Improved Source, Improved Quality? Estimating the Water Quality Gains from Groundwater Expansion in Rural India

(Job Market Paper)

Katrina Jessoe*

December 12, 2008

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 negligibly affect water quality. Using cross-sectional data from a 1998 survey in rural India, I detect and quantify the effect of improved drinking water sources on water quality. A household production function explains that improvements in source quality will reduce demand for in-home water treatment, and limit the quality benefits from improved sources. Hydrological data that exogenously measure the price and supply of improved sources empirically predict household demand for an improved source. I then estimate demand for in-home treatment using a variety of discrete choice models to examine the sensitivity of estimates to model specification. An improved source reduces the probability of in-home treatment between 0.37 and 0.47, with households shifting from chemical treatment and boiling to no treatment. Counterfactual estimates suggest that these behavioral choices offset coliform abatement from improved sources by 6.5 percent or 105 coliform counts per 100 ml.

*Yale University. Email: katrina.jessoe@yale.edu. 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. All remaining errors are my own. 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. EPA has not officially endorsed this publication and the views expressed herein may not reflect the views of the EPA.

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).[1] This initiative increased the availability of safe water sources in under-supplied communities by investing heavily in improved 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.[2] Despite improved water supplies, recent research in developing countries suggests that improvements in the water supply may produce negligible reductions in diarrheal disease (Bennett 2008, Kremer et al. 2007).

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 improved sources. This paper detects whether households trade source quality for private expenditure on quality, and then quantifies in coliform abatement (counts per 100 ml) and in Rupees the damages of these behavioral choices.

Epidemiological and public health studies have established the physiological links between water supply, sanitation, hygiene, and disease. These studies suggest that improvements in source quality significantly reduce exposure to fecal coliform, and some of these studies attribute quality improvements to measurable reductions in diarrheal and other waterborne disease (Checkley et al. 2004, Fewtrell et al. 2005). In the late 1800s and early 1900s, water quality improvements were largely responsible for a rapid reduction in U.S. mortality rates (Cutler 2005). Recently Fewtrell et al. (2005) reviewed all epidemiological studies assessing the impacts of water, sanitation, and/or hygiene on diarrhea in developing countries. Their results, which echoed previous meta-analyses (Esrey and Habicht 1986, Esrey et al. 1991), demonstrated that sanitation, hygiene, and a safe drinking water supply each independently reduce diarrheal incidence.

Another strand of literature in economics and public health suggests that a safe water supply results in minimal health improvements (Brick et al. 2004, Checkley et al. 2004, Kremer et al. 2007). These studies cite sanitation, culture or knowledge as major barriers to the realization of health benefits and conclude that 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). A recent study in a city in the Philippines finds that household substitution between piped drinking water supplies and sanitation eliminates the health gains from the expansion of piped water supplies (Bennett 2008). Due to the scarcity of sanitation in rural India, the minimal health benefits from source quality improvements are unlikely to occur from tradeoffs between sanitation and drinking

[1]On a global scale, the Millennium Development Goals propose to halve the number of people without sustainable access to safe drinking water.

[2]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).

water source. In rural settings, the links between source improvements and quality gains remain unclear (Jalan and Ravallion 2003, Kremer et al. 2007). I investigate whether tradeoffs between improvements in source water quality and in-home treatment limit the quality benefits of source protection, and subsequently overstate the health benefits of source protection.

This paper relies on cross-section data from a 1998 survey in rural India to quantify the effect of household behavior on the water quality benefits from access to improved drinking water sources. I create a simple theoretical model that demonstrates how improved drinking water sources reduce demand for in-home treatment and subsequently limit the quality benefits associated with improved sources. I also examine demand for in-home treatment technologies when households have heterogeneous preferences for water quality. The model demonstrates that improvements in source quality cause households to allocate less time and resources to point-of-use water treatment, though this effect is mitigated for households with strong preferences for water quality.

To empirically test the tradeoff between in-home treatment and source quality improvements, I estimate household demand for a primary drinking water source and then estimate demand for in-home treatment conditional on source. 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, the endogenous placement of improved drinking water sources, 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 supply and price of primary water sources. As shown in Figure 1 geological data spatially characterize the feasible locations of dug wells, tube wells, and bore wells.[3] To generate unbiased estimates of the effect of an improved source on demand for in-home treatment, I estimate 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.[4]

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 and multinomial probit models. As an extension, I empirically test the effect of heterogeneous preferences for drinking water quality on demand for in-home treatment.

Results from all models support the hypothesis that demand for in-home treatment declines with improvements in drinking water source, as households substitute source quality for private expenditure on quality. In the logit model, the probability of any mode of in-home treatment

[3]The data comprise information from seven sources and exploit household, village, tehsil, subdistrict, 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.

[4]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.

decreases by 0.07 for households with an improved source and reduces by 0.47 once I control for the endogeneity of source.

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. Preliminary results suggest that with improvements in source household demand for in-home treatment declines by 22 percent for cloth filters, 12 percent for ceramic filters, 26 percent for boiling water and approximately 42 percent for chemical treatment. Once I control for the endogeneity of drinking water sources, the effect of improved sources on demand for in-home treatment disproportionately impacts high cost technologies, though the effect is only significant for boiling treatment. In terms of water quality, boiling and chemical treatment technologies eliminate coliforms from drinking water, whereas cloth and ceramic filters reduce concentrations, but not to zero. My results suggest that improved source quality induces households to spend less on in-home treatment, with a relatively greater reduction in the most effective modes of in-home treatment.

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 provides a starting point to estimate water quality gains from source improvements. Simulation results from the instrumental variables multinomial model suggest that tradeoffs between source improvements and in-home treatment offset the coliform abatement from improved sources by 7 percent. Under the assumption that households do not respond to improvements in source, the coliform abatement from providing every household with an improved source averages 630 coliform counts per 100 mL for a water quality gain of 40 percent. When demand for in-home treatment depends upon a household's drinking water source, groundwater expansion provides a 33 percent improvement in drinking water quality. Despite substantial gains in coliform abatement from the expansion of improved drinking water sources, my results suggest that coliform counts in drinking water continue to exceed the 10 counts per 100 ml standard imposed by the the Government of India.

2 Related literature

This section provides a brief overview of the literature on offsetting behavior from regulatory innovation, household behavior and source protection, and willingness to pay for environmental quality.

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 the literature on HIV in Africa, compensating behavior strongly explains the small changes in risky sexual activity associated with knowledge of HIV infection (Oster 2006). A recent study related to drinking water and health finds that the introduction of piped drinking water supplies in an urban city in the Philippines reduces participation in community-based sanitation programs, thus eliminating the health benefits from a piped drinking water supply (Bennett 2008).

No previous studies have measured the effects of compensating behavior on the water quality gains from improved drinking water sources. However, there is a significant literature on behavioral choices, source protection and welfare. Earlier studies attribute the infrequency of point-of-use treatment to the high cost of in-home treatment technologies such as boiling (Jalan et al. 2006, Jalan and Somanathan 2004, Mintz et al. 2001, Quick et al. 1999). A study in urban India documents low levels of even low-cost filtration, suggesting that information, education and income also play a role (Jalan et al. 2006). The finding in Jalan and Somanathan (2004) that 39 percent of water collected from clean sources tested dirty after purification suggests common in-home treatment technologies may be ineffective, which may serve as a deterrent to using them.

Recent literature has explored the linkages between source quality improvements, household behavior and health. One particular study estimating the effects of randomized spring source protection in Kenya finds that after the addition of spring protection, households initially relying on a bundle of water sources increase their frequency of trips to the protected springs (Kremer et al. 2007). This study observes no change of in-home treatment with spring protection.[5]

Using revealed and stated preference techniques, economics research on willingness to pay for drinking water consistently demonstrates that households in India and in other developing countries actually pay or are willing to pay high prices for improved water quality and improved water service (Briscoe et al. 1990, Jalan et al. 2006, Johnson and Baltodano 2004, McPhail 1994, Whittington et al. 1991, Kremer et al. 2007). Other studies have exploited travel cost measures such as distance walked to reveal willingness to pay for drinking water (Kremer et al. 2007). One study in India that models the choice between two water sources as a function of time, price and other covariates observes that on average unskilled households are willing to spend approximately half a day's salary to obtain clean drinking water (Asthana 1997). This paper expands on the valuation literature by quantifying changes in willingness to pay for drinking water quality from public improvements in source water quality.

[5] They also observe no health effects from source protection which leaves unanswered the question of why household health does not improve with improvements in source quality.

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).[6] 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.[7] 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 polluted surface water to protected groundwater. While taps and tube wells provide a comparatively safer supply of drinking water, both sources are largely unregulated (Brick et al. 2004, Jalan and Somanathan 2004). In India, the government requires water suppliers to adhere to microbial standards for drinking water. According to the drinking water legislation, the detectable coliform concentration in drinking water must be less than 10 counts per 100 ml before drinking water is deemed suitable for human consumption (Dept. Drinking Water Supply 2007).[8] Despite these standards, without monitoring and enforcement of regulations, both taps and tube wells receive little to no treatment at the source, and have been shown to contain a high incidence of zooplankton, coliform and other bacteria (Islam et al. 2007, Planning Commission 2002). In addition, some deep ground water sources expose households to the separate health risk of unsafe concentrations of arsenic (Chaterjee et al. 1995, Madajewicz et al. 2005).[9] 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

[6] Unable to develop safe as a quantifiable and enforceable standard for drinking water, the World Health Organization (WHO) replaced the the term "safe" with "improved" to describe the best measurable standard for drinking water. An improved source of drinking water 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 household connections, public standpipes, bore wells, protected dug wells, protected springs and rainwater collection (WHO 2000).

[7] 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).

[8] 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* are bacteria that 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.

[9] Persistent exposure to arsenic, a carcinogen, may cause skin, lung and liver cancer and cardiovascular disease.

risk of exposure to other contaminants, this, along with compensating behavior, will offset the water quality gains of reduced microbiological contamination at the source.

While an improved source presents a safer (as compared to an unimproved source) supply of drinking water, the public health literature demonstrates that significant contamination occurs during the transport and storage of water.[10] A recent meta-analysis of field studies measuring bacterial contamination at the source and in the household demonstrates that all water sources incur substantial bacterial contamination during transport and storage, with this contamination effect exacerbated for high quality drinking sources (Wright et al. 2005). Epidemiological research suggests that storage duration and storage material 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).

By contrast, Kremer et al. (2007) suggest contamination during transport and storage is minimal, arguing that mean reversion bias in previous measurements of stored water contaminant concentrations overestimate *E. coli* loads. Within a lab setting, Brick et al. (2004) find similar results; they report that water contaminant loads decrease with storage duration. However in a field setting, they find the reverse contamination pattern - *E. coli* levels increase over time when storage occurs in the household. This suggests that water stored within a household may be exposed to recontamination from unsafe water removal practices (e.g. household members dip hands into stored drinking water upon removal).

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 guarantee clean drinking water consumption (Brick et al. 2004, Fewtrell et al. 2005). Chemical treatment and boiling treatment can eliminate coliform in drinking water. In the villages I visited, households were clearly aware of the benefits of boiling. If a child was ill, households responded that they would boil drinking water as a cure for the illness. However, many in-home treatment methods are costly and thus infrequently practiced by households (Quick et al. 1999). For example, a kilogram of firewood is necessary to boil one liter of water for a minute (Mintz 1995). 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 remove all zooplankton, most phytoplankton and particulates $> 20\mu$m, 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). Another deterrent to in-home treatment may be knowledge or perception. Qualitative interviews

[10] The vast majority of households in rural India travel outside the dwelling to obtain their supply of drinking water (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). To ensure a within-home continuous water supply, households store water and directly consume water from storage containers (Brick et al. 2004, NSS 1999). The health outcomes associated with safe drinking water depend on water quality at the point of consumption (Mintz et al. 1995).

suggest that households believe improved sources to be "pure" because the water flows up from the ground and is untouched by rain runoff or animal waste. By contrast, households consistently respond that open wells provide dirty water since households can observe storm water and waste directly entering the source. As a result, households often treat open water sources but rarely treat water from improved sources.

4 Theoretical framework

I construct a simple model to explore the effect of improvements in source quality on demand for in-home treatment. Results of the model illustrate that improved source quality induces households to reduce private expenditure on water quality, subsequently limiting the water quality and health benefits from improved sources. Additionally, I consider demand for in-home treatment when households have exogenous heterogeneous preferences for water quality. The model predicts that reductions in demand for in-home treatment will be mitigated for households with strong preferences for water quality. Results derived from the model form the hypotheses examined empirically in subsequent sections.

4.1 Setup

Consider a representative household that is both a consumer and a producer. The utility function is characterized by

$$U(Q, X). \tag{1}$$

A household's utility depends on the the consumption of drinking water quality (Q) and all other goods (X) where utility is assumed to be increasing and concave in Q and X.

Households produce water quality and income for consumption. Household income, wL, is produced from a linear function of labor where w denotes the wage rate and L denotes the number of hours worked. Drinking water quality is produced from expenditure on in-home treatment (T_x), time spent on in-home treatment (L_t) and a household's primary drinking water source (S). A household's primary drinking water source and in-home treatment only enter household utility through their effect on water quality.[11] Thus,

$$Q = Q(T_x, L_t, S). \tag{2}$$

Water quality increases at a decreasing rate with treatment; that is $\partial Q/\partial T_x > 0$ and $\partial^2 Q/\partial T_x^2 < 0$, and any level of water quality (Q) is achievable via T_x and L_t irrespective of source. I also assume

[11] There are instances where expenditure on risk averting behavior may directly enter a household's utility function. Most households store drinking water in a container made of a specific material. Materials such as brass and copper remove coliform from the drinking water. Expenditure on brass and copper storage containers will affect drinking water quality. Brass and copper containers may serve another function. Often, containers made of brass and copper form part of a daughter's dowry, and thus storage containers may directly enter a household's utility function.

that primary water source and risk averting behavior are partial substitutes. For two sources S_c and S_u, if quality of source $S_c \geq S_u$, then $\frac{\partial Q}{\partial T_x}|_{S_c} \leq \frac{\partial Q}{\partial T_x}|_{S_u}$.[12]

In terms of in-home treatment, a household can produce water quality by devoting time to in-home treatment (L_t) or purchasing in-home treatment technologies (T_x). For example, a household may spend time collecting firewood and boiling drinking water or a household may purchase chemical water purifiers at a market. Chemical purifiers and boiling function as perfect substitutes, but other in-home treatment technologies such as a cloth filter and a ceramic filter are partial substitutes, since they only reduce coliform concentrations.

Households face a budget constraint and a time constraint. The household budget constraint is given by

$$p_x X + p_t T \leq wL, \tag{3}$$

where w denotes the wage rate, L describes the number of hours worked, p_t denotes the price of treatment and p_x denotes the price of consumption goods. The household also faces a time constraint

$$L_t + L = \overline{L}, \tag{4}$$

where L_t describes time spent on in-home treatment, L denotes hours worked and $\overline{L}$ characterizes the endowment of labor. Combining the time and budget constraints yields

$$p_x X + p_t T_x \leq w(\overline{L} - L_t). \tag{5}$$

Conditional on source, the optimization problem faced by the household is to maximize household utility subject to the production function and the budget constraint. Formally, the household solves

$$\max U(Q, X) \tag{6}$$

$$\text{s.t. } Q(T_x, L_t, S) = Q$$

$$p_x X + p_t T_x \leq w(\overline{L} - L_t).$$

Before considering the optimal allocation of consumption and quality, I must evaluate a household's decision to allocate resources to T_x and L_t. Given that labor and expenditure on in-home treatment are substitutes, the allocation of resources to in-home treatment will depend on the price of treatment technologies relative to the opportunity cost of time,

$$\frac{\partial Q}{\partial T_x}/p_t = \frac{\partial Q}{\partial L_t}/w_f. \tag{7}$$

If labor and expenditure are not perfect substitutes, then a household will equate the marginal product of labor and expenditure. The opportunity cost of time (w_f) is defined as the female

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

wage rate.[13] In the case of perfect substitutes, conditional on in-home treatment, a household's opportunity cost of time will determine whether a household chooses to expend money or to expend labor on in-home treatment.[14] Since expenditure and labor function as substitutes for in-home treatment, I assume for the remainder of the section that households treat drinking water by engaging in labor, L_t.

Returning to the maximization problem, I now assume that an unimproved drinking water source (S_u) located outside the household is the only drinking water source available to a household. Solving the maximization problem presented in equation 6, I find that the optimal level of consumption and in-home treatment is such that

$$\frac{\frac{\partial U}{\partial x_i}}{p_x} = \frac{\frac{\partial U}{\partial Q}\frac{\partial Q}{\partial L_t}}{w}|_{S_u}. \tag{8}$$

Adjusting for prices, households will equate the marginal benefit from an additional unit of water quality to the marginal benefit of an additional unit of non-water consumption.

In addition, the first order condition for time devoted to risk averting behavior is

$$\frac{\partial U}{\partial Q} = \lambda w / \frac{\partial Q}{\partial L_t}. \tag{9}$$

The marginal utility from an extra unit of time is given by the Lagrange multiplier λ. The left-hand side measures the marginal benefit of an extra unit of quality, while $w/\frac{\partial Q}{\partial L_t}$ measures the marginal cost of an extra unit of treatment. As quality improves, the marginal cost of producing quality increases because there are decreasing returns to labor in the production of water quality. In contrast, the marginal cost of an additional unit of consumption, p_x, remains constant. The opportunity cost of producing an extra unit of water quality is $\frac{w}{p_x}/\frac{\partial Q}{\partial L_t}$ units of consumption and is increasing in L_t.

4.2 Introduction of an improved drinking water source

Suppose an improved water supply (S_c) is introduced to a village and assume that households are equidistant from S_c and S_u.[15] This water source provides water quality, $Q_c = Q(0, 0, S_c)$. Water quality from S_c is at least as high as water quality from S_u; that is, for any L_t and T_x, $Q_c = Q(T_x, L_t, S_c) \geq Q(T_x, L_t, S_u) = Q_u$.

[13] In rural India, a single or a few female family members are generally assigned the task of managing water collection and care. Data on field wage rates in rural India characterize separate wages for men and women. Since women receive a lower wage than men and women are responsible for managing drinking water, I use the female wage rate to measure the opportunity cost of time.

[14] In a few instances, I observe households engaging in boiling and chemical treatment. This combination may yield quality benefits aside from coliform reductions and/or households may not be fully informed as to the quality benefits of boiling and/or chemical treatment.

[15] Kremer et al. (2007) find that after spring protection households walk greater distances to access a safe drinking water source.

With the improved source, the household's production function becomes,

$$Q = Q(T_x, L_t, S_c) \tag{10}$$

The impact of this improved source may be decomposed into an income and a substitution effect. Consider the income effect of an improvement in source quality; one can imagine an in-kind gift $Q_c - Q_u$ that provides households with improved water quality conditional on any choice of X, T_x and L_t. Next, consider the substitution effect. Since the marginal product of labor (on in-home treatment) is decreasing in quality, the price to produce an additional unit of water quality has increased.[16]

The optimal choice of consumption and in-home treatment labor is such that

$$\frac{\frac{\partial U}{\partial x}}{p_x} = \frac{\frac{\partial U}{\partial Q}\frac{\partial Q}{\partial L_t}}{w}|_{S_c}. \tag{11}$$

Given the concavity of the utility function, an improvement in water quality, Q_c, decreases the marginal gain in utility from in-home treatment. Additionally, the price to consume an additional unit of water quality increases with an improvement in source quality. To maintain the equimarginal condition of (11), households increase consumption and decrease in-home treatment until the marginal utility of consumption (adjusted for prices) equals the marginal utility from water quality. If a primary drinking water source and in-home treatment are partial substitutes, then given an equidistant (and in-kind) improvement in source quality either L_t or T_x must decrease.

4.3 Heterogeneous preferences for quality

Allow households to have heterogeneous preferences for drinking water quality A. The parameter A describes an exogenous constant that varies from 0 (low quality preferences) to 1 (high quality preferences). Once again, assume that households can only produce water quality through labor on in-home treatment (L_t) or a primary drinking water source (S). A household's maximization problem is now characterized by

$$\max U(AQ, X) \tag{12}$$

$$\text{s.t. } Q(T_x, L_t, S) = Q$$

$$p_x X + p_t T \leq w(\overline{L} - L_t).$$

Conditional on source S, the first order condition is defined by

$$\frac{\frac{\partial U}{\partial x_i}}{p_i} = A\frac{\frac{\partial U}{\partial Q}\frac{\partial Q}{\partial L_t}}{w}|_{S}. \tag{13}$$

[16] Since expenditure on in-home treatment and labor devoted to in-home treatment are substitutes, the marginal product of expenditure is also decreasing in quality.

An exogenous increase in A will increase demand for in-home treatment. Conditional on source, if households have heterogeneous preferences for drinking water quality, then households with strong preferences for drinking water quality will demand more in-home treatment as compared to households with low preferences for drinking water quality.

5 Estimation strategy

The theoretical framework demonstrates that if households are behaving rationally, then optimal household behavior may compromise the water quality benefits from improvements in source water quality. This section empirically tests the tradeoff between source quality improvements and in-home treatment. To allow drinking water source to exogenously determine a household's choice of in-home treatment, hydrological characteristics are used to predict household demand for a drinking water source. Demand for in-home treatment is estimated with a variety of discrete choice models to test the sensitivity of the results to model specification. To test the effect of heterogeneous preferences on demand for in-home treatment, I include the presence of a village water utility as a proxy variable for strong preferences for water quality.

5.1 Demand for drinking water source

In the theoretical framework, a household derives utility from the consumption of drinking water quality and the consumption of all other goods. Let's now examine the sub-utility that a household derives from the consumption of a primary drinking water source.[17] Consider a simple utility function $U(X_i, Y_i, G_j, D_j, \gamma_l)$ where utility from the consumption of a drinking water source depends upon observed household characteristics (X_i), measures of income (Y_i), hydrological characteristics of the district (G_j), other observed district characteristics (D_j) and unobserved (γ_l) sub-state characteristics.[18,19] Household demand for a primary drinking water source is characterized by the latent utility function,

$$U_i = \beta X_i + \eta Y_i + \delta G_j + \psi D_j + \gamma_l. \tag{14}$$

where each household i will choose an improved source if, and only if, household utility from the improved source (S_c) exceeds household utility from the unimproved source (S_u); that is if and only if, $U(S_c) > U(S_u)$. Household demand for drinking water source is estimated using a linear probability model with the error term (v_i) clustered at the village level,

$$S_i = \beta X_i + \eta Y_i + \delta G_j + \psi D_j + \gamma_l + v_i. \tag{15}$$

[17]Recall that a household's primary drinking water source only enters utility through its effect on drinking water quality

[18]I also allow hydrological characteristics to vary at the sub-district level. In these specifications I control for unobserved district heterogeneity and omit observed district controls.

[19]The NSS divides each state into four to six sub-states; I control for this unobserved heterogeneity using fixed effects.

Demand for a primary source is a function, in theory, of prices, income, preferences, and the price of substitutes and complements. The price of a primary drinking water source is measured using hydrological characteristics (G_j) that vary at the district level. These include the proportion of land allocated for dug wells, the percent of land allocated for tube wells or bore wells, the percent of land topographically hilly, the minimum tube well depth, the maximum tube well depth, the minimum discharge rate, and the maximum discharge rate. The availability of groundwater directly determines the price of a drinking water source. For example, as the minimum depth to the water table increases, the price to access a source increases. Measures of income (Y_i) include an index of durable goods, 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. 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) control for unobserved heterogeneity at the sub-state level such as government programs or state regulations.

5.1.1 Tap Users

Improved drinking water sources include taps, tube wells, hand pumps and bore wells. By definition, tap sources can come from either surface or groundwater sources and must adhere to treatment requirements as mandated by a local municipality, government or water supplier. Groundwater characteristics directly influence the likelihood that a household has access to a tube well, hand pump or bore well. As the price to access a groundwater source increases, demand for these sources will decrease. In contrast, since tap water can come from either surface or groundwater sources, 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.

5.2 Demand for in-home treatment

I estimate demand for in-home treatment and examine how demand changes in response to improved drinking water sources. To analyze the sensitivity of my results to model specification, I estimate five distinct discrete choice models of in-home treatment. I am particularly interested in the extent to which my results depend upon the restrictions imposed on the stochastic error term.

Consider a simple utility function $U(X_i, S_i, Y_i, D_j, \gamma_l)$ where utility from the consumption of treatment technology k depends upon household and village preferences (X_i), the price of fuel wood (also included in X_i), household and village income (Y_i), a household's primary drinking water source (S_i), district preferences (D_j) and unobservable district characteristics(γ_l).[20] A household's

[20]The price of fuel wood measures the price of boiling treatment. 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

primary drinking water source describes the price of a substitute good for in-home treatment. A household's choice of in-home treatment is discrete; a household chooses among K in-home treatment technologies.

I characterize the systematic component of utility as

$$U_{ik} = \beta_k X_i + \alpha_k S_i + \eta_k Y_i + \psi_k D_j + \gamma_l, \tag{16}$$

where utility is assumed to be linear in parameters. Conditional on the income and time constraint, each household purchases the risk averting technology that gives it the highest utility. That is, household i will purchase bundle k if, and only if, for all other technology choices $m \geq 0$ and $m \neq k$, $U(T_k) > U(T_m)$.

I assume that this latent utility function characterizes the observed demand for different risk averting technologies (T_k). Demand for in-home treatment is estimated using binary logit, multinomial logit, nested logit, ordered probit and multinomial probit models with substate fixed effects and standard errors clustered at the village level, [21]

$$T_{ik} = \beta_k X_i + \alpha_k S_i + \eta_k Y_i + \psi_k D_j + \gamma_l + u_i. \tag{17}$$

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 improvements in source quality then $\alpha < 0$.

5.2.1 Identification

A potential problem with this discrete choice framework is that households also demand a primary drinking water source. Unobservables captured in the error term such as health endowments, the endogenous placement of improved drinking water sources, or preferences for water quality will likely be correlated with a household's choice of primary source. Formally,

$$Cov(v_i, u_i) \neq 0. \tag{18}$$

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 instances, unobservables at the household level will attenuate the estimated effect of an improved source. By contrast, if taps and tube wells were purposefully placed in locations characterized by low income or low preferences for water quality, then the negative correlation between primary water source and unobservable characteristics will overestimate the effect of improved sources on demand for in home treatment.

serve as substitutes for boiling treatment. An increase in the price of fuel wood may therefore induce households to choose no treatment or choose substitute treatment technologies.

[21] Where possible, I estimate demand for in-home treatment using district fixed effects.

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 allocated to dug wells and tube wells, and percent of district characterized by hilly topography. Minimum tube well depth measures the average minimum depth to reach the water table at the district level. As the minimum tube well depth increases, the costs of installing and accessing an improved source increase. The minimum discharge rate measures the average minimum volume of water provided by a tube well or bore well at the district level. With an increase in the minimum discharge rate, more households can rely on a fixed supply of water or a fixed number of households can use a larger volume of drinking water, subsequently lowering the price of an improved source. Groundwater data provide spatial information on the feasibility of wells. Specifically, the data indicate the area on which dug wells, hand pumps, tube wells, bore wells or a combination thereof can be constructed. These characteristics vary at the district level and will strongly influence whether a household uses an improved or an unimproved drinking water source.

Two recent studies demonstrate that both district industrial growth and district industrial composition in rural India depend upon mineral deposits and tube well depth (Badiani 2008, Keskin 2008). 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 include the district unemployment rate, the district literacy rate, and the composition of the district labor force as controls in all specifications. Additionally, I control for village per capita expenditure.

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. The validity of these instruments is based on the assumption that households are unaware of the geographical distribution of aquifers at the district level, yet the spatial distribution of aquifers determines whether households have access to covered groundwater sources.

To generate unbiased and consistent estimates of α_k, I include a new error term (ϵ_i) in (17), where

$$\epsilon_i = \widehat{v_i} + u_i \tag{19}$$

This error term is composed of the estimated residuals from demand for source ($\widehat{v_i}$), as well as the stochastic component of demand for in-home treatment (u_i).[22,23]

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

$$T_{ik} = \beta_k X_i + \alpha_k S_i + \eta_k Y_i + \psi_k D_j + \gamma_l + \epsilon_i. \tag{20}$$

[22]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).

[23]In this specification of demand for an improved source, I include the price of fuel wood.

5.2.2 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, I assume that that the error term is independently and identically distributed extreme value, where the density from the unobserved component of utility is

$$f(u_{ik}) = e^{-u_{ik}} e^{-e^{-u_{ik}}}.$$

The difference between the two extreme value variables has a logistic distribution,

$$F(u_{ik} - u_{ij}) = \frac{e^{u_{ikj}}}{1 + e^{u_{ikj}}}.$$

The systematic component of utility V_{ik} can be written as

$$V_{ik} = \beta_k X_i + \alpha_k S_i + \eta_k Y_i + \psi_k D_j + \gamma_l.$$

Demand for in-home treatment is estimated as

$$T_{ik} = \beta_k X_i + \alpha_k S_i + \eta_k Y_i + \psi_k D_j + \gamma_l + u_i \tag{21}$$

5.2.3 Nested logit

A multinomial logit requires that the stochastic components of utility be independent across alternatives. In a nested logit, I partition the set of alternatives into M nonoverlapping nests $B_1, ..., B_M$. I continue to impose the independence of irrelevant alternatives (IIA) assumption for alternatives within a nest but I relax the IIA property for alternatives in separate nests. The error term, u_{ij}, is now correlated for any two alternatives in the same nest, but uncorrelated for alternatives in separate nests. The cumulative distribution of the error term is

$$\exp(-\sum_{m=1}^{M} (\sum_{j \in B_m} (e^{-u_{ij}/\lambda_m})^{\lambda_m}))$$

where $cov(u_{ij}, u_{ik}) = 0$ for any $j \in B_m$ and $k \in B_l$ and $l \neq k$, and λ_k measures the degree of independence in unobserved utility for alternatives in nest k.

Demand for in-home treatment is then estimated by

$$T_{ik} = \beta_m X_{1i} + \beta_j X_{2i} + \alpha_j S_i + \eta_j Y_i + \psi_m D_l + \gamma_l + u_{ik}, \tag{22}$$

where subscript j denotes coefficients that vary over alternatives within a nest and m describes coefficients that differ between nests but are constant within a nest. For robustness, I consider a variety of specifications that differ in the variables included in X_1 and X_2. As another robustness test, I estimate a variety of nesting structures.

5.2.4 Ordered probit

I relax the IIA restriction imposed by independently and identically distributed extreme value error terms and now assume that the error terms are normally distributed. 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 four 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. Demand for in-home treatment is now estimated as

$$T_i = k \text{ if } \alpha_{k-1} < U_i \leq \alpha_k \text{ , where } \alpha_0 = -\infty \text{ and } \alpha_K = \infty. \tag{23}$$

An ordered probit generates thresholds (α_k) that confine a household's technology choice. The probability of choosing alternative k is given by

$$\Pr[T_i = k] = F(\alpha_k - \beta X_i + \alpha S_i + \eta Y_i + \psi D_j + \gamma_l) - F(\alpha_{k-1} - \beta X_i + \alpha S_i + \eta Y_i + \psi D_j + \gamma_l), \tag{24}$$

where the error term is now normally distributed,

$$u_i \sim N(0,1),$$

$$F(X,S,Y,D) = \Phi(X,S,D) = \int_{-\infty}^{X,S,Y,D} \phi(X,S,Y,D)dv, \text{ and}$$

and

$$\phi(X,S,Y,D) = (2\pi)^{-1/2} \exp(-(X,S,Y,D)^2/2).$$

This model estimates a total of $k-1+R$ parameters, where $k-1$ denotes the number of threshold parameters and R denotes the number of regressors. The normally distributed errors are clustered at the village level.

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 substate generates aymptotically consistent and unbiased fixed effects estimators.

5.2.5 Multinomial probit

An ordered probit 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.

The systematic component of utility V_{ik} can be described by

$$V_{ik} = \beta_k X_i + \alpha_k S_i + \eta_k Y_i + \psi_k D_j. \tag{25}$$

5.2.6 Heterogeneous preferences

In the theoretical framework, I show that households with strong preferences for drinking water quality will demand more in-home treatment, conditional on source. To empirically test the effect of heterogeneous preferences for water quality on demand for in-home treatment, I exploit sorting by water utilities. Water utilities will locate to areas characterized by a high willingness to pay for water quality. I now include an indicator variable, M_m, that denotes the presence of a billable water utility in village m. I estimate a multinomial logit model of demand for in-home treatment,

$$T_{ik} = \beta_k X_i + \alpha_k S_i + \eta_k Y_i + \rho_k M_m + \psi_k D_j + \gamma_l + u_i. \tag{26}$$

If the presence of a billable water utility measures strong preferences for drinking water quality, then $\rho_k > 0$.

Though the location decision of water utilities will be strongly correlated with unobservable household characteristics, this correlation will lead to an underestimate of ρ. This occurs because the presence of a water utility does not strictly capture preferences for drinking water quality. In particular, water municipalities may also measure drinking water quality and/or an income effect. If water utilities proxy for water quality then households with higher drinking quality will demand less in-home treatment, and $\rho < 0$. Similarly, if water utilities measure an income effect, then the existence of a billable water utility will decrease demand for in-home treatment. Demand for in-home treatment will decrease because households must pay for the service provided by a water utility.

The proposition derived in the theoretical model predicts that households with strong preferences for water quality will demand more in-home treatment, conditional on source. Though the empirical test of this prediction will produce biased estimates, the effect of strong preferences for water quality on demand for in-home treatment will be underestimated rather than overestimated.

6 Data and descriptive results

The data comprise information collected from seven sources and vary in resolution from the household to the district level. Table 1 provides descriptive results including the units of measurement and the level of variation in the data. The household data consist of 26,250 rural households from 1,645 villages in 5 states of rural India sampled between January and June 1998. The data represent a cross-sectional subset of data collected by the National Sample Survey Organization

(NSS) in Round 54 Schedules 1, 31 and 33.[24] On average, 16 households were surveyed in each village. Groundwater data were obtained from the Central Groundwater Board of India. These data comprise hydrological characteristics collected between 1998-2002. Water quality data were gathered during fieldwork in Madhya Pradesh during fall 2007. These data vary by drinking water source and in-home treatment technology.

Household and village interviews were conducted in fall 2007 with 50 households from 6 separate villages in the Bundelkhand region of Uttar Pradesh and the state of Haryana.[25] The detailed results of the small-scale household survey will clarify and complement the information collected in the 1998 NSS household survey.

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. Roughly one-third of the households with an improved source use taps as the primary source of drinking water. Just over 30 percent of households use an in-premise source and households without an in-premise source walk on average 0.20 kilometers to the primary drinking water source. Households with an in-premise source consist of (i) households with an in-home water source and (ii) households with a water source located outside the housing structure but within the property boundaries. Twelve percent of households report using an in-home tap. Other households with an in-premise source rely on dug wells and tube wells located outside the household but within the property.

In-home treatment technologies are grouped into five categories: no treatment, cloth filter, other filter types, boil and chemical treatment.[26] Households that report regular participation in multiple categories are grouped into the boil category if the response includes boil, and are grouped into the the chemical category if the responds includes chemical treatment but does not include the boil treatment. In total, 22 percent of households engage in some form of in-home treatment, with cloth, candle filters, boil, and chemical treatment constituting 86 percent, 8.6 percent, 3.5 percent and 2 percent of the market for in-home treatment, respectively.

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 durables good index describes a measure of income that is calculated as a linear combination of five durable goods - bicycle, motorcycle, car, telephone and television - owned by a household. Information on household composition includes the percentage of young children in a household and

[24] Household surveys of drinking water, sanitation and hygiene, and household expenditure were conducted in Schedule 31 and Schedule 1, respectively. Data in Schedule 31 are comprised of village public goods.

[25] A representative sample of villages was chosen based upon the existence of development projects; three villages received extensive development investment aid and three villages received no external investment aid.

[26] 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. Only one technology for each in-home treatment category was supplied in each market. For example in all the frequented markets, the only filter available aside from a piece of malmal cloth was a candle filter.

the ratio of females to males. My household measure of education is restricted to current students. Household education denotes the highest level of schooling 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. Using data on the monthly bill for a piped drinking water source, I construct a lower bound for the availability of piped water in a village. Piped water providers exist in about 16 percent of the sampled villages. Villages with water utilities characterize areas with potentially high preferences for water quality and with a substitute good for in-home treatment.

6.2 Groundwater and Census data

Data collected between 1998-2002 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.[27] These data were geo-referenced using the 2001 Census of India map, and attributes including rock type, feasible wells, well depth, discharge rate were geo-coded. The groundwater data and the household data were aligned at the district level using 1991 and 2001 Census Maps of India. Variation in groundwater characteristics is at the district level, where well and discharge rates are calculated as the 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 district preferences. These include literacy rates, discriminated social groups, field wage rates, population density, employment rates, and the composition of the labor force. Monthly district field wage data were collected by the Department of Agriculture and Cooperation (2000) between January and June 1998 and serve as a proxy for income.

6.3 Field data on water quality and market costs

Drinking water samples were collected from two representative 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.[28] The water quality data are not representative of source water quality

[27]The CGWB could only provide groundwater data for 5 states. The CGWB selected the five states based upon variation in hydrological characteristics. I am especially grateful to Mr. Swaroop for assistance in accessing and interpreting groundwater data.

[28]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.

throughout India. These tests allow for some "ground-truthing" of source water quality in the context of rural India. All four open drinking water wells tested positive for both total coliform and *E. coli*. Of the six tube wells/hand pumps, three tested positive for total coliform and *E. coli*.[29] These findings are consistent with epidemiological studies that suggest (i) unimproved sources are more susceptible to coliform contamination and (ii) improved sources may test positive for total coliform (Colwell et al. 2003, Helmer 1999, Wright et al. 2005).

To classify the effectiveness of in-home treatment technologies, the coliform concentration of drinking water sources was measured after the application of an in-home treatment technology. Each of the 12 in-home treatment technologies listed in the NSS survey were performed on an improved and an unimproved source. 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 laboratory to quantify coliform concentrations.[30] A total of 48 water quality tests were conducted - 2 sources were treated using 12 separate in-home treatment technologies and each test was implemented on 2 separate occasions.[31] These tests are not representative of household conditions throughout rural India and were conducted in a controlled environment. Nonetheless, this fieldwork provides an estimated qualitative and quantitative index on the effectiveness of in-home treatment technologies.

Coliform concentration is defined as the number of coliform counts per 100 mL drinking water and ranges from 0 to $\geq$1600, where 0 indicates no coliform contamination and 1600 is the maximum detection of limit of Colilert. The initial coliform count in the unimproved source and improved source were 1600 and 270 counts per 100 mL, respectively. My results indicate that within a laboratory setting each in-home treatment technology reduces coliform concentrations and that the level of abatement depends upon the technology and the drinking water source.

To collect market cost data on each risk averting technology, I visited one semi-urban and two urban markets in Delhi, surveyed all vendors in each market 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. In particular, I visited the nearest market surrounding each surveyed village to collect market data on in-home treatment technologies and I surveyed each village panchayat (mayor or local leader) on the price and supply of each in-home treatment technology. Market cost defines the monthly market cost of an in-home treatment technology.[32] Market cost data collected in fall 2007 vary over technology choice and range from 0 to 249 Rupees.

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

[30] Given time and budget constraints, the aforementioned procedure provides the only feasible methodology to quantify coliform concentrations. 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.

[31] Colilert protocol requires water quality tests within 28 hours of collection. Due to time and resource constraints, I was limited to 48 water quality tests.

[32] The minimum daily quantity of water is based upon a World Bank report that states each person requires 3L/day of drinking water to survive (World Bank 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 abatement of coliform, where a value of 0 indicates that no abatement occurs. Each point-of-use treatment technique accomplishes some abatement with more expensive technologies yielding larger reductions. Cloth treatment yields the smallest absolute reductions, with coliform abatement of 700 counts per 100 mL in the unimproved source and coliform abatement of 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. Coliform tests also confirm that conditional on no treatment, improved sources are less contaminated than unimproved sources.

7 Estimation results

7.1 Demand for source

I begin by estimating equation 15, the linear probability model of demand for an improved drinking water source. Table 2 reports results; Columns 1 and 2 report estimates using sub-state fixed effects and district groundwater characteristics and Columns 3 and 4 describe the estimates using district fixed effects and within-district groundwater characteristics. Coefficient estimates presented in Columns 1 and 3 apply to the entire population and the estimates reported in Columns 2 and 4 exclude tap users. I find that hydrological characteristics are significant in predicting a household's choice of drinking water source, though the coefficient estimates differ in sign and magnitude depending on whether I use district or within-district groundwater variation.

Coefficient estimates from Column 1 suggest that demand for improved sources decreases as the price of groundwater increases. A 1 percent increase in the area covered by hills or the area in which only dug well sources are feasible lowers the probability that a household chooses an improved source by 6 percent and 34 percent, respectively. As the minimum depth to a tube well increases by 100 meters, there is a 10 percent reduction in the probability that a household chooses an improved source. A reduction in price, as measured by the minimum discharge rate and the percent of land where only improved sources are available, raises the probability that a household chooses an improved source by 2 percent and 0.07 percent.

Results from the district fixed effects model in Column 3 present mixed results for the marginal effect of price on demand. 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. My other measures of price are consistent with economic theory (demand for source decreases with increases in price), but differ in magnitude when compared to the estimates presented in Column 1. The magnitude of the coefficients on the minimum well depth, the percent improved and the minimum discharge rate increase by factors of 1.5, 1.75 and 9, respectively.

A dominant concern of the analysis is whether tap households drive the results on demand for source, and subsequently drive the results on demand for in-home treatment. The disparate results in Columns 1 and 3 can be partially reconciled when I eliminate tap users from the demand models.

As shown in Columns 2 and 4, once I exclude tap users, all measures of price, with the exception of topographically hilly land, reduce demand for improved sources. However, the marginal effects of price continues to differ in the two specifications - the effect of price is two to three times larger in the district fixed effects specification.[33]

Coefficient estimates on income in Table 2 present mixed and at times counterintuitive results, suggesting that a household's primary source may be partially determined by product placement. My measures of income indicate that demand for improved sources increases with durable goods, but drops with gains in village per capita expenditure and hectares of land owned. Endogenous product placement offers an explanation to the negative effect of land owned and village per capita expenditure on demand for improved sources.[34] Research on groundwater expansion in rural India (Black and 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.

7.2 Demand for in-home treatment

An identifying assumption of my instrumental variables approach is that unobserved tehsil, and in the case of the sub-state fixed effects model, unobserved district characteristics, are uncorrelated with the error term. The main specifications used throughout the analysis control either for unobserved sub-state heterogeneity and observable district characteristics or unobserved district heterogeneity. To ensure that the results of the instrumental variables models are not driven by unobserved village heterogeneity, I estimate equation 17, a binary logit model on whether a household chooses in-home treatment or not, using multiple fixed effects specifications. Table 3 reports results; Column 1 presents estimates using village fixed effects, Columns 2 and 3 present results using district fixed effects and Columns 4 and 5 present estimates using sub-state fixed effects.

If households with improved surface water sources drive the substitution between improved and unimproved user groups, then the limited health gains from groundwater expansion cannot be explained by compensating behavior. To test the effect of a piped water supply on demand for in-home treatment, a dummy variable for villages with a billable water utility is included in Columns 3 and 5.

Estimation results from Column 1 indicate that an improved source reduces demand for in-home treatment by a probability of 0.07; this represents a 18 percent reduction in demand.[35] Controlling for unobserved village heterogeneity, I find that households with higher levels of income, more hectares owned, and in non-disadvantaged social groups are more likely to treat drinking water. Results suggest that the marginal effect of an improved source on demand for in-home treatment is

[33]Given that tap sources can come from either surface or groundwater sources, groundwater characteristics such as dug wells and hilly areas can predict either the presence of unimproved sources or the presence of taps.

[34]The negative effect of privately owned hectares on the demand for improved source may occur due to the strong correlation between land owned and the existence of an open well within the the premise. Households that own more land tend to also possess an open well within the premise, and likely rely on this open well as the primary drinking water source.

[35]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 $N^{-}1\sum_{N}^{i=1}[G(\beta_k X_i+\alpha_k+\eta_k Y_i+\psi_k D_j+\gamma_l)-G(\beta_k X_i+\eta_k Y_i+\psi_k D_j+\gamma_l)]$.

robust to all fixed effects specifications; in both the district fixed effects model (controlling for some village characteristics) in Column 2 and the sub-state fixed effects model in Column 4, the presence of an improved sources reduces demand for in-home treatment by a probability of 0.07. Results suggest that the sub-state specification is a close approximation for a district fixed effects model. Most importantly, as shown in Columns 3 and 5, improved source coefficients (and the marginal effects of an improved source) do not attenuate after controlling for the presence of a water utility. These results suggest that the compensating behavior results are not driven by households with a piped water supply.

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 behaviors, or the endogenous placement of drinking water sources. To consistently estimate demand for in-home treatment, I estimate equation 20, a binary logit model, which includes the predicted residuals from equation 15. Results are reported in Table 4; Column 1 reports estimates using sub-state fixed effects and district groundwater characteristics, and Column 2 reports estimates using district fixed effects and within-district groundwater characteristics.

Results from the instrumental variables model support the hypothesis that a positive relationship between drinking water source and unobservables underestimates the coefficient on improved source. The presence of an improved source reduces the probability of in-home treatment by 0.15 or 50 percent in the sub-state fixed effects model, though this effect is not statistically significant. Coefficient estimates in Column 1 are similar in sign to the estimates presented in Column 4 of Table 3, though the coefficient estimates on improved source, land owned and village expenditure are no longer significant. In Column 2, the coefficient estimate on improved source is significant, approximately triple in magnitude to the estimate in Column 1 and indicates that the presence of an improved source decreases the probability of in-home treatment by 0.50 or 76 percent. Coefficients estimates on income in Columns 1 and 2 indicate that the probability of in-home treatment increases with durable goods and per capita village expenditure.[36] Preferences as measured by village per capita education and non-discriminated social group significantly increase demand for in-home treatment.

In Table 5, I consider a household's choice over various in-home treatment technologies and how a household's primary drinking water source affects demand for each in-home treatment technology. A binary model is restrictive because it only considers a household's decision to treat drinking water. In actuality, 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.[37] I estimate equation 21, a multinomial logit model on demand for in-home treatment,

[36] However, in the district fixed effects model (Column 2), I observe a significant and negative effect of land ownership on in-home treatment. Households that own large tracts of land are more likely to have a privately owned drinking water source within the premise. As a result, these sources cannot be contaminated by other users and may be perceived as cleaner, when compared to shared sources. The NSS survey asks households to describe the quality of the primary drinking water source. Ninety-three percent of households with an unimproved source within the premise respond that the drinking water source is of high quality.

[37] Cloth filtering and boiling represent high opportunity cost of time technologies. By comparison, ceramic filters

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 in Table 5.[38] The presence of an improved source reduces the probability that a household chooses in-home treatment relative to no treatment by 22 percent for cloth filters, 12 percent for other filters, 26 percent for boiling and 37 percent for chemical purification, though the effect is not significant in the case of ceramic filters.

To generate unbiased coefficient estimates of source, I estimate equation 20, a multinomial logit model on demand for in-home treatment that includes the predicted residuals from equation 15. Table 6 reports results, where estimates are reported as the log odds ratio of choosing technology k relative to no treatment. The presence of an improved source only significantly impacts a household's decision to boil drinking water. A comparison of the log odds ratios in Table 6 suggests that demand for boiling and chemical treatment reduces relatively more with source quality improvements, when compared to cloth and ceramic filters which describe less effective and less costly technologies. On average, the presence of an improved source reduces the probability that a household boils by 0.36 or chemically treats by 0.09.

Coefficient estimates on income and preferences suggest that these variables strongly determine the mode of in-home treatment selected by a household. The probability of in-home treatment increases by 0.02 to 0.04 with a one standard deviation increase in income, as measured by the durable goods index. Income, as measured by hectares owned and village per capita expenditure, is only significant in increasing demand for chemical treatment. Preferences differentially affect demand for each mode of in-home treatment; this occurs in part because these treatment technologies vary in quality, market cost, time cost, and availability.

7.3 Robustness tests

To test the independence of irrelevant alternatives assumption, a variety of nesting structures are estimated using equation 22. In a first specification, I impose a nesting structure in which a household makes a binary choice to treat or not treat and then 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 or filter 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 X_1 and X_2 of equation 22. Under all specifications, I find

---

and chemical technologies are characterized by a high market cost and low opportunity cost of time. Households must purchase ceramic or chemical technologies in a formal market, whereas fuel wood for boiling can be collected by a household and cloth filters can be made of clothing. In terms of time costs, households must boil or filter each liter of water prior to drinking. By contrast, with chemical and ceramic filters households either add chemical tablets to the daily supply of drinking water or pour a daily supply of drinking water into a container with an installed filter.

[38] 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. Due to the imprecision of the standard errors, I report results from the sub-state fixed effects model

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 24, an ordered probit model. I assume that the error term is normally distributed and allow the stochastic component of utility to be correlated across alternatives. Table 7 reports results; Columns 1 and 2 present results using substate fixed effects and district controls, and Columns 3 and 4 report results using district fixed effects. Columns 2 and 4 report second stage results from instrumental variables models. An improved source increases the probability of no treatment by 0.06 in both Columns 1 and 3, respectively. Though not statistically significant, the coefficient estimates on improved source in Columns 2 and 4 support the compensating behavior hypothesis, and suggest that failure to control for the endogeneity of source attenuates the effect of compensating behavior. The sign and significance of coefficients estimates on income and preferences closely approximate the estimates of the binary logit models in Tables 3 and 4. As a separate robustness test, I estimate equation 25, a multinomial probit model.[39] Coefficient estimates indicate that the presence of an improved source significantly reduces demand for all in-home treatment technologies.

7.4 Heterogeneous preferences for quality

To incorporate heterogeneity in preferences for water quality, I estimate equation 26, a multinomial logit model of demand for in-home treatment, where in-home treatment is defined as no treatment, cloth filter, other filter, boiling or chemical treatment. Results are reported as the log odds ratio of choosing technology k relative to no treatment in Table 8. Demand for all modes of in-home treatment decreases with an improved source - 23 percent for cloth, 14 percent for ceramic, 32 percent for boiling, and 41 percent for chemical treatment. The presence of a water utility significantly increases demand for cloth filters and boiling. Though the marginal effects are underestimated, the probability of cloth treatment or boil treatment increases by 0.02 for households located in villages with a water utility, irrespective of source.

To generate unbiased coefficient estimates of improved sources, I estimate equation 26 and substitute the residual defined in equation 15 for u_i. Table 9 presents results from the second stage of the multinomial instrumental variables model; estimates are reported as the log odds ratio of choosing technology k relative to no treatment. The presence of a water municipality is significant in increasing the probability of cloth treatment and boiling by 0.02 and 0.03, respectively. Similar to the results in Table 6, the presence of an improved source significantly reduces the probability that a household boils drinking water by 0.36. Coefficient estimates on income and preferences suggest that these variables strongly affect demand for in-home treatment, and mirror the results in Table 6. Results from the multinomial models suggest that while compensating behavior reduces the water quality gains from improved sources, the magnitude of compensatory behavior depends on a household's preference for drinking water quality.

[39]Sub-state fixed effects are not included in the multinomial probit model.

8 Estimating Gains from Source Improvements

Groundwater expansion in rural India was designed to reduce household exposure to waterborne disease. However, improvements in source quality do not necessitate proportional increases in household drinking water quality. The results from the previous section suggest that improvements in source quality reduce private expenditure on drinking water quality, 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 quantify the coliform abatement offset by household behavior. I also estimate the per capita reduction in monthly private expenditure on drinking water quality that occurs from improvements in source quality.[40]

A first counterfactual assumes that households do not exchange source quality for private expenditure on in-home treatment technologies. To provide a benchmark estimate of coliform abatement, I constrain the coefficient on drinking water source to a value of zero and assume that all households use an unimproved water source. Both binary and multinomial logit models allow for closed form solutions and I estimate standard errors using the delta method. To derive coliform abatement estimates for the binary models, I weight the quantity of coliform abatement provided by each in-home treatment technology by the observed frequency of the technology in the population.[41] I next simulate a policy in which all households receive access to improved drinking water sources. In this scenario improved sources do not influence in-home treatment through household behavior. However, a household's primary drinking water source differentially impacts the coliform abatement supplied by in-home treatment technologies.

As shown in the previous section, results suggest that improvements in source quality reduce private expenditure on drinking water quality. This basic result is robust to all discrete choice specifications though the magnitude of the effect varies across specifications. In a second set of counterfactuals, I allow a household's choice of in-home treatment to depend upon source quality. I then estimate the quantity of coliform abatement offset from behavioral choices using the binary logit model with district fixed effects, the mulitinomial logit model with sub-state fixed effects and the instrumental variables models. Using market cost data, the amount of per capita private expenditure offset by household behavior is also estimated.

Table 10 examines the abatement and expenditure implications from groundwater expansion and the importance of behavioral choices in measuring these effects. Columns 1 and 2 provide estimates on the quantity of coliform abatement per 100 mL drinking water from an unimproved source and an improved source, respectively. These columns also provide estimates on monthly per capita expenditure if households use an unimproved and improved source. Column 3 measures the change in coliform abatement (expenditure) from the expansion of improved sources, and Column 4

[40] To measure monthly per capita in-home treatment costs, I assume that each person requires 3L of drinking water per day (World Bank 2008). I then estimate the cost of providing a monthly supply of each technology to an individual.

[41] For example, summary statistics reveal that 79 percent of the in-home treatment population use cloth filtration. As a result, I weight the coliform abatement provided by cloth treatment by 0.79.

reports the percent change in abatement (expenditure) from the introduction of improved sources.[42] Results described in Table 10 suggest that behavioral choices offset the coliform abatement from improved drinking sources by 1 percent to 6 percent. Expenditure estimates reveal that private expenditure on in-home treatment reduces by 3 percent to 69 percent with improvements in source water quality.

In the constrained (no behavioral choice) logit model, quality estimates indicate that the introduction of improved sources increases coliform abatement by 1124 counts per 100 ml or 70 percent. These sizable gains in water quality occur because improved sources provide all households with a coliform abatement quantity of 1320 counts per 100 mL. Once I allow households to substitute between source quality and in-home treatment, results from the logit model predict that source improvements will reduce demand for in-home treatment in the sample population from 30 percent to 23 percent. Allowing for behavioral choices, groundwater expansion produces a 69 percent improvement in water quality. Coefficient estimates from the multinomial logit model predict that source improvements will reduce demand for in-home treatment from 27 percent to 22.5 percent of the sampled population. In the multinomial model, source quality improvements will increase coliform abatement by 1137 counts per 100 mL in the constrained model and 1123 counts per 100 mL in the unconstrained model. Behavioral choices offset coliform abatement by 1 percent in both the logit model and multinomial model.

Despite the extensive gains in drinking water quality from the expansion of improved sources, the measured concentration of coliform in drinking water exceeds the coliform standards established by the Government of India. In my estimates coliform counts in drinking water total at approximately 215 counts per 100 ml, following the universal expansion of protected groundwater sources; these coliform counts far exceed the maximum of 10 counts per 100 ml standard mandated by the Government of India.

Second stage results of the logit and multinomial logit models suggest that estimates of the endogenous model strongly underpredict the effect of source on treatment. After controlling for the endogeneity of source, the logit model predicts that if all households drink from unimproved sources, then 62.5 percent of households will choose some mode of in-home treatment. Coliform abatement from groundwater expansion decreases from 70 percent to 57 percent in the constrained model and 69 percent to 51 percent in the unconstrained model, once I instrument for a household's primary drinking water source. Behavioral choices offset the coliform abatement from groundwater expansion by 6 percent. Failure to control for the endogeneity of source produces substantial overestimates of the coliform abatement provided by source protection.

Even after accounting for behavioral choices and the endogeneity of source, the logit model overestimates the effect of improved sources on coliform abatement. In the logit model, I assume that all modes of in-home treatment produce the same level of coliform abatement and I evaluate changes in the overall frequency of in-home treatment. However, results from the multinomial model

[42] 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, I measure a 100 percent change in abatement as 1600.

indicate that the presence of an improved source differentially affects demand for each in-home treatment technology. Since different modes of in-home treatment provide various quantities of coliform abatement, estimates that allow coefficients to vary by technology more accurately predict quality changes from groundwater expansion. The instrumental variables model predicts that coliform abatement will increase by 632 coliform counts per 100 ml or 39 percent in the constrained multinomial logit model, and by 525 coliform counts per 100 ml or 33 percent if households are allowed to substitute between source quality and mode of in-home treatment. Behavioral choices offset the water quality gains from improved drinking water sources by 6.5 percent.

Using second stage estimates from the multinomial logit model, I disaggregate the quality impact of source protection by mode of in-home treatment. Households that do not participate in treatment pre-policy benefit the most from source improvements. While improved sources supply households with a relatively cleaner source of drinking water, in some cases coliform concentrations in improved drinking water supplies continue to exceed the government mandated 10 counts per 100 ml standard.[43] Estimates indicate that on average coliform abatement lessens by 36 percent for households that boil drinking water before the expansion of groundwater sources.[44] This reduction in abatement occurs because boiling and chemical treatment purify drinking water of coliform and provide households with a safe supply of drinking water. Households that respond to source improvements by shifting from boiling treatment or chemical treatment to no treatment will consume higher concentrations of coliform in drinking water from improvements in source quality.

9 Discussion and Policy Implications

In rural India, improvements in source water quality reduce demand for in-home treatment, and subsequently limit the the quality benefits of source protection. To empirically estimate demand for an improved source, I use hydrological data that exogenously measure the price and supply of improved sources. Estimates from the instrumental variables logit model on demand for in-home treatment suggest that improvements in drinking water source reduce demand for any mode of in-home treatment by a probability of 0.50. Results from the instrumental variables multinomial model indicate that demand for in-home treatment shifts from boiling and chemical treatment to no treatment, though the effect of an improved source is only significant for boiling. The findings in the multinomial logit model are robust to a variety of discrete choice specifications and to the inclusion of heterogeneous preferences for drinking water quality. The effects of compensatory behavior are attenuated for those households with strong preferences for drinking water quality. Results from discrete choice models of demand for in-home treatment suggest that improved source quality induces households to decrease expenditure on in-home treatment, with a relatively greater reduction in the most effective (and most costly) modes of in-home treatment. Substitution between

[43] In the sampled water source coliform counts totaled at 280 counts per 100 ml.

[44] Demand for boiling treatment responds most strongly to improvements in source quality, when compared to other modes of in-home treatment. The percent of households that boil drinking water drops from from 38 to 1.5 percent with the introduction of improved sources.

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.

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. From the household perspective, a reduction in private expenditure on water quality is welfare-increasing and efficient.

Drinking water regulations set by the Government of India require the detectable coliform concentration in drinking water to be less than 10 counts per 100 ml before drinking water is suitable for human consumption (Dept. Drinking Water Supply 2007). To estimate coliform abatement from groundwater expansion, I introduce a policy that provides all households with access to improved drinking water sources. I then quantify expenditure on in-home treatment and coliform abatement under various behavioral assumptions to estimate the expenditure and the coliform abatement offset by household behavior. Results from the policy simulations indicate that behavioral choices offset coliform abatement by 6.5 percent or by 105 counts per 100 ml. Counterfactual estimates also reveal that improvements in source quality lead to sizable improvements in drinking water quality. These large gains in coliform abatement occur because most households in rural India do not participate in treatment. Despite the gains in water quality from a national drinking water policy that provides each household with improved drinking water sources, microbial counts in drinking water continue to exceed regulatory standards.

While I estimate substitution between in-home treatment and source quality improvements, 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. My counterfactual estimates on coliform abatement from source quality improvements describe an upper-bound estimate on coliform abatement from the expansion of protected groundwater sources. A policy scenario that provides universal access to improved drinking water sources in rural India will increase drinking water quality and household welfare; nonetheless upper bound estimates of coliform abatement suggest that microbial counts in drinking water will continue to exceed microbial standards and these improved drinking water sources will remain unsuitable for human consumption.

References

[1] Athana, Anand N., (1997), "Where the Water is Free but the Buckets are Empty: Demand Analysis of Drinking Water in Rural India", *Open Economies Review* **8**:137-149.

[2] Badiani, Reena, (2008), "Industrial Location, Growth and Mineral Resources in India", working paper.

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

[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] 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.

[14] 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.

[15] 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.

[16] 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.

[17] 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.

[18] 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.

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

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

[21] 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.

[22] Jalan, Jyotsna, Somanathan, E., and Chaudhuri, S., (2006), "Demand for Environmental Quality: Survey Evidence on Drinking Water in Urban India" Working Papers UWEC-2007-09, University of Washington, Department of Economics.

[23] Jalan, Jyotsna and Somanathan, E., (2004), "Being informed matters: Experimental evidence on the demand for environmental quality", Indian Statistical Institute, Planning Unit, New Delhi Discussion Papers 04-08, Indian Statistical Institute, New Delhi, India.

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

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

[26] Keskin, Pinar, (2008), "Thirsty Factories, Hungry Farmers: Intersectoral Impacts of Industrial Water Demand", working paper.

[27] Kremer, Michael, Leino, Jessica, Miguel, Edward and Zwane, Alix, (2007), "Spring Cleaning: A Randomized Evaluation of Source Water Quality Improvement", working paper.

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

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

[30] Madajewicz , Malgosia, Pfaff, Alexander, van Geen, Alexander, Graziano, Joseph, Hussein, Iftikhar, Momotaj, Hasina, Syliv, Roksana and Ahsan, Habibul, (2005), "Can Information Alone Change Behavior? Arsenic contamination of groundwater in Bangladesh," working paper, Columbia University.

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

[32] 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.

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

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

[35] 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.

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

[37] 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.

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

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

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

[41] 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.

[42] 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.

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

[44] 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.

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

[46] 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.

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

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

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

[50] 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.

[51] 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	Source	Year	Obs	Mean	Std. Dev.	Min	Max
Improved source (1=yes)	NSS 54.31	1998	26222	0.696	0.450	0	1
Tap (1=yes)	NSS 54.31	1998	26250	0.233	0.423	0	1
Tubewell/Borewell (1=yes)	NSS 54.31	1998	26250	0.463	0.499	0	1
In-premise source (1=yes)	NSS 54.31	1998	26222	0.305	0.460	0	1
In-home source (1=yes)	NSS 54.31	1998	26222	0.127	0.333	0	1
Distance to source (km)	NSS 54.31	1998	26217	0.113	0.205	0	2
In-home treatment (1=yes)	NSS 54.31	1998	26219	0.223	0.462	0	1
Plain cloth (1=yes)	NSS 54.31	1998	26212	0.177	0.382	0	1
Boil (1=yes)	NSS 54.31	1998	26216	0.0198	0.140	0	1
Chemical treatment(1=yes)	NSS 54.31	1998	26221	0.0100	0.0968	0	1
Durables index	NSS 54.31	1998	26176	0.546	0.768	0	5
Land owned (Ha)	NSS 54.31	1998	21439	1.16	214	0	4800
Household size	NSS 54.31	1998	26221	5.17	2.73	1	38
Share females	NSS 54.31	1998	26222	0.477	0.195	0	1
Share children < 5	NSS 54.31	1998	26222	0.131	0.164	0	1
Years school eldest child	NSS 54.31	1998	26222	1.38	3.50	0	14
Social group	NSS 54.31	1998	26218	0.352	0.478	0	1
Agricultural labor (1=yes)	NSS 54.31	1998	26218	0.694	0.461	0	1
Village variables							
Per cap expend (Rs/mth)	NSS 54.1	1998	1645	4381	1909	1199	21233
Per cap educ (school)	NSS 54.1	1998	1645	2.14	1.13	1	6.5
Piped system (1=yes)	NSS 54.1	1998	1645	0.163	0.423	0	1
Price fuelwood (Rs/kg)	NSS 54.31	1998	1192	0.923	0.426	0.1	4.62
District variables							
Field wage (Rs/hr)	AC	1998	104	58.2	13.4	19	118
Min well depth (m)	CGWB	2001	134	64.3	59.5	30	578
Max well depth (m)	CGWB	2001	162	239	185	75	1800
Max discharge (L/min)	CGWB	2001	162	1132	1133	76	3571
Min discharge (L/m)	CGWB	2001	139	428	458	15.3	2400
Percent hilly (km^2)	CGWB	2001	163	0.0813	0.122	0	1
Percent improved (km^2)	CGWB	2001	163	0.082	0.219	0	1
Percent dug wells (km^2)	CGWB	2001	163	0.019	0.083	0	0.735

Table 2: Demand for Improved Drinking Water Sources

Variable	(1) All hh	(2) No tap	(3) All hh	(4) No tap
Min well depth (100m)	-0.101***	-0.110***	-0.152**	-0.104
	(0.0175)	(0.0200)	(0.0778)	(0.0830)
Min Discharge rate (100L)	0.0158***	0.0135**	0.0268**	0.0263*
	(0.00452)	(5.34e-03)	(0.0129)	(0.0139)
Percent improved only	0.0740***	0.143***	0.659***	0.404**
	(0.0216)	(0.0259)	(0.137)	(0.167)
Percent dug well	-0.348***	-0.114*	0.278*	-0.336**
	(0.0523)	(0.0662)	(0.163)	(0.174)
Percent hilly	-0.0621	-0.240***	0.301***	0.372**
	(0.0663)	(0.0815)	(0.116)	(0.145)
Durable goods	0.0303***	0.0252***	0.0321***	0.0259***
	(0.00403)	(0.00517)	(0.00408)	(0.0516)
Land owned	-1.08***	-7.63***	-0.0988***	-0.00776***
	(2.11e-03)	(2.42e-03)	(2.19e-03)	(2.46e-03)
Social group	0.0288***	0.00870	0.0232***	0.00916
	(0.00716)	(0.00832)	(0.00742)	(0.00847)
Share females	0.0423**	0.0208**	0.0519***	0.0318
	(0.0169)	(0.0203)	(0.0171)	(0.0200)
Village expenditure (1000Rs)	-1.22e-03	-5.17e-03	-4.41e-03**	-4.67e-03**
	(1.87e-03)	(2.31e-03)	(1.93e-03)	(2.40e-03)
Village education	0.0147***	0.00435	0.0259***	0.0200***
	(0.00334)	(0.00409)	(0.00366)	(0.00449)
Fixed Effects	sub-state	sub-state	district	district
District Controls	yes	yes		
Observations	18135	14176	16747	13148
R^2	0.15	0.15	0.19	0.26
F-stat	26.22	27.52	9.14	8.90

Notes: The dependent variable is whether (1) or not (0) a household chooses an improved drinking water source. Columns 1-4 report estimates from linear probability models, with robust standard errors in parentheses. Models in columns 1 and 2 use district groundwater characteristics and models in columns 3 and 4 use within-district groundwater characteristics. Models in columns 1 and 3 exclude households with an in-home tap. Additional variables included in Columns 1-4 are the max well depth, the max discharge rate, the household share < 5, and the max school year. Asterisks indicate statistical significance; *** p<0.01, ** p<0.05, * p<0.1. The F-statistic is reported for the seven instruments.

Table 3: Logit Models of Demand for In-Home Treatment

Variable	(1)	(2)	(3)	(4)	(5)
Improved source	-0.439***	-0.583***	-0.628***	-0.600***	-0.636***
	(0.090)	(0.060)	(0.0605)	(0.102)	(0.102)
Durables index	0.542***	0.470***	0.465***	0.449***	0.443***
	(0.0412)	(0.0350)	(0.0348)	(0.0420)	(0.0416)
Land owned (10 ha)	0.346**	0.200	0.213	0.228	0.248*
	(0.144)	(0.129)	(0.129)	(0.17)	(0.145)
Social group	0.515***	0.325***	0.317***	0.301***	0.295***
	(0.0803)	(0.0633)	(0.0635)	(0.0874)	(0.0873)
School year student	0.00748	0.00854	0.00928	0.00671	0.00743
	(0.0094)	(0.00768)	(0.00769)	(0.00903)	(0.00899)
Share children < 5	-0.144	-0.179	-0.175	-0.216	-0.207
	0.187	(0.168)	(0.168)	(0.150)	(0.151)
Share females	0.00038	-0.00186	-0.0186	0.0106	-0.00214
	(0.157)	(0.139)	(0.139)	(0.138)	(0.137)
Price fuelwood		0.101	0.104	0.0199	0.0130
		(0.0711)	(0.714)	(0.137)	(0.138)
Village expend (1000 Rs)		0.0413**	.0347**	0.0278	0.0225
		(0.0167)	(0.0167)	(0.0323)	(0.0324)
Village education		0.160***	0.147***	0.164***	0.153***
		(0.0275)	(0.0277)	(0.0666)	(0.0619)
Water utility			0.431***		0.351**
			(0.0763)		(0.170)
Fixed Effects	village	district	district	sub-state	sub-state
District Controls				yes	yes
Observations	8313	12945	12945	15047	15047

Notes: The dependent variable is whether (1) a household engages in any in-home treatment or not (0). Columns 1-6 report estimates from logit models, with robust standard errors clustered at the village in parentheses. Models in columns 3 and 5 include a dummy that indicates whether a water utility exists in a village. Additional variables included in Columns 4 and 5 are district controls.
Asterisks indicate statistical significance; *** p<0.01, ** p<0.05, * p<0.1.

Table 4: IV Logit Models of In-Home Treatment

Variable	(1)	(2)
Improved source	-1.31	-3.96**
	(1.66)	(2.04)
Durables index	0.474***	0.532***
	(0.0622)	(0.0721)
Land owned (10 ha)	0.161	-0.139*
	(0.256)	(0.284)
Social group	0.431***	0.417***
	(0.105)	(0.102)
School year student	0.00904	0.0127
	(0.0094)	(0.00990)
Share children < 5	-0.148	-0.0303
	0.162	(0.173)
Share females	0.0196	0.163
	(0.162)	(0.201)
Price fuelwood	0.0565	0.0949
	(0.138)	(0.148)
Village expend (1000 Rs)	0.0210	0.0167
	(0.0363)	(0.0401)
Village education	0.122***	0.249***
	(0.0669)	(0.0835)
Fixed Effects	sub-state	district
District Controls	yes	
Observations	13633	10700

Notes: The dependent variable is whether (1) a household engages in any in-home treatment or not (0). Columns 1 and 2 report estimates from logit models, with robust standard errors clustered at the village in parentheses. Instruments in column 1 are district groundwater characteristics. Instruments in column 2 are within-district groundwater characteristics. Column 2 includes district controls. Asterisks indicate statistical significance; *** p<0.01, ** p<0.05, * p<0.1.

Table 5: Multinomial Logit Model of Demand for In-Home Treatment

Variable	(1) Cloth	(2) Ceramic	(3) Boil	(4) Chemical
Improved source	-0.628***	-0.305	-0.466**	-0.841**
	(0.121)	(0.213)	(0.185)	(0.409)
Durables index	0.333***	0.476***	0.791***	0.625***
	(0.0507)	(0.100)	(0.0816)	(0.135)
Land owned (10 ha)	0.224	0.454	0.168	0.682
	(0.150)	(0.299)	(0.490)	(0.524)
Social group	0.379***	-0.168	-0.0251	0.655**
	(0.0977)	(0.183)	(0.195)	(0.287)
School year student	0.0183*	-0.00756	-0.00213	-0.0566
	0.0111	(0.0205)	(0.0165)	(0.0371)
Share children < 5	-0.184	-0.209	-0.317	-0.0576
	0.177	(0.410)	(0.428)	(0.537)
Share females	-0.0357	0.0294	0.259	-0.395
	(0.167)	(0.405)	(0.366)	(0.459)
Price fuelwood	0.0549	0.0427	-0.108	-0.607
	(0.166)	(0.292)	(0.186)	(0.631)
Village expend (1000 Rs)	0.0323	0.136**	-0.0475	0.141
	(0.0367)	(0.0531)	(0.0600)	(0.141)
Village education	0.242***	-0.115	0.0700	0.250
	(0.0721)	(0.134)	(0.117)	(0.267)

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 village are in parentheses. The base outcome in all columns is no treatment. The number of observations is 15,046 households. Asterisks denote significance; *** p<0.01, ** p<0.05, * p<0.1

Table 6: IV Multinomial Logit Model of In-Home Treatment

Variable	(1) Cloth	(2) Ceramic	(3) Boil	(4) Chemical
Improved source	-1.93	2.88	-8.09**	-6.28
	(2.13)	(3.38)	(4.20)	(10.1)
Durables index	0.360***	0.329***	0.988***	0.776***
	(0.0792)	(0.133)	(0.147)	(0.286)
Land owned (10 ha)	0.0434	1.10*	-0.800	0.0257
	(0.303)	(0.593)	(0.795)	(1.32)
Social group	0.500***	-0.0658	0.236	0.716*
	(0.122)	(0.190)	(0.229)	(0.405)
School year student	0.0248**	-0.0257	-0.00775	-0.0756*
	0.0116	(0.0215)	(0.0179)	(0.0397)
Share children < 5	-0.0887	-0.274	-0.171	0.0647
	0.192	(0.439)	(0.433)	(0.567)
Share females	0.00185	-0.152	0.659	-0.0314
	(0.200)	(0.469)	(0.440)	(0.566)
Price fuelwood	0.0287	0.184	-0.256	-0.627
	(0.173)	(0.365)	(0.211)	(0.713)
Village expend (1000 Rs)	0.0207	0.169***	-0.0930	0.148
	(0.0438)	(0.0614)	(0.0646)	(0.172)
Village education	0.195**	-0.196	0.190	0.219
	(0.0804)	(0.150)	(0.125)	(0.330)

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 village are in parentheses. The base outcome in all columns is no treatment. Instruments are district groundwater characteristics.
The number of observations is 13,632 households. Asterisks denote significance; *** p<0.01, ** p<0.05, * p<0.1

Table 7: Ordered Probit Models on Demand for In-Home Treatment

Variable	(1)	(2) IV	(3)	(4) IV
Improved source	-0.278***	-0.816	-0.253***	-1.28
	(0.0468)	(0.824)	(0.0468)	(0.891)
Durables index	0.255***	0.274***	0.267***	0.275***
	(0.0207)	(0.0312)	(0.0211)	(0.0333)
Land owned (10 ha)	0.0671	0.0286	0.0720	-0.0295
	(0.0616)	(0.129)	(0.0579)	(0.126)
Social group	0.111***	0.177***	0.113***	0.155***
	(0.0403)	(0.0498)	(0.0383)	(0.0470)
School year student	0.000	0.000	0.000	0.00110
	(0.00372)	(0.00448)	(0.00439)	(0.00471)
Share children < 5	-0.120	-0.107	-0.0944	-0.0642
	0.0733	(0.0812)	(0.0747)	(0.0861)
Share females	-0.0189	-0.00183	-0.00608	0.0225
	(0.0680)	(0.0841)	(0.0712)	(0.0969)
Price fuelwood	0.0214	0.0330	0.0184	0.0281
	(0.0648)	(0.0672)	(0.0667)	(0.0722)
Village expend (1000 Rs)	0.0176	0.0103	0.0202	0.0159
	(0.0150)	(0.0174)	(0.0148)	(0.0181)
Village education	0.0530*	0.0422	0.0600**	0.0757*
	(0.0290)	(0.0316)	(0.0301)	(0.0396)
Cut Points				
Cut 1	1.23	0.968	2.36	2.29
Cut 2	2.40	1.97	3.60	3.37
Cut 3	2.61	2.18	3.81	3.59
Fixed Effects	sub-state	sub-state	district	district
District Controls	yes	yes		

Notes: The dependent variable is the mode of in-home treatment where in-home treatment technologies are ordered in ascending quality. Columns 1-4 report estimates from an ordered probit model. Instruments in column 2 are district groundwater characteristics. Instruments in column 4 are within-district groundwater characteristics. Robust standard errors clustered at the village are in parentheses. N=15,046 in columns 1 and 3, N=13,632 in column 2 and N=12,578 in column 4. Asterisks denote significance; *** p<0.01, ** p<0.05, * p<0.1

Table 8: Multinomial Logit Model on Demand for In-Home Treatment

Variable	(1) Cloth	(2) Ceramic	(3) Boil	(4) Chemical
Improved source	-0.660***	-0.334*	-0.583***	-0.827**
	(0.121)	(0.204)	(0.180)	(0.415)
Durables index	0.328***	0.473***	0.784***	0.629***
	(0.0503)	(0.100)	(0.0831)	(0.134)
Land owned (10 ha)	0.241*	0.476	0.156	0.683
	(0.148)	(0.299)	(0.480)	(0.533)
Social group	0.377***	-0.179	-0.0885	0.673**
	(0.0980)	(0.184)	(0.190)	(0.293)
School year student	0.0189*	-0.00674	-0.00200	-0.0589
	0.0111	(0.0208)	(0.0169)	(0.0393)
Share children < 5	-0.176	-0.201	-0.278	-0.0675
	0.175	(0.409)	(0.432)	(0.535)
Share females	-0.0474	0.0193	0.206	-0.402
	(0.167)	(0.403)	(0.355)	(0.463)
Price fuelwood	0.0567	0.0300	-0.129	-0.523
	(0.167)	(0.306)	(0.209)	(0.571)
Water utility	0.313*	0.467	0.890***	-0.171
	(0.181)	(0.417)	(0.244)	(0.629)
Village expend (1000 Rs)	0.0285	0.130**	-0.0782	0.150
	(0.0366)	(0.0532)	(0.0608)	(0.142)
Village education	0.231***	-0.120	0.0558	0.261
	(0.072)	(0.121)	(0.118)	(0.274)

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 village are in parentheses. The base outcome is no treatment.
The number of observations is 15,046 households. Asterisks denote significance; *** p<0.01, ** p<0.05, * p<0.1

Table 9: IV Multinomial Logit Model on In-Home Treatment

Variable	(1) Cloth	(2) Ceramic	(3) Boil	(4) Chemical
Improved source	-1.99	2.83	-7.82**	-6.16
	(2.13)	(3.39)	(3.92)	(10.13)
Durable index	0.355***	0.411***	0.769***	1.69***
	(0.0786)	(0.132)	(0.288)	(0.860)
Land owned (1000 hectares)	0.0607	1.12*	-0.751	0.272
	(0.302)	(0.592)	(0.752)	(1.13)
Social group	0.500***	-0.073	0.160	0.731*
	(0.123)	(0.191)	(0.216)	(0.408)
Max schooling student	0.0256**	-0.0255	-0.00773	-0.0780*
	0.0116	(0.0217)	(0.0183)	(0.0423)
Share children < 5	-0.0857	-0.270	-0.127	0.0514
	0.192	(0.437)	(0.437)	(0.565)
Share females	-0.0151	-0.158	0.586	-0.470
	(0.200)	(0.468)	(0.420)	(0.565)
Price Fuelwood	0.0366	0.178	-0.259*	-0.528
	(0.174)	(0.371)	(0.225)	(0.658)
Water Municipality	0.393**	0.391	0.969***	-0.139
	(0.196)	(0.487)	(0.243)	(0.673)
Per capita expend (1000 Rs)	0.0165	0.165***	-0.121*	0.160
	(0.0436)	(0.0615)	(0.0660)	(0.173)
Village education	0.179**	-0.205	0.171	0.230
	(0.0810)	(0.146)	(0.125)	(0.338)

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 village are in parentheses. The base outcome in all columns is no treatment. Instruments are district groundwater characteristics.
The number of observations is 13,632 households and pseudo-R^2 is 0.36.
Asterisks denote significance; *** p<0.01, ** p<0.05, * p<0.1

Table 10: Water Quality and Expenditure from Behavioral Choices

	(1)	(2)	(3)	(4)
	Unimproved	Improved	Change	Percent Change
Logit				
Constrained Abatement	261 (50.8)	1385 (169)	1124 (176)	70.2%
Choices Abatement	261 (50.8)	1370 (159)	1109 (167)	69.3%
Choices Expend	13.5 (2.64)	10.3 (2.49)	3.20 (3.62)	3.07%
Second Stage Logit				
Constrained Abatement	539 (175)	1454 (580)	916 (606)	57.3%
Choices Abatement	539 (175)	1353 (135)	814 (221)	50.9%
Choices Expend	27.9 (9.08)	6.82 (2.12)	21.1 (12.8)	20.3%
Multinomial Logit				
Constrained Abatement	242 (58.5)	1380 (139)	1137 (154)	71.1%
Choices Abatement	242 (58.5)	1366 (142)	1123 (151)	70.2%
Choices Expend	13.2 (2.88)	9.71 (1.87)	3.53 (3.43)	3.39%
Second Stage Multinomial Logit				
Constrained Abatement	845 (3397)	1477 (5573)	632 (6527)	39.5%
Choices Abatement	845 (3397)	1370 (396)	525 (3420)	32.8%
Choices Expend	82.7 (112)	10.3 (15.4)	72.4 (113)	69.6%

Notes: Quality and expenditure estimates are calculated using estimates from the logit, IV logit, multinomial, 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 and costs are in $ per person. In the constrained model, the coefficient on source (α) equals zero. Choices describes the compensating behavior model. Standard errors presented in parentheses were estimated using the delta method.

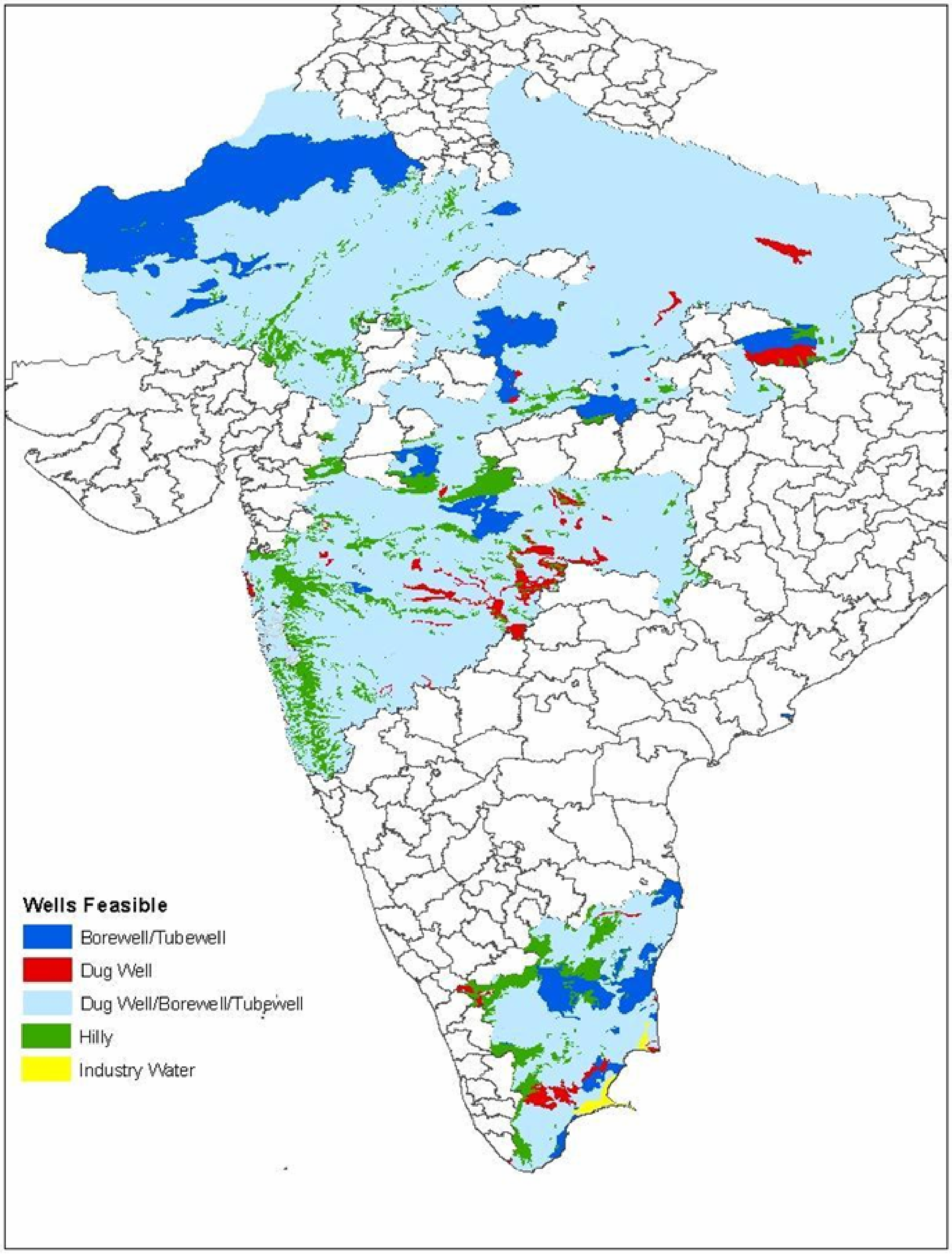

Figure 1: Distribution of groundwater characteristics by well type

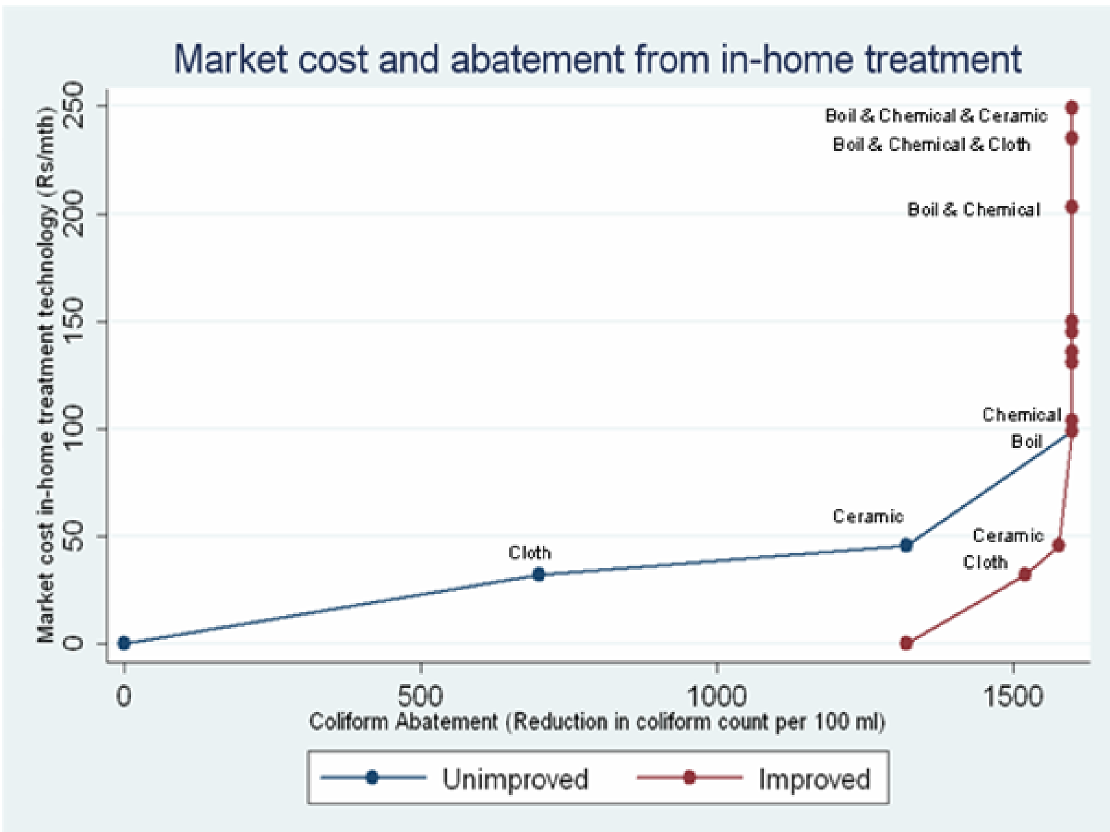

Figure 2: Price and Effectiveness of Treatment Technologies