EVALUATING THE BENEFITS OF A WATER QUALITY INTERVENTION IN RURAL INDIA: A LONGITUDINAL STUDY

By

Juliana Chen
Dr. Subhrendu K. Pattanayak, Advisor
April 2010

Masters project submitted in partial fulfillment of the requirements for the
Master of Environmental Management degree in the
Nicholas School of the Environment of
Duke University
2010

MP Advisor's signature

ABSTRACT

Child diarrhea is one of the primary causes of infant death in the world. It affects poor populations in developing countries who do not have access to clean water or sanitation. Due to the limited resources that can be allocated to its solution in developing countries, new methods try to be sustainable and scalable in cost-effective ways. One such intervention funded by the Acumen Fund is a market-based, community-level water, water quality intervention in Andhra Pradesh, India. This intervention utilizes ultraviolet disinfection to provide safe drinking water at an affordable price of one Rupee for 12 liters. The objective of this longitudinal study is to quantify the economic benefits of this intervention resulting from the reductions in coping costs of diarrhea. In order do this, household averting behaviors were identified and their costs calculated using revealed preferences, specifically the averting cost and cost of illness method. This study is part of larger impact evaluation conducted by RTI International that uses a quasi-experimental research design. The data utilized in this study was gathered from 25 treatment and 25 control villages, matched using propensity score matching, over the course of a year through bi-weekly household surveys. The resulting panel data consists of 100 households observed in 26 rounds. Regression analysis using fixed effects to account for household characteristics that are time-invariant was employed to determine the effects of using clean water from the treatment plant are on averting costs. This study finds that averting costs decrease as the percentage of the household's water that comes from the clean water source increases. For the average household purchasing clean water, monthly savings due to reductions in averting costs is about 580 Rupees, or 32% of their monthly income. Thus, providing clean water at an affordable price can help reduce household coping costs of diarrhea.

TABLE OF CONTENTS

List of Graphs, Plots, and Tables

Graphs

Plots

Tables

1. Introduction

The objective of this thesis is to quantify the economic benefits of a community-level water, sanitation, and hygiene intervention that provides affordable safe drinking water in rural India using longitudinal data. The economic benefits considered here are only those due to reductions in disease and coping costs of diarrhea and not other health benefits, such as reduction in child mortality. This thesis will primarily focus on valuing the benefits of clean water by quantifying the reduction in household's averting costs and costs of illness.

Due to the gravity of the diarrhea problem, one of the targets of the Millennium Development Goals is to reduce the population of people without access to clean drinking water and basic sanitation to half the previous levels (WHO, 2006). However, due to limited resources, decisions have to be made regarding the type of health intervention that should be used in order to accomplish this target in a sustainable and cost-effective way. Some of these interventions can have considerable socio-economic impacts in addition to health improvements. However, these welfare benefits are not always considered, and "unfortunately, the evidence base for quantitative benefits is weak or non-existent" (Pattanayak et al, forthcoming). Benefit estimates, such as the ones measured in this study, "are critical for allocating investments between different health interventions" (Pattanayak et al., forthcoming).

The intervention I will be examining is a market-based community owned and maintained water supply intervention in Andhra Pradesh, India (Poulos et al., 2007). It relies on increasing access to water by providing affordable safe drinking water filtered in a water treatment plant that is maintained by the community. This intervention provides clean water from the source at the community level and not from point of use at the individual households. It is also market-based because clean water is sold to individuals rather than given for free.

2. BACKGROUND

2.1 Child Diarrhea

Child diarrhea is one of the primary causes of infant death in the world resulting in approximately 1.8 million child deaths each year (UNDP, 2006). It affects poor populations in developing countries who do not have access to clean water or sanitation. Half of these infant deaths occur in India. Unfortunately, 1.1 billion people in the developing world do not have clean water and 2.6 billion people do not have access to basic sanitation (UNDP, 2006). Additionally, the burden of diarrhea falls "disproportionally on the poor, women and children" (Pattanayak et al, forthcoming). Diarrhea occurs due to both poor environmental conditions such as high levels of fecal contamination in drinking water sources and poor sanitary practices such as open defecation. Child health research has focused on either the biomedical or the socioeconomic side of infant diarrhea, but recent studies have developed a new approach that incorporates both into a coherent framework, such as The Cebu Study Team (1991).

Attempts to reduce diarrhea occurrence in children have resulted in three major categories of health interventions aimed at diarrhea prevention, namely improving the water supply, increasing sanitation facilities, promoting better hygiene practices, or a combination of all three. Studies have then focused on evaluating the effectiveness of these interventions in terms of reducing diarrhea morbidity or mortality. A recent meta-analysis reviews several of these studies and finds that these interventions significantly reduce the risk of diarrhea and consistently do so with the same effectiveness (Fewtrell et al., 2005). Also, the meta-analysis finds that water quality interventions are more effective than previously thought, and multiple interventions are not more effective than single focus interventions (Fewtrell et al., 2005). Another recent review confirms that household level behavioral interventions, such as hand washing and point-of-use water treatment can be effective in reducing diarrhea. In contrast, community-level interventions that increase water supply do not substantially reduce diarrhea, nor are they effectively maintained (Zwane and Keremer, 2007).

Besides evaluating the effectiveness of clean water supply and sanitation interventions, it is also of interest to determine what the welfare effects of the intervention are. Several studies have looked at the welfare effects of environmental health interventions, including Pattananyak et al. (forthcoming) and Pattanayak and Pfaff (2009). One such study was conducted on an environmental health intervention in Maharashtra, India that was community-demand-driven and provided improvements to both the water supply and sanitation services (Pattanyak et al., 2009). Using a revealed preference valuation method of environmental health improvements, they were able to determine the behavioral changes caused by the intervention implied a benefit of as much as 5% of monthly income (Pattanayak et al., forthcoming). This study seeks to expand on what has been previously done by using revealed preference methods to calculate household averting costs and costs of illness, and analyzing the effects of this intervention on those costs using a fixed effects strategy on longitudinal data. This model is discussed in more detail in Section 3.2.

2.2 Intervention

2.2.1 Description

This intervention consisted of establishing a Community Water Treatment Project centered on a Community Water System that purifies water using filtration and ultra-violet technology up to World Health Organization and US Environmental Protection Agency drinking water standards. The project is supported by the Water and Sanitation Funding Facility, established by the Acumen Fund with the support of the Gates Foundation. The system draws water from a community source, purifies it, and sells it to households for an affordable price of one Rupee for 12 liters. The fees collected are used to maintain and operate the system. These water treatment plants were located in villages across Andhra Pradesh, India. For more details on the intervention, see Pattanayak & Poulos (2009).

2.2.2 Survey Data

The data necessary for this paper was collected by RTI international from panel surveys conducted to evaluate the impact of this intervention using a quasi-experimental research design (Pattanayak & Poulos, 2009). Small tracking surveys were conducted bi-weekly on 100 randomly selected households with children under five from 10 villages during the course of a year, beginning in March 2008 and continuing until February 2009 (Pattanayak & Poulos, 2009). Treatment villages were communities that had an operational Community Water System throughout the study, while control villages were those without a Community Water System. Treatment communities were matched to similar control communities using propensity score matching, a process in which treatment villages were matched to non-treatment villages based on a vector of community characteristics (Poulos et al., 2007). The reason for this was to try to minimize the differences in pre-treatment characteristics between treatment and control villages.

The survey collected information on household and individual behaviors and illnesses. The survey contains questions on household characteristics, such as the number of household members and the number of children under five. It also contains questions on household water, sanitation, and hygiene behaviors, such as whether or not they purchased and used the clean water, hand washing, safe water handling, time spent collecting water, and water storage behaviors. The survey also collected information on individual characteristics, such as age and gender. Because of the nature of the study, there are several issues that must be considered in order to correctly estimate the intervention benefits. These issues are discussed in more detail in Section 3.3.

2.2.3 Previous Results

A previous report of this project studied the health and socioeconomic impacts of this intervention using a difference in differences analysis. This rigorous program evaluation studied focused on the outputs, outcomes, and impacts of the intervention (Pattanayak & Poulos, 2009). The results show that the Community Water System (CWS) project has not reduced child diarrhea over the short timeframe and users have higher diarrhea rates. However, the difference in difference analysis shows that CWS program causes an average of 133 Rupees net savings in coping costs associated with contaminated water. These coping costs include time and money spent on water collection, water treatment, and water storage (Pattanayak & Poulos, 2009).

Another previous study on the smaller tracking survey panel data determined the prevailing trends, and conducted bivariate tests of differences and multivariate regressions of key indicators including diarrhea in children under five and water related behaviors (Poulos et al., 2009). The results from the graphs and bivariate tests show that there are no differences in diarrhea prevalence between households that purchase clean water (users) and those that do not (non-users). The results from the multivariate regressions, however, suggest that diarrhea prevalence is lower in households that purchase clean water after accounting for household characteristics including health status, health-related behaviors, and seasonal influence (Poulos et al., 2009).

3. METHODS

In order to quantify the benefits and the effectiveness of this intervention, I will proceed in the following manner. First, I will determine which averting behaviors could potentially change due to the intervention. Then, I will determine what these behaviors and resulting behavioral changes, if any, tell us about how much costs households incur to obtain clean water, and how these costs may change due to the intervention. Furthermore, I will test to see if there were any significant changes in behaviors and costs between control and treatment groups. Finally, I will determine to what these changes in behaviors, and therefore costs, are attributed to using a multivariate regression analysis. As such, I follow the methodology developed in Pattanayak et al. (2005).

3.1 Valuation

In order to measure the economic benefits of this water, sanitation, and hygiene (WSH) intervention, non-market valuation methods of revealed preferences must be used. Revealed preferences studies use household choices to measure economic impacts (Champ et al., 2003). Specifically, the defensive expenditures (or averting costs) and damage costs (or cost of illness) methods allow us to measure the cost savings due to water and sanitation improvements (Pattanayak et al., forthcoming). Thus, it is possible to determine the benefits gained by measuring how much households spend on averting, mitigating, and illness costs, using the defensive behavior method combined with the damage cost method. The framework used to develop a household production function that incorporates behavior, environment, and health in India is based on a paper by Pattanayak & Pfaff (2009). According to this framework, households have the ability to produce levels of outcomes based on inputs, including environmental quality (i.e. levels of pollution). Thus, households will choose to engage in certain behaviors and incur their costs in order to reduce their exposure to pollution and increase their health. In addition, to calculate the cost of certain behaviors, estimates and assumptions on wages, prices, and time savings were used from other WSH intervention studies in India. These include: a study in Orissa, India (Pattanayak et al., 2009 and Whittington et al., 2009) and a study in Maharashtra, India (Pattanayak et al. forthcoming). Averting costs include all the costs associated with the averting behaviors described previously. Costs of illness include the out of pocket medical expenses, such as clinic or doctor fees and medicines, and the lost income due to lost workdays because of diarrhea episodes. For details on the variables and how they were constructed, see Appendix A.

3.1.1 Averting Behaviors

The first step to valuing the benefits of this intervention consists of finding the behaviors that households engage in order to avoid diarrhea. These behaviors and their associated costs tell us about household's optimization choices based on their preferences and constraints. In this survey, households were asked about several behaviors. From these, the ones that concern averting choices are: time spent collecting water (including time spent walking to and from a water source and time waiting for water), purchasing water (including water from the CWS or other), water treatment (including the use of filters and boiling), and water storage (including purchasing the storage and transportation vessels). These variables were constructed using the responses to the certain key survey questions, for details see Appendix B.

3.1.2 Averting Costs

The first averting cost is the time costs of collecting water. This value is calculated using the time spent collecting water and an estimate of the daily wage. For this, I assumed the daily wage of unskilled labor is 60 Rupees and the value of time savings is 30% of the market wage. Both of these values are taken from Whittington et al. (2009).

The second averting cost is the cost of purchasing water. Households chose to purchase water either from the purification plant from this intervention or other sources. Theoretically, households would be willing to pay for water instead of obtaining it for free because they believe the purchased water is safer or cleaner than the free water. However, there might be other aesthetic reasons for purchasing water. For this calculation, I obtained whether or not a household had purchased water and how much money it spent from the survey questions.

The third averting cost is the cost of treating water. This value is calculated for households that treated their own water, by either filtering or boiling. For this, I assumed households either filtered or boiled only one fourth of all the water collected, and I excluded the costs of other treatments, such as chemicals. The cost of filtering was calculated using a previous estimate of 2 Rupees per month and the cost of boiling was calculating using a previous estimate of 27 Rupees for 5 liters of water. These values were obtained from Pattanayak et al. (forthcoming). Both of these treatment costs were added into a total treatment cost.

The fourth averting cost is the cost of storing water. For this calculation, I obtained whether or not households had purchased a new transportation and storage container and what were the costs of these from the survey questions. Therefore, this value is a onetime fixed cost of a purchase. Since the survey only looks at data from one year, I neglected any decrease in the value and quality (depreciation) of the container and excluded the costs of cleaning the container.

3.1.3 Costs of Illness

The first cost of illness is the out-of-pocket medical expenses. For this calculation, I obtained whether or not a household had paid for any treatment for diarrhea and how much money they spent on these treatments from the survey questions.

The second cost of illness is the time cost for being sick or taking care of the sick. For this calculation, I obtained how many days of work were missed due to personal diarrhea or by taking care of others who had diarrhea. I only included the days lost of work, not school. Although, days missed from school is an important factor that contributes to the negative externalities of diarrhea and reinforce the cycle between poverty, lack of education, and health problems such as diarrhea and malnutrition, I excluded it from this calculation due to the difficulties in determining the opportunity cost of children. Once again, I assumed the daily wage of unskilled labor is 60 Rupees and the value of time savings is 30% of the market wage. Both of these values are taken from Whittington et al. (2009).

3.2 Cost Regressions

It is important to understand what accounts for the averting costs and costs of illness. Certain household characteristics (e.g. size) as well as exogenous factors (e.g. weather) influence the household's decision to incur defensive and damage costs. In order to observe the magnitude of these effects, if any, we can use the following ordinary least squares linear regression model with household fixed effects:

$$Y_{i,t} = a_i + b(X_{i,t}) + c(CWS\%_{i,t}) + d(Non\text{-}user_{i,t}) + g(Monsoon) + h(Diarrhea_{i,t-1}) + e_{i,t}$$

$$\text{where } a_i = a_0 + a_1HH_1 + a_2HH_2 + \ldots + a_nHH_n$$

Where Y is the outcome of interest (costs) for each observation (household, *i*) and time period (round, *t*); X are covariates (household characteristics, such as size and number of children under five); CWS% is the percentage of the household's water that comes from the clean water source; Non-user is a dummy indicator for whether it is in a treatment village but does not purchase clean water; Monsoon is a dummy for whether these costs are incurred in the rainy or dry season; Diarrhea is a dummy for whether the household had diarrhea in the previous period (*t*-1), and e is the error term. The time-invariant intercept (a_i) is a combination of the intercept (a_0) and the coefficients of the dummy variables for each household (a_n). Adding household fixed effects is made possible because of the large number of degrees of freedom in the panel data, which consists of multiple observations (in this case households) in multiple time periods (in this case survey rounds). Fixed effects can help control for endogeneity and omitted variable bias (Stock & Watson, 2007). Thanks to the multiple period observations for each household, we can group the unobservable (or confounding) household characteristics that are time-invariant for each household, creating parallel linear regressions (with the same slope but different intercept).

The variables used in this regression model are expected to have an effect on household costs. The number of household members (including children under five) implies that households consume more water, and therefore purchase or treat more water. We expect that averting costs should increase as household size and number of children under five increases (both are expected to have positive coefficients), especially for the costs of purchasing and storing water. Since diarrhea is more frequent in children, we expect costs of illness to be higher in households that have more children under five (a positive coefficient), especially for costs of lost workdays by parents staying home and taking care of the young. The CWS% variable should have a negative effect (and coefficient) on averting costs. This is because we expect households to engage in fewer averting behaviors, such as boiling water, the more clean water they purchase. The Non-user dummy variable was included in order to separate the intervention effects on non-users and the control villages. It is possible that non-users have some effects from being in the treatment village, even if they do not purchase the clean water. These spillover effects could potentially be beneficial. The monsoon dummy was added in order to control for seasonal effects. Diarrhea is most frequent during the rainy season, so we expect both averting costs and costs of illness to increase during the monsoon. However, the dry season also makes water more scarce, which may increase the time cost of collecting and storing water. We also expect to see averting costs and costs of illness increase if the household experienced diarrhea in the previous period (lag diarrhea dummy). This is because individuals might change their behaviors if they experienced diarrhea recently and consequently engage in more averting behaviors. Also, frequent diarrhea could cause more serious long term illnesses that require more expenditure.

3.3 Possible Sources of Error

There are several potential threats to validity in this quasi-experimental design. The main concern is internal validity due to self-selection bias and unobservables (confounds). This is especially true given the complicated nature of health and behavioral outcomes, which rely on many observable characteristics and individual preferences that tend to be unobservable. The problem of self-selection bias arises from the fact that users and nonusers chose to be either. This decision could be attributed to observable characteristics, such as income or education. However, it could also be due to unobservable characteristics, such as increased awareness of diarrhea or higher willingness to pay for clean water. This intervention might be creating awareness for cleaner water prompting more averting behavior and therefore increasing the costs. This is especially true of time costs collecting water. Users may be more willing to spend time, and thus money, collecting water they believe is clean and will reduce the risk of diarrhea (alternatively, they may value clean water for aesthetic reasons). It is possible that these results reflect the lack of access to water sources. Therefore, households that travel far to obtain water might do so out of necessity and not because they value clean water that much more. All of the previous concerns are household level characteristics that may not very considerably over time. If this is the case, then we can address these issues by controlling for observable (e.g. income and education) and unobservable (e.g. aesthetic preferences) household characteristics by using household fixed effects (Stock & Watson, 2007).

Another important concern is external validity. The results of this intervention could be unique to the geographical and cultural setting of India and not applicable to other developing countries. This is especially true when considering the importance of continual exposure to diarrhea. In areas where diarrhea is more prevalent, this same fact might have led households to adopt more averting behaviors than others households in villages with less frequent diarrhea.

Construct validity is a problem because of the measurement of the outcome variables. These would be constructed from responses to household surveys. This kind of self-reported data could be unreliable because people might lie or forget previous incidents and behaviors. However, this does not appear to be case because there are few behavioral differences between groups that are not explainable by the variables in the model. Another important factor when considering the construct validity of these results is the assumptions made when calculating the costs. For example, non-users clean their water storage and transport containers more often than users. If the costs of cleaning containers would have been included, then this would increase their averting costs beyond that of the user group, that use containers provided by the CWS.

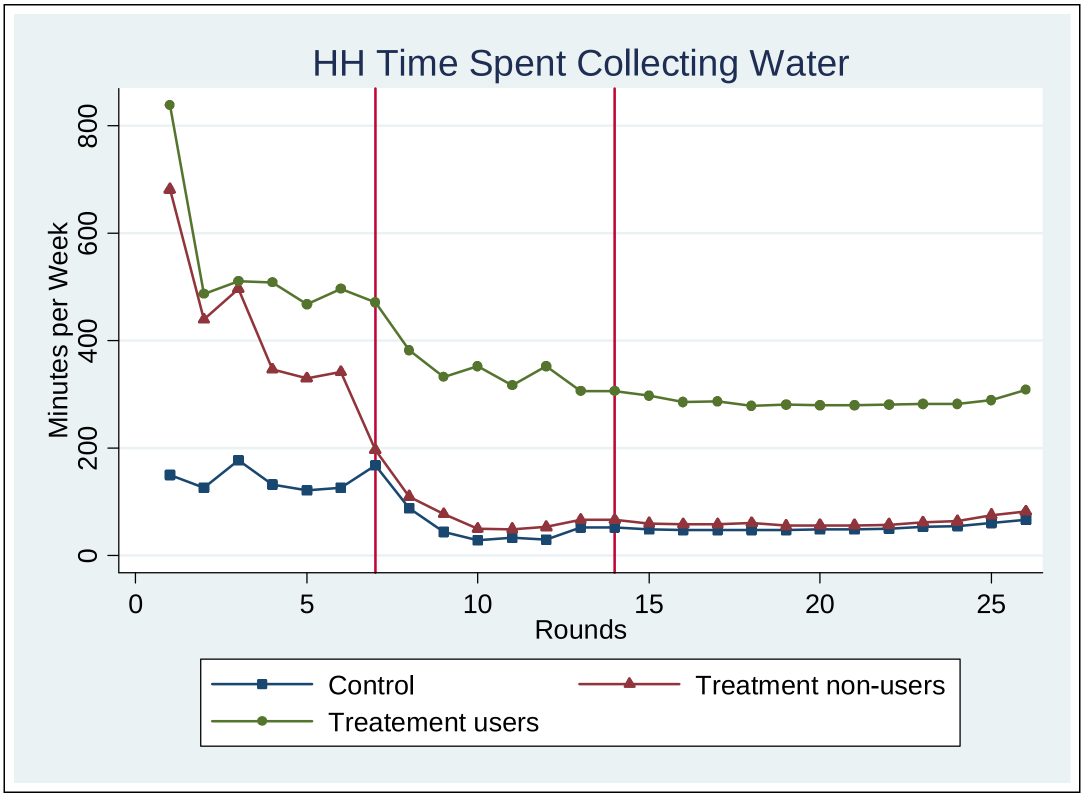
4. Results

4.1 Behavior Trends

The four averting behaviors changes are presented in the following plots that illustrate the trends of these behaviors over the course of the study. The vertical lines indicate the approximate start and end of the monsoon.

4.1.1 Time Spent Collecting Water

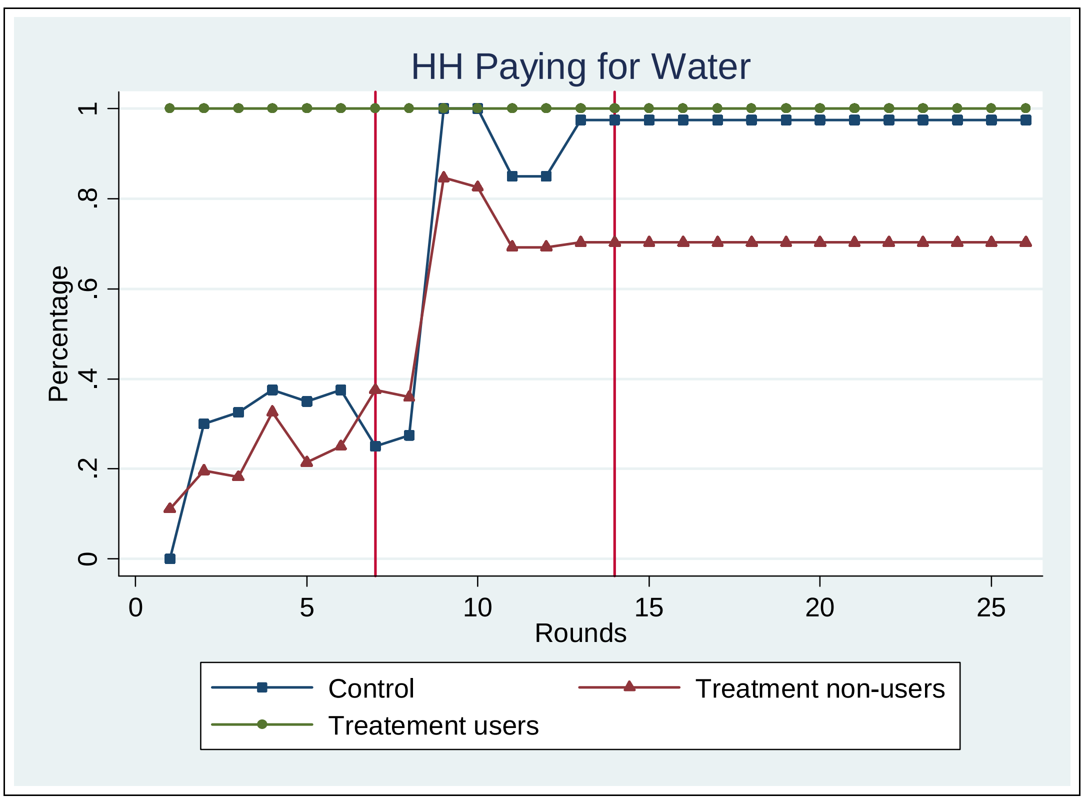
Plot 1. Time (minutes per week) Spent Collecting Water by Groups Trend



The plot above illustrates that treatment users spend significantly more time collecting water than both the non-users and the control group. This can be due to several reasons including: (1) the fact that users tend to have more than one water source and therefore have to travel to more places to collect water, and (2) the Community Water System from which they obtain their clean water might be located far away from their household. Non-users spend slightly more time collecting water than households in control villages.

4.1.2 Purchasing Water

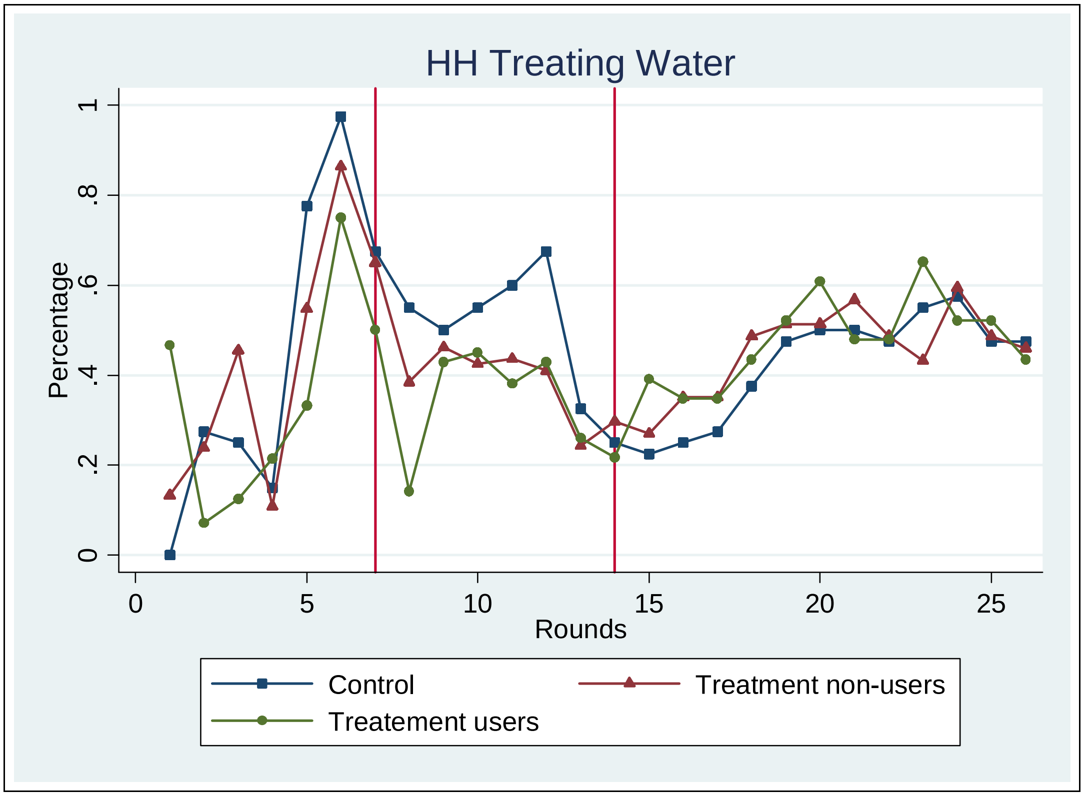
Plot 2. Percentage of Households Purchasing Water by Groups Trend



The plot above illustrates that users consistently purchase water, as expected. It also shows that households in control villages also purchase water but not from the CWS plant, and due so more than non-users. It is evident that non-users and control groups also purchase potentially clean water from other sources, such as bottled water.

4.1.3 Water Treatment

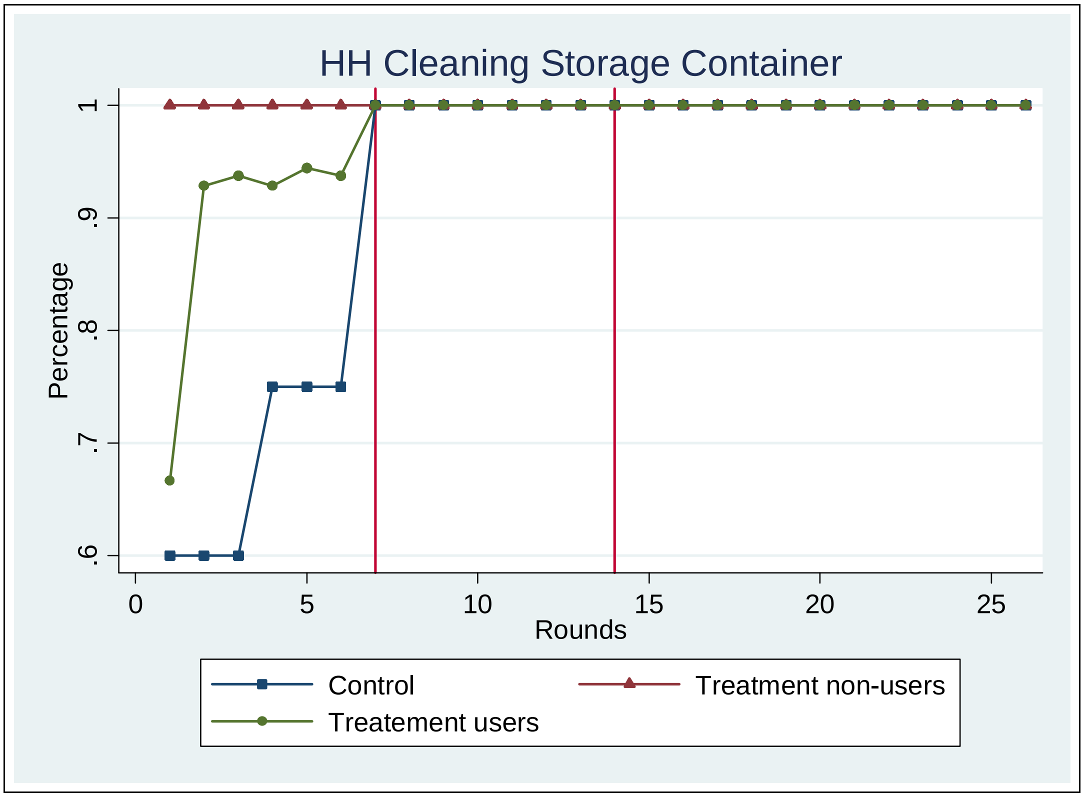
Plot 3. Percentage of Households Treating Water by Groups Trend



The plot above shows that all three groups treat their water and there do not seem to be any apparent differences between the groups. However, the plot also shows that there might be some seasonal effects on water treatment behaviors. As seen in the above graph, water treatment behaviors increase at the onset of the monsoon. This could be due to household's expectations of increasing diarrhea in the rainy season.

4.1.4 Cleaning Water Storage Containers

Plot 4. Percentage of Households Cleaning Water Storage Container by Groups Trend



The plot above illustrates that non-users clean their water storage containers consistently and more often than users and households in control villages. Although, for the most part, all groups clean their storage container frequently.

4.2 Calculating Costs

The averting costs and costs of illness are summarized in the following tables and graphs by groups and season. Then, averting costs and costs of illness between the three groups are compared and their differences in means tested.

4.2.1 Averting Costs

Time Cost of Collecting Water: The average monthly cost of collecting water for all groups in both seasons is 25 Rupees. The monthly average cost in the dry season for all groups is 28 Rupees, while the monthly average cost in the rainy season for all groups is 20 Rupees. The monthly averages for each group across seasons are: 11 Rs for control, 25 Rs for non-users, and 53 Rs for users.

Expenditures on Water: The average monthly cost of purchasing water for all groups in both seasons is 51 Rupees. The monthly average cost in the dry season for all groups is 51 Rupees, while the monthly average cost in the rainy season for all groups is 50 Rupees. The monthly averages for each group across seasons are: 41 Rs for control, 30 Rs for non-users, and 108 Rs for users.

Water Treatment Costs: The average monthly cost of treating water for all groups in both seasons is 128 Rupees. The monthly average cost in the dry season for all groups is 147 Rupees, while the monthly average cost in the rainy season for all groups is 86 Rupees. The monthly averages for each group across seasons are: 138 Rs for control, 143 Rs for non-users, and 81 Rs for users.

Water Storage Costs: The average monthly cost of storing water for all groups in both seasons is 59 Rupees. The monthly average cost in the dry season for all groups is 84 Rupees, while the monthly average cost in the rainy season for all groups is 5 Rupees. The monthly averages for each group across seasons are: 104 Rs for control, 21 Rs for non-users, and 46 Rs for users.

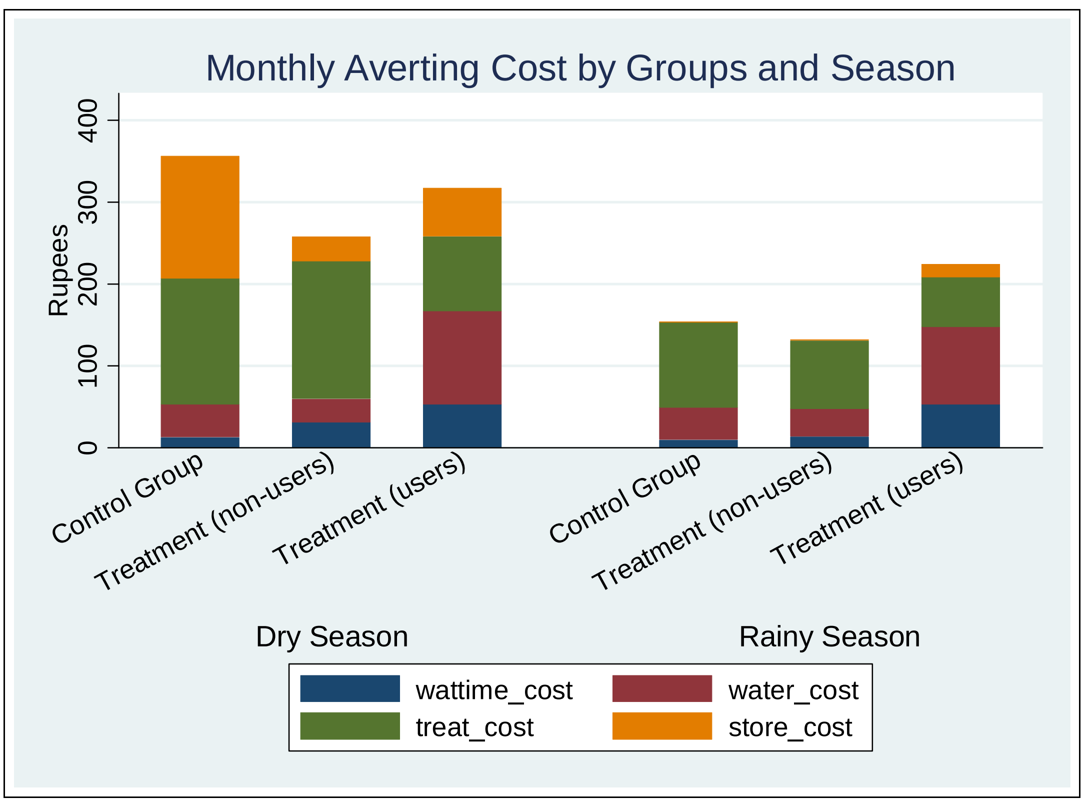
Total Averting Costs: The total average monthly averting costs for all groups across the two seasons is 283 Rupees. The total monthly average cost in the dry season for all groups is

309 Rs, while the total monthly average cost in the rainy season for all groups is 160 Rs. The total monthly averages for each group across seasons are: 294 Rs for control, 220 Rs for non-users, and 288 Rs for users.

Table 1. Monthly Averting Costs (Individual and Total) by Seasons and Groups

Season	Group	Wattime	Watcosts	Wattreat	Storecosts	Total
Both	Control	11	41	138	104	*294*
	Non-Users	25	30	143	21	*219*
	Users	53	108	81	46	*288*
	All	*25*	*51*	*128*	*59*	***283***
Dry	Control	12	41	153	150	*356*
	Non-Users	31	29	169	30	*258*
	Users	53	114	90	60	*317*
	All	*28*	*51*	*147*	*84*	***309***
Rainy	Control	9	40	103	2	*154*
	Non-Users	13	35	83	2	*132*
	Users	53	95	61	16	*224*
	All	*20*	*50*	*86*	*5*	***160***

Graph 1. Monthly Averting Costs (Individual and Total) by Seasons and Groups



4.2.2 Costs of Illness

Time Cost of Lost Workdays: The average monthly time cost of lost workdays due to diarrhea for all groups in both seasons is 22 Rupees. The monthly average cost in the dry season for all groups is 11 Rupees, while the monthly average cost in the rainy season for all groups is 46 Rupees. The monthly averages for each group across seasons are: 19 Rs for control, 23 Rs for non-users, and 25 Rs for users.

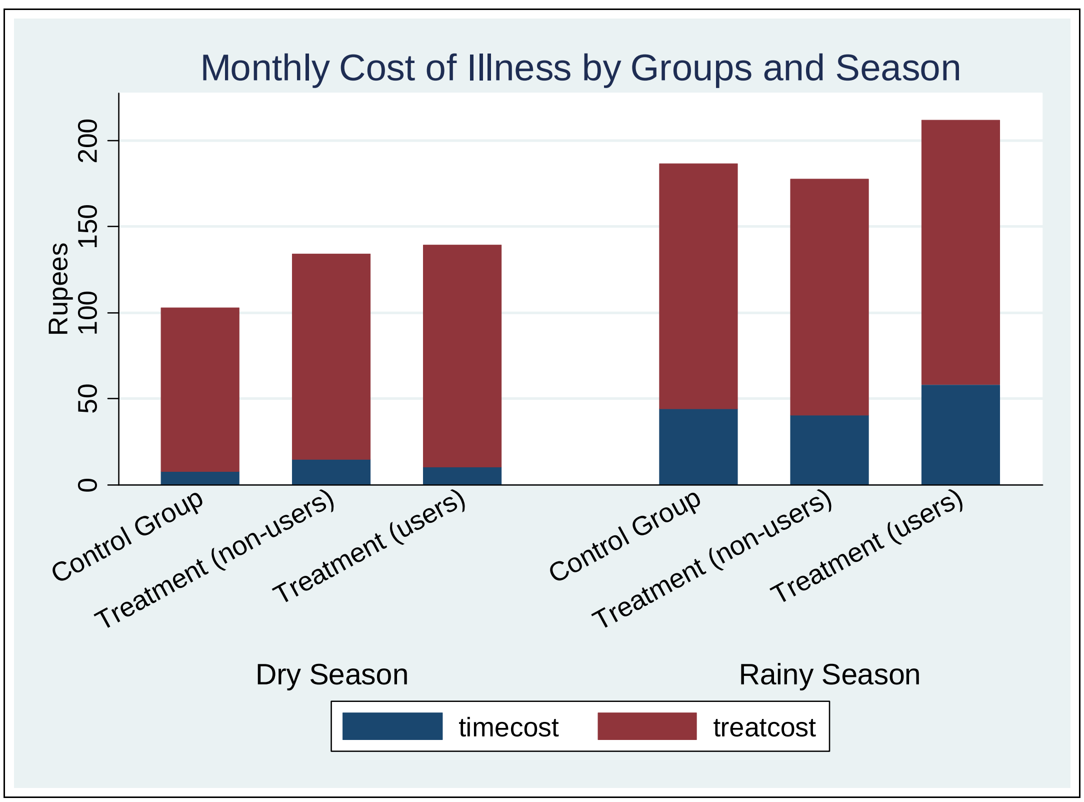
Treatment Cost: The average monthly cost of treatment for diarrhea for all groups in both seasons is 121 Rupees. The monthly average cost in the dry season for all groups is 112 Rupees, while the monthly average cost in the rainy season for all groups is 143 Rupees. The monthly averages for each group across seasons are: 110 Rs for control, 125 Rs for non-users, and 137 Rs for users.

Total Illness Costs: The total average monthly costs of illness due to diarrhea for all groups across the two seasons is 143 Rupees. The total monthly average cost in the dry season for all groups is 123 Rs, while the total monthly average cost in the rainy season for all groups is 188 Rs. The total monthly averages for each group across seasons are: 129 Rs for control, 147 Rs for non-users, and 143 Rs for users.

Table 2. Monthly Costs of Illness (Individual and Total) by Seasons and Groups

Season	Group	Timecost	Treatcost	Total
Both	Control	19	110	*129*
	Non-Users	23	125	*147*
	Users	25	137	*162*
	All	*22*	*121*	***143***
Dry	Control	8	95	*103*
	Non-Users	15	119	*134*
	Users	10	129	*139*
	All	*11*	*112*	***123***
Rainy	Control	44	142	*186*
	Non-Users	40	137	*177*
	Users	58	154	*212*
	All	*46*	*143*	***188***

Graph 2. Monthly Costs of Illness (Individual and Total) by Seasons and Groups



4.2.3 Cost Differences

In order to evaluate whether the changes in behaviors and their respective costs are due to the intervention, we must determine if these costs are different among the three groups. To do this, we do tests of differences in means among three groups. The results are presented below.

Table 3. Averting Costs Differences Among Groups by Seasons

Season	**Mean**			**Std. Dev.**			**Statistic & Probability**	
	Control	Non-Users	Users	Control	Non-Users	Users	Chi-squared	p-value
Both	294	219	288	575	452	378	133	0.000
Dry	356	258	317	671	525	431	100	0.000
Rainy	154	132	224	187	178	214	8	0.020

Graph 3. Averting Costs Means with 95% Confidence Intervals by Groups and Seasons

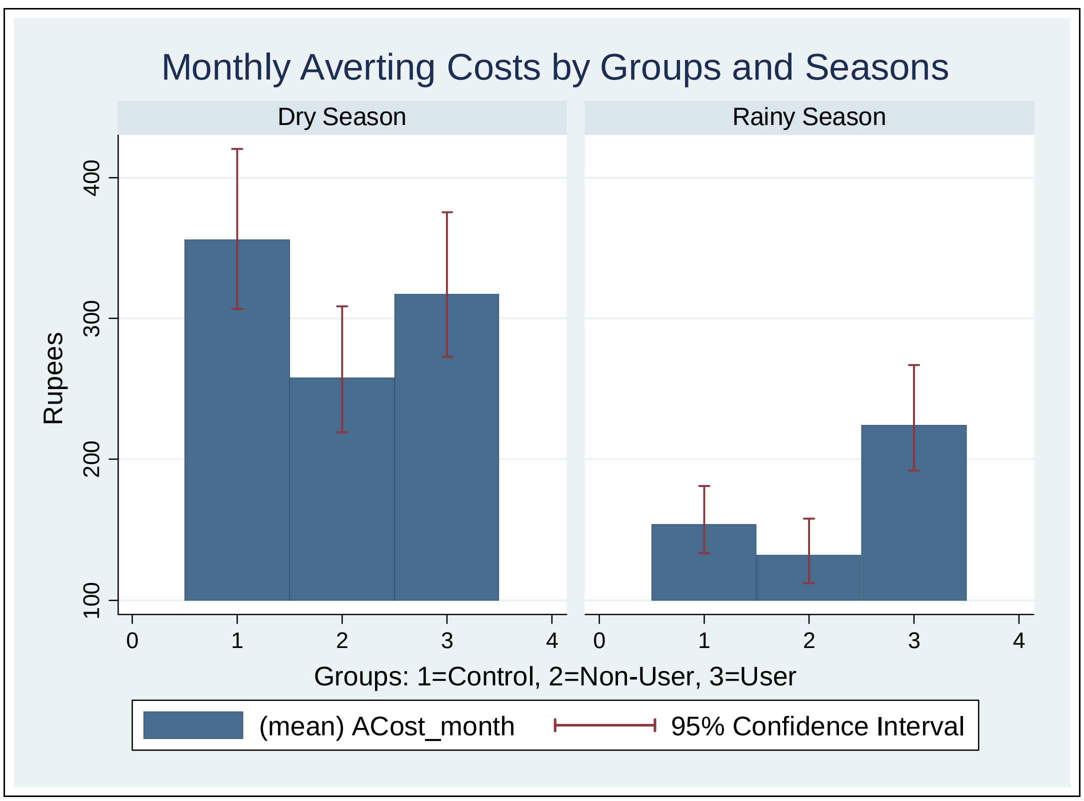
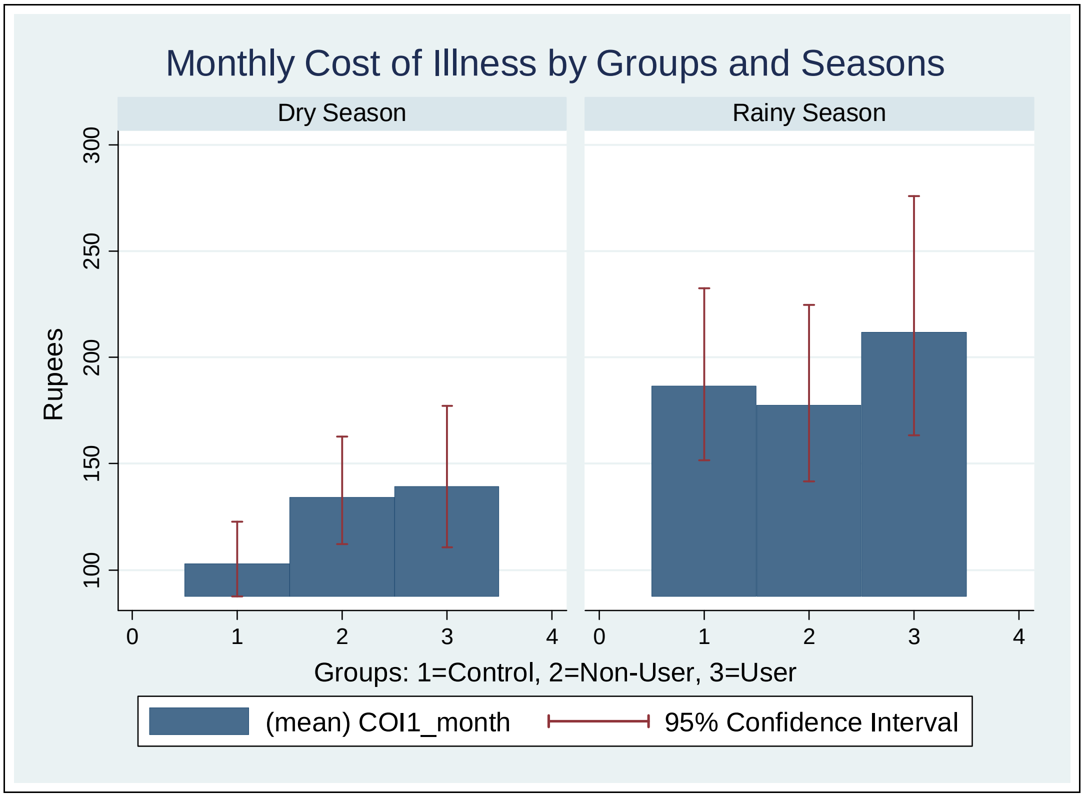


Table 4. Costs of Illness Differences Among Groups by Seasons

Season	**Mean**			**Std. Dev.**			**Statistic & Probability**	
	Control	Non-Users	Users	Control	Non-Users	Users	Chi-squared	p-value
Both	129	147	162	249	305	195	45	0.000
Dry	103	134	139	206	297	281	98	0.000
Rainy	182	177	212	317	320	320	0.04	0.981

Graph 4. Costs of Illness Means with 95% Confidence Intervals by Groups and Seasons



4.3 Cost Regressions

Performing a multivariate analysis will allow us to determine what factors contribute to the averting behavioral changes and their respective costs and the households' costs of illness. Therefore, I ran several ordinary least squares linear regressions of the different costs and household characteristics, such as household size and previous diarrhea occurrence, and controlled for household level variation using fixed effects and seasonal variation with a monsoon dummy. The household level fixed effects allow us to control for household level characteristics that were not included in the model (e.g. income and education) or unobservables that do not vary over the course of one year in order to isolate the effects of the variables included in the model, such as CWS%. The results are summarized in the following tables.

The variables used in the regressions are: hhsize (size of household), numu5 (number of children under 5), CWS_percent (variable indicating what percentage of the household's water comes from the clean water source), nonuser (dummy variable for whether the household is in a treatment village but is not an user), L.hhdiarr (lag variable of diarrhea in the previous round), and monsoon (dummy variable for whether it is the rainy or dry season).

Table 5. Regression Results for Total Averting Costs and Total Costs of Illness

	(1)	(3)
VARIABLES	ACost	COI
CWS_percent	-301.9***	-42.62
	(39.78)	(26.23)
nonuser	-140.7***	-1.293
	(21.55)	(14.21)
monsoon	-18.04***	15.98***
	(4.389)	(2.894)
L.hhdiarr	4.651	17.39***
	(6.623)	(4.367)
hhsize	-45.06	-4.040
	(49.25)	(32.47)
numu5	56.49	1.810
	(55.09)	(36.32)
Constant	279.9*	48.61
	(158.3)	(104.3)
Observations	2,464	2,464
R-squared	0.234	0.123

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 5a. Regression Results for All Averting Costs Components

VARIABLES	(1) wattime_cost	(2) water_cost	(3) treat_cost	(4) store_cost
CWS_percent	-24.70***	-132.3***	-126.2***	-18.68
	(2.076)	(3.004)	(35.49)	(13.13)
nonuser	-11.78***	-83.94***	-34.36*	-10.62
	(1.124)	(1.627)	(19.22)	(7.112)
monsoon	-0.729***	0.257	-7.699**	-9.874***
	(0.229)	(0.331)	(3.916)	(1.449)
L.hhdiarr	-0.598*	0.446	5.421	-0.619
	(0.346)	(0.500)	(5.908)	(2.186)
hhsize	-0.322	11.63***	-45.63	-10.73
	(2.570)	(3.719)	(43.93)	(16.26)
numu5	-2.674	-10.94***	57.38	12.72
	(2.875)	(4.160)	(49.15)	(18.18)
Constant	17.81**	21.03*	190.7	50.34
	(8.259)	(11.95)	(141.2)	(52.24)
Observations	2,464	2,464	2,464	2,464
R-squared	0.635	0.705	0.204	0.220

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 5b. Regression Results for All Costs of Illness Components

VARIABLES	(1) timecost1	(2) treatcost1
CWS_percent	-8.091	-34.53
	(5.933)	(21.96)
nonuser	2.363	-3.656
	(3.213)	(11.90)
monsoon	8.645***	7.331***
	(0.655)	(2.424)
L.hhdiarr	6.467***	10.93***
	(0.988)	(3.657)
hhsize	1.981	-6.021
	(7.344)	(27.19)
numu5	-0.722	2.532
	(8.216)	(30.42)
Constant	-5.822	54.44
	(23.60)	(87.38)

Observations	2,464	2,464
R-squared	0.191	0.103

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The regression coefficients for the averting costs in Table 5 show that averting costs decrease as the percentage of the household's water that comes from the clean water source increases, as compare to the control group. The coefficient for the CWS_percent variable is negative and statistically significant. The results also show that being a non-user decreases averting costs, as compare to the control group. The coefficient for the nonuser dummy is negative and statistically significant. For averting costs, the only other statistically significant variable is the monsoon dummy, which indicates that averting costs decrease in the rainy season. Since clean water use had a statistically significant impact on averting costs, I conducted the same regression for the individual components of averting costs in order to more closely examine this relationship. The regression coefficients for Table 5a show that an increase in the percentage of clean water use decreased all averting costs, except for storage costs, significantly as compared to the control group. Also, being a user decreased all averting costs, except for storage costs, significantly as well. However, CWS_percent had a greater decreasing effect on averting costs compared to nonuser, as indicated by their larger negative coefficients. It is important to note that although averting costs decrease as more of the user's water comes from the clean water source, users use multiple water sources. Therefore, we must evaluate the effects of CWS_percent on users based on how much of their water actually comes from the clean water source. This is discussed in the following section.

The regression coefficient for costs of illness in Table 5 show that if the household had diarrhea in the previous period, their costs of illness would increase. The coefficient for the

L.hhdiarr variable is positive and statistically significant. Also, costs of illness are increased during the rainy season. The coefficient for the monsoon dummy is positive and statistically significant. These results are consistent with previous results, which indicate no differences in diarrhea or its associated costs among groups and increased diarrhea prevalence or its associated costs during the monsoon. These results also indicate that users that experienced diarrhea in previous periods will spend more money coping with the illness than those that did not have diarrhea recently. I conducted the same regression for the individual components of costs of illness in order to more closely examine this relationship. The coefficients in Table 5b further this relationship: time costs of illness and treatment expenditures both increase during the rainy season and if the household experience diarrhea in the previous period.

5. DISCUSSION

5.1 Behavior Trends

It is important to observe the trends due to the large frequency of data collection that could cause a Hawthorne effect or respondent fatigue. The Hawthrone effect is, “the idea that being in an experiment influences the subject behavior” (Stock & Watson, 2007). However, there is no way to test this with the data used in this study. There is some indication that respondent fatigue occurred because the trends show that after several rounds of surveys, we see identical answers for self-reported behaviors. “Fatigue may set in as respondents tire of the frequent visits, the repetitive questions and the moderate demands on her time” (Kremer, 2009). However, this repetitiveness is most likely due to habits and not respondent fatigue, since the tracking survey was relatively short and not very time consuming.

5.2 Behavior Costs

In general, the results from these cost calculations demonstrate that the majority of a household's averting costs are derived from the time cost of collecting water and the majority of a household's costs of illness come from treatment costs. Higher cost of illness in users is probably due to the fact that users were found to have more diarrhea than control and non-users (Poulos et al., 2009). By simply looking at the mean averting costs among groups, it appears that treatment users incur higher averting costs than non-user and households in control villages. Thus, in order to determine whether users actually incur higher averting cost, all else equal, we must look at the results from the multivariate linear regression analysis utilizing fixed effects.

The regressions results indicate that after controlling for household characteristics and other factors, such as weather, the monthly averting costs decrease by 12 Rupees as the percentage of household water that comes from CWS increases by 1%. This value comes from the coefficient for CWS_percent for the total averting cost regression (c = -302). On average, 48% of the users' water comes from the clean water source. Thus, the expected value of the benefits, or reduction in costs, for treatment users compared to the control group is: 0.48(-302) = 145 Rupees each week. This is approximately 580 Rupees per month, or 32% of monthly income (assuming a daily wage of 60 Rupees). The regressions results also indicate that costs of illness increase by as much as 64 Rupees a month in during the rainy season, and by as much as 17 Rupees a week if the household experience diarrhea in the previous week.

6. CONCLUSION

This study provided several insights into the behavioral changes and the associated costs incurred by households in villages that received a Community Water System. Using revealed preferences methods of averting costs and costs of illness, we were able to measure the welfare benefits to households in treatment villages. We were able to isolate the effects of purchasing clean water from the treatment plant by using longitudinal data. Due to the large number of degrees of freedom, we were able to add household fixed effects to better control for time invariant confounding variables.

The results from this study provide only limited information on the economic impact of this intervention to households. This paper only calculated the costs reductions, especially averting costs, associated with the purchase of clean water. Based on the results of this paper, providing clean water at an affordable price can reduce the household's averting costs by an average of 32% of their monthly income. However, it is important to note that the amount of clean water purchased should be considered. Because purchasing water incurs significant amounts of time costs, while not completely replacing all other water sources, users continue to have treatment costs. Therefore, the averting costs for users do not decrease until enough of their water comes from the clean water source. Thus, the goal should be to reduce the cost of access to enough clean water in order to reduce treatment costs. These can be achieved in several ways, including reducing time cost of collecting water by placing more distribution centers.

7. References

Champ, P.A., Boyle, K.J., and Brown, T.C. 2003. "Introduction to Revealed Preference Methods" and "Defensive Behavior and Damage Cost Methods". *A Primer on Nonmarket Valuation*. Dordrecht: Kluwer Academic.

The Cebu Study Team. 1991. Underlying and Proximate Determinants of Child Health: The Cebu Longitudinal Health and Nutrition Study. *American Journal of Epidemiology*, 133:185-201

Fewtrell, L., Kaufmann, R.B., Kay, D., Enanoria, W., Haller, L. & Colford, J.M. 2005. Water, Sanitation, and Hygiene Interventions to Reduce Diarrhoea in Less Developed Countries: A Systematic Review and Meta-Analysis. *The Lancet Infectious Diseases*, 5(1), 42-52.

Kremer, M., Miguel E., Null, C., Van Dusen, E., Peterson Zwane, A. 2009. "Measuring Diarrhea: Quantifying Hawthorne effects in frequently collected data". The Bill and Melinda Gates Foundation.

Patil, S.R. and Pattanayak, S.K. 2010. Behaviors exposed: Panel data evidence on environmental health externalities in rural India.

Pattanayak, S.K. and Pfaff, A. 2009. Behavior, Environment, and Health in Developing Countries: Evaluation and Valuation. *Annual Review of Resource Economics*, 1:183-222.

Pattanayak, S.K. and Poulos, C. 2009. Monitoring and Evaluation of Health and Socio-economic Impacts of Water and Sanitation Initiatives. Report on Short-Term Impacts Based on Analysis of Round 1 and Round 2 Data. *Research Triangle Park, NC, RTI International*.

Pattanayak, S.K., Poulos, C., Yang, J.C., & Patil, S.R. Forthcoming. How valuable are environmental health interventions? Economic evaluation of a water and sanitation project in Maharashtra, India. *Bulletin of the World Health Organization.*

Pattanayak, S.K., Poulos, C., Yang, J.C., Patil, S.R., & Wendland, K.J. 2009. Of Taps and Toilets: Quasi-experimental protocols for evaluating community-demand driven projects. *Journal of Water and Health*, 7(3): 434-451.

Pattanayak, S.K., Yang, J-C, Dickinson, K.L., Poulos, C., Patil, S.R., Mallick, R.J., Blistein, J.L., & Praharaj, P. 2009. Shame or Subsidy Revisited: Social Mobilization for Sanitation in Orissa, India. *Bulletin of the World Health Organization*, 87: 580–587.

Poulos, C., Patil, S.R., Pattanayak, S.K., & Chen, J. 2009. Monitoring and Evaluation of Health and Socio-economic Impacts of Water and Sanitation Initiatives. Final Report on Tracking Study. *Research Triangle Park, NC, RTI International*.

Poulos, C., Pattanayak, S.K., Yang, J.C., Blitstein, J., & Bubna, A. Monitoring and Evaluation of Health and Socio-economic Impacts of Water and Sanitation Initiatives. Interim Report on Baseline Data. *Research Triangle Park, NC, RTI International*; 2007.

Stock, J. H., and Watson, M.W. 2007. "Chapter 10: Regression with Panel Data." *Introduction to Econometrics*. Boston: Pearson/Addison Wesley.

UNDP, 2006. Beyond scarcity: power, poverty and the global water crisis. *Human Development Report 2006*. New York: United Nations Development Program.

Whittington, D., Hanemann, M., Sadoff, C., & Jeuland, M. 2009. The Challenge of Improving Water and Sanitation Services in Less Developed Countries. *Foundations and Trends in Microeconomics*, Vol. 4, Nos. 6-7: 469-609.

WHO, 2006. Meeting the MDG drinking water and sanitation target : the urban and rural challenge of the decade. *WHO Library Cataloguing-in-Publication Data*. World Health Organization and UNICEF. ISBN 978 92 4 156325 3.

Zwane and Kremer. 2007. What Works in Fighting Diarrheal Diseases in Developing Countries? A Critical Review. *World Bank Research Observer*, 22:1–24.

APPENDIX A: ANALYSIS DO-FILE

```
set more off
set memory 100m
cd "H:\MP\FinalMP"

*******************************************
use "MP_Tracking1.dta", clear


*****************************************************
*************** CALCULATING THE COSTS ***************
*****************************************************


***** AVERTING COSTS *****

gen wattime_cost = (wattime/60)*(60/8)*(30/100)
label variable wattime_cost "cost of time collecting water Rs/Week"
gen wattime_cost_month = wattime_cost * 4
label variable wattime_cost_month " monthly costs of time collecting water in
Rupees"

gen water_cost = watpay*watprice
label variable water_cost "cost of buying clean water in Rs/Week"
gen water_cost_month = water_cost * 4
label variable water_cost_month "montly cost of buying water in Rupees"

gen boilcost = boil_freq*(other_water/4)*(27/5)
gen filtercost = filter_freq*(other_water/4)*(2/4)
gen treat_cost = boilcost + filtercost
label variable treat_cost "cost of treating water in Rs/Week"
gen treat_cost_month = treat_cost * 4
label variable treat_cost_month "montly cost of treating water in Rupees"

gen store_cost = newcont*contprice
label variable store_cost "cost of storing/transporting water container in
Rs/Week"
gen store_cost_month = store_cost * 4
label variable store_cost_month "montly cost of storing/transporting water
container in Rupees"

gen ACost =  wattime_cost + water_cost + treat_cost + store_cost
label variable ACost "total averting cost (weekly)"
gen ACost_month = wattime_cost_month + water_cost_month + treat_cost_month +
store_cost_month
label variable ACost_month "total monthly cost of averting behavior"


***** COST OF ILLNESS *****

gen timecost = worklost * 60 * (30/100)
label variable timecost "cost of workdays lost by sickness or taking care of
sick in 1 week"
gen timecost_month = timecost * 4
```

```
label variable timecost_month "monthly cost of workdays lost by illness or
taking care of sick"

gen treatcost = trtcost
label variable treatcost "expendiure on treatment by the HH in 1 week"
gen treatcost_month = treatcost * 4
label variable treatcost_month "monthly HH expenditure on treatment"

gen COI = timecost + treatcost
label variable COI "total cost of illness (weekly)"
gen COI_month = timecost_month + treatcost_month
label variable COI_month "total cost of illness"

*** for the population as a whole:
gen timecost1 = timecost
replace timecost1 = 0 if timecost == .
gen timecost1_month = timecost1 * 4
gen treatcost1 = treatcost
replace treatcost1 = 0 if treatcost == .
gen treatcost1_month = treatcost1 * 4
gen COI1 = COI
replace COI1 = 0 if COI == .
gen COI1_month = COI1 * 4


save "MP_Tracking_Costs1.dta", replace

*******************************************
log using "CalculatingCosts.log", replace
use "MP_Tracking_Costs1.dta", clear


************************************************
*************** SUMM & TEST DIFF ***************
************************************************


***** AVERTING COSTS BY GROUPS AND MONSOON *****
des wattime_cost_month water_cost_month treat_cost_month store_cost_month
ACost_month
sum wattime_cost_month water_cost_month treat_cost_month store_cost_month
ACost_month
tabstat wattime_cost_month water_cost_month treat_cost_month store_cost_month
ACost_month, statistics(mean) by(groups) columns(variables)
by monsoon, sort: tabstat wattime_cost_month water_cost_month
treat_cost_month store_cost_month ACost_month, statistics(mean) by(groups)
columns(variables)


***** COSTS OF ILLNESS BY GROUPS AND MONSOON *****
des timecost_month treatcost_month COI_month
sum timecost_month treatcost_month COI_month
tabstat timecost_month treatcost_month COI_month, statistics(mean) by(groups)
columns(variables)
by monsoon, sort: tabstat timecost_month treatcost_month COI_month,
statistics(mean) by(groups) columns(variables)
```

```
*** for the population as a whole:
des timecost1_month treatcost1_month COI1_month
sum timecost1_month treatcost1_month COI1_month
tabstat timecost1_month treatcost1_month COI1_month, statistics(mean)
by(groups) columns(variables)
by monsoon, sort: tabstat timecost1_month treatcost1_month COI1_month,
statistics(mean) by(groups) columns(variables)


***** AVERTING COSTS DIFFEREMCE BY GROUPS AND MONSOON *****
oneway ACost_month groups, bonferroni tab
oneway ACost_month groups if monsoon == 0, bonferroni tabulate
oneway ACost_month groups if monsoon == 1, bonferroni tabulate


***** COST OF ILLNESS DIFFERENCE BY GROUPS AND MONSOON *****
oneway COI_month groups, bonferroni tab
oneway COI_month groups if monsoon == 0, bonferroni tabulate
oneway COI_month groups if monsoon == 1, bonferroni tabulate

*** for the population as a whole:
oneway COI1_month groups, bonferroni tab
oneway COI1_month groups if monsoon == 0, bonferroni tabulate
oneway COI1_month groups if monsoon == 1, bonferroni tabulate


***** BEHAVIOR CHANGES BY GROUPS AND MONSOON *****
des wattime watpay d1 cleanctr
sum wattime watpay d1 cleanctr
tab round groups, sum(wattime) mean
tab round groups, sum(watpay) mean
tab round groups, sum(d1) mean
tab round groups, sum(cleanctr) mean


log close


******************************************
*************** TIME PLOTS ***************
******************************************

sort round


*** Generate Mean Values of Behaviors:
egen mean_wattime = mean(wattime), by (round groups)
egen mean_watpay = mean(watpay), by (round groups)
egen mean_d1 = mean(d1), by (round groups)
egen mean_cleanctr = mean(cleanctr), by (round groups)


*** Plot Mean Values of Behaviors:
twoway (connected mean_wattime round if groups == 1, msymbol(square)
msize(small)) (connected mean_wattime round if groups == 2, msymbol(triangle)
msize(small)) (connected mean_wattime round if groups == 3, msymbol(circle)
msize(small)), xline(7 14) ytitle("Minutes per Week") xtitle("Rounds")
```

```
title("HH Time Spent Collecting Water") legend(label(1 "Control") label(2
"Treatment non-users") label(3 "Treatement users"))
graph save Graph "plot_wattime.gph", replace
twoway (connected mean_watpay round if groups == 1, msymbol(square)
msize(small)) (connected mean_watpay round if groups == 2, msymbol(triangle)
msize(small)) (connected mean_watpay round if groups == 3, msymbol(circle)
msize(small)), xline(7 14) ytitle("Percentage") xtitle("Rounds") title("HH
Paying for Water") legend(label(1 "Control") label(2 "Treatment non-users")
label(3 "Treatement users"))
graph save Graph "plot_watpay.gph", replace
twoway (connected mean_d1 round if groups == 1, msymbol(square) msize(small))
(connected mean_d1 round if groups == 2, msymbol(triangle) msize(small))
(connected mean_d1 round if groups == 3, msymbol(circle) msize(small)),
xline(7 14) ytitle("Percentage") xtitle("Rounds") title("HH Treating Water")
legend(label(1 "Control") label(2 "Treatment non-users") label(3 "Treatement
users"))
graph save Graph "plot_d1.gph", replace
twoway (connected mean_cleanctr round if groups == 1, msymbol(square)
msize(small)) (connected mean_cleanctr round if groups == 2,
msymbol(triangle) msize(small)) (connected mean_cleanctr round if groups ==
3, msymbol(circle) msize(small)), xline(7 14) ytitle("Percentage")
xtitle("Rounds") title("HH Cleaning Storage Container") legend(label(1
"Control") label(2 "Treatment non-users") label(3 "Treatement users"))
graph save Graph "plot_cleanctr.gph", replace


*****************************************
*************** BAR GRAPHS ***************
*****************************************


***** AVERTING COSTS *****
graph bar (mean) wattime_cost_month (mean) water_cost_month (mean)
treat_cost_month (mean) store_cost_month, over(groups) over(monsoon) stack
ytitle(Rupees) legend(label(1 "wattime_cost") label(2 "water_cost") label(3
"treat_cost") label(4 "store_cost")) title(Monthly Averting Cost by Groups
and Season)
graph save Graph "averting_cost.gph", replace


***** COST OF ILLNESS *****
graph bar (mean) timecost_month (mean) treatcost_month, over(groups)
over(monsoon) stack ytitle(Rupees) legend(label(1 "timecost") label(2
"treatcost")) title(Monthly Cost of Illness by Groups and Season)
graph save Graph "illness_cost.gph", replace
graph bar (mean) timecost1_month (mean) treatcost1_month, over(groups)
over(monsoon) stack ytitle(Rupees) legend(label(1 "timecost") label(2
"treatcost")) title(Monthly Cost of Illness by Groups and Season)
graph save Graph "illness_cost_pop.gph", replace


***** AVERTING COSTS WITH 95% CI *****
use "MP_Tracking_Costs1.dta", clear
collapse (mean) meanACost_month=ACost_month (sd) sdACost_month=ACost_month
(count) n=ACost_month, by(groups monsoon)
generate hiACost_month = meanACost_month + invttail(n-
1,0.005)*(sdACost_month/sqrt(n))
```

```
generate loACost_month = meanACost_month - invttail(n-
1,0.025)*(sdACost_month/sqrt(n))
graph twoway (bar meanACost_month groups) (rcap hiACost_month loACost_month
groups), by(monsoon) ytitle(Rupees) legend(label(2 "95% Confidence
Interval"))
graph save Graph "averting_cost_95CI.gph", replace


***** COST OF ILLNESS WITH 95% CI *****
use "MP_Tracking_Costs1.dta", clear
collapse (mean) meanCOI_month=COI_month (sd) sdCOI_month=COI_month (count)
n=COI_month, by(groups monsoon)
generate hiCOI_month = meanCOI_month + invttail(n-
1,0.005)*(sdCOI_month/sqrt(n))
generate loCOI_month = meanCOI_month - invttail(n-
1,0.025)*(sdCOI_month/sqrt(n))
graph twoway (bar meanCOI_month groups) (rcap hiCOI_month loCOI_month
groups), by(monsoon) ytitle(Rupees) legend(label(2 "95% Confidence
Interval"))
graph save Graph "illness_cost_95CI.gph", replace

*** for whole population:
use "MP_Tracking_Costs1.dta", clear
collapse (mean) meanCOI1_month=COI1_month (sd) sdCOI1_month=COI1_month
(count) n=COI1_month, by(groups monsoon)
generate hiCOI1_month = meanCOI1_month + invttail(n-
1,0.005)*(sdCOI1_month/sqrt(n))
generate loCOI1_month = meanCOI1_month - invttail(n-
1,0.025)*(sdCOI1_month/sqrt(n))
graph twoway (bar meanCOI1_month groups) (rcap hiCOI1_month loCOI1_month
groups), by(monsoon) ytitle(Rupees) legend(label(2 "95% Confidence
Interval"))
graph save Graph "illness_cost_pop_95CI.gph", replace


*******************************************
log using "Regressions.log", replace
use "MP_Tracking_Costs1.dta", clear


***************************************************
*************** REGRESSION ANALYSIS ***************
***************************************************

ssc install outreg2, replace


*** Set as Panel Data:
sort hhid round
xtset hhid round

*** User Dummy: use variable "user"

*** Non-User Dummy:
generate nonuser = group == 2
replace nonuser = 0 if nonuser == .
move nonuser groups
```

```
*** testing for colinearity:
pwcorr hhsize numu5 user nonuser L.hhdiarr monsoon, sig
pwcorr hhsize numu5 CWS_percent nonuser L.hhdiarr monsoon, sig

*** large oneway anova:
loneway ACost hhid
loneway COI1 hhid


***** ACOST REGRESSIONS *****

areg ACost CWS_percent nonuser monsoon L.hhdiarr hhsize numu5, absorb(hhid)
outreg2 using "reg_costs.doc", replace
xtreg ACost CWS_percent nonuser monsoon L.hhdiarr hhsize numu5, fe
estimates store fixed1
xtreg ACost CWS_percent nonuser monsoon L.hhdiarr hhsize numu5, re
estimates store random1
hausman fixed1 random1

areg wattime_cost CWS_percent nonuser monsoon L.hhdiarr hhsize numu5,
absorb(hhid)
outreg2 using "reg_Acosts.doc", replace
xtreg wattime_cost CWS_percent nonuser monsoon L.hhdiarr hhsize numu5, fe
xtreg wattime_cost CWS_percent nonuser monsoon L.hhdiarr hhsize numu5, re

areg water_cost CWS_percent nonuser monsoon L.hhdiarr hhsize numu5,
absorb(hhid)
outreg2 using "reg_Acosts.doc", append
xtreg water_cost CWS_percent nonuser monsoon L.hhdiarr hhsize numu5, fe
xtreg water_cost CWS_percent nonuser monsoon L.hhdiarr hhsize numu5, re

areg treat_cost CWS_percent nonuser monsoon L.hhdiarr hhsize numu5,
absorb(hhid)
outreg2 using "reg_Acosts.doc", append
xtreg treat_cost CWS_percent nonuser monsoon L.hhdiarr hhsize numu5, fe
xtreg treat_cost CWS_percent nonuser monsoon L.hhdiarr hhsize numu5, re

areg store_cost CWS_percent nonuser monsoon L.hhdiarr hhsize numu5,
absorb(hhid)
outreg2 using "reg_Acosts.doc", append
xtreg store_cost CWS_percent nonuser monsoon L.hhdiarr hhsize numu5, fe
xtreg store_cost CWS_percent nonuser monsoon L.hhdiarr hhsize numu5, re


***** COI REGRESSIONS *****

areg COI CWS_percent nonuser monsoon L.hhdiarr hhsize numu5, absorb(hhid)
outreg2 using "reg_costs.doc", append
xtreg COI CWS_percent nonuser monsoon L.hhdiarr hhsize numu5, fe
estimates store fixed2
xtreg COI CWS_percent nonuser monsoon L.hhdiarr hhsize numu5, re
estimates store random2
hausman fixed2 random2
```

```
areg timecost CWS_percent nonuser monsoon L.hhdiarr hhsize numu5,
absorb(hhid)
outreg2 using "reg_COI.doc", replace
xtreg timecost CWS_percent nonuser monsoon L.hhdiarr hhsize numu5, fe
xtreg timecost CWS_percent nonuser monsoon L.hhdiarr hhsize numu5, re

areg treatcost CWS_percent nonuser monsoon L.hhdiarr hhsize numu5,
absorb(hhid)
outreg2 using "reg_COI.doc", append
xtreg treatcost CWS_percent nonuser monsoon L.hhdiarr hhsize numu5, fe
xtreg treatcost CWS_percent nonuser monsoon L.hhdiarr hhsize numu5, re

*** for the population as a whole:
areg COI1 CWS_percent nonuser monsoon L.hhdiarr hhsize numu5, absorb(hhid)
outreg2 using "reg_costs.doc", append
xtreg COI1 CWS_percent nonuser monsoon L.hhdiarr hhsize numu5, fe
estimates store fixed3
xtreg COI1 CWS_percent nonuser monsoon L.hhdiarr hhsize numu5, re
estimates store random3
hausman fixed3 random3

areg timecost1 CWS_percent nonuser monsoon L.hhdiarr hhsize numu5,
absorb(hhid)
outreg2 using "reg_COIpop.doc", replace
xtreg timecost1 CWS_percent nonuser monsoon L.hhdiarr hhsize numu5, fe
xtreg timecost1 CWS_percent nonuser monsoon L.hhdiarr hhsize numu5, re

areg treatcost1 CWS_percent nonuser monsoon L.hhdiarr hhsize numu5,
absorb(hhid)
outreg2 using "reg_COIpop.doc", append
xtreg treatcost1 CWS_percent nonuser monsoon L.hhdiarr hhsize numu5, fe
xtreg treatcost1 CWS_percent nonuser monsoon L.hhdiarr hhsize numu5, re


log close
```

APPENDIX B: DATASET DO-FILE

```
set more off
set memory 100m

use "H:\RTI\acumen\survey\data\hhpanel_indicators.dta", clear


*****************************************************
*************** CREATING THE DATA SET ***************
*****************************************************


***** create village id
rename a11 villageid


***** create monsoon dummy
gen monsoon = 0
replace monsoon = 1 if (round >= 7 & round <= 14)
label variable monsoon "dummy for monsoon"
label define monsoon 0 "0 Dry Season" 1 "1 Rainy Season"
label values monsoon monsoon


***** liters of water (per day and per week)
egen watquantd = rowtotal (c5as1 c5as2 c5as3)
label variable watquantd "total liters of water hh collected in a day"
egen watquantw = rowtotal (c5bs1 c5bs2 c5bs3)
label variable watquantw "total liters of water hh collected in 1 week"


***** create CWS percent
gen CWS1 = 0
replace CWS1 = c5bs1 if c1s1 == 8
gen CWS2 = 0
replace CWS2 = c5bs2 if c1s2 == 8
gen CWS3 = 0
replace CWS3 = c5bs3 if c1s3 == 8
gen CWS = CWS1 + CWS2 + CWS3
gen CWS_percent = CWS/ watquantw
gen other_water = watquantw - CWS


***** create boil frequency
gen boil_freq1 = 0
replace boil_freq1 = (8-d4s1)/7 if dsns1 == 1
gen boil_freq2 = 0
replace boil_freq2 = (8-d4s2)/7 if dsns2 == 1
egen boil_freq = rowtotal (boil_freq1 boil_freq2)


***** create filter frequency
gen filter_freq1 = 0
replace filter_freq1 = (8-d4s1)/7 if dsns1 <= 6 & dsns1 >= 2
gen filter_freq2 = 0
```

```
replace filter_freq2 = (8-d4s2)/7 if dsns2 <= 6 & dsns1 >= 2
egen filter_freq = rowtotal (filter_freq1 filter_freq2)


***** paid for water?  (per week)
gen watpay = 0
replace watpay = 1 if (c7s1 == 1 | c7s2 == 1 | c7s3 == 1)
label variable watpay "did the hh pay for water"


***** how much did you pay for water?  (per week)
egen watprice = rowtotal (c8as1 c8as2 c8as3)
label variable watprice "how much HH paid for water"


***** minutes collecting water = walking to source + waiting at source (per
trip)
gen triptimes1 = (rc6s1*2)+ rc7s1
label variable triptimes1 "minutes spent collecting water per trip, source 1"
gen triptimes2 = (rc6s2*2)+ rc7s2
label variable triptimes2 "minutes spent collecting water per trip, source 2"
gen triptimes3 = (rc6s3*2)+ rc7s3
label variable triptimes3 "minutes spent collecting water per trip, source 3"


***** number of trips to collect water (per week)
gen wattrips1 = c5bs1/10
label variable wattrips1 "number of trips per week, source 1"
gen wattrips2 = c5bs2/10
label variable wattrips2 "number of trips per week, source 2"
gen wattrips3 = c5bs3/10
label variable wattrips3 "number of trips per week, source 3"


***** time spent collecting water = minutes per trip * trips per week (per
week)
gen wattimes1 = triptimes1*wattrips1
label variable wattimes1 "minutes spent collecting water per week, source 1"
gen wattimes2 = triptimes2*wattrips2
label variable wattimes2 "minutes spent collecting water per week, source 2"
gen wattimes3 = triptimes3*wattrips3
label variable wattimes3 "minutes spent collecting water per week, source 3"
egen wattime = rowtotal (wattimes1 wattimes2 wattimes3)
label variable wattime "total minutes spent collecting water per week"


***** bought new storage/transport container?
gen newcont = 0
replace newcont = 1 if (e4aa > 0 | e4ba > 0)
label variable newcont "hh bought new container"


***** cost of storage/transport container
gen contprice = e4ae+ e4be
label variable contprice "hh cost of container"
```

```
***** total HH workdays lost = being sick + taking care of sick (per week)
gen worklost = b12bds1 + b12bds2 + (b12bhs1 + b12bhs2)/8
replace worklost = . if b1 != 1
label variable worklost "days of work lost by the HH in taking care of the
sick members or by being sick in 1 week"


***** total treatment cost (per week)
*egen trtcost = rowtotal(b11s1 b11s2)
*replace trtcost = . if b1 != 1
*label variable trtcost "expendiure on treatment by the HH"


sort hhid round
save "H:\MP\FinalMP\MP_Tracking1.dta", replace
```