Environmental and Resource Economics (2019) 74:1563–1584
https://doi.org/10.1007/s10640-019-00380-5

Household Demand for Water in Rural Kenya

Jake Wagner[1] · **Joseph Cook**[1] · **Peter Kimuyu**[2]

Accepted: 21 October 2019 / Published online: 2 November 2019

Abstract
To expand and maintain water supply infrastructure in rural regions of developing countries, planners and policymakers need better information on the preferences of households who might use the sources. Using data from 387 households in rural Kenya, we model source choice and water demand using a discrete-continuous (linked) demand model. We find that households are sensitive to the price, proximity, taste, and availability in choosing among sources, but are not sensitive to other source qualities including color, health risk, and risk of conflict. Estimates of the value of time implied by our model suggest that households value time spent collecting water at one third of unskilled wages. We use the linked demand framework to estimate own-price elasticities in the rural setting. These estimates range between − 0.13 and − 1.33, with a mean of − 0.56, and are consistent with other elasticity estimates from small and large cities.

Keywords Rural water supply · Water source choice · Value of travel time · Water quality · Kenya · Household water demand · WASH · Water collection · Discrete-continuous demand

1 Introduction

Access to a basic[1] water service has increased globally from 81 to 89% between 2000 and 2015, and the Millennium Development Goal regarding global water supply was achieved. Much of the remaining gap in access is in rural parts of the global South: approximately 80% of the estimated 844 million people without access to a basic water service live in rural areas, mostly in sub-Saharan Africa (WHO/UNICEF 2017). Closing this gap requires not only the expansion of systems of public taps and small piped networks, but also the proper maintenance of existing infrastructure. The rural water sector has a poor history of project sustainability. Much was learned from the mistakes of the 1980's "Decade of

[1] A basic water service is a source within 30 min roundtrip of the household which, by nature of its design and construction, has the potential to deliver safe water (WHO/UNICEF 2017).

✉ Jake Wagner
jake.wagner@wsu.edu

1 School of Economic Sciences, Washington State University, Hulbert Hall 101, PO Box 646210, Pullman, WA 99164-6210, USA

2 Commission for Revenue Allocation, Nairobi, Kenya

Water and Sanitation", including a focus on meaningful participation of women in key water committee leadership roles, the importance of availability of spare parts and training to repair, and the need for "demand-led" planning approaches. Nevertheless, collection of user fees and a lack of cash on hand continue to be challenges (Koehler et al. 2015), and at any given time, one in three handpumps in sub-Saharan Africa are predicted to be out of service (Rural Water Supply Network 2013).

How will water users react if fees for a protected borehole are increased to bolster cash-on-hand? Will they reduce the amount of water collected, or switch to a lower-cost water point or even a free but polluted surface water source? Will they combine the two strategies and collect less from the improved source and use it only for drinking and cooking? A household is in fact making two inter-related decisions here: which source or sources should we collect from (source choice), and how much water should we collect (water demand)? Household water demand may depend on source choice: households who prioritize quality, but live far from an improved source, may demand less water due to the high cost of collection. Similarly, source choice may depend on household demand: households with high demand might be forced to collect from cheap sources, regardless of water quality (Whittington et al. 1990).

The same questions would apply to a rural water supply agency planning new investments in a region. It could concentrate on building relatively few new water points but heavily subsidizing them, requiring low user fees. Or it could build a dense network of new water points, bringing more improved points closer to more homes; doing so would require less subsidy per water point and higher user fees. How do households trade off the value of their time carrying water home with higher financial user fees? Hiring tap attendants would allow a source to be available during more hours of the day, but would require more in user fees. How do households value the availability of the source? Finally, in areas with plentiful surface water sources, an agency focused on meeting Sustainable Development Goals for improved basic water use might be concerned with how households value the cleaner water from improved sources *ceteris paribus*, given that they may choose to treat drinking water separately with chemical or biological means.

Many of these questions also apply to "tap vs. non-tap" choices in small towns and medium- to large-sized cities in the global South, and have been studied extensively in those contexts. These studies typically use cross-sectional household surveys, sometimes in combination with municipal billing data. Several studies have examined the source choice decision, generally finding that price, distance to source, quality and reliability are important determinants (Briscoe et al. 1981; Mu et al. 1990; Madanat and Humplick 1993; Persson 2002; Larson et al. 2006; Nauges and Strand 2007; Basani et al. 2008; Cheesman et al. 2008; Nauges and Van Den Berg 2009; Boone et al. 2011; Kremer et al. 2011; Onjala et al. 2014; Coulibaly et al. 2014; Uwera and Stage 2015; Gross and Elshiewy 2019). A smaller number of studies estimate water demand, generally finding that own-price elasticities range from -0.3 to -0.6 (Acharya and Barbier 2002; Strand and Walker 2005; Larson et al. 2006; Nauges and Strand 2007; Basani et al. 2008; Cheesman et al. 2008; Nauges and Van Den Berg 2009; Coulibaly et al. 2014; Gross and Elshiewy 2019) (see Nauges and Whittington 2010 for a helpful review). There have been surprisingly few empirical investigations in rural areas. Only four of the source choice studies have been in a rural setting (Briscoe et al. 1981; Mu et al. 1990; Kremer et al. 2011; Gross and Elshiewy 2019), where distances to water sources are typically longer, and time costs of collection may be more salient. Only one study examines water demand in rural areas (Gross and Elshiewy 2019). This is in part explained by Nauges and Whittington (2010): such studies need information on the sources *not* chosen, information not captured in large national surveys.

We captured just this type of information in a purpose-built, face-to-face household survey of 387 households in rural Meru County, Kenya. Using this data, our paper makes three contributions to the literature. First, we add to the sparse literature on how households in rural Africa choose which source to collect from. Results from a random-parameters logit model show that households are sensitive to the financial price charged per water container and the (self-reported) travel time from their house to the source, as expected. Households are also sensitive to the availability and taste of the water source, but are not sensitive to other source attributes including, color, health risk, and risk of conflict. The financial and time cost parameters of the model can be used to calculate a value of travel time. Such estimates are still rare in low- and middle-income countries (Whittington and Cook 2019); our results—a second contribution—imply that households value time spent collecting water, on average, at one-third of the unskilled wage rate.

The third contribution of the paper is to estimate water demand in a rural area, and in an innovative way. Borrowed from the recreation demand literature, we adopt the discrete-continuous (linked) demand model (Bockstael et al. 1987; Creel and Loomis 1992) to model source choice and demand. The model uses information from the source choice model to generate a "choice quality" measure that enters an OLS demand equation. As expected, we find strong effects of household size on total water demand, implying each household member increases demand by 26 liters per day, consistent with our descriptive water use statistics. We also find that households in the highest wealth quintile (based on an asset index) use almost three times more water than households in the lowest quintile. Using information from the two stages, we aggregate demand across households for each source and generate elasticity estimates. They are strikingly consistent with estimates from small and large cities in the global South as well as meta-analysis results from industrialized countries (Dalhuisen et al. 2003). Own-price elasticities range between -0.13 and -1.33, with a mean of -0.56.

The remainder of the paper is organized as follows. Section 2 describes our study site and profiles the socioeconomic characteristics of the households interviewed. Section 3 describes household water collection patterns, including summary statistics on households' perceptions of water source characteristics. In Sect. 4 we provide an overview of approaches used in modeling water demand before describing our implementation of the linked demand model. To conclude we summarize results, and discuss limitations and opportunities for future work.

2 Study Site and Household Demographics

We interviewed a total of 387 households near the small market town of Kianjai in September 2013, the dry season. Kianjai is approximately 20 miles from the city of Meru, in north-central Kenya. The study site was chosen purposefully because of the large number of existing water source options available, but households were chosen randomly based on a transect approach.[2] The study site, including a depiction of all available sources, is shown

[2] Since our target sample was 400 households and the most recent census indicated a population of 3005 households in our study site, we targeted approximately 20% of the total population, or every fifth household. In 23 sampled households, the respondents in the household were unavailable so that call backs had to be scheduled. In 15 of these 23, an interview was later completed. Six households declined to be interviewed. Therefore, of the 402 households contacted, 387 were interviewed giving a response rate of 96%.

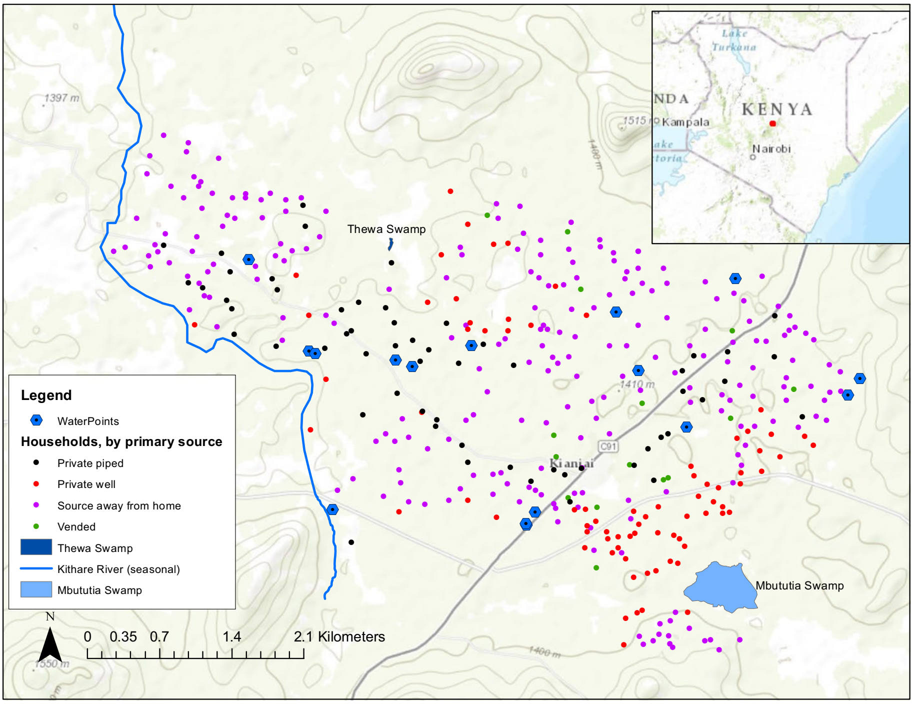

Fig. 1 Study site

in Fig. 1. Each dot on the map represents a sampled household; households are color coded by the type of their primary source. Each blue hexagon marks the location of a public water source.

A team of seven trained enumerators asked households a number of detailed questions in Kimeru (the local language). We interviewed the household member “who is mostly responsible for water-related decisions such as where to get water and how much to collect”; this person was also the person “who collected the most water in the past 7 days” in three-quarters of the cases. Seventy-nine percent of respondents were women.

Enumerators asked about household demographics and socioeconomic status (Table 1). A typical sample household has five members. The household is led by a married couple, both of whom are around 40 years old and have each completed 7 years of education. They own their house and two acres 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 18,374 Ksh or 214 USD (1 USD = 86 Ksh at the time of the survey), and average monthly food expenditure is 9283 Ksh (114 USD). The most common source of income is farming. Thirty-nine percent of households, however, had at least one household member who earned income from full-time employment, part-time or seasonal employment, or business and self-employment. Roughly 10% of households had more than one member earning income from these sources. Typical household assets include a cellphone, bicycle, radio, and livestock. Most households walk to collect water, but 29% of households report using bicycles for water

Table 1 Household demographics

	Mean (SD)
Household size	5.48 (2.19)
Number of children under 15	1.84 (1.43)
Respondent is female	0.79 (–)
Years of education of female head of household	7.23 (3.68)
Total monthly income (Ksh)	18,374 (22,233)
Monthly food expenditure (Ksh)	9283 (5668)
Household has working electricity connection	0.11 (–)
Uses a bike to collect water at least some of the time	0.29 (–)
Uses a wheelbarrow to collect water at least some of the time	0.05 (–)
Uses a cart to collect water at least some of the time	0.03 (–)
N	384

We drop three households from our sample; two households had invested in sufficiently large rain water storage to last throughout the dry season, and one household only listed one source that they could use

collection. Very few households report using a wheelbarrow (5%) or cart (3%) for water collection.

3 Water Sources and Collection Behavior

A piped distribution network operated by a formerly-public, now-private water company (Imetha Water and Sanitation Company) serves the area. The system supplied working tap connections to many households until the distribution network fell into disrepair in the 1990's and the raw water supply became over-allocated. About 10% of our sample still has a working private tap connection to the distribution network, though many of the households in our sample *without* water supply at home were once served by this system and showed us their yard taps that were no longer working. Another group of 28 households (7% of our sample) have private tap connections to what is locally called "project" water. These are self-organized, self-financed distribution networks that typically divert untreated river water. Households contribute labor and some cash for the construction and operation of these schemes. These piped systems also have connections that are each staffed by an operator and are available for the public to use. We refer to these as "public taps".

Private wells are common in the areas of the study site where groundwater is relatively accessible. About 20% of households in our sample have a private well on premises. These are almost all hand-dug wells, rather than machine-bored. Some wells were covered with sturdy metal hatches while others were covered with loose material or brush, or left completely uncovered.

Households also travel to collect water at a number of different public sources in the area (Fig. 1). Public sources included drilled boreholes, shallow wells, the public taps mentioned earlier as well as onselling from private tap connections.[3] There are also free surface water sources available in the area: a seasonal river and two natural springs/swamps where the groundwater surfaces during the wetter periods but recedes during the dry season. Water vendors are active in the area during the worst months of the dry season (July through September). We were told anecdotally that most vendors source their water from the river.

Enumerators asked households which water sources they *could* use (including those they *do* use), during both the dry season and the rainy season, though we focus here on dry season water collection only.[4] For each source, respondents reported the distance to the source (walking, one-way with a full container), and the price charged (if any). We multiply these one-way walk times by 1.75 (since one can walk faster with an empty container) to estimate roundtrip collection times reported below. Respondents assessed the taste of each source ("sweet", "normal", "poor", or "varies"), its color ("clear", "brown", "cloudy", or "varies"), and the health risk from drinking it ("no risk", "some risk", or "serious risk"). They also assessed the source's availability for collecting ("good", "fair", or "poor") and the likelihood of conflict in using it ("not likely at all", "somewhat likely", or "very likely"). Several assumptions made where data were not collected are noted in the footnotes of Table 2.

Using self-reported source attributes, rather than objective measures, incites some concern of measurement error. Households, however, make their source choices based on their perception of source attributes, not on objective measures collected by researchers. In other words, households may be poorly informed, but this poor information drives their choices. Recreation site choice models that use households' perceptions of site attributes moderately outperform models that use objectively measured site attributes (Adamowicz et al. 1997; Deely et al. 2019). Notably, we find that self-reported walk times are plausible and fairly precise.[5]

Public wells and public taps cost on average 2 Ksh, while vended water costs 10 Ksh (Table 2). Households collecting water from public taps or wells reported an average

[3] Many households reported collecting from their "neighbor's" well or private tap. Often these households reported walking significant distances to these "neighbors" and paying financial costs to collect, so we assume that many respondents were referring to the public sources just described.

[4] We do this for two reasons. First, the survey was conducted in the dry season and we have more confidence in respondents' recall of water collection choices and trips when asked about the prior 7 days, rather than representative numbers for an "average week during the rainy season" several months prior. Second, many households collect rainwater in small containers during the rainy season and rely less on collection outside the home. Ninety-six percent of households collect rain water during the rainy season; the average household uses 230 l of rain water per week (during the rainy season). Although this seasonality is important for water supply planning, during the rainy season it is more difficult to observe households making the sorts of tradeoffs that help us identify their preferences.

[5] Based on the set of household-source pairs where we collected GPS locations as well as self-reported one-way walk times (we did not anticipate so many households to report collecting from their "neighbors", and did not take GPS measurements at these sources), we estimate an implied walking speed of 1.7 miles per hour (results available on request). This is in line with other estimates (White et al. 1972; Calvo 1994; Tanser et al. 2006), suggesting the household's self-reported walk times are plausible. We also calculate the difference in a household's reported walk time to a given public source and their nearest neighbor's reported walk time to the same source (we thank a referee for this suggestion). Among the 211 observations for which a household and their nearest neighbor reported having access to the same public source, the average difference (absolute value) in their reported walk times is 7.68 min (median: 2.91, SD 19.64). This represents on average a 12% difference in reported walk times between nearest neighbors (median: 5%), suggesting self-reported walk times are fairly precise.

Table 2 Household reported source characteristics in the dry season, grouped by source type

	Private well	Private tap	Public well	Public tap	Vendor	Surface	Overall
Price (Ksh)	0.00	0.00	2.13	2.57	10.45	0.03	3.78
	(0.00)	(0.00)	(1.66)	(2.02)	(3.43)	(0.27)	(4.42)
Walk time (mins.)[a]	0.57	0.00	43.14	48.75	0.00	96.89	31.35
	(0.54)	(0.00)	(28.62)	(46.91)	(0.00)	(71.37)	(41.28)
Wait time (mins.)[b]	1.00	1.00	57.61	52.65	1.00	31.00	33.97
	(0.00)	(0.00)	(49.68)	(48.55)	(0.00)	(50.52)	(47.18)
Color: clear	0.80	0.83	0.69	0.87	0.55	0.56	0.68
Color: brown, cloudy or varies	0.20	0.17	0.31	0.13	0.45	0.44	0.32
Health risk: none	0.32	0.39	0.36	0.31	0.22	0.29	0.32
Health risk: some	0.43	0.49	0.44	0.61	0.51	0.29	0.47
Health risk: serious	0.25	0.12	0.19	0.08	0.27	0.41	0.21
Taste: sweet	0.05	0.11	0.10	0.14	0.03	0.08	0.08
Taste: normal or varies	0.64	0.71	0.60	0.75	0.66	0.49	0.63
Taste: poor	0.32	0.18	0.30	0.11	0.32	0.42	0.28
Conflict: not likely[c]	1.00	1.00	0.39	0.41	1.00	0.54	0.63
Conflict: somewhat likely	0.00	0.00	0.26	0.44	0.00	0.29	0.19
Conflict: very likely	0.00	0.00	0.35	0.14	0.00	0.16	0.18
Availability: good	0.39	0.16	0.78	0.50	0.13	0.98	0.54
Availability: fair	0.36	0.46	0.20	0.44	0.27	0.02	0.26
Availability: poor	0.25	0.38	0.02	0.05	0.60	0.00	0.20
N	88	76	562	147	307	85	1267

Standard deviations are in parenthesis. *N* is the number of times households reported having access to a source of that type. Some households reported having access to more than one source of the same source type, hence $N = 562$ for public wells because households often reported being able to use more than one public well

[a]Vendor one-way walk times are assumed to be zero. Private well one-way walk times are calculated using a reported distance and an estimated average walking speed

[b]Wait and fill times are assumed to be 1 min for private taps, private wells, and vended water

[c]Conflict is assumed to be "not likely" for private and vendor sources

roundtrip walk time of 45 min, and a wait time of 55 min. Surface and vended water is more likely to be rated as "brown", "cloudy", or "varies". Public and private taps were rated the best tasting sources, while surface sources were rated the worst. Sixty-one percent of respondents with working private tap connections thought that drinking water from their tap connection to the piped network posed at least some health risk, roughly similar to households' perceptions of public tap connections to the piped network. For public taps and wells, approximately sixty percent of respondents said using the source was somewhat or very likely to cause conflict. Public wells were rated highly for availability, while vended water was rated poor.

The survey also asked about households' water collection behaviors (Table 3). The average number of sources that households reported they could use is 3.3 sources (median 4, max 6). Households reported that they had actually used an average of 1.4 sources in the past week, and 2.0 sources in the past year. Households' primary source is defined as the source from which they collected the most water in the past week. Thirty-seven percent of households in

Table 3 Water collection from all sources, organized by the household's primary source

	Private well	Private tap	Public well	Public tap	Vendor	Surface	Overall
Weekly water expenditure (Ksh)	39	18	135	85	546	45	126
	(257)	(61)	(147)	(95)	(454)	(90)	(247)
Weekly time spent collecting (minutes)	252	357	5496	2691	1730	3515	2950
	(554)	(969)	(6571)	(3258)	(3439)	(1186)	(5151)
# of sources households said they could use	3.39	3.79	3.08	3.46	3.12	2.75	3.30
	(1.07)	(0.89)	(0.84)	(0.87)	(0.95)	(0.50)	(0.94)
# of sources households used in past week	1.09	1.27	1.47	1.32	1.68	1.75	1.37
	(0.33)	(0.48)	(0.67)	(0.53)	(0.47)	(0.96)	(0.59)
# of sources households used in past year	1.59	1.64	2.14	2.38	2.29	2.25	1.98
	(0.73)	(0.82)	(0.87)	(0.76)	(0.76)	(0.50)	(0.86)
Total liters collected in the past week	2160	1501	1112	861	1190	815	1366
	(1844)	(1024)	(661.6)	(463)	(1055)	(190)	(1156)
Total 20 l jerricans collected in the past week	108	75	56	43	59	41	68
	(92)	(51)	(33)	(23)	(53)	(9.5)	(58)
% of water collected from their primary source	0.99	0.94	0.88	0.93	0.83	0.88	0.91
	(0.07)	(0.12)	(0.19)	(0.13)	(0.16)	(0.21)	(0.16)
N	76	66	167	37	34	4	384

Standard deviations are in parenthesis. *N* refers to the number of households that reported that this type of source was their primary source. For example, 76 households said a private well was their primary source

our sample have their primary source at home: either private tap connections (17%), or private wells (20%). Ten percent of households reported using a public tap as their primary source; 43% a public well. Nine percent of households reported vended water as their primary source, and the remaining one percent of households use surface water as their primary source.

Households collect on average 1365 liters per week, or 38 liters per capita per day. Households whose primary source is at home consume on average 52 liters per capita per day, while households whose primary source is away from home consume on average 31 liters per capita per day. Households spend an average of 49 hours per week collecting water. Households whose primary source is at home spend on average 5 hours per week collecting water, while households whose primary source is away from home spend on average 75 hours per week collecting water, or 41 hours per week per water collector (households whose primary source is away from home have on average 2 water collectors). Ten households were dropped from analysis because their reported total daily collection time exceeded a plausible maximum of 8 hours per water collector per day.

Sixty-nine percent of households collect all of their total water demand from their primary source, and households collect on average 91% of their total water demand from their primary source. A subset of our sample does, however, collect from more than one source to meet their weekly demand; approximately one sixth of households collect at least 15% of their total water demand from secondary sources.

Households may collect water from different sources to serve different purposes (Nauges and Whittington 2010), so we asked which water source the household primarily uses for different purposes during the dry season, including drinking, washing around the house, cooking, bathing/personal hygiene, watering animals, and other productive activities. The vast majority (97.2%) of households reported the same primary water source for all purposes, indicating that most households rely primarily on one source and use others as occasional or back-up sources.

4 Model of Water Collection

The majority of water source choice and demand studies use a Heckman-style two-step model (Heckman 1979; Shonkwiler and Yen 1999) to estimate conditional demand equations (Larson et al. 2006; Basani et al. 2008; Cheesman et al. 2008; Nauges and Van Den Berg 2009; Uwera and Stage 2015; Gross and Elshiewy 2019). First, a source choice model is used to estimate the likelihood that a household collects from a given source. Then, conditional on positive collection, demand from the given source is estimated using OLS (one conditional demand equation for each source alternative).[6] Since observing positive collection from a source is non-random, a bias correction parameter from the source choice model is included in each of the conditional demand equations.

To make estimation feasible, implementations of the two-step model typically aggregate individual sources into source-types (private tap, public well, surface water, etc.). Aggregating individual sources into source-types restricts substitution between sources of the same source-type to zero, which may induce bias.[7] This two-step model also imposes assumptions on households' water collection behaviors and their decision process: (1) the model assumes that all collection is from a single source (source-type), and (2) the model assumes that households first choose where to collect from, and then decide how much to collect.

We introduce an alternative model that does not require aggregation into source-types, and does not limit source choice to a single source. This model also assumes a different household decision process. Borrowed from the recreation demand literature, the linked demand model decomposes source choice and demand into a two-stage decision process (Bockstael et al. 1987). In the first stage, households make the macro decision of how many collection trips to make in a given week. In the second stage, households make the micro decision of how to allocate their water demand across their choice set of alternative sources. (This decision process is similar to that of the popular Almost Ideal Demand System (Deaton and Muellbauer 1980; Heien and Wessells 1990; Coulibaly et al. 2014).) The fact that source choice and household demand are likely to be interrelated is explicitly accounted for through flexible source choice modeling and a linking function included in the household demand equation.

Total household demand, q_i, is measured by the total number of 20 l collection trips made by household i in a given week, and is a function of household characteristics, H_i, choice set quality, $\delta_i(\hat{\beta}_i, X_i)$, and a random error, ω_i:

[6] In practice, conditional demand equations have sometimes been misspecified in the water source choice and demand literature. Using either the Lee correction (Lee 1983) or Dubin-McFadden correction (Dubin and McFadden 1984) methods, analysts should estimate one conditional demand equation for each choice alternative (see original papers, or Mannering (1986), or Wu and Babcock (1998) for examples).

[7] Researchers in the recreational demand literature face similar issues. Here the discrete-continuous choices are which site to visit and how many visits to make. Suppose sites were fishing destinations, and the choice set included a salt-water site and three freshwater lakes. Data aggregation would collapse the choice set to a salt-water site and a freshwater lake, and fail to make use of any information on heterogeneity in lake characteristics like fish stocking, boat ramps, or water quality.

$$q_i = \gamma H_i + \eta \delta_i(\hat{\beta}_i, X_i) + \omega_i. \tag{1}$$

Choice set quality is a function of source attributes, X_i, for each source in the household's choice set, and the household's preferences over those attributes, $\hat{\beta}_i$. Including choice set quality in the household demand equation links the source choice and household demand decisions, and explicitly controls for the effect of source choice on household demand (Madanat and Humplick 1993).

Household source choice is modeled using random utility theory. The indirect utility, V_{ij}, of household i collecting from source j on any given collection trip is a function of source attributes, X_{ij}, and the household's preferences over those attributes, β_i. Characteristics of the household, which may influence the choice of sources, enter the indirect utility function through the vector Z_i and corresponding taste parameter vector ν_i. Remaining factors that are unobservable to the researcher, but are known to respondents, are in the error term ϵ_{ij}. Making the standard assumption that these components are additive and separable (Haab and McConnell 2002; Nauges and Whittington 2010) yields the following indirect utility function:

$$V_{ij} = \beta_i X_{ij} + \nu_i Z_i + \epsilon_{ij}. \tag{2}$$

Combining results from the source choice and household demand models allows us to construct an aggregate demand function for each source in our sample. Results from the source choice model yield estimates of the probability, $\hat{Pr}_{ij}$, that source j is household i's primary source; $\hat{Pr}_{ij}$ can be interpreted as the share of trips taken by household i allocated to source j (Bockstael et al. 1987). Results from the household demand equation yield estimates of the total number of collection trips made by household i in a given week, $\hat{q}_i$. The product of $\hat{Pr}_{ij}$ and $\hat{q}_i$ represents the predicted share (in 20 l collection trips) of household i's total demand allocated to source j in a given week. Then, predicted aggregate demand at source j, $\hat{Q}_j$, is the sum of $\hat{Pr}_{ij} \times \hat{q}_i$ across households: $\hat{Q}_j = \sum_{i=1}^{N} \hat{Pr}_{ij} \times \hat{q}_i$. This aggregate demand function is used to calculate elasticities.

4.1 Estimating Source Choice

We estimate households' choice of primary source using a random parameters logit (Revelt and Train 1998; McFadden and Train 2000). The random parameters logit allows for heterogeneity among households' preferences, and relaxes the restrictive independence of irrelevant alternatives assumption. By allowing preferences to vary across households, we flexibly allow household demand (and other household characteristics) to affect household preferences and source choice (e.g. individual parameters allow households with high demand to be potentially more sensitive to price and distance, but less sensitive to quality).

Household preferences are allowed to vary across households, but are assumed to follow an analyst-specified distribution, $f(\beta|\Omega)$ (sometimes called the mixing distribution), where Ω represents the mean and standard deviation of the mixing distribution to be estimated. Commonly used mixing distributions include the normal and lognormal distributions (Hensher and Greene 2003). We assume a lognormal distribution for the price coefficient to ensure our model does not predict illogical preferences for higher prices. All other random parameters are assumed to be normally distributed. Due to computational burden, parameters on source-type dummies are not allowed to vary across individuals, and are instead restricted to population-level coefficients.[8]

In our data, each household i has a unique choice set given by J_i. Existing studies have generally structured choices *a priori* in the way that households were asked about sources: researchers asked a household about the nearest kiosk, the nearest public well, surface source, etc. Our approach was open-ended and simply asked households which sources they could use, and which they do use. We categorize them *ex post*, and use the choice set exactly as reported by households.

Each source j has observable (self-reported) source attributes in the vector X_{ij} that provide utility or disutility from using the source, including health risk, taste, color, risk of conflict, source type, and the full cost of collection. Household characteristics, which do not vary across choice alternatives, fall out of the estimation unless interacted with attributes that do vary.

In a typical travel cost model, the full cost of collection is given by $FC_{ij} = P_{ij} + \psi_i T_{ij}$, where P_{ij} is the financial cost per collection trip, T_{ij} is the sum of travel time and wait time per collection trip, and ψ_i is household i's shadow value of time. Often analysts assume a shadow value of time, but we allow it to be estimated by the model. We also allow for unique shadow values of time for walking and waiting given by ψ_i^{walk} and ψ_i^{wait}; households might value their time spent walking and waiting differently, especially considering the disutility associated with carrying 20 kg of water. The revised full cost of collection is given by $FC_{ij} = P_{ij} + \psi_i^{walk} walk_{ij} + \psi_i^{wait} wait_{ij}$.

The probability that household i collects from source j on any collection trip is given by:

$$Pr_{ij} = \int_{-\infty}^{\infty} \frac{e^{V_{ij}}}{\sum_{k \in J_i} e^{V_{ik}}} f(\beta|\Omega) d\beta \tag{3}$$

Given source attributes, the households' choice of primary source, and the above probability statement, we can estimate the mean and standard deviation of household preferences (Table 4).

The reported estimates in Table 4 are the estimated population means and standard deviations of individual household preferences (Ω in $f(\beta|\Omega)$). The mean coefficients on price and walk time are both negative and significant: households prefer closer and cheaper sources. Households also prefer sources that taste "sweet" to those that taste "normal or variable", or "poor", and sources that have "fair" or "good" availability to those with "poor" availability. Households' perceptions of other quality attributes do not seem to affect their choice of primary source, a result we return to in the conclusions.[9] Estimated standard deviations of the mixing distributions suggest households have heterogeneous

[8] We failed to achieve convergence when we allowed coefficients on the source type dummies to be random parameters.

[9] One explanation might be that perceived quality variables like taste, color and health risk are highly correlated, resulting in poor parameter estimates. We estimated several models that controlled for the modest correlation in our data (less than 0.5) between these quality variables. We also estimated models that added correlated quality variables one by one and in combinations. Finally, we estimated a model using a water quality index generated by principal component analysis. Results are generally consistent with our main model and are available upon request. Another concern might be that source-type dummies absorb the effects of quality attributes if there is little variation of quality attributes within source-types, though it is apparent from the raw results in Table 2 that there is substantial variation in source attributes within source-types.

Table 4 Primary source choice model (random parameters logit)

	Mean	Standard Deviation
Price	−0.37** (0.15)	0.54*** (0.20)
Walk time	− 0.06*** (0.02)	0.003 (0.01)
Wait time	0.01 (0.01)	0.001 (0.004)
Color: clear	0.61 (0.64)	2.73*** (0.73)
Health risk: none	− 0.44 (0.54)	0.81 (0.87)
Health risk: serious	− 0.89 (0.77)	2.28*** (0.87)
Taste: sweet	1.40** (0.65)	0.77 (1.30)
Taste: poor	0.19 (0.53)	0.84 (0.70)
Conflict: not likely	1.12 (0.70)	2.19** (1.05)
Conflict: very likely	0.16 (0.73)	2.41*** (0.90)
Availability: good	0.10 (0.60)	1.42 (1.63)
Availability: poor	− 1.98** (0.92)	3.27** (1.40)
Private well	7.65*** (1.96)	– (–)
Private tap	7.89*** (2.19)	– (–)
Public well	3.99*** (1.33)	– (–)
Public tap	2.95** (1.31)	– (–)
Vendor	1.87 (1.71)	– (–)

Number of observations: 1206; Number of households: 366; Log-likelihood: −224.30; AIC: 506.59; BIC: 654.35

*p value < .10; **p value < .05; ***p value < .01. Standard errors are in parenthesis. Eight additional households were dropped from estimation because they had missing source attributes, hence number of households = 366

preferences over price, "clear" color, "serious" health risk, "poor" availability, and conflict (standard deviations statistically different from zero).

The shadow value of walk time, in Kenyan shillings per hour, is given by the ratio of the price and walk time coefficients, or $60 \times \frac{\hat{\beta}_i^{walk}}{\hat{\beta}_i^{price}}$. The population mean for the value of travel (walk) time (VTT) is 11 Ksh/h, and the upper and lower 5-percentiles of the empirical distribution of individual VTT estimates are 6.30 and 23.24 Ksh/h.[10] This mean value of travel time estimate is approximately one-third of the local unskilled wage rate of 35 Ksh/h, and is lower than those found in a companion paper that uses responses to a hypothetical choice among new water sources. That paper found a mean estimate of 18 Ksh/h using a random parameters logit model (Cook et al. 2016).

Since the coefficient on wait time is not statistically significant, the estimated shadow value of time spent waiting is also zero. The disutility of waiting in line (the opportunity cost) might be offset by the ability to rest or socialize with others. It may also be the case

[10] The reported coefficients for the random parameters logit are the mean of each individual parameter estimate. Note that the mean shadow value of walk time is given by $60 \times mean\left(\frac{\hat{\psi}_i^{walk}}{\hat{\beta}_i^{price}}\right)$ which is not the same as $60 \times \frac{mean(\hat{\psi}_i^{walk})}{mean(\hat{\beta}_i^{price})}$.

that collectors can hold their place in line with their jerrican, allowing them to do other tasks and come back. We did not ask how water queues at the site work.

Dummies for the type of water source, with the exception of vended water, are positive and statistically significant: households are more likely to choose these sources over a surface water source *ceteris paribus*. Wald tests indicate that households prefer private taps and wells over public taps and wells (p values < 0.01), are indifferent between private taps and private wells (p value = 0.84), and prefer public wells over public taps (p value = 0.04). Vended and surface water are the least-preferred source types after controlling for observable attributes like price and walk time; households are willing to pay an additional 9.5 Ksh (or walk an additional 52 min) to avoid collecting from a surface source and collect from a public tap instead.

4.2 Estimating Household Demand

As described above, household demand is a function of household characteristics and choice set quality. Although a number of choice set quality measures have been used (Herriges et al. 1999; Phaneuf et al. 2005), the most common is the maximum expected utility of a trip (Hanemann 1982; Bockstael et al. 1987):

$$E[V_i] = ln\left(\sum_{j \in J_i} e^{\hat{\beta}_i X_{ij}}\right) + C = \delta_i(\hat{\beta}_i, X_i) + C. \tag{4}$$

In words, the maximum expected utility of a trip is the sum of the utility obtained from visiting each source weighted by the probability of visiting that source (Creel and Loomis 1992). The intuition is that households with a higher choice set quality (i.e. cheaper, closer, sweeter, and more available sources) may collect more water. Using estimates from the site choice model, we calculate $\delta_i(\hat{\beta}_i, X_i)$ for each household.

This choice set quality measure can then be included directly as a regressor in the OLS household demand function of water collection trips. Choice set quality may, however, be simultaneously determined with household demand. For example, a household that wants to collect more water and have better control over water quality at the source might choose to install a private well, in turn affecting choice set quality, in which case quality would be endogenous in Eq. 1. We therefore instrument for choice set quality using the choice set quality of the household's nearest neighbor, and dummies for sublocation (neighborhood). These instruments are correlated with the household's own choice set quality through locational characteristics (the accessibility of the piped network, the depth of surface water, and the quality and proximity of public sources), but are otherwise unrelated to household demand. The first stage results are presented in Table 7 in the "Appendix". Both nearest neighbor's choice set quality, and sublocation dummies, are statistically significant predictors of a household's own choice set quality. The effective F-statistic (Olea and Pflueger 2013) is 25.03, which means we can reject the null of weak instruments.

As household size increases by one member, households make on average 9 additional collection trips per week (Table 5). This corresponds to 180 liters per week, or 26 liters per day. This marginal increase is plausible and consistent with WHO minimum recommendations for drinking, cooking, and some personal washing (World Health Organization 2013). Because household members share the total water collected and because there are likely economies of scale in the use of water for cooking and cleaning, the average water use for an additional household member will be larger than this marginal increase for

Table 5 Total household demand model

	Trips (q_i)
Choice set quality (Eq. 4)	4.50** (1.89)
Household size	8.85* (4.87)
Household size squared	− 0.04 (0.36)
Number of children under 15	− 3.28 (2.44)
Wealth index	7.00*** (1.89)
Constant	1.07 (19.41)
R-squared	0.22
N	366

*p value< .10; **p value< .05; ***p value< .01. Standard are errors in parenthesis. Since choice set quality is a predicted regressor, we use corrected errors following Murphy and Topel (1985) (IV approach)

the household. Wealthier households, as proxied by our PCA-derived wealth index, collect more water: the bottom wealth quintile collects 114 liters per day, the middle quintile collects 169 liters per day and the top quintile collects 300 liters per day.[11] As expected, the coefficient on choice set quality is positive and statistically significant: households with higher quality choice sets (cheaper, closer, sweeter, and more available sources) consume more water than households with low quality choice sets.

4.3 Aggregate Demand

With both stages now estimated, predicted aggregate demand at source j is given by:[12]

$$\hat{Q}_j = \sum_{i=1}^{N} \hat{Pr}_{ij} \times \hat{q}_i = \sum_{i=1}^{N} \frac{e^{\hat{V}_{ij}}}{\sum_{k \in J_i} e^{\hat{V}_{ik}}} \times \left(\hat{\gamma} H_i + \hat{\eta} \delta_i(\hat{\beta}_i, X_i) \right). \tag{5}$$

$\hat{Pr}_{ij}$ is the predicted share of trips that household i allocates to source j (Eq. 3) and $\hat{q}_i$ is predicted household demand (Eq. 1). The inner product represents household i's allocation of total demand to source j. Both $\hat{Pr}_{ij}$ and $\hat{q}_i$ are functions of price (and other source attributes), and changes in price will affect aggregate demand along both margins. An increase in price at a particular source will decrease the likelihood that households will collect from that source ($\frac{\partial \hat{Pr}_{ij}}{\partial Price_{ij}} \leq 0$). An increase in price will also decrease total household demand ($\frac{\partial \hat{q}_i}{\partial Price_{ij}} = \eta \frac{\partial \delta_i(\hat{\beta}_i, X_i)}{\partial Price_{ij}} \leq 0$).[13]

Using Eq. 5 we can calculate own-price elasticity estimates (Table 6).[14] Average own-price elasticities range from − 1.33 at Nkomo kwa Gerald to − 0.13 at the Nchoro boreholes

[11] The wealth index is calculated following Filmer et al. (2001) and Filmer and Scott (2012). It includes data on durable assets, electricity connection, sanitation, number of rooms, number of buildings, and main cooking fuel. More information on construction of the wealth index is available on request.

[12] $E[Pr_{ij} \times q_i] = \hat{Pr}_{ij} \times \hat{q}_i$ if $Cov(Pr_{ij}, q_i) = 0$. This can be tested empirically: in our sample we find $Cov(Pr_{ij}, q_i) = -0.05$, which is nominal compared to $\hat{Pr}_{ij} \times \hat{q}_i$ (mean: 20.88).

[13] Since changes in attributes of sources not included in the household's choice set are irrelevant to the household's source choice and demand decision, these inequalities are not strict.

[14] $\epsilon_j = \sum_{i=1}^{N} \left(\frac{\partial}{\partial P_j} \left(\hat{Pr}_{ij} \right) \times \hat{q}_i + \hat{Pr}_{ij} \times \frac{\partial}{\partial P_j} \left(\hat{q}_i \right) \right) \times \frac{P_j}{Q_j} = \sum_{i=1}^{N} \left(\beta_i^{price} Pr_{ij} \left(\eta Pr_{ij} + q_i (1 - Pr_{ij}) \right) \right) \times \frac{P_j}{Q_j}$

Table 6 Own-price elasticities by source

	Own-price elasticity	5th percentile[a]	95th percentile
Vended water	– 0.50	– 0.52	– 0.27
Neighbor's well	– 0.17	– 0.17	– 0.07
Neighbor's borehole	– 0.19	– 0.24	– 0.12
Neighbor's tap connection	– 0.37	– 0.35	– 0.18
Kianjai borehole	– 0.31	– 0.43	– 0.21
Nchoro boreholes	– 0.13	– 0.20	– 0.07
Nchoro kwa murugu	– 0.29	– 0.41	– 0.16
Nkomo kwa Gerald	– 1.33	– 2.59	– 0.88
Kithare River	– 0.97	– 1.16	0[b]
Mbuya Lifelink/Redcross	– 0.40	– 1.52	– 0.13
Dairy farm borehole	– 0.84	– 2.57	– 0.66
Lubunu MCK Compassion	– 0.84	– 1.11	0
Rehema polytechnic	– 0.32	– 0.85	– 0.11
Nkomo group project	– 0.67	– 1.50	0
Kirindine well	– 1.21	–[c]	–
Machako tap	– 0.34	– 0.68	– 0.23
Nkundi private wells	– 0.28	– 0.54	– 0.16
Kambeeria water project	– 0.36	– 1.25	– 0.23
Mituntu Karithiria tap water	– 1.10	–	–
Overall	– 0.56	– 0.91	– 0.34

[a]Percentile estimates are calculated from 1000 bootstrap samples. We present non-centered standard errors because the distribution of elasticity estimates is negative and skewed to the left (as observed from the bootstrap samples)

[b]For some bootstrap samples no households were observed using some sources, resulting in an elasticity estimate of zero

[c]Some intervals are missing because the source had too few users to generate reliable bootstrap estimates

with an average of – 0.56. Average own-price elasticities among common source types are: – 0.50 for public taps, – 0.47 for public wells, and – 0.50 for vended water.

These are the first own-price elasticity estimates for public sources in rural areas of middle- or low-income countries calculated at the level of individual sources rather than aggregated source types. Gross and Elshiewy (2019) find more inelastic own-price elasticities of source-types in rural Benin: – 0.11 for public taps, – 0.10 for public wells with hand/foot pumps, and – 0.10 for protected wells. Our results are consistent with estimates in urban areas of middle- and low-income countries: Strand and Walker (2005) estimate own-price elasticities of – 0.3 for households with private tap connections and – 0.1 for households without tap connections. Nauges and Strand (2007) estimate own-price elasticities at – 0.58 for private connections, – 0.66 for public connections, and – 0.41 for tanker water. Nauges and Van Den Berg (2009) estimate own-price elasticities at – 0.15 for households that rely exclusively on a private tap connection, and – 0.37 for households that supplement their demand from their tap connection with water from public sources. Coulibaly et al. (2014), however, finds more elastic own-price elasticities using an Almost Ideal Demand system: – 1.33 for public sources, – 2.90 for tanker water, – 1.43 for treated water, and – 0.62 for bottled water. Our estimates are also roughly consistent with the central estimate of – 0.4

found in a meta-analysis of 124 price elasticity estimates of residential water in the United States (Espey et al. 1997; Dalhuisen et al. 2003). Inelastic own-price estimates suggest that service providers could increase revenues, and thus the financial sustainability of existing water sources, by raising prices. This is true for 16 of the 19 sources that charge a positive price in our sample.

5 Discussion

To expand and maintain rural water supply, planners and policymakers need information on the preferences of households who might use the sources. Results from our source choice analysis show households, as expected, prefer sources that are cheaper and have lower roundtrip walk times. They prefer private sources to public ones, public wells to public taps, and all sources to vended and surface water. Households dislike sources with "poor" availability, but are indifferent between sources with "fair" and "good" availability. They prefer water that tastes "sweet" to water that tastes "normal or varies" and water that tastes "poor". Households on average are indifferent about the level of health risk, risk of conflict, the color of the water, and wait time.

Our estimate of the value of travel time spent collecting water is somewhat lower than a commonly-used benchmark of one-half of the after-tax wage rate (Boardman et al. 2018; Von Wartburg and Waters 2005). Whittington and Cook (2019) surveyed eleven empirical time valuation studies from low-income countries. The studies surveyed used stated preference or revealed preference methods, and valued time savings in transportation, water collection, and health behavior. Whittington and Cook (2019) find support for the 50% benchmark: nine of the eleven studies report mean estimates that fall in the range of 25–75% of income or wages. Our estimation is, however, in line with a similar "rule of thumb" used in the recreation demand literature that values time savings at one-third of wages (Phaneuf and Requate 2017).

Results from our source-level aggregate demand analysis can inform rural water planners about water demand at each source, revenues, and financial sustainability. For each source we estimated own-price elasticities. Own-price elasticities estimates range from -1.33 to -0.13, with an average of -0.56. To explore the variation in our elasticity estimates across sources, we computed the correlation between elasticity estimates and variables of interest (price, quality, wealth of households, etc.)(Tables 8, 9). We caution readers that these correlations are calculated across only 19 sources, too few to reliably comment on any statistical relationships. Five variables have a correlation coefficient of at least 0.3 in absolute value: price ($\rho = -0.30$), color ($\rho = 0.32$), taste ($\rho = 0.34$), health risk ($\rho = 0.63$), and the number of substitutes ($\rho = -0.44$). These correlations imply that demand is more inelastic at sources that are priced low, and (with the exception of taste) are of good quality. Demand is also more inelastic at sources with few substitutes. Because we estimate demand to be inelastic at most sources, our results show that water managers can increase revenues by increasing price for most sources in our study site, though we recognize the political difficulties in doing so.

Our focus was not on the health impacts of improved water supply, and we did not measure household health outcomes, though our results may still have implications for household health. Simple univariate statistics and results from our linked water demand model both show that households with higher quality choice sets (primarily closer and cheaper sources) collect more water. Studies on water, sanitation, and hygiene show that increased water consumption is an important driver of household health (see Stelmach and Clasen 2015 for review).

One surprising result from our source choice model is that households were not sensitive to the perceived health risk of drinking low quality water. Most studies have found quality matters in source choice, though Kremer et al. (2011) found willingness to pay for quality to be surprisingly low. Our results need not imply that households do not value clean drinking water. Rather, it may be that point-of-use treatment (boiling, chlorinating, filtering, etc.) is more cost-effective for households than walking farther or paying more, and may reduce the risk of recontamination in transport containers. Brouwer et al. (2015) finds that households in rural Kenya are willing to pay 2.5% of their disposable income for a new drinking water filter technology. In our data, 57% of households in our sample reported treating their drinking water (primarily by boiling) at least some of the time, though a possible pro-social bias makes us cautious to over-interpret this data. In this respect, one might think of water collection decisions as three-part: (1) how much water to collect, (2) which source to collect from, and (3) whether to treat the water for specific types of uses. Future research should consider each of these parts to better characterize households' water collection decisions.

Acknowledgements We thank Annalise Blum, Josephine Gatua, Mark Mwiti, and John Wainana for valuable assistance in the field and in data analysis. We also thank Dale Whittington, Celine Nauges, and two anonymous reviewers for helpful comments and suggestions. Funding for the project was provided by https://www.efdinitiative.org/kenya Environment for Development-Kenya with support from the Swedish International Development Cooperation Agency.

Appendix: Materials

See Tables 7, 8 and 9.

Table 7 First stage results for household demand model

	Household choice set quality
Nearest neighbor's choice set quality	0.12** (0.05)
Sublocation: Nairiri	− 3.57*** (0.57)
Sublocation: Kianjai	− 0.65 (0.48)
Sublocation: Mutionjuri	− 2.37*** (0.49)
Household size	0.06 (0.26)
Household size squared	− 0.02 (0.02)
Number of children under 15	− 0.01 (0.13)
Wealth index	0.66*** (0.08)
Constant	7.03*** (0.93)
R-squared	0.32
N	366

*p value< .10; **p value< .05; ***p value< .01. Standard are errors in parenthesis, and do not correct for the two-stage nature of the choice set quality estimates

Table 8 Correlation between average source attributes, and estimated own-price elasticity

	Elasticity	Price[a]	Walk time	Wait time	Risk of conflict: some or serious	Availability: poor	Availability: good
Vended water	– 0.50	10.0	0.00	1.00	0.00	0.59	0.14
Neighbour's well	– 0.17	2.00	32.01	51.12	0.67	0.03	0.76
Neighbour's borehole	– 0.19	2.00	31.23	48.00	0.50	0.04	0.68
Neighbour's piped connection	– 0.37	2.00	22.15	39.55	0.53	0.05	0.35
Kianjai borehole	– 0.31	2.50	55.07	61.92	0.63	0.00	0.73
Nchoro boreholes	– 0.13	2.00	49.84	62.97	0.57	0.03	0.70
Nchoro kwa murugu	– 0.29	1.00	43.22	36.25	0.75	0.00	1.00
Nkomo kwa Gerald	– 1.33	2.00	48.46	76.15	0.54	0.00	0.92
Kithare River	– 0.97	2.50	210.00	60.00	0.00	0.00	1.00
Mbuya Lifelink/Redcross	– 0.40	2.00	36.47	39.40	0.17	0.00	1.00
Dairy farm borehole	– 0.84	4.50	75.84	54.17	0.75	0.00	0.25
Lubunu MCK Compassion	– 0.84	2.50	48.14	10.00	1.00	0.00	1.00
Rehema polytechnic	– 0.32	2.00	30.41	110.00	1.00	0.00	0.67
Nkomo group project	– 0.67	2.25	52.50	50.00	0.50	0.00	1.00
Kirindine well	– 1.21	2.00	70.00	5.00	0.00	0.00	1.00
Machako Tap	– 0.34	2.50	81.00	72.30	0.75	0.02	0.65
Nkundi private wells	– 0.28	2.50	69.44	87.40	0.81	0.03	0.89
Kambeeria water project	– 0.36	1.50	170.63	127.50	1.00	0.25	0.50
Mituntu Karithiria tap water	– 1.10	5.00	52.50	30.00	1.00	0.00	0.00
Correlation coefficient		– 0.30	– 0.02	0.21	0.09	0.08	0.19

Each column lists source-level average attributes for each household that said they could use the source

[a]The reported price for each source is the median price reported among households who said they could use the source

Table 9 Correlation between average source attributes, and estimated own-price elasticity

	Elasticity	Taste: sweet	Taste: poor	Color: clear	Health risk: some or serious	Wealth	# of substitutes[a]
Vended water	− 0.50	0.03	0.31	0.45	0.77	− 0.00	2.49
Neighbour's well	− 0.17	0.02	0.44	0.47	0.85	0.07	2.53
Neighbour's borehole	− 0.19	0.18	0.32	0.25	0.58	− 0.24	2.48
Neighbour's piped connection	− 0.37	0.24	0.05	0.09	0.64	− 0.14	2.67
Kianjai borehole	− 0.31	0.21	0.11	0.07	0.25	0.19	2.91
Nchoro boreholes	− 0.13	0.10	0.17	0.27	0.40	− 0.33	2.90
Nchoro kwa murugu	− 0.29	0.00	0.75	1.00	1.00	− 0.63	2.75
Nkomo kwa Gerald	− 1.33	0.23	0.15	0.08	0.23	− 0.31	3.15
Kithare River	− 0.97	1.00	0.00	0.00	0.00	− 1.48	4.00
Mbuya Lifelink/Redcross	− 0.40	0.67	0.00	0.00	0.00	− 0.13	3.17
Dairy farm borehole	− 0.84	0.00	0.25	0.25	0.50	1.41	3.13
Lubunu MCK Compassion	− 0.84	0.00	1.00	1.00	1.00	− 1.77	2.00
Rehema polytechnic	− 0.32	0.00	0.67	0.33	0.67	− 0.12	3.00
Nkomo group project	− 0.67	0.00	1.00	0.00	0.50	− 0.88	2.50
Kirindine well	− 1.21	0.00	0.00	0.00	0.00	0.31	4.00
Machako Tap	− 0.34	0.02	0.18	0.18	0.86	0.11	2.47
Nkundi private wells	− 0.28	0.01	0.32	0.35	0.88	0.22	2.25
Kambeeria water project	− 0.36	0.00	0.00	0.00	0.75	0.12	2.50
Mituntu Karithiria tap water	− 1.10	0.00	0.00	0.00	0.00	− 0.70	4.00
Correlation coefficient		0.11	0.34	0.32	0.63	− 0.09	− 0.44

Each column lists source-level average attributes for each household that said they could use the source

[a]Number of substitutes is defined by the average number of sources households said they could use minus one

References

Acharya G, Barbier E (2002) Using domestic water analysis to value groundwater recharge in the Hadejia–Jama' are floodplain, Northern Nigeria. Am J Agric Econ 84(2):415–426

Adamowicz W, Swait J, Boxall P, Louviere J, Williams M (1997) Perceptions versus objective measures of environmental quality in combined revealed and stated preference models of environmental valuation. J Environ Econ Manag 32(1):65–84

Basani M, Isham J, Reilly B (2008) The determinants of water connection and water consumption: empirical evidence from a Cambodian household survey. World Dev 36(5):953–968

Boardman AE, Boardman AE, Greenberg DH, Vining AR, Weimer DL (2018) Cost-benefit analysis: concepts and practice. Cambridge University Press, Cambridge

Bockstael NE, Hanemann MW, Kling CL (1987) Estimating the value of water quality improvements in a recreational demand framework. Water Resour Res 23(5):951–960

Boone C, Glick P, Sahn DE (2011) Household water supply choice and time allocated to water collection: evidence from madagascar. J Dev Stud 47(12):1826–1850

Briscoe J, Chakraborty M, Ahmed S (1981) How Bengali villagers choose sources of domestic water. Water Supply Manag 5:165–181

Brouwer R, Job FC, van der Kroon B, Johnston R (2015) Comparing willingness to pay for improved drinking-water quality using stated preference methods in rural and Urban Kenya. Appl Health Econ Health Policy 13(1):81–94

Calvo CM (1994) Case study on the role of women in rural transport: access of women to domestic facilities (English). Sub-Saharan Africa Transport Program (SSATP) working paper series; no. 11. Washington, DC: World Bank. http://documents.worldbank.org/curated/en/910381468742552233/Case-study-on-the-role-of-women-in-rural-transport-access-of-women-to-domesticfacilities

Cheesman J, Bennett J, Hung Son TV (2008) Estimating household water demand using revealed and contingent behaviors: evidence from Vietnam. Water Resour Res 16:470

Cook J, Kimuyu P, Blum A, Gatua J (2016) A simple stated preference tool for estimating the value of travel time in rural Africa. J Benefit Cost Anal 7(2):221–247

Coulibaly L, Jakus PM, Keith JE (2014) Modeling water demand when households have multiple sources of water. Water Resour Res 50(7):6002–6014

Creel M, Loomis J (1992) Recreation value of water to wetlands in the San Joaquin Valley: linked multinomial logit and count data trip frequency models. Water Resour Res 28(10):2597–2606

Dalhuisen JM, Florax RJGM, De Groot HLF, Nijkamp P (2003) Price and income elasticities of residential water demand : a meta-analysis. Land Econ 79(2):292–308

Deaton A, Muellbauer J (1980) An almost ideal demand system. Am Econ Rev 70(3):312–326

Deely J, Hynes S, Curtis J (2019) Are objective data an appropriate replacement for subjective data in site choice analysis? J Environ Econ Policy 8(2):159–178

Dubin JA, McFadden D (1984) An econometric analysis of residential electric appliance holdings and consumption. Econometrica 52(2):345–362

Espey M, Espey J, Shaw WD (1997) Price elasticity of residential demand for water: a meta-analysis. Water Resour Res 33(6):1369–1374

Filmer D, Scott K (2012) Assessing asset indices. Demography 49(April):359–392

Filmer D, Scott K, Prichett L (2001) Estimating wealth effects without expenditure data-or Tears: an application to educational enrollments in states of India. Demography 38(1):115–132

Gross E, Elshiewy O (2019) Choice and quantity demand for improved and unimproved public water sources in rural areas: evidence from Benin. J Rural Stud 16:688

Haab TC, McConnell KE (2002) Valuing environmental and natural resources. Measurement 8:326

Hanemann WM (1982) Applied welfare analysis with qualitative response models. Department of Agricultural & Resource Economics, UCB CUDARE Working Papers

Heckman JJ (1979) Sample selection bias as a specification error. Econometrica 47(1):153

Heien D, Wessells CR (1990) Demand systems estimation with microdata: a censored regression approach. J Bus 8(3):365–371

Hensher DA, Greene WH (2003) The mixed logit model: the state of practice. Transportation 30(2):133–176

Herriges JA, Kling CL, Phaneuf DJ (1999) Corner solution models recreation demand: a comparison of competing frameworks. In: Herriges JA, Kling CL (eds) Valuing recreation and the environment: revealed preference methods in theory and practice. Aldershot: Edward Elgar

Koehler J, Thomson P, Hope R (2015) Pump-priming payments for sustainable water services in Rural Africa. World Dev 2015(74):397–411

Kremer M, Leino J, Miguel E, Zwane AP (2011) Spring cleaning: rural water impacts, valuation, and property rights institutions. Q J Econ 126(1):145–205
Larson B, Minten B, Razafindralambo R (2006) Unravelling the linkages between the millennium development goals for poverty, education, access to water and household water use in developing countries: Evidence from Madagascar. J Dev Stud 42(1):22–40
Lee L-F (1983) Generalized econometric models with selectivity. Econometrica 51(2):507
Madanat S, Humplick F (1993) A model of household choice of water supply systems in developing countries. Water Resourc Res 29(29):1353–1358
Mannering FL (1986) A note on endogenous variables in household vehicle utilization equations. Transp Res Part B Methodol 20(1):1–6
McFadden D, Train K (2000) Mixed MNL models for discrete responses. J Appl Econ 15:447–470
Mu X, Whittington D, Briscoe J (1990) Modelling village water demand: a discrete choice approach. Water Resourc Res 26(4):521–529
Murphy KM, Topel RH (1985) Estimation and inference in two-step econometric models. J Bus Econ Stat 3(4):370–379
Nauges C, Van Den Berg C (2009) Demand for piped and non-piped water supply services: evidence from southwest Sri Lanka. Environ Resour Econ 42(4):535–549
Nauges C, Whittington D (2010) Estimation of water demand in developing countries: an overview. World Bank Res Obs 25(2):263–294
Nauges C, Strand J (2007) Estimation of non-tap water demand in Central American cities. Resour Energy Econ 29(3):165–182
Olea JLM, Pflueger C (2013) A robust test for weak instruments. J Bus Econ Stat 31(3):358–369
Onjala J, Ndiritu SW, Stage J (2014) Environment for development risk perception, choice of drinking water, and water treatment evidence from Kenyan Towns. J Water Sanit Hyg Dev 4(2):268–280
Persson TH (2002) Household choice of drinking water in the Phillipines. Asian Econ J 16(4):303–316
Phaneuf DJ, Requate T (2017) A course in environmental economics: theory, policy, and practice. Cambridge University Press, Cambridge
Phaneuf DJ, Requate T, Smith VK (2005) Chapter 15 recreation demand models. In: Karl-Gran M, Jeffrey RBT (eds) Handbook of environmental economics vincent. Valuing environmental changes, vol 2. Elsevier, New York, pp 671–761
Revelt D, Train K (1998) Mixed logit with repeated choices: households' choices of appliance efficiency level. Rev Econ Stat 80(4):647–657
Rural Water Supply Network (2013) Myths of the rural water supply sector
Shonkwiler JS, Yen ST (1999) Agricultural & applied economics association two-step estimation of a censored system of equations two-step estimation of a censored system of equations". Am J Agric Econ 81(4):972–982
Stelmach RD, Clasen T (2015) Household water quantity and health: a systematic review. Int J Environ Res Public Health 12(6):5954–5974
Strand J, Walker I (2005) Water markets and demand in Central American cities. Environ Dev Econ 10(3):313–335
Tanser F, Gijsbertsen B, Herbst K (2006) Modelling and understanding primary health care accessibility and utilization in rural South Africa: an exploration using a geographical information system. Soc Sci Med 63(3):691–705
Uwera C, Stage J (2015) Water demand by unconnected households in Rwanda. Water Econ Policy 1(1):145002
Von Wartburg M, Waters WG (2005) Congestion externalities and the value of travel time savings, chapter 2. In: Zhang A, Boardman AE, Gillen D, Waters WG (eds) Towards estimating the social and environmental costs of transportation in Canada : a report for Transport Canada. Centre for Transportation Studies, University of British Columbia, Vancouver, B.C.
White GF, Bradley DJ, White AU (1972) Domestic water use in East Africa, Drawers of water. University of Chicago Press, Chicago
Whittington D, Cook J (2019) Valuing changes in time use in low- and middle-income countries. J Benefit Cost Anal 10:51–72
Whittington D, Cook J, Xinming M, Roche R (1990) Calculating the value of time spent collecting water: some estimates for Ukunda, Kenya. World Dev 18(2):269–280
WHO/UNICEF (2017) Progress on drinking water, sanitation and hygiene. Technical Report, WHO/UNICEF Joint Monitoring Programme for Water Supply and Sanitation
World Health Organization (2013) How much water is needed in emergencies. Tech Notes Drink Water Sanit Hyg Emerg 9:1–4

Wu JJ, Babcock BA (1998) The choice of tillage, rotation, and soil testing practices: economic and environmental implications. Am J Agric Econ 80(3):494–511