RURAL WATER SOURCE CHOICE:
A CHOICE EXPERIMENT FROM MERU, KENYA

Annalise G. Blum

A thesis submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Master of Science in the Department of Environmental Sciences and Engineering in the Gillings School of Global Public Health.

Chapel Hill
2014

Approved by:

Dale Whittington

Joseph Cook

Greg Characklis

ABSTRACT

Annalise G. Blum: Rural water source choice: A choice experiment from Meru, Kenya
(Under the direction of Dale Whittington)

A stated preference choice experiment is used to investigate factors important to rural households when selecting a primary water source. In particular, the guiding research questions are: (1) how do rural Kenyans trade-off time and price when selecting a water source? (2) how do rural Kenyans value time spent collecting water? (3) how are household characteristics relevant to water source choice? The choice experiment was administered to 388 respondents in rural Kenya. Source price and collection time are important to respondents without an at-home source. Neither income nor education is found to affect sensitivity to source price. Valuation of time is estimated to be 37% of the local unskilled wage rate on average. This choice experiment illustrates a relatively simple method of identifying water source preferences of households that can be used in other locations, however there are challenges in collecting high quality preference data and analyzing the data appropriately.

ACKNOWLEDGEMENTS

I would like to thank Dr. Dale Whittington, my advisor, for his advice and guidance throughout this process. I am also grateful to the other members of my committee: Dr. Joe Cook, who worked patiently with me both in the field and during data analysis, and Dr. Greg Characklis for his time and feedback. I thank Dr. Peter Kimuyu and Josephine Gakkii Gatua for their hard work leading the fieldwork and data collection. I am grateful to all of the household questionnaire participants and others we spoke to during site selection and piloting of the study. This project could not have been completed without funding from Environment for Development-Kenya. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE-1144081. Finally, I thank my friends and family for their support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

asc	alternative specific constant
CL	Conditional logit
CM	Choice modeling
CV	Contingent valuation
F	Fahrenheit
HEV	Heteroscedastistic extreme value
HH	Household
IIA	Independence of Irrelevant Alternatives
JMP	Joint Monitoring Program
Kg	Kilogram
Ksh	Kenyan Shillings
L	Liters
MNL	Multinomial logit
MXL	Mixed logit
n	sample size
NL	Nested logit
UNICEF	United Nations Children's Fund
USD	United States Dollars
WHO	World Health Organization
WTP	Willingness To Pay

1. Introduction

Of the 768 million people globally without access to improved drinking water, 83% live in rural areas (WHO, 2013). To extend access to rural populations, it is important to understand their preferences. For example, are people willing to walk to public taps? Would they rather pay more for water and not have to walk as far? Or would they prefer to pay for at-home piped connections? Hope (2006) explains, "policy that fails to respond to the preferences of the target beneficiaries is likely to allocate resources, capacity, and funds inefficiently and ineffectively." In Kenya, only 54% of rural people have access to improved sources of drinking water compared to 83% of urban dwellers (WHO, 2013). This makes Kenya a particularly important region to study the water source preferences of rural people. I use a stated preference choice experiment to investigate how water source and household attributes affect water source selection. The primary research questions guiding this work are:

(1) How do rural Kenyans trade-off time and price when selecting a water source?

(2) How do rural Kenyans value time spent collecting water?

(3) How are household characteristics (such as income and education) relevant to water source choice?

This work is part of a larger study investigating revealed water source choice and water demand carried out in partnership with Environment for Development – Kenya. The stated preference component of the data is the focus of this thesis. The paper is organized as follows. First, I present the findings from a systematic literature review on water source choice. I then lay out the theoretical framework for the analysis followed by the research design and fieldwork.

Next I present the results including demographics of the study population, current water source practices, the choice experiment findings, and estimated coping costs. Finally, I conclude with a discussion of these results.

2. Literature review

The first study of water source choice in developing countries, *Drawers of Water: Domestic Water Use in East Africa*, was published in 1972. Despite increased international funding for improved water supply in developing countries in the years since (Thompson et al., 2001), surprisingly few studies have focused on household water source choice. General consensus in the literature is that both water source and household characteristics are relevant to household source choice (Nauges & Whittington, 2009). However, there is little agreement regarding the relative importance of varying source attributes and household characteristics in these decisions. Both stated and revealed preference methods have been used to study water source choice. Stated preference methods, such as contingent valuation and choice experiments, present respondents with a hypothetical question or set of choice tasks and have been used widely in developing countries (Whittington, 2010). In contrast, revealed preference methods, such as the travel cost method, estimate valuation of non-market goods based upon observed behavior.

2.1 Systematic review methods

The goal of this systematic literature review is to identify which water source attributes and household characteristics have been found to be important to household water source choice, as well as weaknesses in the methods used in existing literature. Published peer-reviewed literature, dissertations, and working papers focused on water sources for domestic use were included in the review. Titles, abstracts, and keywords were searched for the specified search terms. Appendix A provides the exact search terms. Inclusion criteria included a focus on domestic water source choice and choice modeling based on a variety of water source attributes. Studies on water used for agriculture were excluded from the review. Literature focused on assessing water demand was not included (see Nauges & Whittington, 2009 for a detailed review of this literature).

Information sources for the review included four databases with varying temporal coverage: Web of Science (1955-present), EconLit (1969-present), Academic Search Complete (1975-present), and Scopus (1996-present). There were no geographic, year, or language restrictions on the search. One study selected for the review was not written in French and was translated to English for review using Google Translate. All searches were conducted on March 20, 2014.

2.2 Identified studies

The primary search identified eighty-five studies focused on water source choice after duplicates were removed. Out of the four databases, the Web of Science database identified the most articles (n=42). Based on a review of the titles and abstracts, twenty-one articles on source choice met the inclusion criteria for methods (choice modeling) and topic (water source).

A full review of the twenty-one source choice studies resulted in the selection of fourteen articles. An additional seven studies were identified from the bibliographies of the selected articles, yielding a total of twenty-one studies selected. Figure 1 illustrates the study identification process.

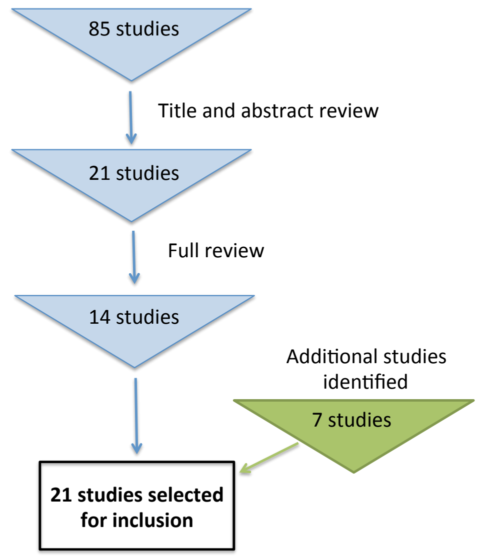


Figure 1. Number of studies at each phase of the screening process

Table 1 provides characteristics of the twenty-one studies included in the review organized by developed or developing country and method (choice experiment, revealed preference, or other).

Table 1. Characteristics of studies included in the systematic review

Study	Ref	Location	Rural or Urban	Piped or unpiped	Sample Size
Developed countries					
Choice experiment					
Blamey et al. (1999)	1	Australia	Urban	Piped	294
Haider & Rasid (2002)	2	Canada	Urban	Piped	100
Scarpa et al. (2012)	3	Italy	Urban	Piped	576
Developing countries					
Choice experiment					
Abramson et al. (2011)	4	Zambia	Rural	Unpiped	403
Anand et al. (2001)	5	India	Urban	Both	148
Dutta & Verma (2009)	6	India	Urban	Both	1100
Echenique & Seshagiri (2009)	7	India	Urban	Both	384
Hope (2006)	8	South Africa	Rural	Unpiped	80
Nam and Son (2005)	9	Vietnam	Urban	Both	1,872
Snowball (2008)	10	South Africa	Urban	Piped	71
Tarfasa & Brouwer (2013)	11	Ethiopia	Urban	Both	145
Revealed Preference					
Asthana (1997)	12	India	Rural	Both	490
Boone et al. (2011)	13	Madagascar	Both	Both	2190
Briand et al. (2009)	14	Senegal	Urban	Both	301
Briscoe et al. (1981)	15	Bangladesh	Rural	Unpiped	180
Kremer et al. (2011)	16	Kenya	Rural	Unpiped	1354
Madanat & Humplick (1993)	17	Pakistan	Urban	Piped	588
Mu et al. (1990)	18	Kenya	Rural	Unpiped	69
Persson (2002)	19	Philippines	Rural	Both	769
Other					
Thompson et al. (2001)	20	Kenya, Tanzania, Uganda	Both	Both	1015
White et al. (1972)	21	Kenya, Tanzania, Uganda	Both	Both	723

2.3 Developed country results

Three of the identified studies meeting the inclusion criteria were from developed countries, all in urban areas, focused on piped water supplies, and using choice experiment methods (Refs 1-3). Scarpa et al. (2012) studied preference heterogeneity within Italian couples with regards to the attributes of taste, smell, cost, turbidity, and staining of pipes. The authors found only small preference differences between men and women within couples and concluded that the gender of the household member interviewed should not greatly affect willingness to pay estimates. Studies of source choice often rely on an interview with only one member of a household to represent preferences of the household unit, so this finding is reassuring. The other two developed country studies, conducted in Australia and Canada, assessed preferences regarding new municipal water supply sources. Blamey et al. (1999) found that Australian respondents supported water recycling for outdoor use but not for indoor uses. In Canada, Haider & Rasid (2002) found cost to not be a significant predictor of source choice; instead respondents were more concerned about improved taste and pressure.

2.4 Developing country results

The remaining eighteen articles selected for the review focus on water source choice in developing regions. Slightly over half of the studies were conducted in Africa (Ethiopia, Kenya, Madagascar, Senegal, South Africa, Tanzania, and Uganda) and the remaining studies were conducted in Asia (Bangladesh, India, Pakistan, the Philippines, and Vietnam). Seven were conducted in rural areas, eight in urban areas, and three included respondents in both rural and urban areas. Over half of the studies (n=11) included respondents with piped water connections as well as respondents without piped connections (or “unpiped households”). Of these studies, eight employed choice experiment methods based on stated preferences and eight used revealed preference methods to inform the choice modeling. The two *Drawers of Water* books (Refs 20-21) relied on data from in-person interviews to calculate an attractiveness index for water sources and are categorized as using an “other” method. Two studies compared different methods: Nam & Son (2005) used both choice modeling and contingent valuation methods in Vietnam and Kremer et al. (2011) compared revealed preferences with stated preference contingent valuation in Kenya.

Table 2 presents the modeling method used, a summary of the findings, and strengths and weaknesses of the research design for each selected study from a developing country. The *Drawers of Water* studies and another early study did not use discrete choice models and instead used pairwise comparisons or indexes based on respondent rankings (Refs 15, 20, 21). The most common modeling method is the conditional logit model (n=10). Conditional logit models rely on the assumption of Independent Irrelevant Alternatives (also known as IIA, which is discussed in more detail in the theoretical framework). Half of the studies using conditional logit models do not check whether the assumption of IIA is valid; the other half use another type of model in addition to conditional logit that relaxes this assumption of IIA, such as a nested logit model. A weakness of all but one of the studies is that they are cross-sectional and source choices may be correlated with unobserved characteristics. Kremer et al. (2011) avoids this problem through randomization of natural spring protection, however, self-recall of travel time may have biased the results. The hypothetical nature of the choice experiment design can be problematic. One of the choice experiment studies reported that ordering of the choice tasks seemed to influence results (Ref 7) and another found that respondents' preference parameters varied during the choice sequence (Ref 11). Two of the most recent studies employed mixed logit models, which are not based upon the assumption of IIA and allow for preference heterogeneity of respondents (Refs 11, 16).

Table 2. Modeling method, findings (source attributes and household characteristics), and strengths and weakness of research design and modeling for selected studies.
Abbreviations used in the table: HH= household, IIA = irrelevance of independent alternatives. Models: MNL = multinomial logit, CL = conditional logit, HEV = heteroscedastistic extreme value, NL = nested logit, CM = choice modeling, MXL = mixed logit

Study	Ref	Model	Source attributes	HH characteristics	Design strengths	Design weaknesses
Stated choice experiments						
Abramson et al. (2011)	4	CL	price (-) quality (+) time (-)	*Sig:* income *not Sig:* education, HH size, gender	Compared WTP with willingness to borrow and to work	Hypothetical options; Recommend a revealed preference study; Did not check IIA
Anand et al. (2001)	5	CL, NL	price (-) convenience	*Sig:* current source; satisfaction with current source	Nested logit relaxes IIA assumption	Found that nested models not appropriate; Other modeling challenges
Dutta & Verma (2009)	6	CL, NL	price (-) quality (+)	*Sig:* HH size, consumption, education	Nested logit relaxes IIA assumption	Parametric assumptions; Confounding from unobserved factors
Echenique & Seshagiri (2009)	7	CL	price (-) quality (+) availability (+) pressure (+)	*Sig*: income, HH size, literacy, toilet ownership *not Sig:* education	Estimated WTP for specific attributes	Assump. that valuation is sum of attributes; Ordering of choice tasks influenced results
Hope (2006)	8	MNL, latent class	convenience (+) quantity over quality	--	Latent class modeling relaxes IIA assumption; Fewer assumptions about data	Did not look at HH characteristics; Specific to the poor, rural study sites
Nam and Son (2005)	9	CL	price (-) quality (+) pressure (+)	*Sig:* income *not Sig*: age, gender	Compared CV and CM; Estimate WTP for specific attributes	Did not check assumption of IIA
Snowball (2008)	10	CL, HEV	price (-) quality (+) discolor (-) interruptions (-)	*Sig*: HH experienced service disruptions	Also used HEV which allows the distribution of error terms to vary across attributes	Small sample (n=71); Only wealthier HHs
Tarfasa & Brouwer (2013)	11	MXL	price (-) quality (+) quantity (+) reliability (+)	*Sig*: income, gender *not Sig:* education, age, HH composition	MXL allows for heterogeneous preferences	Estimated preference parameters vary during choice sequence

Table 2, continued

Study	Ref	Model	Source attr.	HH characteristics	Strengths of design	Weaknesses design
Revealed preference studies						
Asthana (1997)	12	CL	price (-) distance (-)	*Sig (for safe source):* % female, female edu; *Sig (for yard tap)*: income, HH size *not Sig:* male education	Compared HHs with and without piped water relatively large sample size	IIA assump. not checked; No preference heterogeneity with CL
Boone et al. (2011)	13	CL	distance (-)	*Sig:* education, HH asset ownership	Large sample size Both rural and urban population	IIA assump. not checked; No preference heterogeneity with CL; No data on quality or quantity
Briand et al. (2009)	14	probit	price (-) quality (+)	*Sig:* wealth, widow is head of house, HH size, sources available to HH, literacy	Took advantage of policy aimed to extend service to poor	Only studied private connection vs public standpipe decision
Briscoe et al. (1981)	15	pairwise compar-isons	quality (+) *for poor HH:* distance (-), conflict (-)	*Sig:* wealth *didn't look at other HH characteristics*	Early study in applying consumer theory to water source choice	No discrete choice model used; Assumed same preferences for wealth groups
Kremer et al. (2011)	16	CL, MXL	quality (+) distance (-)	*Sig*: latrine ownership, mothers education *not Sig:* diarrhea rates, asset ownership	Spring protection randomly assigned; Compared revealed pref with CV	Possible recall error for self-reported travel times; Focus on one type of source (natural springs)
Madanat & Humplick (1993)	17	MNL	*for drinking:* quality (+) *for bathing*: pressure (+), reliability (+)	*Sig*: male education, ownership of storage facilities *not Sig*: HH size	Choice analyzed by water use type; Also analyzed connection decisions	IIA assump. not checked; No preference heterogeneity; Sequential estimation biased sig upwards
Mu et al. (1990)	18	MNL	price (-) time (-)	*Sig:*% female *not Sig:* income, education	Developed discrete choice model for water; Compared to traditional water demand model	Attributes are all source-specific; Small sample (n=69); IIA assump. not checked
Persson (2002)	19	CL, NL	price (-) distance (-)	*Sig*: HH size *not Sig:* income	Nested logit relaxes IIA assumption; Checked for consistency with utility maximization	Not enough data to estimate the 'full' nested model; No pref heterogeneity
Other						
Thompson et al. (2001)	20	index from ranking	quality (+) convenience (+)	*Sig:* wealth, education, HH size, urban	Longitudinal study compared to *White et al.* from 30 years earlier	No discrete choice model used, instead semi-structured interviews
White et al. (1972)	21	index from ranking	price (-) quality (+)	*Sig:* wealth, HH size	First study of water source choice in developing countries; Three countries studied	No discrete choice model used, instead semi-structured interviews

Source attributes

All of the articles that looked at price found households to prefer water sources with lower prices. Quality of the water was also an important attribute to both piped and unpiped households (Refs 4, 7, 9-11, 14-16). Water quality was defined in most studies as likelihood of health risk due to microbial contamination. In rural Kenya, Kremer et al. (2011) found that households using multiple sources began collecting a greater fraction of their water from springs when those springs were protected to reduce contamination of the water. The relative importance of quality may depend on wealth or whether the household has a piped connection. In South Africa, Hope (2006) found that households without piped connections valued quantity of water over quality improvements whereas Snowball et al. (2008) found "bacteria count" to be the most important attribute for urban households with piped connections.[1] In Bangladesh, Briscoe et al. (1981) also found quality concerns, in this case including taste, smell, and color, to be a factor in water source choice for wealthier households.

For households without piped water connections, distance to the source is of primary concern. Studies in Kenya, India, the Philippines, and Madagascar found distance or walking time to the source to be the main predictor of household source choice (Refs 12-14, 18, 19). The one study that investigated whether possible conflict with other users had an impact on source choice found it to be important in Bangladesh, particularly for the poor (Ref 15). In South Africa and India, studies have found that many households prioritize the convenience of having a piped connection in their compound (Refs 5, 8).

For households that have a piped connection, reliability is a major concern (Refs 7, 11). Pressure is also important for households with piped connections (Ref 9), but may vary in

[1] Hope (2006) did not specify to respondents what type of quality improvement.

importance by type of water use. In Pakistan, Madanat & Humplick (1993) found reliability and pressure to be the prioritized attributes in sources used for bathing.

Household characteristics

Household income or wealth was an important factor in source choice in eight of the studies (Refs 7, 9, 11, 12, 14, 15, 20, 21). However, three studies did not find income or asset ownership to be associated with household water source choice (Refs 16, 18, 19). The financing mechanism to obtain an improved source may influence whether income is a factor in water source choice. Abramson et al. (2011) found that wealthier households in Zambia preferred to pay for water source improvements in cash, while lower-income households preferred "loan and labor financing". Studies in India, Pakistan, and Senegal found households with piped water systems to have higher incomes (Refs 12, 14, 17). There is a relatively extensive literature on willingness to pay for improved water sources, including piped connections. Abramson et al. (2011) conducted a meta-analysis of twenty-one contingent valuation studies estimating willingness to pay for improved water sources. The main findings of the meta-analysis are that willingness to pay for improved water sources is lower in rural compared to urban areas and cost recovery of rural water service improvements is usually infeasible.

Education level is often positively correlated with income (wealthier people are more educated) and has been found to influence source choice (Refs 6, 13, 14, 16). In Madagascar, Boone et al. (2010) found that more highly educated households were more likely to choose public taps over wells. However, studies in Zambia, India, and Ethiopia found no relationship between education and source choice (Refs 4, 7, 11, 12). Some studies looked at education levels of men and women separately. Two studies found female or mothers' education levels to be statistically significantly associated with water source choice (Refs 12, 16). One study in Pakistan found male education to impact water source choice (Ref 17), but another in India found that male education was not a factor (Ref 12).

In households without a piped connection, studies in India and Kenya have found the fraction of women in the household to influence water source choice (Refs 12, 18). This relationship is generally attributed to the fact that women do the majority of water hauling, so additional female household members provide the household with greater carrying capacity. In Senegal, Briand et al. (2009) found that households led by widows were more likely to be connected to the piped system. The proposed explanation for this finding was that women were interested in the convenience of an at-home piped connection and widows made the household decision about whether to connect. Household size has also been found to be positively associated with the likelihood of obtaining a piped connection (Refs 12, 14), as has a greater proportion of men household members (Refs 12).

2.5 Discussion of the literature

The limited literature on water source choice provides some broad findings regarding how households value water source attributes and the importance of household characteristics to source choice. Consistent with economic theory, households prefer sources that are lower-priced and more convenient to use, including at-home piped connections and closer public sources. The microbial quality of the water, particularly for drinking, is also important, though quality may be a secondary concern for those who walk to their water source. Wealthier households are better able to afford improved water sources, however the literature has conflicting findings regarding the whether income and education are important to water source choice. As women are often responsible for the majority of water collection, the fraction of females within a household may influence source choice in providing greater carrying capacity.

In their review of water demand in developing countries, Nauges & Whittington (2009) write that, “the literature on household water source choice, especially in rural areas, is still in its infancy”. This systematic review confirms that there is much to be learned regarding how households decide which water source to use. A major challenge in the study of water source choice work is that findings are usually location-specific. It is often not appropriate to transfer source choice findings to other areas or draw overarching policy conclusions from a study in one location. For this reason, a meta-analysis of these results would most likely not be appropriate. Instead, studies in the particular location where a policy has been proposed may be necessary.

Revealed preference studies on water source choice can provide detailed data on water practices and preferences. However, this method involves the collection of information on all water sources available to each household, including sources that the household does not use, which can be labor intensive. Stated preference choice experiments provide an alternative method of gathering this information. Respondents can be given multiple choice tasks, allowing researchers to gather more data about respondent preferences, however the hypothetical nature of stated preference methods presents a disadvantage. Methodologically, mixed logit models are just beginning to be used in choice modeling. This method has the advantage over conditional logit in that it relaxes the assumption of IIA and allows for heterogeneous preferences. If administered carefully, choice experiments can be a relatively simple, cost-effective way of gaining information about water user preferences. Hope (2006) demonstrates how the choice experiment method is a useful tool for evaluating water policy options for poor, rural households in South Africa. However, stated preference methods have their own challenges and best practices should be employed to reduce bias (see Whittington, 2010). Additionally, without randomization of water source prices or other attributes, cross-sectional studies can be critiqued as not accounting for potential bias due to unobservable attributes associated with household choice (Null et al., 2012). Next, I present the theoretical framework for the study.

3. Theoretical framework

Choice modeling is based on utility maximization, or the theory that consumers maximize their personal utility subject to their budget constraint. In the context of a choice experiment, respondents are expected to select the alternative within the choice set that maximizes their utility. Varying levels of attributes are bundled into alternatives and respondents are assumed to derive utility based upon the attribute levels within each bundle (Lancaster, 1966). The choice set must: (1) have mutually exclusive alternatives; (2) be exhaustive; and (3) have a finite number of alternatives (Train, 2009). The choice set used in this research meets these criteria in that respondents were required to select one preferred water source from a choice set including two hypothetical sources and their current source (the status quo).

The household member responsible for the majority of water decisions was selected as the respondent for the household questionnaire and choice experiment. I thus assume that the respondent interviewed can represent the preferences of the household (as in Mu et al., 1990). Following McFadden (1974), if a household *h* has *J* water sources available to it, the household chooses source *j* if and only if the utility of source *j* is higher than that of source *i:*

$$U_{hj} \geq U_{hi} \qquad \text{assuming } i \neq j \;\&\; i, j \in J \qquad (1)$$

Since the household's complete utility function is unobservable, the utility of household *h* is decomposed into an observed component and a random component:

$$U_{hi} = V_{hi} + \varepsilon_{hi} \qquad i \in J \qquad (2)$$

And the probability that household *h* selects water source *j* can be written:

$$P_{hj} = \text{Prob}(V_{hj} + \varepsilon_{hj} \geq V_{hi} + \varepsilon_{hi}) \qquad \text{for all } i \neq j \ \&\ i, j \in J \qquad (3)$$

If the error term (ε_{hj}) is assumed to be identically, independently distributed extreme value (or Gumbel), the conditional logit model for the probability that household *h* selects water source *j* is:

$$P_{hj} = \frac{e^{(V_{hj})}}{\Sigma_i e^{(V_{hi})}} \qquad (4)$$

The conditional logit model is based upon the assumption that the property of independence from irrelevant alternatives (IIA) is met. IIA is often explained by the red-bus blue-bus problem. When a second bus of a different color is added to a transportation choice set, the probability that a decision maker selects one of the other alternatives should not change, assuming that the buses are identical other than in color. However, in this case, the conditional logit model will overestimate the probability that the decision maker will select one of the buses and under-estimate the probability of selecting a non-bus transportation option. This red-bus blue-bus problem illustrates a scenario in which IIA is not a valid assumption.

The Hausman specification test can be used to determine whether the conditional logit model meets the assumption of IIA. The null hypothesis in this test is that there is no systematic difference between the coefficients estimated in a model including all of the alternatives available compared to a model excluding one of the available alternatives (Hausman, 1978). If IIA is not met, a mixed logit model can be used. In addition to relaxing the assumption of IIA, the mixed logit model also allows respondents to have heterogeneous preferences. As given by Train (2009), if utility in β is linear, then $V_{hi} = \beta' x_{hi}$ and the choice probability for mixed logit can be written:

$$P_{hj} = \int \left(\frac{\exp(\beta' x_{hj})}{\Sigma_i \exp(\beta' x_{hi})}\right) f(\beta)\, d\beta \qquad (5)$$

This represents a weighted average of the logit form based on the weights in f(β). The coefficients can vary, which permits preference heterogeneity.

Returning to the observable utility, V_{hi}, we consider it to be a function of water source attributes, represented by the vector X_{hi}, as well as household characteristics, represented by the vector Z_h. Assuming that the utility function is additive in terms of source and household characteristics, the observable utility V_{hi} can be written:

$$V_{hi} = \gamma X_{hi} + \alpha_j Z_h \qquad (6)$$

where γ represents the coefficients on the water source attributes (constant across all sources and households) and α_j represents source-specific coefficients on the household characteristics. In this work, the main water source attributes of interest (X_{hi}) are price and time. Other source attributes such as health risk, likelihood of conflict, and reliability also likely affect household water source choice. However, the choice experiment was meant to be simple for enumerators to administer and for respondents to answer, so price and time were the only water source attributes varied for the hypothetical choices. Additional attributes are modeled to the extent possible given that these attributes varied only across households' current sources. The household attributes (Z_h) modeled are monthly income, whether the respondent has at least a primary education, and the proportion of women in the household.

Table 3 provides the variables and expected signs on model coefficients. Based upon findings in the source choice literature, I expect cheaper sources and those with lower collection times to be preferred. I also expect a negative coefficient on health risk, as I predict that households will prefer sources that are high quality. The coefficient on reliability is hypothesized to be positive, as I expect households to prefer more reliable water sources. I expect to find that wealthier and more educated respondents are less price sensitive. Mu et al. (1990) found that households with a greater proportion of women were more likely to select sources associated with greater amount of collection time in Ukunda, Kenya. I thus expect the number of women in the house to be positively associated with the selection of water sources with longer collection time.

Table 3. Independent variables in the models, expected sign, and explanation

Variable	Expected sign	Explanation
Price	-	Lower priced sources will be preferred
Time	-	Lower collection time will be preferred
Health risk	-	No significant risk to health will be preferred
Reliability	+	More reliable sources will be preferred
Monthly income x price	+	Households with higher monthly income will be more likely to select higher priced sources
Primary education x price	+	Respondents with at least a primary education will be more likely to select higher priced sources
Proportion of women x time	+	Households with a greater proportion of women will be more likely to select sources with longer collection time

Given that utility is assumed to be additive based upon source attributes, I expect households to reject sources that have a clearly "worse" bundle of attributes compared to another alternative in the choice set. I refer to "preference inconsistency" as instances in which a household selects a source that appears worse based upon observable attributes. This may be due to attributes of a households' current source that we have not observed (such as smell or pressure), bias (status-quo or refusal to participate in the experiment), or lack of understanding of the experiment. Because this is a cross-sectional study, it is possible that unobservable household or source characteristics may affect household water source choice and bias the results (Null et al., 2012). This is a problem for most of studies on water source choice because experimental methods that randomize variations in water source attributes require more complex and expensive fieldwork.

4. Research design

4.1 Fieldwork

Fieldwork was conducted in the areas (most commonly referred to as "sub-locations") of Kianjai, Mutionjuri, Machaku, and Nairiri in Meru County, Kenya. Approximately 140 miles north east of Nairobi, the Kenyan capital, Meru County borders Mount Kenya national park. The elevation is approximately 5,000 feet and average annual temperatures range from 62-69°F. Considered one of the most fertile parts of Kenya, this agricultural area produces staple crops, such as wheat, potatoes, and maize, as well as cash crops including tea, coffee, and bananas. Rice is sold for 85 Ksh (~1 USD)[2] per kilogram and the price of maize is 30 Ksh (~ 0.35 USD) per kilogram (or 2.2 pounds). Average annual rainfall is fifty-four inches and there are a variety of surface and ground water sources. Local government and non-governmental organizations, such as the Red Cross, are currently expanding access to piped systems and public taps, making this area a particularly interesting site to study rural water source choice.

These four sub-locations were selected based on site visits that found a diverse array of water source options. The household questionnaire was informed by discussions with local residents and piloted in areas neighboring the study sites. See Appendix B for photographs of the study region taken during site selection and piloting. Piloting of the survey instrument aimed to identify potential problems with the survey or flag confusing sections to be clarified. The final questionnaire included sections on water sources used by respondents, other water sources available in the area, sanitation options, demographics, income sources, and the choice experiment. The questionnaire was asked of the household member mostly responsible for water-

[2] Based on an exchange rate of 86 Ksh/USD, which is used throughout the paper.

related decisions so that the respondents would be knowledgeable about the water sources available and used by the household.

During August 2013, a team of six enumerators and two field managers completed 388 household surveys. The household interviews were conducted in Meru, the local language. Pairs of enumerators were matched with a village elder within each sub-location for help finding respondent households. Sampling was cross-sectional based on transect walks of each area. If no one was available in a selected house, callbacks were scheduled. Enumerators made three attempts to interview selected households, after which the household was replaced with the next closest household. Table 4 provides the number of households interviewed per sub-location and the total households in the sub-location. Unfortunately, information on how many households were replaced was not collected, nor was the number of households that refused to participate in the study.

Table 4. Interviews conducted and total number of households in each sub-location

Sub-location	Interviews conducted	Total Households
Kianjai	141	1091
Mutionjuri	129	992
Machaku	44	341
Nairiri	74	581

The average interview lasted just over one hour. Data entry from the paper surveys was completed in Kenya using EpiData 3.1 (The EpiData Association, Odense, Denmark) and then exported into Stata11 (StataCorp LP, College Station, Texas) for analysis. Appendix C provides the full household survey with univariate statistics for each of the questions asked.

4.2 Choice experiment design

The following hypothetical scenario was read to each respondent to introduce the choice experiment:

> *Now I would like you to imagine that a group is planning to install several new water points in your area to improve your access to water. The group could be the government or it could be a non-governmental organization. These water points could be boreholes or public standpipes from the piped network. If they install only a few water points, people might have to walk further and wait longer to collect water. If they install more, people might walk shorter distances and have to wait less. Installing these water points is expensive, however. Suppose <the group> will need to charge people who use the water points to recover their costs and properly maintain the water points. If they install more points, they may need to charge more per jerrican.*
>
> *You just told me that the primary source for most purposes right now was <primary source from previous question>. In addition to that source, I want you to imagine you have two new water points available for you to use. You should assume that quality of the water from the new water point is excellent and safe for drinking. You should also assume that the reliability of the new water point would be excellent: it would always have good pressure and you could collect from it whenever it is convenient for you. Finally, you should assume that using the source would not cause any conflict with other water users.*
>
> *The two new water points differ only in the cost you would have to pay per jerrican, and the total amount of time it would take you to walk to the source, wait, fill your container and return. Here is the first task I would like you to think about.*

The enumerator then showed the respondent the choice task card, explained the attributes associated with each hypothetical new water point, and asked if the respondent had any questions. An example choice task card is shown in Figure 2.

TASK 99

	New water point A	**New water point B**	**Your current source**
Total time to walk to source, wait, fill container and return home	**10 minutes**	**5 minutes**	
Cost per 20L jerrican	**1 Ksh** per 20L jerrican	**0.25 Ksh** per 20L jerrican	

Figure 2. Example choice card translated into English

Finally, the enumerator asked:

If these three sources were available to you right now, which source would you ***most prefer to use****? Remember that the two new sources have excellent quality, reliability, and using them would not cause conflict. Which source would you* ***least prefer to use****?*

The enumerator marked on the questionnaire which of the three sources the respondent most preferred and which source was least preferred. The baseline situation was the status quo. A hypothetical baseline was avoided given the difficulties associated with administering these surveys well and comprehension challenges for respondents (Whittington, 2010). The term "preferred source to use" was intended to mean the source that the respondent would use exclusively as their primary water source and this is how the question was posed to respondents in the Meru language.[3]

Each household completed four choice tasks: a "block" of three tasks presented in random order and one task answered by all households. The experiment was based on a full factorial design of two three-level attributes: price of 0.25, 1, or 3 Ksh and total water collection time of 5, 10, or 30 minutes. These attribute levels were chosen to be close but slightly lower than average current source prices and collection times so that they would be tempting to respondents. Because the choice experiment was not the focus of the survey effort, unfortunately these attribute levels were not pre-tested.[4] From the full factorial, obvious choices in which one source dominated the other source with regards to both price and time were eliminated. For example, if Source A had a price of 1 Ksh and a time of 5 minutes and Source B was priced at 3 Ksh and took 30 minutes, Source A was considered to dominate Source B. We then eliminated symmetric duplicates to yield the final nine choice tasks, which were divided equally into the three "blocks". In addition to a randomly selected block, all respondents were presented with a

---

[3] The Meru translation of "If these three sources were available to you right now, which source would you most prefer to use?" is: *"Guntu kuu kuthatu gwa gutaa ruuji gukeethirwa gukionekana kirigwe onendi, ni kuriku ugichaalaa kuruki ya kungi?"*

[4] As a result, the hypothetical times chosen (5, 10, 30 minutes) were less than half of average current source collection times which were all over 60 minutes for households without at-home sources.

task including one source with the lowest time and lowest price and another source with the middle time and middle price (as shown in Figure 2 above.) With one of the two hypothetical sources dominating the other in both time and price, this task served as a simple comprehension check for the choice experiment. The task was also intended to determine whether the most attractive hypothetical source might tempt households with at-home sources, due to the high reliability and excellent water quality of the hypothetical sources. The four choice tasks were presented to respondents in random order. Table 5 provides the ten choice tasks and associated attribute levels.

Table 5. Full choice experiment design

task ID	Source A		Source B	
	Price (Ksh)	**Time (min)**	**Price (Ksh)**	**Time (min)**
11	0.25	10	1	5
12	0.25	30	1	10
13	3	5	1	30
21	0.25	10	3	5
22	0.25	30	3	5
23	3	10	0.25	30
31	0.25	30	1	5
32	3	5	1	10
33	3	10	1	30
99	1	10	0.25	5

5. Results

5.1 Respondent and household characteristics

Table 6 provides average respondent and household characteristics by sub-location. A typical sample household is Catholic and has five members, approximately half male and half female. The household is led by a married couple, both of whom are around forty years old and have each completed seven years of education. They own their house and one acre of land. The household has a private pit latrine, but does not have electricity. Kerosene is used for lighting and firewood is used for cooking and heating. There are two rooms in the main house and three other buildings in the compound. Monthly household income from all sources is approximately 35,000 Ksh or 407 USD. The most common sources of income are farming and full-time wage labor. Average food expenditure is 430 Ksh (5 USD) per household member per week or a total of 14,924 Ksh (174 USD) per month. Household assets include a cell phone, bicycle, and radio. In terms of livestock, the household has four chickens, two goats, and two cows. A larger proportion of respondents have an at-home piped water connection (29%) than have electricity (11%).

Table 6. Socioeconomic and demographic characteristics of sample households by sub-location (mean with standard deviation below in parentheses or percentage of the sample)

Household or respondent characteristic	Kianjai n=141	Mutionjuri n=129	Machaku n=44	Nairiri n=74
Household size	5.3 (2.3)	5.6 (2.2)	4.8 (1.6)	6 .0 (2.2)
Years of respondent education	7.6 (3.9)	7.9 (3.8)	6.75 (3.5)	7.1 (43.9)
Monthly Income (Ksh)	44,598 (36,174)	63,572 (114,847)	35,143 (28,830)	54,451 (73,319)
Weekly per capita food expenditures	554 (414)	378 (223)	323 (238)	376 (235)
Acres of land owned	2.0 (2.3)	2.1 (2.8)	1.8 (1.0)	2.0 (2.3)
% of respondents with job for wages	26%	22%	5%	5%
% of households owning livestock	88%	90%	95%	96%

Ninety-seven percent of respondents have a sanitation facility on their compound and, of these, 13% share this facility with others outside of their household. Nearly all households use a pit latrine. Only two respondents reported to have a water-sealed flush toilet, which were reported to cost between 11,000-30,000 Ksh (128-349 USD). About half of respondents' pit latrines have a slab and about a quarter are ventilated. Three quarters of respondents are somewhat or very satisfied with their pit latrine and the most frequent complaint is the smell.

5.2 Water sources

The most commonly used primary sources were at-home shallow wells (20%) and neighbor's shallow wells (20%). At-home piped connections (15%), public wells (12%), and neighbors' boreholes (11%) were also common primary water sources. The WHO/UNICEF Joint Monitoring Program (JMP) defines "improved" water sources to be at-home piped connections, public taps, boreholes, protected wells, and rainwater. Based on this definition, about half of sample respondents used an "improved" primary source. Unprotected hand-dug wells, water purchased from vendors, and surface water sources are considered “unimproved” sources. Fieldwork was conducted in the dry season so rainwater was generally not available. Ninety five percent of households reported that they use the same primary source during the dry season for drinking, washing in the home, bathing, cooking, watering outside the home, and other productive uses. Women are primarily in charge of water collection; in over three quarters of households, a woman collected the most water in the last week.

Sixty one percent of households have a primary source outside of their compound (or “not at-home”) including neighbors’ sources and public wells, boreholes, and taps. Approximately one third of households have a primary source located on their compound or “at-home”, which includes shallow wells, piped connections, and rainwater. Households without an at-home source have a lower average monthly income (44,411 Ksh or 516 USD) compared to households that have at-home sources (63,174 Ksh or 735 USD).

The remaining 3% of respondents report that water purchased from vendors is their primary source of water. While vended water is not a common primary source, 80% of households reported that it is possible to purchase water from a vendor who delivers to the home. Two thirds of households had purchased water from a vendor at some time in the past and one fifth reported to have purchased water from a vendor in the last week.

Table 7 presents average dry season collection time and price per 20 liter jerrican for each type of primary source in descending order of frequency. Households purchasing water from vendors could be considered part of either the at-home or not at-home primary source group, but are classified as “not at-home” because they do not have a source located on their compound. Collection time for households using off-compound sources was estimated by summing the waiting time (queuing and filling) and twice the one-way walk time. For households purchasing vended water or collecting rainwater, collection time was assumed to be zero. Collection time for households with at-home wells or piped connections was assumed to be two minutes for filling the jerrican. Prices per 20 liter jerrican were given by respondents for the off-compound sources and for water purchased from vendors. For households with at-home wells or using rainwater, the price per 20 liters (L) was assumed to be zero. Average price per 20 L for households with at-home piped connections was estimated based upon their reported previous payment and water use in the last week scaled to the previous payment period (valid based upon the assumption that household use in the last week was typical of use during the previous payment period). Payment systems are generally the same in both the dry and rainy seasons and 67% of all respondents report to pay for their water.

Table 7. Number of households using each type of primary source (*n*), primary source dry season collection time, and price per 20 liter jerrican (including households that do not pay for their water)

Primary source	n	Collection time (min) mean (median)	Price/ 20L jerrican (Ksh) mean (median)
Off-compound			
Neighbor's well	76	86 (66)	1.4 (2)
Public well	47	194 (70)	3.6 (5)
Neighbor's borehole	44	66 (50)	2.4 (2)
Neighbor's piped	26	55 (47)	1.8 (2)
Public borehole	16	103 (70)	2.3 (2)
Vended	12	0* (0)	9.9 (10)
Public tap	8	149 (145)	3.3 (2.5)
Other public sources	7	151 (120)	2.6 (1)
At-home			
Well	78	2* (2)	0* (0)
Piped connection	59	2* (2)	1.2 (0.98)
Rainwater	2	0* (0)	0* (0)

* Values that are assumed rather than based on household questionnaire responses

Households using a neighbor's source spend about half as much time collecting water as those households using public sources. In particular, households using a neighbor's well as their primary source spend an average of 86 minutes in total collection time per trip during the dry season compared to 194 minutes on average for households that gather water from a public well. Figure 3 shows the distribution of collection times for households with primary sources not at-home.

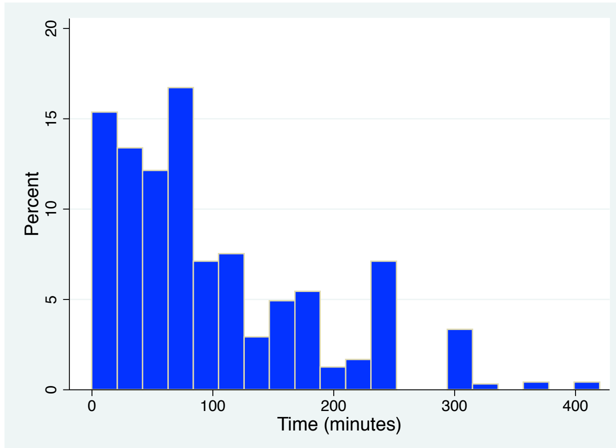


Figure 3. Estimated total water gathering times in minutes from primary sources for households without an at-home source (excluding an outlier of 800 minutes)

The average price for a 20 L jerrican from a neighbor's well is 1.4 Ksh (0.02 USD). Public sources have higher average prices: per 20 L jerrican, the average price for public wells and public taps are both over 3 Ksh. Water purchased from vendors who deliver to the household costs 9.9 Ksh per 20 liter jerrican on average.

Table 8 presents reliability, quality perceptions, and frequency of water treatment for each type of primary source. Reliability is calculated as the average hours per day the source is available across a week, censored at a maximum of 12 hours/day.[5] Vendors are assumed to be available 12 hours/day and rainwater is assumed to be available 0 hours/day given that data was collected during the dry season. At-home wells are assumed to be available all the time, unless the respondent reported that the well did not have water in August, the month that the survey was conducted. Twelve percent of households using at-home wells reported that the well water was not available and reliability was thus coded as 0 hours/day. Respondents were asked about perceived serious health risk for all the sources. Table 8 also presents the percent of respondents using a given primary source that believe that drinking from the source poses a serious risk to

[5] Censoring at 12 hours per day was done to avoid confusion about availability of sources at night time. For example, some households may have interpreted the question such that they reported neighbor's source to be available 24 hours/day while others reported 12 hours/day.

their health. The final column in table 8 presents the percent of households using each primary source who report to treat the water before drinking. Unfortunately, households were not asked about treatment of water from at-home piped connections or purchased from vendors.

Table 8. Primary source dry season reliability (censored at 12 hrs/day), perceived serious health risk from drinking, and water treatment

Primary source	n	Reliability (avg hrs/day) mean (median)	Serious health risk % perceiving serious risk	% treating water before drinking (dry season)
Off-compound				
Neighbor's well	76	11.3 (12)	31%	64%
Public well	47	11.4 (12)	21%	67%
Neighbor's borehole	44	10.0 (12)	6%	52%
Neighbor's piped	26	9.0 (10)	0%	64%
Public borehole	16	10.4 (12)	10%	62%
Vended	12	12* (12)	17%	--
Public tap	8	8.7 (10)	50%	86%
Other public sources	7	11.5 (12)	57%	80%
At-home				
Well	78	10.62 (12)	25%	82%
Piped connection	59	5.8 (5)	70%	--
Rainwater	2	0.0* (0)	0%	100%

* Values that are assumed rather than based on household questionnaire responses

At-home piped connections are by far the least reliable, providing water on average less than six hours per day (averaged across the week). Most other sources, with the exception of rainwater, are available close to twelve hours/day on average. Figure 4 shows the distribution of reliability of primary water sources for households without an at-home primary water source. About two thirds of these respondents have a primary source that is available close to at least 12 hours/day on average.

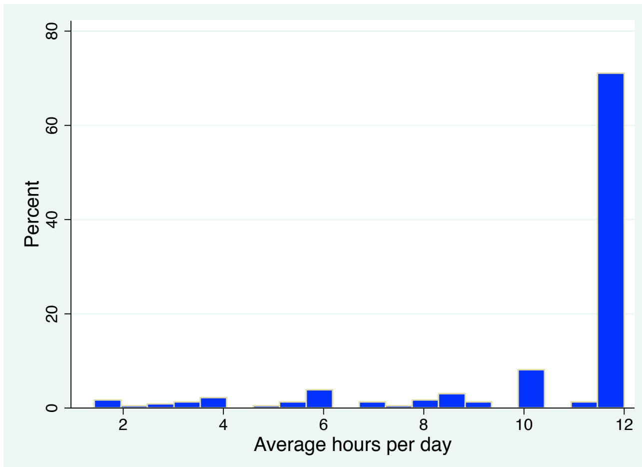


Figure 4. Average hours per day that water is available from primary sources for households without an at-home source (averaged across a week and censored at 12 hours/day)

Perceptions of serious health risk vary considerably across the sources but the majority of respondents report to treating their water before drinking. About a third of respondents using a neighbors' well as their primary source perceive a serious health risk from drinking this water. About half of respondents using public taps or other public water points as their primary source believe that water from these sources poses a serious risk to health. Neither of the households reporting rainwater as their primary source perceive the water to be a health risk but both treat the water with chlorine. Few households perceive the public and neighbors' boreholes to pose a serious health risk. It is interesting that none of the households using a neighbor's piped system perceived a serious health risk from drinking the water while 70% of households using their own piped connection perceive a health risk from the water.

Table 9 gives the frequency of different methods of water treatment by monthly income groups: low (median 15,500 Ksh or 180 USD), middle (median 35,000 Ksh or 407 USD), and high (median 69,640 Ksh or 810 USD). Unfortunately the questionnaire did not include questions about treatment of at-home piped connections or water purchased from vendors, so this table excludes the seventy-one respondents who used this type of primary source. Treatment rates are relatively similar across these three income groups, though middle- and high-income households are more than three times as likely to both boil their water and add chlorine compared to the low-income group.

Table 9. Fraction of respondents by income group using water treatment methods

Treatment type (n=303)	Monthly Income Classification		
	low	middle	high
no treatment	33%	31%	31%
chlorine	11%	11%	12%
boil	50%	43%	46%
stand and settle	2%	3%	0%
boil and add chlorine	3%	10%	11%

Table 10 presents the volume of water collected in the last week from each primary source, as well as the average fraction of water used from that primary source, relative to water used from all sources in the last week. For households using at-home wells and piped connections as a primary source, over 90% of the water used was from their primary source on average. Households using public boreholes and other public sources obtained about half of their water from other sources.

Table 10. Volume of water (L) collected in the last week from primary sources and as a percent of all water used, by primary source type

Primary source	n	L collected in last week from primary source mean (median)	Average percent of water used in last week collected from primary source
Off-compound			
Neighbor's well	76	879 (840)	73%
Public well	47	763 (700)	74%
Neighbor's borehole	44	686 (570)	69%
Neighbor's piped	26	740 (750)	83%
Public borehole	16	774 (700)	53%
Vended	12	795 (700)	82%
Public tap	8	703 (570)	68%
Other public	7	497 (480)	48%
At-home			
Well	78	2983 (1810)	96%
Piped connection	59	1422 (1120)	95%
Rainwater	2	0 (0)	0%

Table 11 presents the number of water collection trips made by households in the last week, excluding households that did not walk to collect water. Some households with at-home sources or that purchased vended water collected water from additional sources outside the home. These households with at-home wells or piped connections using other sources made many trips per week on average. The percent of trips on foot relative to total trips (including using a bicycle, cart, car, motorbike, or wheelbarrow) on average is also presented.

Table 11. Water collection trips made in the last week, and fraction of these trips made on foot

Primary source	n	Total trips to collect water (HHs making >0 trips/wk) mean (median)	Average percent of trips that are walking
Off-compound			
Neighbor's well	71	40 (35)	68%
Public well	42	30 (21)	55%
Neighbor's borehole	42	28 (26)	53%
Neighbor's piped	26	37 (35)	82%
Public borehole	27	35 (35)	55%
Vended	6	9 (6.5)	98%
Public tap	7	22 (25)	55%
Other public	5	29 (28)	40%
At-home			
Well	8	31 (36)	69%
Piped connection	10	20.5 (8.5)	62%
Rainwater	1	6 (6)	0

Sample households report to have used a median of two water sources in the last seven days. Table 12 shows the percent of households with each primary source that have used at least one other source in the last week. Households with at-home wells are the least likely to have used an additional source in the last week, which makes sense because at-home wells are generally convenient, reliable, and have no per jerrican price. Both of the households that reported to use rainwater as their primary source used alternate sources in the last week as we would expect in the dry season. The majority of households purchasing vended water as their primary source report to have used at least one alternate source in the last week.

Table 12. Percent of households using at least one other source in the last week

Primary source	n	% using at least one other source in the last week
Not at-home		
Neighbor's well	76	39%
Neighbor's borehole	44	45%
Neighbor's piped	26	27%
Public water point	91	51%
Vended	12	58%
At-home		
Private piped	59	24%
Private well	78	14%
Rainwater	2	100%

Households were also asked what back-up source they would use if their primary source was unavailable. Table 13 provides the frequency of back-up source types reported by households with the three most common primary sources (neighbor's well, well at-home, piped connection at-home).[6] A public borehole or neighbor's borehole are the most common back-up sources for households using a neighbor's well as their primary source. For households with wells at home, a neighbor's well is the back-up source for a majority of households. The most common back-up source for households with a piped connection at home is a public well (32%) followed by a neighbor's well (23%) or public borehole (15%).

[6] The question about back-up sources was phrased as follows: "Now I want you to imagine that your primary water source was unavailable for some reason. Which source would you use as a back-up for each of the following purposes?"

Table 13. Back-up sources for three most common primary sources

	Primary source		
Secondary source	Neighbor's well (n=73)	Well at home (n=73)	Piped at home (n=53)
Bottled water	0%	1%	0%
Neighbor's borehole	16%	10%	4%
Neighbor's piped	11%	4%	8%
Neighbor's well	11%	59%	23%
Other public	14%	4%	2%
Public Borehole	18%	7%	15%
Private well	0%	0%	6%
Public tap	0%	0%	8%
Public well	1%	3%	32%
Rainwater	7%	3%	2%
Surface water	1%	0%	2%
Vended	21%	10%	9%

5.3 Choice experiment and validity checks

A total of 1,550 choice tasks were completed by 382 different respondents. Households with at-home sources (piped connections, wells, and rainwater) are less likely to be interested in one of the hypothetical sources given the convenience, short collection time, and low price associated with at-home water sources. However, the majority of respondents using an at-home piped connection as their primary source perceive the water to be low quality and unreliable; one quarter of households using at-home wells perceive the water to be a serious health risk. It is possible that the reliable and high quality hypothetical sources might tempt these households with at-home current sources. Table 14 shows the percentage of respondents without an at-home water source that always chose their primary source, the cheapest source, or the source with the lowest time, as well as the percent always rejecting their primary source, the most expensive source, or the source with the highest time, in every choice task completed. Table 15 presents these statistics for households that use at-home piped connections, wells, or rainwater as their primary source.

Table 14. Respondents without at-home sources that always selected or rejected certain types of sources in all choice tasks*

	Kianjai n=76	Mutionjuri n=106	Machaku n=11	Nairiri n=56	All n=249
Always **chose** current source (status quo)	20%	5%	0%	0%	8%
Always **chose** cheapest source	28%	12%	20%	16%	18%
Always **chose** lowest time source	3%	2%	0%	0%	2%
Always **rejected** current source (status quo)	48%	32%	20%	76%	46%
Always **rejected** most expensive source	15%	8%	0%	44%	18%
Always **rejected** highest time source	43%	32%	20%	75	44%

*note: data on least preferred source were missing for some households without at-home sources. For the "always rejected" sample sizes are: Kianjai n=75, Mutionjuri n=104, Machaku n=10, and Nairiri n=55.

Table 15. Respondents with at-home sources that always selected or rejected certain types of sources in all choice tasks

	Kianjai n=65	Mutionjuri n=22	Machaku n=33	Nairiri n=18	All n=138
Always **chose** current source (status quo)	75%	50%	30%	28%	54%
Always **chose** cheapest source	72%	55%	21%	6%	49%
Always **chose** lowest time source	74%	50%	33%	33%	55%
Always **rejected** current source (status quo)	8%	14%	0%	22%	9%
Always **rejected** most expensive source	11%	5%	3%	6%	7%
Always **rejected** highest time source	43%	32%	21%	0%	30%

Only 8% of households without an at-home source selected their current source in every choice task compared to 54% of households with an at-home source. It is surprising that 9% of households with at-home sources always ranked their current source last; about half of these households had at-home wells and about half had at-home piped connections. Because it makes the most sense for households without at-home sources (including those whose primary source is water purchased from vendors) to switch to a hypothetical source, subsequent analysis focuses on this group of households. Excluding households with at-home sources (wells, piped connections, or rainwater) as their primary source results in the analysis of 973 choice tasks completed by 244 different respondents.

Inconsistent preferences

Three types of inconsistent preferences were possible. First, respondents could select the dominated hypothetical source in the choice task completed by all respondents. Twelve respondents selected this dominated source, which suggests that these respondents may have not understood the choice experiment.

Second, respondents could select their current source even though one of the hypothetical sources appears to have a "better" bundle of attributes. In 722 of the 973 total choice tasks at least one of the hypothetical sources dominated the respondent's current source based upon the attributes studied. In 7% of these tasks (forty-nine tasks representing twenty-six different households) the respondent selected their current source even though the hypothetical option appeared to be "better". In three-quarters of these "preference inconsistencies", the respondent's current source was a neighbors' source. In 417 choice tasks both of the hypothetical sources appeared to dominate the current source. In 5% of these instances (nineteen choice tasks, representing eleven different households) respondents selected their current source. Although these preference choices appear inconsistent within the choice experiment, the respondent's choice could be due to status quo bias or some other characteristic of his or her current source that we did not observe.

Third, respondents could prefer a hypothetical source even though their current source appeared to dominate this hypothetical source. However, none of the households with primary sources outside the home had a source that dominated both hypothetical sources.

Robustness checks

To test the reliability and accuracy of the choice experiment data, I conducted four types of robustness checks: (1) sub-set analysis of only the choice tasks in which there were no preference inconsistencies; (2) analysis of a choice between each hypothetical source and the household's current source based on the ranking data; (3) comparison of the two hypothetical source options based on the ranking data; and (4) sensitivity of the results to assumptions about collection time. Detailed results from the four robustness checks summarized below are provided in Appendix D.

First, I conducted a sub-set analysis excluding "preference inconsistencies", as defined in the previous section. This excluded sixty-one choice tasks: twelve tasks in which the respondent selected the dominated hypothetical source in the choice task completed by all respondents and forty-nine tasks in which at least one of the hypothetical sources appeared to dominate the current source, yet the respondent preferred their current source. The coefficients and statistical significance are very similar to the model that includes these sixty-one "inconsistent preference" tasks. The magnitude of the coefficient on price is slightly smaller and magnitude of the coefficient on time slightly larger compared to the mixed logit model including choice tasks that revealed inconsistent preferences.

Second, I analyzed the data as a choice between each hypothetical source and the household's current source based on the ranking data. Each choice task completed by a respondent led to two observations in this data set, Hypothetical Source A versus current source and Hypothetical Source B versus current source, resulting in a total of eight choices per respondent. Price and collection time attributes were re-coded as the difference between the respondent's current water source and the relevant hypothetical source. Finally, the health risk and reliability of the respondent's current source were also included as independent variables. In this model, the coefficients on price and collection time are negative and statistically significant, as expected. When reliability and health risk are added to the model, neither of the coefficients on these variables is statistically significant and the coefficient on price is less statistically significant. While these logit models do not account for the full structure of the data, it is reassuring that the sign and significance of the price and time variables remain relatively consistent.

Third, I used the ranking data collected to compare relative rankings of the two hypothetical sources. Similar to the logit model used in the previous robustness check, I calculated independent variables of price difference and collection time difference between the two hypothetical sources. Other source attributes (reliability and health risk) were constant across both hypothetical sources and could not be included in the model. This logit model yields coefficients on both price and time that are negative and highly statistically significant, as expected ($p < 0.01$).

Finally, I evaluated the sensitivity of the results to assumptions made about collection time. Respondents without at-home sources were asked how long it took to walk one way with a full container, rather than the round-trip walk time. To make these estimates comparable to the total collection time for the hypothetical sources (round-trip walking, waiting, and filling), I doubled this one-way walk time for the analysis. However, walking with a full container is slower than with an empty container and doubling these times may have resulted in artificially high walk time estimates. To check this assumption, a conservative estimate for walk time assuming that walking with an empty container took half as long as with a full container (so 1.5 times the reported one-way walk time) was estimated. In this model, sign and statistical significance of the coefficients does not change. The coefficient on price is smaller but the coefficient on time is virtually unchanged.

5.4 Variables and coding

Including highly correlated independent variables can be problematic in multinomial regression models,[7] so I examined the correlation coefficients between source attribute variables. Appendix E provides the correlation matrices of these independent variables. The alternative specific constant (asc), equal to one for a respondent's current source, and the variables of reliability and likelihood of conflict are all highly correlated because all of the hypothetical sources (asc=0) were very reliable and associated no chance of conflict. I also find also high correlation between collection time and asc because the hypothetical times presented in the choice experiment were substantially lower than most households' current round-trip collection time. I would expect to see high correlation with health risk, which was assumed to be zero for the hypothetical sources. This correlation is somewhat lower, which can be explained by the low frequency of occurrence relative to the other variables. Because of the particularly high correlation between likelihood of conflict and the other independent variables, it is excluded from the multinomial models.

Table 16 describes the coding of the variables used in subsequent multinomial logit regression models. The hypothetical sources were coded according to the time and price presented on the task card. As given in the hypothetical scenario, the attribute of health risk was

[7] If there is multicollinearity between variables, one of the variables must be dropped to run the model or the matrix is singular. Similarly, highly correlated independent variables can also cause estimation problems.

coded as 0 and reliability as 12 hours/day. Characteristics of the household's primary source were coded according to the respondent's responses in the questionnaire.

Table 16. Description of primary and hypothetical source attribute variables for households without an at-home source

Variable	Description	Hypothetical source coding	Primary Source Coding
Time	Round trip walk time and waiting	5, 10, or 30 minutes	(2*1 way walk time with full container) + (wait time)
Price	Price of 20L jerrican	Ksh 0.25, 1, or 3	0 if doesn't pay otherwise, Ksh/20 L jerrican price
Health Risk	Perceived risk from drinking	0	0 = reported no or some health risk 1=reported serious health risk
Reliability	Average hours/day in a week (censored at 12 hrs/day)	12 hours/day	Average hours/day across the week Vended: assumed 12 hours/day
asc	alternative specific constant	0	1

5.5 Multinomial regression models

As found in the systematic literature review, conditional logit models are used in the analysis of most water source choice studies. However, if the assumption of IIA is rejected based upon Hausman's specification test, mixed logit models are preferred. Estimated using maximum simulated likelihood, mixed logit models generate a distribution of coefficients for the sample based upon respondent choices (Hole, 2007). The shape of these distributions must be specified, leading to the need for additional assumptions.

Source attributes

Table 17 presents conditional and mixed logit models of preferred source for households without at-home sources. The independent variables included in this table are asc, price per 20 L jerrican, and collection time. The conditional logit model is based on the assumption that all respondents have the same preferences and thus each coefficient is constant across the sample. For the two mixed logit models (Models 2-3), the mean and standard deviations of predicted coefficients are provided. Each maximum likelihood simulation used 500 Halton draws.[8]

The conditional logit model (Model 1) yields negative and statistically significant coefficients on collection time and asc. It is unexpected that the coefficient on price is not statistically significant. The negative coefficient on asc indicates that respondents prefer the hypothetical sources to their current source, which may reflect the importance of source attributes other than price or time, such as quality or reliability. However, based on the Hausman test, the assumption of IIA does not hold: I reject the null that the difference in coefficients is not systematic when the current source alternative is dropped from the model ($p<0.0001$). This indicates that conditional logit models are not appropriate for these data, so I move onto mixed logit models.

The mixed logit models, assuming a normal distribution for all variables (Model 2) or a lognormal distribution for price (Model 3), yield negative, statistically significant coefficients on both price and time, as hypothesized. Under the assumption of a normal distribution for all variables (Model 2), the model estimates positive price and/or time coefficients for some of the respondents, but it is implausible that households would prefer higher prices or collection times. Lognormal distributions eliminate the chance of positive coefficients. The mixed logit model with a lognormal distribution for price (Model 3) has coefficients close in magnitude to those of Model 2, but the standard deviation of the price coefficients is an order of magnitude larger.

[8] Halton draws have been found to give more uniform coverage compared to pseudo-random sequences (Hensher, 2005).

Unfortunately, the maximum likelihood function for models assuming a lognormal distribution for time would not converge. [9]

Table 17. Multinomial logit models for selected source

distributions:		normal	price lognormal[10]
	(1)	(2)	(3)
VARIABLES	Conditional logit	Mixed logit	Mixed logit
Price - mean	-0.0964***	-0.882***	-1.003***
	(0.0321)	(0.125)	(0.338)
Price - SD		-1.089***	1.843***
		(0.161)	(0.120)
Time - mean	-0.0230***	-0.182***	-0.225***
	(0.00277)	(0.0218)	(0.0310)
Time - SD		0.113***	0.163***
		(0.0151)	(0.0233)
asc - mean	-0.458***	-2.590*	-2.775*
	(0.150)	(1.381)	(1.605)
asc - SD		10.03***	11.679***
		(1.380)	(2.070)
Observations	2,886	2,870	2,870

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

All of these models rely on assumptions that can be questioned. The conditional logit results (Model 1) rely on the assumption of IIA. The mixed logit with all normal distributions (Model 2) permits respondents to have preferences for higher prices and collection times. The mixed logit model with a lognormal distribution for price still assumes a normal distribution for collection time, although it is unlikely that a person would prefer a greater collection time. Lognormal distributions also assume that there is a long tail, or in this case, large price coefficients for some respondents. Nevertheless, the mixed logit model with a lognormal

[9] The model would not converge for a lognormal distribution for time, despite setting starting values manually, specifying a different stepping algorithm for non-concave regions, or increasing the draws.
[10] The coefficient on price is not very stable. The maximum likelihood function consistently converges at two different log likelihoods with different mean price coefficients (0.75 versus 1.00). Here I present the model with a higher log likelihood value.

distribution for price is based upon the most plausible assumptions, so I proceed with the analysis using this model.

Table 18 and figure 5 show the average marginal effects of increases in price for households without at-home sources.[11] An increase in price from 1 to 3 Ksh per 20 liter jerrican decreases the probability of selection by 11% on average.

Table 18. Average marginal effects of increases in price of 20 L jerrican for households without an at-home source

Price (Ksh)	Change in probability
0-1	-0.09
1-3	-0.11
3-5	-0.04
5-10	-0.04

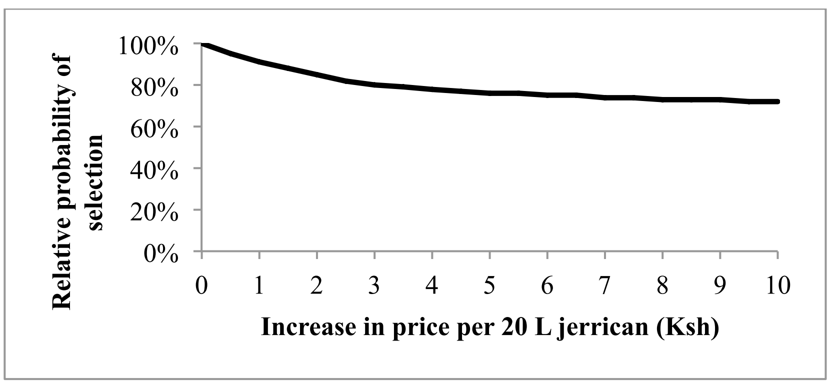


Figure 5. Cumulative change in probability of source selection as price per 20L jerrican increases (with *time* and *asc* variables held constant)

Table 19 and figure 6 show the average marginal effects of increases in collection time for households without at-home sources. An increase in collection time from 20 to 40 minutes decreases the probability of selection by 15% on average.

[11] To calculate marginal effects, I took the difference in probabilities that a source would be chosen at two different attribute levels. Because the choice experiment data set is at the alternative-level, a random alternative within each choice task was selected to obtain the average probability for a given attribute level.

Table 19. Average marginal effects of increases in collection time for households without an at-home source

Time (minutes)	Change in probability
5-10	-0.10
10-20	-0.14
20-40	-0.15
40-60	-0.04

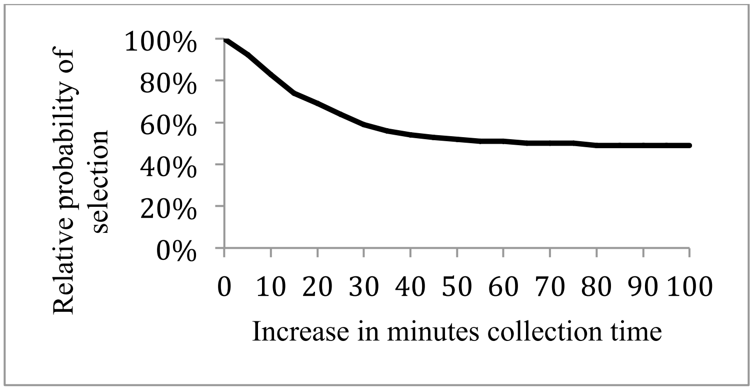


Figure 6. Cumulative change in probability of source selection as collection time increases (with *price* and *asc* variables held constant)

In addition to price and time, I hypothesized that source choice decisions would be influenced by the attributes of health risk and reliability. However, the simple choice experiment only varied the attributes of price and time. Respondents were told to assume no health risk and complete reliability of all hypothetical sources. The variables of quality and reliability are thus fixed for a given household: their current source remains the same and all hypothetical sources are the same for these attributes. To include these variables in a multinomial model, I created interaction terms with the alternative specific constant (asc). Table 20 provides the results of a mixed logit model (with lognormal price distribution) including these asc-health risk and asc-reliability interaction terms for households without at-home sources. Neither of the interaction terms are significant in this mixed logit model, however, the conditional logit model yields a negative and statistically significant coefficient on the asc-health risk interaction term (see Appendix F).

Table 20. Mixed logit model for selected source including source attributes of reliability and health risk interacted with asc, normal distributions except lognormal price distribution

VARIABLES	selected source
Price - mean	-0.752***
	(0.274)
Price - SD	1.905***
	(0.154)
Time - mean	-0.257***
	(0.0353)
Time - SD	0.200***
	(0.0287)
asc - mean	-12.48
	(8.590)
asc - SD	-6.656***
	(1.790)
asc*Reliability - mean	0.819
	(0.902)
asc*Reliability - SD	1.677***
	(0.261)
asc*Health risk - mean	-2.698
	(2.341)
asc*Health risk - SD	4.233***
	(1.502)
Observations	2,399

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

Value of time

The value of time can be estimated by dividing the coefficient on time by the coefficient on price (Jeuland et al., 2010). Based on the parameters estimated in the mixed logit model with a lognormal distribution assumed for price (Table 17, Model 3), respondents value their time on average at 0.15 USD/hour (assuming an exchange rate of Ksh 86/USD). The hourly unskilled wage rate in our study region was reported to be 0.41 USD/hour (based on 300 Ksh/8.5 hour work day). Thus, on average, time spent collecting water is valued at 37% of the unskilled wage rate for the region. This is close to convention in the literature to value time at approximately 50% of the wage rate (Boardman et al., 2011).

Few studies have estimated the value of time in Africa. A study in Mozambique, which used the travel cost method, estimated that respondents value time at 20-50% of the median hourly wage (Jeuland et al., 2010). Two previous studies have estimated the value of time spent collecting water in Kenya. Whittington et al. (1990) found that Kenyans in the town of Ukunda valued time at approximately the unskilled wage rate. In contrast, in rural Kenya, Kremer et al. (2011) estimated time to be valued at 7% of the casual labor wage rate. However, the wage rate assumed in that study was very small (1.26 USD/eight hour day), about a third of the wage rate in our study area. The average value of time estimated for this sample thus falls between these two more extreme estimates of the value of time in Kenya.

Household characteristics

Based on the literature, I hypothesized that wealthier and more educated households would be less sensitive to price. Because household characteristics are fixed for a respondent, each characteristic is interacted with one of the other independent variables that varies over the alternatives. Table 21 presents the results of mixed logit models including each of the hypothesized relevant household characteristics. Neither monthly income level (low, medium, or high, Model 1) nor primary education (Model 2) interactions with price yield a statistically significant coefficient.

Model 3 includes an interaction term between collection time and proportion of total household members who are female and between the ages of sixteen and fifty-seven (the tenth and ninetieth percentile ages for female household members who have collected the most water in the last week). The coefficient on this interaction term is negative and statistically significant, contrary to my hypothesis.[12] This finding suggests that households with a greater proportion of women are more likely to select water sources with lower collection time. A possible explanation is that these households with a greater proportion of women also have greater interest in using a primary water source with a shorter collection time. The studies that found a higher proportion of women to be associated with higher collection time were both revealed preference studies. It is plausible that this finding might not emerge from a stated preference task in which respondents are considering hypothetical source alternatives.

[12] Number of women in the house interacted with collection time also yields a negative, statistically significant coefficient.

Table 21. Mixed logit models with household characteristics (lognormal distribution for price, normal distributions for other variables)

VARIABLES	(1)	(2)	(3)
Price - mean	-0.310	-0.786***	-0.809***
	(0.341)	(0.240)	(0.230)
Price - SD	1.358***	-2.099***	2.033***
	(0.232)	(0.0982)	(0.119)
Time - mean	-0.213***	-0.248***	-0.155***
	(0.0294)	(0.0354)	(0.0236)
Time - SD	0.123***	0.166***	0.159***
	(0.0195)	(0.0244)	(0.0225)
asc - mean	-2.242**	-2.277**	-2.515**
	(1.097)	(0.965)	(1.140)
asc - SD	9.455***	11.47***	11.21***
	(1.378)	(1.696)	(1.671)
income*price - mean	0.120		
	(0.0931)		
income*price - SD	-0.0648		
	(0.0683)		
primary edu*price - mean		0.0350	
		(0.110)	
primary edu*price - SD		0.185	
		(0.116)	
frac women*time - mean			-0.328***
			(0.0628)
frac women*time - SD			0.188***
			(0.0318)
Observations	2,870	2,870	2,870

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

Appendix G presents mixed logit regression results for households with at-home primary sources (piped, well, or rainwater).

Reported prioritized source attribute

Following the choice experiment, respondents were asked: "When you are deciding which water source to use, which factor would you say is most important?" Table 22 presents the frequency of responses among households without an at-home source. Sixty one percent report that distance to the compound is the most important factor and 31% cite price. Since respondents answered this question after completing the choice experiment that prompted them to focus on trade-offs between time and price, they may have been particularly focused on these two source attributes. Had this question been asked before the choice experiment, the responses may have been different.

Table 22. Responses to the question "When you are deciding which water source to use, which factor would you say is most important?"

Response	**Frequency of response**
Distance from compound	62%
Cost	31%
Quality/safety	4%
Reliability	2%
Potential for conflict	1%

5.6 Coping costs

Sample households engage in a number of coping behaviors that result in additional water costs. Monthly coping costs for each household are estimated based upon the mean valuation of time (0.22 Ksh/minute or 0.15 USD/hour) and household-specific coping behaviors.[13] Time spent traveling to a water source outside the home and waiting to fill a jerrican represents a time cost. Treatment of water represents another coping cost. Sixty three percent of households without at-home primary sources reported to treat their drinking water and boiling was reported to be the most popular form of treatment (53% of all households). Six percent of households without-at home primary source use chlorine treatment and an additional

[13] The mixed logit model provides household-specific time valuation estimates, however coefficients depend on where the likelihood function converges and are unstable, so mean valuation of time is used.

3% both boil and use chlorine products to treat their drinking water. The equation for calculating coping costs for households without at-home sources is:

Coping costs = (price of vended water purchased in addition to primary source) + (walking time costs) + (other travel time costs) + (treatment costs)

Treatment costs were estimated by multiplying the amount of water used (in liters) by the cost of treatment per liter. For boiling, it was assumed that, as estimated by the WHO (2014), one kg firewood is needed to boil one liter of water and this was multiplied by the price of firewood per kg to obtain cost per liter. Time costs were estimated by multiplying the minutes spent collecting water by the mean valuation of time in the sample. Figure 7 provides a frequency distribution for the estimated monthly coping costs for households without at-home sources.

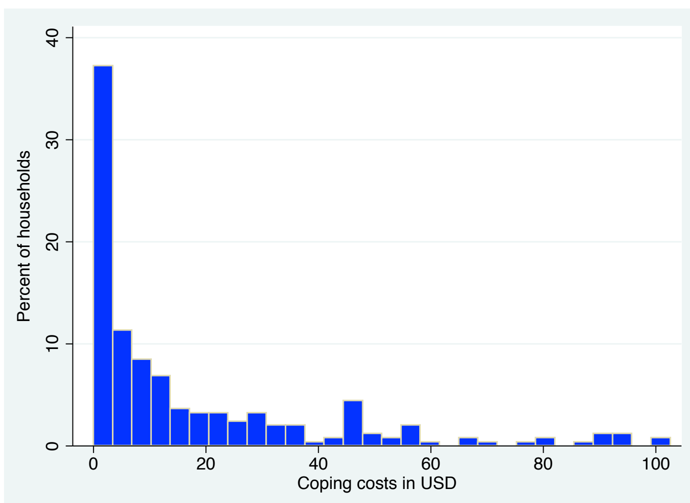


Figure 7. Monthly coping costs frequency distribution for sample households without at-home sources (excluding outliers of 179 and 278 USD/month)

Total cost is the sum of coping costs and price paid for water used from the household's primary source. Based on this calculation, the average monthly coping cost for sample households without at-home sources is 19 USD and the median is 8 USD. Including price paid, the average monthly total cost of water to households is 27 USD and the median 13 USD. Average water costs per 20 L jerrican, including coping costs, are remarkably consistent across low-, middle-, and high income groups (24, 27, and 23 Ksh respectively). Figure 8 shows the composition of monthly water costs by cost-type for each income classification for households without an at-home source. Households in the lowest and highest income groups spent the largest fraction of monthly coping costs on treatment of drinking water. For the middle-income group, time cost was the biggest component of cost, on average. The price paid for purchased vended water was a larger fraction of total monthly water cost for the wealthiest households compared to the other income groups.

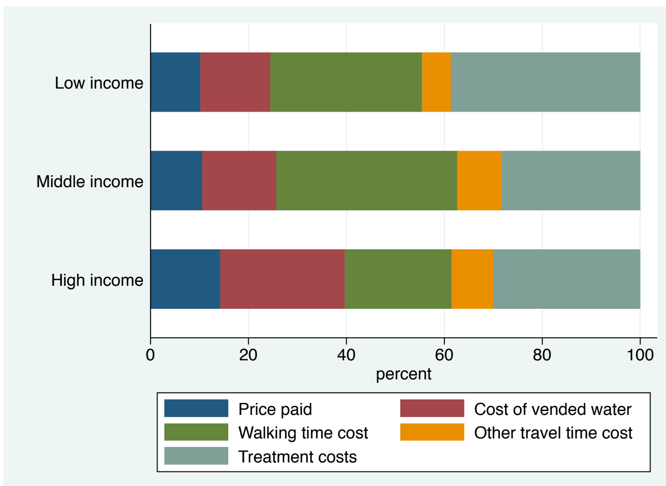


Figure 8. Composition of water costs by income group for households without at-home sources

Figure 9 shows the composition of monthly costs of water by source type for households without at-home water sources. For most primary source types, walking time and/or treatment costs represent the largest average water costs to households. Households purchasing vended water as their primary source are omitted because they were not asked about treatment of this source.

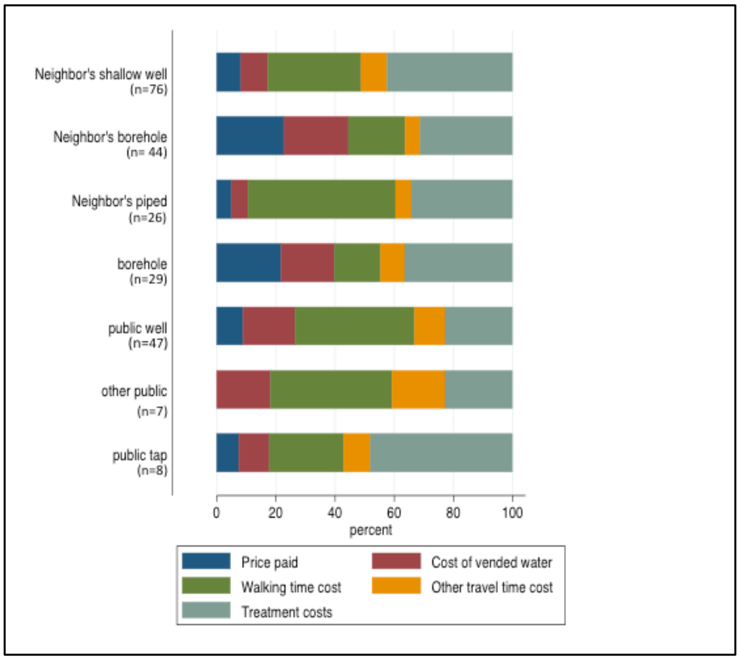


Figure 9. Composition of water costs by not at-home primary source type

Figure 10 shows mean monthly coping costs for households without at-home sources by primary source type. The eight public tap users reported an average of 40 USD/month in coping costs, however, the small sample size suggests that this may not be a reliable estimate. Boreholes are the source type associated with the lowest coping cost at about 10 USD/month.

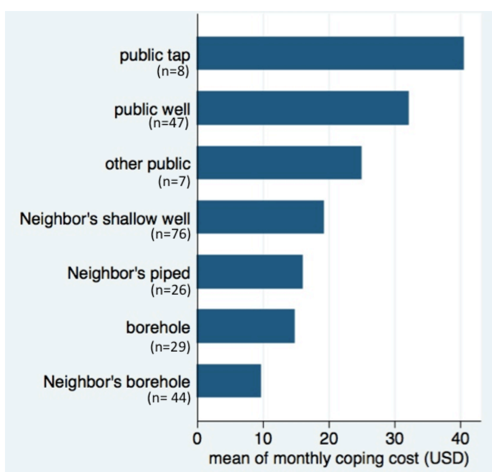


Figure 10. Average coping costs for households walking to their primary source

Figure 11 shows the relationship between coping cost per liter of water and monthly water use by households without at-home primary sources for the three income groups. The trend lines suggest that households with higher coping costs per liter use fewer liters of water, although the majority of households are clustered near the y-axis.

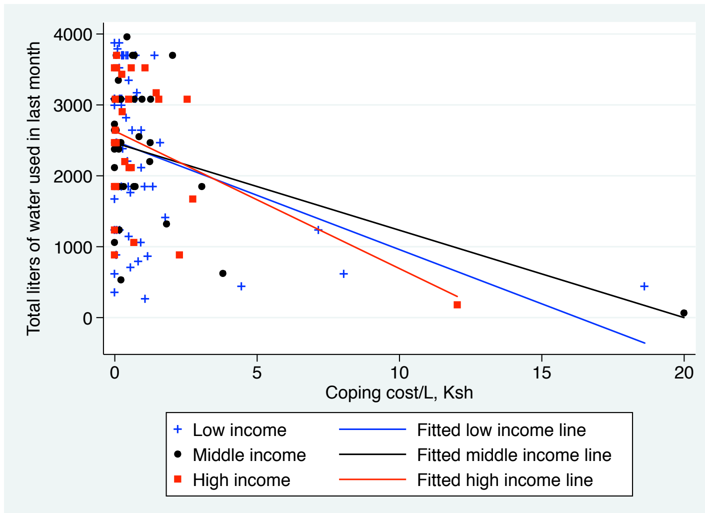


Figure 11. Coping cost per L and quantity of water used by households without at-home primary sources (excluding households with vended water as their primary source, with outliers removed)

Table 23 gives the linear regression model predicting liters of water used last month for households without at-home primary sources. Households with greater coping costs per liter and those charged a higher price per liter for their primary source reported to use statistically significantly fewer liters of water in the last month. Households with a higher monthly income and with a greater number of household members used statistically significantly more liters of water in the last month. The R-squared value for the model is 0.167.

Table 23. Linear regression of total liters of water used in the last month for households without at home primary water sources

VARIABLES	L water used last month
Coping cost per L (USD)	-288.4***
	(45.25)
Monthly household income (USD)	0.412**
	(0.179)
Number of Household members	326.2***
	(62.50)
Primary source price per L (USD)	-282,364***
	(104,947)
Constant	2,585***
	(455.3)
Observations	225
R-squared	0.167

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Even households with at-home piped connections incur coping costs; nearly all of these households have invested in storage tanks due to the intermittent delivery of water supply. Following Pattanayak et al. (2005), I use a real annual interest rate of 15% to amortize the capital investment of purchasing storage tanks into monthly payments and assume a storage tank lifetime of thirty years. (See Appendix H for calculations.) Based on these assumptions, households with piped connections spend an average of 102 Ksh (1.18 USD) per month in coping costs associated with the investment of purchasing storage tanks. Additional coping costs for households with piped connections include the use of secondary sources (24%) and the treatment of water before drinking. Although households were not asked about treatment of water from their at-home piped connections, most judged the water to pose a serious health risk and likely incurred treatment costs.

6. Conclusions

Consistent with economic theory and water source choice literature, I find that households prefer lower-priced water sources that take less time to collect. For households without a source at-home, an increase in price from 1 to 3 Kenyan shillings decreases the likelihood that the source will be selected by 11% on average. An increase in collection time from 20 to 40 minutes decreases the likelihood that a source will be selected by 15% on average. While the magnitude of these estimates is not transferable to other locations, similar methods can be used to assess user preferences. I am not able to draw conclusions about valuation of additional source attributes such as reliability or health risk as these variables were not varied across the hypothetical sources.

Neither monthly income nor primary education is found to impact sensitivity to price. In Kenya, Kremer et al. (2011) and Mu et al. (1990) also found no statistically significant relationship between income and source choice in communities without household-level piped connections. However, Kremer et al. (2011) did find that more highly educated mothers were more likely to prefer springs protected from contamination. In this sample, the price paid for water represents only 2% of monthly income on average (excluding households that did not pay for water). The low fraction of income spent paying for water may help to explain why I do not find a relationship between income and sensitivity to price. However, monthly coping costs are greater than the price paid for water on average. Households may have been particularly interested in reducing high time costs incurred during the dry season when collecting water. Unexpectedly, I find that households with a greater proportion of female members are less likely to select sources with higher collection times. A possible explanation for this finding is that households with a greater proportion of women have more members who prefer sources with lower water collection times.

On average, respondents value time spent collecting water at 37% of the casual labor wage rate. This falls within the range of estimated time valuations from two other studies that valued the time spent collecting water in Kenya. In the small town of Ukunda, Whittington et al. (1990) estimated the value of time at close to the unskilled wage rate, whereas in rural and more remote districts (Busia and Butere-Mumias), Kremer et al. (2011) estimated the valuation of time to be 7% of the local casual labor wage rate. The field site for this study was rural, but relatively near urban areas. It is plausible that people in more remote and rural areas would place a lower value on their time, given the difficulty of finding work.

Coping costs are large, particularly walking time and treatment costs. It is surprising that households are not purchasing more vended water, given the high time costs associated with water collection (particularly from public sources). The median price of a 20 L jerrican delivered by vendors, 10 Ksh, is less than half the mean total cost, including coping costs, (24 Ksh/20 L jerrican), but higher than the median cost (6.6 Ksh/20 L jerrican). However, vended water is generally perceived to be of similar quality compared to other sources, so many households would still incur treatment costs, a major coping cost. In addition, one fifth of households in the sample reported that vended water was not available in their area. A 2011 report by the International Finance Corporation estimated the average per capita cost for rural water supply infrastructure in Kenya to be 3,342 Ksh (40 USD). Given that the average sample household has five members, the cost of at-home piped services can be estimated at 200 USD per household. This is less than eight times average monthly water cost, including coping costs.

In testing a simple choice experiment that varied only two attributes, I aimed to determine whether the method could gather useful source preference data. High correlation between source attributes made estimating the relative importance of each attribute difficult. In addition, I find that the selection of a particular multinomial logit model (conditional versus mixed logit, as well as mixed logit distributional assumptions) may impact the magnitude of estimated parameters. This illustrates the importance of trying a variety of modeling techniques to determine whether the findings are robust across different specifications.

A limitation of this study is that fieldwork was conducted exclusively in the dry season. Respondents report similar water prices during both seasons, however the wait time in the rainy season is a small fraction of the dry season wait time. Given that 96% of households report to use rainwater, I would expect that households would be less interested in new hypothetical sources during the rainy season. Of potential concern is that collection times were based on self-reported walk and wait times, which may not be accurate. Since respondents were only asked about one-way walk times, round-trip times had to be estimated. Robustness checks on these time assumptions did not find more conservative walk time estimates to substantially change findings. Finally, including other stated choice best practices such as giving respondents time to think would give greater confidence in the results (Whittington, 2010).

Although about half of respondents had access to an "improved" primary water source (WHO, 2013), many were still interested in highly reliable and good quality hypothetical sources. Another key finding is that households in rural Kenya are incurring substantial coping costs, particularly time and treatment costs. For households without primary sources on their compound, these coping costs are greater than the price paid for water on average. As coping costs per liter of water increase, households use fewer liters of water, which may have health implications. The systematic literature review found that many gaps remain regarding how households make decisions about water source options. Simple choice experiments such as this one can be a useful tool for understanding water source preferences, however, careful design, piloting, implementation, and analysis of the data is essential.

APPENDIX A - WATER SOURCE CHOICE LITERATURE REVIEW SEARCH TERMS

("water source" OR "water supply" OR "sources of water" OR "source of water" OR "sources of drinking water" OR "source of drinking water" OR "water services" or "piped water" or "tap water" or "in-house water" or "domestic water" or "source water")
AND
("choice experiment" OR "discrete choice" OR "household choice" OR "choose sources" OR "conjoint analysis" OR "source choice" OR "stated preference" OR "choice model" OR "choice modeling" OR "choice of water")

APPENDIX B - PHOTOGRAPHS OF THE STUDY AREA

Figure 12. Households where the questionnaire was piloted neighboring the study region

Figure 13. Water sources: Thewa swamp (left), borehole (middle), and a shallow well (right)

Figure 14. Terrain off the main road (left) and town located along the main road (right)

APPENDIX C - FULL QUESTIONNAIRE WITH UNIVARIATE STATISTICS

Household Questionnaire: Kenya Water Choice

New text not in the original survey looks like this

1.1_____________ Questionnaire Number

Hello. My name is (*add name*). I am part of a team conducting a survey about water. Your household has been chosen at random to participate in the survey. May I and tell you more about the survey? (*Enumerator: please be sure you have the respondent's attention before continuing*)

Thank you for letting me tell you more about the survey. The survey is part of a research project conducted by Professor Peter Kimuyu at the University of Nairobi. This study is supported by Environment for Development - Kenya. The survey is for research purposes only.

I will ask you questions about sources of water in your community and how much water you use. Some questions are about your household, including your economic situation, and all the people who live here. If you are uncomfortable answering any question, you may skip that question and go on to the next one.

Your participation is voluntary; that means you do not have to participate unless you want to. If you accept to participate in the survey, we will need about 45 minutes of your time. The interview might be somewhat shorter or longer, but should not take more than one hour.

The results of this survey will be kept confidential. If you agree to participate, we will keep your questionnaire in a safe place at the University of Nairobi and use it for research only. We will not share your individual answers with other persons. We will combine your answers with the answers from all the other households we interview and produce a report that we hope will be useful to leaders in your community and Kenya's policy makers.

If you agree to participate, I will need to write your name and location on this form so that my supervisor can check that our interview was done properly. We will keep the link between your name and your responses for five years, after which we will erase any link between your name or location and your responses.

Professor Peter Kimuyu can address any concerns you may have about the study and his telephone contact is (omitted)

Do you have any questions about the study? (*Enumerator: try to answer the questions as best you can. If you are unsure of the answer, contact your supervisor*)

Do you agree to participate in this study? (1) _____Yes (2) _____No

Are you the person who is mostly responsible for water-related decisions? For example, are you the person who would decide how much water is collected and which source of water to collect from?(*Enumerator: If not, ask them to get that person to be the respondent.)*

(*Enumerator: if the respondent looks* ***younger than age 30****, ask them politely what year they were born. If the respondent is not yet 18 years old, thank them for their consideration and end interview*)

1.2 Respondent Name__________________ 1.3Enumerator Code |___|___|

1.4 Sub- village

Sub-location	**# of interviews**
Kianjai	141
Mutionjuri	129
Machaku	44
Nairiri	74

1.5 Start time (end time : Q121)

Calculated duration of interviews (*excluding 13.15 and 13.5 hour interviews –seem to be errors)*

mean	1 hour 12 min
median	1 hour
min	33 min
max	3 hours
n	380

1.6 Date of interview (all 2013):

Date	Frequency	Percent
Aug 5	11	2.84
Aug 6	24	6.19
Aug 7	23	5.93
Aug 8	17	4.38
Aug 9	22	5.67
Aug 10	20	5.15
Aug 12	20	5.15
Aug 13	23	5.93
Aug 14	21	5.41
Aug 15	21	5.41
Aug 16	26	6.7
Aug 17	23	5.93
Aug 18	1	0.26
Aug 19	30	7.73
Aug 20	28	7.22
Aug 21	32	8.25
Aug 22	22	5.67
Aug 23	9	2.32
Aug 24	15	3.87

Enumerator*: throughout the survey, please mark "don't know" responses with the code -98*

SECTION 2. WATER***

2.1 Just to confirm, are you the person who is mostly responsible for water-related decisions such as where to get water and how much to collect?

Freq	Response
387	Yes
1	No

2.2 Are you also the person who collected the most water in the past seven days, that is from last _<*today's day of week*>_ until today?

Freq	Response
301	Yes
87	No

IF NO, *ask if it is possible for the person who collected the most water to join the respondent for the water-related questions.*

2.3. Is the person who collected the most in the past seven days present?

Freq	Response
73	Yes
13	No

PIPED CONNECTIONS ***

2.4 Do you have a piped connection to a network in the compound, even if it is not currently working? Do not include pipes that connect a private well or borehole to your compound.

Percent (n=388)	Response
27%	Yes
73%	No

By sub-location

	Kianjai n=141	Mutionjuri n=129	Machaku n=44	Nairiri n=74
Yes	22%	26%	55%	24%
No	78%	74%	45%	76%

2.5 Is the tap inside your house, or outside? *(**Enumerator**, inside = under the roof)*

Percent (n=106)	Response
2%	Inside the home
97%	Outside the home
1%	Both inside and outside

2.6 Has any water come out of your tap in the past six months?

Percent (n=106)	Response
72%	Yes
28%	No

If yes, (*skip to question 2.8*)

2.7 Is that because you are behind on a water payment, or because the system has not been functioning properly? (*skip to "Rainwater" on page 7*)

Percent (n=30)	Response
10%	Behind on payment
90%	System not functioning properly

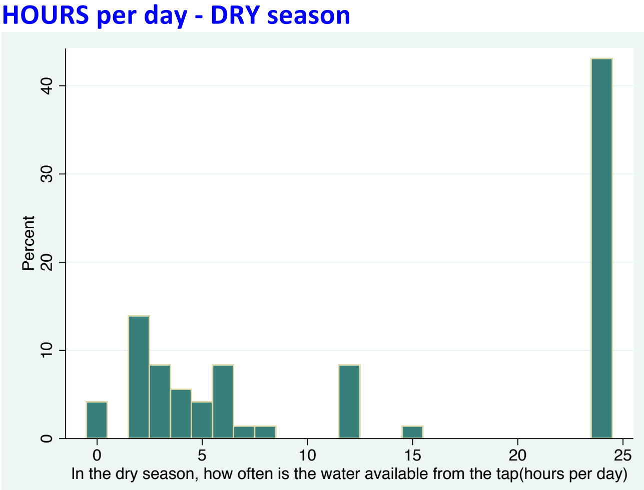
2.8 Where does the water from your tap come from?

Number of respondents providing each answer by sub-location

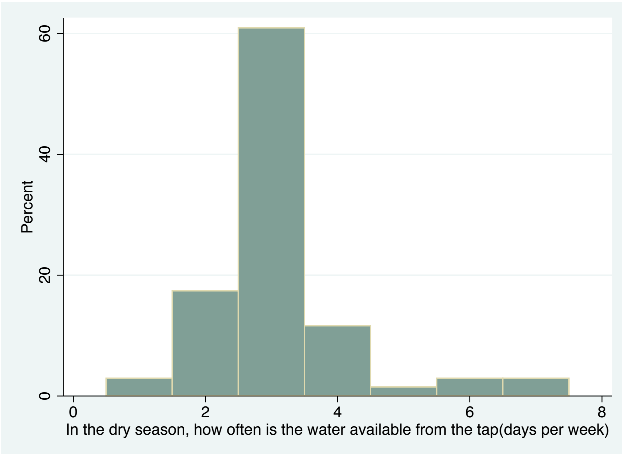
Origin of tap water	Kianjai	Mutionjuri	Machaku	Nairiri
Molem	4	5	0	0
Imetha WASCO	9	7	1	0
Mawea project water	0	2	22	16
Nyambene	0	4	0	0
Don't know	128	112	21	58

2.9 In the **dry** season, how often is water available from the tap? (For households with a functioning tap)

HOURS per day - DRY season

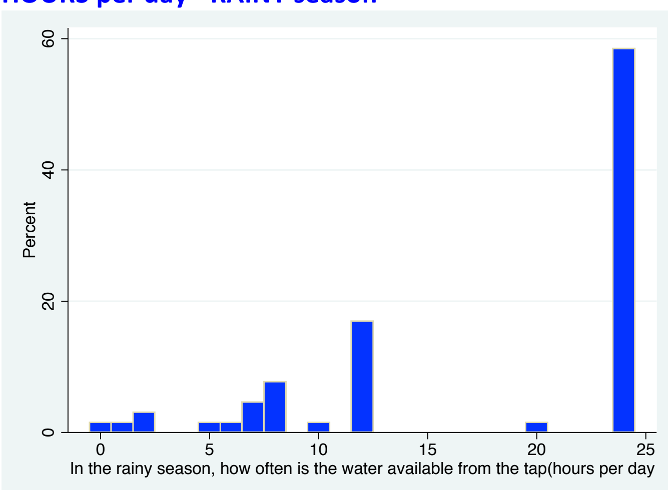


DAYS per week - DRY season

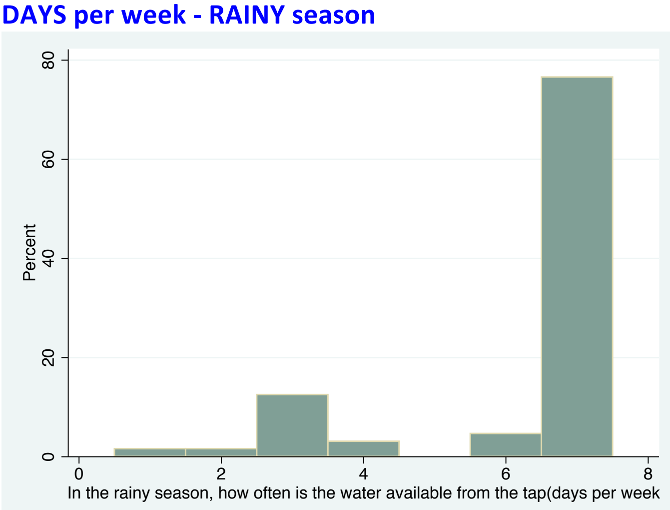


2.10 In the **rainy** season, how often is water available from the tap?

HOURS per day - RAINY season



DAYS per week - RAINY season



Number of households with 24 hrs/day, 7 days per week piped service in dry and rainy seasons

	Kianjai n=19	Mutionjuri n=18	Machaku n=23	Nairiri n=16
dry season	2	0	0	0
rainy season	7	7	13	6

	DRY SEASON n=75		RAINY SEASON n=56	
2.11 During the dry (rainy) season, how does the water from the tap **taste**? (*read responses*)	11%	Sweet	61%	Normal
	63%	Normal	27%	Poor
	19%	Poor	7%	Varies
	4%	Varies	5%	DK
	4%	DK		
2.12 During the dry (rainy) season, what **color** is the water from the tap? (*read responses*)	79%	Clear	11%	Clear
	13%	Cloudy	32%	Cloudy
	4%	Brown	57%	Brown
	4%	DK		
2.13 During the dry (rainy) season, how would you judge the **health risk** of water from the tap (*read responses*)	7%	No risk	12%	No risk
	32%	Some risk	52%	Some risk
	49%	Serious risk	36%	Serious risk
	12%	DK		
2.14 During the dry (rainy) season, how would you judge the **reliability** of water from the tap? (*read responses*)	1%	Very Regular	59%	Very Regular
	35%	Regular	30%	Regular
	44%	Irregular	5%	Irregular
	16%	Unreliable	4%	Unreliable
	4%	DK	2%	DK
2.15 Are any of these characteristics different in the dry season than the rainy season, (in terms of color, taste, health risk, or reliability)?	74%	Yes		
	11%	No		
	15%	DK		

2.13 – Perceived health risk by sub-location – frequency of responses (not percentage)

	Kianjai	Mutionjuri	Machaku	Nairiri
No Risk	10	9	4	1
Some risk	7	4	16	10
Serious Risk	1	1	3	4
Don't know	1	3	0	1
Total	19	17	23	16

2.16 Do you have a water meter?

Percent (n=76)	Response
28%	Yes
72%	No

2.17 Is the meter working?

Percent (n=22)	Response
82%	Yes
18%	No

2.18 Do you pay regularly for water from this connection?

Percent (n=76)	Response
80%	Yes
18%	No

2.19 Do you receive a water bill, or do you contribute informally to a group that is responsible for the piped system?

Percent (n=61)	Response
13%	Receive a water bill
87%	Pay informally to a group

Count that receive a bill versus pay to a group by sublocation

	Kianjai	Mutionjuri	Machaku	Nairiri
Receive a bill	3	1	0	4
Pay to group	4	16	23	10

2.20 How much do you pay for your connection? ____________Ksh every _____month(s)

Percent (n=53)	Response (Ksh)
8%	250
58%	300
2%	350
30%	500
2%	2000

2.21 Could I see a copy of your most recent water bill?

Percent (n=8)	Response
0%	Yes
100%	No

2.22 How much was your water bill the last time you received one? _____________Ksh

Percent (n=7)	Response
14%	250
29%	300
14%	400
29%	500
14%	4000

2.23 How many months did it cover? _________months

Percent (n=7)	Response
86	1
9	6

2.24 How is your bill determined? N=8 because only 8 respondents receive a water bill

(a) _3 respondents____ Fixed payment ---------------> __300 or 500 (n=2)_Ksh per month

(b) __5 respondents___Volumetric charge -------------> ___5 (n=1)__Ksh per cubic meter

(b) ___1 respondent__ Combination of fixed and volumetric -------> (1) __250 (n=1)__Ksh per month
+ (2)__-9.8 (n=1)____Ksh per cubic meter

*This seems to be an attempted "Don't know" response?

2.25 Has your household contributed in any other way to **operations and maintenance** of the piped water system?

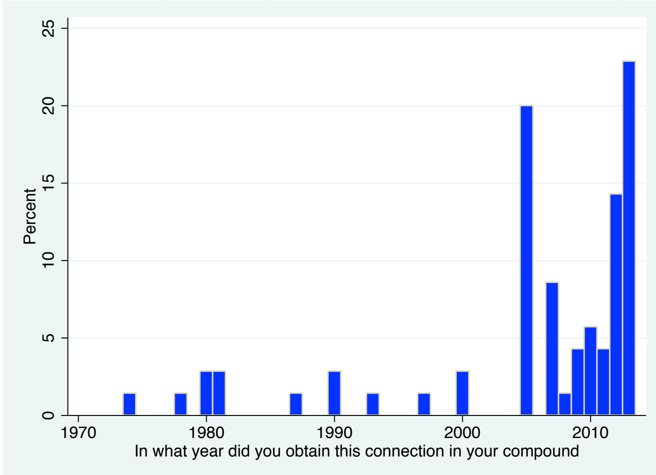
Percent (n=76)	Response
55%	Yes
45%	No

2.26 How has your household contributed? (*check all that apply*)

Percent (n=41)	Response
76%	Someone in my household volunteered
71%	The household contributed money

Sums to more than 100% because respondents listed multiple responses

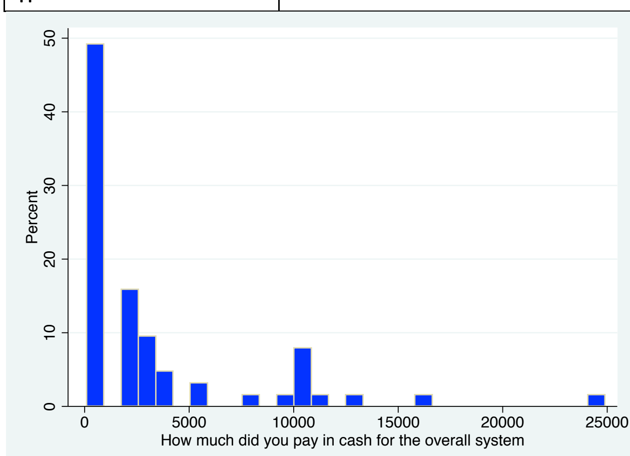
2.27 In what year did you obtain this connection in your compound? ___________ year



2.28 How much did you pay in cash and in labor for your connection? (*mark all that apply*)

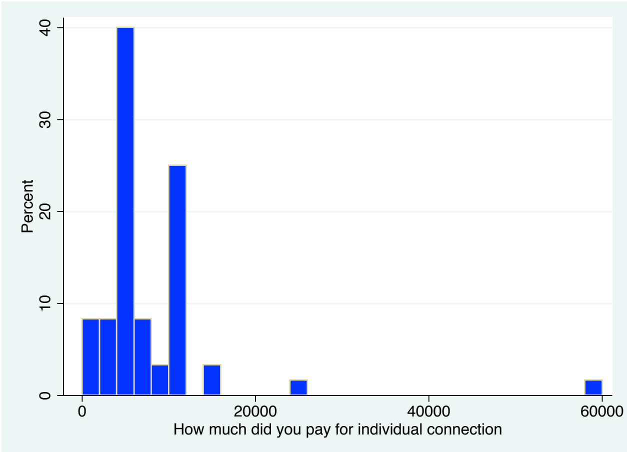
Overall System

mean	3271 (excluding outlier 24900, mean = 2922)
median	2000
min	100
max	24900
n	63



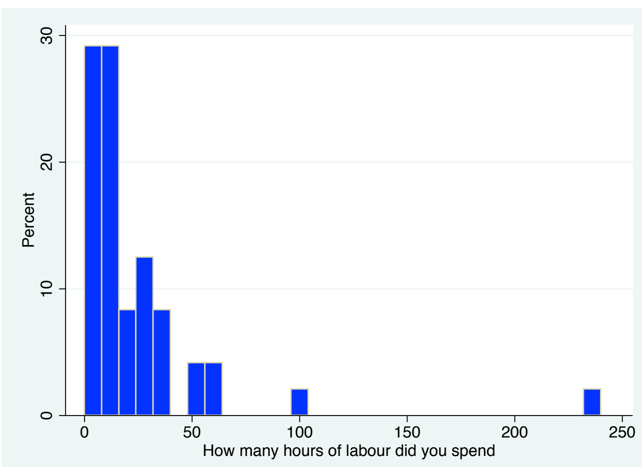
Individual connection

mean	7509
median	5000
min	0
max	60000
n	60



Hours of Labor

mean	22
median	12
min	0
max	240
n	48



2.29 Do you store water from your piped connection?

Percent (n=76)	**Response**
89%	Yes
11%	No

28 out of 31 (95%) of respondents who said that they get 24 hour piped water in the dry season said yes, they store water. But most of these people only receive water 3 days a week (only 2 respondents receive water 7 days per week in the dry season.) 31 out of 33 (94%) respondents who have 24 hour/day 7-days/week service in the rainy season store water (but this could be in the dry season, the question wasn't season-specific), and 1 out of 2 (50%) respondents who have 24 hour/day 7-days/week service in the dry season reports to store water.

2.30 How do you collect and store water from the piped system? *(read responses)*

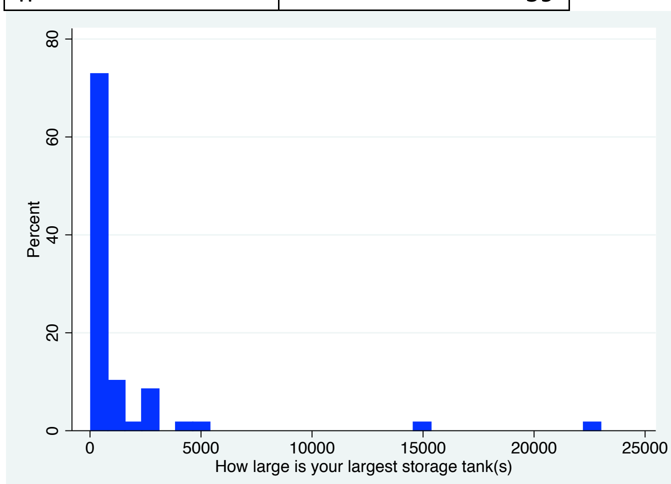
Percent (n=68)	Response
13%	Collect in many small buckets
16%	Collect and store in large
71%	Both

2.31 How many storage tanks do you have? ______________tanks

Percent (n=59)	Response
51%	1
32%	2
8%	3
7%	4
2%	7

2.32 How large is your largest storage tank(s)? ____________ liters

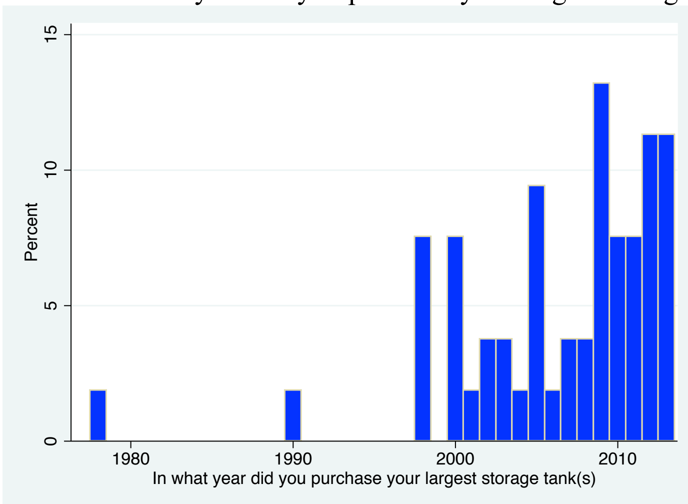
mean	1345
median	210
min	50
max	23000
n	59



2.33 Are your large storage tank(s) covered?

Percent (n=59)	Response
78%	Yes
22%	No

2.34 In what year did you purchase your largest storage tank? ___________Year



2.35 How much would it cost today to buy a storage tank of this size?______________ Ksh

mean	5995
median	2000
min	500
max	35000
n	41

2.36 How many liters of water are you able to store in your house in total from your piped connection? ______________Liters

mean	1019 (mean =691 after dropping the outlier 23000)
median	300
min	25
max	23000
n	68

2.37 How many liters of piped water have you used in the past seven days?

L per day		L per week	
mean	232	mean	1132
median	170	median	890
min	1	min	240
max	1000	max	4180
n	46	n	24

2.38 How many liters of piped water do you use on an average week in the **dry** season?

L per day / L per week

mean	230	mean	1229
median	160	median	930
min	10	min	100
max	1350	max	3200
n	51	n	16

2.39 How many liters of piped water do you use on an average week in the **rainy** season?

L per day / L per week

mean	81	mean	458
median	60	median	410
min	20	min	40
max	300	max	1050
n	42	n	12

2.40 What purposes do you use water from your tap for? (*check all that apply*)

Yes No (%)

(1) |_99%_|_1%_| Drinking
(2) | 100%_|_0_| Domestic uses (Washing dishes, bathing, personal washing, cooking)
(3) |_92%_|_8%_| Watering animals, or plants or trees near the home

2.41 Do you allow your neighbors to regularly use water from your connection?

Percent (n=76)	**Response**
50%	Yes
50%	No

2.42 How many households regularly use water from your connection?_______ households

mean	4.6
median	3
min	1
max	20
n	38

2.43 Has sharing water from your connection ever led to conflicts with your neighbors?

Percent (n=38)	**Response**
26%	Yes
74%	No

2.44 Do they pay anything to use water from your connection?

Percent (n=38)	Response
24%	Yes
76%	No

It appears that households with more neighbors using their connection are more likely to be paid for the water. The mean number of number of neighbors using free connections is 3.4 (median 3) and the mean number of neighbors paying for water is 8.4 (median 6).

2.45 How much do they pay? *(mark the quantity as appropriate)*

(a)_2 or 5 (n=2)_ Ksh - Flat rate per month **OR**

(b)__2, 2.5, or 5 (n=4)_ Ksh – Per 20 L jerrican

(c)__0_Different prices for different users

2.46 Do you sell water to anyone else besides your neighbors? 100% said No

2.48 In general, how satisfied are you with your piped service? (*Read responses*)

Percent (n=76)	Response
16%	Very satisfied
58%	Somewhat satisfied
16%	Somewhat dissatisfied
11%	Very dissatisfied

Average responses by satisfaction level

	Very satisfied	Somewhat satisfied	Somewhat dissatisfied	Very dissatisfied
Dry season avg hours/day	17.1	12.8	12.2	10.6
Dry seasons avg days/week	3.4	2.9	3.8	3.3
Rainy avg hours per day	20.3	17.3	20.2	12.7
Rainy avg days/week	6.9	6.1	5.9	6.2
% reporting at least some risk (either season)	0.8	0.9	0.6	0.5
Poor taste in either season	0.2	0.4	0.3	0.25
Brown color in either season	0.2	0.5	0.5	0.5
Avg year obtained connection	2007	2007	2000	1997
Avg cash for connection	1572	2540	8465	4625
Avg payment for connection	7240	8445	4583	4260

SECTION 3. *RAINWATER* **
Now I want to ask you about collecting rainwater.

N=387	RAINY SEASON	
3.1 How does stored rainwater **taste**? (*read responses*)	66%	Sweet
	27%	Normal
	4%	Poor
	2%	Varies
	1%	DK
3.2 How would you judge the **health risk** of drinking rainwater? (*read responses*)	70%	No risk
	27%	Some risk
	1%	Serious risk
	2%	DK

3.3 Does your household collect rainwater?

Percent (n=388)	Response
96%	Yes
4%	No

3.4 How do you collect and store rain water? *(read responses)*

Percent (n=371)	Response
33%	Collect in many small buckets or jerricans
10%	Collect and store in large (>=50) tanks
57%	Both

3.5 How many storage tanks do you have? _______________tanks

mean	1.6
median	1
min	1
max	8
n	249

3.6 How large is your largest storage tank(s)? ____________liters

mean	1389
median	200
min	50
max	90000
n	249

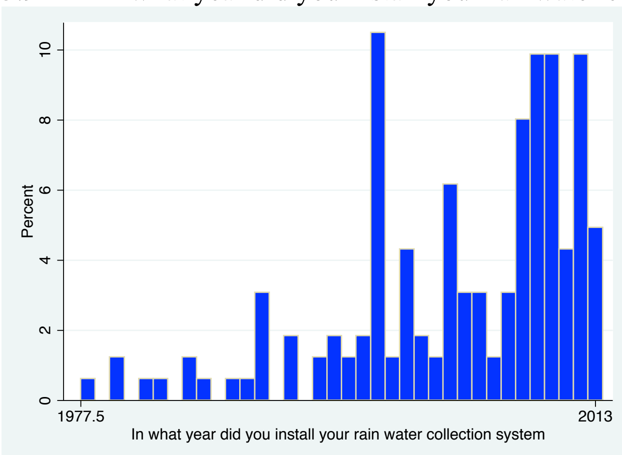
3.7 Are your large storage tank(s) covered?

Percent (n=249)	Response
79%	Yes
21%	No

3.8 Do you have a rainwater collection system?

Percent (n=251)	Response
70%	Yes
30%	No

3.9 In what year did you install your rainwater collection system? _____________Year



3.10 How much would it cost today to build this collection system (including purchasing equipment)? Ksh

mean	8897
median	700
min	150
max	500000
n	130

3.11 How much water can you store in total? All in Liters

mean	1216 (mean is 975 excluding outlier 90340)
median	200
min	6
max	90340
n	371

3.12 What purposes do you use the collected rainwater for? (*check all that apply*)

Yes No

(1) |_98%_|_2%_| Drinking

(2) |_98%_|_2% | Domestic uses (Washing dishes, bathing, personal washing, cooking)

(3) |_80%_|_20%_| Watering animals, or plants or trees near the home

3.13 Do you boil or treat stored rainwater before drinking it? (*check all that apply*)

Percent (n=366)	Response
63%	No boiling or treatment
12%	Add chlorine/Aquaguard/Pur
25%	Boil before drinking
1%	Let stand and settle

3.14 How many liters of rainwater do you use on an average week in the **rainy** season?

L per DAY		L per WEEK	
mean	140	mean	734
median	120	median	680
min	20	min	5
max	1000	max	3240
n	315	n	56

3.15 For how many weeks does your stored rainwater last when the rains stop in the dry season?

Percent (n=371)	Response
73%	Less than one week
11%	1 week
5%	2 weeks
11%	3 or more weeks

SECTION 4: BOTTLED WATER ************************************

Now I want to ask you about bottled water that some people purchase.

n=388	YEAR-ROUND	
4.1 How does bottled water **taste**? (*read responses*)	27%	Sweet
	32%	Normal
	14%	Poor
	4%	Varies
	22%	DK
4.2 How would you judge the **health risk** of drinking bottled water? (*read responses*)	61%	No risk
	9%	Some risk
	3%	Serious risk
	27%	DK

4.3 Have you or someone in your household purchased bottled water in the past seven days?

Percent (n=387)	Response
17%	Yes
82%	No
1%	DK

4.4 How many liters have you or someone in your household purchased in the last seven days?

Percent (n=67)	Response
52%	1 L
21%	2 L
27%	3 or more liters

SECTION 5: HAND-DUG SHALLOW WELLS ***************************

5.1 Does your household have a hand-dug well on your compound?

Percent (n=388)	Response
23%	Yes
77%	No

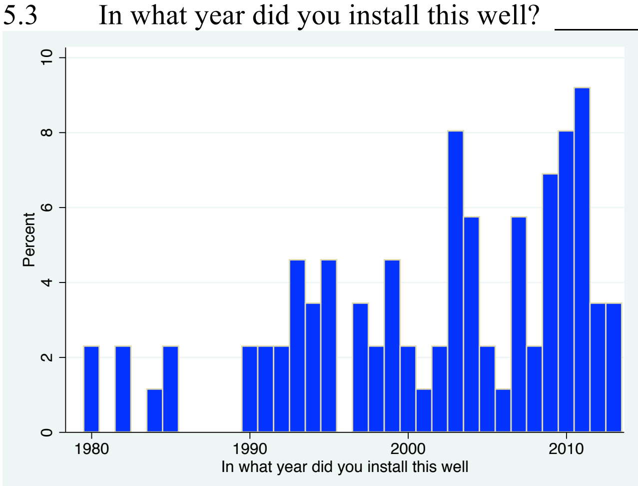
Count of households with and without hand-dug wells and piped connections

	piped connection	no piped connection
Hand-dug well	12	77
No hand-dug well	94	205

5.2 How far away is the private well from the main house, in in meters?

mean	43 (mean without 500 m outlier is
median	15
min	1
max	500
n	89

5.3 In what year did you install this well? ____________Year



5.4 How deep is your well?

mean	33
median	30
min	7
max	100
n	87

5.5 How much would it cost today to build this type of well? _______________ Ksh

mean	34789
median	25000
min	1500
max	200000
n	79

5.6 Does the well have a concrete pad around it or packed dirt?

Percent(n=77)	Response
57%	Concrete
35%	Logs and/or timber, wood
3%	Mud
4%	No cover

5.7 How do you get water from the well?

Percent (n=87)	Response
91%	Bucket and rope
6%	Electric pump
3%	Handpump

	DRY SEASON n=89		**RAINY SEASON n=66**	
5.8 During the dry (rainy) season, how does the water from the well **taste**? (*read responses*)	4%	Sweet	3%	Sweet
	56%	Normal	59%	Normal
	33%	Poor	32%	Poor
	6%	Varies	5%	Varies
	1%	DK	2%	DK
5.9 During the dry (rainy) season, what **color** is the water from the well? (*read responses*)	79%	Clear	46%	Clear
	10%	Cloudy	21%	Cloudy
	10%	Brown	33%	Brown
	1%	DK		
5.10 During the dry (rainy) season, how would you judge the **health risk** of drinking water from the well? (*read responses*)	30%	No risk	24%	No risk
	44%	Some risk	41%	Some risk
	25%	Serious risk	32%	Serious risk
	1%	DK	3%	DK
5.11 During the dry (rainy) season, how would you judge the **reliability** of water from the well?	38%	Very Regular	98%	Very Regular
	37%	Regular	2%	Regular
	16%	Irregular		
	9%	Unreliable		
5.12 Are any of these characteristics different in the dry season than the rainy season, (in terms of color, taste, health risk, or reliability)?	74%	Yes		
	22%	No		
	3%	DK		

Percent of respondents reporting some or serious risk from drinking water from each each source

	Kianjai	Mutionjuri	Machaku	Nairiri
piped water (2.13) dry or rainy	66%	71%	96%	93%
rainwater (3.2) just rainy	23%	38%	10%	32%
bottled water (4.2) year-round	20%	15%	0%	19%
hand-dug wells (5.10) dry or rainy	66%	79%	81%	100%
water vending (7.4) year round	78%	81%	100%	96%

5.13 What purposes do you use the well water for (*check all that apply*)

Yes No

(1) |_93%_|_7%_| Drinking

(2) |_97%_|_3%| Domestic uses (Washing dishes, bathing, personal washing, cooking)

(3) |_93%_|_7%_| Watering animals, or plants or trees near the home

5.14 Do you boil or treat water from the well before drinking it? (*check all that apply*)

Percent (n=85)	Response
22%	No boiling or treatment
13%	Add chlorine/Aquaguard/Pur
56%	Boil before drinking
2%	Let stand and settle
16%	Add chlorine to well

** sums to >100% because 28 respondents selected multiple responses*

5.15 In the last 12 months, was there water in your well during the month of .....(*read months, mark an X if water is normally available*)

% that said yes (n=89):

|_98%__| January |_93%__| July

|_98%__| February |_85%__| August

|_98%__| March |__67%_| September

|_97%__| April |_72%__| October

|__96%_| May |_88%__| November

|_96%__| June |_92%__| December

5.16 How many liters did you obtain from the private well **in the past seven days**? (*Enumerator: fill in the table in the way that is easiest for the respondent. If you record amounts per day, confirm that this amount is usually collected every day during the week*)

Per day		PerWeek	
mean	503	mean	2565
median	240	median	2020
min	40	min	540
max	3880	max	8760
n	47	n	33

5.17 How many liters do you obtain from the private well on an average week in the **dry** season?

Per day		Per Week	
mean	427	mean	2663
median	240	median	2220
min	60	min	200
max	3960	max	8760
n	55	n	26

5.18 How many liters do you obtain from the private well on an average week in the **rainy** season?

Per day		Per Week	
mean	110	mean	900
median	100	median	560
min	20	min	160
max	400	max	3000
n	52	n	13

5.19 Do you share the water from your well with any other people on a regular basis?

Percent (n=89)	**Response**
85%	Yes
15%	No

To investigate whether households obtaining water from their neighbors are poorer, I separated the sum of monthly income (11.7) into 3 groups with equal numbers of respondents (low, mid, high income). I then found the number of HHs reporting that a neighbor's well was their overall primary source 8.48-8 (This is source 21 from section 8, sources away from compound.)

Frequency of a neighbor's well as a primary source

	low income	mid income	high income
# HHs using neighbor's well as primary water source	31	25	17
mean monthly income	KES 14,090	KES 34,926	KES 105,729
total HHs in income group	130	130	128

A total 73 HHs reported neighbor wells as their primary source overall.

5.20 Number of HH using water from your hand-dug well?

mean	7
median	4
min	1
max	60
n	75

5.21 Has sharing water from your well ever led to conflicts with your neighbors?

Percent (n=75)	**Response**
28%	Yes
72%	No

SECTION 6: PRIVATE BOREHOLE *******************************

6.1 Does your household have a borehole on your compound? ***100% said No***

SECTION 7: WATER VENDING ***********************************

7.1 Is it possible for households in your area to buy from a water vendor <u>who delivers to your household</u>?

Percent (n=388)	Response
80%	Yes
20%	No

7.2 How does the water from vendors **taste**? (*read responses*)	3%	Sweet
	35%	Normal
	32%	Poor
	22%	Varies
	9%	DK
7.3 What **color** is the water from vendors? (*read responses*)	48%	Clear
	35%	Cloudy
	10%	Brown
	7%	DK
7.4 How would you judge the **health risk** of drinking water from vendors? (*read responses*)	14%	No risk
	51%	Some risk
	27%	Serious risk
	9%	DK
7.5 How would you judge the **reliability** of vendors who deliver water?	6%	Very Regular
	27%	Regular
	43%	Irregular
	17%	Unreliable
	7%	DK

7.6 How much do water vendors charge per 20L jerrican during the dry season? In Ksh

mean	10
median	10
min	2
max	30
n	297

7.7 Have you ever paid anyone to deliver water to your household?

Percent (n=309)	Response
64%	Yes
36%	No

7.8 How many liters did you purchase from water vendors in the past seven days)
(*Enumerator: fill in the table in the way that is easiest for the respondent.*)

Per day		Per Week	
mean	273	mean	458
median	120	median	360
min	20	min	40
max	1400	max	1800
n	21	n	59

If we assume that only those who answered question 7.8 are the ones to have bought water in the last week, then the percent of all sample households who have purchased vended water in the last 7 days is only 21%.

Percent (n=388)	Response
21%	Yes
79%	No

7.9 How many liters do you purchase from water vendors on an average week in the **dry** season?

Per day		Per Week	
mean	167	mean	525
median	120	median	420
min	20	min	21
max	1600	max	1460
n	96	n	93

7.10 What purposes do you use the water you purchase from vendors? (*check all that apply*)

Yes No

(1) |__96%__|__4%__| Drinking
(2) |_100%__|___0%_| Domestic uses (Washing dishes, bathing, personal washing, cooking)
(3) |__91%__|__9%__| Watering animals, or plants or trees near the home

SECTION 8: WATER SOURCES AWAY FROM THE COMPOUND ******

****Summary statistics for this table not included in this document**

Now I will ask you about other types of water sources like boreholes, public water points, rivers, and neighbors. I want you to think about all the sources of water **it is possible for households in your area** to collect water from during both the rainy and the dry season. These could be sources that your neighbors use even if you do not, in addition to sources that you have collected from in the past 12 months. If so, I will ask you several questions about that source. **Please also include getting water from your neighbor (either for free or by paying).**
Which of these sources is it possible for households in your area to use?

SOURCE CODE TABLE

	Generic		**Kianjai and Nearby**		**Mutionjuri and nearby**		**Machako and Nairiri**
21	Neighbor's hand-dug shallow well	41	Thewa Swamp	61	Methodist Church	81	Machako public tap from Mwea Water Project
22	Neighbor's borehole	42	Kianjai Borehole aka "Polytechnic"	62	Mbuya "LifeLink"/Redcross water point	82	Nkundi/Muchena/Boniface private wells
23	Neighbor's piped connection	43	Nchoro boreholes (multiple)	63	Dairy borehole on main tarmac road	83	Loria River (seasonal)
		44	Moturi Deep Hand-dug well	64	Mutionjuri paid borehole	84	Nairiri Primary School tank

(Write Source ID in the table below. Write all possible IDs across the first row first, and then complete all questions for the first source before asking about the next source. For NEW sources (not on map), turn to last page of survey and complete. Write pre-assigned source ID for new water sources in 3.10. Tell your supervisor today about the new source)

	Source A	Source B	Source C	Source D	Response Codes
8.1 *Source ID*	|__|__|	|__|__|	|__|__|	|__|__|	
8.2 How many **minuteswalk** is this source from your house, **one-way with the full container**?					Minutes
8.3 How many minutes would you spend **waiting** at this source queuing or filling your container during the **dry** season?					Minutes
8.4 How many minutes would you spend **waiting** at this source queuing or filling your container during the **rainy** season?					Minutes
8.5 How does the water from this source **taste** during the **dry** season? *(read responses)*					(1) Sweet,(2)Normal, (3) Poor, (4) Varies, (-98) Don't know
8.6 What **color** is the water during the **dry** season? (*read responses*)					(1) Clear (2) Cloudy (3) Brown (4) Other
8.7 During the dry season, how would you judge the **health risk** of water from this source? (*read responses*)					(1) No risk, (2) Some risk, (3) Serious risk, (-98) Don't know
8.8 How likely is it that there could there be **conflict** if you collected from this source in the **dry** season? (*read responses*)					(1) Not likely at all (2) Somewhat likely, (3) Very likely, (-98) Don't know
8.9 Are any of those characteristics (taste, color, health risk, conflict) different during the rainy season? (*If NO, skip to 8.14*)					(1) Yes (2) No
8.10How does the water from this source **taste** during the **rainy** season? (*read responses)*					(1) Sweet, (2) Normal, (3) Poor, (4) Varies, (-98) Don't knowrink
8.11 What **color** is the water during the **rainy** season? (*read responses*)					(1) Clear (2) Cloudy (3) Brown (4) Other
8.12During the rainy season, how would you judge the **health risk** of water from this source (*read responses*)					(1) No risk, (2) Some risk, (3) Serious risk, (-98) Don't know

8.13 Could there be conflict if you collected from this source in the **rainy** season? (*read responses*)					(1) Not likely at all (2) Somewhat likely, (3) Very likely
8.14How many days per week is this source available?					Days/week *(should be in 1-7 range)*
8.15How many hours per day is this source available when it is open?					Hours/day *(should be in 1-24 range)*
Payment					
8.16 Would you have to **pay** anything to use this source during the **dry** season? (*if* ***NO****, skip to* ***8.21***)					(1) Yes (2) No (-98) Don't know
8.17 How would you pay during the dry season? (*read responses*)					(1) By jerrican or container; (2) per time period (*skip to 8.19*)
8.18 (*If by jerrican*): How much do you pay per jerrican during the dry season?					(Ksh per 20 L jerrican)*(Enumerator: calculate amount/20L if necessary)*
8.19 *(If per time period):* How much do you pay?					Shillings
8.20 (Time period)	(1)__wk (2)__mo (3)__yr	(1)_wk (2)_mo (3)__yr			*Check one per source*
8.21 Is the payment system different during the rainy season? (*If NO, skip to 8.26*)					(1) Yes (2) No (-98) Don't know
8.22 How would you pay during the rainy season? (*read responses*)					(1) By jerrican or container; (2) per time period
8.23 (*If by jerrican*): How much would you pay per jerrican during the rainy season?					(Ksh per 20L jerrican)
8.24 *(If per time period):* How much would you pay during the rainy season?					Shillings
8.25 *(Time period)*	(1)__wk (2)__mo (3)__yr	(1)_wk (2__mo (3)_yr			*Check one per source*
**************************	******	*****	*****		*******
8.26Have you used this water source in the past twelve months? *(if****NO****: turn to beginning of table and ask about the next possible source).*					(1) Yes (2) No
**************************	******	*****	*****		*******
8.27 Do you use water from this					(1) Rainy season only,

source during the rainy season only, the dry season only, or both?					(2) Dry season only, (3) Both
8.28 *If used during the DRY season,* what do you use the water from this source for during the **dry** season? (*read responses*; *check all that apply*)					*Check all that apply*
(1) Drinking					(1) Yes, (2) No
(2) Domestic uses (Washing dishes,cooking, bathing)					(1)Yes, (2) No
(3) Watering animals, or plants and trees near the home					(1)Yes, (2) No
8.29 *If used during the RAINY season,* do you use water from this source for different purposes during the rainy season (*if NO, skip next question*)					(1) Yes (2) No
8.30 What do you use the water from this source for during the **rainy** season? (*read responses*; *check all that apply*)					*Check all that apply*
(1) Drinking					(1)Yes, (2) No
(2) Domestic uses (Washing dishes,cooking, bathing)					(1)Yes, (2) No
(3) Watering animals, or plants and trees near the home					(1)Yes, (2) No
8.31 (*If the source is used for drinking*) Do you treat water from this source during the **dry** season before you drink it? (*If No skip to 8.34)*					(1) Yes, always (2) Yes, sometimes (3) No
8.32 How do you treat water from this source?					(1) Boil; (2) Filter; (3) Let it stand and settle;(4)Chlorine/ Waterguard/Pur; (5) Solar disinfection (-95) Other
8.33 (*If the source is used for drinking*) Do you treat water from this source during the **rainy** season before you drink it?					(1) Yes, always (2) Yes, sometimes (3) No
8.34 Do you collect water from this source on foot? (***IF NO,***					(1) Yes, (2) No

SKIP TO 8.39)					
8.35 How many trips **on foot** have all the members of your household made to this source in the past seven days?					Trips
8.36 During the **dry** season, how many trips per week would all the members of your household make to this source **on foot**?					Trips
8.37 During the **rainy**season, how many trips per week would all the members of your household make to this source **on foot**?					Trips
8.38 How many liters in total did all members of your household collect from this source on foot in the past seven days? (*Calculate using daily collection)*					
8.39 Do any household members use a donkey, a cart, a wheelbarrow, bicycle, motorbike, or vehicle to collect water from this source?(*if **NO**, skip to 8.44*)*Check all that apply*					
(1) Donkey					(1)Yes, (2) No
(2) Cart					(1)Yes, (2) No
(3) Wheelbarrow					(1)Yes, (2) No
(4) Bicycle					(1)Yes, (2) No
(5) Motorbike					(1)Yes, (2) No
(6) Vehicle					(1)Yes, (2) No
8.40 How many trips by any of these means have all of the members of your household made in the past seven days (in total)? (*use daily trips asguide*)					Trips
8.41 How much water in total have you collected from this sources during the past seven days using these means (not on foot)?					
8.42 During an average week in the **dry** season, how much water would all the members of your household collect from this source using those means? (**not on foot)**?					Liters

8.43 During an average week in the **rainy** season, how much water would all the members of your household collect from this source using those means? (**not on foot**)?					Liters

Ask about all water sources before continuing.

IF RESPONDENT USES A GRUNDFOS "LIFELINK" BOREHOLE, ASK

8.44 Do you have a keyfob to use the Grundfos borehole?

Percent (n=388)	Response
2%	Yes
98%	No

8.45 Do you ever let anyone else collect water using your keyfob?

Percent (n=7)	Response
71%	Yes
29%	No

8.46 Would you be willing to share your LifeLinkkeyfob number?

3 respondents said yes

8.47 Do you borrow someone else's keyfob?

(1)______Yes
(2) _____ No

8.48 Thinking back over all the water sources you have told me your household uses, which source would say is the **primary** source you use during the **dry** season for:

ENUMERATOR: WRITE IN SOURCE CODES; write "99" if household does not use water for that purpose.

(1)|__|__| Drinking
(2)|__|__| Washing dishes, washing in the home
(3)|__|__| Bathing, personal washing
(4)|__|__| Cooking
(5) |__|__|Watering animals
(6) |__|__| Watering plants or trees near the home
(7) |__|__|Other productive uses (e.g. brickmaking, beer-brewing)
(8) |__|__|**OVERALL, "most uses"**

source code	Source Name	# of HHs	% of total
	At Home		
13	Private hand dug shallow well	78	20%
11	Piped connection	59	15%
15	Vended water	12	3%
12	Rainwater	2	1%
16	Bottled water	0	0%
	Outside home		
21	Neighbor's hand-dug shallow well	76	20%
82	Nkundi/Muchena/Boniface private wells	47	12%
22	Neighbor's borehole	44	11%
23	Neighbor's piped connection	26	7%
42	Kiangia Borehole aka "Polytechnic"	13	3%
43	Nchoro boreholes (multiple)	12	3%
81	Machako public tap from Mwea Water Project	8	2%
62	Mbuya "LifeLink"/Redcross water point	4	1%
41	Thewa Swamp	3	1%
45	Kabaibui stream	2	1%
47	Mbututia swamp	1	0%
66	Rehema polytechnic	1	0%

8.49 Now I want you to imagine that your primary water source was unavailable for some reason. Which source would you use as a back-up for each of the following purposes?
ENUMERATOR: WRITE IN SOURCE CODES; write "99" if household does not use water for that purpose.

(1)|__|__| Drinking
(2)|__|__| Washing dishes, washing in the home
(3)|__|__| Bathing, personal washing
(4)|__|__| Cooking
(5) |__|__|Watering animals
(6) |__|__| Watering plants or trees near the home
(7) |__|__|Other productive uses (e.g. brickmaking, beer-brewing)

8.50 Thank you for your patience in answering my questions. I know they are very detailed but they are important for our research. Now I want you to think about the total amount of water you collected over the past seven days. How many 20 liter jerricans of water did you use in the past seven days for ......(*Enumerator: fill in the table in the way that is easiest for the respondent to remember. If you record amounts per day, confirm that this amount is usually collected every day/week*)

Water use	per day mean(range) n	per week mean(range) n
(1) Drinking	8 (1-60) n=149	12(1-140) n=235
(2) Washing dishes, washing in the home	34 (5-820) n=278	195 (20-880) n=109
(3) Bathing, personal washing	70 (10-570) n=267	403 (20-1460) n=121
(4) Cooking	28(10-600) n=304	150 (20-560) n=83
(5) Watering animals	59 (3-1400) n=265	325 (40-1120) n=66
(6) Watering plants or trees near the home	264 (5-2300) n=60	400 (20-2195) n=27
(7) Other productive uses (e.g. brickmaking, beer-brewing)	281(20-1080) n=19	334 (20-2010) n=20

Hypothetical scenario
Now I would like you to imagine that a group is planning to install several new water points in your area to improve your access to water.The group could be the government or it could be a non-governmental organization. These water points could be boreholes or public standpipes from the piped network.If they install only a few water points, people might have to walk further and wait longer to collect water. If they install more, people might walk shorter distances and have to wait less. Installing these water points is expensive, however. Suppose <the group> will need to charge people who use the water points to recover their costs and properly maintain the water points. If they install more points, they may need to charge more per jerrican.

(Enumerator: remind yourself which source the respond said was their primary source in the dry season for "most purposes", in 8.48, option 8)

You just told me that the primary source for most purposes right now was <source>. In addition to that source, I want you to imagine you have two new water points available for you to use. You should assume that quality of the water from the new water point is excellent and safe for drinking. You should also assume that the reliability of the new water point would be excellent: it would always have good pressure and you could collect from it whenever it is convenient for you. Finally, you should assume that using the source would not cause any conflict with other water users.

The two new water points differ only in the cost you would have to pay per jerrican, and the total amount of time it would take you to walk to the source, wait, fill your container and return. Here is the first task I would like you to think about. (*Enumerator: Shuffle the choice cards to randomize their order. Hand respondent first choice card*)

In this card, you can see the two new water points. The first new water point (*point to source A*) would require a total time to walk to the source, wait, fill your container and walk home of <***X***> minutes. If you used this source, you would have to pay <***X***>Ksh per 20L jerrican. The second new water point (*point to source B*) would require a total time to walk to the source, wait, fill your container and walk home of <***X***>

minutes. If you used this second new source, you would have to pay <***X***>Ksh per 20L jerrican. Of course, you can still continue to use your current primary source :<*mention source*>.

S1. Do you have any questions?

(Enumerator: Ask the following questions for each of the choice cards after pointing out the collection time and cost of the two new options).

If these three sources were available to you right now, which source would you **most prefer to use**? (*Enumerator: Mark source with "1".*) Remember that the two new sources have excellent quality, reliability, and using them would not cause conflict.

Which source would you **least prefer to use**? (*Enumerator: Mark source with "3", and mark "2" in the other source. In case of a tie, mark "2" for both sources*)

For the source you would most prefer to use, about how many jerricans do you think you would collect from this source during an average week this month? (*If preferred source is the current source, write -93 in table below under "number jerricans collected"*)

<table>
<tr><td rowspan="2">TASK NUMBER</td><td colspan="3">RANK (1 = best, 2=middle, 3 = worst)</td><td rowspan="2"></td><td rowspan="2">Number of jerricans collected from best source per day (-93 = current source)</td></tr>
<tr><td>New Source A</td><td>New Source B</td><td>Current Source</td></tr>
<tr><td>a. |__|__|</td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>b. |__|__|</td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>c. |__|__|</td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>d. |__|__|</td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>e. |__|__|</td><td></td><td></td><td></td><td></td><td></td></tr>
</table>

Summary statistics of ranked sources A, B, and current source

task_ID	Source A Price	Source A Dist	Source B Price	Source B Dist	% selecting source A as #1	% selecting source B as #1	% selecting current source as # 1	n
11	0.25	10	1	5	33%	31%	35%	121
12	0.25	30	1	10	11%	54%	35%	122
13	3	5	1	30	44%	17%	39%	122
21	0.25	10	3	5	25%	45%	30%	133
22	0.25	30	3	5	21%	41%	38%	133
23	3	10	0.25	30	39%	23%	39%	133
31	0.25	30	1	5	16%	60%	24%	127
32	3	5	1	10	34%	37%	30%	127
33	3	10	1	30	42%	21%	37%	126
99	1	10	0.25	5	4%	70%	27%	381

S2. Did the respondent think carefully about the choices? N = 388

(1) _76%___Yes, very carefully
(2) _22%___Yes, some
(3) __2%__No

S.3 When you are deciding which water source to use, which factor would you say is most important? *(Read responses – Select one)*

Percent (n=388)	Response
3%	Quality/ safety
61%	Distance from the compound
30%	Cost
5%	Reliability
1%	Potential for conflict

SECTION 9 : SANITATION ***
Now I have a few questions about your health.

9.1 Did any one in your household have diarrhea **in the past seven days**?

Percent (n=388)	Response
9%	Yes
90%	No
0.3%	DK

9.2 Did your household spend any money on special foods, medicines, or any other treatment for the diarrhea in the **past seven days**? Answered by only those who said yes to the previous question.

Percent (n=37)	Response
84%	Yes
13%	No
3%	DK

9.3 If yes, how much have you spent in the past seven days? _________________ Ksh

mean	408
median	280
min	20
max	1500
n	29

Average total water expenditures in the last week

mean	74
median	0
min	0
max	900
n	388

9.4 Do you have a sanitation facility for disposal of human excreta on your compound?

Percent (n=388)	Response
97%	Yes
3%	No

9.5 Do you share this facility with anyone outside your household?

Percent (n=376)	Response
13%	Yes
87%	No

9.6 How many other households use your facility on a regular basis? __________households

Percent (n=48)	Response
65%	1
15%	2
10%	3
10%	4 or more

9.7 Whom do you share this toilet with? (*check all that apply*)

Percent (n=48)	Response
17%	Neighbors who are not relatives
83%	Neighbors who are relatives

9.8 What facility do you use? For the 12 households who don't have a facility on the premises

Percent (n=12)	Response
75%	Neighbor's facility
8%	Public latrine
17%	Bush/ no facility

9.9 How many other households use this facility on a regular basis?
5 respondents said 1 other household, 4 respondents said 2 other households

9.10 How many minutes does it take to walk there, one-way?
7 respondents said 1 minute, 1 said 2 minutes, 1 said 3 minutes

9.11What kind of sanitation facility is this? (*read options*)

Percent (n=386)	Response
1%	Flush/ water-sealed toilet
99%	Pit latrine

WATER-SEALED TOILET

**3 respondents answered these questions even though only 2 answered that they had a flush toilet in question

9.12Do you flush the toilet by pouring water by hand or with a tank?

(1) _2 respondents__ Hand pour flush
(2) _1 respondent__ Tank

9.13 Overall, how satisfied are you with this toilet?

(1)_ 1 respondent ____Very satisfied
(2)_ 2 respondents_____Somewhat satisfied
(3)_____Less than satisfied/Somewhat dissatisfied
(4)_____ Not satisfied at all

9.14 How much would it cost to have this type of flush toilet installed today?
responses: 11000, 15000, 30000 Ksh

9.15 Where are the wastes from the toilet discharged?
(1)__100%___Septic tank

9.16 How much would it cost to have the septic tank cleaned today?
responses: 5000, 6000, 10000 Ksh

9.17 How much would it cost to have a septic tank like that one installed today?
responses: 15000, 30000, 30000 Ksh

skip to " Demographics"

PIT LATRINE

9.18 Does the pit latrine you use have a slab?

Percent (n=386)	Response
54%	Yes
46%	No

9.19 Is the pit latrine you use ventilated?

Percent (n=388)	Response
23%	Yes
77%	No

9.20 Overall, how satisfied are you with the pit latrine you use?

Percent (n=386)	Response
40%	Very satisfied
35%	Somewhat satisfied
16%	Somewhat dissatisfied
9%	Very dissatisfied

9.21 What do you like least about pit latrine you use? (*Spontaneous response, choose only one*)
(1)__1%___Nothing, it is completely satisfactory
(2)__3%___Dirty
(3)___3%__Dark
(4)__2%___Far from the house
(5) ___8%__ Need to keep digging new pits
(6) ____25%_ Dangerous risky
(7) __15%___ No privacy
(8) ___31%__ Smell
(-95) ___Other, specify: __14 said "want it cemented"_________

SECTION 10: DEMOGRAPHICS
(***Enumerator: read the following***) Now I would like to ask you some questions about you and the people who live in this household. When I say "household", I mean the people whom you regularly share meals with on a daily basis and who sleep here in this compound most of the time.

10.1 What is your relationship to the household head?

Percent (n=388)	Response
23%	Head
68%	Wife/Husband
5%	Child/ Adopted child
0.5%	Grandchild
1%	Father/ Mother
0.3%	Sister/ Brother
0.8%	Son/ Daughter-in-law
0.5%	Brother/ Sister-in-law
0.3%	Not related

10.2 What is highest school grade you completed? (*If Std 3 or higher, skip next question*)

Percent (n=388)	Response
9%	None
19%	Std 1-6
49%	Std 7-8
18%	Form 1-6
5%	Vocational Diploma
2%	Some University

10.3 Can you read in any language?
Assumed literate if they had completing Std 3 or higher

Percent (n=388)	Response
89%	Yes
11%	No

10.4 What is your marital status?

Percent (n=388)	Response
77%	Married
4%	Separated
11%	Widow/Widower
9%	Never Married

10.5 What is highest school grade your spouse completed? (*If Std 3 or higher, skip next question*)

Percent (n=298)	**Response**
7%	None
16%	Std 1-6
54%	Std 7-8
17%	Form 1-6
4%	Vocational Diploma
2%	Some University

10.6 Can your spouse read in any language? *(Read responses)*

Assumed literate if they had completing Std 3 or higher

Percent (n=298)	**Response**
91%	Yes
9%	No

I would like you to tell me about all of the members of your household. If it easier, you can tell me their names, but you don't have to if you don't want to.

ID	10.7 Name (First name only, *optional*)	10.8 Gender	10.9What is their age, or what year were they born? (*Enumerator: record whichever is easiest for respondent)*		10.10 Has this person collected water outside the household in the last seven days? (*Codes : 1 = Yes, 2 = No)*
			Age	**Year**	
1 (*resp.*)		☐M☐F			

Cont.

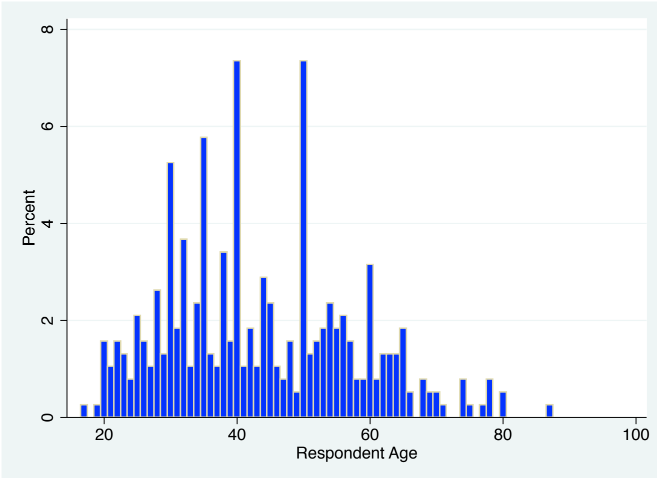
SKIP IF PERSON HAS NOT COLLECTED WATER	
10.11 Is[NAME] the person who collected the most water in the past seven days? (*Codes : 1 = Yes, 2 = No)*	10.12 Does [NAME] have a job for salary or wages? (*Codes : 1 = Yes, 2 = No)*

Respondent characteristics

10.8 Gender

Percent (n=46)	**Response**
69%	Female
31%	Male

10.9 Age of respondent



10.10 Has [*the respondent*] collected water outside the household in the last seven days?

Percent (n=46)	Response
51%	Yes
59%	No

10.11 Is [*the respondent*] the person who collected the most water in the past seven days?

Percent (n=46)	Response
58%	Yes
42%	No

10.12 Does [*the respondent*] have a job for salary or wages?

Percent (n=46)	Response
18%	Yes
82%	No

Household characteristics

10.8 Number of household members (by gender and total, excluding respondent)

Males		Females		Total HH members	
mean	2.7	mean	2.8	mean	5.5
median	3	median	3	median	5
min	0	min	0	min	1
max	10	max	9	max	15
n	388	n	388	n	388

10.13 How many of these people who lived in the house in the last seven days are visitors?

Percent (n=388)	Response
83%	None
9%	1
5%	2
3%	3 or more

SECTION 11: SOCIOECONOMICS ************************************

11.1 What is your religion?

Percent (n=388)	Response
83%	Catholic
9%	Protestant
5%	Other Christian
1%	Other

11.2 Do you own this compound or rent it?

Percent (n=388)	Response
97%	Own
3%	Rent

11.3 How much rent do you pay per month? Of the 7 respondents who pay rent:
Responses: 1, 5, 85, 300, 1500 Ksh

11.4 How many bedrooms does the main house have?

Percent (n=388)	Response
35%	1
37%	2
17%	3
11%	4 or more

11.5 How many other buildings or structures are there in this compound?*(must have 4 walls and roof)*

mean	3.5
median	3
min	0
max	17
n	387

11.6 Do you have a working connection to the main electricity grid?

Percent (n=385)	Response
11%	Yes
89%	No

Frequency of households with type of piped connection compared to electricity

	24-7 rainy season piped connection	intermittent piped connection	non-functioning piped	no piped connection
Electricity	6	13	6	18
No electricity	27	30	23	262

11.7 What was your most recent electricity bill? __________Ksh

mean	585
median	430
min	2
max	2005
n	39

11.8 How many months did it cover? 39 respondents said 1 month, 1 said 2 months, 1 said DK

11.9 What is your main source of energy for **lighting**?

Percent (n=388)	Response
11%	Main-grid electricity
72%	Kerosene
17%	Solar panel/ lamp
1%	other

11.10 What is your main source of energy for **cooking/ heating**?

Percent (n=388)	Response
1%	Main-grid electricity
14%	Biomass
80%	Firewood
4%	Charcoal

11.11 Now I would like to ask you about assets that your household owns. Can you tell me if anyone in this household owns any of the following items?

Yes

(1) 93%_Mobile telephone
(2) 75% Bicycle
(3) 16% Cart
(4) 82% Radio
(5) 34% TV
(6) 13% Motorbike
(7) 7% Vehicle

11.12 Does your household own any livestock? n = 387

(1) __91%___ Yes
(2) ___9% __ No (*skip next question*)

11.13 How many of the following does your household own? (*Enumerator: record number*)

	Cattle	Goats	Sheep	Chickens
mean	2	2	1	6
median	2	2	0	4
min	0	0	0	0
max	22	20	12	50
n	352	352	352	352

11.14 How much has your household spent on **food** in the **past seven days (Ksh)**?

mean	2133
median	1841
min	0
max	10500
n	388

Food expenditures last week per capita by household

mean	435
median	350
min	0
max	3500
n	388

11.15 How much land does your household own? In acres

mean	2.0
median	1.0
min	0.1
max	17.0
n	388

INCOME

Now I would like to know the sources of income for your household. Does anyone in your household currently earn money from...

	11.16			11.17
Source (n= 388 for y/n questions)	Yes	No		(*Ask about all sources of income before asking this question)*How much do you earn in a normal month from this source? (Ksh per month)
a. Full time employment	16%	84%		Mean 29,263, median 7500 range (1000, 700,023) n=52
b. Part-time or seasonal wage labor	53%	47%		Mean 4919, median 3500 range (250, 27,000) n=195
c. Business or self-employment	48%	52%		Mean 9138, median 5000 range (0, 105,000) n=169
d. Merry-go-rounds or ROSCAs	85%	15%		Mean 2269, median 1525 range (200, 20,000) n= 322
e. Remittances	23%	77%		Mean 3125, median 1500 range (200,50,000) n= 87
f. Rental income	5%	95%		Mean 8056, median 5000 range (100, 32,000) n= 17
g. Animal produce	14%	68%		Mean 2908, median 1520 range (0, 16140) n= 53
				How much did you earn in revenues over the **last harvest**?
h. Farming	96%	4%		Mean 39,104, median 25,200 range (940, 685,000) n= 373

(**If household has farming, ask11.18 and 11.19**:)

11.18 Did you rent any of the land that you cultivated? n=326

(1) __35%__Yes

(2) __65%__No (*skip next question*)

11.19 How much in total did you pay to rent that land that you cultivated?

*Note: strange that average and median rent per month is more than per year

	per month	per season	per year
mean	4222	5000	3606
median	4000	3000	3000
min	1500	2000	400
max	10000	10000	12000
n	55	3	61

11.20Are there any other sources of income I have not mentioned? I won't ask you **what** that source of income is. n=387 (1) _5%___Yes

(2) __95%__No (*skip to Section 11)*

11.21 How much does your household earn from that source in a normal month? Ksh

mean	8395
median	4500
min	2
max	30,000
n	19

Total monthly income (sum of all sources)

	Ksh
mean	51,713
median	34,850
min	0
max	730,400
n	388

SECTION 12: Wrap – up

This is the end of the interview. Thank you very much for your participation. I have to do a few more things outside your house, including taking a photo of your house and recording its location so that my supervisor can return here and make sure that the survey was conducted properly. Is that OK?

(***If the household has a private well or borehole, ask)*** May I see your well/borehole and take a picture of it for our research?

12.1 Finish time _____: _____

* * * * * * * * * * * * * * ***FOR THE ENUMERATOR**** * * * * * * * * * * * * * *

12.2 Was the person who answered the questions irritated or nervous during the interview? n=388

(1) __4%__ Yes

(2) __96%__ No

12.3 Do you think the respondent made an effort to tell the truth? n=388

(1) _99%___ Yes

(2) __1%__ No

12.4 How would you rate the overall quality of the interview? n=388

(1)_78%__ Good

(2) 20%__ Fair

(3)_1%__ Poor

12.5 Please note specific concerns or comments

APPENDIX D - ROBUSTNESS CHECKS

Table 24. Sub-set analysis excluding choice tasks with inconsistent preferences - mixed logit model with a lognormal distribution for price and normal distributions for collection time and asc for households without at-home sources

VARIABLES	Sub-set	Full sample
	mixed logit	mixed logit
Price - mean	-0.631**	-0.752***
	(0.252)	(0.249)
Price - SD	2.357***	11.68***
	(0.173)	(2.070)
Time - mean	-0.334***	-0.225***
	(0.0504)	(0.0310)
Time - SD	0.201***	0.163***
	(0.0319)	(0.0233)
asc - mean	-7.876***	-2.775*
	(1.967)	(1.605)
asc - SD	11.52***	1.843***
	(1.947)	(0.120)
Observations	2,702	2,870

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 25. Logit regression for preferring current source over a hypothetical source for households without an at-home primary source

VARIABLES	(1)	(2)
Price difference (current - hypothetical)	-0.107**	-0.0923*
	(0.0461)	(0.0534)
Time difference (current - hypothetical)	-0.0112***	-0.0102***
	(0.00214)	(0.00265)
Current source serious health risk		-0.477
		(0.374)
Average hours per day current source		-0.00743
		(0.0515)
Constant	-0.324*	-0.250
	(0.171)	(0.588)
Observations	1,898	1,158

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 26. Logit regression for preferring hypothetical source A over hypothetical source B

VARIABLES	Selection of Hypothetical source A
Price difference (Hyp A – Hyp B)	-0.225***
	(0.0678)
Time difference (HypA – Hyp B)	-0.0667***
	(0.00748)
Constant	-0.471***
	(0.0630)
Observations	973

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 27. Mixed logit models with a conservative estimate of walk time (1.5*one way with full container) for households without at-home sources. *Distributions for mixed logit*: normal except lognormal for price.

VARIABLES	(1) conservative time estimate mixed logit	(2) normal time estimate mixed logit
Price - mean	-0.826***	-1.044***
	(0.223)	(0.231)
Price - SD	1.700***	2.401***
	(0.0958)	(0.0889)
Time - mean	-0.220***	-0.226***
	(0.0297)	(0.0309)
Time - SD	-0.826***	0.178***
	(0.223)	(0.0249)
asc - mean	-3.256**	-3.976**
	(1.620)	(1.802)
asc - SD	14.96***	17.05***
	(2.144)	(2.680)
Observations	2,870	2,870

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

APPENDIX E – CORRELATION BETWEEN INDEPENDENT VARIABLES

Table 28. Correlation between independent variables - households without an at-home source

	Price	Time	asc	Health Risk	Reliability	Conflict
Price	1.00					
Time	0.21	1.00				
asc	0.40	0.65	1.00			
Health Risk	0.40	0.61	0.85	1.00		
Reliability	-0.09	-0.26	-0.39	-0.30	1.00	
Conflict	0.29	0.64	0.76	0.63	-0.39	1.00

Table 29 below shows the correlation coefficients for households with an at-home source. Likelihood of conflict is omitted because it is zero for all at-home sources.

Table 29. Correlation between independent variables - households with an at-home source

	Price	Time	asc	Health Risk	Reliability
Price	1.00				
Time	-0.11	1.00			
asc	-0.33	-0.53	1.00		
Health Risk	-0.23	-0.41	0.77	1.00	
Reliability	0.036	0.28	-0.52	-0.33	1.00

APPENDIX F – ADDITIONAL REGRESSION MODEL

Table 30. Conditional logit model for asc interactions with health risk and reliability

VARIABLES	clogit selected_source
Price_CE	-0.202***
	(0.0274)
Time_CE	-0.0427***
	(0.00301)
asc_risk	-1.011***
	(0.174)
asc_hrsday	0.0278
	(0.0192)
asc	0.853***
	(0.216)
Observations	4,120

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

APPENDIX G – MULTINOMIAL REGRESSION MODELS FOR HOUSEHOLDS WITH AT-HOME SOURCES

The coding of the variables for households with at-home sources is given in Table 31 below.

Table 31. Description of primary and hypothetical source variables for choice experiment for households using an at-home source

Variable	**Description**	**Card source coding**	**Primary Source Coding**
Time	Round trip walk time and waiting	5, 10, or 30 minutes	WELL or PIPED: 2 minutes (to fill the container) RAIN: 0
Price	Price of 20L jerrican	Ksh 0.25, 1 or 3	WELL: 0 PIPED: 0 if don't pay. If pay: estimated Ksh/jerrican: cost per month divided by L/month (assuming amount was constant throughout the month)/(20 L/jerrican) RAIN: 0
Health Risk	Perceived risk from drinking	0	0 = reported no or some risk 1=reported serious health risk from drinking water
Reliability	Avg hrs/day in a wk (censored at 12 hrs/day)	12 hours/day	WELL: 12 hours/day or 0 if no water in well PIPED: average hours/day over a week reported RAIN: 0 in dry season
asc	alternative specific constant	0	1 for the current primary source

Table 32 presents mixed logit models for preferred source among households using an at-home source. The coefficient on price is not statistically significant, which is somewhat surprising. Since the median price per 20 L jerrican for this group is 0, households with at-home sources that are interested in the hypothetical sources likely are prioritizing a factor other than cost.

Table 32. Mixed logit model (lognormal price distribution) of preferred source for households using an at-home source

VARIABLES	Mixed logit
Price - mean	0.169
	(0.256)
Price - SD	-1.254***
	(0.239)
Time - mean	-0.385***
	(0.0724)
Time - SD	0.211***
	(0.0542)
asc - mean	1.390
	(1.151)
asc - SD	8.765***
	(1.698)
Observations	1,609

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

Table 33 shows the average marginal effects of increases in collection time. Similar to households without at-home sources, as collection time increases from 20 to 40 minutes, households are 11% less likely to select the source.

Table 33. Average marginal effects of increases in collection time for households with an at-home source

Time (minutes)	Change in probability
5-10	-0.10
10-20	-0.13
20-40	-0.11
40-60	-0.04

For households with at-home sources, current sources dominated both hypothetical sources in 72 choice tasks. In 20 of these tasks, the respondent did not select their primary source although it appears to have been a "better" choice. In all of these cases, the respondent had a well at-home, which suggests that assumptions made about the time associated with gathering water from at-home wells may have been a low estimate.

As a robustness check for collection time assumptions for households with at-home wells, as well as piped connections, I increased assumed collection time from 2 to 5 minutes and re-ran the multinomial logit models (See Table 34). I found marginally smaller (less than 10% different) coefficients on the time variable. I also checked whether this change reduced the number of households appearing to have "inconsistent preferences" in that they preferred a hypothetical source even though their current source appeared better. Assuming 5 minutes collection time for at-home piped connections and wells reduced the number of households with seemingly inconsistent preferences from six to one.

Table 34. Mixed logit models with collection time assumed to be 5 minutes (instead of 2 minutes) for households with at-home sources. *Distributions for mixed logit*: normal except lognormal for price.

VARIABLES	mixed logit
Price - mean	0.0944
	(0.242)
Price - SD	1.045***
	(0.161)
Time - mean	-0.346***
	-0.0697
Time - SD	-0.207***
	-0.0602
asc - mean	2.306**
	(1.093)
asc - SD	9.550***
	(2.054)
Observations	1,609

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

Table 35 shows mixed logit models including household characteristics. None of the household characteristic interaction terms (income, education, proportion of women) were found to be statistically significant for these households with an at-home primary source.

Table 35. Mixed logit regression models for preferred source for households using an at-home source including household characteristics. *Distributions for mixed logit*: normal except lognormal for price.

	(1)	(2)	(3)
VARIABLES	Mixed logit	Mixed logit	Mixed logit
Price - mean	0.288	0.105	0.114
	(0.412)	(0.348)	(0.211)
Price - SD	0.924***	-1.286***	1.021***
	(0.230)	(0.321)	(0.120)
Time - mean	-0.376***	-0.417***	-0.361***
	(0.0664)	(0.0847)	(0.0945)
Time - SD	0.196***	0.229***	0.201***
	(0.0489)	(0.0681)	(0.0534)
asc - mean	1.320	0.746	1.298
	(1.149)	(0.906)	(1.074)
asc - SD	9.987***	-9.188***	10.23***
	(2.175)	(1.997)	(2.242)
Inc_class*Price - mean	0.0786		
	(0.185)		
Inc_class*Price - SD	0.0153		
	(0.0829)		
Primary ed*Price - mean		0.00566	
		(0.375)	
Primary ed*Price - SD		0.156	
		(0.454)	
Fraction women*time - mean			-0.0622
			(0.245)
Fraction women*time - SD			-0.0696
			(0.191)
Observations	1,753	1,753	1,753

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

APPENDIX H – COPING COST CALCULATIONS

Walking time costs were estimated by multiplying the sum of round-trip walk and wait time by the average value of time. Respondents were not asked about travel time collecting water using other means including cars, motorbikes, bicycles, carts, and wheelbarrows. For these methods of transport, I assumed that wait time remained the same, but the total travel time was half of total walk time (so equivalent to one-way walk time). Monthly expenditure on chlorine was estimated by multiplying the per L cost of WaterGuard, a popular chlorine product in Kenya, by the quantity of water used for drinking by the household for a month (based on reported use in the last day or week). Eighty percent of households reported to use firewood to boil water so boiling costs were calculated based on the local price of firewood, reported to be approximately 800 Ksh/70 kgs. The WHO estimates that 1 kg of firewood is required to boil 1 L of water (WHO, 2014). I expect that some respondents gather their firewood, so their boiling costs likely are lower than these estimated costs.

For households with an at-home piped connection, I amortize the capital investment of storage tanks into monthly payments. I assume a real annual interest rate of 15% and a storage tank lifetime of thirty years (Pattanayak et al., 2005). The calculations are based on the equation:

$$\text{Periodic payment} = \text{Principal investment}*\frac{(i(1+i)^n)}{(1+i)^n - 1}$$

where n= 30, i=0.15, and principal is the capital investment (cost of storage tank) as reported by the household. The periodic payment was calculated annually and divided into equal monthly payments across the year.

Half of households with an at-home piped connection had more than one storage tank, but since respondents were asked only about the cost of their largest tank. I assumed that the cost of additional tanks was half the cost of the largest tank.

REFERENCES

Abramson, A., Becker, N., Garb, Y., & Lazarovitch, N. (2011). Willingness to pay, borrow, and work for rural water service improvements in developing countries. *Water Resources Research*, *47*(11), n/a–n/a. doi:10.1029/2010WR010147

Anand, P. B. (2001). Consumer preferences for water supply? An application of choice models to urban India. *WIDER Discussion Paper*, 1–40.

Asthana, A. (1997). Where the Water is Free but the Buckets are Empty: Demand Analysis of Drinking Water in Rural India. *Open Economies Review*, *8*(2), 137–149.

Blamey, R., Gordon, J., & Chapman, R. (1999). Choice modeling: assessing the environmental values of water supply options. *The Australian Journal of Agricultural and Resource Economics*, *43*(3), 337–357.

Boardman, A. E., Greenberg, D. H., Vining, A. R. & Weimer, D. L. (2011) *Cost-Benefit Analysis: Concepts and Practice*. Fourth Edition. Upper Saddle River: Prentice Hall.

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.

Briand, A, Nauges, C. & Travers, M. (2009). Les determinants du choix d'approvisionnement en eau des menages d.e dakar. *Revue d'Economie du Developpement, 9*(3): 83-108.

Briscoe, J., Chakroborty, M., & Ahmed, S. (1981). How do Bengali Villagers Choose Sources of Domestic Water. *Water Supply & Management*, *5*, 165–181.

Dutta, V., & Verma, N. M. P. (2009). Water Pricing, Affordability and Public Choice: An Economic Assessment from a Large Indian Metropolis, 1–15.

Echenique, M., & Seshagiri, R. (2009). Attribute-based willingness to pay for improved water services: a developing country application. *Environment and Planning B: Planning and Design*, *36*(3), 384–397.

Haider, W., & Rasid, H. (2002). Eliciting public preferences for municipal water supply options. *Environmental Impact Assessment Review*, 337–360.

Hensher, D. A., Rose, J. M., & Greene, W. H. (2005). *Applied choice analysis: a primer*. Cambridge: Cambridge University Press.

Hole, A R. (2007). Fitting mixed logit models by using maximum simulated likelihood. *The Stata Journal*, *7*(3), 388–401.

Hope, R. A. (2006). Evaluating water policy scenarios against the priorities of the rural poor. *World Development*, *34*(1), 167–179. doi:10.1016/j.worlddev.2005.07.014

International Finance Corporation. 2011. “The Market for Small-Scale Piped Water Systems in Kenya.” *SSAWA Market Brief.* World Bank Group.

Jeuland, M., Lucas, M., Clemens, J., & Whittington, D. (2010). Estimating the private benefits of vaccination against cholera in Beira, Mozambique: A travel cost approach. *Journal of Development Economics*, *91*, 310–322. doi:10.1016/j.jdeveco.2009.06.007

Kremer, M., Leino, J., Miguel, E., & Zwane, A. P. (2011). Spring Cleaning: Rural Water Impacts, Valuation, and Property Rights Institutions. *The Quarterly Journal of Economics*, *126*(1), 145–205.

Lancaster, K. J. (1966). A New Approach to Consumer Theory. *Journal of Political Economy*, *74*(2), 132–157.

Madanat, S., & Humplick, F. (1993). A Model of Household Choice of Water Supply Systems. *Water Resources Research*, *29*(5), 1353–1358.

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (ed.), *Frontiers in Econometrics* (105-142). New York: Academic Press.

Mu, X., Whittington, D., & Briscoe, J. (1990). Modeling Village Water Demand Behavior: A Discrete Choice Approach. *Water Resources Research*, *26*(4), 521–529.

Nam, P. K. (2005). Household Demand for Improved Water Services in Ho Chi Minh City: A Comparison of Contingent Valuation and Choice Modeling Estimates, 1–31.

Nauges, C., & Whittington, D. (2009). Estimation of Water Demand in Developing Countries: An Overview. *The World Bank Research Observer*, *25*(2), 263–294.

Null, C., Kremer, M., Miguel, E., Hombrados, J. G., Meeks, R., & Zwane, A. P. (2012). Willingness to pay for cleaner water in less developed countries: systematic review of experimental evidence. *The International Initiative for Impact Evaluation (3iE)*.

Pattanayak, S. K., Yang, J.-C., Whittington, D., & Bal Kumar, K. C. (2005). “Coping with unreliable public water supplies: Averting expenditures by households in Kathmandu, Nepal.” *Water Resources Research* 41(W02012) 1-11.

Persson, T. H. (2002). Household Choice of Drinking-Water Source in the Philippines*. *Asian Economic Journal*, *16*(4), 303–317.

Scarpa, R., Thiene, M., & Hensher, D. A. (2012). Preferences for tap water attributes within couples: An exploration of alternative mixed logit parameterizations. *Water Resources Research*, *48*(1), W01520.

Snowball, J. D., Willis, K. G., & Jeurissen, C. (2008). Willingness to pay for water service improvements in middle-income urban households in South Africa: A stated choice experiment. *South African Journal of Economics*, *76*(4), 1–16.

Tarfasa, S., & Brouwer, R. (2014). Estimation of the public benefits of urban water supply improvements in Ethiopia: a choice experiment. *Applied Economics*, *45*(9), 1099–1108.

Thompson, J., Porras, I. T., Tumwine, J. K., Mujwahuzi, M. R., Katui-Katua, M., Johnstone, N., & Wood, L. (2012). Drawers of Water II, 1–122.

Train, K. (2009). Chapter 6 Mixed Logit. In *Discrete Choice Methods with Simulation* (2nd ed., pp. 1–17). New York: Cambridge University Press.

Whittington, Dale, X. Mu, and R. Roche (1990) "Calculating the Value of Time Spent Collecting Water: Some Estimates for Ukunda, Kenya,'' *World Development* 18, 269–280.

Whittington, D. (2010). What Have We Learned from 20 Years of Stated Preference Research in Less-Developed Countries? *Annual Review of Resource Economics*, *2*(1), 209–236.

White, G. F., Bradley, D. J., & White, A. U. (1972). Drawers of Water. *Bulletin of the World Health Organization*, 1–11.

World Health Organization, UNICEF. (2013). Progress on Sanitation and Drinking Water: 2013 update. *Joint Monitoring Program*, 1–40.

World Health Organization. (2014). "Managing water in the home: Accelerated health gains from improved water supply." Available at http://www.who.int/water_sanitation_health/dwq/wsh0207/en/index4.html (5-1-14)