ESSAYS ON THE TIME USE AND BEHAVIORAL PATTERNS OF WOMEN'S ACCESS TO HOUSEHOLD WATER IN RURAL INDIA

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Shiuli Vanaja

May 2020

ESSAYS ON THE TIME USE AND BEHAVIORAL PATTERNS OF WOMEN'S ACCESS TO HOUSEHOLD WATER IN RURAL INDIA

Shiuli Vanaja, Ph. D.

Cornell University 2020

In this dissertation, three independent research papers are joined together by the common research theme of 'Women's access to household water' in India. The first two papers are based on the self-collected data from selected villages of Jharkhand, India. In these papers, I am studying the role of behavioral factors, like water quality perceptions and water handling practices in determining source choices, drinking water quality, and risk of diarrhea for women in rural households. The third paper uses the data provided by the Indian Human Development Survey and analyses the effect of access to piped water on women's time use and absenteeism from school for children in rural India.

In the first paper, I use panel data collected for 30 villages in Jharkhand, under a household fixed effect model to study the reasons for low drinking water quality and its implications for health costs. The E. coli test results for water samples collected from the field are used as a proxy for water quality. We find that choosing safe water sources improves drinking water quality, but contamination can still happen at home. Thus, the households that use water filters and wash hands tend to have significantly better water quality and lower risk of diarrhea.

In the second paper, I use revealed and stated preference data from Jharkhand to explore the role of perceptions of water quality in drinking water source choices when the physical quality of water is not known. A mixed logit model is used for this analysis. We find that taste and color based quality perceptions are significant determinants of drinking water source choices. Furthermore, households are willing to pay between 4 to 68 and between 2 to 40 U.S. Dollars per month for improvements in the taste and color of their drinking water.

In the third paper, I study the impact of households' access to piped water on time savings and school attendance for children in rural India by using child and household fixed effects along with the non-self-community ratio as an instrumental variable for piped water. The results indicate that there are time savings for families that have access to piped water, which in turn reduces the number of days children miss school annually by 35 percent. We do not find any differences in school absenteeism by the gender of the child in rural India.

BIOGRAPHICAL SKETCH

Shiuli Vanaja was born in Mumbai, India, but grew up in a tribal village in the Hoshangabad district of the Indian state of Madhya Pradesh. Both her parents worked as social-political activists advocating the rights of the indigenous population in that region. After completing her primary education in the government school of the village, she passed the qualifying exam for Jawahar Navodaya Vidyalaya (a prestigious residential school in the district) and left home to study there. She studied in this school until her senior secondary examination (high school). She was fortunate to study under some great mathematics teacher in this school, who helped her develop a strong foundation in the subject. This proved useful while studying economics later during her undergraduate and graduate education. In high school, Shiuli chose to study sciences as her specialization. However, towards the end of her school years, she realized that her interests lay somewhere else, and decided to major in economics in her undergraduate studies. She joined Banaras Hindu University for pursuing a B.A.(Hons) degree in economics. Shiuli's performance in her B.A. was impressive throughout. She received a gold medal for scoring the highest marks in the social science department of the entire university in her senior year. After that, she joined the Center for Economic Studies and Planning at Jawaharlal Nehru University for her master's degree (M.S. or M.A. in economics) in Delhi. During her master's, she was exposed to the various schools and the basic premises of Economics. She learned that, while it is essential to achieve a good grasp of econometrics and mathematical tools to carry out rigorous empirical research, at the same time, it is also necessary to understand the social impact or utility of that research for real-life economics. After her M.A., she decided to pursue an M.Phil. degree in economics from the same university. She spent the next two years researching the effect of deforestation linked to coal mining on the livelihoods of indigenous societies in the Sundargarh district of Odisha, a state in western India. During her M.Phil., she realized that she needed exposure to better research techniques and methods, which led her to join Cornell University for a Ph.D. in applied economics. Joining Cornell University and specifically, Tata Cornell Institute at Dyson school of applied economics and management proved to be life-changing for her academic career. Here she has spent some of her best years learning how to conduct high-quality research, work with peers involved in exciting projects, and learn from some of the leading economists in development economics. The best part of her six-year graduate life experience at Cornell has been the time spent in conducting fieldwork in villages of Jharkhand in India. The eighteen months which she spent in the field made her realize that for engaging in meaningful research in development economics, researchers must spend significant time on the field and incorporate the lessons learned there in their study. This entire fieldwork experience has been possible due to the continuous support of the Tata Cornell Institute for Agriculture and Nutrition at Cornell University. The most significant takeaway of being part of the Cornell community and finishing her Ph.D. degree is that they imbibed her with an innovative research aptitude that will stay with her wherever she goes. It has enriched Shiuli's life experience and prepared her for a career in teaching and research in development economics in the future.

To my father who inspired me to study economics and to my mother who taught me to be a good human being: Sunil & Smita

ACKNOWLEDGMENTS

To begin with, I would like to acknowledge the contribution of the chair of my Ph.D. committee, Prof. Prabhu Pingali, and the other committee members Professors David Just and David Sahn, towards my dissertation work. They helped me with their expert advice, suggestions, and guidance at every step of the way, right from forming the research questions and designing the fieldwork to data collection from the field, analyzing the data, writing papers, and the final defense of my dissertation. They were always available to answer questions, gave me useful advice, critical insights, and pointed out the scope for improvement. I was fortunate to have them in my Ph.D. committee and got a chance to learn from them the finer points of research in applied economics.

It would have been impossible to finish this dissertation without the support of the Tata Cornell Institute for Agriculture and Nutrition. I got funding and all-round support from the Tata-Cornell Institute to conduct an 18-month long field survey in India, which formed the basis for the two papers of my dissertation. The support, aid, and feedback on my research, which I received from TCI members, post-doc students, and administrative staff, played an enormous role in sharpening my research and improving my dissertation work over the years. I would like to thank TCI assistant director Mathew Abraham, post-doc students Anaka Aiyyar and Andaleeb Rahman, administrative staff Mary-Catherine French, Jessica Ames, Sandra A. Stickler, and Brenda Daniels-Tibke in Ithaca for their support and help. TCI provided me with a platform to present and discuss research and a peer group that helped me during the hard times and kept me on my toes when I started to get lax. I am especially grateful to Professor Prabhu Pingali, who, apart from being an outstanding mentor and advisor for me, showed compassion and understood my constraints and emotional strain during the last year of my Ph.D. when my mother was diagnosed with cancer.

There are a lot of people who made the fieldwork experience in India a success. I would like to thank my entire team in Jharkhand, including my field coordinator Thangkhanlal Thangsing (Lal), field supervisor Ashish Paul, survey enumerators Rosemary Ekka, Rajiv Kumar Poddar, Kundan Kumar, and many others. They were the foot soldiers who helped in carrying out this vast field survey exercise. The India experience would be incomplete without mentioning the institutional support I received from the Tata Institute of Social Sciences (TISS), Mumbai, and especially from Dr. Bhaskar Mitra and Maya Nair. I am also thankful to local non-governmental organizations PRADAN and SUPPORT in Jharkhand and their members. They facilitated our initial introduction to different villages in the Khunti and Ramgarh districts. I am very grateful to them for their support, particularly to Satya, Prem, Biglal, Kuldeep Kumar, and Bhawani Shanker Gupta. I cannot thank enough Dr. Karuna Jha and Chandrabhooshan Choudhary (family friends) and the entire family of Sumitra aunty (my next-door neighbor) who provided me with a home away from home in Ranchi, India.

At Cornell, there are quite a lot of people who touched my life in different ways, both academically and personally. At Dyson school, I am grateful to Professors Nancy Chau, Mark Constas, and Shanjun Li for the advice and support which I got from them at different times. I am also thankful to my classmates, juniors, and other students at

Dyson school, with whom I enjoyed many formal and informal encounters over the years. It would be remiss not to admit the exceptional help which I got from my classmate and friend Avralt, especially during the first two years of course work. I have made some great friends during my stay in Ithaca. I am thankful to Xin, Margaret, Maysoon, and Anna for their friendship and unconditional support in times of need.

My best friend from India, Priyanka, was my anchor during my six years in Ithaca, with innumerable phone calls at odd hours bearing witness to a friendship that stood the test of distance and time. I will be forever thankful to my uncle and aunt in New York (later in Boston), Rileen, and Dolly. They provided me with a second home where I could go anytime to relax and reenergize before jumping back into research. I am also thankful to my uncle Amitabh in Florida, who provided me with emotional support during my stay in Ithaca. I would like to thank my parents Sunil and Smita, who have been the pillars of my life and the source of my strength, my brother Iqbal and my extended family in India. Lastly, I would like to thank and fondly remember Dr. Swati (Chotinani), a perfectionist who inspired and motivated me to work hard and to give my best in every sphere of my life.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

QUALITY PERCEPTIONS, AVAILABILITY OF SOURCES AND DRINKING WATER QUALITY IN RURAL INDIA

1.1 Introduction

Diarrhea is the second leading cause of death for children under the age of 5 in the world, with every year, 525000 children dying due to diarrhea (WHO, 2017). Diarrhea and other water-borne diseases could be a major problem for people who do not have access to safe drinking water. Almost one-third of the world population concentrated in the rural areas of South Asia and Africa do not have access to clean drinking water[1]. In the absence of access to safe drinking water, people drink contaminated water. If they fall ill, then they may incur economic costs in terms of health expenses, loss of productive time due to illness, and inability to work. Incurring time and health costs on water-borne illness could lead to the diversion of resources from other household expenses, including education and food. To ensure safe drinking water, both the policymakers and consumers need to know the quality of water, which requires tests for the presence of different disease-causing microorganisms. One proxy for good quality could be the absence of fecal indicator bacteria like E.coli in drinking water. The test results for the presence of E. coli in water can be used for calculating the health risk associated with unsafe drinking water and water-borne diseases.

[1] In 2015, 29 percent of global population lacked access to safe drinking water (Sustainable Development Goals Report, 2018).

The lack of access to good quality water is attributed to the lack of proper water infrastructure liked piped water taps. The type of water source available in a community, whether it is open source or closed, underground or surface, will determine the water quality at the source and ultimately at home. But the drinking water quality at home in many cases could also be dependent on the type of water handling, and water storage practices adopted, including the hygiene and sanitation environment of the house. The studies in the literature that address the question of access to good water quality as a question of lack of proper infrastructure fail to stress the importance of behavioral practices of water handling and home treatment of water that might play an important role in determining the drinking water quality. Furthermore, in many developing countries, there are no established standards for drinking water quality. The tests for drinking water quality are few and are far between. In such cases, people's perception of water quality and preferences over different types of water sources may affect drinking water quality at home.

In this paper, we address three aspects of the problem of access to good quality drinking water and its implications for associated health costs. Firstly, we study the impact of both available water infrastructure and water handling practices on drinking water quality in a household. In the earlier literature, different papers when studying the problem of health costs associated with unsafe drinking water have focused either on the type of sources or behavioral practices adopted inside a house. We conduct a comprehensive study researching the impact of both of these aspects. Secondly, in this paper, we analyze the role of perception of quality in determining objective water quality by affecting the selection of a source from available sources in the community.

This paper is a unique contribution to the literature as it uses a biomatrix for drinking water quality, links water infrastructure with water handling practices and measures of water quality, and calculates the risk of illness in areas where people have underlying perceptions about water quality[2]. In the literature, it is assumed that access to certain source types provides a good quality of water. Instead of making *a priori* assumptions, which may or may not be accurate, we collect water samples from both the source and the household to test it for the presence of E. coli bacteria. We use these results to calculate the health risk associated with water-borne diseases for the households consuming contaminated water.

The structure of the rest of the paper is as follows. The literature review for this paper is being provided in section 1.2. In section 1.3, the pathways for contamination during the process of water collection, storage, and drinking in the Indian rural context are being discussed. The conceptual framework behind this analysis is being explained in section 1.4. In section 1.5, we present the estimation strategy and equations used in this paper. The descriptive statistics and estimation results are being provided in sections 1.6 and 1.7. In section 1.7, we discuss the risk of illness associated with the presence of E. coli in the water for our data set. The last section, 1.9, contains the conclusion and policy implications of the paper.

[2] We find that to the best of our knowledge, this is first such study in the literature.

1.2 Literature Review

United Nations have included 'access to clean drinking water and sanitation' as one of its sustainable development goals (SDG) that are essential for social and economic development. In the literature, studies have found that access to drinking water sources like piped water taps tends to reduce the diarrhea incidence especially for children under the age of 5 in many developing countries (Alves & Belluzzo, 2004; Galiani et al., 2005; Jalan & Ravallion, 2003; Jessoe, 2013; Kremer et al., 2009; Lamichhane & Mangyo, 2011; Naser et al., 2015; Novak, 2014). Thus, economists and policymakers have often advocated building improved water infrastructure like piped water systems in the rural areas for reducing the disease burden and health costs associated with drinking unsafe water. The presence of improved water sources in a community does not necessarily mean that people will use these water sources.

Furthermore, even when households end up choosing an improved water source, behavioral factors inside a house like handwashing and home treatment of water may play a role in determining final water quality before drinking (Allcott & Rogers, 2012; Augier et al., 2016; Ma. et al., 2014; Clasen et al., 2007; Jessoe, 2013; Novak, 2014; Pérez-Flores et al., 2017; Sorenson et al., 2011). In this paper, we investigate the role of improved water sources and behavioral factors like handwashing and home treatment in determining drinking water quality

In the literature, most of the other studies are inconclusive in terms of establishing the link between source improvement and home water quality except by Kremer based in Kenya (Kremer et al., 2009). In some studies, access to improved water sources like piped water taps is shown to be associated with reduced diarrhea and better health outcomes (Galiani et al., 2005; J. Jalan & Ravallion, 2003; Lamichhane & Mangyo, 2011; Naser et al., 2015). But in these studies, the water quality is not measured either at the source or at home. Improved source types are assumed to have better water quality. If there are better health outcomes, then it is assumed that it indicates better water quality at home as a result of access to better types of water sources.

People's perception of water quality often influences the selection of a water source for drinking. In the literature, we find a few papers where the perception of quality plays a role in people's decision to choose between tap water and bottled water[3] for drinking purposes (Etale et al., 2018; Lanz & Provins, 2016). The perception of water quality also plays a role when people decide whether they need to engage in home-treatment of water before drinking (Lanz & Provins, 2016; Pérez-Flores et al., 2017). This focus of this set of literature is on urban areas with semi or fully developed markets for drinking water. This paper fills the gap in the literature by studying the role of quality perceptions of people in a rural area and its influence on their water source choices when there is no developed market for drinking water.

The contribution of this paper to the existing literature is that it provides a comprehensive analysis of the relationship between source types, behavioral practices relating to water storage, home treatment, and sanitation environment inside a house with drinking water quality at home. Instead of piecemeal studies looking at one aspect of this complex relationship, we combine all elements while controlling for the pathways of contamination of water inside and outside a house. Additionally, we

[3] Bottled water is purchased from the market where a market for drinking water exists.

incorporate the role of perception of quality based on color and taste in the choice of a water source when the objective quality of water is not known.

1.3 Pathways for Contamination of Drinking Water in Rural India

In rural India, we find that there is the incidence of large-scale water-borne diseases like diarrhea, absence of access to improved water sources, and widespread norms and perceptions about water quality, making it an ideal setting for this study. Every year in India, about 13 percent of the total child deaths under the age of five are due to diarrhea. The central and state governments have tried to increase access to improved water sources, especially after 2005. However, there are still many people without access to clean drinking water in rural areas. In 2011, about 83 percent of the rural population had access to improved water sources, with about 31 percent having access to piped water taps at home (Census[4], 2011). The water sources available in rural India like handpumps, tube wells, and piped water taps are defined as improved sources as they are covered and draw water from deep underground aquifers through pipes. The title 'improved sources' generally imply that these water sources are safe for drinking. But in practice, it need not be the case (Novak, 2014)[5]. Even if water sources defined as 'improved sources' are safe with no microbial contamination, it can still lead to lower water quality at home depending on the water handling practices, hygiene environment, and water treatment techniques adopted inside a house.

[4] Census of India, 2011. http://censusindia.gov.in/2011-Common/CensusData2011.html

[5] If there is any leakage in underground pipes it may lead to possible contamination of water sources that are generally assumed to be safe

The water sources that are available in rural areas are generally public sources, with very few households having their own private drinking water sources. Most of the time, there are very few water quality tests undertaken to measure and maintain the drinking water quality of these water sources in terms of bacterial contamination of water. Under these circumstances, people make decisions about choosing a water source based on their perceptions of quality, which may or may not match with the objective measures of water quality (Pérez-Flores et al., 2017).

India is a developing country with almost 70 percent of the population living in rural areas. One part of the problem of lack of access to clean drinking water at home in rural India is the lack of water infrastructure providing good quality water. The water sources which are generally available to rural households for drinking purposes are either surface water sources like well, river, streams, and ponds, or underground water sources like hand pumps, borewells. Piped water systems with public or private taps can get water either from a surface water source or underground water source[6].

One common assumption in the literature is that people will always use conveniently located water sources in comparison to the location of their houses inside a village. But people in rural areas, especially in India, have beliefs, perceptions, and social norms around drinking water that may influence their preferences over water sources available in the community.

After the selection of a water source for drinking, households' members fetch water from this source and store it inside the house[7]. The burden of water collection typically

[6] Piped water systems most of the time access water from underground water tables.
[7] In the absence of household water sources at home, the family members travel large distances outside their home to fetch water for drinking purpose

falls on women in rural households. They often make multiple trips during the entire day, depending on the need for drinking water[8] and the storage capacity in their house. The water is generally brought in open containers, carried on their heads, or in hands. The family member fetching water may stop on their way to the house to talk to other people in the village. It is possible that stopping on the way to the house may lead to contamination of water if they leave water containers on the ground, and it comes in contact with animals or small children. People store water in containers inside the house, which are often kept on the ground. There could be multiple pathways for contamination of this stored water. If small children or animals are roaming inside the house, it can increase the possibility of contamination of this water. They are taking out water from the container for drinking using different utensils like glass, mug, jug, or ladle (a bowl attached to a long handle). Except for ladle, all other utensils which are used for taking out water from the container can lead to contamination as it increases the possibility of dirty hands going inside the water. The presence of toilets and the practice of handwashing can affect the cleanliness of hands and thus the quality of drinking water. One way of improving quality is the home treatment of water via boiling or filtering with a water filter. The success of either of these methods depends on the timing and effectiveness of these treatments. If people undertake home treatment of water before storage and the other pathways of contamination relating to storage, water handling, and sanitation are not controlled, then it can still lead to contamination.

[8] The drinking water is also used for cooking in most of the rural houses.

To better understand the relationship between water source choice, water quality at home, and health outcomes, it is important to know the pathways for contamination that generally exists in a rural Indian household. The chart diagram provided below in Figure 1.1 depicts this process from source choice to the final act of drinking.

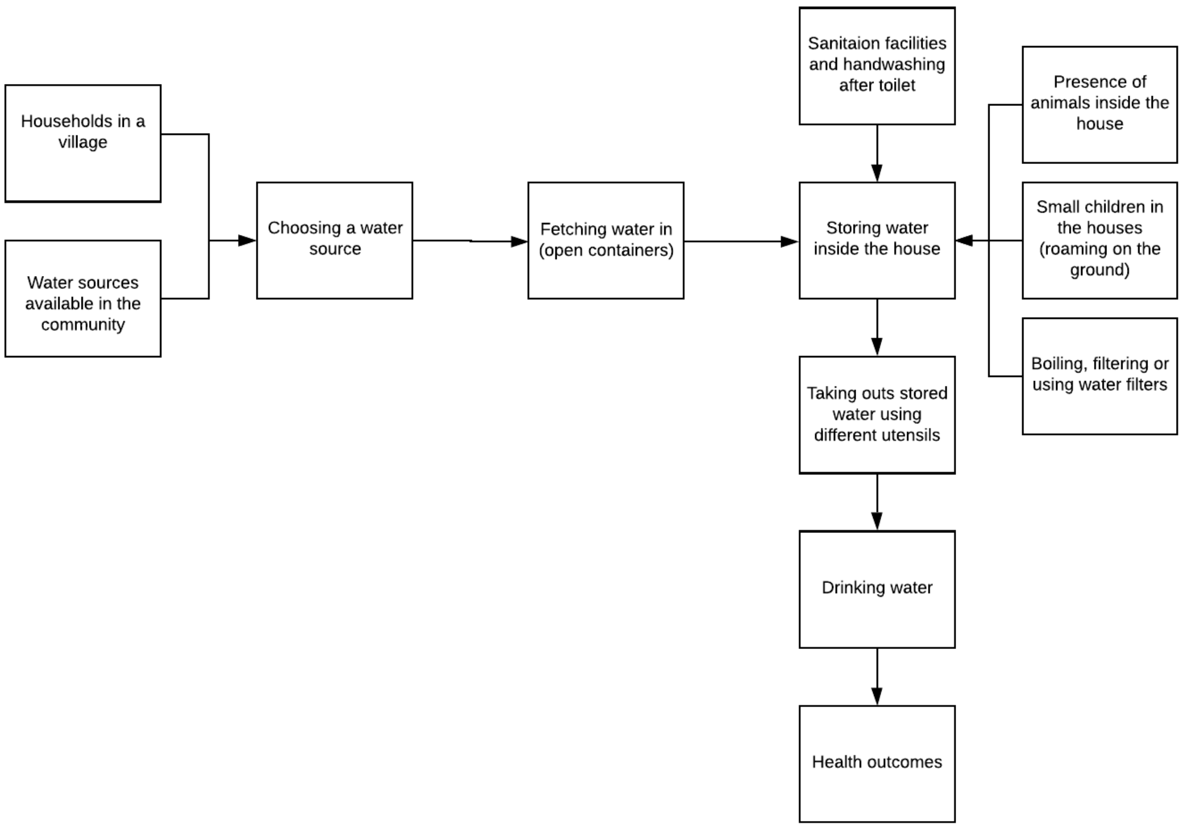

Figure 1.1 Pathways for Contamination of Drinking Water in Rural India

1.4 Conceptual Framework

In this paper, we use a conceptual framework based on utility maximization by a household under certain simplifying assumptions. We consider each household as a unit and ignore intrahousehold decisions and the allocation of resources within a household. For each household, there is a women representative that makes decisions regarding the allocation of resources and time in a way so that it can maximize the total utility for that household. Furthermore, we assume that there is no uncertainty; there is separability of consumption and production functions and the absence of dynamic relationships from period one to another. We also assume that the utility function is concave.

Let's assume that for each household, the representative woman has a utility function as follows.

$$U = u(X, H, q, l)$$

In this function, X is numeraire good (price normalized to one) representing purchase and consumption of different goods by that household, H is the health function of that household, q is the perceived quality of the water[9] consumed by the household and l is the leisure time consumed by the household. The utility function is positive and concave in all (first differential positive and second-order differential is negative) its four components.

Furthermore, we have additional equations defining different elements of the above-defined utility function.

The health of the household H is a function of the objective quality of water Q and other health-related components defined by Z (which we assume to be constant for this

[9] We assume that the perceived quality of water at the source is same as the perceived quality at home and q* is the desired level of water quality at home which could be different from the q or could be the optimum value of q that maximizes the household's utility.

analysis). If the objective quality of water Q increases, the overall health of the household H increases, provided there are no changes in Z.

$$H = H(Q, Z)$$

The objective quality of water Q, in turn, is a function of source quality (or source type-safe or unsafe) S, home treatment of water T, and water handling practices inside a house W_h. We assume that if there is an improvement in source quality, it will improve home water quality Q, keeping everything else constant in this equation. Alternatively, if there is the adoption of home treatment techniques, it will improve water quality at home Q.

$$Q = Q(S, T, W_h)$$

Source quality S, itself, is a function of the distance to the source D_s and perception of water quality q. Distance to the source D_s implies a time cost of water collection. We can denote the time spent on water collection l_w. Larger is the distance to the source higher will be the time spent in water collection.

$$S = S(\ q, D_s)$$

We also define home treatment as a function of the perception of water quality q.

$$T = T(q)$$

Using all these equations, we find that finally, the utility function in this framework is a function of q, l, and l_w,

$$U = U(X, f(q, l_w, C), l)$$

The budget constraint for a household facing above stated utility function would be

$$Y = X + T_c * q + W * l_w - W * l$$

For this analysis, we assume that the labor time spent in water collection l_w , in productive activities l_p , and leisure activities l are provided in such a way that it adds up to the total labor hour available for a household L_t.

$$L_t = l_w + l_p + l$$

The household can make two choices, either to home treat the water incurring the treatment cost of T_c or shift to a safe water source incurring the time spent in water collection l_w with the average opportunity cost of local wages W.

For this analysis, we assume that the household is trying to maximize the utility function by maximizing the leisure or by minimizing the time spent in water collection, given the optimal values of perceived water quality and composite goods X.

So, we write the utility function as

$$U = U(X, f(q, l_w, C), l)$$

$$U = U(X, f(q, l_w, C), (L_t - l_p - l_w))$$

We can also rewrite the budget constraint, so we have

$$Y = X + T_c * q + W * l_w - W * (L_t - l_p - l_w)$$

$$W * l_p = Y$$

$$Y - W *_{l_p} + W * L_t = X + T_c * q + 2W * l_w$$

Let us assume that the right-hand side a monetary constant representing the maximum income, which the household could have if used all the labor hours productively. Then,

$$Y - W * l_p + W * L_t = M$$

$$M = X + T_c * q + 2W * l_w$$

The langrage equation for (if the budget constraint is binding) for this utility maximization problem would be

$$L = u(X, q, l_w) + + \lambda(M - X - \mathrm{T_c} * \mathrm{q} - 2W * l_w)$$

The first-order conditions will provide the following relationships in equilibrium

$$\frac{u_q}{u_x} = T_c$$

$$\frac{u_{l_w}}{u_x} = 2W$$

$$\frac{u_q}{u_{l_w}} = \frac{T_c}{2W}$$

From the first-order equation, we solve to get the optimum values,

$$q^* = \frac{M - X}{2T_c}$$

$$l_w^* = \frac{M - X}{4W}$$

Or

$$\frac{q^*}{l_w^*} = \frac{2W}{T_c}$$

The q^* optimum value of q, is the desired level of perceived quality by a household in this framework. Each household chooses a source and decides on the adoption of home treatment techniques to achieve this desired level of perceived quality q^*. There is a trade-off between the cost of treatment T_c and the opportunity cost of time spent in water collection W. The time spent in water collection l_w is a function of the distance to the source, and higher the distance to the source, higher will be the time spent in water collection.

We observe from the above equilibrium condition that if the cost of treatment T_c goes up, then the perceived quality at home will fall. The household will have to increase the time spent in water collection, maybe by shifting to a water source that is safe but at a larger distance from the concerned household to maintain the same level of perceived quality q^* at home.

On the other hand, if the monetary wages W, goes up, then it will tend to substitute the time spent in water collection towards other productive activities reducing l_w. It will mean shifting to a water source that is nearer to the house, reducing the time spent in water collection. If we want to maintain the desired level of water quality (which can fall by shifting to a nearer but unsafe source), then q must increase presumably by home treatment of water.

If we differentiate the equation of objective quality Q

$$Q = Q(S, T, W_h)$$

with perceived quality q, we get the following relationship,

$$\frac{\partial Q}{\partial q} = S\frac{\partial Q}{\partial T}\frac{dT}{dq} + T\frac{\partial Q}{\partial S}\frac{dS}{dq}$$

By construction, we have $\frac{\partial Q}{\partial T}$ a positive relationship. As the home treatment (assuming that it is effective) takes place, it would tend to increase the drinking water quality at home. Similarly $\frac{\partial Q}{\partial S}$, it is also positive; if we shift to safer sources, i.e., source quality increases, we will tend to have a better objective quality of water at home. If the perceived quality of water fetched from a source goes up, that would mean that the households will be less likely to engage in-home treatment of water before drinking. It means that $\frac{dT}{dq}$ it will have a negative sign. The sign $\frac{dS}{dq}$ is less clear. It will depend on whether households are correctly perceiving the water quality at the source or not. If the perception of quality is based on some factor that is systematically associated with lower source quality (S), then $\frac{dS}{dq}$ it will have a negative sign. On the other hand, if the perception of quality is based on some factor that is systematically associated with higher source quality (S), then $\frac{dS}{dq}$ it will have a positive sign. If $\frac{dS}{dq}$ it has had a negative sign, then $\frac{\partial Q}{\partial q}$ it will be negative, indicating that the objective quality of water will fall with improvement in the perceived quality of water. Alternatively, if $\frac{dS}{dq}$ it has had a positive sign, then $\frac{\partial Q}{\partial q}$ it will be positive or negative depending on the relative strength of the effect of home treatment on water quality. It will happen because home treatment will fall with an increase in q as compared to the effect of source quality on water quality at home.

1.5 Estimation Strategy

The chart diagram[10] presented in Figure 1.2 indicates the relationships that are being analyzed in this paper. These relationships link the perception of quality to source choice and objective measure of water quality with implications for associated health costs.

[10] Only the first three boxes in this chart diagram are being analyzed in this paper.

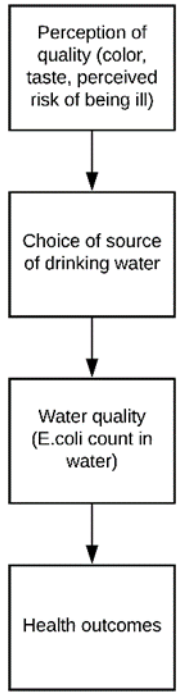

Figure 1.2 Relationship between Perception, Source and Water Quality

There are two sets of equations that define the relationship between quality perceptions, water source type, and drinking water quality at home. The first reduced form equation is based on the estimation of the selection of water sources (safe or unsafe sources based on the E. coli count in water) by a household using a household and year fixed effect model.

$$S_{it} = \beta_0 + \beta_1 q_{ijt} + \beta_2 X_{it} + \beta_3 D_{it} + \lambda_t + \mu_i + \epsilon_{it} \quad (1)$$

In this equation S_{it} represents source type proxied by water quality at the source[11]. The other variables are, q_{ijt} representing different quality perception-based indicators, D_{it} distance to the source, and X_{it} other household-specific control variables like wealth (asset level) of the household, female education level, number of children in the family, etc.

The second equation in this set up is based on the estimation of drinking water quality using household and year fixed effect model.

$$Q_{it} = \alpha_0 + \alpha_1 S_{it} + \alpha_2 T_{it} + \alpha_3 P_{ijt} + \alpha_4 X_{it} + \alpha_5 q_{ijt} + \alpha_6 D_{it} + \lambda_t + \mu_i + \varepsilon_{it} \qquad (2)$$

In this equation Q_{it} represents the objective water quality measures at stored water at home[12], S_{it} source type measured by source quality, T_{it} adoption of home treatment techniques at home before drinking water, P_{ijt} represents a set of water handling practices including handwashing, q_{ijt} representing different quality perception-based indicators, D_{it} distance to the source and X_{it} different households specific control variables, as explained above.

The underlying hypothesis tested in this estimation framework of two interrelated reduced form equations are as follows. Does the perception of quality affects the selection of water sources for a household and if source choice (water quality at source) affects the drinking water quality at home along with water handling practices? We use the household and year fixed effect model, controlling for the

[11] S is measured here as a continuous variable indicating the presence of E. coli count at the source. If it is zero that indicates safe water sources, if it is greater than zero then indicating choice of unsafe water source.

[12] Objective measure of quality is also measured by the E.coli count in the stored water at home. If it is zero that indicates safe water sources, if it is greater than zero then indicating choice of unsafe water source.

household-specific and year specific unobserved variables that are not changing over time[13] to estimate these two equations.

1.6 Data Collection and Field Survey Design

An eighteen-month long field survey was conducted across 30 villages in two districts of Jharkhand state in India to collect panel data for two years. Two districts Khunti and Ramgarh were chosen that were about 60 Km from the state capitol Rachi for this field survey. Khunti is predominantly an indigenous population dominated district while Ramgarh has a more mixed social composition (caste wise). After the selection of these two districts, 15 random villages were chosen from each district. In each village, 30 households were randomly selected and surveyed. The field surveys were conducted by trained enumerators who knew the local language. In each household, a representative woman was interviewed to get details about the entire household. From each household, water samples were collected for testing the presence of E. coli bacteria in the water. Similarly, water samples were collected from each water source in the village that is being used for drinking purposes. In the second year (2017-18), the households that were surveyed in the year 2016-17 were revisited to conduct interviews and collect water samples. The timeline for the data collection is being provided below by Figure 1.3.

[13] If there are any other time and household specific unobserved variables that might affect the source choice by a household, then it may create bias in the system.

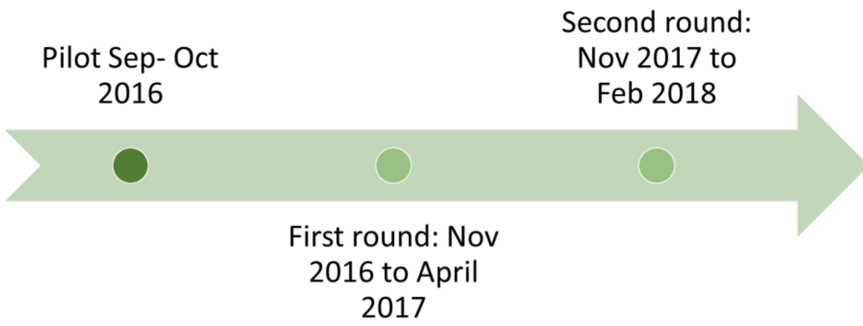

Figure 1.3 Timeline of Field Data Collection

1.7 Estimation Results

1.7.1 Descriptive Statistics

In Table 1.1, we present the summary statistics for households in years 1 and 2[14] and the difference of means between years 1 and 2. From Table 1.1, we can observe that the proportion of households accessing safe water sources is around 50 percent in our data sets, and it remains the same across years 1 and 2.

On the other hand, the awareness about the quality of water has changed between the years. About 50 percent of households reported that they understood the meaning of good quality water in the year 2016-17, and this percentage changed to about 90 percent in the next year. The percentage of households engaging in-home treatment has gone up from 60 to 70 percent. At the same time, the proportion of households having toilets at home increased from 40 percent to 70 percent in the year 2016-17. The average family size is the same across the two years, but the proportion of

[14] For this discussion, we refer year 2016-17 (the first round of survey) as year 1 and year 2017-18 (the second round of survey) as year 2.

children under the age of five shows a small difference. The presence of animals inside a home has gone down with only 30 percent of sample households reporting the presence of animals inside their house in year two as opposed to 50 percent in year one. We find a significant deterioration in average drinking water quality from year one to year two, as represented by the increase in the E. coli count found in stored water at home.

Table 1.2 provides the summary statistics of the sample data after dividing it into two groups of households accessing safe and unsafe water sources[15]. As per expectation, the water quality measured at home is significantly better (low E. coli count) in households using safe drinking water sources. Among the households selecting safe water sources, about 60 percent of the households adopt home treatments, while about 70 percent of the households choosing an unsafe source do the same. The higher percentage of households reporting home treatment techniques while choosing unsafe sources could be due to two reasons.

Either they know that they are using an unsafe source and want to compensate for that by adopting the home treatment method. Another explanation could be that filtering by cloth is adopted as a method of home treatment by the households choosing unsafe sources like well[16]. This method of home treatment may harm water quality instead of improving it.

[15] We divide households in safe source or unsafe source category based on the E. coli count found at the samples collected from their primary source of drinking water.

[16] In the field, a lot of people reported that they prefer filtering with cloth for the water collected from wells due to the possible presence of leaves, etc.

Table 1.1 Summary Statistics Based on the Survey Rounds (Years)

Variable names	Year = 1	Year = 2	Difference	$t-stat$	$p-value$
Ecoli colonies per 100ml water(quality)	47.2	58.5	−11.28***	−4.79	0.00
Safe source,=1 yes	0.5	0.5	−0.00	−0.10	0.92
Understands quality,=1 if yes	0.5	0.9	−0.38***	−18.09	0.00
Q perception (color), =1 if satisfied with color	0.7	0.9	−0.23***	−11.04	0.00
Q perception (taste), =1 if satisfied with taste	1.0	0.9	0.02	1.58	0.11
Q perception (disease risk), =1 if believe less risk	0.3	0.3	−0.01	−0.25	0.81
Using ladle	0.2	0.2	−0.02	−1.31	0.19
Home treatment	0.6	0.7	−0.08***	−3.34	0.00
Wealth	5.0	5.2	−0.13	−1.02	0.31
Education (female representative)	5.3	5.6	−0.35	−1.40	0.16
Family size	5.8	5.8	−0.05	−0.46	0.65
Proportion of children under 5	0.1	0.1	0.02*	2.28	0.02
Animals inside home, if yes=1	0.5	0.3	0.18***	7.43	0.00
Toilet at home	0.4	0.7	−0.25***	−10.12	0.00
Total number of water sources in each village	11.9	13.4	−1.48***	−6.93	0.00
Distance to primary source of water	102.1	89.4	12.74*	2.02	0.04
Well (open)	0.4	0.4	−0.02	−0.83	0.41
Hand pump	0.5	0.5	0.01	0.56	0.57
Piped water at home	0.1	0.1	0.01	0.78	0.44
Public tap	0.0	0.0	0.00	0.00	1.00
River	0.0	0.0	0.00	0.00	1.00
Borewells	0.0	0.0	−0.00	−0.66	0.51
Type of source (=1 private)	0.3	0.2	0.04	1.87	0.06
Caste category,= 1 upper	3.2	3.2	0.00	0.00	1.00

Note:*se* are in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% level respectively.
Difference indicates the difference of mean values of households in year 2 from year 1.

There is no significant difference in awareness about drinking water quality between the people choosing safe or unsafe drinking water sources in Table 1.2[17]. Still, the percent of households with the perception of quality based on the color of the water is significantly higher in the group of households choosing unsafe water sources (90 percent as opposed to 70 percent). It indicates a mismatch between quality perception based on color and objective measure of quality based on E. coli count in the water. We will discuss this in more detail in the next section of this paper.

[17] In both the groups about 70 percent of the households reported that they understand meaning of water quality.

Table 1.2. Summary Statistics Based on the Quality of the Water at the Source

Variable names	Safe = 0	Safe = 1	Difference	$t-stat$	$p-value$
Ecoli colonies per 100ml water(quality)	73.4	31.6	41.75***	19.73	0.00
Understands quality,=1 if yes	0.7	0.7	−0.04	−1.59	0.11
Q perception (color), =1 if satisfied with color	0.9	0.7	0.21***	10.07	0.00
Q perception (taste), =1 if satisfied with taste	1.0	0.9	0.01	1.16	0.24
Q perception (disease risk), =1 if believe less risk	0.3	0.3	−0.05*	−2.25	0.02
Using ladle	0.2	0.2	−0.02	−1.24	0.22
Home treatment	0.7	0.6	0.09***	3.81	0.00
Wealth	5.1	5.1	−0.07	−0.54	0.59
Education (female representative)	5.3	5.6	−0.30	−1.20	0.23
Family size	5.9	5.7	0.19	1.60	0.11
Proportion of children under 5	0.1	0.1	0.01	0.75	0.46
Animals inside home, if yes=1	0.4	0.4	−0.00	−0.02	0.98
Toilet at home	0.6	0.6	0.00	0.16	0.87
Total number of water sources in each village	12.6	12.8	−0.27	−1.26	0.21
Distance to primary source of water	123.5	67.3	56.21***	9.13	0.00
Well (open)	0.8	0.0	0.77***	48.64	0.00
Hand pump	0.1	0.9	−0.77***	−47.76	0.00
Piped water at home	0.1	0.1	0.01	1.04	0.30
Public tap	0.0	0.0	0.00	0.98	0.33
River	0.0	0.0	0.01**	2.81	0.01
Borewells	0.0	0.0	−0.02***	−3.36	0.00
Type of source (=1 private)	0.3	0.2	0.16***	7.54	0.00
Caste category,= 1 upper	3.3	3.1	0.19***	3.70	0.00

Note:*se* are in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% level respectively.
Difference indicates the difference of mean values of safe source household from unsafe ones.

In the literature, it is often assumed that the water sources defined as 'improved' will have better quality. We test this hypothesis by regressing source quality (measured by the E. coli count at the water sample collected from the source) against different source types available in our field area. The results of this regression are provided in Table 1.3. We find that the water sources assumed to be unsafe, like well and river, on average tend to have higher E. coli count and lower water quality. But for the sources assumed to be 'improved' like handpumps, piped water taps at home and borewells, there are cases when water from these sources are found to have high E. coli count and low quality. One reason for this is that sometimes there is fecal contamination of these sources due to leakage from toilet tanks into underground tables and the pipes drawing water from these underground tables.

We found that, on average improved sources like hand pump and piped water taps at home have lower E. coli count compares to wells (constant term in Table 3). But the overall effect of the hand pump and piped water on the water quality (E. coli count) is not as expected[18]

Table 1.3. Effect of Source Types on Quality of Water at Source

Dep. variable = Ecoli count for water at source (log)	FE
Hand pump	−3.798***
	(0.068)
River	−0.443
	(0.374)
Piped water at home	−2.218***
	(0.135)
Public tap	−0.759
	(0.513)
Borewells	−3.462***
	(0.235)
Constant	4.131***
	(0.234)
Village fixed effects	Yes
N	1530.00
R^2	0.71

Note:*se* are in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% level respectively. Reference category (Constant) is open well.

In the next part of this section, we present five graphs depicting the relationship between water source types, water handling, and treatment practices with the E. coli based measure of water quality. In Figure 1.4, we show that accessing different kinds

[18] The overall effect of Hand pump is calculated by adding the constant term with the coefficient on Hand pump. As the constant term is bigger than the coefficient on Hand pump, the overall effect of Hand pumps on E. coli count at the source is positive. The same analysis is valid for piped water taps at home and borewells.

of water sources for drinking purposes has a disparate impact on the probability of good quality water at home. Using open sources like well have a low probability, while using handpumps have a high probability of resulting in good quality water at home. In Figure 1.5, we differentiate between the effect of different water sources on the good quality water at home between the households that engage in handwashing practices and those that do not. We find that the adoption of handwashing practices increases the probability of that household having better quality water at home. It is true across all types of water sources.

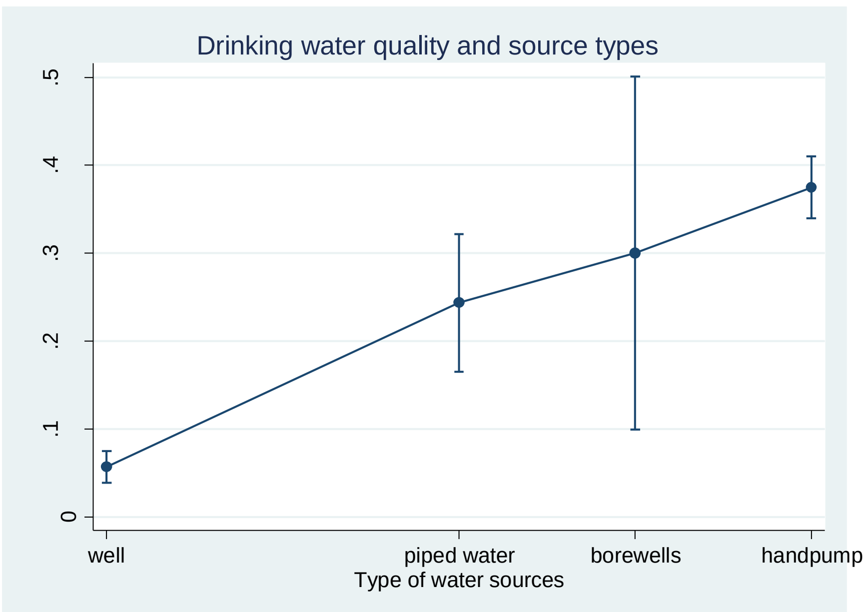

Figure 1.4. Effect of Types of Water Sources on Drinking Water Quality

In Figure 1.6, we include the use of ladle along with handwashing practices to analyze the effect of these practices on the quality of water across different water sources. It can be observed from Figure 1.6 that the households that have adopted both handwashing and use of ladle in their house have the highest probability of having good quality water at home. On the other hand, households that do not engage in handwashing practices or use a ladle for taking out water tend to have a low quality of water at home. This effect can be observed across all water sources, but it is more pronounced in relatively safer water sources like handpumps. The effect of different home treatment methods on water quality at home is being provided in Figure 1.7. It can be observed from this graph that among different methods of home treatment filtering by machine has the highest probability of resulting in good quality water at home. The method of boiling water before drinking also improves the chances of having good quality water as compared to no treatment of water.

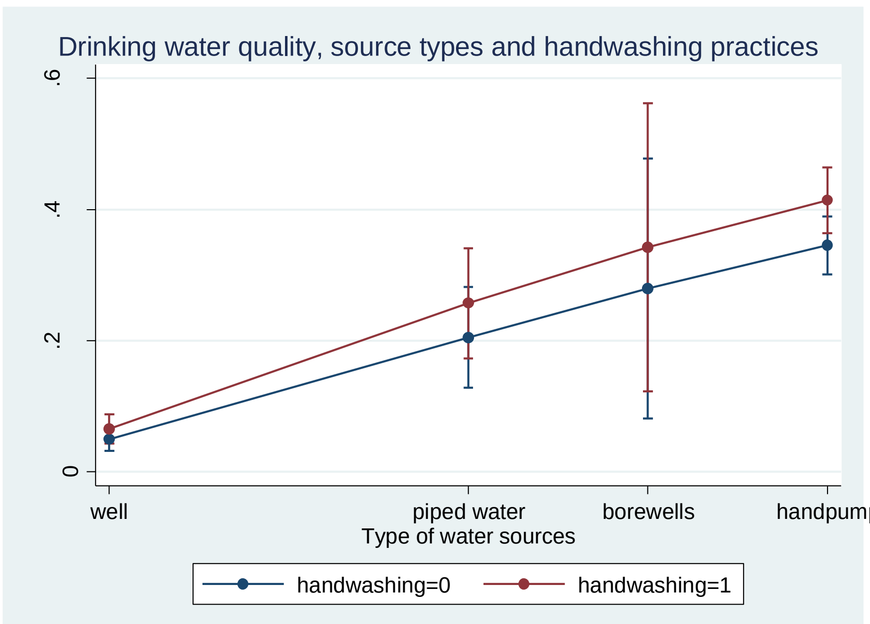

Figure 1.5. Effect of Handwashing Practices on Water Quality across Different Water Sources

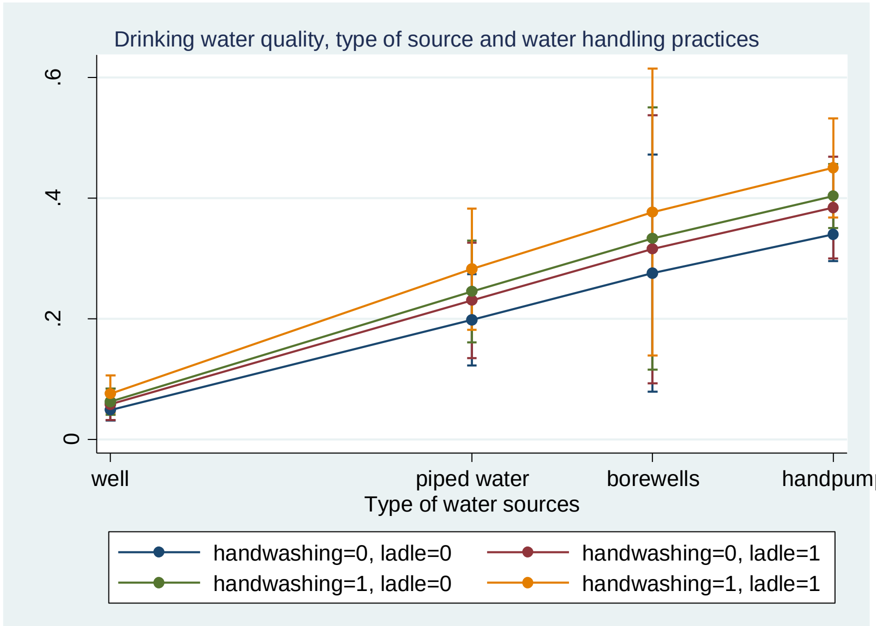

Figure 1.6. Effect of Water Handling Practices on Water Quality Across Different Water Sources

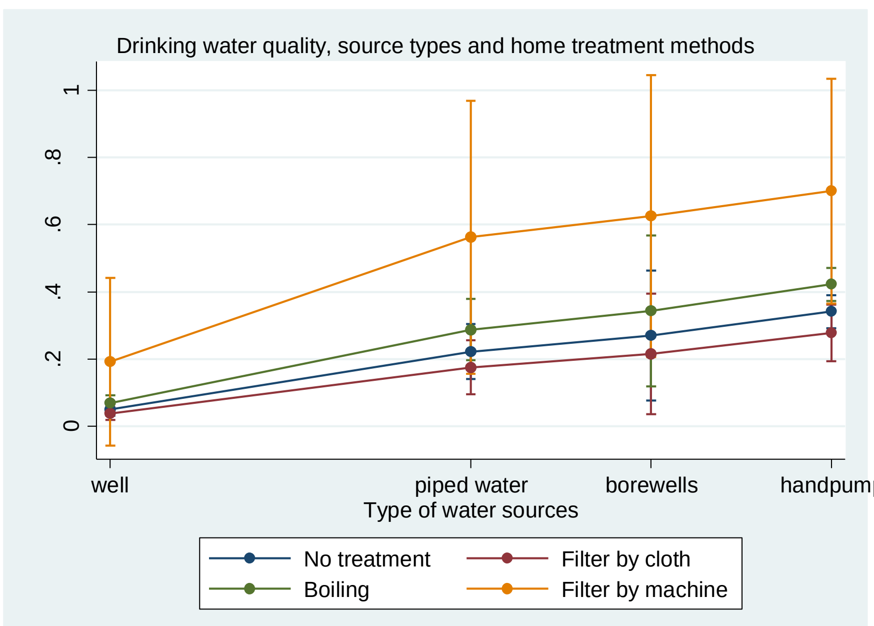

Figure 1.7. Effect of Home Treatment Methods on Water Quality Across Different Water Sources

Interestingly, we observe that filtering with a cloth reduces the probability of having good quality water at home as compared to the households without any treatment of water. It suggests that it would have been better for the households to not engage in any treatment than use the filtering of water with a cloth. These results and trends are valid across different types of water sources available in a community.

In Figure 1.8, the effect of distance to the source on the quality of water at home is being presented. We can observe that shorter distance to the source increases the probability of having good quality water at home. There could be two possible reasons for this result.

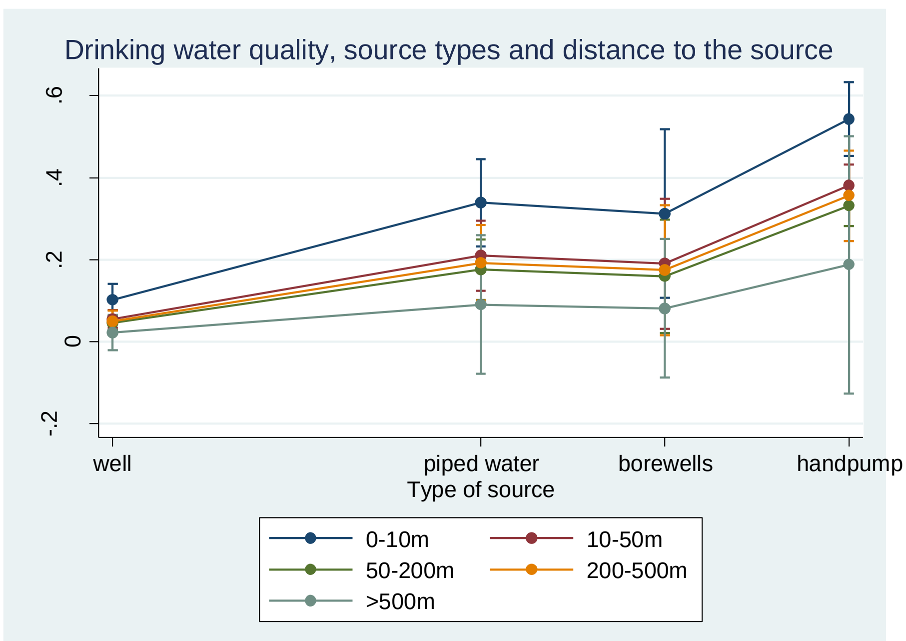

Figure 1.8 Effect of Distance to the Source on the Quality of Water Across Different Water Sources

A shorter distance to the source implies fewer chances of contamination of water during the travel from the source to the house. Secondly, a shorter distance to the source might encourage people to make frequent trips to the source. That may lead to frequent handwashing practices as well as changing the water stored inside the house, thereby reducing the contamination of water. Observing the trend across different source types, we find that the positive effect of shorter distance on the good quality of water is higher in relatively safer sources like handpumps. Additionally, the difference in the effect of distance to the source does not vary much across different sources when the distance to the source is high, i.e., more than 500 meters.

1.7.2 Regression Results: Determinants of Drinking Water Quality

The estimation results for the regression of source type (proxied by source quality) on different quality perception indicators, distance to the source, and several household level control variables are provided in Table 1.4. In the first two columns, we provide household fixed effects, and in the last two columns, village fixed effects results are being provided. It can be observed that as the distance to the source of drinking water increases, the E. coli count at the source also increases, making it unsafe. This significant positive relationship remains there even when we control for other variables. One possible reason for this positive relationship is that often sources like wells and rivers are situated outside the village near the agricultural fields. At the same time, hand pumps, borewells, and piped water taps are located inside the village. Since well and river generally have a higher E. coli count (as can be observed from Table 1.3), this could explain why the increase in distance tends to imply that the household is choosing a relatively unsafe water source.

We use three quality perception indicators in this analysis, based on the color of water, the taste of water, or the perceived risk of not falling ill (diarrhea). All three of them are binary indicators. We are interested in the interaction of these variables with the variable measuring quality awareness[19].

From Table 1.4, it can be observed that the households that understand water quality and whose perception of water quality is based on the color of water tend to have, on average, a higher E. coli count at the source. It means that the households who decide their primary source of water based on the color of the water tend to choose relatively unsafe water sources. The reason behind this lies in the fact that in rural Jharkhand, there is little turbidity in water collected from wells[20][21] while the pipes used in hand pumps or piped water systems are often of low quality. The pipes used are often old and get rusted easily. It could create a red color in the water from hand pumps and piped water taps.

Also, if the pipes in the piped water system or borewells get broken, it can lead to a mixing of mud into pipes creating a brownish color in the water. Under these circumstances, households choosing water sources based on the color of the water may tend to choose water sources like well that is relatively unsafe. On the other hand, the households that understand the meaning of water quality and choose water sources based on the taste of water tend to choose water sources with lower E. coli count at the source. We do not find any impact of the perception variable based on the perceived risk of being ill on the quality of water at the source.

---

[19] Labeled as 'Understands quality' in the above Table 1.4.
[20] Except maybe in some short period of Monsoon.
[21] As our data was collected over winter and spring seasons in both the survey rounds, we do not face this problem.

Table 1.4: Effect of Perception of Water quality on Source Choices

Dep. variable = Ecoli count at source (log)	Model A	Model B	Model C	Model D
Q perception (color), =1 if satisfied with color	0.008	0.263	0.721***	0.068
	(0.212)	(0.584)	(0.214)	(0.577)
Q perception (taste), =1 if satisfied with taste	1.484**	0.629	1.407**	2.543**
	(0.458)	(1.034)	(0.459)	(0.784)
Q perception (disease risk), =1 if believe less risk	0.442	0.836	0.161	0.476
	(0.282)	(0.573)	(0.285)	(0.532)
Asset index	−0.016	0.524	−0.845	−0.565
	(1.022)	(1.024)	(0.750)	(0.940)
Education (female representative)	0.062	0.083	−0.006	0.091
	(0.064)	(0.077)	(0.024)	(0.048)
Distance to primary source of water	0.002**	0.002**	0.003***	0.003***
	(0.001)	(0.001)	(0.001)	(0.001)
Understands quality,=1 if yes	0.976*	0.988*	0.964*	0.931
	(0.479)	(0.482)	(0.479)	(0.483)
Understands quality*Q perception (color)	0.636**	0.655**	0.525*	0.552*
	(0.237)	(0.239)	(0.248)	(0.250)
Understands quality*Q perception (taste)	−1.512**	−1.543**	−1.359**	−1.344**
	(0.478)	(0.480)	(0.481)	(0.483)
Understands quality*Q perception (disease_risk)	−0.108	−0.190	−0.181	−0.196
	(0.253)	(0.255)	(0.252)	(0.253)
Asset index*Q perception (color)		−0.466*		−0.011
		(0.203)		(0.198)
Asset index*Q perception (taste)		0.375		−0.021
		(0.224)		(0.223)
Asset index*Q perception (disease_risk)		−0.011		−0.000
		(0.171)		(0.168)
Household fixed effects	Yes	Yes	No	No
Village fixed effects	No	No	Yes	Yes
Interaction terms included	*No*	*Yes*	*No*	*Yes*
N	1453.00	1453.00	1453.00	1453.00

Note:*se* are in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% level respectively.

The result from the estimation of equation (2) is being provided in Table 1.5. In this equation, the drinking water quality at home[22] is being regressed on the source type (proxied by source quality), water handling practices like handwashing, quality perception indicators, and distance to the source. We find that in all the four

[22] As measured by the presence of E. coli in drinking water stored at home.

formulations of household and village fixed effects[23], the source quality has a positive impact on drinking water quality at home. In other words, if there is lower E. coli count at the source water (better water quality at the source), this tends to lower E. coli count at stored water at home (better water quality at home).

Additionally, we find that the practice of handwashing after using a toilet tends to lower E. coli count, i.e., improve water quality at home. Different households can adopt different home treatment techniques to improve drinking water quality at home. In our data, households adopt either boiling of water, filtering with a cloth, or using filters before drinking water if they choose to conduct home-treatment of water. It can be observed from Table 1.5 that in the household fixed effect model home-treatment variable has a positive sign indicating that it tends to increase the E. coli count (reducing water quality at home), but the relationship is not significant.

In column 3 of Table 1.5, in a village fixed effect model, home treatment tends to weakly reduce the E. coli count at stored water at home. In columns 2 and 4 of Table 1.5, we look at the effect of different types of home treatment techniques like boiling water, filtering with cloth, and using filters on E. coli count in the water at home. We find that in both household and village fixed effect models, using filters tends to significantly reduce the E. coli count at home (improve drinking water quality). But only in the village fixed effect model, the treatment method of boiling water reduces E. coli count in stored water at home significantly.

[23] The results for these 4 formulations are provided in the 4 columns of Table 5.

Table 1.5. Effect of Source Types (Quality) and Water Handling Practices on Water Quality at Home

Dep. variable = Ecoli count at home (log)	Model A	Model B	Model C	Model D
Asset index	1.282	−1.062	−0.456	−0.289
	(1.564)	(1.479)	(0.763)	(0.934)
Education (female representative)	0.001	−0.069	0.014	0.006
	(0.094)	(0.115)	(0.024)	(0.049)
Distance to primary source of water	0.004	0.004	0.002	0.002
	(0.002)	(0.002)	(0.001)	(0.001)
Proportion of children under 5	1.117	1.040	0.496	0.476
	(0.609)	(0.613)	(0.261)	(0.261)
Understands quality,=1 if yes	−0.470	−0.664	−0.028	−0.194
	(0.715)	(0.737)	(0.495)	(0.507)
Understands quality*Q perception (color)	0.101	0.142	−0.143	−0.120
	(0.345)	(0.349)	(0.243)	(0.246)
Understands quality*Q perception (taste)	0.744	0.907	0.186	0.304
	(0.708)	(0.725)	(0.496)	(0.505)
Using ladle	0.007	0.012	0.017	0.022
	(0.229)	(0.230)	(0.133)	(0.132)
Ecoli count at source(log)	0.321***	0.311***	0.408***	0.405***
	(0.058)	(0.059)	(0.027)	(0.027)
Washing hands after toilet	−0.560**	−0.552**	−0.309**	−0.309**
	(0.192)	(0.192)	(0.107)	(0.107)
Home treatment	0.140		−0.208*	
	(0.170)		(0.105)	
Fetching water more than once	0.323	0.382	0.452	0.419
	(0.423)	(0.427)	(0.245)	(0.247)
Distance*Fetching water more than once	−0.002	−0.002	−0.001	−0.001
	(0.002)	(0.002)	(0.001)	(0.001)
Boiling drinking water		0.064		−0.286**
		(0.179)		(0.110)
Filetring water with cloth		0.390		0.035
		(0.247)		(0.151)
Using filters (containers)		−2.643**		−1.830*
		(1.010)		(0.712)
Asset index*Q perception (color)		−0.474		−0.418*
		(0.294)		(0.197)
Asset index*Q perception (taste)		0.540		0.223
		(0.321)		(0.219)
Asset index*Q perception (disease_risk)		0.226		0.051
		(0.245)		(0.167)
Household fixed effects	Yes	Yes	No	No
Village fixed effects	No	No	Yes	Yes
Interaction terms included	*No*	*Yes*	*No*	*Yes*
Q indicators included	*Yes*	*Yes*	*Yes*	*Yes*
N	1362.00	1362.00	1362.00	1362.00

In columns 2 and 4 of Table 1.5, we present results relating to the three quality perception indicators (based on color, taste, and risk of being ill) and their interactions with the wealth and education level of the female representative from sample households. All the three quality perception indicators, as well as their interaction with quality awareness variable, are not significant in this regression[24]. If we combine these results with the results in Table 1.4, it seems that the effect of the quality perception indicators on water quality at home operates indirectly via source choice (source quality).

1.8 Health Risk Associated with Unsafe Drinking Water

Different water-borne diseases are often a result of different types of microorganisms in the drinking water. Ideally, one would like to measure pathogen levels of different disease-causing microorganisms to calculate the risk of illness associated with drinking contaminated water. Unfortunately, it is not always feasible to identify the presence of different pathogens in drinking water[25]. According to the world health organization guidelines, the presence of a fecal indicator bacteria like E. coli can be used as a proxy for drinking water quality as it is easy to test and cost-effective in terms of testing. The presence of fecal indicator bacteria like E. coli often indicates the presence of other pathogens in water. The presence of disease-causing microorganisms in the water increases the probability of infection and disease, but the outcome

[24] We have not presented the coefficient on these terms due to space restricts but they are included in this regression.
[25] It is often found that the presence of pathogenic micro-organisms in water is erratic and testing for it is costly and not straight forward.

depends on the immune system of a particular individual. The risk of illness lies in the probability distribution of the hazard represented by disease-causing microorganisms in water.

The acceptable value of E. coli in drinking water is zero. The presence of E. coli in water above the levels of zero can cause infection and may constitute a health risk for the concerned individual. We assume the single hit and independent action theory of infection (which means that even a single microorganism can cause infection and no threshold limit is needed). For proper Quantitative microbiological risk assessment (QMRA), we need to conduct tests in labs for different pathogens in the drinking water. As we do not have this data, we use measured values of E. coli[26] in the water as a proxy for the possible presence of pathogens to get an idea about the risk of illness. We use a dose-response function as often used in the epidemiology literature. The dose is calculated by multiplying the drinking water consumed per day with the value of E. coli found in the water sample collected from a household. We divide this value of dose with the total number of family members in a household to get the average dose consumed in a day[27]. After calculating the value of the average daily dose (intake of E. coli), we run a regression of disease incidence on the dose[28]. The predicted value of y variable (disease incidence as a binary variable) is the probability of illness that is being plotted in Figure 1.9.

[26] These values of E. coli bacteria in the water are calculated using the most portable number (MPN) method.

[27] There are also consideration of weight, age group and source types while calculating dose. We are ignoring all these for the time being.

[28] We are using linear regression, but we can try exponential, beta-poisson and logisting regression for further analysis.

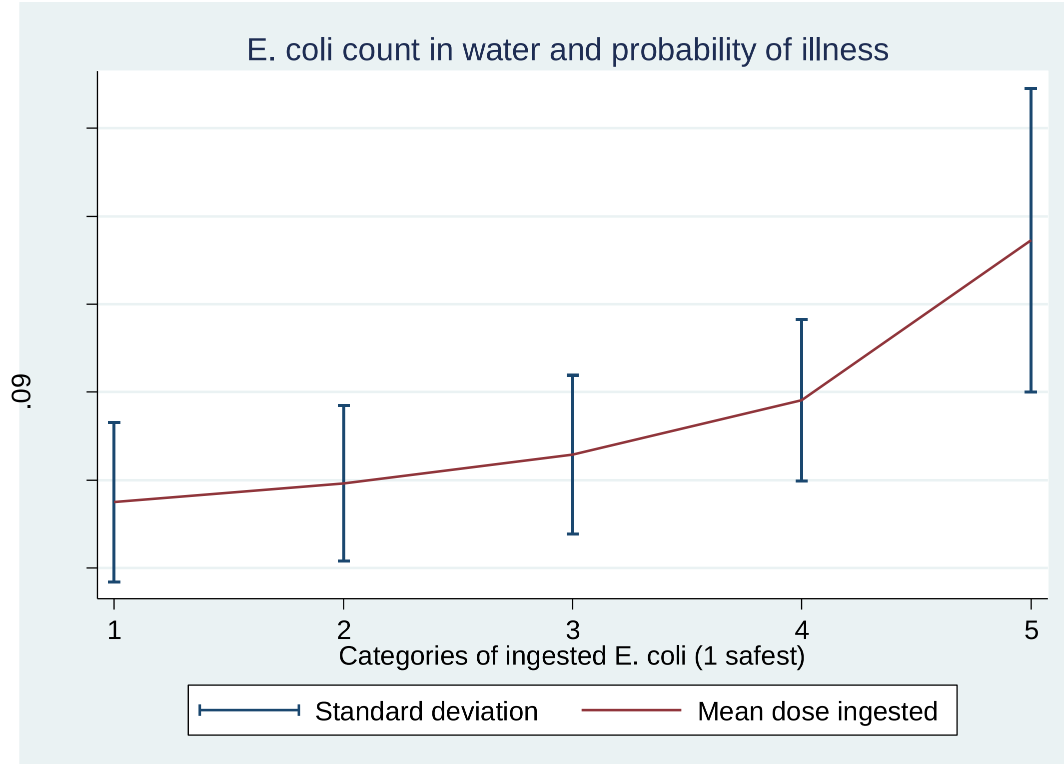

Figure 1.9 E. coli in Drinking Water and Risk of Diarrhea

For this analysis, we have divided the value of E. coli consumed (dose) into five different categories where 1 means the lowest category of E. coli intake and 5 indicates the highest E. coli in the sample data. It can be observed from the following graph that as the dose of E. coli consumed increases, the probability of illness increases. The risk of illness (standard deviation of falling sick) is more or less the same at levels from 1 to 4 and increases only at level 5 of E. coli dose intake[29].

[29] The categories for dose of E. coli in the graph are as follows. The category 1 is for zero intake of E. coli, 2 is for daily dose between 0 to 100, 3 is for daily dose between 100 to 500, 4 is for daily dose between 500 to 1000 and category 5 is for daily dose above 1000.

1.9 Conclusion

This paper analyzes the determinants of drinking water quality in a rural setting of India where the objective water quality is not known, and the selection of water sources for drinking is based on people's perception of water quality. Using a two-year panel data collected from 30 villages in Jharkhand in India, we find that the earlier method of defining improved sources by source types without undertaking tests for water quality is not correct. In our field survey, we collected water samples from different water sources and tested them for the presence of E. coli bacteria. We found that the sources that are considered unsafe, like well and rivers, are indeed unsafe with high E. coli count, but the sources that are considered improved water sources like hand pumps, bore wells and piped water taps are not always safe and sometimes indicate a high count of E. coli[30]. Thus, contrary to the literature, instead of defining safe sources by source types, we use the absence of E. coli at the source as an indicator of safe sources with good water quality.

Additionally, the estimation results reveal that people's perceptions about water quality, specifically relating to the color and taste of the water affect water source choices. Households that understand the meaning of quality of water, and base their quality perceptions on the color of water, choose the water sources that looks colorless or clean. Due to the nature of public infrastructure in rural Jharkhand, these households tend to end up selecting water sources with higher E. coli count like wells.

[30] This will happen if there is underground fecal contamination of these sources due to leakage from underground pipes and septic tanks of home toilets.

Thus, the perceptions based on the color of water tend to leave a household with unsafe water source choices compared to those households whose quality perceptions are not based on the color of the water. Furthermore, households that understand the meaning of water quality and have perceptions based on the taste of water tend to end up selecting relatively safer water sources. The analysis in the paper indicates that the quality of the water at the source, selected for drinking by a household tends to have a significant positive impact on the water quality at home[31]. The water handling practices and hygiene practices inside a house affect drinking water quality at home. We find that the practice of handwashing after using a toilet[32] tends to significantly reduce the E. coli count in stored water at home.

Among the different home treatment techniques used in the field area, we find the use of water filters is the only method that tends to improve drinking water quality at home. We do not find any significant impact of the use of ladle[33] on drinking water quality at home, contrary to our expectations. On preliminary analysis, we find that there a positive relationship between the probability of falling ill and consuming E. coli contaminated water. Households with higher intake of unsafe water (E. coli contaminated water) have a higher probability of falling ill compared to those that drink safe water. The risk of illness is highest for the households that are consuming more than 1000 liters of E. coli colonies on an average daily intake.

The results and analysis of this paper suggest that we need to build water infrastructure in the villages of India. The investment in building water sources like

[31] The water sources that have higher E. coli count at the source (lower water quality at source) tend to have higher E. coli count at home (lower water quality at home).
[32] Conditional on the presence of toilet inside a house.
[33] For taking out water from the storage container.

piped water taps at home, hand pumps, or borewells can help improve the source quality and, ultimately, the home water quality. But this can only happen if the underground water tables from which these sources draw water are not contaminated. To ensure this, the construction of these systems, as well as the maintenance afterward, needs to be done carefully. The building of good quality water infrastructure is necessary but not sufficient to ensure good quality water inside the home. We also need to maintain hygiene and sanitation environment inside a house.

Along with building more toilets, we need behavior change programs centered around handwashing and the adoption of water filters for home treatment before drinking. We found in this paper that the perceptions of water quality based on taste and color influence the selection of water sources for drinking by a household. In such situations, it is important to change people's perceptions of the quality of water through behavioral programs in a way that the perception of quality matches the objective quality of water at the sources. A system of frequent measurements of objective quality of water for microbial contamination and dissemination of this knowledge among the villagers can also help improve drinking water quality at home and reduce the associated health risk[34].

[34] In the meantime, the pipes used for handpump, piped water systems and borewells can be changed so that there is no color in the water from these sources. This can ensure that the color-based quality perceptions do not make people choose wells over these water sources in rural Jharkhand.

REFERENCES FOR CHAPTER 1

Alelign, T., Asegidew, W. & Abera, A., 2016. A cross-sectional study on the incidence and risk factors of diarrheal illness among children under-five years of age in Debre Berhan town, Ethiopia. *J Health Med Econ.*, 2(2), 1–6.

Allcott, H., & Rogers, T. (2012). The short-run and long-run effects of behavioral interventions : Experimental evidence from energy conservation. *NBER Working Paper Series*, 18492.

Alves, D., & Belluzzo, W. (2004). Infant mortality and child health in Brazil. *Economics and Human Biology*, 2, 391–410.

Ashraf, N. et al., 2017. Water, health and wealth. *NBER Working Paper Series*, 23807, 1–24.

Augier, P., Dovis, M. & Lai-Tong, C., 2016. Better access to water, better children's health: A Mirage? *Oxford Development Studies*, 44(1), 70–92.

Azizullah, A. et al., 2011. Water pollution in Pakistan and its impact on public health - A review. *Environment International*, 37(2), 479–497.

Baisa, B., Davis, L. & Salant, S.W., 2008. The welfare costs of unreliable water service, *MPRA Munich personal RePEc archive*, 9812.

Batram, J. et al., 2005. Focusing on improved water and sanitation for health. Millennium project. *Lancet*, 365, 810–812

Behrman, J.R. & Deolalikar, A.B., 1988. Chapter 14 Health and nutrition. *Handbook of Development Economics*, 1, Elsevier science publishers, 631–711.

Bontemps, C. & Nauges, C., 2016. The Impact of perceptions in averting-decision models: An application of the special regressor method to drinking water choices. *American Journal of Agricultural Economics*, 98(1), 297–313.

Brown, J.M., Proum, S. & Sobsey, M.D., 2008. Escherichia coli in household drinking water and diarrheal disease risk: Evidence from Cambodia. *Water Science and Technology*, 58(4), 757–763.

Carson, R.T. et al., 2002. Estimating willingness to pay for improved drinking water quality using averting behavior method with perception measure. *Environmental and Resource Economics*, 33(3), 277–283.

Chandra, A. & Staiger, D., 2016. Sources of inefficiency in healthcare and education. *American Economic Review*, 106(5), 383–387.

Clasen, T. F. et al., 2018. Reducing diarrhea through the use of household-based ceramic water filters: A randomized, controlled trial in rural Bolivia. *The American Journal of Tropical Medicine and Hygiene*, 70(6), 651–657.

Charrois, J.W.A., 2010. Private drinking water supplies: Challenges for public health. *Cmaj*, 182(10), 1061–1064.

Dada, A., 2009. Sachet water phenomenon in Nigeria : Assessment of the potential health impacts. *African Journal of Microbiology Research*, 3(1), 15–21.

Dang, A., Likhar, N. & Alok, U., 2016. Importance of economic evaluation in health care: An Indian perspective. *Value in Health Regional Issues*, 9(6), 78–83.

Doria, M. de F., Pidgeon, N. & Hunter, P.R., 2009. Perceptions of drinking water quality and risk and its effect on behaviour: A cross-national study. *Science of the Total Environment*, 407(21), 5455–5464.

Doria, M. de F., 2010. Factors influencing public perception of drinking water quality. *Water Policy*, 12(1), 1–9.

Dorner, S.M. et al., 2007. Pathogen and indicator variability in a heavily impacted watershed. *Journal of Water and Health*, 5(2), 241–257.

Dumas, C.E., Schuhmann, P.W. & Whitehead, J. c., 2005. Measuring the economic benefits of water quality improvement with benefit transfer : An introduction for noneconomists. *American Fisheries Society Symposium*, 1–18.

Dupas, P. & Miguel, E., 2016. *Impacts and Determinants of Health Levels in Low-Income Countries*. Chapter 2, Handbook of Economic Field Experiments, Volume II, Elsevier Ltd.

Dupont, D., Adamowicz, W.L. & Krupnick, A., 2010. Differences in water consumption choices in Canada: The role of socio-demographics, experiences, and perceptions of health risks. *Journal of Water and Health*, 8(4), 671–686.

Dutta, V., Chander, S. & Srivastava, L., 2005. Public support for water supply improvements: Empirical evidence from unplanned settlements of Delhi, India. *Journal of Environment and Development*, 14(4), 439–462.

Easter, K.W. & Konishi, Y., 2006. Estimating economic health costs of not controlling toxic water pollution. *Presented at IAAE 2006 Australia,* August 12-18 2006.

Edberg, S.C. et al., 2012. Escherichia coli: the best biological drinking water indicator for public health protection. *Journal of Applied Microbiology*, 88(S1), 106S–116S.

Etale, A., Jobin, M. & Siegrist, M., 2018. Tap versus bottled water consumption: The influence of social norms, affect and image on consumer choice. *Appetite*, 121, 138–146.

Galiani, S. et al., 2005. Water for Life: The Impact of the Privatization of Water Services on Child Mortality. *Journal of Political Economy*, 113(1).

Gamper-Rabindran, S., Khan, S. & Timmins, C., 2010. The impact of piped water provision on infant mortality in Brazil: A quantile panel data approach. *Journal of Development Economics*, 92(2), 188–200.

Grady, C.A. et al., 2015. Microbial quality of improved drinking water sources: Evidence from western Kenya and southern Vietnam. *Journal of Water and Health*, 13(2), 607–612.

Gundry, S., Wright, J. & Conroy, R., 2004. A systematic review of the health outcomes related to household water quality in developing countries. *Journal of Water and Health*, 2(1), 1–13.

Havelaar, A. & Melse, J.M., 2003. Quantifying public health risk in the WHO guidelines for drinking water quality. *WHO RIVM report 734301022*, 1–49.

Himayatullah Khan, Faiza Iqbal, Imranullah Saeed, I.K., 2010. Estimating willingness to pay for improvements in drinking water quality : evidence from Peshawar, Northern Pakistan. *Environmental Economics*, 1(2),38–43.

Hunter, P.R., Zmirou-Navier, D. & Hartmann, P., 2009. Estimating the impact on health of poor reliability of drinking water interventions in developing countries. *Science of the Total Environment*, 407(8), 2621–2624.

Husain, I., Husain, J. & Arif, M., 2015. Socio-economic effect on socially-deprived communities of developing drinking water quality problems in arid and semi-arid area of central Rajasthan. *Proceedings of the International Association of Hydrological Sciences*, 364(June), 229–235.

Hutton, G. & Haller, L., 2004. Evaluation of the costs and benefits of water and sanitation improvements at the global level. *World Health Organization*, 1–87.

Hyun-Jung, J., Choi, Y.J. & Ka, J.O., 2011. Effects of diverse water pipe materials on bacterial communities and water quality in the annular reactor. *Journal of Microbiology and Biotechnology*, 21(2), 115–123.

Ishii, S. & Sadowsky, M.J., 2008. Escherichia coli in the environment: Implications for water quality and human health. *Microbes and Environments*, 23(2), 101–108.

Jalan, J. & Ravallion, M., 2003. Does piped water reduce diarrhea for children in rural India? *Journal of Econometrics*, 112(August), 153–173.

Jessoe, K., 2013. Improved source, improved quality? Demand for drinking water quality in rural India. *Journal of Environmental Economics and Management*, 66(3), 460–475.

Khan, J.A.M. et al., 2018. Economic costs of hospitalized diarrheal disease in Bangladesh: a societal perspective. *Global Health Research and Policy*, 3(1), 1–12.

Koolwal, G. & van de Walle, D., 2012. Access to water, women's work, and child outcomes. *Economic Development and Cultural Change*, 61(2), 369–405.

Kremer, M. et al., 2009. Spring Cleaning: Rural water impacts, valuation and property rights institutions.*NBER Working paper series* 15280.

Kumar, A. & Das, K.C., 2014. Drinking water and sanitation facility in India and its linkages with diarrhoea among children under five: Evidence from Recent Data. *International Journal of Humanities and Social Science Invention*, 3(4), 50–60.

Kuo, Y.F., Wu, C.M. & Deng, W.J., 2009. The relationships among service quality, perceived value, customer satisfaction, and post-purchase intention in mobile value-added services. *Computers in Human Behavior*, 25(4), 887–896.

Lanz, B. & Provins, A., 2016. The demand for tap water quality: Survey evidence on water hardness and aesthetic quality. *Water Resources and Economics*, 16(October), 52–63.

Lanz, B. & Provins, A., 2017. Using averting expenditures to estimate the demand for public goods: Combining objective and perceived quality. *Resource and Energy Economics*, 47, 20–35.

Ma, C. et al., 2014. Behavioural factors associated with diarrhea among adults over 18 years of age in Beijing, China. *BMC Public Health*, 14(1), 1–7.

Mangyo, E., 2008. The effect of water accessibility on child health in China. *Journal of Health Economics*, 27(5), 1343–1356.

McGarvey, S. T. et al, 2008. Community and household determinants of water quality in coastal Ghana. *Journal of Water Health*, 6(3), 339–349.

Martínez-Santos, P., 2017. Does 91% of the world's population really have "sustainable access to safe drinking water"? *International Journal of Water Resources Development*, 33(4), 514–533.

Merkel, L., Bicking, C. & Sekhar, D., 2012. Parents' perceptions of water safety and quality. *Journal of Community Health*, 37(1), 195–201.

Montgomery, M.A. & Elimelech, M., 2008. Water and sanitation in developing countries: Including health in the equation. *Environmental Science & Technology*, 41(1), 17–24.

Mosler, H.J., 2012. A systematic approach to behavior change interventions for the water and sanitation sector in developing countries: a conceptual model, a review, and a guideline. *International Journal of Environmental Health Research*, 22(5), 1–19.

Moyo, S. et al., 2004. Realising the maximum health benefits from water quality improvements in the home: A case from Zaka district, Zimbabwe. *Physics and Chemistry of the Earth*, 29(15–18 SPEC.ISS.), 1295–1299.

Naser, A.M. et al., 2015. Effects of source- versus household contamination of tubewell water on child diarrhea in rural Bangladesh: A randomized controlled Trial. *Plos One*, 10(3), 0121907.

Ngoran, S.D., 2015. The dynamism between urbanization, coastal water resources and human health: A case study of Doula, Cameroon. *Journal of Economics and Sustainable Development*, 6(3).

Novak, L., 2014. The impact of access to water on child health in Senegal. *Review of Development Economics*, 18(3), 431–444.

Nygren, B.L. et al., 2016. The relationship between distance to water source and moderate-to-severe diarrhea in the global enterics multi-center study in Kenya, 2008-2011. *American Journal of Tropical Medicine and Hygiene*, 94(5), 1143–1149.

Olmstead, S.M., Hanemann, W.M. & Stavins, R.N., 2007. Water demand under alternative pricing structures. *National Bureau of Economic Research Working Paper Series*, (1), 1–5.

Onjala, J., Ndiritu, S.W. & Stage, J., 2014. Risk perception, choice of drinking water and water treatment: evidence from Kenyan towns. *Journal of Water, Sanitation and Hygiene for Development*, 4(2), 268–280.

Pérez-Flores, M.E. et al., 2017. Perceived versus actual water quality: Community studies in rural Oaxaca, Mexico. *Science of The Total Environment*, 622–623, 626–634.

Pierce, G. & Gonzalez, S., 2017. Mistrust at the tap? Factors contributing to public drinking water (mis)perception across U.S. households. *Water Policy*, 19(1), pp.1–12.

Polasky, S. et al., 2012. Linking water quality and well-being for improved assessment and valuation of ecosystem services. *Proceedings of the National Academy of Sciences*, 109(45), 18619–18624.

Price, J.I. & Heberling, M.T., 2018. The effects of source water quality on drinking water treatment costs: A review and synthesis of empirical literature. *Ecological Economics*, 151(April), 195–209.

Rehman, S. et al., 2012. Drinking water quality and human health risk in Charsadda district, Pakistan. *Journal of Cleaner Production*, 60, 93–101.

Rygaard, M., Arvin, E. & Binning, P.J., 2009. The valuation of water quality: Effects of mixing different drinking water qualities. *Water Research*, 43(5), 1207–1218.

Schmidt, W. & Cairncross, S., 2009. Comment on "Household water treatment in poor populations : Is there enough evidence for scaling up now ?". *Environmental Science and Technology*, 43(4), 986–992.

Sorenson, S.B., Morssink, C. & Campos, P.A., 2011a. Safe access to safe water in low-income countries: Water fetching in current times. *Social Science and Medicine*, 72(9), 1522–1526.

Stauber, C. et al., 2014. Evaluation of the compartment bag test for the detection of Escherichia coli in water. *Journal of Microbiological Methods*, 99(1), 66–70.

Tang, C. & Zhang, Y., 2016. Using discrete choice experiments to value preferences for air quality improvement: the case of curbing haze in urban China. *Journal of Environmental Planning and Management*, 59(8), 1473–1494.

Violette, W., 2017. Optimal pricing and informal sharing : Evidence from piped Water in Manila. , *Job Market Paper, Brown University*, 1–86.

Wade, T.J. et al., 2003. Do U.S. Environmental Protection Agency water quality guidelines for recreational waters prevent gastrointestinal illness? A systematic review and meta-analysis. *Environmental Health Perspectives*, 111(8), 1102–1109.

Wang, A. et al., 2017. Household microbial water quality testing in a Peruvian demographic and health survey: Evaluation of the compartment bag test for Escherichia coli. *American Journal of Tropical Medicine and Hygiene*, 96(4), 970–975.

Wedgworth, J.C. et al., 2014. Associations between perceptions of drinking water service delivery and measured drinking water quality in rural Alabama. *International Journal of Environmental Research and Public Health*, 11(7), 7376–7392.

WHO, 2011. Guidelines for Drinking-water Quality 4th edition., WHO, Geneva, p. 340. *World Health Organization*.

Wright, J.A., Gundry, S.W. & Conroy, R., 2004. Household drinking water in developing countries: A systematic review. *Tropical Medicine and International Health*, 9(1), .106–117.

Zhou, J. et al., 2007. Estimating water quality pollution impacts based on economic loss models in urbanization process in Xi'an, China. *Journal of Urban Planning and Development*, 133(3), 151–160.

Zhu, X. & van Ierland, E.C., 2012. Economic modelling for water quantity and quality management: A welfare program approach. *Water Resources Management*, 26(9), 2491–2511.

CHAPTER 2

ARE PEOPLE MAKING CORRECT CHOICES? DRIVERS OF WATER SOURCE CHOICES IN RURAL JHARKHAND

2.1 Introduction

In our daily life, many of us make choices about almost everything from selecting the clothes to wear to buying products for consumption. Our perceptions and preferences govern these choices. In many countries of western Europe, clean and safe water is being provided at home through piped water, and people drink it by filling a glass from the tap. They believe that the water supplied from piped water tap at their home is safe for drinking. Regular tests for quality of water and public disbursement of the quality test results help in supporting and maintaining this belief. One can say that the perception of quality coincides with objective measures of quality in an ideal situation. But even in developed countries, there are times when people doubt the quality of water being supplied from piped water at their homes. In these situations, it has been observed that people start buying bottled water from the market for drinking. They choose between using the piped water provided at their home or purchasing water from the market for drinking based on their perception of water quality. The perceptions of water quality could be based on several factors that would affect water source choices, and these factors generally differ in a developing and developed country.

This desire to have clean drinking water is widespread, from the poorest of regions to the wealthiest. The difference is that people who live in the less developed parts of the world suffer from lack of proper water infrastructure, do not have systems for regular water quality tests, and water markets are absent or not fully developed. In a rural community of a developing country like India, people choose water sources from the available sources in the community. If there is only one source available in the community, then there is no question of a choice, people will have to use the available source for drinking purposes. But in cases where more than one source is available (in numbers as well as in terms of types of sources), people choose among available sources. The sources available in a rural community in India are generally public, which are supplied by the government or are common property sources like a river. There is another option of building a private source by independent households for their personal use. However, this option is viable only for those households that can and are ready to invest in their private sources of water.

There should be periodic testing of water quality and disbursement of this knowledge among the public. If the objective quality of water is known to everyone, then one would expect the water quality test results to influence people's choices (keeping in mind people's trust in the credibility of these sources). But in the rural areas of India, there are almost no tests for water quality. Even if there are some random occasional water quality tests, they are not repeated, standards are not maintained, and the results are not made public. In the absence of actual information about the objective tested quality of water, people's perception of the quality determines their water source choices. Distance to the source and travel time spent in the collection of water is an important factor that affects whether people's choice of a water source. But people also travel a distance to fetch water by leaving the nearest sources if they feel that a source has better color, better taste or low risk of making them ill. In this paper, we use a revealed preference data and a stated preference data to analyze the impact of perception indicators of water quality on water source choices as well as their willingness to pay for an improvement in these indicators.

There is an acute lack of safe water sources in Indian villages, and we need government or other developmental agencies to invest in building water infrastructure. But we also want people to use these sources for drinking purposes. People don't necessarily use better quality water sources if they are made available to them. It depends a lot on their perception of water quality. Thus, there is a need to understand these quality perceptions and how they drive water source choices.

In Jharkhand, we observed that sometimes instead of choosing a water source nearer to the house or sources which are considered safer, people select other water sources. These are the sources which they perceive to be a better source. It may seem puzzling or strange behavior. However, their perceptions drive these decisions. These perceptions are formed by their acquired knowledge, practices, social and cultural norms, and peer effects of their neighbors in the community. Instead of assuming or taking it for granted that once provided access to people will choose the safest or nearest sources, it is essential to understand people's behavior and their perceptions. Different countries, different communities have different practices and norms related to drinking water, which may also affect their choices. A better understanding of people's behaviors and their perceptions can allow us to make better policy decisions.

It will guide us in the making of a drinking water policy with a clear understanding of where to invest, how to invest, and what kind of behavioral programs should be accompanied by to make it more effective.

The structure of the rest of the paper is as follows. After the introduction, sections 2.2 and 2.3 contain a literature review and a conceptual framework for the methods used in the analysis. Section 2.4 has an explanation of the revealed preference and stated preference data used in the paper. In section 2.5, we have the descriptive statistics of these data sets. Section 2.6 contains the estimation results and willingness to pay calculations for both data sets, and we conclude in section 2.7.

2.2. Literature Review

In the literature, there is a set of papers that use discrete choice models to explain the choice of a particular water source for a specific purpose (Madanat & Humplick, 1993; Mu & Briscoe, 1990; Persson, 2002; Whittington et al., 1990). These papers are concentrated in rural areas of developing countries. The paper by Mu and Briscoe (Mu & Briscoe, 1990) estimated the value of time spent in collecting water from different sources of water in a multinomial logit model. It was based in Kenya, where people chose water from the available choices in the community like the kiosk, well, and vendors. The price and time spent in the collection of the water are determining the choice of water source in Mu and Briscoe (Mu & Briscoe, 1990) 's paper. In another paper by the same authors, they use the above-stated variables along with the number of women in the family, income, and female education for estimating the choice of water source for a household.

In the Madanat and Humplick (Madanat & Humplick, 1993) paper, a joint analysis of the choice of a water source choice and decision to take out a piped water connection is undertaken using a sequential maximum likelihood estimation method. They use cross-sectional survey data from Pakistan. Compared to Mu and Briscoe (Mu & Briscoe, 1990), this paper distinguishes between the sources of water used for different household purposes. The assumption of sequential choices when these two decisions (of taking piped water connection and selection of a water source in the community) seem to be a joint decision is the one major limitation of this paper.

Using the nested conditional logit model, Persson (Persson, 2002) analyses the choice of water sources for drinking purposes in the rural Philippines. In this paper, input prices, income (as a proxy for taste), and household size are the factors that are used for calculating the choice probabilities. Intercepts corresponding to each water source choice are included in the regression to account for different utility from the selection of different water sources in the data. As the choice set could be different for different families living in different communities, the nested conditional logit model design was proposed. But only the conditional part of the model is estimated due to lack of enough observations in the data. The results of this paper indicate that the price (time cost) of any alternative reduces the choice probability of that alternative. At the same time, income (a proxy for taste) has no significant effect on it.

Another study in recent years on water source choices is by Sahn et al. (Boone, Glick & Sahn, 2010). This paper explores the choice of a water source by households in

Medagsaker using a conditional logit model. In their analysis, the choice of water source is a function of the distance, time spent in water collection and household characteristics like the education of family members. It also estimates the determinants of time spent in the collection of water for adult family members and children, regressing it on the distance to the water source and other household characteristics. The results in the paper indicate that the time spent in the collection of water is positive and significantly associated with the distance to the water source. Furthermore, it finds that female family members spend more time in the collection of water in general. The children spend more time in the collection of water in rural areas, and having more adult members in the family reduces the time spent by any individual in the collection of water.

Another paper in the literature is by Kremer and others (Kremer et al., 2011) based on spring water treatment at the water source in Kenya. In their paper, they found that the contingent valuation estimate was much higher than the travel cost and experiment-based calculation of willingness to pay for the water quality improvements (Kremer et al., 2011).

The second set of literature is focused on a more urban setting with semi or fully developed markets for drinking water. In this literature, especially in the recent years, there have been papers that have found perception of drinking water quality as one of the main factors influencing water source choices (Cheabu & Ephraim, 2014; Doria et al., 2009; Sajjadi et al., 2016; Kinyuru et al., 2014; Onjala et al., 2014). But there have not been many studies in this literature looking at the effect of quality perception on water source choices that would be relevant in a rural setting.

In the literature, we also find some discrete choice experiments on social norms, information, quality of water, health, and household behavior. One such study is the Opower electricity company-related experiments(USA)(Allcott & Rogers, 2012). In this study, the effect of social norms on household behavior was researched. Letters were sent to the users along with their electricity bill, indicating the usage of electricity relative to their neighbors to understand the effects of social norms on individual/household behavior (Allcott & Rogers, 2012). Similar to the electricity usage experiments, there were experiments conducted on water usage, which are known as the Cobb County experiment(USA)(Ferraro et al., 2011; Ferraro & Price, 2013). In these experiments, letters were sent to water users along with their bills, indicating the usage by the neighbors. Both research studies reported a decline in the usage by treatment households compared to the control after receiving the social comparison messages.

Another factor that might affect source choices for water is the taste preferences for water. Taste preferences might be based on beliefs formed through experience, information and knowledge about associated health effects, the risk associated with water, social norms and behavior of the peers in selecting water sources (Doria et al., 2009; Espinosa-García et al., 2015; Huber & Mosler, 2013). Then there are also chemical and sensory parts relating to taste preferences of water (Puget et al., 2010). There is also a part due to advertisement and brand names, which is more critical when there are developed markets for drinking water. In rural areas of developing countries like India, these markets most of the time do not exist, so brand names and advertisements may not have much impact on taste preferences. But the other factors like experience, information, and belief about associated health risk and social norms might affect the taste preferences for drinking water in rural India.

In most of the earlier literature set in rural areas, the focus was on the distance to the source and household characteristics like income in the determination of water source choices. But it did not include the role of perception-based variables in water source choices. On the other hand, the literature that did include perception-based factors was based in an urban setting and developed water markets and did not take into account the role of distance and travel cost in fetching water.

In an urban setting, people are making choices between piped water supplied at home or bottled water from the market. That is quite different from a rural setting with a non-existent drinking water market and people making choices between mostly public sources available in the community. In this paper, we analyze the role of perception of quality and taste preference along with the distance to the source in choice of a drinking water source. It is the first such study (to the best of our knowledge) that combines these two sets of literature. The contribution of our research lies in the fact that we aim to use more explanatory variables and a mixed logit model allowing for household heterogeneity in the discrete choice model for drinking water source choice. We use a revealed and stated preference data to analyze the role of perception based indicator in water source choices and compare the willingness to pay measures calculated for both types of data sets. It is also the first study (to the best of our knowledge) conducting choice analysis for water sources in rural India.

2.3 Conceptual Framework

Let us suppose that there are n alternative water sources available to a household in a community. These are exclusive, exhaustive, and finite number of alternative water sources from which the household chooses one at a given point of time for drinking purposes. [35]

The representative individual for household r chooses a water source alternative j only if it maximizes the utility for that household. There are attributes of the water sources and the household characteristics which jointly determine the choice of a particular alternative water source for any household in a village. The basic underlying assumption is that a household maximizes its utility given the set of choice of water sources available to it.

Suppose there are total N alternative water sources, then a representative individual in household r will choose water source j if

$$U_{jr} \geq U_{ir} \; for \; j, i \in \{N\} \; and \; i \neq j$$

If Y is a variable representing the choice of water source j then

$$Y_{jr} = 1$$

$$if \quad U_{jr} \geq U_{jr} \quad for \; j,i \in \{N\} \; and \; i \neq j$$

[35] Even though people might be using more than one source at a time, in theory, our observations from the field provide evidence against that. So, we assume that at one point in time, any household chooses only one source from the available sources to be used for drinking purposes. Thus, by construction, the choice set in this analysis consists of all finite water sources available in the community but not their combinations.

$$Y_{jr} = 0 \ otherwise$$

The household's unobserved utility from choosing any source of water is assumed to be an additively separable function under a random utility model. It consists of an observed term (indirect utility which is observed and can be calculated from the data) and a random error term.

$$U_{jr} = V_{jr} + \varepsilon_{jr}, \quad for \ j \in N$$

It means that

$$Y_{jr} = 1$$

$$if \ V_{jr} + \varepsilon_{jr} \geq V_{ir} + \varepsilon_{ir} \quad for \ j,i \in \{N\} \ and \ i \neq j$$

$$Y_{jr} = 0 \ otherwise$$

Any discrete choice model allows us to estimate the probabilities of choosing a particular source of water given the attributes of the water source and the household characteristics. If the utility function for a household is as defined above, then in terms of the random utility model the probabilities associated with choosing water source j would be

$$Prob(Y_{jr} = 1) = P(V_{jr} + \varepsilon_{jr} \geq V_{ir} + \varepsilon_{ir}) \ for \ j,i \in \{N\} \ and \ i \neq j$$

$$Prob(Y_{jr} = 1) = P(V_{jr} - V_{ir} \geq \varepsilon_{ir} - \varepsilon_{jr}) \ for \ j,i \in \{N\} \ and \ i \neq j$$

For the identification of this discrete choice model and estimation of the actual utilities and the choice probabilities associated with it, the distribution of the error term needs to be defined. The variables which enter into the determination of the indirect utility function also have to be specified.

In different discrete choice models, there are different specifications for the variables which enter into the indirect utility function for a household. The variables which enter the utility function, in this case, would be the household characteristics and the attributes of water sources. Generally, a multinomial logit model incorporating both the attributes of the water sources and the characteristics of the household takes the form of a conditional logit model as follows.

If X is the household characteristics for a household r and Z is the vector of attributes of particular water source j for household r then the indirect utility (observed part of the utility function) would be

$$V_{jr} = \beta_j X_r + \alpha Z_{jr}$$

In the conditional logit model, the characteristics of water sources Z vary for each water source j for household r. Still, the household characteristics X remain the same for a household r across all water sources j.

Let's assume that the indirect utility or the observed part of the utility is a function of demand for water and composite goods

$$V_{jr} = V_{jr}\ (W_d, C)$$

Another way of writing it would be,

where the water quantity demanded is a function of the characteristics of water source choices and the specific household characteristics in the following manner.

$$W_d = W_d\ (Z_{jr}, X_r)$$

The constraint for this utility maximization problem would be

$$C * p + W_d * P_w \leq Y_r$$

Where Y_r is the income of household r, p is the price of composite good and P_w is the price of collecting water (time cost of collection of water converted into monetary equivalents using wage rates).

The household or the individual tries to maximize its utility given the constraints on the water demand and the consumption of the other composite goods. The water source alternative, which leads to the maximization of the household's utility given the constraints and the household and water source characteristics, is the one which is chosen by the household. Since the dependent variable is discrete, the demand functions and the choice probabilities cannot be derived simply by using the first-order conditions of a utility maximization problem. Instead, the utility functions will be directly applied to find out the demand function and choice probabilities.

Under the assumption of Independence of Irrelevant Alternatives (IIA) and homogeneity of the population's preferences lack of variation of household characteristics across different water source choices within a household, a conditional logit model can be used to estimate the household choices of water.

In this model, it can be assumed that the error term follows extreme value distribution and the model, and the estimation procedure becomes similar to the standard multinomial logit models. In the simplest case, it can also be assumed that both indirect utility and the water demand functions are linear in their parameter.

Under these assumptions, the following equation can be derived for household r choosing water source j.

$$Y_{jr} = \beta_j X_r + \alpha Z_{jr} + \varepsilon_{jr}$$

where the utility U_{jr} is maximized when water source j is chosen, and the value of Y_{jr} is equal to one.

If the assumption of IIA does not hold and there is household heterogeneity, then we cannot use the conditional logit model. In that case, we use a mixed logit for the choice of a water source, which allows for household heterogeneity in preferences as well as the violation of the independence of irrelevant alternative criteria.

If the preferences of people vary across a village, that would mean that instead of one constant value of β_j in the sample data, it follows a probability distribution with a density function as $f(\beta/\theta)$ that explains the heterogeneous preferences driving water source choices in the data set.

If we assume that f(beta) follows a normal distribution, we can have a mixed logit model, where the probability of choosing a particular source will be given by

$$\int \frac{\exp\left(x_{jr}\beta j\right)}{\sum_{i=1}^{N} \exp(x_{ir}\beta i)} f(\beta/\theta) d\beta$$

2.4 Fieldwork and Data Construction

There are two sets of data that are being used in this paper, revealed preference, and stated preference data. In this section, we explain in detail the fieldwork and data construction method used for creating these two data sets used in this paper.

2.4.1 Revealed Preference Data

We collected a two-year panel data from 30 villages in the Jharkhand state of India. The 30 villages in the data have different sets of available water sources from which a household chooses a water source for drinking each year. As the water sources in these villages are generally not constructed by one development agency, there are differences in the mode of construction as well as the final source type. Due to this reason, it is not realistic to assume that one water source from one village will be the same as a water source of that type from another village. For example, we cannot say that for an individual choosing a water source well in the village one is the same as another well in village two.

For this reason, we create a universal choice set consisting of all possible sources comprising of all sources in different villages. A household in a village chooses from the sources available in their village. The distance between villages makes it almost impossible for a household to select a source from another village. It is a varying choice set in which households in each village are choosing sources from the same set of water sources, and households in different villages are choosing from a different set of sources. However, by construction, each household is exposed to the same universal choice set of sources.

We collected data from 15 villages in each district Khunti, and Ramgarh of Jharkhand state in India, with 25 randomly chosen households from each village for two years, 2016-17 and 2017-18. Data were collected through the interview-based survey of a representative woman from each household. The distance from each sample household to all the water sources available in the community was measured during the field data collection. Using this distance data, we calculate travel costs to each source in the village for sample households irrespective of whether that is the source chosen by that household or not. We use these travel costs and the information about the selected source to construct a discrete choice data set for each household surveyed in a village. We then repeat this exercise for sample households in different villages and combine this data to construct a choice data based on the universal choice set. In each household, we use the revealed data about quality perception indicators (based on taste, color, and no perceived risk of falling ill; these are all binary indicators with 1-0 values)[36] to construct village averages for each of the water sources in the community. We use this village averages as the values of different quality based perception indicators associated with various sources for a household. Since we do not have individual reported values on individual perception indicators except for the chosen source by a household, this is the best alternative method that we could use in this analysis. Using village level averages for perception-based indicators for different water sources can also control to some extent, the possible endogeneity in individual quality perceptions. People's experiences, like the history of illness, may affect people's perceptions as well as their source choices. Using the village averages indicates using community-based quality perceptions that can act as a proxy for individual perceptions. Still, at the same time, one would expect them to be uncorrelated with personal experiences and old illnesses in a household.

[36] Likert scale indicators did not work as people were not able to respond to them in pilot study.

In the revealed preference data set, there are alternative specific constants for different types of water sources available in the community like well, hand pumps, Chua[37], and river. These are publically available sources which anyone can use in the village[38]. But each household also has the option of choosing to build a private source of its own, which other people in the community are not allowed to use. There are also piped water taps in households that get water from a piped water system in the village. These connections are also private most of the time but can be public too when instead of individual taps, there are public taps in the village linked with a piped water system. To differentiate between the nature of the public and private source and its effect on water source choices, we have constructed two separate discrete choice data. In the public data set, it has only publically available sources, and the households that choose private sources are removed from the data set. The second data set contains both types of households, public sources, along with privately-owned water sources. Each household should be exposed to the same set of choices. To ensure that we assume that a household which is choosing public sources now can, in theory, choose to build a private source of its own. We introduce a private source choice for these households and create discrete choice data for analysis[39].

[37] A smaller well built around natural water fountains (springs) in Jharkhand.

[38] Provided that there are not unstated caste based restrictions on open access of public sources which we do not observe in our field villages in Jharkhand.

[39] We understand the limitation which income and other households specific variables, observed or unobserved variables may place on this assumption, but for the sake of data construction, we assume it anyway.

2.4.2 Stated Preference Data

The stated preference data is collected from 200 randomly chosen sample households, by selecting 20 respondents each from 10 villages. These households and villages are a subset of those from which we had collected panel data used for the construction of the revealed preference data set. The same woman representatives who had been interviewed in the panel data survey were asked to take part in the stated preference data collection. Each respondent was provided with a sheet with pictures[40] containing a description of two alternative water sources A and B (with no alternative specific constant). These sources are defined by three attributes, distance to the source, color of the water, and taste of the water. Based on the description of these two sources, the women were asked to choose the source which they would prefer. Their answers were recorded on a separate sheet by the survey enumerator. Each respondent was presented with eight such choice questions. All of the eight choice questions which the woman was asked to answer contained different choice sets of sources A and B defined by varying levels of the three attributes – distance, taste, and color.

The data for one household contains answers for eight choice questions. Each answer constitutes one data point in the construction of stated preference data. So, we have multiple data points for each household. Since there are no alternative specific constants in this data sets, the only way one choice differs from other is through the level of attributes and the cost of travel (defined by the distance to the source).

[40] A sample of the sheet containing choice sets used in the collection of stated preference data is being provided in the appendix.

There are three levels for each attribute of water source constructed and used in a choice set, i.e., taste, color, and distance. These levels for each attribute are ranked from low to high (or high to low check; it is rated best to worst). The levels are constructed based on the information collected in the pilot study conducted before the actual discrete choice data collection. This entire exercise of data collection was done from January 2018 to February 2018 in Ramgarh district of Jharkhand state of India.

2.5 Descriptive Statistics and Estimation Results

2.5.1 Summary Statistics: Revealed Preference Data

The summary statistics for the public sources revealed preference data set are being provided in Table 2.A.1 in the appendix. The mean values of different variables are provided in separate columns, depending on water source choices made by different households in the data. We observe that, on average, some of the variables have similar values across households choosing different types of water sources. Some of the other variables have entirely different values across sources. For example, the proportion of upper caste and lower caste people choosing a water source is more or less the same across the sources except in Chua. We find that compared to other sources, a slightly higher proportion of lower caste people choose Chua. The percentage of children under five and the proportion of married women among survey respondents are distributed similarly between different sources. The age of the female respondents is similar across various alternative sources, with those choosing Chua being slightly younger. The asset index is a ratio of the number of assets available in a household as with a total of 22 possible assets. The mean value for the asset index is similar across different source types with a slightly higher value for people using piped water taps. It indicates that better-off families in a village tend to choose piped water connections. The average levels of female education show interesting trends. The households that choose the river for drinking purposes have the lowest completed education levels for female survey respondents. In contrast, those choosing Chua are the ones with the highest level of completed education compared to others. The four factors determining source choices, travel cost, taste perception of water, color-based quality perception, and perceived risk of falling ill differ between different source types. It is expected as these differences are the factors driving source choices for households in a village.

In Table 2.A.2 in the appendix, we provide the summary statistics for the revealed preference data, which include privately-owned sources. The results in different columns correspond to various source alternatives. The results of the summary statistics are more or less similar to the earlier table (Table 2.A.1). The only difference is having an additional source type of privately-owned water source for households in the data.

Distance to the source determines the time cost of traveling and influences a choice of water source for people living in rural communities. But distance is not the only factor determining this choice. If the distance to the source were the only factor determining people's choice, then they would choose the water source nearest to their house. We plot the percentage of people choosing their closest water sources in the sample villages, and the result is provided in Figure 2.1. From Figure 2.1, it can be observed that in the communities in our sample, there is a substantial proportion of households that are choosing sources situated far away from their house. About 40 percent of the total households in the data decided to fetch water from sources, which are further away from their home.

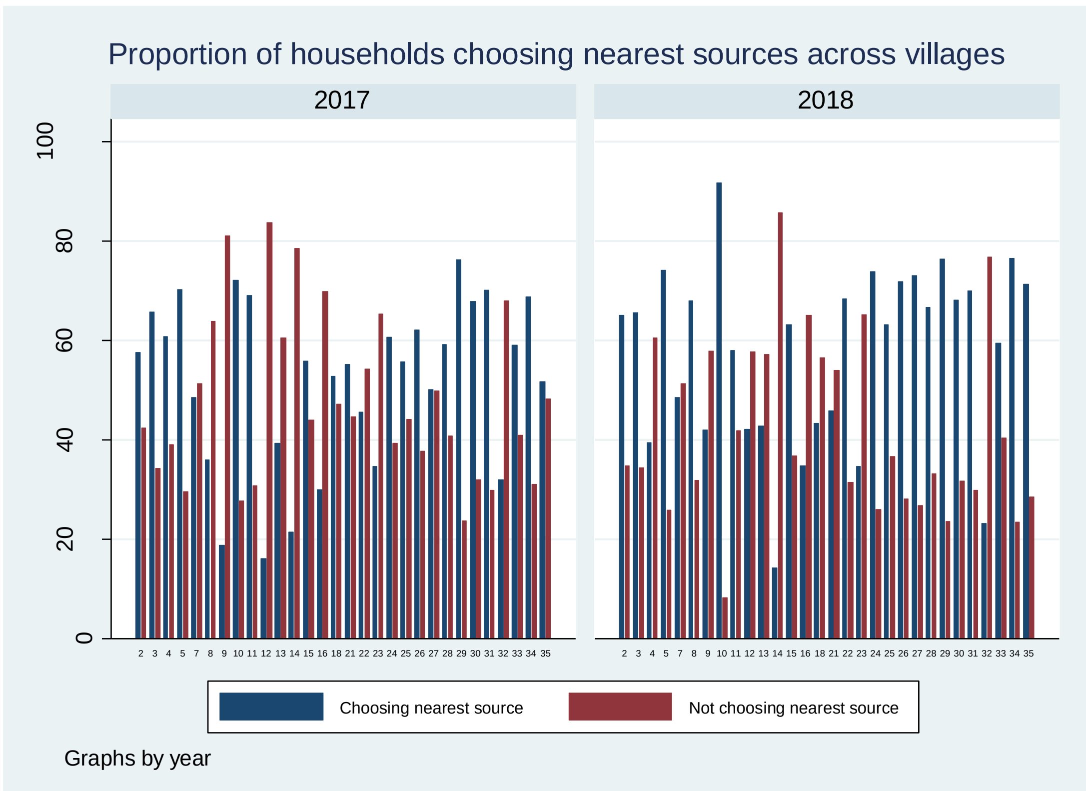

Figure 2.1: Distribution of Houses Choosing the Nearest Source for Drinking Across Different Villages

It indicates that there might be other factors apart from the distance to the source and the time cost of traveling, which might affect people's water source choices.

In Figure 2.2, we plot the households in each village for which the quality perception of the chosen source matches the objective quality of water. In the panel data set, we had also collected water samples from the sources and tested them for the presence of E. coli bacteria as a proxy for drinking water quality. From Figure 2.2, we can observe the proportion of households in each village for which the perception of quality matches and non-matches with the tested water quality of the source.

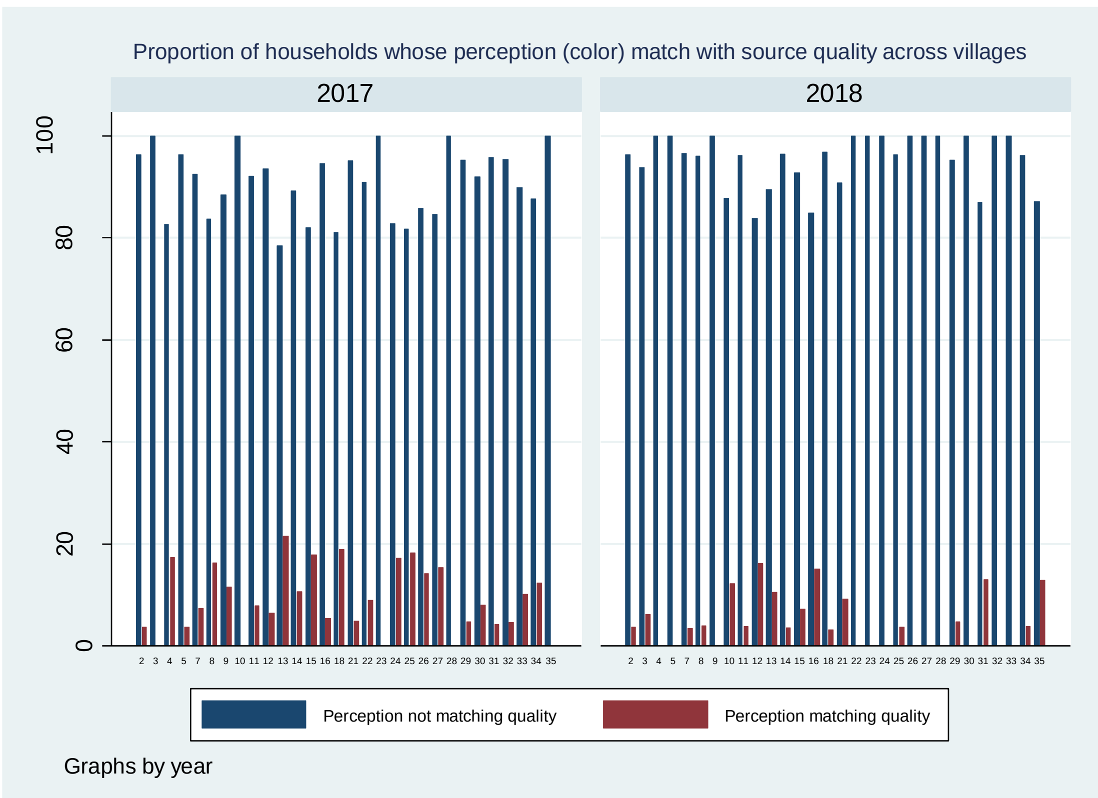

Figure 2.2 Distribution of Chosen Sources in Different Villages Showing a Match between Perception and Objective Measures of Water Quality

In the sample villages, at most, 20 percent of the household had a match between the perception of quality with the measured objective quality.

In other words, in these houses, the chosen source had a good quality of water (safe source with zero E. coli count) with coincided with their perception of quality. But there are also villages with not a single sample household for which objective quality of water matched with people's perception of quality. In the majority of the households, perception of quality does not match with objective water quality. Since people do not know their quality of water (objective tested values), the perception of quality seems to be the main factor determining source choices. We will explore the role of perception of quality in source choices in more detail in the next section of regression results.

2.5.2 Summary Statistics: Stated Preference Data

The summary statistics for the stated preference data sets are provided in the appendix Table 2.A.5. We can observe that all the household-specific variables have precisely the same average values for unlabeled sources A and B. The values of time cost of traveling to the source, taste, and color based quality perception indicators are almost the same on average. The questionnaire and choice sets for state preference data were designed in a way to ensures that these average values should be the same for all the variables of the model.

The households which were interviewed for stated preference data collection had the following characteristics. On average, the women respondents were 31 years old, were educated till 5th standard, and came from families with six members. In our sample, there was an almost equal proportion of households came from upper and lower castes groups. Furthermore, we find that on average, each household had about six assets (out of 22), and the ratio of children under 5 in the family was 0.12.

2.5.3 Estimation Results: Revealed Preference Data

2.5.3.1 Effect of Perception on Source Choices

A mixed logit approach is used where there is random heterogeneity in consumer preferences. For our data set, we assume that there are continuous unobserved variations in people's perception of drinking water quality. The four parameters of the model (travel cost, taste, color-based quality perception, and perceived risk of illness) follow a continuous normal distribution. We take into account the panel data structure[41] and use random draws for simulation of the preference parameters of the model. In Table 2.1, we present results from mixed logit estimation of source choices for two models with alternative specific constants and different attributes of water sources. In Model A, we assume that the travel cost is fixed while the other three perception-based indicators have a random normal distribution. In model B, we assume that travel cost is distributed normally. The estimated coefficients provided in Table 2.1 are the mean and standard deviation of the distribution of different variables, indicating the presence of heterogeneous preferences. From the two models in Table 2.1, it can be observed that model B has a grater log-likelihood and is a better fit for

[41] It means that each observation of a household is treated as independent point (case) for analysis.

the data. Thus, the other results for this data set (public source choice revealed preference data) are calculated using the estimation results from model B of Table 2.1. Using this model, we calculate the willingness to pay and the average marginal effects of different attributes of source choices. These results are provided in Table 2.2 and Table 2.3. The estimated coefficients from column 2 of Table 2.1 represents the relationship of choice of a source with its attributes.

This relationship is valid for the households that have preferences corresponding to the mean value of the distribution. From column 2 of Table 2.1, we can say that there is a significant negative relationship between travel cost and source choices. There is also a significant positive association between different quality perception indicators and source choices for households whose preferences lie on the average part of the distribution of these attributes.

We repeat the entire analysis for the revealed preference data set, including the choices of privately owned sources. The mixed logit regression coefficients for this data set are provided in Table 2.2. The comparison of the two models and their log-likelihood values indicate that model B with randomly distributed travel cost is a better fit to the data. We can observe from column 2 of Table 2.2 that an increase in travel cost significantly reduces the probability of choosing a source. In contrast, improvement in the perception-based indicators increases this probability. These results are valid for the households that have the preferences corresponding to the average values for travel costs and other quality perception indicators. The willingness to pay calculations and marginal effects of different perception based indicators on the choice of a source for this data set, which includes private source choices are being provided in columns 2 of Tables 2.3 and 2.4.

Table 2.1 Effect of Perception Indicators on the Choice of a Source (Revealed Preference Data with Public Sources)

Dep. variable = Choice Dummy	Model A	Model B
Mean		
Travel cost	−0.313***	−0.516***
	(0.0661)	(0.0355)
Taste of water	3.822***	4.359***
	(0.338)	(0.232)
Quality perception(color)	2.159***	2.720***
	(0.300)	(0.316)
Perceived risk of being ill	2.204*	2.036***
	(1.089)	(0.420)
SD		
Taste of water	0.160	0.688***
	(1.293)	(0.135)
Quality perception(color)	0.978	1.803***
	(0.729)	(0.351)
Perceived risk of being ill	1.325	0.513
	(2.969)	(0.358)
Travel cost		0.233***
		(0.0171)
Number of Obs.	18824	18824
Log likelihood	−775.643	−698.807
Wald (χ^2)	804.037	560.498
p-value for Wald test	0.000	0.000

Note:*se* are in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% level respectively. Alternative specific constants for different source choices included in the estimation with river as base alternative.

2.5.3.2 Willingness to Pay (WTP)

The willingness to pay for improvement in perception based water quality indicators is calculated for both sets of revealed preference data sets (without privately-owned source choices and with it). These results are provided in Table 2.3. From column 1 of

Table 2.3, we can note that a household is willing to pay 8 Indian rupees to have an improvement in the taste of the drinking water. This result is valid for the household, which has average coefficients for travel cost and taste-based perception indicator as provided in Column 2 of Table 2.1.

Table 2.2 Effect of Perception Indicators on the Choice of a Source (Revealed Preference Data including Private Sources)

Dep. variable = Choice Dummy	Model A	Model B
Mean		
Travel cost	−0.198***	−0.473***
	(0.0481)	(0.0307)
Taste of water	3.946***	4.280***
	(0.244)	(0.231)
Quality perception(color)	1.622***	1.765***
	(0.339)	(0.298)
Perceived risk of being ill	1.458***	1.241**
	(0.427)	(0.414)
SD		
Taste of water	−0.0621	0.185**
	(0.0378)	(0.0634)
Quality perception(color)	1.387	1.659***
	(0.772)	(0.224)
Perceived risk of being ill	1.630***	1.383**
	(0.493)	(0.493)
Travel cost		0.200***
		(0.0133)
Number of Obs.	25030	25030
Log likelihood	−1190.822	−988.981
Wald (χ^2)	433.172	779.147
p-value for Wald test	0.000	0.000

Note:*se* are in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% level respectively. Alternative specific constants for different source choices included in the estimation with river as base alternative.

In other words, to have a taste of water that the household like (dummy variable with value 1), they would be willing to pay 8 Rs extra per day or 240 Indian rupees per month (about 3 U.S. dollars in a month)[42]. A household with average preferences (in the distribution of cost and perception indicators) will be willing to pay about 5 Indian rupees per day (150 Indian rupees per month or about 2 U.S. dollars per month) to have a water source with better color (or no color). They will also be willing to pay 4 Rs per day (120 Indian rupees per month or 1.5 US dollars per month) to have a better source with no perceived risk of illness. The willingness to pay for revealed preference data, including privately owned sources, is provided in column 2 of Table 2.3.

Table 2.3: Willingness to Pay Estimates for Revealed Preference Data

	Willingness to pay (in Indian Rs per day)	
Variables	Public Data	Data including private source
Taste preference	8.45	9.06
Color (quality perception)	5.28	3.74
No perceived risk of illness	3.95	2.63

We find that while the willingness to pay has gone up slightly for improvements in the taste of water while it has gone down for color and no perceived risk of illness compared to data with no privately owned source choices.

[42] This may seem quite small but compare it with a daily wage of 150-200 Indian rupees for people in villages in Jharkhand. In order to have an improvement in the taste of their drinking water people are willing to give up one day's wages.

2.5.3.3 Average Marginal Effects

Table 2.4 Marginal Effects for Revealed Preference Data

Marginal effects (with no interaction terms)		
	Public data	Data including private source
Taste preference	0.36	0.41
Color (quality perception)	0.16	0.12
No perceived risk of illness	0.13	0.08

The average marginal effects of different quality perception indicators for the revealed preferences data set is being provided in Table 2.4. On average, if a household likes the taste of water, the probability of choosing that source increases by 35 percent in the data set with only public source choices and by 41 percent when privately own source choices are included in the data. If they like the color of the water, then the probability of choosing that source increases by 16 and 12 percent respectively in public and public-private revealed preference data sets. If they believe that there is no perceived risk of illness associated with a source, then the probability of choosing that source goes up by 13, and by 8 percent in public and public-private revealed preference choice data sets.

2.5.3.4 Choice Probabilities

The average choice probabilities associated with different water sources are provided in Table 2.5. These results are calculated as a comparison to the base alternative, which is the river in this specification[43]. From Table 2.5, it can be observed that the highest choice probability compared to the choice of the river is associated with Chua. It is an unexpected result but particularly relevant for this region of India. In rural Jharkhand, Chua is a small well, which is built around a natural fountain of water from in the ground. In places where water comes out from the earth naturally as a fountain or small stream (typically in low land areas in the farms in the outskirts of a village), villagers put a small structure of stones and bricks around. Compared to a well, the water level is quite high in Chua, and it tends to have a smaller circumference. During the pilot survey and group discussion with villagers, it was found that if there is a Chua in the village, people prefer to fetch water from it for drinking even if it is far away from the village.

[43] River is taken as the base alternative as we have observed that river is the most unlikely source to be chosen (most undesirable).

Table 2.5 Choice Probabilities of Different Water Sources in the Revealed Preference Data Set

Choice Probabilities (compared to the river as a base alternative)		
	Public data	Data including private sources
Hand pump	0.122	0.130
Well	0.133	0.150
Chua	0.204	0.215
Piped	0.036	0.065
Private source		0.007

In Table 2.5, we find that the average choice probabilities attached with piped water tap compared to those of the river are small. It seems that people do not prefer choosing a piped water tap for drinking purposes compared to other sources in the community. It is an interesting result that may have policy implications for providing access to safe drinking water sources in rural India. To make piped water as the primary source for drinking, we need to have campaigns or behavior change programs that could influence people's perception in favor of piped water taps.

Interaction effects

The effect of different perception based indicators on water source choices may vary based on the age and income (wealth) of the households in the data set. To check these effects, we include the interaction of age, wealth (and other control variables) with the quality perception indicators. The results of a mixed logit analysis of source choices, including the interaction terms with household-specific variables, are being provided in Tables 2.A.3 and 2.A.4 in the appendix.

2.5.4 Estimation Results: Stated Preference Data

2.5.4.1 Effect of Perception on Source Choices

In the stated preference data collection, each respondent was presented with choices between two unlabeled sources A and B, which were defined by different levels of distance to the source, taste, and color of the water. The recorded choices presented a discrete choice data set. The estimation results of a mixed logit estimation for the choice of un unlabeled water source for this stated preference data sets are presented in Table 2.6. The time cost of travel is calculated using the distance to the source and uniform speed of walking. We can observe that it has a negative relationship with source choice. Greater the distance to the source less likely is a household to choose that source for drinking. But the significance of the relationship goes away when we allow travel costs to be random instead of fixed. Since the value of the log-likelihood is higher with random cost, taste, and color-based perception of water, we would prefer the model B in Table 2.6 for our analysis. It can be observed from Table 2.6, column 2, that both taste and color based quality indicators have a significant positive effect on source choices. It is a mixed logit analysis with random preference parameters. The results presented in Table 2.6 are for the people whose preference lies at the average part of the distribution of the parameters of the model. Both taste and color based parameters of the model have three levels, 1, 2, and 3. Level 1 refers to the worst (least preferred option), and level 3 refers to the most preferred option.

Table 2.6: Effect of Perception-based Indicators on the Choice of a Source (Stated Preference Data)

Dep. variable = Choice Dummy	Model A	Model B
Mean		
Time cost	−0.0117*	−0.00913
	(0.00507)	(0.00634)
Taste of water	0.766***	0.882***
	(0.0487)	(0.0795)
Quality perception(color)	0.405***	0.588***
	(0.0539)	(0.0848)
SD		
Taste of water	0.376***	0.519***
	(0.0566)	(0.0779)
Quality perception(color)	0.270**	0.309**
	(0.0878)	(0.0957)
Time cost		0.0660***
		(0.0108)
Number of Obs.	3872	3872
Log likelihood	−1066.227	−1049.191
Wald (χ^2)	398.057	180.107
p-value for Wald test	0.000	0.000

Note:*se* are in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% level respectively. For Taste and Quality (color) variables there are 3 levels, 3 being the best taste/quality and 1 being worst taste/quality.

These levels were determined after conducting a thorough pilot survey. In the case of the taste based indicator, level 3 is for no taste of water, and level 1 refers to a salty taste of water. In the case of color based indicator, level 3 is for transparent water with no color, and level 1 is for muddy water with brown color. So, the positive sign of the estimated coefficients for taste and color based perception indicators show that if a water source has better taste or color of water, chances of choosing that water source increases. We also include household-specific control variables in the estimation by interacting them with our alternative specific variables like taste and color of the water. The results of the mixed logit estimation with the interaction effects are provided in Appendix Table 2.A.6.

2.5.4.2 Willingness to Pay (WTP)

We use the results from column 2 of Table 2.6 to calculate the willingness to pay for quality improvements for people who have their preferences at the average part of the distribution of the parameters in a heterogeneous population. These willingness to pay estimated are being provided in Table 2.7. The time cost of traveling to a source is calculated by assuming that people work for eight hours for a daily wage of 200 Indian rupees. We find that households are willing to pay 159 Indian rupees per day or 4800 Indian Rupees in a month (about 70 U.S. dollars) for improvement in the taste of the water of the source. For improvement in the color of water, they are willing to pay 94 Rs per day or about 2800 Indian rupees per month (about 40 U.S. dollars)[44].

If we compare the willingness to pay for taste and color-based quality improvement in water in the stated preference data and revealed preference data, there is a vast difference. The willingness to pay measures estimated in the stated preference data seems more than the monthly income of many rural households. While in revealed preference data, people seem willing to pay in a month amounts that are equivalent to the daily wage of a laborer. It is important to note that the willingness to pay estimates in the revealed preference data, though seemingly small, is still more than what people

[44] These calculation of time cost of travel is done under the assumption of daily average wage of 200 Indian Rupees per day and working time of 8 hours per day.

pay in villages of Jharkhand for getting piped water for a month (about 30-60 Indian rupees a month). In the literature, it is stated that the actual willingness to pay estimates would lie somewhere in between the revealed preference and stated preference estimates. Irrespective of the exact amount, the positive non-zero values indicate that households have a desire to see improvements in both the taste and the color of the water. This desire can be seen reflected in the willingness to pay estimates calculated using the reveled preference and stated preference choice data sets.

Table 2.7: Willingness to Pay Estimates for Stated Preference Data

Willingness to pay estimates (in Indian Rupees per day)		
	Taste of water	Color of water
WTP	159.19	93.89

2.6 Conclusion

Correctly choosing a drinking water source in a rural community is essential for ensuring that safe water sources are being chosen. If there are safe water sources available in a community and people are choosing other sources, then it could be a cause for concern due to the risk of water-borne diseases and implied health costs. In this paper, we found that in the villages of Jharkhand, perception-based indicators of water quality play an important role in determining source choices. Perception of water quality based on the taste of water, the color of water, or perceived risk of illness is driving the choice of water sources in rural areas along with the cost of traveling to the source. Households seem to not know about objective tests for water quality, and we found a large mismatch between the perception of quality and objective measure of water quality. From the analysis of the revealed preference data, we find that Chua (similar to a small well) has the highest choice probability among all the water sources. Chua is one of the unsafe sources of drinking water[45] , but people prefer it because they believe it has better quality.

On the other hand, the choice probability associated with piped water taps (when available in the villages) is quite low. It indicates that people's perception of quality is making them choose sources that may not be safe for drinking. The willingness to pay estimates shows the degree of people's desire to see improvements in the quality of water sources. In our field area, the quality is based on perception-based indicators of taste preference and the color of the water. Households are willing to pay about a minimum of 150 Indian Rupees per month and a maximum of 4800 Indian Rupees per month for improvements in the taste of water. If we find a way to match the objective quality of water with the perception of quality for people in rural Jharkhand, then these estimates will also tell us people's willingness to pay for providing safe water sources in their communities.

[45] We have also tested drinking water quality from different sources in the field area using presence of E. coli bacteria in the water as a proxy for quality. In all the sample collected from Chua, we have found really high presence of E. coli in water.

APPENDIX TO CHAPTER 2

Table 2.A.1: Summary Statistics for Revealed Preference Data (with Public Sources)

	River	*Chua*	*Well*	*Piped*	*HP*
Taste	0.183	0.149	0.0625	0.0409	0.0750
	(0.453)	(0.289)	(0.187)	(0.148)	(0.205)
Quality(color)	0.183	0.145	0.0583	0.0409	0.0547
	(0.453)	(0.286)	(0.182)	(0.148)	(0.179)
Perceived health risk	0.0215	0.0390	0.0190	0.00770	0.0306
	(0.130)	(0.154)	(0.106)	(0.0673)	(0.134)
Travel cost	19.43	39.58	40.60	31.98	32.52
	(24.45)	(127.6)	(79.95)	(43.32)	(69.94)
Age of female head	36.79	31.03	34.35	34.02	36.75
	(1.351)	(7.133)	(6.532)	(3.161)	(7.173)
Asset index	0.212	0.215	0.218	0.249	0.218
	(0.104)	(0.113)	(0.101)	(0.106)	(0.102)
Education	3.585	6.693	5.083	4.735	5.447
	(4.880)	(4.795)	(4.919)	(5.136)	(4.832)
Nearest source	0.525	0.366	0.493	0.447	0.518
	(0.501)	(0.482)	(0.500)	(0.498)	(0.500)
Married(=1 if yes)	0.966	0.834	0.939	0.969	0.933
	(0.182)	(0.372)	(0.239)	(0.174)	(0.250)
Upper caste	0.280	0.126	0.289	0.279	0.326
	(0.451)	(0.332)	(0.453)	(0.449)	(0.469)
Lower caste	0.720	0.874	0.710	0.721	0.673
	(0.451)	(0.332)	(0.454)	(0.449)	(0.469)
Child under 5	0.144	0.146	0.126	0.125	0.134
	(0.145)	(0.191)	(0.174)	(0.135)	(0.174)
Female head	0.0678	0.139	0.121	0.168	0.136
	(0.252)	(0.346)	(0.326)	(0.375)	(0.343)

Note: Mean values with respect to different source types. *se* are in parentheses.
HP stands for Handpump and Chua is a smaller well.

Table 2.A.2: Summary Statistics for Revealed Preference Data (with Private Sources)

	River	*Chua*	*Well*	*Piped*	*HP*	*Own*
Taste	0.282	0.148	0.0577	0.0704	0.0682	0.180
	(0.657)	(0.288)	(0.175)	(0.227)	(0.191)	(0.358)
Quality(color)	0.282	0.144	0.0533	0.0671	0.0497	0.160
	(0.657)	(0.284)	(0.169)	(0.223)	(0.166)	(0.347)
Perceived health risk	0.0413	0.0389	0.0174	0.0368	0.0282	0.0598
	(0.169)	(0.154)	(0.0988)	(0.178)	(0.125)	(0.229)
Travel cost	22.40	39.95	53.20	52.24	47.93	31.05
	(27.89)	(127.2)	(195.5)	(91.41)	(222.2)	(82.66)
Age of female head	36.80	31.14	34.56	34.29	36.55	35.92
	(1.325)	(7.203)	(6.370)	(3.725)	(6.942)	(6.728)
Asset index	0.230	0.217	0.231	0.259	0.233	0.233
	(0.114)	(0.112)	(0.109)	(0.112)	(0.110)	(0.112)
Education	3.694	6.622	5.308	5.786	5.721	5.448
	(5.031)	(4.819)	(4.947)	(5.183)	(4.866)	(4.908)
Nearest source	0.552	0.370	0.546	0.581	0.566	0.559
	(0.499)	(0.484)	(0.498)	(0.494)	(0.496)	(0.497)
Married(=1 if yes)	0.955	0.831	0.938	0.937	0.936	0.940
	(0.208)	(0.376)	(0.242)	(0.243)	(0.245)	(0.238)
Upper caste	0.306	0.130	0.340	0.401	0.371	0.350
	(0.463)	(0.336)	(0.474)	(0.491)	(0.483)	(0.477)
Lower caste	0.694	0.870	0.659	0.599	0.628	0.649
	(0.463)	(0.336)	(0.474)	(0.491)	(0.483)	(0.477)
Child under 5	0.155	0.144	0.127	0.120	0.132	0.131
	(0.248)	(0.190)	(0.181)	(0.142)	(0.185)	(0.177)
Female head	0.0896	0.143	0.126	0.144	0.139	0.135
	(0.287)	(0.350)	(0.332)	(0.352)	(0.346)	(0.341)

Mean values with respect to different source types. *se* are in parentheses.
HP stands for Handpump, Chua is a smaller well and Own stands for privately constructed sources.

Table 2.A.3: Effects of Perception Indicators and Household-specific Variables on Source Choice (Revealed Preference Data with Public Sources)

Dep. variable = Choice Dummy	Model C	Model D
Mean		
Travel cost	−0.326***	−0.522***
	(0.0716)	(0.0361)
Taste	3.351*	3.813*
	(1.370)	(1.615)
Quality(color)	3.579*	4.037*
	(1.727)	(2.034)
Perceived health risk	−0.627	−0.706
	(2.262)	(2.371)
Age*Color	−0.0226	−0.0442
	(0.0407)	(0.0483)
Assets*Color	−0.378	−0.299
	(2.822)	(3.341)
Age*Taste	0.00369	0.00663
	(0.0263)	(0.0307)
Age*Risk of illness	0.0263	0.0303
	(0.0527)	(0.0595)
Assets*Taste	2.121	2.544
	(2.110)	(2.443)
Assets*Risk of illness	8.287	4.022
	(7.160)	(5.695)
SD		
Taste	0.00893	−0.132***
	(0.0394)	(0.0390)
Quality(color)	0.117**	0.215***
	(0.0434)	(0.0527)
Perceived health risk	0.126*	0.165
	(0.0544)	(0.141)
Age*Color	0.00727***	0.0170**
	(0.00183)	(0.00587)
Assets*Color	−2.422***	0.0325
	(0.460)	(0.177)
Age*Taste	−0.00764***	−0.00142
	(0.00129)	(0.000990)
Age*Risk of illness	0.00375	0.00982
	(0.00205)	(0.00585)
Assets*Taste	1.796***	−0.911***
	(0.203)	(0.215)
Assets*Risk of illness	3.494*	2.758**
	(1.574)	(1.059)
Travel cost		0.235***
		(0.0205)
Number of Obs.	18824	18824
Log likelihood	−739.983	−668.516
Wald (χ^2)	1654.523	1379.357
p-value for Wald test	0.000	0.000

Note: *se* are in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% level respectively. Alternative specific constants included in the estimation with river as base alternative.

Table 2.A.4: Effects of Perception Indicators and Household-Specific Variables on Source Choice (Revealed Preference Data including Private Sources)

Dep. variable = Choice Dummy	Model C	Model D
Mean		
Travel cost	−0.197***	−0.495***
	(0.0499)	(0.0329)
Taste	4.543***	4.940**
	(1.213)	(1.649)
Quality(color)	3.158*	3.855
	(1.509)	(2.027)
Perceived health risk	−0.587	0.0667
	(2.197)	(3.662)
Age*Color	−0.00571	−0.0341
	(0.0353)	(0.0385)
Assets*Color	−1.984	−2.182
	(2.362)	(2.629)
Age*Taste	0.000568	0.00596
	(0.0245)	(0.0282)
Age*Risk of illness	0.0331	0.00701
	(0.0457)	(0.0587)
Assets*Taste	0.167	0.371
	(1.748)	(2.091)
Assets*Risk of illness	1.541	1.216
	(3.774)	(4.206)
SD		
Taste	−0.295***	−0.0189
	(0.0650)	(0.0567)
Quality(color)	0.225***	0.0283
	(0.0595)	(0.0714)
Perceived health risk	0.207*	0.0791
	(0.0981)	(0.187)
Age*Color	0.0123***	0.0101
	(0.00353)	(0.00736)
Assets*Color	1.788*	2.166***
	(0.901)	(0.469)
Age*Taste	0.00278**	0.00177
	(0.000864)	(0.00209)
Age*Risk of illness	0.0177***	−0.00649
	(0.00496)	(0.00442)
Assets*Taste	−0.474*	−0.0960
	(0.210)	(0.274)
Assets*Risk of illness	3.259*	−2.711
	(1.305)	(1.985)
Travel cost		0.209***
		(0.0144)
Number of Obs.	25030	25030
Log likelihood	−1087.841	−888.816
Wald (χ^2)	1554.769	1242.553
p-value for Wald test	0.000	0.000

Note:se are in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% level respectively. Alternative specific constants included in the estimation with river as base alternative.

Table 2.A.5: Summary Statistics for Stated Preference Data

	SourceA	*SourceB*
Taste of water	1.952	2.023
	(0.866)	(0.851)
Quality perception(color)	2.202	2.054
	(0.869)	(0.708)
Time cost	12.86	12.40
	(8.140)	(8.214)
Age of female head	31.29	31.29
	(8.245)	(8.245)
Asset index	0.259	0.259
	(0.113)	(0.113)
Education of female head	5.277	5.277
	(5.109)	(5.109)
Married(=1 if yes)	0.975	0.975
	(0.156)	(0.156)
Upper caste	0.475	0.475
	(0.500)	(0.500)
Lower caste	0.525	0.525
	(0.500)	(0.500)
Family size	6.087	6.087
	(2.349)	(2.349)
Child under 5	0.125	0.125
	(0.133)	(0.133)

Note: Standard deviations are in parentheses.
For Taste and Quality (color) variables there are 3 levels, 3 being the best taste/quality and 1 being worst taste/quality.

Table 2.A.6: Effects of Perception-based Indictors and Household-specific Variables on Source Choice (Stated Preference Data)

Dep. variable = Choice Dummy	Model C	Model D
Mean		
Time cost	−0.0116*	−0.00910
	(0.00506)	(0.00647)
Taste of water	1.218***	1.449***
	(0.208)	(0.283)
Quality perception(color)	0.635**	0.855**
	(0.240)	(0.288)
Female education*Taste	−0.0238*	−0.0285*
	(0.0106)	(0.0128)
Female education*Quality	0.00303	−0.000282
	(0.0117)	(0.0135)
Age*Taste	−0.0167**	−0.0206**
	(0.00548)	(0.00718)
Age*Quality	−0.00632	−0.00836
	(0.00642)	(0.00763)
SD		
Taste of water	0.340***	0.470***
	(0.0675)	(0.0874)
Quality perception(color)	0.235*	0.235
	(0.0918)	(0.137)
Female education*Taste	−0.00332	−0.00365
	(0.00822)	(0.0159)
Female education*Quality	−0.00271	0.00630
	(0.0141)	(0.0181)
Age*Taste	−0.000242	0.000699
	(0.00203)	(0.00368)
Age*Quality	−0.00358	0.00479
	(0.00228)	(0.00882)
Time cost		0.0677***
		(0.0111)
Number of Obs.	3872	3872
Log likelihood	−1058.840	−1042.500
Wald (χ^2)	444.402	186.903
p-value for Wald test	0.000	0.000

Note:*se* are in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% level respectively. For Taste and Quality (color) variables there are 3 levels, 3 being the best taste/quality and 1 being worst taste/quality.

Table 2.A.7 Attributes of Water Sources for the Construction of Choice Sets used in Stated Preference Data Collection

	Attributes of Water Sources		
Levels of Attributes	Distance	Color of water	Taste of water
1	>300 meters	Muddy (brown)	Bitter (Hard water)
2	50-300 meters	Yellow (somewhat clear)	Salty
3	<50 meters	Clear (no color)	No Taste (sweet)

Choice set	**Water source A**	**Water source B**
Distance to the source from house (in meters)	50-300 meters	0-50 meters
Color of water	Muddy water	No color
Taste of water	Hard water	No taste
Choice (by the respondent)		

Figure 2.A.1 Choice Set Example used in the Stated Preference Data Collection

REFERENCES FOR CHAPTER 2

Adamowicz, W., Jennings, S. & Coyne, A., 2019. A sequential choice model of recreation behavior. *Western Journal of Agricultural Economics*, 15(1), 91–99.

Adler, D. & Loewenstein, G.,1995. A bias in the prediction of tastes. *Royal Economic Society*, 105, 929–937.

Allcott, H. & Rogers, T., 2012. The short-run and long-run effects of behavioral interventions : Experimental evidence from energy conservation, *National Bureau of Economic Research Working Papers*, 18492.

Andersson, T. et al., 2018. Sequential school choice with public and private schools. *Department of Economics, Lund University, Working papers 2018*, 39.

Apesteguia, J. & Ballester, M.A., 2013. Choice by sequential procedures. *Games and Economic Behavior*, 77(1), 90–99.

Arcidiacono, P. & Miller, R.A., 2011. CCP estimation of dynamic discrete choice models with unobserved heterogeneity. *Econometrica*, 79(6), 1823-1867.

Ashraf, B.N., Berry, J. & Shapiro, J.M., 2010. Can higher prices stimulate product use ? Evidence from a field experiment in Zambia. *The American Economic Review*, 100(5), 2383-2413

Bastin, F. & Cirillo, C., 2010. Reducing simulation bias in mixed logit model estimation. *Journal of Choice Modeling*, 3(2), 71–88.

Benson, A.R., Kumar, R. & Tomkins, A., 2018. A discrete choice model for subset selection. *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, 37-45.

Birch, L.L., 2016. Effects of peer models' food choices and eating behaviors on preschoolers' food preferences. *Child Development*, 51(2), 489–496.

Blackstock, K.L. et al., 2010. Understanding and influencing behaviour change by farmers to improve water quality. *Science of the Total Environment*, 408, 5631–

5638.

Boone, C., Glick, P. & Sahn, D.E., 2011. Household water supply choice and time allocated to water collection: Evidence from Madagascar. *Journal of Development Studies*, 47(12), 1826-1850.

Bowles, S., 1998. Endogenous preferences : The cultural consequences of markets other economic institutions. *Journal of Economic Literature,* 36(1), 75–111.

Brent, D.A., Cook, J.H. & Olsen, S., 2015. Social comparisons, household water use, and participation in utility conservation programs: Evidence from three randomized trials. *Journal of the Association of Environmental and Resource Economists*, 2(4), 597–627.

Briscoe, J. et al., 1981. How Bengali villagers choose sources of domestic water. *Water Supply and Management*, 5, 165-181.

Brumfield, R.G., Adelaja, A.O. & Lininger, K., 1993. Consumer tastes, preferences, and behavior in purchasing fresh tomatoes. *Journal of American Society of Horticulture Sciences*, 118(3), 433–438.

Buis, M. L., 2011. The Consequences of unobserved heterogeneity in a sequential logit model. *Research in Social Stratification and Mobility*, 29, 247-262.

Cairncross, S., 2003. Editorial: Water supply and sanitation: Some misconceptions, *Tropical Medicine and International Health,* 8(3), 193-195.

Carlsson, F., García, J.H. & Löfgren, Å., 2010. Conformity and the demand for environmental goods. *Environmental Resource Economics*,47, 407–421.

Cidu, R., Frau, F. & Tore, P., 2011. Drinking water quality : Comparing inorganic components in bottled water and Italian tap water. *Journal of Food Composition and Analysis*, 24(2), 184–193.

Clasen, T. et al., 2007. Interventions to Improve Water Quality for Preventing Diarrhoea: Systematic Review and Meta-analysis. *British Medical Journal*, 334(7597), 782–785.

Cook, J., Kimuyu, P., Blum, A.G., et al., 2015. A Simple Stated Preference Tool for Estimating the Value of Travel Time in Rural Africa, *Environment for Development Discussion Paper Series*.

Cook, J., Kimuyu, P. & Whittington, D., 2015. The Costs of Coping with Poor Water Supply in Rural Kenya, *Environment for Development Discussion Paper Series*.

Daniel, C., 2016. Economic constraints on taste formation and the true cost of healthy eating. *Social Science & Medicine*, 148, 34–41.

Day, B. & Mourato, S.,1998. Willingness to pay for water quality maintenance in Chinese rivers. *Centre for Social and Economic Research on the Global Environment, University College London, and the University of East Anglia.*

Devesa, R. et al., 2012. Influence of minerals on the taste of bottled and tap water : A chemometric approach. *Water Research*, 47, 693-704.

Devoto, F. et al., 2012. Happiness on tap: Piped water adoption in urban Morocco. *American Economic Review: Economics Policy*, 4(4), 68–99.

Dhar, R. et al., 2000. Consumer choice between hedonic and utilitarian goods. *Journal of Marketing Research*, 37(1), 60–71.

Dinelli, E. et al., 2012. Comparative study between bottled mineral and tap water in Italy. *Journal of Geochemical Exploration*, 112, 368–389.

Dinelli, E. et al., 2012. Major and trace elements in tap water from Italy. *Journal of Geochemical Exploration*, 112, 54–75.

Doria, M. de F., Pidgeon, N. & Hunter, P.R., 2009. Perceptions of drinking water quality and risk and its effect on behaviour: A cross-national study. *Science of the Total Environment*, 407(21), 5455–5464.

Drewnowski, A., 1997. Taste preferences and food intake. *Annual Review of Nutrition*, 17, 237-253.

Dubé, C. & Malmberg, K.J., 2018. A Bayesian approach to discriminating between biased responding and sequential dependencies in binary choice data. *Decision*, 5(1), 16-41.

Dupas, P. et al., 2013. Micro-ordeals, targeting, and habit formation. *Working Paper*.

Eliza, M. et al., 2008. Comparison of the bacteriological quality of tap water and bottled mineral water. *International Journal of Hygiene and Environmental Health*, 211, 504–509.

Eriksen, S. & Anker, O., 2008. *Sample selection and taste correlation in discrete choice transport modeling*. Thesis DTU Transport, Technical University of Denmark.

Espinosa-García, A.C. et al., 2015. Drinking water quality in a Mexico city university community: Perception and preferences. *EcoHealth*, 12(1), 88–97.

Etale, A., Jobin, M., & Siegrist, M. (2018). Tap versus bottled water consumption: The influence of social norms, affect and image on consumer choice. *Appetite*, 121, 138–146.

Falahee, M. & Macrae, A.W., 1995. Consumer appraisal of drinking water: Multidimensional scaling analysis. *Food Quality and Preferences*, 6(95), 327–332.

Falahee, M. & Macraeb, A.W., 1997.Perceptual Variation among Drinking Water: The Reliability of sorting and Drinking data for Multidimensional Scaling. *Food Quality and Preferences*, 8(5), 389–394.

Faruk, G. & Pesendorfer, W., 2005. The revealed preference theory of changing tastes. *The Review of Economic Studies*, 72(2), 429–448.

Fehr, E. et al., 2011. Tastes, castes, and culture : The influence of society on preferences. *The Economic Journal*, 121(556), F396-F412.

Ferraro, P.J., Miranda, J.J. & Price, M.K., 2011. The persistence of treatment effects with norm-based policy instruments: Evidence from a randomized environmental policy experiment. *American Economic Review*, 101(3), 318–322.

Ferraro, P.J. & Price, M.K., 2013. Using nonpecuniary strategies to influence behavior: Evidence from a large-scale field experiment. *Review of Economics and Statistics*, 95(1), 64–73.

Freidin, E., Aw, J. & Kacelnik, A., 2009. Sequential and simultaneous choices : Testing the diet selection and sequential choice models. *Behavioural Processes*, 80, 218–223.

Fu, X. & Juan, Z., 2015. Estimation of multinomial probit-kernel integrated choice and latent variable model : Comparison of one sequential and two simultaneous approaches. *Transportation*, 535.

Hensher, D.A. & Greene, W.H., 2006. *The Mixed Logit Model : The State of Practice and Warnings for the Unwary*.

Hibbard, B.J.H. et al., 2012. An experiment shows that a well-designed report on costs and quality can help consumers choose high-value health care. *Health Affairs*, 31(3), 560–568.

Hindman Persson, T., 2002. Household choice of drinking-water source in the Philippines, *Asian Economic Journal*, 16(4).

Hole, A.R. & Kolstad, J.R., Mixed logit estimation of willingness to pay distributions : A comparison of models in preference and WTP space using data from a health-related choice experiment. *Empirical Economics*, 42, 445–469.

Huber, A.C. & Mosler, H.J., 2013. Determining the differential preferences of users of two Fluoride-free water options in rural Ethiopia. *Journal of Public Health (Germany)*, 21(2), 183–192.

Inauen, J. & Mosler, H.-J., 2016. Mechanisms of behavioural maintenance: Long-term effects of theory-based interventions to promote safe water consumption. *Psychology & Health*, 31(2), 166–183.

Jalan, J. & Ravallion, M., 2003. Does piped water reduce diarrhoea for children in rural India. *Journal of Econometrics*, 112(August), 153–173.

Jalan, J. & Somanathan, E., 2008. The importance of being informed: Experimental evidence on demand for environmental quality. *Journal of Development Economics*, 87(1), 14–28.

Jessoe, K., 2013. Improved source, improved quality - demand for drinking water quality in rural India. , *Journal of Environmental Economics and Management*, 66(3), 460–475.

Joshua, B. et al., 2016. Water quality violations and avoidance behavior : Evidence from bottled water consumption, *The American Economic Review*,101(3), 448-453.

Kinyuru, R., Muathe, N., Makau, S., & Kerre, F. P. (2014). The moderating effect of industrial context on the relationship between brand equity and consumer choice in branded bottled water Nairobi, Kenya. *European Scientific Journal*, *10*(4), 1857–7881.

Kremer, M. et al., 2011. Spring cleaning: Rural water impacts, valuation, and property rights institutions. *Quarterly Journal of Economics*, 126(1), 145–205.

Kremer, M., Ahuja, A. & Peterson-zwane, A., 2010. Providing safe water : Evidence from randomized evaluations, *Harvard Environmental Economics Program Discussion papers*, 10-23.

Lantagne, D., Quick, R. & Mintz, E., 2006. Household water treatment and safe storage options in developing countries: A review of current implementation practices, *Woodrow Wilson Quarterly*, 17–38.

Lanz, B., & Provins, A. (2016). The demand for tap water quality: Survey evidence on water hardness and aesthetic quality. *Water Resources and Economics*, 16(October), 52–63.

Larson, D.M. & Shaikh, S.L., 1999. Empirical specification requirements for two-constraint models of recreation demand. *Selected paper for the 1999 Annual Meeting of the American Agricultural Economics Association.*

Lee, C. et al., 2007. Valuing drinking water quality improvement using a Bayesian analysis of the multinomial probit model. *Applied Economic Letters*, 14, 255-259.

Lipscomb, M. & Mobarak, A.M., 2007. Decentralization and water quality: Evidence from the re-drawing of county boundaries in Brazil. *BREAD working paper*, (159).

Logan, T.D. & Rhode, P.W., 2008. Movable feasts : A new approach to endogenizing tastes.*Working paper.*

Louviere, J.J., Flynn, T.N. & Carson, R.T., 2010. Discrete choice experiments are not conjoint analysis. *Journal of Choice Modeling*, 3(3), 57–72.

Madanat, S. & Humplick, F., 1993. A model of household choice of water supply systems in developing countries. *Water Resources Research*, 29(5), 1353–1358.

Mak, V. et al., 2018. A network ridesharing experiment with sequential choice of transportation mode. *Theory and Decision*, 85(3), 407–433.

Marcussen, H. et al., 2013. Sensory properties of Danish municipal drinking water as a function of chemical composition. *Food Research International*, 54(1), 389–396.

Mccauley, A.P. et al., 1990. Changing water-use patterns in a water-poor area: Lessons for a Trachoma intervention project. *Sm. Sci. Med*, 31(11), 1233–1238.

Mu, X. & Briscoe, J., 1990. Modeling village water demand behavior' A discrete choice approach. *Water Resource Research*, 26(4), 521–529.

Nagakura, B.D. & Kobayashi, M., 2009. Testing the sequential logit model against the nested logit model. *The Japanese Economic Review*, 60(3), 345–361.

Neelamegham, R. & Jain, D., 1999. Consumer choice process for experience goods : An econometric model and analysis. *Journal of Marketing Research*, 36(3), 373–386.

Null, C. et al., 2012. Safe drinking water. Who is willing to pay the price? *International Initiative for Impact Evaluation*, Systematic Review 6.

Onjala, J., Ndiritu, S.W. & Stage, J., 2014. Risk perception, choice of drinking water and water treatment: Evidence from Kenyan towns. *Journal of Water, Sanitation,* *and Hygiene for Development*, 4(2), 268–281.

Pérez-Flores, M. E. et al., 2017. Perceived versus actual water quality: Community studies in rural Oaxaca, Mexico. *Science of The Total Environment*, 622–623, 626–634.

Persson, T.H., 2002. Household choice of drinking–water Source in the Philippines. *Asian Economic Journal*, 16(4), 303–316.

Puget, S. et al., 2010. Tap water consumers differ from non-consumers in chlorine flavor acceptability but not sensitivity. *Water Research*, 44, 956–964.

Raveau, S. et al., 2010. Sequential and simultaneous estimation of hybrid discrete choice models: Some new findings. *Transportation Research Record*, 2156, 131–139.

Revelt, D. & Train, K., 1997. Mixed logit with repeated choices: Household's choices of appliance efficiency level. *Review of Economics and Statistics*., 1–33.

Rosen, S. et al., 1999. Household water resources and rural productivity in Sub-Saharan Africa : A review of the evidence. *Harvard Institute for International Development.*

Rotterdam, G. & Bo, L., 2017. Elderly travel frequencies and transport mode choices in Greater Rotterdam, the Netherlands. *Transportation*, 44, 831–852.

Sajjadi, S. A. et al., 2016. Consumer perception and preference of drinking water sources. *Electronic Physician,* 8(11), 3228–3233.

Shaheed, A. et al., 2014. Water quality risks of 'improved' water sources : evidence from Cambodia. *Tropical Medicine and International Health*, 19(2), 186–194.

Sipos, L. et al., 2012. Discrimination of mineral waters by electronic tongue, sensory evaluation, and chemical analysis. *Food Chemistry,* 135, 2947–2953.

Sydney, N. & Krueger, R., 2017. Random taste heterogeneity in discrete choice models : Flexible nonparametric finite mixture distributions. *Transportation*

Research Part B: Methodological, 106, 76–101.

Teillet, E. et al., 2010. Consumer perception and preferences of bottled and tap water. *Journal of Sensory Studies*, 25, 463–480.

Teillet, E. et al., 2010. Sensory methodologies and the taste of water. *Food Quality and Preference*, 21(8), 967–976.

Viscusi, W.K., Huber, J. & Bell, J., 2015. The private rationality of bottled water drinking. *Contemporary Economic Policy*, 33(3), 450–467.

Völker, S., Schreiber, C. & Kistemann, T., 2010. Drinking water quality in household supply infrastructure — A survey of the current situation in Germany. *International Journal of Hygiene and Environmental Health*, 213, 204–209.

Whittington, D., Briscoe, J., Mu, X., & Barron, W. (1990). Estimating the Willingness to Pay for Water Services in Developing Countries: A Case Study of the Use of Contingent Valuation Surveys in Southern Haiti. *Economic Development and Cultural Change*, 38(2), 293.

Whittington, D., Mu, X. & Roche, R., 1990. Calculating the value of time spent collecting water: Some estimates for Ukunda, Kenya. *World Development* 18(2), 269–280.

Whittington, D. et al., 2009. How well is the demand-driven, community management model for rural water supply systems doing? Evidence from Bolivia, Peru, and Ghana. *Water Policy*, 11(6), 696–718.

WHO, 2007. Household Water Treatment and Safe Storage Technologies. 1–13.

Zeithaml, A. V., 1988. Consumer perceptions of price, quality, and value : A means-end model and synthesis of evidence. *Journal of Marketing*, 52(3), 2–22.

CHAPTER 3
ACCESS TO PIPED WATER, TIME SAVINGS, AND ABSENTEEISM IN SCHOOL: EVIDENCE FROM INDIA

3.1 Introduction

School attendance is essential for enrollment rates, completed grades, and overall school performance of children. It can affect the school dropout rates and ultimately may affect children's earning potential as an adult. The problem of school absenteeism is more prevalent for children in rural areas of developing countries like India. Schools are sometimes the only place where these children can study. They often come from households with illiterate parents, have no study environment at home, and have no money to pay for tuition and coaching classes outside schools. In such a situation, absenteeism from school can have a tremendous impact on the school performance of these children. One prominent cause for school absenteeism for children could be the burden of water collection from water sources situated far away from their houses. If the children are engaged in water collection from outside sources, then the time spent in water collection may cause delays and absenteeism from schools. As mostly the women and girls in the household are engaged in water collection, the effect of water collection could be more on the girls as compared to boys. Additionally, if the adult women in a family are engaged in water collection, it may result in the children, especially girls, in these households undertaking other chores in place of their mothers, sisters, or aunts. The burden of these additional chores may also indirectly cause them to miss school.

In rural India, the process of acquiring access to in-house piped water taps works at two levels. The government selects a village for establishing a piped water system. In rural Indian society, communities do not seem to make a group decision for accessing piped water inside a village. The decision to bring a piped water scheme in a village is imposed from above by the government or other development agencies. This decision is generally not driven by pressure from the community. Once the piped water system has come into the village, different households self-select themselves into the piped water system. The establishment costs and user charges for obtaining a piped water connection inside a community along with the preferences of the head of a household for clean water, and better child school outcomes drive the self-selection into piped water.

We control for the household level selection into piped water by using an instrumental variable based on the non-self-community ratio[46] of access to piped water. The non-self community ratio is the proportion of households in a village that choose to get a piped water connection, excluding the household under consideration. Under the assumption that most of the variation in access to piped water lies at the household level and not at the community level, the non-self-community ratio will affect the time spent in the collection of water and absenteeism from school only through household access to piped water. An instrumental variable based on the non-self-community ratio

[46] Non-self-community ratio instrumental variable for household access to piped water is constructed by undertaking the following calculations. Total number of households that have access to piped water in a village are calculated with deduction of 1 from it if the concerned household have access to piped water. A ratio of this number with the total number of households in a village provides us with the non-self-community ratio for a household.

seems to work well in predicting the household access to piped water after controlling for observed variables.

In this paper, we analyze the effect of access to piped water on absenteeism from school for children who were in the primary and secondary school in 2005 in rural India. In India, only about 40 percent of rural households have access to in-house piped water taps. In the absence of access to piped water, family members travel outside their homes to fetch water for drinking, cooking, bathing, and other household purposes. The time spent in collecting water from outside sources entails a loss of time for other activities like going to school. We use panel data of two years from the Indian Human Development Survey (IHDS) (with a gap of seven years) to find out the effect of access to piped water on school absenteeism for children in rural India.

In the literature, there are papers on the effect of absenteeism of teachers from schools on school outcomes of children. But not many studies have focused on the impact of child absenteeism from school on school outcomes. Furthermore, until very recently, the literature on access to piped water had focused primarily on its health effects. This paper is the first study connecting access to piped water with school absenteeism for children in the Indian context. It makes an essential contribution to the literature by incorporating both the time-saving effects of piped water and school absenteeism from school for children in one study. We find that investment in piped water infrastructure will not only reduce the time spent in water collection; it can also affect the absenteeism for school going children in houses with access to pipe water.

In section 3.2, we provide a brief literature review for this study. Section 3.3 of the paper talks about the drinking water policy in India. Section 3.4 contains a description of the data used in the analysis. In section 3.5, we provide estimation methodology and reduced-form equations. In sections 3.6, we provide the estimation results and their discussion. Section 3.7 presents the conclusion of this paper.

3.2 Literature Review

The literature on the effect of household access to piped water concentrates on child health and water-borne diseases like diarrhea (Arnold & Colford, 2007; Clasen et al., 2007; Devoto et al., 2012; Galiani et al., 2005; Gamper-Rabindran et al., 2010; Jalan & Ravallion, 2003; Kremer et al., 2011; Lamichhane & Mangyo, 2011; Mangyo, 2008; Merrick, 1985; Timmins et al., 2008). Only in recent years, the focus has shifted on the effect of access to piped water on time spent in the collection of water and substitution of this saved time for other activities, including going to school for children.

Some studies in recent years have analyzed the labor time use of women in a developing country and the way infrastructural changes like piped water can affect the distribution of time of women between different activities (Crow et al., 2012; Ilahi & Grimard, 2000; Koolwal & Walle, 2013; Meeks, 2017; Null et al., 2012). The labor time use pattern of women in a rural community in a developing country is often restricted by the existing labor markets and home production technologies. The home production technologies are dependent on the state of existing infrastructure and the institutional structure of the rural community (Fontana & Natali, 2008). Lack of necessary infrastructure like piped water taps implies that women spend a lot of time on household chores and less time on market-based production activities (Koolwal & Walle, 2013). In his paper, Meeks (Meeks, 2017) finds that access to piped water tends to reduce the time spent in the collection of water at the household level by 170 minutes on average per day. Furthermore, this saved time in water collection leads to an increase in the time spent on leisure activities by 80 minutes and in farm labor by 90 minutes per day. He uses a panel data of Kyrgyzstan to derive these results.

One of the papers on the time-saving effects of access to piped water is by Ilhai and Grimrad (Ilahi & Grimrad, 2000). In this paper, they use a cross-sectional data of 1991 to analyze the effect of access to piped water on time allocated to other activities by the woman in Pakistan. They find that the distance to the water source affects both the decision to collect water from outside and the time spent in water collection by a woman. The results from this paper indicate that 'as access to water deteriorates, there is a reduction in the probability of collection of water, but for those women who continue to collect water more time has to be spent in water collection' (Ilahi & Grimrad, 2000). This paper concludes that access to piped water leads to time savings and may increase the time spent in other activities by women, but it does not look at its effect on school outcomes per se.

There could be a substitution of the freed-up time from water collection as a result of access to piped water for other household chores like preparing children for school or child care. In the literature, the effect of improvement in water infrastructure on time spent in child care is ambiguous. It can work in two ways. If the adult women in the family were mainly engaged in child care, then they can use the freed-up time from water collection for child care. It can also mean that they might use freed up time for other market-based labor activities like a farm or non-farm work, which might imply a higher burden on adolescent girls to take care of younger children in the family (Koolwal & Walle, 2013).

Similarly, the effect of access to piped water at home on school outcomes for girls going to school is not very clear (Masuda & Cook, 2013; Nauges & Strand, 2013; Zhang & Colin, 2016). Some papers find that access to piped water reduces the burden of water collection on adolescent girls and thus improves their school outcomes (Nauges & Strand, 2013). But it is also possible that women in a household with access to piped water taps substitute their freed-up time from water collection for farm or non-farm activities, leaving the burden of child care or other household chores on girls in the family. In such cases, the increased burden of child care or other household chores on adolescent girls may adversely affect their school outcomes. In their paper, Nauges and Strand (Nauges & Strand, 2013) use a pseudo panel of four years of Demographic Household Survey (DHS) data for children in the school-going age in Ghana to analyze the effect of time spent in the collection of water on school attendance of girls at the community level. They find that for the communities that spend more than twenty minutes in water collection, there is a significant reduction in the average school attendance of girls. In another paper, Masuda and Cook (Masuda & Cook, 2013) use a quasi-experimental method and find that in Ethiopia, improved water access positively affects boy’s school attendance compared to girls.

Most of the papers in this literature have focused on the short-term effects of piped water on health, time savings, and school outcomes. In contrast, Zhang and Colin (Zhang & Colin, 2016) paper analyses the long-term effects of treated piped water systems on completed years of schooling for youth between 18 to 25 years of age. Their (Zhang & Colin, 2016) paper is a study based in China that conducted surveys after 20 years of operation of the piped water systems. They find that there is a significant impact of access to treated piped water on the completed years of schooling, especially for girls in their data.

The papers that studied school absenteeism for children with improved water sources did not explicitly analyze its time-saving results. The literature on the effect of access to piped water on time use patterns of women did not look at its effect on the school attendance of children. This paper address that gap in the literature by studying both of these effects of access to piped water in one paper, establishing a link between changes in time spent in the collection of water and school absenteeism for children. Furthermore, this paper addresses the issue of endogeneity to some extent by using panel data with the child fixed effect model along with the use of an instrumental variable for access to piped water. It is the first study addressing this problem in the context of South Asia, specifically in rural India.

3.3. Drinking-Water in India and Government Water Policy

In India, there is a central government policy initiative National Rural Drinking Water Program (NRDWP) for providing clean drinking water to households in rural areas. One of the main objectives of the NRDWP is to provide houses in villages with access to the piped water instead of other previously favored water sources like hand pumps. By 2017, this policy directives aim to ‘ensure that at least fifty percent of rural households are provided with piped water supply; at least thirty-five percent of rural households have piped water supply with a household connection; less than twenty percent use public taps, and less than forty-five percent use hand pumps or other safe and adequate private water sources.’

Apart from this central government policy, there are also state-specific policies providing piped water access to rural households in different states of India.

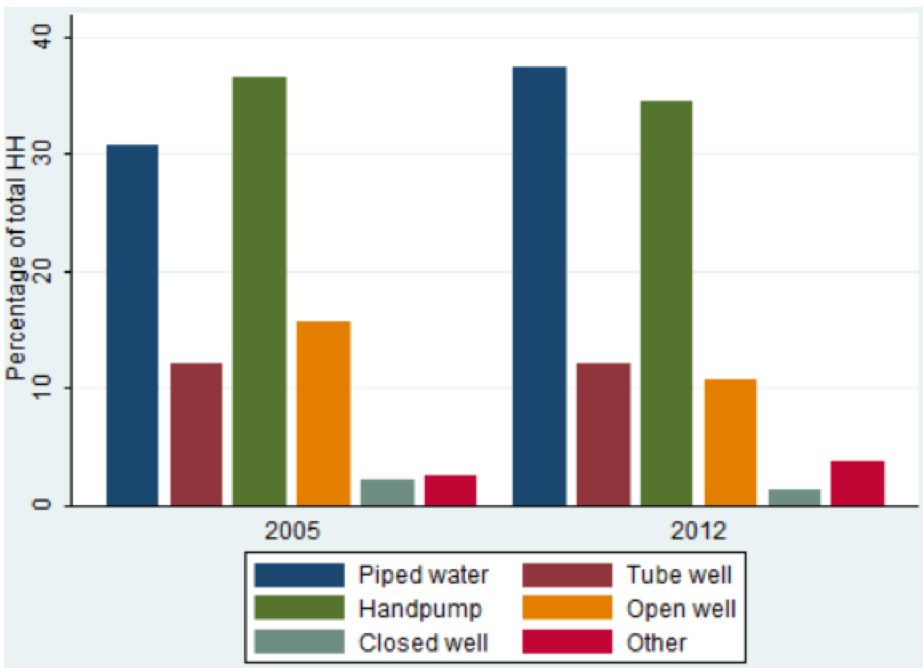

Figure 3.1 Distribution of Sources of Drinking Water in Rural India

The distribution of water sources in rural India for the years 2005 and 2012 is being provided in Figure 3.1. It can be noted down from Figure 3.1 that around 30 percent of the household had access to piped water in 2005. In 2012 this changed to about 37 percent. In 2005, the largest proportion of sample households were using handpumps.

But that changed to piped water taps in the year 2012. Furthermore, Figure 3.1 indicates that there is variation in the sources of water used for drinking purposes in each year and across the years. It enables us to carry out the estimation analysis for access to piped water, time savings, and absenteeism from school.

3.4 Data

In this paper, we use the Indian Human Development Survey (IHDS) data for India. It is a panel data of two years, 2004-05 and 2011-12, collected by the University of Maryland and National Council of Applied Economic Research of India (NCAER). The first round of data was collected in 2004-05, by conducting surveys of a total of 41554 urban and rural households in a way that it was a nationally representative of India. The first round of the study was spread across 33 states of India, in 384 districts, 1503 villages, and 276 towns and cities. This round collected data by surveying 215,754 individuals in 41554 households. About fifteen percent of the households surveyed in 2004-05 were not covered in the second round of the survey in 2011-2012. Additionally, some new households were included in the second round of the study. It was done to compensate for the attrition in the data from the first round of the survey. In this paper, we use the individual and household modules after merging data for the households and individuals who were surveyed in both the rounds.

The household-level data is being used for estimation of the effect of access to piped water on time spent in the collection of water. To estimate the impact of access to piped water on absenteeism from school for children in classes 1 to 8 in the year 2004-05 (first round of the survey), we use the individual-level data for the two rounds of

IHDS survey. Some variables are available only at the household level. For the second part of the analysis, we merge these variables with the individual-level panel data.

3.5 Estimation Methodology

In this analysis, we use an instrumental variable for piped water access binary variable. This instrumental variable is defined as the non-self community ratio or leave-out ratio[47] LR_{ht}.

LR_{ht} is the proportion of households in a village that choose to get a piped water connection, excluding the house under consideration,

$$LR_{ht} = \frac{\mathrm{TP}-1}{\mathrm{Pop}} \text{ if } Pipe_{ht}=1 \text{ for household h in time t}$$

$$LR_{ht} = \frac{\mathrm{TP}}{\mathrm{Pop}} \text{ if } Pipe_{ht}=0 \text{ for household h in time t}$$

Where TP is the total number of households in a village with a piped water connection, and 'Pop' is the total number of households in a village. An instrumental variable LR_{ht} needs to satisfy the inclusion and exclusion criteria. We find from the data analysis that the non-self community ratio meets both inclusion and exclusion criteria for instrumental variables.

We find that there is a strong correlation between piped water binary variable and LR_{ht}. If a more substantial proportion of households in a village take piped water connections (higher value of LR_{ht}), then there is a higher chance for a particular house

[47] In the literature, both names are used. In the rest of the paper, we will be using non-self community ratio and represent it with LR_{ht}.

in that village to get a piped water connection[48]. In India, the selection into piped water is private and based on monthly paid user fees. Under the assumption that most of the variation in access to piped water lies at the household level and not at the community level, the non-self-community ratio (LR_{ht}) will affect the time spent in the collection of water and absenteeism from school only through household access to piped water.

For analyzing the effect of access to piped water on time spent in the collection of water, we use the household fixed effect model of estimation. This model allows us to control for the unobserved time-invariant household-specific variables that might affect the time spent in collecting water at the household level. The estimation equation used for the analysis of time savings is as follows.

$$Time_{ht} = \alpha_0 + \alpha_1 Pipe_{ht} + \mu_h + \lambda_t + \alpha_2 Z_{ht} + \varepsilon_{ht} \qquad (1)$$

In this equation, $Pipe_{ht}$ represents the household 'h' with access to piped water in time t[49], $Time_{ht}$ is the log of total time spent in the collection of water per day for household 'h' in time t, Z_{ht} represents household level control variables like the education level of the head of the household, the proportion of female family members, the ratio of children in the family, household assets, and the number of animals in the household 'h' in year t. In this equation λ_t represents year fixed effects and μ_h controls for unobserved time-invariant household characteristics for household h. In this setup, $'\alpha_1'$ describing the impact of access to piped water on the time spent in the collection of water is the central coefficient of interest.

[48] We also find that LR_{ht} satisfies the checks for weak instruments (value of F statistic is above 10).

[49] It is a dummy variable.

If there are time savings for households with access to in-house piped water taps, then it might influence the absenteeism from school for children in these households. It could happen directly via substitution of time spent in the collection of water by children for preparing and traveling to school from their house on time. The effect would be more prominent if they have to travel a distance to reach schools situated in other villages. Additionally, access to piped water can affect school absenteeism for children indirectly through substitution of time of women in these households from water collection towards child care and household work, thereby freeing school-going children in their homes from these activities. A third pathway could be that in houses with piped water taps, women can divert their time from water collecting activities towards activities that involve preparing breakfast, bathing, and dressing children so that they get ready for school on time.

Using a child fixed-effect model, the estimation equation for absenteeism from school for a child will be as follows.

$$A_{cht} = \alpha_0 + \alpha_1 Pipe_{ht} + \mu_c + \lambda_t + \alpha_2 X_{cht} + \alpha_3 Z_{ht} + \varepsilon_{cht} \quad (2)$$

In the above equation A_{cht} represents the number of days in an academic year child ‘c’ in household ‘h’ and year ‘t’ has been absent from school; $Pipe_{ht}$ represents the access to piped water in household h in year t; λ_t describes year fixed effects and μ_c controls for unobserved time-invariant child characteristics for child c in household h; X_{cht} are the child level control variables like the age of the child, gender of the child, marriage status, school type, school enrollment status and medium of instruction; Z_{ht} are household-level control variables like the proportion of children and female family members in the family, household asset index, education level of the female head of the household, highest education level in the family, and availability of water storage facilities.

The selection into piped water by a household seems to be a decision, based on parental preferences for improved water sources. There could be child-specific unobserved variables related to parental preferences for investment in child education, which could affect the decision to get connected to a piped water system. If that is the case, then the maximum likelihood estimation results in a child fixed effect model for school absenteeism could be biased and inconsistent. The possible endogeneity in the system is addressed by using the non-self-community ratio[50] as an instrumental variable for piped water dummy for analyzing the effect of piped water on absenteeism from school for children.

$$Pipe_{ht} = \gamma_0 + \gamma_1 LR_{ht} + \gamma_2 Z_{ht} + \epsilon_{ht} \tag{3}$$

In this equation, LR_{ht} is the non-self community ratio instrumental variable defined as the proportion of households in a village that choose to get a piped water connection, excluding the house under consideration. For this instrumental variable to work, there has to be a high correlation between variables LR_{ht} and $Pipe_{ht}$.

We use the non-self-community ratio of access to piped water as an instrumental variable for household access to piped water for both the regression equations[51]. For

[50] Non-self-community ratio instrumental variable for household access to piped water is constructed by undertaking the following calculations. Total number of households that have access to piped water in a village are calculated with deduction of 1 from it if the concerned household have access to piped water. After that a ratio of this number with the total number of households in a village provides us with the non-self-community ratio for a household.

[51] In the first regression analysis, the dependent variable i.e. log of time spent in collection of water is a continuous variable. For continuous variables, the estimation procedure using fixed effect model is carried out under the assumption of normal distribution of the error term. In the second regression analysis, days of absence from school for a child in an academic year is a count variable. In case of

the first reduced form equation, as it is a linear estimation process, we use 2 stage least squares (2sls) method of instrumental variable estimation. For this equation (1) and equation (3) are used together in the estimation process. First, we run the regression for equation (3). We then use the estimated values $Pipe_{ht}$ from equation (3) into the estimation of equation (1) and use a household fixed effect model to estimate equation (1).

For the second reduced form equation as school as the number of days a child is absent is a count variable, the error terms follow a negative binomial distribution. Due to the non-linear nature of the equation (2), the control function method of instrumental variation is being used. In this estimation process, equation (2) and equation (3) are used together. First, we estimate equation (3) and calculate the estimated value of residuals for this equation. We substitute these values of residuals back into the estimation equation (2). We then derive the estimated coefficient of interest for piped water binary variables under the child fixed effect estimation model. If we just use equation (1) or (2) for estimation, then we get ordinary least square (OLS) estimation results under household fixed effect and child fixed effect models. If we use equation (3) with equation (1) and equation 2 (separately), then we get the instrumental variation (IV) estimation results.

---

count variables, the error term does not follow normal distribution and another distributional assumption must be made regarding the error term to carry on the regression analysis. In this case, given the over-specification of the data, a negative binomial distribution for the error term is used.

3.6 Estimation Results

3.6.1 Summary Statistics and Regression Results: Access to Piped Water and Time Spent in Collection of Water

The summary statistics for the household level data with mean and standard deviations of the relevant variables used in the analysis are being provided in appendix Table 3.A.1. The average time spent on collecting water at the household level is 88 minutes per day. The household asset index measures the number of household assets a house possesses, ranging from 0 to 30. The average number of household assets (household asset index) in the sample is 13. The average household size is 5, with two children on average per family in the data set.

The Ordinary Least Square (OLS) and instrumental variable (IV) estimation results for the effect of access to piped water on time spent in the collection of water in a household (equation (1) in section 3.5) are provided in Table 3.1. It can be observed from columns 1 of Table 3.1 that the regression coefficient for the access to piped water is negative and statistically significant. It indicates that households with access to piped water spend, on average, significantly less (about 36 percent less) time in collecting water compared to households that do not have piped water in a household fixed-effect model[52]. Some of the other results in Table 3.1 column 1 have expected signs.

[52] For estimation under household fixed effect model the standard errors are clustered at the village level to control for possible correlation between error terms for households within a village.

Table 3.1: Effect of Access to Piped Water on Time Spent in the Collection of Water for a Household

	Model 1		Model 2	
Dep. variable = Log time	OLS	IV	OLS	IV
Piped Water (=1 if HH have access)	−0.358***	−0.394*	−0.307***	−0.429*
	(0.082)	(0.183)	(0.053)	(0.200)
Proportion of female family members	−0.153	−0.152	0.014	0.011
	(0.146)	(0.146)	(0.063)	(0.063)
Proportion of children in the family	0.087	0.086	0.436***	0.435***
	(0.078)	(0.078)	(0.037)	(0.037)
Education level of female head	0.004	0.004	−0.011***	−0.011***
	(0.006)	(0.006)	(0.003)	(0.003)
HH fixed effects	Yes	Yes	No	No
Village fixed effects	No	No	Yes	Yes
N	30848.00	20808.00	30848.00	30817.00

Note: IV for piped water dummy is the leave out ratio calculated as the proportion of households with piped water in the village excluding household under consideration.Village level clustered *se* are in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% level respectively.

In column 2 of Table 3.1, the two stages least square (2sls) estimation results of the regression of time spent in the collection of water on piped water are reported using the leave-out ratio instrumental variable. These results are obtained using a 2sls household fixed effect estimation model. We find that access to piped water reduces the total time spent in the collection of water per day at the household level by 39 percent.

We also conduct both the OLS and IV estimation for the sample households under the village fixed effect model. These results are being provided in the 3rd and 4th column of Table 3.1. We can observe that with a slight difference in the coefficient values, the effect of access to piped water on time spent in the collection of water is similar to those under the household fixed effect model. The similarity in results under household fixed effect and village fixed effect models indicates the robustness of our estimation results.

The results of the first stage regression of piped water on non-self community ratio instrument variables (IV) are provided in Appendix Table 3.A.2[53]. The results in column 1 of Table 3.A.2 are from the household fixed-effect model. The coefficient on the non-self community ratio is positive and statistically significant, indicating a strong correlation between the non-self community ratio IV and access to piped water binary variable for a household.[54] Column 2 of Table 3.A.2 contains the first stage regression results under the village fixed effect model. Again we observe that there is a strong correlation between access to piped water variable and non-self community ratio instrumental variable.

3.6.2 Summary Statistics and Regression Results: Access to Piped Water and Absenteeism from school

Individual-level Indian Human Development Survey (IHDS) panel data is used for analyzing the effect of access to piped water on absenteeism[55] from school for children. The focus group of this analysis is the children who were in primary school (in classes one to five) and secondary school (in classes six to eight) in year one (2005) of the panel data. As we are interested in studying the effect of access to piped water on school absenteeism for children in rural areas, we only use the IHDS data for rural India. There is a two-year panel with a gap of 7 years, in the second year (2012),

[53] The first stage results are calculated using a linear probability model for regression of piped water dummy on the leave out ratio instrumental variable.
[54] The F statistic values for the first stage regression is above 10 negating the possibility of weak instruments in this setting.
[55] Number of days a school going child is absent from school in an academic year (9 months).

these children are in different classes at the secondary or high school level[56]. The summary statistics for the variables of interests in the individual-level IHDS data are provided in Appendix Table 3.A.3. From Table 3.A.3, it can be observed that on average, a child misses three days of school in a month. Assuming a nine-month-long school year, this means that on average, there are 27 days of absenteeism from school for a child in rural India.

The number of days a child is absent from school is a count variable, and a negative bionomical regression model based on the maximum likelihood estimation method is used in this analysis. The results for the ordinary least square (OLS) and instrumental variables (IV) negative bionomical regression of the number of absentee days from school on access to piped water in a child fixed effect model [57] is provided in Table 3.2[58]. The regression results and the incidence rate ratios of the regression coefficients are presented in columns 2 and 4 of Table 3.2. The dependent variable A_{cht} is 'the number of absent days from school in a year.' It is a count variable. In a count model, the effect of a variable on the dependent variable is better understood by using incidence instead of regression coefficients.

[56] The class in which they are in panel year 2, 2012 depends on their repetition rate, dropout rates and the initial age when they started studying.
[57] Robust standard errors are calculated for both child and household fixed effect model under the negative binomial regression using maximum likelihood estimation method.
[58] First stage results for regression of leave ration IV on piped water access are provided in the appendix Table A5. In this regression linear probability model is used.

Table 3.2: Effect of Access to Piped Water on School Absenteeism for Children in Primary and Secondary School

	Primary		Secondary	
Dep. variable = Absent Days(Year)	OLS	IV	OLS	IV
Piped Water (=1 if HH have access)	0.779***	0.654***	0.840	0.767
	(0.031)	(0.033)	(0.090)	(0.110)
Age of the child	1.041***	1.041***	1.089**	1.085*
	(0.011)	(0.011)	(0.035)	(0.035)
Female child(=1 if girl)	0.952	0.954	1.099	1.120
	(0.038)	(0.038)	(0.135)	(0.139)
Education level of female head	0.999	0.999	1.004	1.003
	(0.005)	(0.005)	(0.013)	(0.014)
Distance to the school	1.001	1.001	1.002	1.003
	(0.003)	(0.003)	(0.006)	(0.006)
Completed years of education	0.935***	0.937***	1.011	1.013
	(0.010)	(0.010)	(0.033)	(0.034)
Proportion of female family members	1.533*	1.464	1.558	1.652
	(0.334)	(0.320)	(0.951)	(1.029)
Proportion of children in the family	0.807*	0.808*	0.825	0.826
	(0.075)	(0.075)	(0.247)	(0.252)
Water storage facility(=1 if yes)	0.755***	0.778***	0.626***	0.635**
	(0.030)	(0.030)	(0.086)	(0.089)
Child fixed effects	Yes	Yes	Yes	Yes
N	13158.00	13208.00	1648.00	1648.00

Note: Negative binomial regression analysis is conducted. The coefficients represent incidence rate ratios. IV for piped water dummy is the leave out ratio calculated as the proportion of households with piped water in the village excluding household under consideration.Robust *se* are in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% level respectively.

In Table 3.2, the incidence rate ratios[59] of different variables are provided. If the incidence rate ratios are less than one, then it indicates a negative effect of that variable on the dependent variable. If it is more than one, then it represents a positive effect. The first two columns in Table 3.2 show the estimation results for children who were in primary school (class 1 to class 5) in 2005, while columns 3 and 4 show the results for children in secondary school (class 6 to class 8) in 2005.

[59] Incidence rate ratios are calculated by taking exponential values of the MLE regression coefficients.

It can be observed from columns 1 and 2 of Table 3.2 that the effect of access to piped water on the ‘number of days a child is absent from school’ is negative and statistically significant under both OLS and IV regression analyses for children in a primary school in year 1. Access to piped water reduces the child absenteeism by 35 percentage points under the child fixed effect model for these set of children (column 2 of Table 3.2).

In both the models (OLS and IV), the age of the school-going child has a significant positive effect on the school absenteeism for children who were primary school in the year 2005. If we increase in the age of a child by one year, it results in a 40 percent increase in school absentee days for that child. It could be because the older a child is, the higher is the possibility of him/her in being engaged in water collection or other household chores and miss school. We find from columns 3 and 4 of Table 3.2 that children who were in secondary schools (in classes six to eight) in 2005 experiences no significant reduction in absentee days from school after getting access to piped water in their homes.

The other variable that has a significant impact on absenteeism from school for a child in a primary school in 2005 is the presence or absence of a water storage facility at the house. The household that has a storage facility for water tend to have significantly lower absentee days for a school going child in their house. It is because of two reasons. Firstly, if you have a water storage facility at home, then it means that you have to make fewer trips to the source once you get access to piped water. Households with no water storage facility tend to fetch water in multiple trips to the source of water. Also, for some of the activities like bathing, washing, and cleaning utensils, they often occur at the source if there is no water storage at home. If there is a crowd at the source in the morning, these could mean an additional loss of time for children who have to get ready for school. It may lead to delays and absenteeism from school. Secondly, the presence of a storage facility for water indicates the wealth status of households. Households that are well off in a village can employ workers to undertake household chores, including fetching water and reduce the burden of these activities for school-going children.

The completed years of education for a child also seem to have a small but significant effect on school absenteeism for children in a primary school in year 1. Higher are the average years of completed education for a child, more motivated he/she must be to attend the school, and lower would be the absenteeism from school for that child. The coefficient on the girl child dummy variable is negative (less than one) but not statistically significant. It indicates that on average, in households with access to piped water, a girl child is absent from school fewer days compared to a boy child. Still, the difference between the two is not statistically significant[60]. This result suggests that for children who were in a primary school in year 1 (2004-05), there are no significant gender differences among children in terms of absenteeism from school[61]. It is a surprising result as one would have expected the girl child to have more days of absenteeism from the school. One possible reason for this could be a rising awareness and change in attitudes towards girl child education in rural households in India.

---

[60] This result can be correlated to the results in time savings from access to piped water for girl child which was not statistically significant. Refer to the Table A5 in the Appendix for more detail.

[61] In order to understand the differential impact of access to piped water based on the gender of the child refer to the interaction effects presented in the next section of the paper.

As a robustness check, we conduct the regression analysis of school absenteeism for children on access to piped water using a household fixed effect model with the same set of variables and the same instrumental variable for piped water access. The OLS and IV results for this regression are presented in Appendix Table 3.A.5, with the first stage equation results being presented in Table 3.A.6. The incidence rate ratios are quite similar to the one achieved under child fixed effects for children who were in a primary school in the year 2005, indicating the robustness of our earlier results.

Table 3.3: Effect of Access to Piped Water on School Absenteeism for Households with Only Boy Children

	Primary		Secondary	
Dep. variable = Absent Days(Year)	OLS	IV	OLS	IV
Piped Water (=1 if HH have access)	0.756***	0.644***	0.739	0.521**
	(0.059)	(0.062)	(0.137)	(0.125)
Age of the child	1.005	1.004	1.180**	1.162**
	(0.022)	(0.022)	(0.067)	(0.065)
Education level of female head	0.997	0.997	1.034	1.031
	(0.011)	(0.011)	(0.024)	(0.024)
Distance to the school	1.003	1.002	1.013	1.016
	(0.007)	(0.007)	(0.010)	(0.010)
Completed years of education	0.921***	0.923***	0.968	0.975
	(0.022)	(0.022)	(0.055)	(0.056)
Proportion of female family members	1.404	1.382	0.842	0.801
	(0.560)	(0.547)	(0.716)	(0.685)
Proportion of children in the family	0.713	0.726	0.393	0.398
	(0.133)	(0.135)	(0.202)	(0.205)
Water storage facility(=1 if yes)	0.741**	0.760**	0.401***	0.417***
	(0.068)	(0.069)	(0.103)	(0.106)
Child fixed effects	Yes	Yes	Yes	Yes
N	3070.00	3076.00	556.00	556.00

Note: Negative binomial regression analysis is conducted. The coefficients represent incidence rate ratios. IV for piped water dummy is the leave out ratio calculated as the proportion of households with piped water in the village excluding household under consideration.Robust *se* are in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% level respectively.

For the children in a secondary school in the year 2005, the magnitude of the effect on school absenteeism is similar to that in the child fixed effect model. But there is a difference in the power of the result. Under household fixed effect estimation, access to piped water leads to a significant reduction in school absenteeism for children, under the child fixed effect model, this result was insignificant.

Different types of households may experience a differential effect of access to piped water on school absenteeism. We have divided all families into three categories, houses that have only boy children, only girls, and those that have a mix of the two. The results for the OLS and IV regression for these three types of households under the child fixed effect model are provided in Tables 3.3, 3.4, and 3.5. We find that in these three types of houses, the children in primary school experience about a 35 percent decline in their absenteeism from school (using the IV results from column 2 of Table 3.3, 3.4, and 3.5). These results are similar to those provided in Table 3.2, column 2, for the entire sample without differentiating the type of families based on the gender composition of children.

For the children in a secondary school in the year 2005, only the households with boy children show a significant decline in school absenteeism after access to piped water at home(column 4 of Table 3.3). In families with only girl children and with both girl and boy children (columns 4 of Table 3.4 and 3.5), access to piped water reduces school absenteeism. Still, the effect is not significant for children who were in a secondary school in the year 2005.

Table 3.4: Effect of Access to Piped Water on School Absenteeism for Households with Only Girl Children

	Primary		Secondary	
Dep. variable = Absent Days(Year)	OLS	IV	OLS	IV
Piped Water (=1 if HH have access)	0.824	0.678**	0.448*	0.636
	(0.090)	(0.095)	(0.140)	(0.252)
Age of the child	1.040	1.035	1.003	1.026
	(0.029)	(0.029)	(0.093)	(0.095)
Education level of female head	1.014	1.011	0.993	0.989
	(0.014)	(0.014)	(0.036)	(0.035)
Distance to the school	1.024	1.025	0.968	0.970
	(0.015)	(0.015)	(0.023)	(0.023)
Completed years of education	0.951	0.955	1.224*	1.252*
	(0.029)	(0.029)	(0.116)	(0.118)
Proportion of female family members	2.516	2.516	0.481	0.658
	(1.320)	(1.318)	(0.665)	(0.874)
Proportion of children in the family	0.674	0.699	0.533	0.552
	(0.169)	(0.175)	(0.412)	(0.419)
Water storage facility(=1 if yes)	0.789*	0.862	1.411	1.451
	(0.091)	(0.099)	(0.620)	(0.624)
Child fixed effects	Yes	Yes	Yes	Yes
N	1788.00	1800.00	244.00	246.00

Note: Negative binomial regression analysis is conducted. The coefficients represent incidence rate ratios. IV for piped water dummy is the leave out ratio calculated as the proportion of households with piped water in the village excluding household under consideration.Robust *se* are in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% level respectively.

In rural Indian society, there is a patriarchal difference in the role of a girl child and boy child in different household chores, including water collection, which is often age-dependent. At a young age, a boy child may also contribute to household chores if asked by their mothers as they are not overly conscious of their gender identity. But as they grow older, they become less inclined to engage in household chores, especially water collection, as it is considered women’s work. In contrast, as girls grow and become adults, they are more and more engaged in household chores, including water collection.

Table 3.5: Effect of Access to Piped Water on School Absenteeism for Households with Both Boy and Girl Children (Mix Households)

	Primary		Secondary	
Dep. variable = Absent Days(Year)	OLS	IV	OLS	IV
Piped Water (=1 if HH have access)	0.776***	0.669***	0.885	0.880
	(0.031)	(0.034)	(0.103)	(0.136)
Age of the child	1.042***	1.041***	1.049	1.051
	(0.011)	(0.011)	(0.037)	(0.037)
Education level of female head	0.993	0.994	0.997	0.997
	(0.005)	(0.005)	(0.015)	(0.015)
Distance to the school	1.003	1.003	0.998	0.998
	(0.004)	(0.003)	(0.007)	(0.007)
Completed years of education	0.945***	0.947***	1.063	1.070
	(0.010)	(0.010)	(0.037)	(0.037)
Proportion of female family members	1.283	1.250	0.640	0.611
	(0.304)	(0.295)	(0.485)	(0.466)
Proportion of children in the family	0.855	0.854	1.142	1.181
	(0.082)	(0.082)	(0.372)	(0.385)
Water storage facility(=1 if yes)	0.756***	0.776***	0.621**	0.624**
	(0.030)	(0.031)	(0.094)	(0.095)
Child fixed effects	Yes	Yes	Yes	Yes
N	12532.00	12592.00	1356.00	1358.00

Note: Negative binomial regression analysis is conducted. The coefficients represent incidence rate ratios. IV for piped water dummy is the leave out ratio calculated as the proportion of households with piped water in the village excluding household under consideration.Robust *se* are in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% level respectively.

Furthermore, even after getting access to piped water at home, due to the intermittent supply of piped water, there is still some time spent in water collection in a household. In the families with only boy children in a secondary school in year one (2005), as they become grown-up adults in year two (2012), they probably do not engage in any water collection tasks. In these households, tasks like water collection from outside sources due to intermittent supply are carried out by women members of the family. The boy children are more focused on education and are attending schools more regularly. For households with only girl children, the girls who were in secondary schools in year one are now adult women in the family in year 2 (2012). They are most probably still engaged in water collection from outside sources due to intermittent supply or engaged in other household chores. Thus, we do not observe any significant reduction in their school absenteeism rate as a result of access to piped water taps at home for households that have only girl children or a mix of boy and girl children in the secondary school in the year 2005.

In Table 3.2, we observed the gender of the child on their absenteeism from school. But it would be interesting to understand the differences in the impact of piped water on absenteeism from school by the gender of the child. We run the regression of school absenteeism on the interaction of access to piped water and gender of the child. The marginal effect of access to piped water on absenteeism from school for children of different gender (boy vs. girl child) calculated at different ages is depicted in Figure 3.2. We explain the results in terms of incidence rate ratios. For a boy child in a household with no piped water, there is about a 24 percent chance of being absent from the school if he is of age five. For a girl child of age five, the chances of being absent from school are slightly lower at about 22 percent. As the households get access to piped water, we find that across both the boy and girl children, the marginal effect of access to piped water on school absenteeism is the same, i.e., about 17 percent chance of being absent from school. As the age of the child increases, the chances of being absent from school declines for both a girl child and a boy child. This decline is present in both situations, when households have access to piped water, and when they do not have access to piped water.

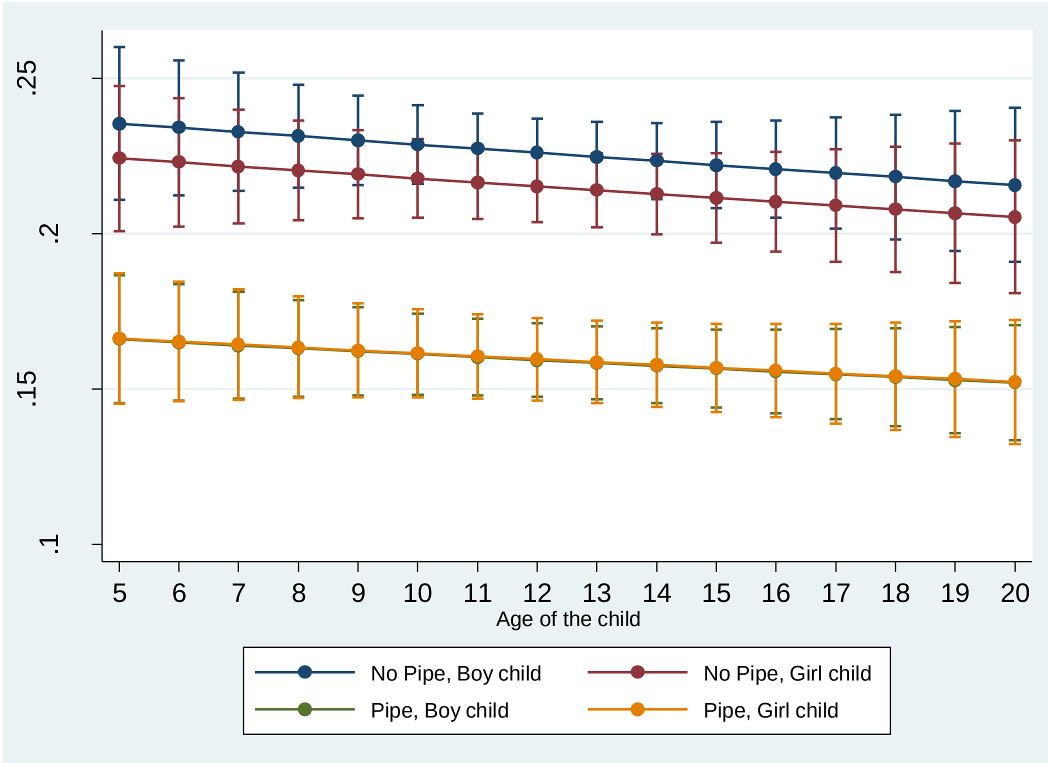

Figure 3.2: Variation in the Marginal Effect of Access to Piped Water on School Absenteeism Based on the Gender of the Child

The decline in the chances of being absent from school at any age is more substantial for a boy child compared to the girl child as a household gets access to piped water, but the difference is quite small. The marginal effect of access to piped water is thus larger for boy child than a girl child in terms of reducing absenteeism from school.

We observe that access to piped water leads to a reduction in school absenteeism, but we do not find any significant differential impact based on the gender of the child. It is contrary to expectations, as one would expect a higher decline in rates of absenteeism for girl children as a result of access to piped water at home. In fact, in absolute terms, the reduction in absenteeism rates is slightly higher for a boy child compared to a girl child. It indicates that there is a rising awareness about the education of girl children in rural India, especially when they are in primary school.

3.7 Conclusion

A nationally representative panel data of rural India is used in this paper to study the effect of access to piped water on time spent in water collection and absenteeism from school for children in primary and secondary schools in 2005. The variation in the non-self-community ratio of access to piped water is used for the identification of household access to piped water taps. We find that access to piped water leads to a significant reduction in time spent in the collection of water at the household level. This analysis is conducted by using 2sls instrumental variable estimation in a household fixed effect model. In the second part of this paper, the effect of access to piped water on school absenteeism for children in primary and secondary school is analyzed under a child fixed effect model. In this analysis, we find that there is a significant negative effect of access to in-house piped water taps on school absenteeism for children who were in a primary school in the year 2005 but not for children in secondary school. In other words, children who were in a primary school in year one (2004-05) and got access to piped water taps in their homes in the year 2012 have a lower rate of absenteeism from school compared to others who did not get piped water connection.

We find that, contrary to expectations, there are no significant differences in the absenteeism rates of girl and boy children in the sample households due to access to piped water taps. We differentiate between different types of families, depending on the gender mix of the children. We find that the results for children in a primary school in the year 2005 in houses with the only girls, only boys or mix households are similar to those achieved in the earlier estimation of the total child data with no differentiation in terms of gender mix of the family. But for those in the secondary school in the year 2005, the effect of access to piped water on school absenteeism is negative and significant only for households with boy children. For houses with only girl children or mixed households, the children who were in a secondary school in the year 2005 do not experience any significant decline in school absenteeism as a result of getting access to piped water. It indicates that for households with only boy children the access to piped water results in a reduction in school absenteeism both for children who were in primary and secondary schools in the year 2005.

In this paper, we have demonstrated that access to in-house piped water taps is associated with savings in the total time spent in the collection of water in a household and reduced absenteeism from school for children in a primary school in the year 2005. The improved attendance in school may have long term implications on dropout rates from school, child education, adult earnings for children in these households. This analysis helps in establishing the necessity of providing in-house piped water taps to rural households. Investment in piped water systems in villages of India will not only save time for family members engaged in water collection; it will also help in reducing school absenteeism for children, for both boys and girls.

APPENDIX FOR CHAPTER 3

Table 3.A.1: Summary Statistics for Household Data

Variable names	Mean	SD	Min	Max
Time in collecting water	87.8	101.8	0.0	1440.0
Piped Water (=1 if HH have access)	0.3	0.5	0.0	1.0
HH Asset Index	13.1	6.3	0.0	30.0
Family size	5.4	2.7	1.0	38.0
Number of children	1.6	1.6	0.0	18.0
Highest education degree	7.8	5.1	0.0	16.0
Education level of female head	4.9	5.1	0.0	16.0
No. of animals	2.7	5.4	0.0	301.0
Proportion of female family members	0.2	0.1	0.0	1.0
Proportion of children in the family	0.3	0.2	0.0	1.0
Number of Obs.	77852			

Table 3.A.2: First Stage Effect of Non-community Ratio IV on Piped Water Access for Household Data

	Model 1	Model 2
Dep. variable =Piped water	First stage	First stage
Leave-out-ratio IV	0.934***	0.792***
	(0.010)	(0.018)
Proportion of female family members	0.001	−0.030*
	(0.020)	(0.013)
Proportion of children in the family	−0.011	−0.008
	(0.011)	(0.007)
Education level of female head	−0.001	−0.001
	(0.001)	(0.000)
HH fixed effects	Yes	No
Village fixed effects	No	Yes
N	47287.00	47287.00

Note: IV for piped water dummy is the leave out ratio calculated as the proportion of households with piped water in the village excluding household under consideration.Village level clustered *se* are in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% level respectively.

Table 3.A.3 Summary Statistics for Individual data

Variable names	Mean	SD	Min	Max
Piped Water (=1 if HH have access)	0.3	0.5	0.0	1.0
Time spent by women(Minutes/day)	54.4	61.0	0.0	720.0
Total time(Minutes/day)	92.0	106.3	0.0	1440.0
Enrolled in school	0.9	0.3	0.0	1.0
Completed years of education	4.2	3.8	0.0	15.0
School type	3.1	1.8	1.0	13.0
Age of child	10.4	5.4	0.0	23.0
Girl child	0.5	0.5	0.0	1.0
Child marriage dummy	0.3	0.5	0.0	1.0
Medium of instruction	3.3	3.4	1.0	14.0
Education level of female head	5.1	5.1	0.0	16.0
Highest education degree	8.1	5.0	0.0	16.0
HH Asset Index	13.5	6.2	0.0	30.0
Proportion of female family members	0.2	0.1	0.0	1.0
Proportion of children in the family	0.3	0.2	0.0	1.0
Water stored in house	0.9	0.3	0.0	1.0
Available water sufficient	0.9	0.3	0.0	1.0
No. of animals	3.0	5.8	0.0	301.0
School hours/week	31.7	8.9	0.0	99.0
Homework hours/week	8.3	6.2	0.0	93.0
Tution hours/week	2.0	4.8	0.0	88.0
Free books	0.5	0.5	0.0	1.0
Free uniform	0.2	0.4	0.0	1.0
School fee govt dummy	0.3	0.4	0.0	1.0
Scholarship	104.9	622.0	0.0	45000.0
School fees/month	1831.8	7260.1	0.0	750000.0
Father dead	0.0	0.2	0.0	1.0
Father away from home	0.0	0.1	0.0	1.0
Mother dead	0.0	0.1	0.0	1.0
Mother away from home	0.0	0.1	0.0	1.0
Gender of hh head	0.9	0.3	0.0	1.0
Age of head of hh	53.7	12.7	12.0	99.0
Age of spouce	47.3	11.7	16.0	93.0
Sick from minor illness	0.4	0.5	0.0	1.0
No. of sick days(month)	3.0	5.0	0.0	30.0
Dropout dummy	0.1	0.2	0.0	1.0
No dropout kids in the family(=1 if yes)	0.8	0.4	0.0	1.0
Proportion of under 5 children	0.1	0.1	0.0	0.8
Non-school going young siblings(=1 if yes)	0.4	0.5	0.0	1.0
Absent days(in a month)	3.3	5.3	0.0	30.0
Absent Days(Year)	29.9	47.8	0.0	270.0
Number of Obs.	293998			

Table 3.A.4: First Stage Effect of Non-community Ratio IV on Piped Water Access in Individual Data (Child Fixed Effect Model)

	Primary	Secondary
Dep. variable = Piped Water	First stage	First stage
Leave-out-ratio IV	2.742***	2.391***
	(0.056)	(0.129)
Age of the child	1.006	1.008
	(0.004)	(0.011)
Female child(=1 if girl)	0.996	0.908
	(0.039)	(0.077)
Education level of female head	1.001	0.994
	(0.001)	(0.004)
Distance to the school	0.999	1.000
	(0.001)	(0.001)
Completed years of education	0.996	1.001
	(0.002)	(0.007)
Proportion of female family members	0.868**	1.046
	(0.043)	(0.153)
Proportion of children in the family	0.974	0.922
	(0.019)	(0.065)
Water storage facility(=1 if yes)	1.024***	1.050
	(0.007)	(0.032)
Child fixed effects	Yes	Yes
N	21160.00	4999.00

Note: IV for piped water dummy is the leave out ratio calculated as the proportion of households with piped water in the village excluding household under consideration.Robust *se* are in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% level respectively.

Table 3.A.5: Effect of Access to Piped Water on School Absenteeism for Children in Primary and Secondary School (Household Fixed Effect Model)

	Primary		Secondary	
Dep. variable = Absent Days(Year)	OLS	IV	OLS	IV
Piped Water (=1 if HH have access)	0.796***	0.705***	0.795*	0.722**
	(0.026)	(0.028)	(0.073)	(0.083)
Age of the child	1.039***	1.038***	1.054*	1.053*
	(0.007)	(0.007)	(0.024)	(0.024)
Female child(=1 if girl)	0.988	0.990	1.052	1.052
	(0.023)	(0.023)	(0.085)	(0.085)
Education level of female head	0.998	0.998	1.009	1.008
	(0.004)	(0.004)	(0.011)	(0.011)
Distance to the school	1.002	1.002	1.002	1.002
	(0.003)	(0.003)	(0.006)	(0.006)
Completed years of education	0.937***	0.939***	1.006	1.007
	(0.008)	(0.008)	(0.029)	(0.029)
Proportion of female family members	0.864	0.820	0.847	0.842
	(0.161)	(0.153)	(0.459)	(0.456)
Proportion of children in the family	0.845*	0.851*	0.657	0.668
	(0.068)	(0.068)	(0.168)	(0.171)
Water storage facility(=1 if yes)	0.737***	0.753***	0.633***	0.646***
	(0.024)	(0.025)	(0.076)	(0.078)
Household fixed effects	Yes	Yes	Yes	Yes
N	15972.00	16044.00	2106.00	2108.00

Note: Negative binomial regression analysis is conducted. The coefficients represent incidence rate ratios. IV for piped water dummy is the leave out ratio calculated as the proportion of households with piped water in the village excluding household under consideration.Robust *se* are in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% level respectively.

Table 3.A.6: First Stage Effect of Non-self Community Ratio IV on Piped Water Access in Individual Data (Household Fixed Effect Model)

	Primary	Secondary
Dep. variable = Piped Water	First stage	First stage
Leave-out-ratio IV	2.733***	2.352***
	(0.071)	(0.134)
Age of the child	1.002	1.003
	(0.001)	(0.004)
Female child(=1 if girl)	0.999	1.009
	(0.002)	(0.010)
Education level of female head	1.001	0.994
	(0.002)	(0.004)
Distance to the school	1.000	1.000
	(0.001)	(0.002)
Completed years of education	0.996*	1.001
	(0.002)	(0.006)
Proportion of female family members	0.860**	1.058
	(0.050)	(0.165)
Proportion of children in the family	0.972	0.933
	(0.022)	(0.067)
Water storage facility(=1 if yes)	1.027**	1.058
	(0.009)	(0.036)
Household fixed effects	Yes	Yes
N	21160.00	4999.00

Note: IV for piped water dummy is the leave out ratio calculated as the proportion of households with piped water in the village excluding household under consideration.Robust *se* are in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% level respectively.

REFERENCES FOR CHAPTER 3

Ahamad, S., Briscoe, J. & Chakraborty, M., 1981. How Bengali villagers choose sources of domestic water?. *Water Supply and Management*, 5, 165-181.

Ahamad, S., Arnold, B. F., & Colford, J. M. Jr., 2007. Treating water with chlorine at point-of-use to improve water quality and reduce child diarrhea in developing countries : A systematic review and meta-analysis. *American Journal of Tropical Medical Hygiene,* 76(2), 354–364.

Asthana, N. A., 1997. Where the water is free but the buckets are empty. *Open Economies Review*, 8(2), 137-149

Alves, D., & Belluzzo, W., 2004. Infant mortality and child health in Brazil. *Economics and Human Biology*, 2, 391–410.

Behrman, J. R., & Wolfe, B. L., 1987. How does mother's schooling affect family health, nutrtiton, medical care usage and household sanitation?. *Journal of Econometrics*, *36*, 185–204.

Boone, C., Glick, P., & Sahn, D. E., 2010. Household water supply choice and time allocated to water collection: Evidence from Madagascar. *Journal of Development Studies*, 47(12), 1826-1850.

Cairncross, S., & Cliff, J. L., 1987. Water use and health in Mueda, Mozambique. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 81, 51-54.

Cairncross, S., 2003. Editorial: Water supply and sanitation: some misconceptions. *Tropical Medicine and International Health*, 8(3), 193-195

Clasen, T., Schmidt, W.-P., Rabie, T., Roberts, I., & Cairncross, S. (2007). Interventions to improve water quality for preventing diarrhoea: systematic review and meta-analysis. *British Medical Journal*, *334*(7597), 782–785.

Cockburn, J., & Dostie, B., 2007. Child work and schooling: The role of household asset profiles and poverty in rural Ethiopia. *Journal of African Economics*, 16(4), 519–563.

Cook, J., Kimuyu, P., & Whittington, D., 2015. The costs of coping with poor water supply in rural Kenya. *Environment for Development, Discussion Paper Series.*

Crow, B. E. N., Swallow, B., & Asamba, I. (2012). Community organized household water increases not only rural incomes but also men' s work. *World Development*, *40*(3), 528–541.

Devoto, B. F., Duflo, E., Dupas, P., Parienté, W., & Pons, V., 2012. Happiness on tap: Piped water adoption in urban Morocco, *American Economic Journal: Economic Policy*, 4(4), 68–99.

Dungumaro, E. W., 2007. Socioeconomic differentials and availability of domestic water in South Africa. *Physics and Chemistry of the Earth*, 32, 1141-1147.

Fontana, M., & Natali, L., 2008. *Gendered Patterns of Time Use in Tanzania: Public Investment in Infrastructure Can Help*. Paper prepared for the IFPRI project on 'Evaluating the Long Term Impact of Gender-focused Policy Interventions.'

Galiani, S., et al., 2005. Water for Life: The impact of the privatization of water services on child mortality. *Journal of Political Economy*, 113(1).

Gamper-Rabindran, S., Khan, S., & Timmins, C., 2006. The impact of piped water provision on infant mortality in Brazil: A quantile panel data approach. *Journal of Development Economics*, 92, 188–200.

Hindman Persson, T. (2002). Household choice of drinking-water source in the Philippines. *Asian Economic Journal*, 16(4).

Ilahi, N. & Grimard, F., 2000. Public infrastructure and private costs: water supply and time allocation of woman in rural Pakistan. *Economic Development and Cultural Change*, 49(1), 45-75.

Ilahi, N., 2001. Children's work and schooling: Does gender matter? Evidence from the Peru LSMS panel data. *Policy Research Report on Gender, World Bank.*

Jalan, J., & Ravallion, M., 2003. Does piped water reduce diarrhea for children in rural India? *Journal of Econometrics*, 112, 153–173.

Jenkins, M. W., & Curtis, V., 2005. Achieving the "good life": Why some people want latrines in rural Benin. *Social Science & Medicine*, 61, 2446–2459.

Kamara, A., & Sally, H., 2003. Water for food, livelihoods and nature: Simulations for policy dialogue in South Africa. *Physics and Chemistry of the Earth*, 28, 1085-1094 .

Koolwal, G., & Walle, D. van de., 2013. Access to water, women's work, and child outcomes. *Economic Development and Cultural Change*, 61(2), 369–405.

Kremer, M., & Peterson-zwane, A., 2010. Providing safe water : Evidence from randomized evaluations. *Harward Kenndy School: Discusson Paper Series* 10-23.

Lamichhane, D. K., & Mangyo, E., 2011. Water accessibility and child health: Use of the leave-out strategy of instruments. *Journal of Health Economics*, 30, 1000–1010.

Lee, L.-F., Rosenzweig, M. R., & Pitt, M. M., 1997. The effects of improved nutrition, sanitation, and water quality on child health in high-mortality populations. *Journal of Econometrics*, 77, 209–235.

Madanat, S., & Humplick, F., 1993. A model of household choice of water supply systems in developing countries. *Water Resources Research*, 29(5), 1353–1358.

Mangyo, E., 2008. The effect of water accessibility on child health in China. *Journal of Health Economics*, 27, 1343–1356.

Mark Blackden, C., & Wodon, Q., 2006. Gender, time use, and poverty in Sub-Saharan Africa. *World Bank Working Paper*, 73.

Masuda, Y. J. & Cook, J., 2013. Does improved water access increase child schooling? A quasi-experimental approach from rural Ethiopia. Working paper.

Mccauley, A. P., Lynch, M., Pounds, M. B., & West, S., 1990. Changing water-use patterns in a water poor area: Lessons for a Trachoma intervention project. *Sm. Sci. Med*, *31*(II), 1233–1238.

Mccauley, A. P., West, S., & Lynch, M., 1992. Household decisions among the Gogo people of Tanzania: determining the roles of men, women and the community in implementing a Trachoma prevention program. *Sm. Sci. Med*, 34(7), 817–824.

Meeks, R. C., 2017. Water works: The economic impact of water infrastructure. *Journal of Human Resources*, online print Feb 6, 2017.

Merrick, T. W., 1985. The effect of piped water on early childhood mortality in urban Brazil 1970 to1976. *Demography*, 22(1).

Mu, X., & Briscoe, J., 1990. Modeling village water demand behavior' A discrete choice approach. *Water Resources Research*, 26(4), 521–529.

Nankhuni, F. J., & Findeis, J. L., 2004. Natural resource-collection work and children's schooling in Malawi. *Agricultural Economics*, 31, 123-134.

NRDWP, National Rural Drinking Water Programme. Movement Towards Ensuring People's Drinking Water Security in Rural India: Framework for Implementation (2013), Ministry of Drinking Water and Sanitation, Government of India. http://mdws.gov.in/sites/default/files/NRDWP_Guidelines_2013.pdf

Null, C., Meeks, R., Miguel, E., & Zwane, A. P., 2012. Safe drinking water. Who is willing to pay the price? *International Initiative for Impact Evaluation, Systematic*.

Root, G. P. M., 2001. Childhood diarrhea in rural Zimbabwe. *Journal of Health, Population and Nutrition*, 1919(22), 73–8273.

Rosen, S. et al., 1999. Household water resources and rural productivity in Sub-Saharan Africa: A Review of the Evidence. *Harvard Institute for International Development.*

Sekhri, S., 2013. Female Literacy and Access to Drinking Water in Rural India Working paper draft: March 2013.

Nauges, C., & Strand, J., 2017. Water Hauling and Girls ’ School Attendance Some New Evidence From Ghana. *Environment and Resource Economics,* 66, 65-68.

Timmins, C. et al., 2008. The impact of piped water provision on infant mortality in Brazil, *NBER Working Paper Series*, 14365.

Whittington, D., Mu, X., & Roche, R., 1990. Calculating the value of time spent collecting water: Some estimates for Uunda, Kenya. *World Development Report,* 18(2), 269–280.

WHO, & UNICEF., 2014. *Progress on Sanitation and Drinking-Water 2014 Update. Monitoring Program for Water Supply and Sanitation,* 1–78.

Zhang, J., 2011. The impact of water quality on health: Evidence from the drinking water infrastructure program in rural China. *Journal of Health Economics*, *31*, 122–134.

Zhang, J., & Colin, L., 2016. The long-run effects of treated water on education : The rural drinking water program in China. *Journal of Development Economics*, 122, 1–15.