protection due to omitted variables bias. Households that use improved sources likely differ from those who do not in income, health endowments and access to health services. Additionally, improved sources were intentionally constructed in villages based on village unobservables such as health services, pubilc infrastructure and child health. Because of these factors, unobservables captured in the error term will be correlated with a hosuehold's primary source; formally,

$$Cov(S_i, u_i) \neq 0. \tag{8}$$

If improved drinking water sources and risk averting behavior are both normal goods, then demand for each should increase with income. Similarly, if households have lower health endowments then they might try to improve health by choosing an improved drinking water source and engaging in in-home treatment. In these two instances, unobservables at the household level will attenuate the estimated effect of an improved source on demand for in-home treatment. By contrast, if taps and tube wells were purposefully placed in disadvantaged locations characterized by poor health, few health services or little public infrastructure then the negative correlation between primary water source and unobservable characteristics will overestimate the effect of improved sources on demand for in home treatment. Research on groundwater expansion in rural India (Black and

[10] In a multinomial model, I anticipate that demand for boiling treatment will decrease with an increase in the price of fuel wood. However, in a binary model demand for in-home treatment may increase with increases in the price of fuel wood. Cloth, other filters, and chemical treatment serve as substitutes for boiling treatment. An increase in the price of fuel wood may therefore induce households to choose no treatment or to substitute between treatment technologies.

Talbot 2005) documents that the government expanded and invested groundwater supplies in rural areas characterized by low levels of income and scarce supplies of pubic infrastructure.

Seven instruments (G_j) identify a household's primary drinking water source - minimum and maximum tube well depth, minimum and maximum discharge rate, percent of district (sub-district) allocated to dug wells and tube wells, and percent of district (sub-district) characterized by hilly topography. I will argue that these instruments measure the price to access a unit of groundwater from an improved source.

Minimum tube well depth measures the average minimum depth to reach the water table. As the minimum tube well depth increases, the cost to access a unit of drinking water from an improved source increases. The minimum discharge rate measures the average minimum volume of water per minute provided by a tube well or bore well. With an increase in the minimum discharge rate, households can access a larger quantity of water per unit of effort, subsequently lowering the price per unit of water from an improved source. Groundwater data provide spatial information on the feasibility of wells. The data indicate the area on which dug wells, hand pumps, tube wells, bore wells or a combination thereof can be constructed. For example, as the percent area on which only dug wells can be constructed increases, the average cost to bring water from an improved source to district j increases. These instruments vary at the district or sub-district level.

Household demand for an improved drinking water source is estimated using a linear probability model,

$$S_i = \beta X_i + \eta Y_i + \delta G_j + \psi D_j + \gamma_l + v_i. \tag{9}$$

To generate unbiased and consistent estimates of α_k, I include a new error term (ϵ_i) in (7), where

$$\epsilon_i = \widehat{v_i} + u_i \tag{10}$$

This error term is composed of the estimated residuals from equation 9 ($\widehat{v_i}$), as well as the stochastic component of demand for in-home treatment (u_i).[11]

Demand for for each in-home treatment is now as estimated as

$$T_{ik} = \beta_k X_i + \alpha_k S_i + \eta_k Y_i + \psi_k D_j + \gamma_l + \epsilon_i. \tag{11}$$

5.2.1 Validity concerns

Recent studies in rural India demonstrate that tube well depth and natural resource endowments affect industrial growth and the composition of industries in districts (Badiani 2009, Keskin 2009). Because of this, I expect that district hydrological characteristics will be correlated with the composition of the district labor force, district income, and district unemployment. To control for potential correlation between hydrological characteristics and the industrial composition of districts, I either estimate a district fixed effects model or in the sub-state fixed effects models I

[11] Though u_i has a logistic distribution, it approximates a normal distribution, thus allowing the use of the estimator described in Rivers and Vuong (1988).

include the rural district unemployment rate, the rural district literacy rate, and the composition of the rural district labor force. Additionally, I control for village per capita expenditure.

In the sub-state fixed effects models, the validity of my identification strategy hinges on the assumption that households do not migrate between districts based upon hydrological characteristics. I allow households to sort within a district, the U.S. equivalent of a county, based on groundwater characteristics. For example, a household can move between 2 villages in a single district if a household prefers the groundwater characteristics in village A to those in village B. However, my identification strategy prevents migration between districts based upon groundwater characteristics. Previous studies exploring the determinants of schooling in rural areas, the impact of dams on district development, and the effects of trade liberalization on district growth patterns find low labor mobility between districts in India, and do not allow for migration between districts in their identification strategies (Duflo and Pande 2007, Foster and Rosenzweig 1996, Keskin 2008).

In the district fixed effects models, I assume that households do not migrate between sub-districts, where a sub-district describes a collection of tehsils (see Appendix B for a description of sub-districts). Household migration based upon groundwater characteristics seems unlikely, unless sub-district groundwater characteristics are correlated with labor markets within a district. As a robustness test, I estimate a district fixed effects model that includes sub-district unemployment rates, literacy rates, and the rural composition of the labor force.

5.2.2 Robustness test: in-home tap users

This identification strategy predicts improved drinking water sources in so much as it predicts demand for covered groundwater sources such as tube wells, hand pumps and bore wells. However, improved drinking water sources include taps, and tap sources can come from either surface or groundwater sources. As a result, the availability of groundwater supplies may either increase or decrease the probability that a household uses a tap. As a robustness check, I exclude tap users and estimate demand for improved drinking water sources.

5.3 Model specification

To analyze the sensitivity of my results to model specification, I estimate five distinct discrete choice models of demand for in-home treatment. I am particulary interested in the extent to which my results depend upon the restrictions imposed on the stochastic error term.

5.3.1 Logit and multinomial logit

The binary logit and multinomial logit models estimate if a household engages in-home treatment and the mode of in-home treatment a household chooses. In the logit and multinomial logit model, the error term is independently and identically distributed extreme value, where the density from the unobserved component of utility is $f(u_{ik}) = e^{-u_{ik}} e^{-e^{-u_{ik}}}$. Demand for in-home treatment is estimated in the multinomial logit model using sub-state fixed effects with standard errors clustered at the district. In the binary model, demand is estimated using district fixed effects with standard errors clustered at the sub-district.

5.3.2 Robustness test: nested logit

A multinomial logit requires that the stochastic components of utility be independent across alternatives. However, households that use a home produced treatment technology will be more likely to choose other "homemade" treatment technologies than technologies only available for purchase in markets. In a nested logit, treatment technologies are partitioned into M non-overlapping nests (for example "homemade" and market produced nests) $B_1, ..., B_M$. The independence of irrelevant alternatives (IIA) assumption is imposed for alternatives within a nest but relaxed for alternatives in separate nests. The cumulative distribution of the error term is $\exp(-\sum_{m=1}^{M}(\sum_{n\in B_m}(e^{-u_{in}/\lambda_m})^{\lambda_m}))$ where $cov(u_{in}, u_{il}) = 0$ for any $n \in B_m$ and $l \in B_{m2}$ and $n \neq l$, and λ_l measures the degree of independence in unobserved utility for alternatives in nest l.

Demand for in-home treatment is then estimated as

$$T_{ik} = \beta_m X_{1i} + \beta_n X_{2i} + \alpha_j S_i + \eta_j Y_i + \psi_m D_j + \gamma_l + u_{ik}, \tag{12}$$

where subscript n denotes coefficients that vary over alternatives within a nest and m describes coefficients that differ between nests but are constant within a nest. Extending the earlier example, in the nested logit model a household first chooses whether to treat a source, then chooses between home produced and market produced technologies and finally decides on a technology. For robustness, I consider a variety of specifications that differ in the variables included in X_1 and X_2.

5.3.3 Robustness test: ordered probit

A probit model allows the stochastic component of utility to be correlated across alternatives. In the ordered probit model, I rank in-home treatment technologies by quality and households choose between three distinct quality bundles. I impose a restriction on the covariance matrix such that the error term of alternative k is correlated with the $k-1$ and $k+1$ technology bundles, but uncorrelated with all other alternatives. An ordered probit generates thresholds (α_k) that confine a household's technology choice. The probability of choosing alternative k is given by

$$\Pr[T_i = k] = F(\alpha_k - \beta X_i + \alpha S_i + \eta Y_i + \psi D_j + \gamma_l) - F(\alpha_{k-1} - \beta X_i + \alpha S_i + \eta Y_i + \psi D_j + \gamma_l), \tag{13}$$

where the error term is now normally distributed, $u_i \sim N(0, 1)$.[12]

[12] The incidental parameters problem leads to inconsistent estimation of probit models with fixed effects (Chamberlain 1984, Wooldridge 2002). Inconsistency arises due to a limited number of observations per L, where L is the number of sub-states. The number of observations per sub-state in my sample ranges between 680-800. This large number of observations in a sub-state generates aymptotically consistent and unbiased fixed effects estimators.

5.3.4 Robustness test: multinomial probit

An ordered probit model is limited in that it assumes demand for in-home treatment can be ranked along a quality gradient. In actuality, in-home treatment technologies can be ranked by market cost, opportunity cost of time, or unobserved preferences. I remove the quality ranking imposed on demand for in-home treatment technologies, and allow the unobserved portion of utility to be correlated across all choices. I use a multinomial probit to estimate demand for five distinct technology bundles. The error term is normally distributed.

6 Data and descriptive results

The data comprise information collected from five sources and vary in resolution from the household to the district level. Table 1 provides descriptive results; means are reported for all households and by primary source type. The household data consist of 26,222 rural households from 1,645 villages in 5 states of rural India sampled between January and June 1998. The data represent a subset of data collected by the National Sample Survey Organization (NSS) in Round 54 Schedules 1 and 31. Groundwater data obtained from the Central Groundwater Board of India comprise hydrological characteristics collected in 2002.

6.1 NSS Data

As shown in Table 1 approximately 70 percent of the sample population selects an improved source as the primary source of drinking water, and roughly one-third of these households use taps. Almost all households, 94 percent of improved users and 91 percent of unimproved users, perceive their drinking water source to be of satisfactory quality, where the difference in perceived quality between the two users groups is statistically significant. Just over 30 percent of households use an in-premise (located within the property boundaries) source and the remaining 70 percent of households walk on average 0.20 kilometers to the primary drinking water source. Five percent of households report using an in-home tap; these households are wealthier, more educated and more likely to engage in in-home treatment.

In-home treatment technologies are grouped into five categories: no treatment, cloth filter, other filter types, boil and chemical treatment.[13] In total, 22 percent of households engage in some form of in-home treatment, with cloth, candle filters, boil, and chemical treatment constituting 79 percent, 8 percent, 9 percent and 3.5 percent of the market for in-home treatment, respectively. If the state of Uttar Pradesh is excluded from the sample since only 1.8 percent of households treat water, the percentage of households engaging in in-home treatment increases to 33 percent, with boiling treatment constituting 10 percent of the market for in-home treatment.

[13] In-home water treatment in the NSS data describes general modes of in-home treatment. To determine the choice set of technologies associated with each in-home treatment category, I visited urban markets in Delhi and 2 rural markets. Conversations with shop owners, sales people, village panchayats (mayor or local leader) and rural households revealed minimal options for in-home treatment technologies. For example in all the visited markets, the only filter available aside from a piece of malmal cloth was a candle filter.

Additional household data capture income and preferences. Measures of income include a durable goods index, hectares of land owned and mean per capita expenditure in a village. The durable goods index describes a measure of income that is calculated using a primary components analysis of seven durable goods - bicycle, motorcycle, car, telephone, television, bathroom and latrine - solely owned by a household (Jalan et al. 2009). Information on household composition includes the percentage of children under the age of 5 and the ratio of females to males. My household measure of education is restricted to current students; it captures the highest level of schooling in years for all students in a household. Social group indicates whether a household belongs to a historically disadvantaged social group, where disadvantaged social group is defined as a scheduled caste, other backwards caste, or scheduled tribe.

Per capita expenditure and per capita education data were collected from the household expenditure survey and align with the household drinking water data at the village level. These two variables are calculated as the mean monthly expenditure and median schooling level of all sampled households in a village. Mean monthly expenditure amounts to 438 Rs and median education is literate with some primary schooling. The village market price of fuel wood measure the price to boil drinking water. Seventy-two percent of the surveyed villages report the market price of fuel wood, where the mean price for a kilogram of fuel wood in 1998 Rupees is 0.923.

6.2 Groundwater, census and wage data

Data collected in 2002 by the Central Groundwater Board of India (CGWB) spatially characterize groundwater resources at a scale of 1:250,000 in the states of Uttar Pradesh, Madhya Pradesh, Maharashtra, Tamil Nadu and Rajasthan. Variation in groundwater characteristics is at the district (sub-district) level, where well and discharge rates are calculated as the district (sub-district) mean. Figure 1 illustrates the spatial variation in feasible wells. The geology and hydrology of India allow dug wells and tube wells to be constructed on most of the land, though on 2 percent of the land improved groundwater sources are not viable and on 8 percent of the land dug wells are infeasible.

District data from the 2001 Census provide information on rural district preferences. These include literacy rates, discriminated social groups, field wage rates, population density, employment rates, and the composition of the labor force.

Information on male wages were collected by the Department of Agriculture and Cooperation (2000) between January and June 1998 for 105 of the 169 districts in the NSS data; the daily wage rate ranges from 19 Rupees to 118 Rupees (in 1998 Rs), with an average wage rate of 46 Rupees.

7 Estimation results

7.1 Demand for in-home treatment

I begin by estimating equation 7, a binary logit model on whether a household chooses in-home treatment or not using multiple fixed effects specifications. Table 2 reports results; column 1 presents estimates using village fixed effects, columns 2-4 present results using district fixed effects and column 5 presents estimates using sub-state fixed effects.

The multiple fixed effects models allow me to test whether my results are driven by unobserved village characteristics, and in the case of the sub-state fixed effects model, unobserved district characteristics. Estimates with village fixed effects indicate that an improved source reduces demand for in-home treatment by 7.5 percentage points from a probability of 0.39 to a probability of 0.31.[14] This result is robust to all fixed effects specifications, though in the district and sub-state fixed effects models it attenuates to 5 percentage points from a probability of 0.25 to 0.20. As shown in column 3, this rebound effect also holds when I regress in-home treatment on improved source only; demand for treatment decreases by 4 percent.

If households with improved surface water sources drive the difference in treatment between improved and unimproved user groups, then the limited health gains from groundwater expansion cannot be explained by behavior. To ensure that households with in-home taps are not driving my results, since tap water can come from either surface or groundwater sources, I exclude these users and estimate equation 7. As shown in column 4, surface water users are not driving the rebound effect; demand for treatment continues to reduce by 5 percentage points from a probability of 0.25 to 0.20 with source protection. Combined the results in Table 2 suggest that demand for in-home treatment reduces with source protection.

In Table 3, I consider household demand for various in-home treatment technologies. A binary model is restrictive because it only considers a household's decision to treat drinking water. Modes of in-home treatment vary in the water quality provided by the technology (as described in Figure 2), the market cost to purchase the good, and the time cost to filter the drinking water. I estimate equation 7, a multinomial logit model (with sub-state fixed effects) on demand for in-home treatment, where in-home treatment is defined as no treatment, cloth filter, other filter, boiling treatment or chemical treatment. Results are reported as the log odds ratio of choosing technology k relative to no treatment.[15] Improved sources only reduce demand for cloth treatment, the least costly and least effective treatment technology; the probability of cloth treatment relative to no treatment reduces by 3.5 percentage points.

In both the logit model with district fixed effects and the multinomial logit model, I find evidence that demand for in-home treatment increases with income and education. In the binary logit model, demand for in-home treatment increases by 4 percentage points with an 1 standard deviation increase in the durable goods index and 1 percentage point with a 100 Rs increase in monthly household expenditure. If the median household in a village or the most educated student in a household completes primary school, demand for in-home treatment increases by 1.5 percentage points and 3 percentage points, respectively. In the mulitnomial logit model, I find that a 1 standard

[14] In the results, the marginal effect of an improved source is calculated across households in the population. In the logit model, the marginal effect is estimated as $N^{-1}\sum_{N}^{i=1}[G(\beta_k X_i+\alpha_k+\eta_k Y_i+\psi_k D_j+\gamma_l)-G(\beta_k X_i+\eta_k Y_i+\psi_k D_j+\gamma_l)]$.

[15] I also estimate multinomial logit models using district fixed effects. Coefficient estimates in the sub-state fixed effects and district fixed effects specifications are qualitatively similar, however in the district fixed effects model, standard errors are not precisely estimated. Imprecise standard errors occur because for some in-home treatment technologies, only a small sample reports using the technology and I have limited degrees of freedom.

deviation increase in the durable goods index increases the probability of treatment by 0.5 to 1.5 percentage points. Education, as measured by both village and household schooling levels, is only significant in increasing the probability of cloth treatment; the lack of explanatory power for the other modes of in-home treatment may occur due to the limited number of households that engage in each of these treatment technologies. In both the binary and multinomial models, in-home treatment increases for households in non-socially disadvantaged households with the exception of boiling treatment, and in villages that report the price of fuel wood, where the ladder variable may partly capture income or whether formal markets exist in a village.

7.2 Instruments: demand for improved drinking water sources

As mentioned in the estimation strategy, a household's primary drinking water source may be correlated with unobservables such as household income, household health endowments, hygienic behavior, or the strategic placement of drinking water sources in poor, less educated and less healthy villages. To identify improved drinking water sources, I estimate equation 9, a linear probability model of the decision to drink from an improved source. Table 4 reports results; columns 1-4 report estimates using district fixed effects and sub-district groundwater characteristics and columns 5-6 report estimates using sub-state fixed effects. Column 3 excludes in-home tap users and column 4 excludes all tap users. To ensure that my results are not driven by my definition of a sub-district, I present results using subdistrict groundwater data in column 5 and district groundwater data in column 6.

I find that demand for improved sources decreases with an increase in price, when price is measured using the hydrological instruments min well depth, min discharge rate and percent land where only tube wells can be constructed. As the minimum depth to a tube well increases by 100 meters, there is a 10 to 13.5 percent reduction in the probability that a household chooses an improved source; qualitatively an increase in the price of an improved source causes a decrease in demand for an improved drinking water source. A reduction in price, as measured by a 100 lpm increase in the minimum discharge rate, raise the probability that a household chooses an improved source by approximately 1.6 to 3 percentage points. Similarly, demand for improved sources increases by roughly 6.5 percent (col. 1) with a 1 percent increase in the percent of land where only improved sources are available, though this effect attenuates to 3.8 percent and 1.4 percent in the model excluding tap users and the sub-state fixed effects model (col. 5), respectively.

As shown in Columns 1-3, for two measures of price - the percentage of topographically hilly land and the percentage of land with dug wells - demand for improved sources increases with an increase in price. By contrast in both the sub-state fixed effects models, demand for improved sources decreases with an increase in these price variables, regardless of whether I use sub-district or district hydrological characteristics. Once I exclude all tap users from the district fixed effects model, as shown in column 4 all measures of price, with the exception of topographically hilly land, reduce demand for improved sources. In columns 4-6, a one percent increase in the percent of land where only dug wells can be constructed decreases demand for improved sources by approximately

3.5 percent.

7.3 Instrumental variables: demand for in-home treatment

To consistently estimate demand for in-home treatment, I estimate equation 11, a binary logit model, which includes the predicted residuals from equation 9. Results are reported in Table 5; columns 1-3 report estimates using district fixed effects and sub-district groundwater characteristics, and columns 4 and 5 report estimates using sub-state fixed effects, and sub-district and district groundwater characteristics, respectively. The presence of an improved source reduces the probability of in-home treatment when groundwater is measured using within district variation. This result is robust to the exclusion of all covariates aside from source protection (col. 2), to the exclusion of in-home tap users (col. 3) and to the inclusion of rural tehsil controls such as the composition of the labor force, literacy rates and population density. Results from the instrumental variables model support the hypothesis that a positive relationship between water source and unobservables underestimates the coefficient on improved source. In the sub-state and district fixed effects models, the probability of in-home treatment reduces by 28 and 39 percentage points if a household drinks from an improved source.[16]

The sizable increase in the marginal effect of source protection (as compared to the marginal effect in the endogenous model) can be explained if we compare households with improved sources to those with unimproved sources. As shown in table 1, 20 percent (and this percentage drops to 18 percent if we exclude households with in-home taps) of households with an improved source as compared to 28 percent of households with an unimproved source engage in in-home treatment. Improved users systematically differ from households using unimproved sources in social group, durable goods owned, child education, village education and village per capita expenditure. On average, improved users are wealthier, more educated and in non-discriminated social groups - all factors that significantly increase the probability of in-home treatment. Similarly, other unobservables such as preferences for drinking water quality and actual income, will likely increase the probability that a household uses an improved source and engages in in-home treatment.

In Tables 6 and 7, I present results from the multinomial logit model on demand for in-home treatment described in equation 11; this model uses the predicted residuals from equation 9 to generate unbiased coefficient estimates on source. In Table 6, groundwater data in the first stage vary at the sub-district and in Table 7 groundwater data vary at the district. Coefficient estimates are reported as the log odds ratio of choosing technology k relative to no treatment. In both tables, the presence of an improved source significantly impacts a household's decision to boil drinking water and in Table 7, an improved source is weakly significant in reducing demand for cloth treatment. On average, the presence of an improved source reduces the probability that a household boils water by 22 to 27 percentage points from a probability of 0.24 to 0.02 and treats

[16] As a robustness test, I estimate demand for in-home treatment with district fixed effects using an OLS model. Similar to the results of the logit model, the marginal effect of an improved source is 38 percentage points in the 2SLS model and 5 percentage points in the endogenous OLS model.

water with cloth by 11 percentage points from 0.19 to 0.08.[17] Since so few households engage in chemical treatment, as a robustness test I exclude chemical treatment users and estimate equation 11; my results are robust to the exclusion of chemical treatment users except now source protection is weakly significant in reducing demand for cloth treatment regardless of the unit of variation in the groundwater data.

Together, these results suggest that households are shifting away from time intensive treatment technologies, particularly boiling, to no treatment. In terms of magnitude, I find that demand for boiling treatment reduces relatively more with source protection, when compared to cloth treatment, a less costly and less effective technology. Source protection induces households to shift from time intensive treatment technologies to no treatment, with a relatively greater reduction in the more costly and more effective treatment technologies.

In the 2SLS logit model with district fixed effects and 2SLS multinomial logit model, coefficient estimates on income and education mirror, in significance and magnitude, those reported in Tables 2 and 3. Importantly, however, in the 2SLS model where I use district groundwater data and in the 2SLS model where we exclude chemical users and use within district groundwater data, I now find that demand for boiling decreases with an increase in the price of fuel wood. A 1 standard deviation increase in the price per kilogram of fuel wood, 0.426 Rs, decreases the probability of in-home treatment by 2.5 percent. Additionally, consistent with the predictions in our theoretical model, I find that probability of in-home treatment decreases with an increase in the opportunity cost of time, as measured by the agricultural field wage rate.

7.4 Robustness tests

To test the independence of irrelevant alternatives assumption, a variety of nesting structures are estimated using equation 12. In a first specification, I impose a nesting structure in which a household makes a binary choice to treat or not treat and then conditional on treatment chooses a particular mode of in-home treatment. In a second nesting structure, I model a household's choice of in-home treatment in three stages - a household first chooses whether to treat drinking water, a household then chooses whether to purify (boil, chemical) or filter (cloth, ceramic) drinking water, and finally a household chooses a particular mode of in-home treatment. To test the robustness of these nesting structures, I estimate a variety of specifications that differ in the covariates included in Z_1 and Z_2 of equation 12. Under all specifications, I find that the nested logit model does not systematically differ from the multinomial logit model.

In another robustness check, I rank in-home treatment technologies by drinking water quality and estimate demand for in-home treatment using equation 13, an ordered probit model with sub-state fixed effects.[18] Table 8 reports results; columns 2-3 present second stage results using

[17]As a robustness test, I extend the data to include districts with missing groundwater data. For districts without groundwater data, I take a weighted average of the groundwater characteristics in the neighboring districts. Hydrological characteristics are weighted by the area of each neighboring district. The marginal effects of an improved source is similar to those reported in Table 7.

[18]Standard errors in the district fixed effects model are questionable

subdistrict and district groundwater characteristics, respectively. Similar to my results in the binary logit models, I find a 4 percentage point decrease in demand for in-home treatment in the sub-state fixed effects model (col. 1) and a 23 to 34 percentage point decrease in demand in-home treatment in the 2SLS model. As a separate robustness test, I estimate a multinomial probit model using groundwater characteristics as instruments for improved sources; coefficient estimates indicate that the presence of an improved source reduces demand for boiling treatment by 27 percentage points.

8 Valuing the Gains from Source Improvements

Groundwater expansion in rural India was designed to reduce household exposure to waterborne disease. However, improvements in source quality may not generate proportional increases in household drinking water quality. As shown in the previous section, improvements in source quality reduce demand for in-home treatment, thereby limiting the quality benefits of source protection. Failure to account for tradeoffs between investment in source quality and in-home treatment will overestimate the benefits of source quality improvements, and subsequently overstate the health benefits from source protection. In this section, I rely on small sample field data collected in rural India to quantify the coliform abatement and averting expenditure offset by household behavior.

8.1 Field data on water quality and market costs

Water quality data were gathered and household and village interviews were conducted during fieldwork in Madhya Pradesh during fall 2007. The water quality data are not representative of source water quality throughout India, but rather allow for some "ground-truthing" of source water quality and the effectiveness of in-home treatment technologies in the context of rural India.

Drinking water samples were collected from two rural villages in the district of Jhansi. Water samples were taken from every drinking water source in the village and measured for the presence or absence of coliform using the Colilert reagent and the Colilert P/A Test procedure in a laboratory setting.[19] All four open drinking water wells tested positive for total coliform, fecal coliform and *E. coli*. Of the six hand pumps, three tested positive for total coliform, fecal coliform and *E. coli*.[20] These findings are consistent with previous studies that suggest (i) unimproved sources are more susceptible to coliform contamination and (ii) improved sources may test positive for coliforms (Colwell et al. 2003, Helmer 1999, Wright et al. 2005).

To classify the effectiveness of in-home treatment technologies, the coliform counts in an improved and unimproved drinking water source were measured before and after the application of each in-home treatment technology. Two of the contaminated water sources (one open well and one hand pump) were randomly selected from the previously sampled drinking water sources. The Colilert reagent and the 15-Tube Most Probable Number Dilution Procedure were used in a labo-

[19] Colilert is a coliform test certified by the U.S. EPA and used by U.S. drinking water suppliers for compliance with the Safe Drinking Water Act.

[20] Coliform tests performed by the NGO, Development Alternatives, in the same villages confirm my results.

ratory to quantify coliform concentrations.[21] A total of 36 water quality tests were conducted - 2 sources were treated using 9 separate in-home treatment technologies on 2 separate occasions.

Coliform concentration is defined as the number of coliform counts per 100 mL drinking water and ranges from 0 to 1600, where 0 indicates no coliform contamination and 1600 is the maximum detection limit of Colilert. The initial coliform counts in the unimproved source and improved source were 1600 and 270 counts per 100 mL, respectively.

To collect market cost data on each risk averting technology, I visited one semi-urban and two urban markets in Delhi, surveyed all vendors on the choice set and price of treatment technologies, and purchased all available in-home treatment technologies. I verified the market price of each technology in rural villages in Haryana and Jhansi.[22] Market costs, defined as the monthly cost of in-home treatment technologies vary over technology choice and range from 0 to 95 Rs; these costs are similar to those found in other studies in India (Clasen et al. 2008).

In Figure 2, I graph the quantity of coliform abatement and market cost of each in-home treatment technology by improved and unimproved source. The horizontal axis measures coliform abatement, where a value of 0 indicates that no abatement occurs. Each point-of-use treatment technique accomplishes some abatement with more expensive technologies supplying larger reductions. Cloth treatment supplies the smallest absolute reduction in coliform counts, with coliform abatement of 700 counts per 100 mL in the unimproved source and 200 counts per 100 mL in the improved source. Boiling, chemical treatment and any combination of the two techniques eliminate all coliform from the drinking water samples. Coliform tests also confirm that conditional on no treatment, improved sources are less contaminated than unimproved sources. Lastly, the marginal product of in-home treatment is larger for unimproved than improved sources.

8.2 Quality gains from source protection

To quantify the water quality from source protection and the abatement offset by the accompanying reduction in demand for in-home treatment, I introduce a simple policy in which every household in the sample shifts from an unimproved to an improved drinking water source. Coefficients estimates from the instrumental variables binary and multinomial logit models are used in the policy simulation. To derive coliform abatement estimates for the binary models, I weight the technology-specific coliform abatement by the observed frequency of the technology in the sample population.[23]

First, I assume that each household drinks from an unimproved source and I measure the coliform abatement supplied by in-home treatment. I then introduce a policy in which each household receives access to an improved drinking water source. In a first counterfactual, households do not substitute source quality for in-home treatment; that is I constrain the coefficient on drinking wa-

[21] The 15-Most Probable Number Dilution Procedure estimates the most probable number of fecal and total coliform counts per 100 mL of water by creating three dilutions (1, 1:10 and 1:100) of the water sample and testing for the presence or absence of coliform in five samples of each dilution.

[22] I visited the nearest market surrounding each surveyed village to collect market data on in-home treatment technologies and surveyed village leaders on the price and supply of treatment technologies

[23] For example, 79 percent of the households engaging in in-home treatment use cloth filters. As a result, I weight the coliform abatement provided by cloth treatment by 0.79.

ter source (α) to be zero. Still, as shown in Figure 2 coliform abatement from in-home treatment depends on a household's primary drinking water source. In a second counterfactual, I allow a household's choice of in-home treatment to depend upon source quality. I then quantity of coliform abatement offset by behavioral choices.

Table 9 examines the abatement provided by source protection and the impact of behavioral choices on coliform abatement. Columns 1 and 2 measure the coliform abatement if households drink from an unimproved source and an improved source, respectively. Column 3 measures the change in coliform abatement from the expansion of improved sources, and Column 4 reports the percent change in abatement from the source protection.[24] Rows labeled "choices" allow demand for in-home treatment to depend on drinking water source, while rows labeled "constrained" do not. Results suggest that the reduction in demand for in-home treatment offsets the coliform abatement from improved drinking sources by roughly 5 percent.

Using estimates from the IV district fixed effects logit model, I find that source protection increases coliform abatement by 982 counts or 61.5 percent in constrained model. These sizable gains in water quality occur because households that forego in-home treatment regardless of source type receive an in-kind abatement gift of 1320 counts per 100 ml. As measured by coliform abatement, these households benefit the most from source protection. Once I allow households to substitute between source quality and in-home treatment, source protection produces 56 improvement in water quality. Using estimates from the IV sub-state fixed effects logit model, source protection leads to larger quality gains, 66 percent in the constrained model and 62 percent in the choices model.

In the remainder of Table 9, I evaluate the change in coliform abatement from source protection using results from the instrumental variables multinomial logit model. In the multinomial model, I can disentangle demand for in-home treatment by technology and assign technology specific abatement to each mode of in-home treatment. On average source protection increases coliform abatement by 847 counts in the constrained model and 770 counts in the unconstrained model. In both the binary model with district fixed effects and the multinomial logit model, behavioral choices offset coliform abatement from source protection by roughly 5 percent.

In the sections of Table 9 labeled "Cloth Treatment w/ Unimproved" and "Boil Treatment w/ Unimproved", I measure the per capita change in water quality from source protection for households that were predicted to engage in cloth treatment or boiling when drinking from an unimproved source. For households that engaged in cloth treatment pre-policy, source protection produces per capita coliform abatement of 705 counts; this amounts to a 44 percent improvement in drinking water quality. Behavioral choices offset abatement by 7.2 percent for these households. For households that chose to boil drinking water when using an unimproved source, I find that source protection on average lowers drinking water quality. Coliform abatement reduces by 16.4 percent or 263 coliform counts per 100 ml with source protection. This reduction in abatement occurs because boiling eliminates coliform from the drinking water source and provides households

[24]Percent change is measured as the difference reported in Column 3 divided by 1600, since coliform counts can vary from 0 counts to 1600 counts per 100 ml.

with a safe supply of drinking water. Households that respond to source improvements by shifting from boiling treatment to no treatment will consume higher concentrations of coliform in drinking water.

Assuming that 50 percent of improved sources are uncontaminated, my results suggest that coliform counts in drinking water total at approximately 125 counts with source protection, far exceeding the 10 counts per 100 ml standard set by the government. In the context of rural India, improved sources do not necessarily provide households with a clean water supply, yet 94 percent of households in the NSS sample and all households in my field survey perceive improved sources to be of satisfactory quality.[25] The reduction in demand for in-home treatment from source protection may occur due to false perceptions about the drinking water quality provided by improved sources. This offsetting behavior compromises some of the quality benefits for households that engaged in in-home treatment when consuming water from an unimproved source and actually lowers water quality for households that engaged in boiling or chemical treatment when drinking from an unimproved source.

8.3 Willingness to pay for source protection

Table 10 reports the change in averting expenditure in 1998 Rs using results from the instrumental variables binary logit models with district fixed effects (rows 1-2) and substate fixed effects (rows 3-4), and the instrumental variables multinomial logit model. In column 1 prices are measured as a weighted average of the monthly cost of treatment technologies in peri-urban Delhi and the village price of fuel wood, assuming that each individual requires a minimum 3 liters of drinking water per day (World Bank 2008). When measuring the price to boil water, I rely on results from a recent study in semi-urban India which finds it takes 100 grams of fuel wood and 6.42 minutes to boil a liter of water (Clasen et al. 2008). In Column 2 expenditure results are calculated using the opportunity cost of time (OCT), where the value of time is assumed to equal one-third the female district wage rate for field labor (Englin and Shonkwiler 1995).[26] However, since female wage data for most districts are missing, I use the male wage rate and impose a male-female wage differential of 1.3 (Bhan 2001). Based upon qualitative surveys, I assume that cloth treatment requires half the amount of time as boiling treatment, and ceramic and chemical treatment require neglible amounts of time.

As shown in row 1, the per capita change in expenditure on in-home treatment from an improvement in source water quality totals at 3.1 Rs or 3.7 Rs per month, depending on whether in-home treatment is evaluated at the market price or the OCT. This change amounts to a 0.85 percent savings in monthly expenditure when price is evaluated using retail prices, and a 1 percent savings when price is evaluated at the OCT. Using estimates from the IV sub-state fixed effects

[25] During field interviews, households responded that improved sources were "pure" because the water flows up from the ground and is untouched by rain runoff or animal waste. By contrast, these households consistently responded that open wells provided dirty water since households can observe waste directly entering the source.

[26] Opportunity cost of time ranges between 1/4 and 1/3 wage rate, though recently Kremer et al. find the value of time equals 6.2% of the wage rate.

logit model, the savings in averting expenditure from source protection reduces to 0.63-0.83 percent of monthly expenditure. The monthly per capita reduction in expenditure on cloth treatment is valued at 0.87 Rs or 1.08 Rs depending on whether retail or labor prices are used. This amounts to a 0.22 to 0.3 percent savings in monthly expenditure. The per capita change in expenditure on boiling from source protection is 2.15 Rs or 0.6 percent of monthly expenditure when price is measured using market cost data. The change in monthly expenditure increases to 5.2 Rs or 1.3 percent of total expenditure when price is evaluated at the OCT.

When price is measured using market cost data, I find that wealthier, non-discriminated and more educated households display a higher willingness to pay. Individuals in non-discriminated social groups are willing to pay 38 Rs per year for an improved source, as compared to 36.75 Rs per year for households in discriminated social groups; the difference in expenditure between the two social groups is statistically significant. Similiar to findings in other studies (Jalan et al. 2009), willingness to pay increases with education; when compared to households in the lowest schooling quartile, those in the highest quartile are willing to spend 35 percent more for source protection. Willlingess to pay also increases with wealth - households in the highest wealth tercile are willing to pay 18 percent more for source protection.

9 Source protection, behavior, and childhood diarrhea

So far, I have assumed that source protection leads to minimal reductions in the incidence of childhood diarrhea. In this section of the paper, I estimate the causal effect of source protection on childhood diarrhea using a linear probability model with village (district) fixed effects,

$$D_{ijl} = \alpha + \beta_1 H_l + \beta_2 F_{jl} + \beta_3 C_{ijl} + \gamma_k + u_i \tag{14}$$

The probability that child i with mother j in household l suffers from diarrhea depends on household characteristics H_l, mother characteristics F_{jl} and child characteristics C_{ijl}. Household characteristics include the total value of assets owned, the primary drinking water source and household size; female characteristics include caste, literacy and whether the father of the child is literate; child characteristics include gender and age. Village (district) fixed effects control for unobservables constant for all households in village (district) k. In an alternative model, I estimate the effect of in-home taps (I_l) on the probability of childhood diarrhea. To estimate these models, I use REDS data from 1999.

This estimating equation will generate biased coefficient estimates of source protection (or in-home taps) due to omitted variables bias. Households that use improved sources and in-home taps likely differ from those who do not in income, health endowments and private expenditure on drinking water quality. In the district fixed effects model, I am also concerned about omitted bias at the village level since improved sources were intentionally built in villages with low public infrastructure, low income and high child mortality rates. To control for this bias I estimate a 2SLS model, where in the first stage I use tehsil hydrological characteristics as instruments for whether a child drinks from an improved or unimproved source. These groundwater characteristics measure the price to access a unit of water from an improved source and also serve as an objective proxy for drinking water quality. Compared to unimproved sources, improved sources on average contain lower concentrations of total coliforms, fecal coliforms and E. coli. In the second stage, I estimate the probability of childhood diarrhea controlling for district unobservables. Since our variable of interest varies at the tehsil, standard errors are clustered at the tehsil.

Results are presented in Table 11; Columns 1-4 report results from the LPM using village fixed effects (col. 2 and 4) and district fixed effects (col. 1 and 3). In Columns 1 and 2, I find that in general source protection does not reduce the probability of childhood diarrhea. There is one exception - consistent with other economic studies in India, I find (in columns 3 and 4) that the probability of diarrhea reduces systematically by roughly 1 percentage point for households with an in-home tap (Jalan and Ravallion 2003). As shown in Column 5, results from the 2SLS models suggest that if source protection was introduced exogenously, then I should observe a 6 percentage point reduction in the probability of diarrheal disease. This finding echoes the results from a randomized study in rural Kenya that attribute a 33 percent reduction in the probability of diarrheal disease to spring protection (Kremer et al. 2009).

The endogenous placement of improved sources and a household's decision to use an improved source may be negatively correlated with unobservables captured in the error term, thereby negating the health benefits from source protection. As shown in previous sections, households with improved drinking water sources demand less in-home treatment, with a greater shift away from the most effective modes of in-home treatment. This reduction in private expenditure may also extend to other drinking water quality behaviors - hand washing, storage, in-home treatment, water removal and sanitation. I show that while exogenous improvements in drinking water quality should lower the incidence of diarrhea, to some extent behavioral choices that are negatively correlated with demand for improved sources offset these quality benefits.

10 Discussion and Policy Implications

In rural India, exogenous improvements in drinking water quality reduce the probability of childhood diarrhea. However when households are able to choose their drinking water source I find that unobservables correlated with a household's water source negate the reduction in diarrhea. Limited health benefits may occur because households reduce private expenditure on drinking water quality in response to source protection. This paper finds that households decrease expenditure on in-home treatment and in particular boiling treatment, a costly but effective mode of treatment in response to source protection. Using field data on drinking water quality, I show that reductions in demand for in-home treatment offset coliform abatement by 5 percent. Still, these counterfactual estimates also reveal that source protection leads to sizable gains in drinking water quality with the largest gains accruing to households that engage in no in-home treatment. Substitution between source quality improvements and private investment in water quality offers a partial explanation to the limited health gains from the expansion of protected drinking water sources. However, since no dose-response links coliform counts to human health, I cannot translate microbial concentrations to the incidence of diarrhea.

Demand for other water quality enhancing behaviors may also decrease with improvements in source water quality. Failure to account for tradeoffs between source quality and private expenditure on water storage or hygiene will underestimate the effect of behavioral choices and overstate the effect of source quality improvements on drinking water quality. Estimates on coliform reductions from groundwater expansion are further exaggerated since I do not consider the possibility that a drinking water source may be recontaminated during the transport and storage of the water from the source to the home. Thus, my counterfactual estimates from source quality improvements describe an upper-bound estimate on coliform abatement.

While behavioral choices offset the water quality gains from source protection, this compensatory behavior is welfare enhancing. The introduction of an improved source will cause utility maximizing households to reallocate time and money from water quality to other welfare-enhancing activities. On average, I find that the per capita savings from source protection amounts to roughly 1 percent of total monthly expenditure.

References

[1] Badiani, Reena, (2009), "Industrial Location, Growth and Mineral Resources in India", working paper.

[2] Bennett, Daniel, (2008), "Clean Water Makes You Dirty: Water Supply and Sanitation Behavior in the Philippines", working paper.

[3] Bhan, Gautam, (2001), *India Gender Profile*, Report 62, Sussex, U.K.: Institute of Development Studies.

[4] Black, Maggie and Talbot,Rupert, (2005), *Water: A Matter of Life and Health*, New Delhi, India: Oxford University Press.

[5] Brick, Thomas, Primrose, B., Chandrasekhar, R., Roy, S., and Muliyil, J., (2004), "Water Contamination in Urban South India: Household Storage Practices and their Implications for Water Safety and Enteric Infections", *International Journal of Hygiene and Environmental Health*, **207**:473-480.

[6] Briscoe, John, Furtado de Castro, Paulo, Griffin, Charles, North, James, and Olsen, Orjan, (1990), "Toward Equitable and Sustainable Rural Water Supplies: A Contingent Valuation Study in Brazil", *The World Bank Economic Review*, **4**(2):115-134.

[7] Cameron, Colin A. and Trivedi, Pravin K., (2005), *Microeconometrics: Methods and Applications*, New York: Cambridge University Press.

[8] Chaterjee, A., Das, D., Mandal, B.K., Chowdhury, T.R., Samanta, G., and Chakraborti, D., (1995), "Arsenic in ground water in six districts of West Bengal, India: the biggest arsenic calamity in the world. Part I. Arsenic species in drinking water and urine of the affected people", *Analyst*, **120**:643-650.

[9] Census of India, (1991), "Census Atlas Series 21, Rajasthan", New Delhi: Directorate of Census Operations.

[10] Census of India, (2001), "India Map", New Delhi: ML InfoMap Pvt. Ltd..

[11] Chamberlain, Gary, (1984), "Panel Data", in Zvi Griliches and Michael Intriligator (eds.), *Handbook of Econometrics*, Amsterdam: North-Holland.

[12] Checkley, William, Gilman, Robert, Black, Robert, Epstein, Leonardo, Cabrera, Lilia, Sterling, Charles and Moulton, Lawrence, (2004), "Effect of Water and Sanitation on Childhood Health in a Poor Peruvian Peri-Urban Community," *The Lancet*, **363**:112-118.

[13] Clasen, Thomas, McLaughlin, Catherine, Nayaar, Neeru, Boisson, Sophie, Gupta, Romesh, Desai, Dolly and Shah, Nimish, (2008), "Microbiological Effectiveness and Cost of Disinfecting

Water by Boiling in Semi-urban India," *American Journal Tropical Medicine and Hygiene*, **79**(3):407-413.

[14] Colwell, Rita R., Huq, Anwar, Islam, M. Sirajul, Aziz, K.M.A, Yunus, M., Khan, N. Huda, Mahmud, A., Sack, R. Bradley, Nair, G.B., Chakaborty, J., Sack, David A., and Russek-Cohen, E., (2003)"Reduction of Choler in Bangladeshi Villages by Simple Filtration."*Proceedings of the National Academy of Sciences of the United States of America*, **100**(3):1051-1055.

[15] Cutler, David and Miller, Grant, (2005), "The Role of Public Health Improvements in the Health Advances: The Twentieth-Century United States", *Demography*, **42**(1):1-22.

[16] Dept. of Drinking Water Supply, (2007), "Guidelines for the Preparation of Legislation for Framing Drinking Water Regulations", Draft, New Delhi: Ministry of Rural Development, Government of India.

[17] Duflo, Esther and Pande, Rohini, (2007), "Dams", *The Quarterly Journal of Economics*, **122**(2):601-646

[18] Englin, J. and Shonkwiler, (1995), "Modeling Recreation Demand in the Presence of Unobservable Travel Costs: Towards a Travel Price Model," *Journal of Environmental Economics and Management*, **29**:308-377.

[19] Esrey, Steven A. and Habicht, Jean-Pierre, (1986), "Epidemiologic Evidence for Health Benefits from Improved Water and Sanitation in Developing Countries", *Epidemiological Review*, **8**(1):117-128.

[20] Esrey, Steven A., Potash, J. B., Roberts, L. and Shiff, C., (1991) "Effects of improved water supply and sanitation on ascariasis, diarrhoea, dracunculiasis, hookworm infection, schistosomiasis, and trachoma", *Bulletin World Health Organization*, **69**(5):609621.

[21] Fewtrell, Lorna, Kaufmann, R., Kay, D., Enanoria, W., Haller, L. and Colford Jr., J., (2005), "Water, Sanitation, and Hygiene Interventions to Reduce Diarrhoea in Less Developed Countries: a Systematic Review and Meta-Analysis", *Lancet Infectious Diseases*, **5**:42-52.

[22] Foster, Andrew and Rosenzweig, Mark, (1996), "Technical Change and Human-Capital Returns and Investments: Evidence from the Green Revolution", *American Economic Review*, **86**:931-953.

[23] Hayashi, Fumio, (2000), *Econometrics*, Princeton: Princeton University Press.

[24] Helmer, R., (1999), "Water quality and health", *The Environmentalist*, **19**:11-16.

[25] Islam, M. Sirajul, Brooks, A., Kabir, M.S., Jahid, I.K., Islam, Shafiqul, M., Goswami, D., Nair, G.B., Larson, C., Yukiko, W., and Luby, S., (2007), "Faecal Contamination of Drinking Water Sources of Dhaka City during the 2004 Flood in Bangladesh and Use of Disinfectants for Water Treatment",*Journal of Applied Microbiology*, **103**(1):80-87.

[26] Jalan, Jyotsna, Somanathan, E., and Chaudhuri, S., (2009), "Awareness and the Demand for Environmental Quality: Survey Evidence on Drinking Water in Urban India", *Environment and Development Economics*, **14**(6):665-692.

[27] Jalan, Jyotsna and Somanathan, E., (2008), "The Importance of Being Informed: Experimental Evidence on Demand for Environmental Quality"', *Journal of Development Economics*, **87**: 14-28.

[28] Jalan, Jyotsna and Ravallion, Martin, (2003), "Does piped water reduce diarrhea for children in rural India?", *Journal of Econometrics*, **112**:153-173.

[29] Johnson, Nancy L. and Baltodano, Maria Eugenia, (2004), "The economics of community watershed management: some evidence from Nicaragua", *Ecological Economics*, **49**:57-71.

[30] Keskin, Pinar, (2009), "Thirsty Factories, Hungry Farmers: Intersectoral Impacts of Industrial Water Demand", working paper.

[31] Kremer, Michael, Leino, Jessica, Miguel, Edward and Zwane, Alix, (2009), "Spring Cleaning: A Randomized Evaluation of Source Water Quality Improvement", working paper.

[32] Lave, Charles A., (1985), "Speeding, Coordination, and the 55 MPH Limit", *The American Economic Review*, **75**(5):1159-1164.

[33] Maddala, G.S., (1983), *Limited-dependent and Qualitative Variables in Econometrics*, New York: Cambridge University Press.

[34] Madajewicz , Malgosia, Pfaff, Alexander, van Geen, Alexander, Graziano, Joseph, Hussein, Iftikhar, Momotaj, Hasina, Syliv, Roksana and Ahsan, Habibul, (2007), "Can Information Alone Change Behavior? Arsenic contamination of groundwater in Bangladesh," *Journal of Development Economics*, **84**:731-754.

[35] Mansur, Erin T., (2008), "Measuring Welfare in Restructured Electricity Markets", *The Review of Economics and Statistics*, **90**(2):369-386.

[36] McKenzie, Dave and Ray, I., (2004), "Household Water Delivery Options in Urban and Rural India", Prepared for 5th Stanford Conference on Indian Economic Development, June 3-5, 2004.

[37] McPhail, Alexander A., (1994), "Why Don't Households Connect to the Piped Water System: Observations from Tunis, Tunisia", *Land Economics*, **70**(2):189-196.

[38] Ministry of Agriculture, (2000), "Agricultural Wages in India 1997-1998", New Delhi: Government of India, Department of Agriculture and Cooperation.

[39] Mintz E., Reiff, F. and Tauxe, R., (1995), "Safe Water Treatment and Storage in the Home: A Practical New Strategy to Prevent Waterborne Diseases", *Journal of the American Medical Association*, **273**:948-953.

[40] Murray, Michael P., (2006), "Avoiding Invalid Instruments and Coping with Weak Instruments", *Journal of Economic Perspectives*, **20**(4):111-132.

[41] National Sample Survey Organization, (1999), "Drinking Water, Sanitation and Hygiene in India: NSS 54th Round", Report 449, New Delhi: Government of India, Department of Statistics.

[42] Oster, Emily, (2006) "HIV and Sexual Behavior Change: Why Not Africa", Preliminary Draft.

[43] Peltzman, Sam, (1975), "The Effect of Automobile Safety Regulations", *Journal of Political Economy*:677-726.

[44] Planning Commission, (2002), "India Assessment 2002, Water supply and Sanitation", A WHO/UNICEF Sponsored Study, New Delhi: Government of India, Planning Commission.

[45] Oettinger, Gerald S., (1999), "The Effects of Sex Education on Teen Sexual Activity and Teen Pregnancy", *The Journal of Political Economy*, **107**(3):606-644.

[46] Quick, R.E., Venczel, L.V., Mintz, E.D., Soleto, L., Aparicio, J., Gironaz, M., Hutwagner, L., Greene, K., Bopp, C., Maloney, K., Chavez, D., Sobsey, M., and Tauxe, R.V., (1999), "Diarrhoea Prevention in Bolivia through Point-of-Use Water Treatment and Safe Storage: a Promising New Strategy", *Epidemiological Infect.*, **122**:83-90.

[47] Rivers, D. and Vuong, Q., (1988), "Limited Information Estimators and Exogeneity Tests for Simultaneous Probit Models", *Journal of Econometrics*, **39**:347-366.

[48] Saka, J.D. Kalenga, (2006), "A Chemical Study of Surface and Groundwater in the Lake Chilwa Basin, Malawai" in *Groundwater pollution in Africa*, eds. Xu, Yongxin and Usher, 229-252, Leiden, The Netherlands: Taylor and Francis.

[49] Small, Kenneth and Van Dender, Kurt Van Dender, (2005), "The Effect of Improved Fuel Economy on Vehicle Miles Traveled: Estimating the Rebound Effect Using U.S. State Data", UC Irvine Economics Working Paper 05-06-03.

[50] Train, Kenneth, (2003), *Discrete Choice Methods with Simulation*, Cambridge, United Kingdom: Cambridge University Press.

[51] Whittington, Dale, Lauria, Donald T. and Mu, Xinming, (1991), "A Study of Water Vending and Willingness to Pay for Water in Onitsha, Nigeria", *World Development*, **19**(2/3):179-198.

[52] World Health Organization(WHO), (2000), "World Health Organization. Global water supply and sanitation assessment 2000 report", WHO: Geneva.

[53] World Health Organization(WHO), (2004), "Water, Sanitation, and Hygiene links to Health: Facts and Figures", WHO: Geneva.

[54] Wooldridge, Jeffrey M., (2002), *Econometric Analysis of Cross Section and Panel Data*, Cambridge: MIT Press.

[55] World Bank, (2008), "World Bank Study on Review of Effectiveness of Rural Water Supply Schemes in India", World Bank Policy Paper, New Delhi: World Bank.

[56] Wright, Jim, Gundry, Stephen, and Conroy, Ronan, (2004), "Household drinking water in developing countries: a systematic review of microbiological contamination between source and point-of-use", *Tropical Medicine and International Health*, **9**(1):106-117.

Table 1: Summary Statistics

Household variables	Obs	Mean	Std.Dev.	Min	Max	**Mean by source type**		
						Unimp	Imp	In-home
Improved source (1=yes)	26222	0.696	0.460	0	1			
Tap (1=yes)	26250	0.233	0.423	0	1			
In-home tap (1=yes)	26222	0.0453	0.208	0	1			
Distance to source (km)	26217	0.113	0.205	0	2			
In-home treatment (1=yes)	26214	0.223	0.416	0	1	0.279	0.185	0.382
Plain cloth (1=yes)	26212	0.177	0.382	0	1	0.23	0.145	0.282
Boil (1=yes)	26212	0.0198	0.139	0	1	0.021	0.0183	0.0455
Chem treatment(1=yes)	26212	0.0080	0.0810	0	1	0.010	0.006	0.022
Durables index	26174	0.284	0.438	0	2.42	0.222	0.281	0.744
Land owned (Ha)	21439	1.16	2.14	0	48	1.42	1.09	1.59
Household size	26221	5.17	2.73	1	38	5.18	5.14	5.50
Share females	26222	0.477	0.195	0	1	0.474	0.480	0.471
Share children < 5	26222	0.131	0.164	0	1	0.134	0.131	0.176
Years school eldest child	26222	1.38	3.50	0	14	1.23	1.42	1.82
Social group	26218	0.352	0.478	0	1	0.342	0.343	0.542
Village variables								
Per cap expend (Rs/mth)	1645	438	191	120	2123	435	435	508
Per cap educ (school)	1645	2.14	1.13	1	6.5	2.02	2.17	2.59
Price fuelwood reported	1645	0.724	0.447	0	1	0.711	0.736	0.639
Price fuelwood (Rs/kg)	1191	0.923	0.426	0.1	4.62	0.685	0.672	0.582
District variables								
Min well depth (m)	126	29.7	16.9	9.14	139	26.8	26.2	25.7
Max well depth (m)	155	118	69	27	450	99.0	111	101
Max discharge (L/min)	155	591	61.9	25.9	3000	466	578	447
Min discharge (L/m)	130	236	314	6.89	2400	181	203	171
Percent hilly (km^2)	157	0.033	0.076	0	0.716	0.036	0.050	0.024
Percent improved (km^2)	157	0.054	0.180	0	1	0.037	0.033	0.033
Percent dug wells (km^2)	157	0.013	0.056	0	0.55	0.013	0.007	0.005

Table 2: Logit Models of Demand for In-Home Treatment

Variable	(1)	(2)	(3)	(4)	(5)
Improved source	-0.506***	-0.447***	-0.334***	-0.486***	-0.472***
	(0.137)	(0.0976)	(0.913)	(0.0990)	(0.0958)
Durables index	1.26***	0.893***		0.855***	0.842***
	(0.0996)	(0.0815)		(0.0832)	(0.0997)
Social group	0.508***	0.286***		0.298***	0.298***
	(0.102)	(0.0797)		(0.0802)	(0.0836)
School year student	0.0106	0.0146*		0.0161*	0.0130
	(0.0112)	(0.00947)		(0.00871)	(0.00950)
Price reported		0.444*		0.513**	0.644**
		(0.302)		(0.255)	(0.239)
Price fuelwood		0.122		0.125	0.0454
		(0.150)		(0.145)	(0.146)
Village expend (1000 Rs)		0.747**		0.474*	0.513
		(0.298)		(0.282)	(0.0355)
Village education		0.132**		0.138**	0.102*
		(0.0620)		(0.0632)	(0.0577)
Fixed Effects	village	district	district	district	sub-state
District Controls					yes
Observations	8323	14753	18740	18353	21396

Notes: The dependent variable is whether (1) a household engages in any in-home treatment or not (0). Columns 1-5 report estimates from logit models, with robust standard errors clustered at the sub-district in (1)-(4) and district (5) in parentheses. The model in column 4 excludes in-home tap users. Additional variables included in Columns 1 and 3-5 are household share < 5, share females and land owned. Additional variables included in Column 5 are district controls. Asterisks indicate statistical significance; *** p<0.01, ** p<0.05, * p<0.1.

Table 3: Multinomial Logit Model of Demand for In-Home Treatment

Variable	(1) Cloth	(2) Ceramic	(3) Boil	(4) Chemical
Improved source	-0.546***	-0.184	-0.288	-0.754
	(0.132)	(0.220)	(0.251)	(0.591)
Durables index	0.592***	0.846***	1.14***	0.924*
	(0.118)	(0.155)	(0.0697)	(0.490)
Social group	0.449***	0.133	-0.318*	0.370**
	(0.118)	(0.193)	(0.176)	(0.268)
School year student	0.0299**	0.00556	0.0104	-0.0429
	0.0144	(0.0192)	(0.0210)	(0.03351)
Price reported	0.838**	0.0945	0.441	1.94**
	0.359	(0.556)	(0.502)	(0.806)
Price fuelwood	0.0420	0.250	-0.226	-0.464
	(0.203)	(0.446)	(0.184)	(0.624)
Village expend (100 Rs)	0.0263	0.110**	-0.0626	0.183**
	(0.0385)	(0.0523)	(0.0498)	(0.0921)
Village education	0.136*	-0.148	0.106	0.0852
	(0.0705)	(0.115)	(0.132)	(0.185)

Notes: The dependent variable is the mode of in-home treatment where results for cloth filters, ceramic filters, boiling and chemical treatment are reported in Columns 1, 2, 3 and 4, respectively. Columns 1-4 report log odds ratios from a multinomial logit model using sub-state fixed effects and district controls. Robust standard errors clustered at the district are in parentheses. The base outcome in all columns is no treatment. Additional variables included are hectares of land owned, share children < 5 and share of females. The number of observations is 18,127 households. Asterisks denote significance; *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 4: Demand for Improved Drinking Water Sources

Variable	(1) All	(2) All	(3) No i/s tap	(4) No tap	(5) All	(6) All
Min well depth(100m)	-0.134*	-0.121*	-0.134*	-0.0802	-0.114***	-0.102***
	(0.0789)	(0.0679)	(0.0808)	(0.0876)	(0.0316)	(0.0185)
Min discharge(100L/min)	0.0253*	0.0175	0.0299**	0.0244	0.0158**	0.0162***
	(0.0141)	(0.0133)	(0.0145)	(0.0150)	(0.00638)	(0.00491)
Percent improved only	0.651***	0.519***	0.639***	0.381**	0.136***	0.0721***
	(0.128)	(0.107)	(0.134)	(0.155)	(0.0297)	(0.0208)
Percent dug well	0.274*	0.0630	0.251	-0.381**	-0.305***	-0.346***
	(0.152)	(0.142)	(0.155)	(0.175)	(0.0737)	(0.0495)
Percent hilly	0.295**	0.248**	0.340***	0.340**	-0.156*	-0.0725
	(0.125)	(0.0996)	(0.129)	(0.158)	(0.0629)	(0.0601)
Durables index	0.0652***		0.0473***	0.0418***	0.0623***	0.0649***
	(0.00819)		(0.00883)	(0.0104)	(0.00797)	(0.00796)
Social group	0.0207***		0.0182**	0.00894	0.0251***	0.0257***
	(0.00745)		(0.00775)	(0.0829)	(0.00728)	(0.00728)
Village expend(100Rs)	-4.34e-03*		-5.21e-03**	-4.83e-03*	-1.06e-03	-1.53e-03
	(2.07e-03)		(2.18e-03)	(2.59e-03)	(1.97e-03)	(1.96e-03)
Village education	0.0242***		0.0242***	0.0186***	0.0136***	0.0143***
	(0.00360)		(0.00372)	(0.00435)	(0.00340)	(0.00339)
Fixed Effects	district	district	district	district	sub-state	sub-state
District Controls					yes	yest
Observations	16816	20838	16115	13247	18133	18133
R^2	0.20	0.18	0.20	0.26	0.15	0.15
F-stat	8.91	8.58	9.11	8.05	21.03	26.18

Notes: The dependent variable is whether (1) or not (0) a household chooses an improved drinking water source. Columns 1-5 report estimates from linear probability models where columns 1-5 use within district groundwater characteristics and the model in column 6 uses district groundwater characteristics. Columns 3 and 4 exclude households with an in-home tap and any tap, respectively. Additional variables included in columns 1-6 are the max well depth, the max discharge rate, the household share < 5, share females, land owned, price of fuel wood, price fuel wood reported and the max school year. District controls are included in columns 5-6. Asterisks indicate statistical significance;*** p<0.01, ** p<0.05, * p<0.1. The F-statistic is reported for the seven instruments.

Table 5: IV Logit Models of In-Home Treatment

Variable	(1)	(2)	(3)	(4)	(5)
Improved source	-3.64*	-4.82**	-4.02**	-2.77*	-1.29
	(1.88)	(2.25)	(2.01)	(1.59)	(2.04)
Durables index	1.09***		0.982***	0.996***	0.901***
	(0.153)		(0.132)	(0.129)	(0.132)
Social group	0.346***		0.366***	0.393***	0.365***
	(0.0943)		(0.0955)	(0.0950)	(0.105)
School year student	0.0201**		0.0213**	0.0175**	0.0178*
	(0.00940)		(0.00990)	(0.00887)	(0.0106)
Price reported	0.478		0.546*	0.696**	0.670**
	(0.311)		(0.318)	(0.255)	(0.285)
Price fuel wood	0.0749		0.115	-0.0168	0.0134
	(0.150)		(0.154)	(0.140)	(0.153)
Village expend (100 Rs)	0.0231		0.0116	0.0162	0.0196
	(0.0315)		(0.0338)	(0.0302)	(0.0318)
Village education	0.184**		0.192**	0.0965	0.0741
	(0.0822)		(0.0850)	(0.0687)	(0.0651)
Fixed Effects	district	district	district	sub-state	sub-state
District Controls				yes	yes
Observations	14753	18740	14155	18128	18128

Notes: The dependent variable is whether (1) a household engages in any in-home treatment or not (0). Columns 1-5 report estimates from logit models, with robust standard errors clustered at the subdistrict in 1-3 and district in 4-5 in parentheses. Instruments in columns 1-4 are within district groundwater characteristics and in column 6 are district groundwater characteristics. Columns 4-5 includes district controls. Additional variables included in 1,2,4 and 5 are hectares of land owned, share children < 5 and share of females. Column 3 excludes in-home tap users. Asterisks indicate statistical significance; *** p<0.01, ** p<0.05, * p<0.1.

Table 6: IV Multinomial Logit Model of In-Home Treatment with sub-district groundwater instruments

Variable	(1) Cloth	(2) Ceramic	(3) Boil	(4) Chemical
Improved source	-2.84	1.66	-6.62**	-2.82
	(2.21)	(2.67)	(3.26)	(5.85)
Durables index	0.739***	0.731***	1.83***	0.702
	(0.184)	(0.263)	(0.232)	(0.606)
Social group	0.498***	0.0930	-0.178	0.285
	(0.130)	(0.195)	(0.215)	(0.289)
School year student	0.0293**	0.00583	0.00779	-0.0411
	0.0145	(0.0191)	(0.0217)	(0.0347)
Price reported	0.882**	0.0481	0.524	1.84**
	(0.357)	(0.710)	(0.511)	(0.852)
Price fuel wood	-0.00485	0.290	-0.358	-0.378
	(0.203)	(0.457)	(0.220)	(0.647)
Village expend (100 Rs)	0.0219	0.113**	-0.0833*	0.186**
	(0.0385)	(0.0523)	(0.0492)	(0.0914)
Village education	0.169**	-0.174	0.211	0.0348
	(0.0719)	(0.123)	(0.160)	(0.186)

Notes: The dependent variable is the mode of in-home treatment where results for cloth filters, ceramic filters, boiling and chemical treatment are reported in Columns 1, 2, 3 and 4, respectively. Columns 1-4 report log odds ratios from a multinomial logit model using sub-state fixed effects and district controls. Robust standard errors clustered at the district are in parentheses. The base outcome in all columns is no treatment. Instruments are sub-district groundwater characteristics.

Additional variables included are hectares of land owned, share children < 5 and share of females.

The number of observations is 18,127 households. Asterisks denote significance; *** p<0.01, ** p<0.05, * p<0.1

Table 7: IV Multinomial Logit Model of In-Home Treatment with district groundwater instruments

Variable	(1) Cloth	(2) Ceramic	(3) Boil	(4) Chemical
Improved source	-2.93*	3.00	-6.11**	-4.41
	(1.77)	(3.14)	(2.37)	(5.64)
Durables index	0.746***	0.643**	1.79***	1.16**
	(0.163)	(0.290)	(0.209)	(0.571)
Social group	0.500***	0.0600	-0.171	0.456*
	(0.128)	(0.200)	(0.196)	(0.278)
School year student	0.0291**	0.00601	0.00727	-0.0450
	0.0143	(0.0192)	(0.0212)	(0.0350)
Price reported	0.896**	0.00497	0.571	2.04**
	(0.360)	(0.717)	(0.513)	(0.866)
Price fuel wood	-0.00896	0.315	-0.346*	-0.558
	(0.200)	(0.452)	(0.196)	(0.647)
Village expend (100 Rs)	0.0223	0.116**	-0.0709	0.176*
	(0.0384)	(0.0523)	(0.0503)	(0.0901)
Village education	0.171**	-0.195	0.193	0.138
	(0.0744)	(0.124)	(0.142)	(0.192)

Notes: The dependent variable is the mode of in-home treatment where results for cloth filters, ceramic filters, boiling and chemical treatment are reported in Columns 1, 2, 3 and 4, respectively. Columns 1-4 report log odds ratios from a multinomial logit model using sub-state fixed effects and district controls. Robust standard errors clustered at the district are in parentheses. The base outcome in all columns is no treatment. Instruments are district groundwater characteristics.
Additional variables included are hectares of land owned, share children < 5 and share of females.
The number of observations is 18,127 households. Asterisks denote significance; *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 8: Ordered Probit Models on Demand for In-Home Treatment

Variable	(1)	(2)	(3)
Improved source	-0.210***	-1.69**	-1.16*
	(0.0425)	(0.763)	(0.672)
Durables index	0.481***	0.589***	0.558***
	(0.0358)	(0.0659)	(0.0645)
Social group	0.0975***	0.162***	0.153***
	(0.0362)	(0.0461)	(0.0451)
School year student	0.00670*	0.00627	0.00651
	(0.00408)	(0.00421)	(0.00417)
Price reported	0.192*	0.263*	0.249*
	(0.0667)	(0.131)	(0.132)
Price fuelwood	0.0372	-0.00412	0.0109
	(0.0648)	(0.0695)	(0.0690)
Village expend (1000 Rs)	0.0995	-0.00756	0.00831
	(0.0123)	(0.144)	(0.143)
Village education	0.0296	0.0397	0.0313
	(0.0270)	(0.0152)	(0.0324)
Cut Points			
Cut 1	1.78	1.16	1.25
Cut 2	3.00	2.15	2.24
Cut 3	3.29	2.42	2.51
Fixed Effects	sub-state	sub-state	
District Controls	yes	yes	
Observations	21225	17987	

Notes: The dependent variable is the mode of in-home treatment where in-home treatment technologies are ordered in ascending quality. Columns 1-3 report estimates from an ordered probit model. Instruments in columns 2 and 3 are within district and district groundwater characteristics, respectively. Robust standard errors clustered at the sub-district are in parentheses. Additional controls include land owned, children < 5, and share of females. District controls are included in all columns. Asterisks denote significance; *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 9: Abatement from In-home Treatment and Improved Sources

	(1)	(2)	(3)	(4)
	Unimproved	Improved	Change	Percent Change
Second Stage Logit District FE				
Constrained Abatement	450	1432	982	61.4%
Choices Abatement	450	1346	896	56.0%
Second Stage Logit Sub-State FE				
Constrained Abatement	349	1407	1058	66.1%
Choices Abatement	349	1345	996	62.2%
Second Stage Multinomial Logit				
All Technologies				
Constrained Abatement	591	1438	847	53.0%
Choices Abatement	591	1364	774	48.3%
Cloth Treatment w/ Unimproved				
Constrained Abatement	700	1520	820	51.2%
Choices Abatement	700	1405	705	44.1%
Boil Treatment w/ Unimproved				
Constrained Abatement	1600	1600	0	0.0%
Choices Abatement	1600	1337	-263	-16.4%

Notes: Quality and expenditure estimates are calculated using estimates from the IV logit and IV multinomial logit models. Columns 1 and 2 describe abatement in the unimproved and improved source. Column 3 is the difference between (2) and (1) and column 4 calculates this difference as a percentage. Abatement is reported in counts per 100 mL of water. In the constrained model, the coefficient on source (α) equals zero. In the choices model, in-home treatment depends on a household's water source.

Table 10: Change in Averting Expenditure from Improved Sources

	(1)	(2)
	Market cost	OCT
IV Binary logit w/ District FE		
Change expend (Rs/mth)	3.09	3.70
Change expend/Month expend	0.855%	0.988%
IV Binary logit w/ Sub-state FE		
Change expend (Rs/mth)	2.31	3.14
Change expend/Month expend	0.63%	0.83%
IV Multinomial logit		
Change cloth (Rs/mth)	0.807	1.08
Change cloth/Monthly expend	0.223%	0.303%
Change boiling (Rs/mth)	2.15	5.20
Change boiling/Month expend	0.568%	1.34%

Notes: The change in expenditure is calculated using a weighted sum of technology prices and the price of fuelwood in Column 1 and a fraction of the district wage in Column 2. The change in expenditure is reported in absolute terms, and as a percentage of monthly expenditure.

A Willingness to pay and risk averting expenditure

Allow household utility to equal $U_{ijk} = B(P(t_k, s), X_i, Y_i) + \lambda(Y - w_j t_k)$ and for ease of notation assume that treatment (t) and drinking water source are continuous (s). The optimal amount of in-home treatment is such that

$$w_j = \frac{\frac{\partial U}{\partial P}\frac{\partial P}{\partial t}}{\lambda}. \tag{15}$$

The household equates the marginal cost of an additional unit of in-home treatment (w_j) to the marginal benefit from an extra unit of in-home treatment. The marginal benefit from in-home treatment is the monetized reduction in disutility from the increase in drinking water quality that results from an additional unit of in-home treatment.

A.1 Willingness to pay for improvements in source water quality

Household willingness to pay for an improvement, s_u to s_c, in source water quality is measured as the change in income Y_u to Y_c that leaves household utility constant; that is $V(s_c, Y_c) = V(s_u, Y_u)$. Household income must adjust to changes in source water quality ($Y^*(s)$) to hold utility constant along indifference curve V_u,

$$\frac{dV}{ds} = 0 = \frac{\partial V}{\partial Y}\frac{dY^*}{ds} + \frac{\partial V}{\partial s} \tag{16}$$

Table 11: Linear Probability Model of Child Diarrhea

Variable	(1)	(2)	(3)	(4)	(5)
Improved	0.000308	-0.00236			-0.0629**
	(0.00276)	(0.00351)			(0.0276)
In-home piped			-.00655*	-0.0102**	
			(0.00371)	(0.00460)	
Assets(10000Rs)	3.81e-05	4.68e-05	5.64e-05	7.06e-05	1.18e-04
	(5.38e-05)	(6.38e-05)	(5.34e-05)	(6.51e-05)	(8.18e-05)
Household size	0.00110***	0.00108***	0.00109***	0.00107***	0.00144**
	(0.000363)	(0.000393)	(0.000361)	(0.000392)	(0.000570)
Mother caste	0.00360	0.000882	0.00380	0.00119	0.00484
	(0.00309)	(0.00381)	(0.00314)	(0.00381)	(0.00411)
Mother literate	-0.000482	0.000685	-0.0000956	0.00105	0.00180
	(0.00271)	(0.00296)	(0.00281)	(0.00301)	(0.00371)
Father literate	-0.00310	-0.000778	-0.00283	-0.000546	-0.00592
	(0.00361)	(0.00282)	(0.00356)	(0.00280)	(0.00356)
Gender	-0.00258	-0.00234	-0.00270	-0.00250	-0.00491
	(0.00285)	(0.00292)	(0.00284)	(0.00291)	(0.00356)
Age (mths)	-5.5e-05***	-5.37e-05***	-5.43e-05***	-5.29e-05***	-5.77e-05***
	(1.48e-05)	(1.32e-05)	(1.47e-05)	(1.32e-05)	(1.62e-05)
Fixed Effects	district	village	district	village	district
Observations	10368	10368	10368	7535	7535
R^2	0.015	0.026	0.015	0.027	
F-stat					58.52

Notes: The dependent variable is whether (1) or not (0) a child incurs diarrhea
Columns 1-4 report estimates from linear probability models and columns 5-6 report results from 2SLS. Instruments in columns 5 and 6 are tehsil groundwater characteristics. Standard errors are clustered at the village and district in the village and district fixed effects models, respectively. Asterisks indicate statistical significance;*** p<0.01, ** p<0.05, * p<0.1.

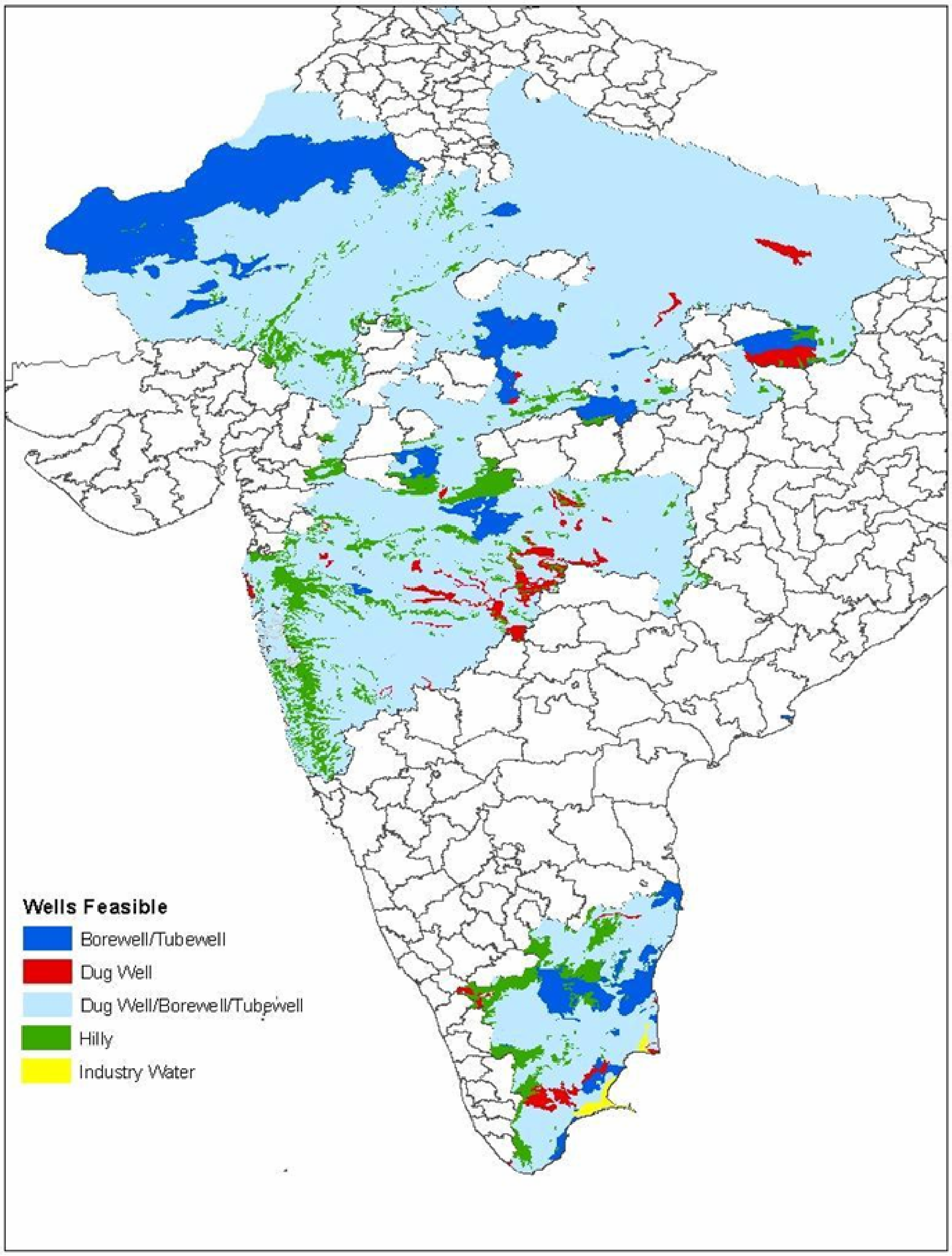

Figure 1: Distribution of groundwater characteristics by well type

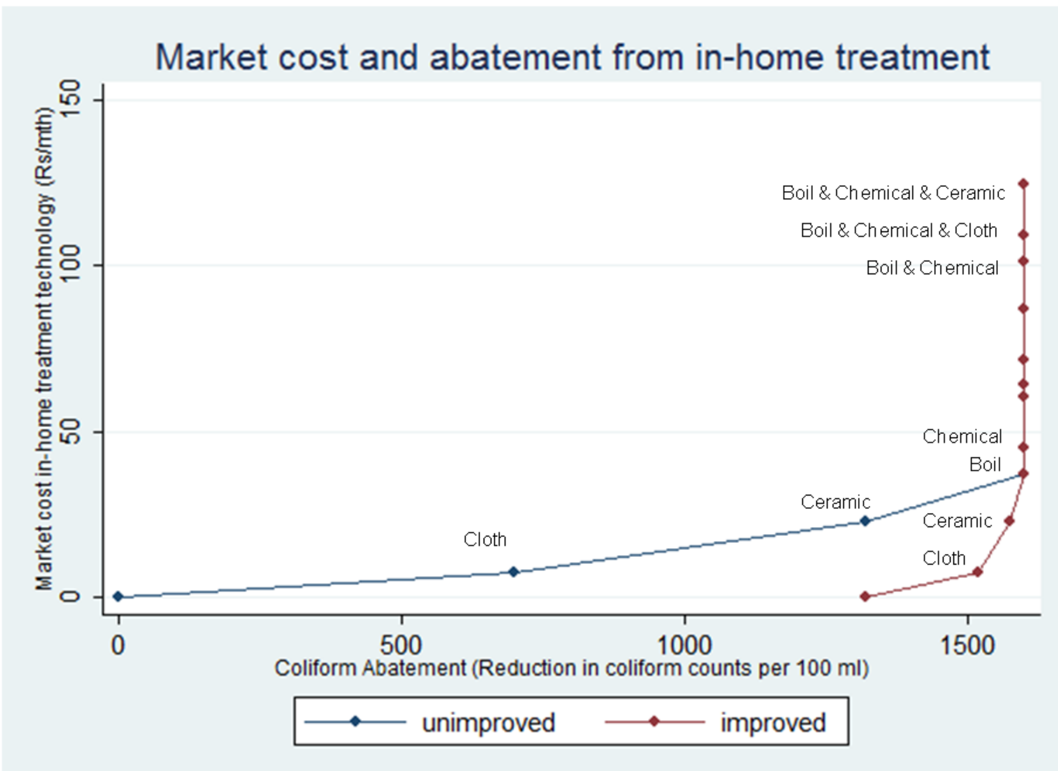

Figure 2: Price and Effectiveness of Treatment Technologies

or

$$\frac{\partial Y^*}{\partial s} = -\frac{\partial V}{\partial s} / \frac{\partial V}{\partial Y}$$

Differentiating the indirect utility function with respect to Y and s, I find that

$$\frac{\partial V}{\partial Y} = \lambda \tag{17}$$

and

$$\frac{\partial V}{\partial s} = \frac{\partial U}{\partial P}\frac{\partial P}{\partial s} \tag{18}$$

Substituting 15 into 18, I show that willingness to pay for an improvement in source water quality is,

$$\frac{\partial Y^*(s)}{\partial s} = -p_t \frac{\partial P}{\partial s} / \frac{\partial P}{\partial t}. \tag{19}$$

A.2 Comparing willingness to pay to averting expenditure

Consider the change in drinking water quality that occurs with an improvement in source water quality $(s_c - s_u)$,

$$\frac{dP}{ds} = \frac{\partial P}{\partial t}\frac{dt}{ds} + \frac{\partial P}{\partial s}. \tag{20}$$

As shown in the latter term of the right-hand side of equation 20, an improvement in source water directly increases water quality. At the same time, an improvement in source quality decreases demand for in-home treatment and this decrease in in-home treatment reduces drinking water quality.

As I will show, the change in averting expenditure may either overestimate or underestimate household willingness to pay for an improvement in drinking water quality. To determine whether I overestimate or underestimate willingness to pay, I compare

$$\frac{\partial P/\partial s}{\partial P/\partial t} \text{ to } \frac{dP}{ds} \tag{21}$$

using equations 19 and 6. If $\frac{dP}{ds} > 0$, then $\frac{\partial P/\partial s}{\partial P/\partial t} > \frac{dt}{ds}$. In words, if there is an increase in drinking water quality from source protection, then risk averting expenditure underestimates willingness to pay. This occurs because if equation $\frac{dP}{ds} > 0$, then for the equality in equation 20 to hold $\frac{dP}{ds} < \frac{\partial P/\partial s}{\partial P/\partial t}$. Using similar intuition, I can also show that if there is no change in drinking water quality from source protection, then the change in risk averting expenditure equals willingness to pay, and if there is a decrease in drinking water quality from source protection, then the observed change in risk averting expenditure overestimates household willingness to pay.

B Sub-district

In the first stage of the 2SLS models, I use hydrological characteristics to predict if a household drinks from an improved or unimproved source. The data provided by the CGWB vary within tehsils but the NSS data can only be identified at the district or sub-district level. To link the NSS data with the hydrological data, I calculate hydrological characteristics as the district mean. In averaging groundwater characteristics at the district level, I reduce the variation in the groundwater data two-fold: first, I mute the overall variation in the instruments and second, I reduce the number of observations. Imagine there are 3 equally sized tehsils labeled A, B and C; tehsil A and B are located in district 1 and tehsil C is located within district 2. The minimum discharge rate is 150 lpm in tehsil A, 50 lpm in tehsil B and 100 lpm in tehsil C. While there is substantial variation between tehsils, there is no variation in discharge rates between districts when I take the mean district discharge rate. Additionally, I also mute the variation that I would like to exploit within a district. One might expect that individuals in tehsil A will be more likely than those in tehsil B to drink from an improved source. However, my measure of discharge rate assumes that households in tehsil A and tehsil B, as well as those in tehsil C, have the same probability of drinking from an improved source.

By contrast in the district fixed effects model, I exploit within district variation and identify household behavior using more information, both in terms of variation and observations. To do this, I construct measures of groundwater characteristics that vary at the sub-district level. The limitation in this approach is in the construction of the sub-district, my unit of observation. Below, I will define a sub-district, describe how I constructed this unit of observation, and present why I believe it is the preferred unit of observation. A sub-district describes a collection of tehsils that are adjacent to each other in 80 percent of the instances and a single tehsil may be counted in multiple sub-districts.

The NSS maintains a list of villages surveyed in Schedule 54 Round 31; these files contain state names, district names, tehsil names, sub-round numbers, sub-sample numbers, village names and village codes. To protect the anonymity of households the NSS randomizes the tehsil and village codes in this master list so that they cannot be linked to the NSS survey data. However, information on states, districts, sub-rounds and sub-samples can be matched to the NSS survey data. A sub-round denotes the timing of data collection - sub-round 1 and 2 indicate that data were collected between January-March and April-June, respectively. In each district, the NSS forms stratums based upon population size using 1991 census data; villages are selected from these stratum in two sub-samples. A sub-district describes a unique state district sub-round and sub-sample. To link the hydrological data to the NSS data, I merge the hydrological data to the NSS village list data using tehsil names, district names and state names. Since tehsil and district boundaries as well as names change between the 1991 Census and the 2001 Census, I use Census maps to reclassify the 2001 groundwater data in terms of the NSS 1998 data. I then take the mean of groundwater characteristics in a sub-district, where the mean is weighted by tehsil area. These sub-district data are then matched to the NSS survey data.

To spatially evaluate what comprised a subdistrict, I looked at the geography of tehsils within a district. Tehsils that constitute a sub-district are in general geographically clustered; within a sub-district, 80 percent of tehsils are adjacent. I also find that tehsils may be located in more than one sub-district; 50 percent of tehsils are grouped into 2 sub-districts. When tehsils are placed in multiple sub-districts I continue to find spatial patterns; for example one sub-district might descibe the eastern part of the district while another describes the central part of a district. There is strong evidence for geographical clustering of sub-districts, and when compared to district, there is less geographical heterogeneity. Since fewer teshils comprise a subdistrict and these tehsil are spatially correlated, sub-district groundwater provide more precise information than groundwater characteristics averaged at the district.