The Journal of Development Studies

Household Water Supply Choice and Time Allocated to Water Collection: Evidence from Madagascar

Christopher Boone [a] , Peter Glick [b] & David E. Sahn [c]

[a] Department of Economics , Columbia University , New York, USA

[b] RAND Corporation , Arlington, USA

[c] Division of Nutritional Sciences, Department of Economics , Cornell University, Ithaca, USA
Published online: 01 Dec 2011.

PLEASE SCROLL DOWN FOR ARTICLE

whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms & Conditions of access and use can be found at http://www.tandfonline.com/page/terms-and-conditions

Routledge
Taylor & Francis Group

Household Water Supply Choice and Time Allocated to Water Collection: Evidence from Madagascar

CHRISTOPHER BOONE*, PETER GLICK** & DAVID E. SAHN†

*Department of Economics, Columbia University, New York, USA, **RAND Corporation, Arlington, USA, †Division of Nutritional Sciences, Department of Economics, Cornell University, Ithaca, USA

ABSTRACT *This article uses household survey data from Madagascar to examine water supply choice and time spent in water collection. We find that the choice of water source is strongly influenced by a number of household characteristics, as well as distance to sources. There are also strong substitution effects across sources. For example, increasing the distance to a public tap by 1 km increases the probability of using a well by 43 per cent in urban areas. With regards to time spent gathering water, we focus on the effects of gender, age, and distance to water. Women and girls spend the most time gathering water. The response to reducing distance to water sources differs in rural and urban areas, as well as by gender and age of household members. Investments to reduce to the distance to water sources will have larger impacts on adults than children, and on men than women.*

I. Introduction

The shortage of plentiful and clean water is a critical problem that threatens the health and well-being of much of the world's population. The lack of adequate and safe water for drinking, bathing and other household tasks is a direct threat to health. Further, agricultural and rural livelihoods are jeopardised by water shortages, as is the environment more generally. Given inadequacies in water infrastructure households in many parts of the world incur large time costs associated with gathering water. These costs are disproportionately borne by women and children, who also are most vulnerable to disease and food shortages that arise from a lack of access to safe and sufficient supplies of water. Improvements to water supply infrastructure can therefore have a range of benefits, with one of the potential benefits being the time savings resulting from reductions in time required to fetch water. One study estimates that time savings would account for 63 per cent of the

total economic benefits from achieving the Millennium Development Goals target for water supply (WHO et al., 2006).[1]

While there is a large literature on the health impacts of inadequate access to safe water[2] and on the critical importance that water plays in agriculture,[3] far less attention has been paid to examining the relationship between water infrastructure and time allocation. Indeed, it is widely assumed that the task of collecting water represents a significant burden for many of the world's poor. However, much of the evidence in this regard is anecdotal, although there are a few empirical articles that quantify the time allocated to water collection. Similarly, while travel time to water sources is presumably an important element in the time burden of water collection, here too the actual survey based empirical evidence is quite sparse. To be able to accurately estimate the benefits of improvements to water supply, it is important to have information about the time people spend gathering water, and how this time is related to the available infrastructure.[4]

One expected result might be that reductions in the distance to a household's water supply, for example, would reduce the time necessary for gathering water, by lowering the travel time to the source. However, there are a number of reasons why we might fail to observe significant time savings benefits as a result of households using closer sources. It is possible that travel time is only one small part of total water collection time, such that the time savings from reductions in distance are not very large. It might also be the case that sources that are closer to many households tend to have higher queuing or filling times, so that the reduced travel time is offset by higher time requirements at the source.[5] Another possibility is that households with more distant water sources may tend to make fewer trips to the source.

In order to provide further insight into the importance of access to water sources for the demand for safe water and the time allocated to water gathering, this article uses a unique dataset from Madagascar to address the following three questions.

- How is the household's choice of water source affected by the distance to available alternatives, wealth, and other household characteristics?
- How are the burdens of water collection distributed across different types of individuals within the household?
- How do the distance to water and the types of sources available affect time spent in water collection – or conversely, what are the potential time benefits of investments in different water source infrastructures?

In regard to the first issue of the choice of water source, we add to a relatively small list of studies examining the choice of water supply source. One early analysis is the article by Briscoe et al. (1981) that looks at the choice of water supply for a sample of about 150 families in a rural village in Bangladesh. The authors find that water quality is an important determinant of the choice of drinking water source, while distance appears much less important. However, distance has an important effect when choosing water sources for non-drinking purposes, such as cooking and bathing. Another factor affecting the choice of water supply is the potential for conflict with other people over the use of a source. Mu et al. (1990) use a multinomial logit model to analyse the choice of water source for a sample of 69 households in Kenya. Households in the sample choose among three types of sources (vendor, well, and kiosk), which differ by travel

time, price, and taste. Later studies include: Madanat and Humplick (1993), who examine the choice of water source for a sample of 588 households in Faisalabad, Pakistan; Asthana (1997), who uses a conditional logit model to examine household choice of water supply for 490 households in rural India; and Persson (2002), who uses a nested conditional logit model to analyse household choice of drinking water source for a sample of 769 rural households in the Philippines. The results of these studies indicate that the characteristics of available water infrastructures strongly affects the household's choice of source.[6] Like these previous studies, in this article we use information on the set of water sources available to the household in order to estimate a discrete choice model of household water supply.

But of potentially greater interest – and less well studied – are the second and third issues above, concerning the burdens of water collection, how these are related to the distance to the water source, and how they are distributed within the household. The question of time use in water collection has been addressed by several previous studies. However, few have focused on the impact of the characteristics of the available infrastructure and of household and individual characteristics on the time allocated to gathering water. Noteworthy among those that do is the study from Malawi by Nankhuni and Findeis (2004), who find that having piped water access in the home significantly reduces the probability of and time spent in water collection among children. They also find that having piped water access is positively associated with a child attending school and not doing any water-collection work, and negatively associated with combining the two (that is, engaging in water collection while attending school). Similarly, Ilahi (2001a) uses panel data from Peru to examine the impact of water infrastructure on total time spent in housework and time spent in income-generating activities by male and female adults. The results indicate that women in households without in-house water supply do not have significantly higher housework burdens than women in households with in-house water supply. Nor does the presence of in-house water supply have any significant impacts on the time that women spend in income-generating activities. For men, however, the presence of an in-house water supply significantly increases time spent in self-employment activities (such as agriculture) and decreases time spent in wage work.

A number of other articles have focused on child time use, particularly the relationship between child work and schooling. Some of these studies include information about water infrastructure in order to examine its impacts on work and schooling. Ilahi (2001b) looks at the impact of in-house water supply on children's schooling and time use. He finds that an in-house water supply has a significant impact on grade-for-age of girls but not boys, a significantly negative impact on child labour in urban areas, and seemingly no significant impact on time in housework. Akabayashi and Psacharopoulos (1999) investigate the impact of distance to water source on children's total work time in Tanzania. They find that distance is significantly associated with hours of work for both boys and girls. Distance is also negatively associated with hours of study for girls, but the results are only marginally significant. Cockburn and Dostie (2007) analyse the choice between schooling, work, and inactivity as the primary activity of children in rural Ethiopia. The number of minutes to the nearest water source reduces the probability of schooling as the main activity. This relationship is strongly significant for boys, but less so for girls. For boys, there is also a marginally

significant positive relationship between the minutes to water source and the choice of work or inactivity as the main activity.

Almost all of these prior studies suffer from a common limitation: that the choice of, or distance to the water source used by the households is treated as exogenous to time use and other household outcomes. Since the household typically chooses its source from among a set of alternatives, this choice may be associated with unobserved variables (preferences, knowledge) that also affect time use. As a result, the type of water source used, and thus also the distance to water, reflects a household decision hence is a potentially endogenous variable. Two articles that at least partially address this crucial econometric issue are those of Ilahi and Grimard (2000) and Glick et al. (2004).[7] In the former, the authors investigate the impacts of water infrastructure on the time allocation of women in rural Pakistan and find that having in-house taps as a primary source in the community significantly reduces the probability of collecting water and time spent in water collection, but is not significantly associated with time in market work or total work time. This community-level variable (indicating the presence of in-house taps in the community) is less likely to be endogenous than a variable indicating the household's actual source type. Similarly, in examining the impact of distance to water on time allocation, Ilahi and Grimard (2000) use what they call the 'leave-out mean cluster distance' instead of the household's actual distance, and in doing so find that distance to water is positively associated with hours spent in water collection.[8]

Glick et al. (2004) investigate the impacts of water infrastructure on time use in Madagascar and Uganda. Like Ilahi and Grimard, they use community-level variables to examine the impact of available water sources. For rural areas of both countries, they find little impact on water collection time of the availability of a well in the community relative to having only natural surface water sources available. (Given the absence of direct community level information, well water is defined as available or present if one or more households in the sample cluster reported using this source.) In Uganda, the presence of exterior taps in the community (defined in a similar way) slightly reduces time spent in water collection. In urban areas of both countries, availability of interior taps in the community also reduces average time spent in water collection, while large reductions in water collection time result, not surprisingly, from the household's predicted use of interior taps.[9] For Madagascar, the authors have information on the distance to the household's primary water source. Using leave-out mean cluster distance as an instrument for distance to the household's water source, they find no impact of distance on water collection time in rural areas; for urban areas, they find a significant positive relationship between distance and time for girls and boys (and a marginally significant relationship for women), but the effects are small.

In this study, we are able to build on these approaches by exploiting an innovative household survey from Madagascar to examine the amount of time people spend gathering water, and how this time varies with characteristics of the water supply infrastructure. We use these unique data to investigate how the distance to water affects the time spent in water collection, how the task of water collection is distributed among members of the household, and how the amount of time spent collecting water varies with other individual and household characteristics. We are able to estimate causal relationships with our data because, unlike previous studies,

we collected information on the entire set of water sources available to each household, along with the distance to each of these sources, regardless of whether the household chose to use that water source. We first use this information to estimate a discrete choice model for the household's choice of water source. We then estimate a reduced form model of individual time spent in water collection, where we regress water collection time on distance and source variables, as well as individual and household characteristics. For the distance variable, we use the leave-out mean cluster distance, but we construct this variable differently from earlier articles in order to take advantage of the additional information in our dataset.

Our results indicate, as expected, that an increase in the distance to one source means that households are less likely to choose that source over the available alternatives. We also find (in common with most studies) that women spend more time collecting water than men. For adults, the gender differences are large. For children, the gender differences, while statistically significant, are much smaller. Further, we find that the distance to a household's water source is positively associated with water collection time both for adults and older children (ages 15 and up), but for younger children (ages six to 14), there is no statistically significant relationship between distance and water collection time.

The remainder of this article is organised as follows. Section II describes the data and the modelling strategy used in our analysis. Our econometric results are presented in Section III, followed by our conclusions in Section IV.

II. Data and Empirical Specification

Data and Descriptive Statistics

Our data comes from the *Etude sur la Progression Scolaire et la Performance Academique en Madagascar* (EPSPAM), a household and school-level survey conducted by Cornell University and Institut National de la Statistique (INSTAT) in 2004–2005 which covered 73 rural and urban communities across Madagascar. The survey consisted of 2190 households, for which we collected detailed data on the characteristics of household members, as well as information on household wealth and assets. The employment and time use module included information on time use in various activities in the last seven days, including specifically hours spent gathering water. Although the survey covered rural and urban areas in all of the six provinces in Madagascar, it cannot be strictly considered as nationally representative since the bulk of the sample design involved returning to 48 communities that were initially part of a larger education study.[10]

The dataset contains information on the type of water source used by the household.[11] We have divided source types into four categories: public tap, private tap, well, and open surface water (such as ponds, streams, rivers or lakes). Regardless of the choice of water source, households provided the distance to each of the four types of water sources, if those sources were located within 10 kilometres of the domicile.[12]

Table 1 shows the number and percentage of households that use the different types of sources. Note that the use of private taps – either interior or exterior – is very low in rural areas, comprising only one per cent of households. In urban areas,

however, the use of private taps is much higher, with 19 per cent of urban households relying on this source of water. There is also a large difference in the use of surface water sources between rural and urban areas. In rural areas, 44 per cent of households collect water from a stream, pond, river or lake; the figure for urban areas is much lower, at eight per cent.

Table 2 presents summary statistics for the distance to the primary water source used by the household. For rural areas, the average and median distances are 243 m and 100 m, respectively. For urban areas, the figures are 110 m and 20 m. In both rural and urban areas, people using surface water sources travel the farthest distance,

Table 1. Primary water source used by the household during the dry season

	Rural		Urban	
Water source	Number of households	Per cent	Number of households	Per cent
Public tap	429	26	171	32
Private tap	18	1.0	84	19
Well	405	26	232	40
Surface water	749	44	46	8.0
Other	40	2.4	5	0.7
Total	1641	100	538	100

Notes: Private taps include interior and exterior private taps. Surface water includes rivers, lakes, ponds, and springs. 'Other' category includes rainwater and water vendors. Source information is missing for nine rural and two urban households.

Table 2. Household availability and distance to water by type of source

	Rural areas				Urban areas			
Source type	# Obs[a]	Mean dist	Median dist	St. dev.	# Obs	Mean dist	Median dist	St. dev.
Source used	1601	243	100	582	533	110	20	244
Avg over all available[b]	1650	753	300	1006	540	474	255	612
Nearest source	1650	193	80	391	540	63	10	141
Households using these sources:								
Public tap	429	147	50	568	171	129	80	180
Private tap	18	49	20	65	84	17	0	60
Well	405	177	80	432	232	69	8	176
Surface water	749	343	100	653	46	447	200	546
Households reporting source as available (within 10 km):								
Public tap	775	1137	100	2055	377	426	100	839
Private tap	164	1799	400	2728	289	236	40	734
Well	883	660	150	1618	398	230	20	682
Surface water	1473	827	300	1308	375	1303	500	1884

Notes: All distances in metres. [a]There are 1650 rural households and 540 urban households, but because of missing data we only know the distance to source for 1601 rural households and 533 urban households. [b]These statistics are for the average distance to all source types available to the household.

and people using private taps travel the shortest (this is essentially zero, of course, for those with interior or courtyard taps). While the percentage of households using surface water is much lower in urban areas, these individuals must travel farther on average than their rural counterparts. The median distance to surface water for individuals using this source is 200 m in urban areas and 100 m in rural areas.[13]

Table 2 also shows statistics for the average distance to all source types available to the household, as well as for distance to the nearest available source type. Note that, on average, the nearest available source is closer than the source actually used. This is because some households bypass the nearest available source. On the other hand, the distance to the source used tends to be much smaller than the average distance to all available sources.[14] The bottom four rows in Table 2 provide statistics on the availability of the different source types. In rural areas, 1473 out of 1650 households report having a surface water source available within 10 km, making this the most commonly available source type. In urban areas, wells are the most commonly available type, with 398 out of 540 households reporting them as available.

Table 3 displays data on the number of hours spent gathering water during the past seven days, disaggregated by type of individual. We report the average and median time for women, men, girls, and boys, in rural and urban areas. People in rural areas spend more time in water collection than people in urban areas. In rural areas, on average, women 15 and older spend 3.3 hours per week gathering water; men spend 1.6 hours per week; girls spend 2.3 hours per week; and boys spend 2.0 hours per week. For urban areas, the averages are: 2.0 hours per week for women; 1.2 hours per week for men; 2.0 hours per week for girls; and 1.7 hours per week for boys. This urban–rural difference in averages has two components: the percentage of people engaged in water collection is higher in rural areas; and, for those individuals actually engaging in water collection, the average weekly collection time is larger in rural areas.[15]

Table 3. Water collection time by type of individual

		All individuals		Only people collecting water	
	% collecting water	mean hours per week	median hours per week	mean hours per week	median hours per week
Rural areas					
Women (15+)	71	3.3	2	4.7	3
Men (15+)	35	1.6	0	4.5	3
Girls (6–14)	68	2.3	1	3.4	2
Boys (6–14)	57	2.0	1	3.6	2
Everybody (6+)	56	2.3	1	4.1	3
Urban areas					
Women (15+)	55	2.0	1	3.6	2
Men (15+)	36	1.2	0	3.4	2
Girls (6–14)	60	2.0	1	3.3	2
Boys (6–14)	47	1.7	0	3.5	2
Everybody (6+)	49	1.7	0	3.5	2

As in other environments, women and girls bear a disproportionate share of the burden of water collection. In both rural and urban areas, the percentage of adult women engaged in water collection is much larger than for adult men. For people ages six to 14, girls have a higher participation rate than boys, but the differences are not as large as for adults; in addition, when we look at only those individuals actually engaged in water collection, boys have a slightly higher average collection time than girls.[16]

In Table 4, we show the average total household collection time by type of water source used. For each source type, we also show the average amount of time spent by all women, all men, all girls, and all boys in the household, as well as the percentage of the total household time that each group contributes. Households using private taps have the lowest collection times, as we might expect. In rural areas, the average number of hours per week is roughly the same for households using public taps, wells, and surface water. In urban areas, households using surface water spend about 40 per cent more time getting water than households using public taps or wells. Given that surface water sources are, on average, much farther away than other source types in urban areas, this result is perhaps not too surprising.

In looking at the shares of time spent by each group, we can see that there are some differences across source types. In rural areas, women in households using wells are responsible for a larger share of time collecting water than women in households using public taps or surface water, while men in households using public taps have a larger share than men in households using wells or surface water. Girls and boys have slightly lower shares in households using wells than in households using public taps or surface water. In urban areas, women's and boys' shares are lowest in households using surface water, while men's and girls' shares are highest in these households.

While the presence of detailed information on time and distance to all water sources, including those not used, makes our dataset unusual, a limitation is the absence of direct measurement of the quality of these water sources, which may be an important factor influencing the household's choice of water source. Furthermore, the data record only the primary source type used for drinking and cooking, despite that some households may use multiple sources.

Econometric Specification: Household Choice of Water Source

We employ a conditional logit[17] to model the household's choice of water source as a function of distance to the source and a set of household characteristics. McFadden (1974) discussed the conditional logit model in the context of a random utility model. Following the notation in Greene (2008: 842), suppose households face a choice among J alternatives (indexed $j = 1, \ldots, J$) and the utility for household i from choosing alternative j is

$$U_{ij} = \mathbf{z}'_{ij}\theta + \varepsilon_{ij}, \qquad j = 1, ..., J, \tag{1}$$

where the vector $\mathbf{z}_{ij}$ contains characteristics describing the alternatives and characteristics of the households. If the random variable Y_i indicates the choice made, then $Y_i = j$ if and only if $U_{ij} > U_{ik}$ for all $k \neq j$. If the error terms ε_{ij} are

Table 4. Household collection time by source type, and shares of household time by type of individual

	Avg total hours for each group				Average hours per household	Percentage of total HH time				
	Women	Men	Girls	Boys		Women	Men	Girls	Boys	Total
Rural areas										
Public tap	4.4	2.8	2.4	2.3	11.8	36%	21%	24%	19%	100%
Private tap[a]	3.5	0.8	1.4	2.4	8.2	36%	18%	18%	28%	100%
Well	5.8	2.5	2.2	2.0	12.5	48%	16%	20%	15%	100%
Surface water	5.3	2.2	3.0	2.5	13.0	40%	15%	25%	19%	100%
All households	5.1	2.4	2.6	2.3	12.3	41%	17%	23%	18%	100%
Urban areas										
Public tap	3.5	2.1	2.2	1.9	9.6	39%	23%	21%	17%	100%
Private tap	1.1	0.7	0.6	0.7	3.1	44%	20%	17%	19%	100%
Well	4.0	1.8	2.2	1.7	9.7	43%	19%	23%	15%	100%
Surface water	4.3	3.6	3.8	1.9	13.6	32%	25%	31%	12%	100%
All households	3.3	1.9	2.0	1.6	8.8	40%	21%	23%	16%	100%

Note: [a]Note that only 1 per cent of rural households use private taps, so this row of statistics should be treated with caution.

independently and identically distributed according to a type I extreme value distribution, then the probability that source j is chosen equals

$$\Pr(Y_i = j) = \frac{\exp\left(\mathbf{z}'_{ij}\theta\right)}{\sum_{k=1}^{J} \exp\left(\mathbf{z}'_{ik}\theta\right)}. \tag{2}$$

Any characteristics in $\mathbf{z}_{ij}$ that do not vary across alternatives will drop out of this equation. In order to include household-specific characteristics such as wealth in the model, it is necessary to interact them with alternative-specific constants.

In our model, there are $J = 4$ water source alternatives: public tap, private tap, well, and surface water, the four categories for which we have distance information. We treat the *household* as the unit of analysis, and assume that each household uses only one source of drinking water. Not all sources are available to all households (based on the criterion above that the source be within 10 km) and 599 have only one local source available. Since we are analysing the choice among alternatives (and not using water is obviously not an alternative), only those households with more than one source available to them are included in the sample.[18] Among these households, there is still considerable variation in the number of sources available, and the conditional logit formulation is easily adjusted to allow for differing numbers of options. In addition to households with a single option, 29 households are excluded from the estimation because their primary water source is not one of the four main types; they use either rainwater, water vendors, or have missing data. This brings the sample size for the conditional logit model to 1562 households. Of these, just over half have only two sources available. Another 31 per cent report that there are three water sources available and 15 per cent have four sources available.

The explanatory variables contain one alternative-specific attribute: distance to each of the four source types, which is entered in quadratic form. The remaining explanatory variables are household-level characteristics that do not vary across source alternatives. These household characteristics are: age and number of years of education of the household head; a dummy for whether the household head is female; an asset index; variables describing the size and composition of the household (number of women age 15–60, number of men age 15–60, number of girls age 6–14, number of boys age 6–14, number of children five and under, number of people over 60); and regional dummies for province. Means and standard deviations are reported in the Online Appendix (Table A1).

In rural areas, wells are the closest available source (660 m), on average, followed by surface water sources (827 m) and public taps (1137 m), while private taps are the most distant (1799 m).[19] In urban areas, wells and private taps are the closest sources (230 m and 236 m, respectively), followed by public taps (426 m) and surface water sources (1303 m). The average years of education of the household head is 4.7 in rural areas and 7.5 in urban areas. The household head is female in 17 per cent of rural households and 22 per cent of urban households. The asset index was created following the procedure outlined in Sahn and Stifel (2003). Rural households on average are considerably less wealthy than their urban counterparts: the mean of the asset index is −0.30 for the rural sample compared with 1.11 for the urban sample, a difference of about 1.5 standard deviations. The average household size is 6.3 individuals in rural areas and 6.0 individuals in urban areas.

Econometric Specification: Time Spent in Water Collection

In modelling the amount of time that individuals within the household spend gathering water, we are interested in determining how the household's choice of water source and the distance to that source affect individual water collection time. We are also interested in determining how household and individual characteristics affect water collection time.

The challenge in estimating this type of model is that the distance to the water source and the type of source used are potentially endogenous variables; that is, they are likely to be correlated with the error term in the regression, thus leading to bias in the estimated parameters. There are a number of reasons why the distance and source variables might be correlated with the error term. One reason is that, for the majority of households with access to multiple water options and who therefore choose their source, this choice may be correlated with unobserved characteristics of the household that also affect time allocation to different activities – for example, preferences for leisure or for higher quality water. With regard to distance, we would expect households with low preference for water gathering work (or high returns to other activities) to have a source that is closer – in other words, the relationship may be from time in water collection to distance rather than the reverse. A different source of bias could result from the fact that both the time in water collection and the distance to water source are reported by the household. Both of these variables are measured with error. If this measurement error is random it will be relatively harmless for the dependent variable but not for the distance regressor. Further, it is possible that these errors are correlated, leading to systematic measurement error. A final issue is that the location of the various sources available to the household may itself be endogenous to household preferences. Households may choose where to live in part based on their desire to be near water sources or more generally, location within the community may be associated with unmeasured factors also influencing water collection times. In this case it is not just the specific choice and its distance that is endogenous to time use outcomes, but potentially the presence or distance of all choices.

To address these problems, we estimate a reduced form model, in which we omit the potentially endogenous variables and instead regress water collection time on a set of (arguably) exogenous variables. This approach was used in several previous studies, including Ilahi and Grimard (2000) and Glick et al. (2004). However, our approach differs from earlier studies in that our leave-out mean distance is calculated using the average distance to all sources available to the household, not merely the distance to the source *chosen* by the household. This is done to further avoid the potential endogeneity of the source selected. Thus, the household level average distance is then averaged over all other households in the cluster.[20] Since households presumably choose their water source to some extent based on distance, the measure based on chosen sources will clearly underestimate the average time to water sources in the community, even though it takes an average over all households.

Our specification is shown in Equation (3). We regress water collection time on the leave-out mean cluster distance along with dummy variables indicating which of the four types of sources are reported available by the household.

$$\begin{aligned}
\text{TIME}_i = \;& \beta_1 \cdot \text{CLUSTERDIST}_i + \delta_1 \cdot \text{AVBLPUBLIC}_i + \delta_2 \cdot \text{AVBLPRIVATE}_i \\
& + \delta_3 \cdot \text{AVBLWELL}_i + \delta_4 \cdot \text{AVBLSURFACE}_i + \gamma \cdot X_i + u_i
\end{aligned} \tag{3}$$

As mentioned, we use the leave-out mean cluster distance instead of the actual distance to the water source because the cluster distance is less likely to be endogenous. Other possibilities would have been to instead include the distance to the nearest available water source, or to the nearest source not chosen by the household, or the average distance to all available water sources. Each of these potential variables addresses the first endogeneity issue discussed above: they are correlated with the distance to the household's source, but otherwise uncorrelated with its choice of water source (among available alternatives) as well as its preferences or behaviour with respect to water collection time. We choose the leave-out mean cluster distance over the other options because of its advantages in dealing with the other sources of bias. One is measurement error, which would afflict household-specific reported distance variables such as distance to the nearest source to the households or to all sources from the household. Another is endogenous placement, if the household location vis-à-vis water sources is related to behaviour regarding time use. As noted by Ilahi and Grimard (2000), the leave-out mean cluster distance, as it averages over all (other) households in the cluster, does not suffer from respondent-specific (non-systematic) measurement error. It also partly addresses the endogenous placement problem, to the extent that this involves the individual household and not all households in the community. That is, while the leave-out mean cluster distance is uncorrelated with unobservable household-specific characteristics that vary within the cluster, it may still be susceptible to endogeneity resulting from variation across clusters. The extent to which certain types of households choose to live in particular clusters or are able to influence placement or availability of sources within the cluster could affect the level of bias in our estimates.

The source availability dummies are included in order to control for effects of the type of water source, which may be systematically related to collection time, perhaps due to differences in queuing or filling time. The vector X_i contains additional individual and household characteristics and a constant term. The individual characteristics include age, age-squared, and a dummy for gender. The additional household characteristics include the education level of the household head, the asset index, and regional dummies for five of the six provinces. As in the choice model above, we include variables describing the composition of the household; however, for this individual-level specification, the numbers of individuals in each age/gender category are exclusive of the individual her or himself.[21] We interact the female dummy with the asset index in order to determine if the effect of assets on water collection time differs by gender. We also interact the female dummy with the number of other people over 60; interactions with the other demographic variables were not significant and were dropped from the regression. Means and standard deviations are reported in the Online Appendix (Table A2). The average number of hours spent in water collection (our dependent variable) is 2.31 in rural areas and 1.69 in urban areas. The average value for the leave-out mean cluster distance is 0.769 km in rural areas and 0.483 km in urban areas.[22]

With regard to functional form of the time use regressions, as we saw in Table 3, a large number of individuals report spending no time gathering water during the

previous seven days. This means that the regression in Equation (3) has a censored dependent variable, making least squares an inappropriate estimation method. To deal with this problem, we use a censored regression – or tobit – model. A potential limitation of using a tobit model in this context is the implicit assumption that the factors determining the decision to participate in water collection also determine the number of hours spent in essentially the same way. It is possible that this assumption is not fully satisfied in our setting.[23]

An alternative approach would be to employ a sample selection model, such as a Heckman model.[24] This sort of model requires additional assumptions in order to be fully identified. Usually, an exclusion restriction is required: we need a variable that influences the decision to participate in water collection, but does not affect the amount of time spent gathering water for those individuals who engage in water collection. Because we have no plausible candidates in our data – indeed such a variable would be very hard to find – we proceed with using the tobit model (following Glick et al., 2004), but the misspecification concern should be kept in mind.

III. Results and Discussion

Household Choice of Water Source

We first consider the results of the conditional logit model for water supply choice. The parameter estimates are found in the Online Appendix (Table A3), while the marginal effects, which are easier to interpret and thus the focus of our discussion, are reported in Table 5. For continuous regressors, the marginal effects represent the change in probability of choosing each source type given a unit change in the regressor. For discrete variables, the effect is the change in probability of choosing each type given a change in the regressor from 0 to 1.

From the coefficient estimates on distance and distance-squared, we compute the total marginal effects of distance, and we compute standard errors using the delta method. The marginal effects (or discrete changes) are dependent on the values of the data. Given our particular interest in the effects of distance, we present the calculations for 10 metres distance, mean distance, and mean plus 1 standard deviation distance. When evaluating at a particular distance for one source type, the distances to other source types are held to their means, as are the values of all other covariates. The marginal effects of distance are highly significant for each source. As expected, increasing the distance to a source decreases the probability of using that source, and increases the probability of using the other types. These findings correspond to the results of previous studies – that increases in collection time, distance, or travel time reduce the likelihood of choosing a particular source. The magnitudes of these probabilities give an indication of the degree of substitution between different source types. In rural areas, the greatest degree of substitution is between public taps and wells and between public taps and surface water. Focusing on the results at the mean distance, we see that an increase of 1 km in the distance to a public tap increases the probability of using a well by 27 per cent and increases the probability of using surface water by 16 per cent. Increasing the distance to surface water by 1 km raises the probability of using a well by only 9 per cent. These results

Table 5. Marginal effects from conditional logit model for choice of water source

	Rural areas				Urban areas			
	change in probability of choosing				change in probability of choosing			
Variables	Public tap	Private tap	Well	Surface water	Public tap	Private tap	Well	Surface water
Distance to source								
at 10 metres								
km to public tap	−0.158***	0.000	0.100***	0.059***	−0.691***	0.230***	0.459***	0.002
km to private tap	0.0002	−0.0004	0.0001	0.0001	0.313***	−0.709***	0.394***	0.0017*
km to well	0.400***	0.000	−0.528***	0.127***	0.449***	0.283***	−0.734***	0.0024*
km to surface water	0.346***	0.000	0.187***	−0.533***	0.062**	0.039**	0.078**	−0.180**
at mean distance								
km to public tap	−0.430***	0.00001	0.271***	0.159***	−0.644***	0.215***	0.428***	0.0018
km to private tap	0.000	−0.00001	0.000*	0.000	0.224***	−0.506***	0.281***	0.0012*
km to well	0.306***	0.000*	−0.403***	0.097***	0.444***	0.280***	−0.727***	0.0024*
km to surface water	0.167***	0.000	0.090***	−0.257***	0.0015	0.001*	0.0019*	−0.004*
at mean plus 1 st dev								
km to public tap	−0.056***	0.000	0.036***	0.021**	−0.115***	0.038***	0.076**	0.0003
km to private tap	0.000**	0.000**	0.000****	0.000***	0.0003	−0.085***	0.047***	0.0002
km to well	0.026***	0.000	−0.034***	0.0083***	0.138***	0.087***	−0.225***	0.0007
km to surface water	0.017**	0.000	0.0092***	−0.026***	0.00003	0.00002	0.00003	−0.0001
Household variables								
Schooling of household head	0.021**	0.000	−0.016**	−0.0044	0.009	0.013	−0.021	0.000
Age of household head	0.0009	0.000	−0.0014	0.0005	−0.0075**	0.0056	0.0019	0.0001
Female-headed household[a]	0.163**	0.000	−0.102	−0.061**	0.147	0.032	−0.178	−0.0009
Assets	0.066	0.00001	0.091	−0.157***	−0.040	0.131***	−0.090	−0.0006
# of women (15–60)	0.037	0.000	−0.0044	−0.033*	−0.016	0.033	−0.016	−0.0005
# of men (15–60)	0.047*	0.000	−0.029	−0.017	−0.019	−0.013	0.032	0.0002
# of girls (6–14)	−0.057**	0.000	0.044	0.013	0.042	−0.060**	0.017	0.0009
# of boys (6–14)	−0.0084	0.000	0.053*	−0.044***	0.016	0.032	−0.048	−0.0006
# of kids (5 and under)	0.018	0.000	−0.043	0.025*	−0.045	0.079***	−0.034	0.0003
# people over 60	0.060	0.00001	−0.037	−0.023	0.147	0.083	−0.228*	−0.001

Table 5. (*Continued*)

| | Rural areas | | | | Urban areas | | | |
| | change in probability of choosing | | | | change in probability of choosing | | | |
Variables	Public tap	Private tap	Well	Surface water	Public tap	Private tap	Well	Surface water
Province 2[a]	−0.214*	0.000	0.302**	−0.089**	−0.038	0.077	−0.040	0.0012
Province 3[a]	−0.040	0.000	0.136	−0.095	−0.141	−0.147*	0.290***	−0.002
Province 4[a]	0.060	−0.00007	0.126	−0.186***	−0.085	−0.150	0.249**	−0.013
Province 5[a]	0.191**	−0.00004*	0.052	−0.242***	0.100	−0.042	−0.059	0.0008
Province 6[a]	0.195	−0.00006*	−0.291***	0.096	−0.073	−0.223***	0.298***	−0.0015

Notes: Standard errors calculated using the delta method and adjusted for clustering. Marginal effects for distance to a particular source type are evaluated at 10 metres, mean distance, and mean plus one standard deviation; other source types are held to their mean values. Marginal effects for household variables evaluated at mean values for all variables. Means and standard deviations are shown in the Online Appendix (Table A1). Coefficient estimates for this model are shown in the Online Appendix (Table A3). Sample size is 1101 rural households and 461 urban households.
[a]Marginal effect for dummy variable is effect of changing from 0 to 1.
***p < 0.01, **p < 0.05, *p < 0.1

indicate that households view wells and taps as relatively similar, but surface sources and wells as quite different – which accords with expectations. The marginal effects of distance for private taps in rural areas are small due to the low number of households using these types.[25] In urban areas, the greatest substitution occurs between public taps and wells. At the mean distance, increasing the distance to a public tap by 1 km increases the probability of using a well by 43 per cent. Looking at the marginal effects at the three different distance values, we can see that for both rural and urban areas, the magnitudes of the marginal effects are generally decreasing as the distance is increased. This means that the effect of a one-unit increase in the distance to a particular source type is larger (in magnitude) for closer sources. All in all, the results show that households' choice of water source is quite sensitive to distance. Therefore, public investments can fairly easily (subject to cost) induce use of specific (and safer) water sources by bringing them closer to households.

A number of household characteristics also have significant effects on the source type chosen. Controlling for distance, wealth, and other factors, years of education of the household head is positively associated with choosing a public tap, and negatively associated with choosing a well. For example, an additional year of school increases the probability of using a public tap in rural areas by two per cent; in urban areas, the effect of schooling is not significant. Interestingly, female-headed households in rural areas are 16 per cent more likely to use public taps and six per cent less likely to use surface water. Since women in such households presumably have greater say in decision-making than in male-headed households, this finding may reflect greater preference of women for better or safer water sources. Household assets are positively associated with using a private tap in urban areas and negatively associated with using surface water in rural areas. There are few demographic variables for which the marginal effects are significant in the water source choice models. Among those that are, there is no clear pattern or story that emerges, and the magnitude of the effects are small. For example, the number of girls is negatively associated with choosing a private tap in urban areas and a public tap in rural areas, while for boys, the only significant marginal effect is in rural areas where their presence reduces the likelihood of using surface water sources. These results seem both difficult to explain and are of little obvious policy relevance.

Time Spent in Water Collection

We now turn to the results of the model of water collection time, where we regress hours in water collection on the leave-out mean cluster distance (to all sources) and other covariates, including indicators for the availability of specific water sources.[26] These results are shown in Table 6. We estimate the model separately for adults (ages 15 and up) and children (ages six to 14). For rural areas, we find that distance is significantly positively associated with water collection time for adults. We also tried a specification with distance in quadratic form: when leave-out mean cluster distance-squared is included as well, the distance variables are jointly significant for both adults and children (p-values of 0.081 for adults and 0.052 for children, results not shown).

For urban areas, our results show that distance is again significant for adults, but not for children (in both the linear and quadratic specifications). The point estimate for adults is positive and large, indicating that higher distance leads to significantly higher time spent in water collection.

We find significant effects of other covariates as well.[27] For adults in both rural and urban areas, the coefficient on the female dummy is large and strongly significant, indicating that women spend much more time gathering water than men and consistent with the descriptive statistics. For children, the female dummy coefficient is much smaller, and it is significant only for the rural sample. The years of education of the household head is negatively associated with water collection time for children in urban areas. Water collection time increases with age for children, but the rate of increase falls with age. For adults, collection time is decreasing in age. In urban areas, a higher level of household assets is associated with less time in water collection for both adults and children. Note that since we include source availability indicators in these regressions, not the actual source used, part of this effect of wealth on collection time is operating through the choice of water source. The results on the interaction between the asset variable and gender indicate that the negative impact of asset level on water collection time in urban areas is larger for adult women than men (though the interaction term is only marginally statistically significant). That is, the time use 'benefits' that come with wealth disproportionately accrue to women.[28] In rural areas, assets also have a strong negative impact on time collecting water, but this is limited to female children six to 14 years of age. For adults, there is no effect of assets on water collection in rural areas.

Most of the household composition variables are strongly significant in rural areas. In general, having additional household members reduces an individual's time in water collection, for both adults and children. This suggests, not surprisingly, that there are economies of scale in this activity. One exception is that a higher number of children ages five and under *increases* the time in water collection for girls and boys. This may be because the care of young children implies a higher demand for water. The point estimates on the number of adult women in the household are larger in magnitude than any other group, indicating that the greatest reduction in individual collection time results from the presence of other adult women. In urban areas, while the effects of the household composition variables are often no longer significant, almost all the point estimates remain negative, suggesting again economies of scale in water collection. The effect of the number of adult women is still significant: having other adult women in the household significantly decreases collection time for all types of individuals.

Comparing the results on the distance variables across estimation samples in Table 6, we see that the effects of distance are larger and more statistically significant for adults than for children. In order to better understand how the effects of distance might differ across individuals, we run a number of additional specifications. We describe our findings here, while the actual estimation results are shown in the Online Appendix (Table A4).

First, to get a better sense of how distance affects collection time for the household as a whole, we estimate the model for the combined sample containing all individuals ages six and older. In this specification, the estimate on distance is positive but not

Table 6. Reduced form tobit model of hours in water collection

Variables	(1)	(2)	(3)	(4)
	Rural areas		Urban areas	
	Adults	Children	Adults	Children
	(15+)	(6–14)	(15+)	(6–14)
Leave-out mean cluster dist (km)	0.758**	0.137	2.434***	1.275
	(0.347)	(0.265)	(0.767)	(1.742)
Available: public tap	−0.145	0.366	0.196	0.327
	(0.466)	(0.378)	(0.945)	(0.473)
Available: private tap	0.111	0.430	0.250	0.848
	(0.452)	(0.368)	(0.517)	(0.694)
Available: well	−0.148	−0.471	1.222**	0.601
	(0.451)	(0.331)	(0.487)	(0.537)
Available: surface	0.0635	0.352	0.448	0.494
	(0.548)	(0.492)	(0.400)	(0.763)
Female	3.814***	0.372**	2.199***	0.503
	(0.435)	(0.153)	(0.499)	(0.351)
Female*assets	−0.997	−0.659**	−0.383*	0.405
	(0.695)	(0.263)	(0.188)	(0.306)
Age	−0.145***	2.440***	−0.118	3.870***
	(0.0371)	(0.325)	(0.0700)	(0.820)
Age-squared	−0.000288	−0.0925***	−0.000445	−0.149***
	(0.000495)	(0.0144)	(0.000943)	(0.0388)
Schooling of household head	0.00885	−0.0245	−0.0418	−0.168***
	(0.0535)	(0.0343)	(0.0474)	(0.0553)
Assets	0.138	0.144	−0.715***	−0.987**
	(0.440)	(0.228)	(0.235)	(0.354)
# other women (15–60)	−0.857***	−0.648***	−0.551***	−0.650***
	(0.135)	(0.1000)	(0.155)	(0.187)
# other men (15–60)	−0.271**	−0.359***	−0.209	−0.310
	(0.120)	(0.0759)	(0.186)	(0.238)
# other girls (6–14)	−0.650***	−0.566***	−0.249	−0.125
	(0.146)	(0.121)	(0.186)	(0.213)
# other boys (6–14)	−0.368***	−0.424***	−0.0409	−0.455
	(0.119)	(0.0997)	(0.138)	(0.308)
# other children (5 & under)	−0.401**	0.345***	−0.148	−0.325
	(0.158)	(0.108)	(0.160)	(0.187)
# other people > 60	0.317	−0.583	−1.056*	−1.938
	(0.567)	(0.373)	(0.605)	(1.196)
(# older than 60)*female	−0.249	1.424***	−0.0823	0.281
	(0.642)	(0.416)	(1.123)	(1.016)
Constant	5.680***	−11.57***	2.466*	−21.41***
	(0.895)	(1.984)	(1.295)	(4.380)
Sigma	5.838***	4.027***	4.983***	4.576***
	(0.398)	(0.258)	(0.337)	(0.513)
Observations	4888	3523	1776	1111

Notes: Dependent variable is hours spent gathering water during the past seven days. Cluster-adjusted standard errors shown in parentheses. Regional dummies included in the model but not shown here. ***p < 0.01, **p < 0.05, *p < 0.1.

statistically significant for individuals in rural areas; for urban areas, the estimate is positive but only marginally significant at the 10 per cent level. These results suggest that there may not be a strong relationship between distance to water and total household collection time.

We then add in interactions between the distance and age variables (both age and age-squared), in order to see how the effects of distance differ across age groups. For rural areas, distance and the distance-age interactions are strongly jointly significant (p-value of 0.0061), suggesting that distance does indeed affect individual collection time, but this effect differs by age. The results are similar for urban areas, but only marginally significant (p-value of 0.0596). These results are consistent with the results shown earlier in Table 6. When we interact distance with both age and age-squared, we find that the effect of distance initially increases with age, then decreases. In both rural and urban areas, the positive effect of distance on collection time is largest in magnitude around age 25 (and then decreases with age, becoming negative around age 48). It may be the case that the distance to the water source affects the distribution of water collection responsibilities within the household. In households with distant water sources, the burden of collection may shift towards young adults. These findings suggest that these individuals are the most likely to benefit (in terms of reduced collection time) from interventions that reduce the distance to water; in contrast, children and older adults may see smaller reductions in collection time from such interventions.

We are also interested in differing effects by gender, so we run specifications where we interact the distance and gender variables, again estimating the model separately for adults and children. The results for the rural and urban samples are similar. For children, neither distance nor the female-distance interaction term is significant. For adults, the estimate on distance is positive and significant, while the female-distance interaction is negative and significant (but smaller in magnitude). These results suggest that higher distance leads to higher collection time for both men and women, but the effect is much larger for men. This means that reductions in the distance to water may lead to larger reductions in collection time for men than for women. Note also that the way in which the time savings are distributed across household members can affect how these savings should be valued (in benefit–cost calculations, for example). In particular, since young adults (and especially men) seem to experience the largest reductions in collection time, this might suggest that the time saved can be valued closer to the prevailing unskilled wage than if the benefits accrued primarily to children or the elderly.[29]

IV. Conclusion

In this article we have explored the complementary questions of the effects of household characteristics and distance to water on the choice of water source, and the determinants of time allocated to the collection of water among different individuals in the household. We use a relatively unique dataset that has information not only on distance to the water source used by the household, but also on other water sources not chosen. This enables us to estimate a discrete choice model of water supply choice. We also estimate a model of individual time involved in water

collection, using a reduced form approach to deal with the potential endogeneity of the distance variable.

Our results indicate that the distance to available water sources has an important impact on the choice of water source, with the expected result that a decrease in the distance to one type of water source makes the household more likely to use that source, and to switch from other sources. Given this sensitivity to distance, infrastructure investments that provide more (and closer) sources of drinking water will lead to changes in household behaviour. More specifically, investments in safer water supply that also increase proximity of supply to households will lead to greater uptake of these sources – apart from the effects of improved water quality itself on demand and of public campaigns to encourage use of these sources. That being said, there are large differences in the magnitude of the substitution between different source types, and these differ by geographical area. We also find that schooling affects the choice of water source in rural areas, where it is associated with increased use of public taps and reduced use of wells. Likewise, assets are important in the choice of water source, and this is particularly so in urban areas where assets are associated with increased use of water from private taps.

While distance seems to have a strong effect on the choice of water source, we find less evidence that distance has an impact on total household water collection time. In the combined sample of individuals of all ages, the effects of distance were not significant for rural areas, and only marginally significant for urban areas. A number of factors affect the amount of time that people spend gathering water, and it appears that the distance to the source is not as strong of a predictor as one might expect. It is also worth keeping in mind that measurement error in the self-reported distance and time variables could be contributing to the lack of strongly significant effects.[30] Nonetheless, we do find evidence that the distance to the water source significantly affects collection time for individual household members. The impact of distance differs across age groups, possibly by shifting the responsibility for water collection among household members. There is also evidence that distance has a larger impact on collection time for men than for women. Therefore, investments in water infrastructure may have distributional impacts on time use; changes in the average distance to water sources or to specific types of water source may impinge differently on different household members with respect to the burden of water collection.

The total amount of time spent in water collection is also strongly influenced by a person's age and gender. On average, adult women spend the most time in water collection, followed by girls, boys, and then adult men. By implication, women and girls will see the greatest benefits from investments that substantially reduce the required collection time (installing taps in the home, for example). However, this does not necessarily mean that women and girls will see larger reductions in collection time from marginal changes in infrastructure (slightly reducing the distance to water, for example).

Finally, this article adds to the literature on water choice and time allocated to water collection, in part due to the fact that we collected information on both time use and on distance to all water sources, not just those used. That being said, a number of interesting additional issues could not be addressed due to limitations of the available data. Specifically, we do not have any information on the quantity and quality of water accessed. Given the importance of both the quantity and quality of

water in affecting health outcomes, gathering such data in future surveys that also look at the range of alternative sources would open a variety of new research opportunities, including how quality, both measured and perceived, affects water source choice and time spent gathering water. Likewise, knowing how distance affects the quantity of water consumed would be very useful in considering the efficacy of investments to make water more accessible.

Notes

1. In order to estimate the economic benefits of improvements in water supply, WHO et al. (2006) assume that expanding access to water supply would result in time savings of 30 minutes per household per day, which the authors consider to be a conservative estimate. To arrive at this number, the authors examine estimates of water collection time from a number of previous studies, but also point out 'the dearth of data on time uses in the literature', as well as 'insufficient global evidence' to alter their estimate of time savings. It is also worth noting that while the authors cite estimates of time spent in water collection in a number of different regions, they do not cite any studies that discuss how collection time is affected by improvements in water supply infrastructure. Also see Churchill et al. (1987), who emphasize the time savings benefits from water supply improvements, and suggest methods for valuing these benefits.
2. See, for example, Cairncross (2003), Clasen et al. (2007).
3. See, for example, World Bank (2007), Comprehensive Assessment of Water Management in Agriculture (2007).
4. The review by Rosen and Vincent (1999: 55–64) surveys the evidence on the time costs of collecting water. Most of the studies that they describe are site-specific, and none of these studies collect national-level data; the evidence on time-related benefits of water investment is even more limited. Thompson et al. (2001) describe the results of a study on changes in water use in 34 sites in three countries in East Africa. While the sites were not randomly selected, the study offers detailed information about the relationship between water infrastructure and household behaviour, including time use. More information on time in water collection and the relationship to water source distance is given below.
5. It is also possible that travel time has a higher disutility than queuing time, since carrying water is so physically taxing; there may also be social benefits to queuing due to the interaction with one's neighbours. An individual choosing between two sources might therefore trade off a decrease in travel time for an even larger increase in queuing time. In such a case, the person could be better off even though total collection time has increased. In an observational study, this tendency would serve to reduce the positive association between collection time and distance or even contribute to a negative relationship.
6. Mu et al. (1990) find that collection time negatively impacts the decision to use a particular water source; Asthana (1997) finds negative impacts of distance on source choice; and Persson (2002) finds negative impacts of travel time on source choice.
7. Rosen and Vincent's (1999) survey of the literature regarding household water resources in rural sub-Saharan Africa did not find studies that focus specifically on the time savings resulting from interventions that reduce the distance to a source of water. However, they discuss a few studies that seem to indicate that using a more distant water source increases the time allocated to water collection. In one study, Cairncross and Cliff (1987) compared two similar villages in Mozambique – one with a close water source and the other with a distant source – and noted that women in the village with the distant source spent more time collecting water. Another study (Huttly et al., 1990) examined the impact of constructing boreholes within three villages in Nigeria and observed that the time devoted to water collection dropped substantially after the intervention.
8. For each household, the leave-out mean cluster distance is the average distance across all other households in the cluster, excluding that particular household. The leave-out mean cluster distance is also sometimes referred to as the non-self mean distance.
9. The authors model the use of interior taps (for urban areas only) using availability of interior taps in the community as an instrument for use of an interior tap by the household.

10. In order to make our results more representative, we construct sampling weights based on province and rural/urban location; these weights are used to calculate all of the statistics and regressions reported in this article. See Glick et al. (2005) for more details on the survey.

11. Specifically, households were asked for the primary water source used for drinking and food preparation. Separate responses were recorded for the wet season and dry season. Most households report using the same primary source in both seasons: only 94 out of 1641 rural households and 37 out of 538 urban households use different sources in the wet season and dry season. In this article, we focus on the water sources used during the dry season. The results are very similar for the wet season.

12. Unfortunately, no other information about these source types was recorded, so, for example, we have no information about water quality, reliability of the source, or whether a fee is required.

13. These statistics for distance are somewhat larger than those found by Glick et al. (2004) in their earlier study in Madagascar (using the results of a 1993 survey). The authors report the average distance to water to be 192 m in rural areas and 92 m in urban areas, compared to 241 m and 126 m, respectively, in our sample. In addition, they find that wells are the most distant source in rural areas, at an average of 335 m, while the average for surface water sources is 167 m. In rural Pakistan, Ilahi and Grimard (2000) report an average round-trip distance of 560 m, corresponding to a one-way distance of 280 m. Akabayashi and Psacharapoulos (1999) find the average distance to be 0.275 miles (440 m) in Tanzania, though this is the individual average for children in the sample, not the household average. For their sites in East Africa, Thompson et al. (2001) report an average distance of 466 m to the water source for rural households, and 204 m for urban households. Many surveys record the distance to water in units of time. In rural Ethiopia, the average number of minutes to the nearest water source is 18.5 minutes for boys in the sample, and 17.5 minutes for girls (Cockburn and Dostie, 2007). In the Philippines, households in rural areas require an average of 6.96 minutes to walk to the water source (Persson, 2002). Cairncross and Valdmanis (2006) report the results of an analysis of MICS survey data from 23 African countries which finds that 44 per cent of households were more than 30 minutes away from their water source. The 2003–2004 DHS survey from Madagascar asks for the amount of time required to travel to the water source, collect water, and return home. The median time required is 9.3 minutes in rural areas and five minutes in urban areas. In addition, 65 per cent of rural households and 73 per cent of urban households have times of less than 15 minutes (INSTAT and ORC Macro, 2005).

14. In rural areas, 218 out of 1600 households (14%) use a source other than the closest available. On average, these households travel 607 m to reach their water source, which is an additional 509 m beyond the nearest available source. In urban areas, 118 out of 532 households (22%) bypass the nearest source, travelling 304 m on average, or an additional 224m beyond the nearest available source.

15. These patterns hold for the disaggregated groups as well, with one exception: men have a slightly higher participation rate in urban areas (36%) than in rural areas (35%).

16. Our results for collection time are similar to those found by Glick et al. (2004) in their earlier study in Madagascar, though the gender differences in their results are slightly larger: compared to our results, they find slightly higher average times for women and girls and slightly lower average times for men and boys. Their results for adults in Uganda are also similar. Ilahi and Grimard (2000) report an average of 11.45 hours per *month* in water collection for adult women in rural Pakistan; our result of 3.4 hours per week for rural women corresponds to 14 hours per month. Wodon and Blackden (2006) find that girls ages six to 14 in rural Benin spend one hour per day on average in water collection, while boys spend 25 minutes. Malmberg-Calvo (1994) reports the results of surveys in four different sites located in Ghana, Tanzania, and Zambia; the average time in water collection for rural adult women in these four areas ranges from 22 to 93 minutes per day, which corresponds to 2.6 to 11 hours per week. A 1996 survey across six states in India (UNDP and ESMAP, 2004) found that rural households spend an average of 0.93 hours per day fetching water, corresponding to 6.5 hours per week; this figure is slightly more than one-half the amount for rural households in our sample (Table 4).

17. The conditional logit model is closely related to (and sometimes referred to as) the multinomial logit model.

18. Among the 528 rural households with only one source available, 76 per cent have only surface water sources available. Of the 71 urban households with only one source available, 46 per cent have access to only public taps, and another 46 per cent have access only to wells.

19. The 'private taps' category includes any interior or exterior private tap to which the household has access, for free or for a fee; the tap does not have to be owned by the household. Of the 164 rural households (10%) reporting private taps as available, 49 of these households report that the tap is located at a distance of 2 km or greater. Access to private taps is very low in rural areas.

20. More precisely, $\text{CLUSTERDIST}_i = \frac{1}{n-1}\sum_{j \neq i} \text{AVGDIST}_j$, for household i and all other households j in the cluster, where n is the number of households in the cluster and AVGDIST_j is the average distance to all source types available to household j.

21. So, for each individual, all but one of these variables (representing the category to which the individual belongs) is equal to the total number of individuals of that type in the household.

22. In addition to the specification shown in Equation (3), we run a specification omitting the distance variables and another one omitting the availability indicators but including distance.

23. In particular, it is conceivable that the distance to the water source might negatively affect the decision to participate in water collection, but conditional on participating, distance could have a positive effect on hours. If this is the case, it means that a tobit model would be a misspecification.

24. This is the procedure used by Ilahi and Grimard (2000). In order to identify the model, they assume that the presence of in-house water taps in the household and in the community affect the decision to participate in water collection, but do not affect the time spent gathering water conditional on participation. In our sample, where the prevalence of in-house taps is quite low, this argument would be less persuasive.

25. Only 18 out of 1650 rural households use private taps. Marginal effects are scaled by the mean probabilities of using the choice, so will be low for these sources.

26. We run the same models without the availability dummies and find that the results are very similar in terms of significance and size of the coefficients.

27. The results on the other covariates are robust to the inclusion of the quadratic distance term.

28. These are not 'benefits' in the sense of additional leisure, necessarily; all that the estimates indicate is that as wealth rises, women increasingly use their time for other, non-water collection purposes. This could include additional work time, and indeed, a higher marginal return to other productive activities as wealth rises (perhaps because labour and productive assets are complements) would lead women to substitute other productive time for water collection time.

29. We thank an anonymous referee for drawing our attention to this point.

30. People may have trouble accurately recalling or computing the total time spent in water collection over the previous seven days, meaning that there is likely to be measurement error in our dependent variable. The distance variable also likely contains error, especially since households were asked the distance not only to the source they used, but to other sources as well. As discussed above, we compute an average distance across the households in the cluster, which may result in a more precise measure of distance, but this leave-out cluster mean distance is itself an imprecise measure of the distance that we might think actually matters for household collection time. In addition, the tobit model is nonlinear, so the implications of measurement error may differ from the conventional results for the linear model. For example, heteroskedastic measurement error in the explanatory variables can potentially lead to upward-biased parameter estimates (Bound et al., 2001).

References

Asthana, A.N. (1997) Where the water is free but the buckets are empty: demand analysis of drinking water in rural India. *Open Economies Review*, 8(2), pp. 139–149.

Akabayashi, H. and Psacharopoulos, G. (1999) The trade-off between child labour and human capital formation: a Tanzanian case study. *The Journal of Development Studies*, 35(5), pp. 120–140.

Bound, J., Brown, C. and Mathiowetz, N. (2001) Measurement error in survey data, in: J.J. Heckman and E. Leamer (eds) *Handbook of Econometrics*, Vol. 5. (Amsterdam: North Holland), pp. 3705–3843.

Briscoe, J., Chakraborty, M. and Ahmed, S. (1981) How Bengali villagers choose sources of domestic water. *Water Supply and Management*, 5(2), pp. 165–181.

Cairncross, S. (2003) Editorial: water supply and sanitation; some misconceptions. *Tropical Medicine & International Health*, 8(3), pp. 193–195.

Cairncross, S. and Cliff, J.L. (1987) Water use and health in Mueda, Mozambique. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 81(1), pp. 51–54.

Cairncross, S. and Valdmanis, V. (2006) Water supply, sanitation, and hygiene promotion, in: D.T. Jamison, J.G. Breman, A.R. Measham, et al. (eds) *Disease Control Priorities in Developing Countries,* 2nd ed. (New York: Oxford University Press), pp. 771–792.

Churchill, A. A., with the assistance of de Ferranti, D., Roche, R., Tager, C., Walters, A. A. and Yazer, A. (1987) Rural water supply and sanitation: time for a change. World Bank Discussion Paper No. 18, World Bank, Washington, DC.

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

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

Comprehensive Assessment of Water Management in Agriculture (2007) *Water for Food, Water for Life: A Comprehensive Assessment of Water Management in Agriculture* (London: Earthscan, and Colombo: international Water Management Institute).

Glick, P., Saha, R. and Younger, S.D. (2004) Integrating gender into benefit incidence and demand analysis. Cornell Food and Nutrition Policy Program Working Paper No. 127, Cornell University, Ithaca, NY.

Glick, P., Rajemison, H., Ravelo, A., Raveloarison, Y., Razakamanantsoa, M. and Sahn, D.E. (2005) The progression through school and academic performance in Madagascar study: preliminary descriptive results. Cornell University, INSTAT, and Ministry of Education and Scientific Research. Cornell Food and Nutrition Policy Program Working Paper No. 166.

Greene, W.H. (2008) *Econometric Analysis,* 6th ed. (Upper Saddle River, NJ: Prentice Hall).

Huttly, S.R.A., Blum, D., Kirkwood, B.R., Emeh, R.N., Okeke, N., Ajala, M., Smith, G.S., Carson, D.C., Dosunmu-Ogunbi, O. and Feachem, R.G. (1990) The Imo State (Nigeria) drinking water supply and sanitation project, 2. Impact on dracunculiasis, diarrhoea and nutritional status. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 84(2), pp. 316–321.

Ilahi, N. (2001a) Gender and the allocation of adult time: evidence from the Peru LSMS panel pata. Background article for *The Policy Research Report on Gender*, World Bank, Washington, DC.

Ilahi, N. (2001b) Children's work and schooling: does gender matter? Evidence from the Peru LSMS panel data. Background article for *The Policy Research Report on Gender*, World Bank, Washington, DC.

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

INSTAT (Institut National de la Statistique) and ORC Macro (2005) Enquête démographique et de santé de Madagascar 2003–2004 (Calverton, MS: INSTAT and ORC Macro).

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

Malmberg-Calvo, C. (1994) Case study on the role of women in rural transport: access of women to domestic facilities. SSATP Working Paper No. 11, Technical Department, Africa Region, World Bank.

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

Mu, X., Whittington, D. and Briscoe, J. (1990) Modeling village water demand behavior: a discrete choice approach. *Water Resources Research*, 26(4), pp. 521–529.

Nankhuni, F.J. and Findeis, J.L. (2004) Natural resource-collection work and children's schooling in Malawi. *Agricultural Economics*, 31(2–3), pp. 123–134.

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

Rosen, S. and Vincent, J. R. (1999) Household water resources and rural productivity in sub-Saharan Africa: a review of the evidence. Harvard Institute for International Development, Harvard University, Cambridge, MA.

Sahn, D.E. and Stifel, D. (2003) Exploring alternative measures of welfare in the absence of expenditure data. *Review of Income and Wealth*, 49(4), pp. 463–489.

Thompson, J., Porras, I. T., Tumwine, J. K., Mujwahuzi, M. R., Katui-Katua, M., Johnstone, N. and Wood, L. (2001) *Drawers of Water II: 30 Years of Change in Domestic Water Use and Environmental Health in East Africa* (London: International Institute for Environment and Development).

UNDP (United Nations Development Programme) and ESMAP (World Bank Energy Sector Management Assistance Programme) (2004) *The Impact of Energy on Women's Lives in Rural India* (Washington, DC: World Bank).

WHO (World Health Organization), Hutton, G., Haller, L., and Bartram, J. (2006) Economic and health effects of increasing coverage of low cost water and sanitation interventions. Occasional Paper 2006/33, Human Development Report Office, UNDP.

Wodon, Q. and Blackden, C.M. (eds) (2006) *Gender, Time Use, and Poverty in Sub-Saharan Africa* (Washington, DC: World Bank).

World Bank (2007) *World Development Report 2008: Agriculture for Development* (Washington, DC: World Bank).