European Bank
for Reconstruction and Development

Household response to improvements in water infrastructure: Evidence from peri-urban Kyrgyz Republic

Elcin Akcura, Leon Bost and J. Michelle Brock

Abstract

Fresh water is an increasingly scarce resource across the world. To prevent excessive water usage, broad participation in appropriately priced municipal water infrastructure and services is necessary. This paper estimates the impact of improvements to water network infrastructure on the take-up of municipal water services in the Kyrgyz Republic. We also assess the intention-to-treat impacts of these infrastructure improvements on households' willingness to pay for improved water services, and on their time use. We find that infrastructure improvements induce take-up of municipal water taps, but do not generate time savings. The adoption of taps is nevertheless associated with an increase in bathing time, which may mean higher water consumption. New customers also update their willingness to pay after gaining experience with municipal tap connections: they become more willing to pay for better water pressure but less willing to pay for fewer and shorter service outages.

Keywords: water access, adoption, willingness to pay, time use

JEL Classification: D12, L95, Q25, O12, O13, O18

Contact details: J. Michelle Brock, Associate Director, Senior Research Economist, European Bank for Reconstruction and Development, London E14 4BG, UK. Email: brockm@ebrd.com.

Elcin Ackura is a Senior Energy Economist at the World Bank; Leon Bost is European Economist at CRU; J. Michelle Brock is at the EBRD and the CEPR.

We benefited from early comments from William Parienté, Ralph De Haas and participants in the the 8th Annual Life in Kyrgyzstan Conference (2022). We thank Adiya Belgibayeva, Francesca Dalla Pozza and Victoria Marino for excellent research assistance. This work was funded by a technical cooperation grant from the EBRD. The pre-analysis plan can be found at EGAP Registry https://doi.org/10.17605/OSF.IO/N9EWH. Views presented are those of the authors and not necessarily those of the European Bank for Reconstruction and Development or the World Bank.

Working Paper No. 277 **Prepared in February 2023**

Household response to improvements in water infrastructure: Evidence from peri-urban Kyrgyz Republic*

Elcin Akcura† Leon Bost‡ J. Michelle Brock§

February 27, 2023

Abstract

Fresh water is an increasingly scarce resource across the world. To prevent excessive water usage, broad participation in appropriately priced municipal water infrastructure and services is necessary. This paper estimates the impact of improvements to water network infrastructure on the take-up of municipal water services in the Kyrgyz Republic. We also assess the intention-to-treat impacts of these infrastructure improvements on households' willingness to pay for improved water services, and on their time use. We find that infrastructure improvements induce take-up of municipal water taps, but do not generate time savings. The adoption of taps is nevertheless associated with an increase in bathing time, which may mean higher water consumption. New customers also update their willingness to pay after gaining experience with municipal tap connections: they become more willing to pay for better water pressure but less willing to pay for fewer and shorter service outages.

JEL classification: D12, L95, Q25, O12, O13, O18
Keywords: water access, adoption, willingness to pay, time use

*We benefited from early comments from William Parienté, Ralph De Haas and participants in the the 8th Annual Life in Kyrgyzstan Conference (2022). We thank Adiya Belgibayeva, Francesca Dalla Pozza and Victoria Marino for excellent research assistance. This work was funded by a technical cooperation grant from the European Bank for Reconstruction and Development. The pre-analysis plan can be found at EGAP Registry https://doi.org/10.17605/OSF.IO/N9EWH. Views presented are those of the authors and not necessarily those of the European Bank for Reconstruction and Development or the World Bank.

†Senior Energy Economist, The World Bank; Email: eakcura@worldbank.org
‡European Economist, CRU; Email: leonbost.lb@gmail.com
§Associate Director, Senior Research Economist, European Bank for Reconstruction and Development and CEPR, London E14 4BG, UK; Email: brockm@ebrd.com

1 Introduction

Currently, 771 million people globally lack basic drinking water services and 2.3 billion lack basic sanitation services. Given their critical role in economic development, time use and public health, water and wastewater services are a key priority in the sustainable development goals (SDG6: Ensure access to water and sanitation for all) (Meeks, 2017; Beach, 2022).[1] While high-income countries have near universal household coverage of water services, this is often not the case in low and middle income countries, which account for over 80% of the global population. For example, in Vietnam only 53% of the population is connected to a piped water network. This figure is 14% in Nigeria and Bangladesh, and 9% in Uganda. Expansion of water-intensive industries, population growth, urbanization and climate change are all likely to exacerbate current water stress and further compromise access.

The financial sustainability of water utilities requires a large customer base that is willing to pay for water services, because significant investment is needed to maintain and operate the infrastructure (e.g. water mains and sewerage pipe networks). But motivating the adoption of individual household connections to municipal water networks in developing areas is a known challenge (e.g. Devoto et al. (2012)). Utility companies in many of these countries have suffered from decades of under-investment, poor management, and weak financial sustainability. Asset deterioration, poor service quality, and an inability to invest in the requisite infrastructure to meet growing demand are common consequences. This can undermine public trust in the utility companies, motivating households to seek alternative (private) water sources to meet their needs, such as bottled water, vendor water and wells (World Bank, 2020).

Improving service quality can be a key strategy for attracting more customers to the water network. However, the relationships between water service improvements, connec-

[1] https://unric.org/en/sdg-6/

tions to the water network and willingness to pay for services are not well understood. Thus, our primary research aim is to identify whether and to what extent water infrastructure improvements cause households to become utility company customers. We also look at how experience with improvements influences willingness to pay (WTP) for water services. WTP indicates which service attributes are valued most by the population and thus informs not only how improvements change preferences (i.e. whether people's WTP changes after their experience) but also what services people value moving forward. If improvements in the network do not motivate higher uptake by customers and/or increase their WTP for improved services, then the financial sustainability and the reach of the network is likely to be jeopardized.

While there is a handful of papers looking at the impact on households from access to taps, this paper uniquely provides a plausible causal link between the quality of municipal infrastructure and tap adoption. Researchers have also widely explored which aspects of municipal water services people are most willing to pay for, but not how changes in services impact uptake and WTP as we do. Finally, we explore the impact on time use to better understand changes in WTP. Time use is a central outcome when looking at household impacts of water infrastructure improvements. The literature typically looks at changes in time use due to tap adoption. Our work takes a step further by considering how WTP for water services, and therefore viability of the network, is linked to that.

Our study takes place in the Kyrgyz Republic, a post-Soviet lower-middle income country where the existing water infrastructure remains inadequate and run-down. By 2015, only two-thirds of Kyrgyz households had access inside their homes to reliable, safe and clean water (World Bank, 2021). Although the country has striven to improve municipal infrastructure and water services, these efforts have been constrained by lack of financial resources and public resistance to tariff increases.[2] As a legacy of the communist

[2] Tariff increases go through a two-step approval process, first by the Anti-Monopoly Agency, and thereafter by the City council.

regime, many households historically received municipal water without paying for it. The combination of historical and present-day conventions means that the belief that water services should be provided by the government, with no additional cost to households, is widespread. The sensitivity of the topic is further evidenced by bouts of civil unrest linked to tariff increases and distrust of government, the most notably in 2010.

Finally, the urban population in the Kyrgyz Republic has grown rapidly since 2010, presenting an additional challenge to cities and towns that must simultaneously maintain services for existing customers while also expanding coverage and capacity. This is further complicated by abundant use of private wells. Wells can provide a reliable source of water. They may be preferred if households distrust the quality of municipal water or proper governance of the utility company. But wells are also technically and financially difficult to maintain. For example, monitoring well water for microbial and chemical contaminants is a complex and expensive task (Foster et al., 2021).[3] Therefore, it is unclear if people will adopt taps and update their willingness to pay for better water service provisions in response to the infrastructure improvements, or whether they will maintain the position that the service, at any level of quality, should be free.

Our data are from a household survey that we ran before and after a water infrastructure investment project by the European Bank for Reconstruction and Development (EBRD) in the city of Talas. The aim of the project was to improve municipal piped water services in Talas and introduce metered billing. Municipal piped water services distribute network water to a conveniently close location for households to access, for example via house and yard taps. The transition to these network-based services require significant investments for the utility company and households alike. The EBRD project

[3] Water in the Aral Sea Drainage Basin, which covers Talas as well as the majority of the Kyrgyz Republic, is contaminated with high concentrations of agricultural and industrial pollutants, including copper, arsenic and nitrite (Törnqvist et al., 2011). Dichlorodiphenyltrichloroethane (DDT) is also found. Groundwater poses a greater health risk than surface water. But our survey finds that households do not perceive a difference in clarity or taste, between well water and water from the utility company.

included a loan to the local utility company to rehabilitate the existing water network, upgrade several water treatment facilities, improve wastewater infrastructure, and prepare the city for metered billing (meters were distributed in a subsequent phase of the project). The costs of these improvements would be recuperated through service tariffs and connection fees. Establishing new domestic connections was voluntary and unsubsidized. The expected direct benefits to households include cleaner, more reliable, and easier access to water for domestic use.

We use doubly robust difference in differences to establish a plausible causal link between infrastructure improvements and our outcomes of interest. We find that the infrastructure improvements do motivate households to become utility company customers, increasing their adoption of both house taps and yard taps. We also observe a 33% reduction in using wells as the main source of water among households that were originally not utility customers (in our baseline survey). In contrast, the impact on those who were already customers at baseline is small but negative, indicating a slight loss of customers in that segment. This small negative impact should be interpreted cautiously – it is informed by only a small number of households and is not robust to alternative specifications.

Regarding impacts on WTP, we find that results diverge between households that were utility customers at baseline and those that were not. The program had no significant impact on WTP for households that were already customers at baseline. Non-decreasing WTP is an indication that households were largely satisfied with the infrastructure improvements and may suggest that they are slowly adjusting beliefs about who should pay for water services.

Meanwhile, households that were not customers at baseline decrease their WTP for reduced cut duration and frequency, while WTP for water pressure increases. At the time of the baseline survey, these households had no experience with water utility services.

Typically, respondents draw on prior experience and beliefs when stating their valuation of service attributes during a survey. If prior beliefs are based on incorrect information, or limited experience, they may lead respondents to over or underestimate their true value for different service attributes. It is likely that this group of households updated their valuations on becoming utility customers. Finally, our results on time use are similar to Devoto et al. (2012). We do not see an impact on labor market participation from the improvements, and instead observe increases in duration of bathing. Taken together, results suggest that network improvements can increase the size of the customer base, even when there is ready access to clean water from private wells and households do not realize time-related economic gains from adoption.

Our data collection occurs after all physical works were completed but before meters were distributed or installed. This allows us to capture preferences for access to metered tap water without the confounding effects of the experience with the metered billing itself. This is important because we want to isolate the relationship between service quality and WTP. Households may develop a positive WTP for higher service quality, but a negative reaction to meters. For example, even if households were indifferent to metered billing, they may experience disutility related to meters if installation were poorly executed or communication insufficient. As a result, we executed data collection before meter installation. We note, however, that observed impacts on tap adoption will encompass households' *anticipation* of receiving a meter. The measured effect on tap adoption will thus indicate preferences for taps that are metered (rather than unmetered taps that rely on consumption norm based billing). Meanwhile, the impacts of water service quality on WTP and time use should not include this anticipation effect.

The literature on the impacts of household infrastructure in low- and middle-income countries typically looks at time savings, gender roles and labor supply (Ilahi and Grimard, 2000; Koolwal and Van de Walle, 2013). Time savings is a common benefit to households who adopt networked utilities. Having these utilities in the home reduces the time it takes to perform essential household tasks. In the case of getting in-home tap water, households have been shown to reallocate the time to labor (Meeks, 2017) and/or leisure (Devoto et al., 2012). Higher WTP for services may then reflect experience with time saved, increased affordability (in case of increased labor market participation) and higher quality of life.

Time savings from access to water infrastructure are largest in rural communities of developing countries, where people (usually women and children) otherwise have to haul water a long way and can obtain considerable health and time gains with a household tap (Agesa and Agesa, 2019; Dinkelman and Ngai, 2022; Gross et al., 2018; Winter et al., 2021). In these settings, installation of municipal water infrastructure can reduce the millions of hours of time spent on daily water-related chores, freeing up time for paid labor, leisure, or both. Saving labor for women is of particular relevance in the Kyrgyz Republic, which has seen significant decline in female labor force participation in recent years (World Bank, 2021).

Our study most directly contributes to a specific subset of this literature, which looks at changes to urban or quasi-urban infrastructure. These studies consider urban or quasi-urban communities that have existing municipal infrastructure (either street tap or private taps) (Devoto et al., 2012; Chen et al., 2019; Meunier et al., 2019). These communities are unlikely to save much time since water is close to the house already, but households can benefit in other ways. The cornerstone paper in that literature is by Devoto and co-authors (2012), which uses an encouragement study to assess the impacts of house tap adoption on household time use and other non-health attributes in urban Morocco. They find no labor market impacts, and instead an increase in time spent on leisure activities. Chen et al. (2019) also study the impact of access to tap water, in the capital city of Nepal, Kathmandu, focusing on benefits to infrastructure reliability.

Though their study is not causal, they find that having a tap with intermittent water supply, as opposed to not having a tap in the house, is correlated with higher water consumption. Our study fills a gap in this literature by producing causal estimates of the impact of better infrastructure on uptake and WTP, rather than just time allocation.

We further build on the existing literature by exploring tap adoption among households that have viable and nearby alternatives. In our setting, households tend to access nearby private wells; 96% of those that collect water outside of their home travel no more than 100 meters to access it. This is an important aspect as we are able to observe whether households value the benefits of municipal water services enough to become customers even when they have viable alternatives. Devoto et al. (2012) also studies tap use in an urban area, where the average distance to the public tap is over 100 meters. They find that residents save only 27 minutes per day when they eliminate their reliance on public taps.

In a complementary study to ours, Meeks (2017) looks at the impact of installing of public street taps in 38 villages of the Kyrgyz Republic. She finds a 12 to 15% increase in the likelihood that households use a domestic water source that is less than 200 meters from their home. This allows households to save up to 136 minutes per day, on average. Our study builds on this research by looking at the next step in development of water infrastructure – installation of metered household taps.

Regarding WTP, several studies find positive WTP for improved water supply in developing countries (Gidey and Zeleke, 2015; Maliva Islam et al., 2018; Aslam et al., 2018), but none have looked at how changes in services impact WTP. For example, in South Korea and Ghana, higher quality of existing services are found to be associated with higher WTP (Kim et al., 2021; Amoah and Moffatt, 2021). The positive link between WTP and service quality may also be strongest among people who need the water the most: Fujita et al. (2005) show that households having greater restrictions to water supply were willing to pay the most for improved services.[4] But previous studies have not analyzed how these preferences may change over time or in response to service improvements. Our study fills this gap.

Our study further improves upon the WTP literature by using a more robust method to assess WTP. We use a stated preference survey methodology to measure WTP. Specifically, we included several discrete choice experiment (DCE) questions in our survey instrument. The literature is dominated by the contingent valuation method (CVM). The CVM approach involves directly asking people through a survey, how much they would be willing to pay for a good or service attribute. One of the most cited shortcomings of the CVM is that it is scope insensitive; the estimated WTP often fail to vary with the size of the good being evaluated (Kahneman and Knetsch, 1992). Choice experiments, which we use to infer WTP from the hypothetical choices or trade-offs people make in the survey, are seen as a more robust approach.

The paper proceeds as follows. Section 2 provides a description of the study context. Section 3 elaborates on the survey design and data, while Section 4 presents the key hypotheses. The empirical methods are described in Section 5. Results and discussion appear in Sections 6 and 7, respectively. Section 8 concludes.

2 Description of context and treatment

The study site, Talas, is a small, agricultural city in the north of the Kyrgyz Republic, situated 300 km north east of the capital, Bishkek. In 2015, the city had an estimated population of 45,000 people, comprising 8,500 households spread over 13 square kilometers (5.0 sq mi). Until 2015, 61% of households in the Kyrgyz Republic still used private

[4] The level of development and climatic context may matter for WTP results. For example, opposite the literature cited here, Hensher et al. (2005)'s choice experiment to assess WTP for avoidance of water disruptions in Australia found that the more interruptions the respondent faced, the less was their WTP. The authors argue that an increase in the number of disruptions increases the likelihood of taking measures to reduce the impact of disruptions, such as storing water. This averting expenditure then reduces dependence on the municipal water, and WTP.

wells as their main source of water. Talas was typical of the national average, with 64% of households relying on wells. Only 38% of Talas households were utility company customers at baseline, with newly developed, outlying areas less likely to be connected to the water network as more centrally located households (see Table 1).[5] Prior to the EBRD infrastructure project, the water utility company in Talas neither charged rates to support cost recovery, nor did they regularly cut service to customers that were in arrears.

At the time of the study, the water pipe network in Talas comprised approximately 37,000 meters of pipes, was around 50 years old and suffering from considerable leakage. Leaks most often occur at junctures between two pipes. Leaking can start as early in the network as the industrial water pumps that pull the water out of the ground, since these pumps consist of a number of pipe sections joined together. Water is passed from the pumps into large mains pipes, which are limited in length by the size of the trucks that transport them. These must be joined using couplers. Each time the network turns a corner it introduces additional junctures. Finally, smaller diameter pipes are run off the mains to carry water to each building.[6] In addition to failing junctures, leaks can can result from corroded pipes, holes caused by tree roots, faulty parts or accidental damage caused by other works.

The water utility company, Talas Water Company, was producing 4.8 million cubic meters per year for around 28,000 people, three times the national per capita average, at baseline. 1.5 million cubic meters of this can be expected to serve the typical daily needs of customers.[7] The remaining 3.3 million cubic meters was unaccounted for and is likely to have been lost to leaks or used to irrigate backyard agricultural plots (which are

[5] Talas residents live in freestanding houses or in multi-storey dwellings. Multi-storey dwellings are more likely to have both water and wastewater connections.

[6] These smaller pipes are flexible and longer, so require fewer junctures.

[7] An average rate of water consumption per capita of 100-165 liters per person per day, a consumer base of 38% of the 8,500 households and three people per household results in 2.5 million to 4.1 million liters per day. Since there are 1,000 liters in one cubic meter, this comes to 912,500 to 1.5 million cubic meters per year.

common throughout the country (Meeks, 2017)). While the precise volume of water lost to leaks could not be calculated, the EBRD estimated that replacement of leaking pipes and pumps could reduce water losses to 12 per cent of the pre-project level.

Leakage in the water network affects service quality as it can result in low water pressure, water shortages and interrupted service. This is because sending water through heavily leaking pipes and joints reduces the volume of water per cubic meter of pipe which lowers overall water pressure in the network. The low flows can also reduce the ability for multiple households to draw from the mains simultaneously, or for households further from the pumping station to access water at all, during peak times. Prior to the EBRD project, in the worst affected areas of the city, water could only be supplied for 2 hours per day.

Treatment

The EBRD project that this paper evaluates involved repairing or replacing leaky water supply pipes and pumps, and updating service equipment. Works included equipment installation, treatment plant upgrades and leak detection and repairs. The equipment included chlorination and laboratory equipment, leak detection and control equipment, bulk water flow meters and domestic water meters. Domestic water meters were to be offered during a second stage of project implementation, after the study period.[8]

Alongside the network repairs, households were offered the chance to have yard taps and/or house taps installed in their property, so that they could directly benefit from this improved water infrastructure. Households had to fund these private installations. The median cost to convert a well into a house tap was 13,000 Kyrgyz som, about three months' income. The corresponding cost to convert a yard tap to a house tap was about half of that.

[8] The official normative standards in place at the time of the study were 100 liters per capita per day for individual houses and 165 liters per capita per day for flats. Households that preferred not to be metered could either choose not to install a tap, or stop using their tap as their main source of water.

In practice, fixing a leak involves first exposing the pipe, which may be under paved road, and then either patching or replacing the broken segment. Mains pipes are typically buried 0.75 to 1.35 meters below ground level, so digging can involve moving a large amount of earth. After the repair the hole is filled in and, if applicable, the street is re-paved. For damage that is caused by tree roots, repairs may also require (re)moving trees or installing a root guard. Pipe repairs may also require that water supply to the surrounding area be temporarily cut, but it is not always necessary.

Households were informed of all planned project phases prior to the start of the project, as part of an awareness campaign launched to ensure resident buy-in. Information was distributed in all geographical areas of the city, as all areas were to be affected by construction activities (treatment group) and/or tariff adjustments (treatment and control groups). This campaign was run by the Public Relations Unit of the Talas City Administration and included announcements in newspapers, on the radio and on local television. These were managed by way of press releases. The project was also discussed at a public hearing and three public community meetings, each in a different areas of the city. They produced hand-outs for distribution at relevant meetings. Finally, households would obtain information from their own observation of the works, which required the closing of streets in order to dig holes big enough to access large underground pipes.

Expected benefits to households that were are already utility company customers before the infrastructure improvements, include higher water pressure, fewer interruptions and a better overall experience with utility company services. For some households in this group, additional benefits will come from not having to fetch water from a yard tap. The main expected benefit to households that were not connected to the network at baseline was time savings and the convenience of not having to collect water from outside the home. A longer-run expected benefit for this group is not having to spend time or money on maintenance of private water pumps and related equipment.

Defining the treated area

We defined the treatment area based on where repairs occurred and the households whose water supply could be impacted by each repair. Figure 2 shows an up-close segment of the water pipe network in Talas where repairs occured.[9] Blue lines indicate water mains and red lines indicate sewers. Repairs were executed on mains pipes (rather than pipes serving individual households). Households located outside of this area saw no changes to their pipes, so will have had the same or worse quality water services at endline than at baseline.

Repaired leaks should primarily impact households whose feeder pipes occur between the leak and the next intersecting mains pipe. No information was available on the exact households that fed off each segment of the water mains, so we used information on how water networks are constructed to define potentially treated households. In dense pipe networks, mains may run along each street, such that one row of houses on each side of the street is fed from that pipe. Ending points for each segment are cross streets, cul-de-sacs or geographic features (i.e. natural ending points). At a minimum, properties on either side of a repaired pipe segment should be considered to be in the treatment group. In lower density areas, like Talas, the distance from a house to the water mains can be longer; the mains pipe that serves the house may be behind the house or on another block. This will also occur when street configurations violate a grid pattern. Therefore, a larger radius of houses around each repaired segment should also be considered.

To accommodate the uncertainty of exposure to the treatment, and taking into account the lower density of Talas, we define a household as being in the treated area if it is either: (1) in the one row parallel to the repaired pipe segment, along both sides of the pipe or; (2) adjacent to the houses in category (1) (households two rows from the repaired pipe segment). Given the density of mains, it is unnecessary to consider a treatment area

[9] We obtained this map from the EBRD team responsible for the project.

of three rows. By including two rows of homes around each repair we minimize the chance for a contaminated control group. Households situated within 2 blocks of a repaired pipe segment are indicated by the shaded area in Figure 1. Our results on adoption are robust to reducing the treatment area to only the first row of households (see Appendix D.) Results on WTP and time use have the same signs in the smaller treatment area, but less significance. This is likely due to a loss of statistical power.

3 Data

In this section we describe the data and define the main variables of interest. Our data are from a household survey implemented in Talas in 2015 and 2019, before and after the infrastructure improvements (referred to as the baseline and endline surveys, respectively).[10] The survey included questions on residents' main sources of water, types of utility connections, perceived water quality, time use and socio-demographics (such as income and employment status). The survey also included a discrete choice experiment, described in more detail below, which allows us to analyze willingness to pay for water service attributes.

We implemented 1,845 face-to-face interviews collected in Talas over two waves. The first and second waves of the survey entailed 945 and 900 interviews, respectively. Households were randomly sampled within each of four strata, defined by their connection to the municipal water infrastructure at baseline. The categories of connection include: both water and wastewater connections inside the home, water connection inside the home with no wastewater connection, water connection outside the home with no wastewater connection and none of the above. In each sampled household, we interviewed the person responsible for, or knowledgeable about, household finances and paying bills. We refer to this person as the "primary respondent". We also interviewed the person responsible for

[10] The first survey ran from February to June 2015. The second survey was implemented from August to October 2019.

childcare and chores. The former is the most appropriate for providing accurate information on water bills and willingness to pay. The latter is most appropriate to provide accurate information about water chores in the household. For households where this is one and the same person, there is only one respondent.

We over-sampled from the "none of the above" category because it is an important group for whom to measure impacts, and would not occur in large enough numbers without stratification and over-sampling, given the total sample size of the study. To ensure estimates are representative of the population in Talas, all analyses are weighted to adjust for stratification.

Panel

For this paper, we use a sub-sample of 349 households that answered the survey in both waves, for a total of 698 observations. Table 1 contains the breakdown of the baseline sample by connection type and treatment group, for the panel and non-panel, respectively. There is one notable difference in the panel and non-panel: households in the area that received improvements were less likely to join the panel if they had no utility company connection and more likely if they already had both water and wastewater connections. We therefore use survey design and attrition weights to help make the estimates representative. Using these weights, we report balance between treatment and control areas for baseline observables and outcome variables (Table A.1 and Table A.2, respectively). Significant differences reported are from univariate mean comparisons between the two groups. At baseline, households in the treated area are more likely to be apartments in multi-unit dwellings situated in higher density neighborhoods. They tend to report worse water pressure (likely linked to the challenge of achieving adequate pressure in higher floors of apartment buildings). Households in the control area tend to have more rooms, a backyard agricultural plot and are less likely to report receiving remittances.

There are also notable differences in baseline values of the outcomes of interest. Households in the treated area are more likely to already have a water network connection, though at 53% of households, it is not universal. These connections tend to be house taps. Wells are subsequently less relied upon, though still prevalent (47% of households in the treated area rely on wells). They are more likely to have wastewater connections and report shorter bathing episodes than those in the untreated area. Balance tables for the entire sample can be found in the appendix.

The aim for the endline was to collect a second interview with each household that answered the baseline survey. Despite efforts to obtain a balanced panel, only 39% of baseline survey respondents also appear in the endline data.[11] 31% of the baseline sample refused to answer the second wave, while 22% had moved out of town (Table A.3).[12]

If attrition is correlated with the treatment, or if specific groups are not represented in the panel, it undermines the internal validity of the results. This is because people missing from the panel may have responded to the treatment in a systematically different way from the group that remains, thus generating inconsistent coefficient estimates. In section 9, we show evidence that attrition in our sample is "missing at random", which means it can be related to observables and thus accounted for in the estimations. Our very rich data set and large number of observed variables should assuage concerns over unobservables. Therefore, while this level of attrition is undesirable, the parameter estimates may still be consistent and representative.

To further reinforce the validity of the results, we report treatment effect estimates under two different attrition weighting regimes (Figure A.2 and Figure A.3). The weights ensure that groups with higher attrition get higher representation in the coefficient esti-

[11] We executed up to four visits to the same household within one month to obtain the second interview. If the person responsible for or knowledgeable about household finance refused to be interviewed, the interview was conducted with another knowledgeable household member.

[12] We attempted to interview the same person at a different address if possible, but finding movers (based on phone numbers collected five years before) was rare. It happened eight times in total.

mates. The key assumption is that people with the same set of observables would have responded to the treatment the same way. The first weight makes the panel more comparable to the baseline data (which includes panel and non-panel observations). Respondents that have a higher predicted probability of attrition, and are thus less represented in the balanced panel, will get a boost. Those with a low predicted probability get weighed down. We call this the "representative weight". The second weight makes the panel look more like the 60% of households that left the sample. The objective of the second weight is to show whether results might have been different if the majority had remained in the sample and the minority had attrited. Favoring observations that are most comparable to the attriting majority minimizes the influence of people that are very different from the majority. Results are robust to these adjustment to attrition weights. This further supports the assertion that, conditional on covariates, attriters are not markedly different from non-attriters. Results and details about the construction of the weights appear in section 9.

Finally, we find that our panel sample is comparable to the 2014 sample of the Kyrgyz Integrated Household Survey, a nationally representative consumption and expenditure survey run by the Kyrgyz National Statistics Committee every year. We look at the national statistics and statistics specific to the Talas region. The Talas region is largely rural, while Talas city is more urban; not all variables will align at the city and regional level.[13] For example, in terms of the proportion of free-standing houses, number of people per household and prevalence of house taps, our panel sample is more comparable to the national averages (and standard deviations). Meanwhile, the proportion of people having worked in the last week is less than the national average, but in line with regional averages. Self-reported health and prevalence of wells (as the main source of water) are also in line with regional averages. The alignment with national averages on housing

[13] The survey includes data from Talas city itself, but sample sizes are too small to analyze in isolation.

characteristics and household size reflects the peri-urban nature of Talas (neither fully urban nor fully rural). Alignment with regional employment, health and wells reflects more geographically linked conditions. The means and standard deviations for each of these variables are shown in Table 2.

3.1 Main variables of interest

Utility company and house tap adoption

The main outcome variables of interest are whether someone is a utility company customer and whether they use a tap as the main source of water. We rely on self-reported presence of house taps instead of administrative data because the utility company records were unreliable. For example, households were not always billed for the connections they had. We take an incremental approach to obtain more accurate information. The survey first asks households if they are charged for water supplied by the utility company, and if so, for which source of water: in-house, in-yard or from a street tap. Following that, we ask households to indicate the full set of water sources to which they have access (e.g. house tap, yard tap, well, etc.).[14] Finally, we ask respondents to tell us their main source of water.

Willingness to Pay

We use a discrete choice experiment (DCE) to assess willingness to pay (WTP). Choice experiments present the respondents with a series of alternative options to their status quo, and respondents are asked to choose the option they prefer the most. Usually respondents are asked to makes a series of such choices. A monetary value is included as one of the attributes. When respondents choose their preferred option they implicitly make trade-offs between the levels of the attributes in different alternatives, and price,

[14] The list includes: In-house water connection provided by the utility company, In-yard water connection provided by the utility company, In-house water connection from a yard tap, with the yard tap provided by the utility company, In-house water connection from my own yard well, In-yard own well/tap, I get water from a street tap provided by the utility company, I get water from some other source

presented in a choice set (Alpizar et al., 2001). Choice experiments thus allow the estimation of values for specific attributes as well as situation changes presented by the choice options. As long as one of the attributes used to describe the alternatives is monetary, it is possible to estimate respondents' willingness to pay for the attributes of the good in question.

In our DCE, each scenario offers a different quality of water services, where quality is defined by five attributes. Attributes include type of connection to the water network, water pressure, water cleanliness, frequency of service disruptions and duration of disruptions (see Table 4). In each scenario, each attribute takes one of several levels. Some service attribute levels are better in one scenario and some are better in the other. For example, a household will choose between one scenario that provides all the levels of each attribute that they currently enjoy, to another scenario where the water pressure is higher, but the water clarity is lower, with other attributes at the status quo level. Respondent choices are treated as indicating how he/she values the service measures in relation to one another, in accordance with established principles of random utility theory. By randomizing attribute levels to scenarios, and scenarios to each respondent, the DCE data provides causal estimates of how (hypothetical) changes in attribute levels impact preferences.

In our study, respondents evaluated six sets of scenarios.[15] The scenarios each person evaluated were variations on their status quo water and wastewater services. Shown in Table 5, the four status quo scenarios are: no water supplies provided by the utility company and no wastewater connection; in-house water connection provided by the utility company but no wastewater connection; water provided by the utility company available outside the home (in-yard water connection or street standpipe access) but no wastewater

[15] The survey design achieves Bayesian efficiency in a MNL model, which makes use of random priors. The experimental design generates 24 cards blocked into four groups of six. Blocks of six are then randomly allocated to households.

connection; in-house water connection provided by the utility company and wastewater connection.

The scenario listed first in each choice set is always the respondent's status quo (i.e. current level of water and wastewater services). Households do not consider any connection types that are significantly worse than their current connection type. For respondents that already had an in-house water connection provided by the utility company, the levels of water connection in all scenarios were fixed to "in-house". Households that already had either street tap access or in-yard water connections considered scenarios with either their status quo level, in-house water connection or in-yard water connection. Households with no access to water from the utility company considered scenarios that included their status quo, access via street, yard tap or house tap.[16]

The levels of other service attributes are randomized without constraints across scenarios. For example, while a household that already has an in-house connection will only evaluate scenarios that include an in-house connection, the water pressure can be at any level.[17]

Price attributes

In addition to water connection, clarity, pressure and reliability, we randomized the cost of obtaining and using water. This is necessary to generate the WTP measures. The prices we consider include water connection fee and household water bill. Water bill amounts that we offer in the DCE were calibrated to household size.

The experiment includes three levels of the water connection fee, based on estimates for the current connection cost and expected future tariff increases: current level (2,350 KGS), a 20% increase (2,820 KGS) and a 50% increase (3,525 KGS).[18] The water bill levels

[16] In-house tap water connections are included in any scenario where wastewater connections are offered. Households that already have a wastewater connection are only offered scenarios that include that connection.

[17] One of the key features of the DCE instrument is to select valid and appropriate measures of services and their levels. Services and levels were piloted and adjusted accordingly.

[18] Expected future tariff increases were chosen based on conversations with the utility company and assessment of tariff increases that had already been implemented in some cities in the Kyrgyz Republic

were presented in percentage terms (of the respondent's current bill) for households that already receive a water bill from a utility company.[19] The percentage levels included are: 10%, 20%, 30%, 50%, 70%, 100% increases from the current bill. The levels are chosen to vary also around 20-30% increase and include an extreme level as 100%, closer to about 150% which would cover operating costs for the utility company. Households without a water connection at baseline do not receive any services from the utility company and therefore do not get any water bills. For this group, the design envisages the same levels of payment as other groups (a 10%, 20%, 30%, 50%, 70%, 100% increase), but the scenarios are presented in currency rather than per cent increase. Bill amounts are determined by applying the percentage increases to a median expected bill amount, calibrated to household size.

Time use

To observe the time people spend on water-related activities, our survey included a time use diary. The diary systematically collects respondent activities from the day prior to the interview. Interview days were randomized, to yield a representative sample of days of the week.

The time use diary questions were administered by the enumerator, who asked about each activity from the point the respondent woke up. They then entered the activity and the length of the episode into the electronic survey form, using a pre-coded list of 33 activities (see Table 3). Following the convention from the Harmonised European Time Use Survey, respondents were asked to report activities in 15-minute long episodes.[20] Respondents could report up to two simultaneous activities (e.g. cooking and watching

at the time of the baseline survey.

[19] We chose to present the water bill levels in percentage terms due to both the complicated structure of the choice experiment and the technical constraints of the computer program used to administer the survey (CAPI software). To simplify calculations for respondents, households were also shown a table to help them relate how percentage changes result in final bill amounts for average bills per household of 100 KGS, 150 KGS, 200 KGS, 250 KGS and 300 KGS. The percentage show card is included in Figure 3.

[20] This means that activities that take fewer than 15 minutes may not be recorded.

T.V.) during any 15-minute period, and identified for the interviewer which was the primary and which the secondary activity. Finally, we asked respondents whether the day in question was a typical day for them. We can therefore control for atypical days in the analysis.[21]

To create the water-related time use variables, we add up the time spent on water-related activities in the course of the respondent's day. For example, to look at all water-related activities, we add together time spent on washing, bathing, collecting and storing water. We also did robustness checks where we consider food preparation a water-related activity – this does not change results. The summation of time across activities includes when the activity is listed as primary and secondary.

4 Hypotheses

In this section we present our hypothesis for each outcome. The first set of hypotheses are focused on uptake. We then discuss subsequent impacts on willingness to pay and time use.

4.1 Adoption of household water infrastructure

We first assess the impact of infrastructure improvements on adoption of different types of household water infrastructure. When a water utility company makes improvements, this changes the cost-benefit calculation of those households that have not yet signed up. Take, for example, households that rely on backyard wells with clean ground water. While drawing water from the well is extra work, they may decide that an unreliable or low pressure water service from the utility company is not worth the connection fee

[21] We also asked respondents if they had collected water at all in the past week, how many trips they made, and how much time each trip took. Such questions are commonly used to supplement the time use diary, providing information on longer-term time use than that captured by the diary day. For example, respondents were asked: 'How many times over the last week did you collect/store water?', and then, for an indication of duration: 'The last time that you collected/stored water, how long did you spend doing it?'

and water bills. Some houses with private wells even have systems that draw water from the well to the house, giving them running water and/or a flushing toilet. But such private infrastructure means that the household bears the cost of maintenance, for which they may not have expertise and cannot benefit from scale. Improvements in the municipal water infrastructure may therefore induce some unconnected households to join the network.

The treatment could also enable households to obtain more reliable information about the cost of installation, reducing uncertainty that may have inhibited them from installing a hook-up previously. Finally, the presence of the physical works on their street may have acted as a nudge to establish their own connection. We do not distinguish between these possible mechanisms, but list them to justify our hypothesis of a net positive impact.

Hypothesis 1 *Infrastructure improvements will increase the portion of households with a private connection to the municipal water network.*

While we hypothesize a net positive impact, we acknowledge that some households may choose to discontinue their relationship with the utility company. This is unlikely to be the dominant effect due to both the positive changes in service quality, as well as the fact that utility company coverage has already been going up in this city over time. Nonetheless, it is a risk because the program included a promise of future installation of household meters (carried out after the study period, but announced before). Prior to 2019, all billing was norm-based. Norm-based billing is a formulaic approach where the households are billed based on consumption norms, rather than actual consumption.[22] Households that wished to avoid potential increases in their water bill that a meter might bring could do so by reverting to more regular use of a private well. Therefore, testing for

[22] Consumption norms used in Talas prior to metered billing were 100 liters per capita per day for individual houses and 165 liters per capita per day for people in apartment buildings.

heterogeneous impacts for those with and without municipal water network connections at baseline may yield impacts in opposite directions.

Changes in uptake should correspond with installation of house and/or yard taps. Existing customers may upgrade from a yard tap to a house tap, generating an increase in one and a drop in the other. Meanwhile, new customers may install either or both, and stop using their wells. Therefore, we expect the impact on house taps to be at least positive and the impact on wells to be negative. The overall average impact on yard taps will depend on new customers' preference for the cheaper yard tap versus the more expensive, but more convenient, house tap.

Hypothesis 2 *Infrastructure improvements will increase the probability that households use a house tap as their main source of water. There will be a corresponding drop in primary use of wells. There will be a non-zero impact on yard taps in at least one of the heterogeneous groups of interest.*

4.2 Willingness to pay and time use

The second set of hypotheses pertains to willingness to pay and time allocation effects. Since households with and without a municipal water network connection at baseline follow different causal pathways from treatment to outcomes, we apply the hypotheses to each group separately, and not the sample as a whole.

4.2.1 Willingness to pay

As we describe in the methods section, WTP is measured using a discrete choice experiment. This approach allows us to capture the preference for marginal changes in service quality, in terms of the household water bill. Assessing the impact on WTP is especially important for Talas, as the utility company plans to increase prices over time to levels required for sustainably maintaining service levels.

A positive impact on WTP may result from households having a positive experience with the improved services, which helps to dissolve distrust and demonstrate that households can receive value for their money. Alternatively, households that become accustomed to better services may instead decrease their willingness to pay. This could occur if the expectation of paying nothing for municipal water is too entrenched and may be especially prevalent among new customers. We assess the impacts on WTP for each of the service attributes separately.

Hypothesis 3 *Infrastructure improvements will change willingness to pay for water service attributes, including water quality, pressure, low frequency of cuts and low duration of service cuts.*

Heterogeneous effects depend on whether trust or feelings of entitlement dominate the response for each group. For example, households who were not connected to the water network at baseline are likely to experience a more observable change in their water use experience. But their WTP at the time of the baseline survey may have already been calibrated to the expected outcomes such that, if expectations are met, this WTP would remain constant. On the other hand, if their expectations are not met, then the WTP should drop, as an indication of their disappointment with the value for money.

Meanwhile, existing customers may be less likely to change their WTP. If they are not readily able to discern changes in their experience with using the tap water, we could fail to see a change in WTP. But if the improvements are discrete improvements, existing customers may increase WTP. It would signal their preference for further improvements and trust that the utility company will carry them out.

4.2.2 Time use

Finally, to better understand the impact of these improvements on WTP, we assess the impact of infrastructure improvements on time allocation. Juxtaposing impacts on WTP with impacts on time use provides evidence of how activities of daily living frame the demand for water infrastructure and services. For example, if time allocation does not change while WTP does, it is suggestive of the value that households put on convenience.

We hypothesize that improvements in water infrastructure will reduce the amount of time households spend on water-related chores. Water-related chores include fetching and managing water, as well as washing floors, dishes and clothes. For existing customers, time savings will come from the improved water pressure. Better water pressure reduces the time it takes to fill pots and kettles, or to run enough water to bathe. For new customers, the time savings will come from eliminating the need to fetch water from a well.

The existing literature on time use from water infrastructure is sparse and estimates of time savings range widely: from 27 to 136 minutes. Since most households in our sample have nearby private wells, any time savings will likely be on the lower end of the range. We expect a corresponding (small) increase in time spent on labor, child rearing and/or leisure.[23]

Regarding heterogeneous effects, we hypothesize that time use impacts will be largest among treated households that did not already have a municipal water network connection at baseline, compared to those who did.

Hypothesis 4 *Infrastructure improvements will decrease time spent on water-related chores.*

[23] Child rearing includes "Playing with, reading to, helping with homework (own children)", "Helping or caring for adults and children who don't live with you (not as voluntary or paid work)". Labor includes any paid work, and looking for work.

Hypothesis 5 *Infrastructure improvements will increase time spent on child-rearing activities, labor or leisure.*

An important water-related activity that may demonstrate effects in the opposite direction is bathing. This is because bathing may combine necessary work (a water-related chore) with leisure. The work of obtaining the water – either by drawing it at low pressure or fetching it from a well or yard tap – and thus total time spent bathing should decrease as a result of the infrastructure improvements (and hypothesized take-up of taps). On the other hand, having reliable tap water makes bathing more convenient. People who have preferences for better personal hygiene, or who simply enjoy bathing, may then increase their time spent bathing as a result of the improvements. For example, Devoto et al. (2012) find an increase in time spent bathing once people obtain a house tap. It is not clear which of the two effects will dominate in our study, but we hypothesize a non-zero impact.

Hypothesis 6 *Infrastructure improvements will change the duration of bathing episodes.*

5 Empirical strategy

To estimate program impacts, we use both generalized difference in differences (also known as two-way fixed effects, TWFE) and doubly robust difference in differences (DR-DiD) (Sant'Anna and Zhao, 2020). Because this is a difference in differences framework, the estimates capture the average effect of the treatment on the treated (ATT). DRDiD has the advantages of the generalized DiD, which corrects for non-random allocation into the treatment groups, but also allows us to correct for violations of common trends, using inverse probability weights as in Abadie (2005). Standard DiD assumes that, in the absence of treatment, the unobserved differences between treatment and control groups are the same over time. If there are time varying traits correlated with the outcome, creating permanent shifts in the unobserved differences, then two-way fixed effects DiD will be biased.

To generate inverse probability weights, the Sant'Anna and Zhao (2020) DRDiD estimator uses the propensity score, which is an estimate of the propensity to be in the treated group based on observables. These are combined with the conditional expectation, from a standard linear model, to achieve consistent estimators of the ATT. Specifically, the panel DRDiD estimator is a product of the propensity score and conditional expectation. Imbens and Wooldridge (2009) refers to this kind of combined estimator as best practice for DiD estimations.

This method is well suited to our case, since we only have data from two points in time and cannot establish parallel trends empirically. We use panel fixed effects, to improve estimate consistency with respect to time invariant traits. But fixed effects do not properly account for differing dynamics between the treatment and control groups during the study period. Using inverse probability weights solves this by giving a higher weight to observations in the untreated area that are most similar to households in the treatment area (and therefore more comparable dynamically).[24] Finally, we cannot use spatial regression discontinuity due to limited density of observations sufficiently close to the treatment boundary.

We estimate the propensity score component of the DRDiD estimate using a logistic model with treatment assignment as the outcome variable. Independent variables used to construct the propensity score are determined iteratively, to achieve balance and reduce bias. The propensity score calculation includes theoretically relevant variables, as well as variables that help improve comparability between groups, even if not theoretically

[24] At the extreme, if everyone in control is an equally poor or equally good match with the treatment group members, in terms of propensity score, then each control observation will have a weight of 1.

meaningful. The latter are included as they may proxy for important unobservables. For this study, the variables include baseline measures of: being a utility company customer; using house tap (vs yard tap) as the main source of water; using a private well as the main source of water; whether the household is connected to the municipal wastewater network; if the home is a free standing house or a flat in a multi-unit dwelling; number of rooms in the household; economic condition (categorical, described below); if they own a backyard plot; number of people in the household; number of people in the household multiplied by economic condition; the interactions of owing a computer with wastewater connection and with receipt of remittances (see Table A.9 for results from this regression). Figure 5 shows the density plots of the propensity scores for subjects by treatment status and baseline connection status. There is considerable overlap in the supports of the distributions, which means that, in each group, there are households with comparable propensity score in each of the treated and untreated areas. Rubin's B and Rubin's R, measures of bias (which together indicate balance) are in acceptable ranges, at 22.8% and 0.65, respectively.

For the conditional expectation component of the DRDiD, we use a linear probability model, with household fixed effects. We cluster Wild bootstrap error terms at the household level. We also include weights that control for stratification and attrition in all analyses, to make results generalizable to the population. Finally, the propensity score weighting will up-weigh specific households of the treatment and control that have similar scores.

To consistently measure the impact of improvements on WTP and time use, we account for the fact that there is a different causal pathway for people with and without water network connections at baseline. For those who had connections at baseline, changes in WTP and time use will come from changes in service quality, and to some extent from people upgrading from yard to house taps. In fact, benefits can vary endogenously, depending on whether someone upgraded their tap from a yard tap to a house tap. For example, households with higher potential time savings are more likely to adopt. Since we observe who upgrades, we can condition on this change in our estimation. The conditional estimates correct for any simultaneity between the outcome of interest (e.g. time use) and choice of infrastructure (e.g. house tap) (Ferraro and Miranda, 2017; Imbens and Wooldridge, 2009). Conditioning on who changes to and from house and yard taps effectively holds all else equal, and improves comparability between treatment and control groups.

Meanwhile, households that did not have a connection to the municipal water network at baseline will only realise benefits if they acquire one – the causal pathway is through adoption. Theoretically, we should be able to condition on the type of tap adopted (the change from nothing to house tap, or the change from nothing to yard tap). But this does not work in our sample since so few households chose a yard tap. Conditioning on change in tap ownership effectively removes the variation between adoption and non-adoption, and produces null results. In case intensity of adoption matters (house tap verses yard tap), we estimate the models dropping households that switch to yard taps. The results hold. We cannot do the same for households that chose house taps, due to the small sample size for yard tap adopters.

5.1 Estimating WTP for water service attributes

As stated earlier we assess WTP for water services using a stated preference discrete choice experiment (DCE). DCEs leverage insights from the theory of random utility (Luce, 1959; McFadden, 1974). In this subsection we specify how we move from the random utility model to the estimation of WTP.

Let the utility that person i experiences from scenario j be

$$U_{ij} = V_{ij} + \epsilon_{ij} \tag{1}$$

where V_{ij} is a deterministic component of the utility equal to the sum of utilities gained from the level of each attribute x in scenario j. ϵ_{ij} is a independently and identically distributed random variable that captures the unobserved random element of individual i's choice.

If $U_{ij} > U_{ik}$, then person i chooses scenario j (over k), when given the choice between them. The probability of choosing j over k can then be expressed as a binary variable, Y, being equal to 1.

$$\begin{aligned} Prob[Y_i = 1] &= Prob[U_{ij} > U_{ik}] \\ &= Prob[V_{ij} + \epsilon_{ij} - V_{ik} - \epsilon_{ik} > 0|\mathbf{x}_i] \\ &= Prob[\mathbf{x}_i'\beta_j + \epsilon_{ij} - \mathbf{x}_i'\beta_k - \epsilon_{ik} > 0|\mathbf{x}_i] \\ &= Prob[\mathbf{x}_i'(\beta_j - \beta_k) + \epsilon_{ij} - \epsilon_{ik} > 0|\mathbf{x}_i] \\ &= Prob[\mathbf{x}_i'\beta + \epsilon_i > 0|\mathbf{x}_i] \end{aligned} \tag{2}$$

where $\mathbf{x}_{ij}'$ is a vector of the collection of attributes in each scenario. This formulation translates directly to a binary maximum likelihood regression. To obtain individual specific WTP estimates, we use the extension of this basic model that allows for random coefficients (Goett et al., 2000). We fit a mixed logit model using Bayesian methods to analyze WTP for the selected water service attributes. We conduct the analysis in WTP space through a transformation of the coefficient on a price variable. We assume that the coefficient on the price variable follows a log-normal distribution.

Also, for those without water network connections at baseline, we measure ITT impacts on these outcomes, instead of ATT. This is because only households in that group who adopted a tap will experience the improved services.[25] Since opting in is endogenous, we partition the data according to pre-treatment status.

For those that already had water network connections prior to treatment, the im-

[25] We cannot use allocation to infrastructure improvements as an instrument because the treatment was not randomized.

provements will have had direct impacts – due to improved water pressure and shorter duration of cuts. Nonetheless, the magnitude of the impact may depend on whether a household in this group chose to upgrade from a yard tap to a house tap, or even reverted to using wells as the main source of water. We therefore condition these estimates on the main source of water.

6 Results

6.1 Adoption

Mean tap uptake in each wave, by treatment and control group, appears in Figure 4. By 2019, 43% of the respondents in the panel were using water provided by the utility company as their main source of water, an 11 percentage point increase since 2015, averaged across both treatment and control areas. There is also a 12 percentage point increase in the proportion of households that use a house tap as the main source of water. Usage of yard taps as the main source of water does not change markedly between waves, overall. The prevalence of households that use wells as the main source of water went down.

All of these changes are significantly larger in the treated area than the untreated area. Table 8 shows regression results. Dependent variables are listed in the first column, with treatment effect coefficients and p-values reported in the table. Coefficients in columns 2-4 are from TWFE estimations, using sample weights and attrition weights. Columns 5-7 report estimates from the doubly robust difference-in-differences specification, with the same weighting regime.[26]

We see that treated households were statistically significantly more likely to adopt a connection to the water utility network (Columns 2 and 5). They were also more likely

[26] See section 9 for details on attrition weights. The results are robust to excluding the attrition weights. The are also robust to dropping observations that lie on the very edge of town and therefore may have reduced access to a water mains pipe.

to install house taps and decrease the use of wells. An average increase in yard tap usage is, meanwhile, not a significant overall program impact.

Results from partitioning the sample by whether or not each household had a water network connection at baseline appear in Columns 3,4,6 and 7 of Table 8. Households that were not connected at baseline demonstrate a clear pattern of switching from using wells to using house or yard taps. They showed a 28% increase in the adoption of municipal connections, and a 31% decrease in reliance on wells. Unlike the overall results, this group has a significant increase in usage of both house and yard taps, although they adopted house taps as the main source of water twice as often as yard taps.

Meanwhile, some households who were already utility company customers at baseline decreased their usage of house taps, as seen by the negative coefficient in the third panel of Table 8. This is likely linked to an aversion to metered billing. While this is not the norm (the coefficient is small and significance does not survive robustness tests) there is a risk that the company may lose some customers once metered billing ensues. However, such an an effect may also be temporary. It is beyond the scope of this study to speculate further, especially given the tenuousness of the result. Overall we see strong average take-up, with a limited trade-off in terms of losing existing customers.

6.2 Willingness to pay and time use

6.2.1 WTP for water service attributes

We now turn to the effect of improvements on willingness to pay and time use. Impacts on WTP reflect how the experience with improved infrastructure, and associated improvements in service quality, impacted households' preference for different service attributes. Table 9 shows that improved infrastructure increased WTP for water pressure, but only among those households that were not already connected (and paying water bills) at baseline. Meanwhile, their WTP for reduced cut duration and frequency went down. Changes in WTP could indicate that households mis-estimated their true willingness to pay for limited cuts to service, based on their experience with constant (albeit laborious) access to clean water through wells. They will have also experienced first hand the limited differences in taste or clarity of water from the well and water from the municipal network. Increased WTP for improved water pressure, meanwhile, likely reflects the increased time spent bathing.

For the connected households there is no significant difference between their WTP at baseline and endline. These households may have had a sense of their WTP for municipal water services based on experience, which was reaffirmed by the improvements (i.e. did not change). Non-decreasing WTP also implies that these households were not dissatisfied with the infrastructure improvements.

6.2.2 Time use

Time use results appear in the bottom half of Table 9. The infrastructure improvements do not appear to have impacted time spent on water-related chores. This null result appears for all households, regardless of baseline connection status. It also does not matter if water chores are partitioned into ones that use water (e.g. washing dishes) and ones that produce water (e.g. fetching water). Correspondingly, we also do not see significant impacts on time spent in paid labor, child rearing or leisure. We do, however see impacts on bathing time. Among households that were not yet customers when the project began, average bathing time increased by 24.86 minutes more in the treated area than it did for comparable peers in the untreated area (first row in column 1 of Table 9).

For all outcomes, coefficient estimates obtained from using DRDiD are very close to those from generalized (TWFE) DiD. According to Sant'Anna and Zhao (2020), when covariate-specific trends are relevant, the estimand associated with TWFE is not the ATT – it is severely biased. DRDiD, on the other hand, reduces bias considerably, even with mis-specified functional forms. The similarity between the TWFE and DRDiD estimates in our study therefore suggests that, while there may be covariate-specific trends, they are not relevant for identifying the ATT. Irrelevance may be because: (1) dynamic trends differ between the treatment and control groups, but are not correlated with the response to the treatment or (2) dynamic trends do not differ between treatment and control groups (i.e. the parallel trends assumption holds).

Coefficients between models diverge more for households that were connected to the water network at baseline. The DRDiD procedure likely mattered more for creating dynamically comparable groups for this sub-sample due to the average differences in wastewater connections at baseline. Treated households from this group where more likely to also have a wastewater network connection. This can set them on a different trend for usage of house taps and willingness to pay for water services. Using the DRDiD procedure, which included wastewater connection in the propensity score model, therefore helped us get closer to the true ATT for this group. That said, and including the added efficiency of panel data, we do not see meaningful impacts among households that were already connected to the water network (columns 4 and 7 from Table 8, columns 4 and 6 from Table 9).

7 Discussion

Our results show that infrastructure improvements can motivate people to switch to municipal water services, even when it involves costly installation of new equipment. There are no differences in the time that treated and untreated households spend on water-related chores. But households in the treated area do spend more time bathing than untreated households. The treatment effect on bathing is accompanied by positive ITT impacts on willingness to pay for water pressure, the attribute most linked to better bathing experiences.

Meanwhile, our analysis indicates that households' WTP is likely informed by their experience with their newly adopted taps. Those with existing connections show no impacts on WTP, suggesting that fixing leaks does not markedly change the user sentiment.

Households that do adopt the higher cost house taps do not receive observable time-associated economic benefits from using them. Treated households do not save time on water-related activities, so no time is freed up that might be used on income generation. Potential time saving from reduced burden of water-related chores was small to begin with, because households already had access to clean water; 52% had access inside their homes and an additional 46% had access within 100 meters of their homes. This is in contrast to studies on impacts of house tap adoption from less developed countries, which almost exclusively focus on agrarian or remote villages. Without house or street taps, residents in those study areas spend hours, rather than minutes, each time they fetch water.

The last key finding is that improvements may also lead to changes in water consumption. Bathing is the only time-use category where we observe treatment impacts. Bathing can require a lot of water and so it is not surprising that people do it more when water is more convenient to obtain. Thus, in the context of this study, households appear to gain convenience more than time. This result is in line with Devoto et al. (2012) and Chen et al. (2019), both set in urban areas where residents have some existing access to water. This reinforces the strength of our result and suggests that it may be generalizable, to households without indoor plumbing, across cultures and institutional contexts.

8 Conclusion

This study looks at the impact of improving water supply infrastructure on household adoption of municipal water network connections. Household access to municipal water networks is constrained in low and middle income countries. Maintaining water networks, and extending them to unconnected areas, entails significant costs. Water utility companies need a broad, fee-paying customer base to fund loan repayments, invest in maintenance of their network and provide regular service. Countries and cities that have not managed this balance well end up with a legacy of poor water services. In turn, households may revert to alternative (private) water sources. The resulting low subscribership in the municipal water network can lead to limited funds being available for upkeep, which further undermines retention of the consumer base. The problem is particularly pronounced when people have reliable access to alternative sources of water.

Our paper is the first to provide evidence of causality in this relationship. We ask whether repairing rundown municipal water infrastructure can draw customers into the network. We use doubly-robust difference in differences to assess the impact of a water infrastructure improvement project in a peri-urban town in the Kyrgyz Republic. In this setting, households have easy access to water via private wells. The town we analyze has a legacy of poor water services. Meanwhile, the country has historically provided free water services to the population. This has impacted the perceptions and attitudes of people who now must accept water as a priced service. The project we analyze included fixing leaks and the promise of modernized billing systems (to switch from non-consumption based billing to billing based on metered consumption). Households also had the option to install a tap directly into their home or install it in the yard, at their own expense.

In addition to overall impact, we assess heterogenous impacts according to households' water network connection status, prior to the improvements. Further, we study what kind of tap households adopt for their primary source of water. The fee for the yard tap connection is half of that for the house tap. Nonetheless, cheaper yard taps are not as widespread as house taps. Assessing which infrastructure is more popular after improvements is important because it changes the magnitude of the impact (how much it changes someone's life) and therefore the social return on investment. To complement this, we use data on willingness to pay and time use, which help us understand benefits from tap adoption, and preferences for specific water service attributes.

We find that the average effect of the infrastructure improvements on the treated households is a 9% increase in the adoption of municipal connections and a 12% decrease in reliance on wells as the main source of water for domestic use. Using a house tap as the main source of water also increases, by 8%. Yard tap connections only increase for people not already connected at baseline. These effects are a combination of a large increase in utility connections among those not connected at baseline and little to no loss of the existing client base.[27]

For households that were not customers at baseline, the program had a mixed impact on WTP. It increased WTP for water pressure and reduced WTP for reducing frequency and duration of service cuts. There are no corresponding impacts in WTP among those who were already customers at baseline. This may be a positive outcome, since existing customers whose expectations of service are not met would likely have decreased their WTP, which we do not observe in the results. Taken together, the results show that new customers adjust their willingness to pay once they are connected to the network.

An important caveat is that WTP (and time use) could continue to evolve, especially once meters are installed. Therefore, the impact on WTP is mostly indicative of the fact that actions taken to improve municipal services can change consumer sentiment. Continued tracking of WTP may be necessary to fully understand project impacts on households' value for improved water services. Pulse surveys can help track population level satisfaction and WTP for different service attributes.

Finally, our paper provides evidence on the limited time use savings from these kinds

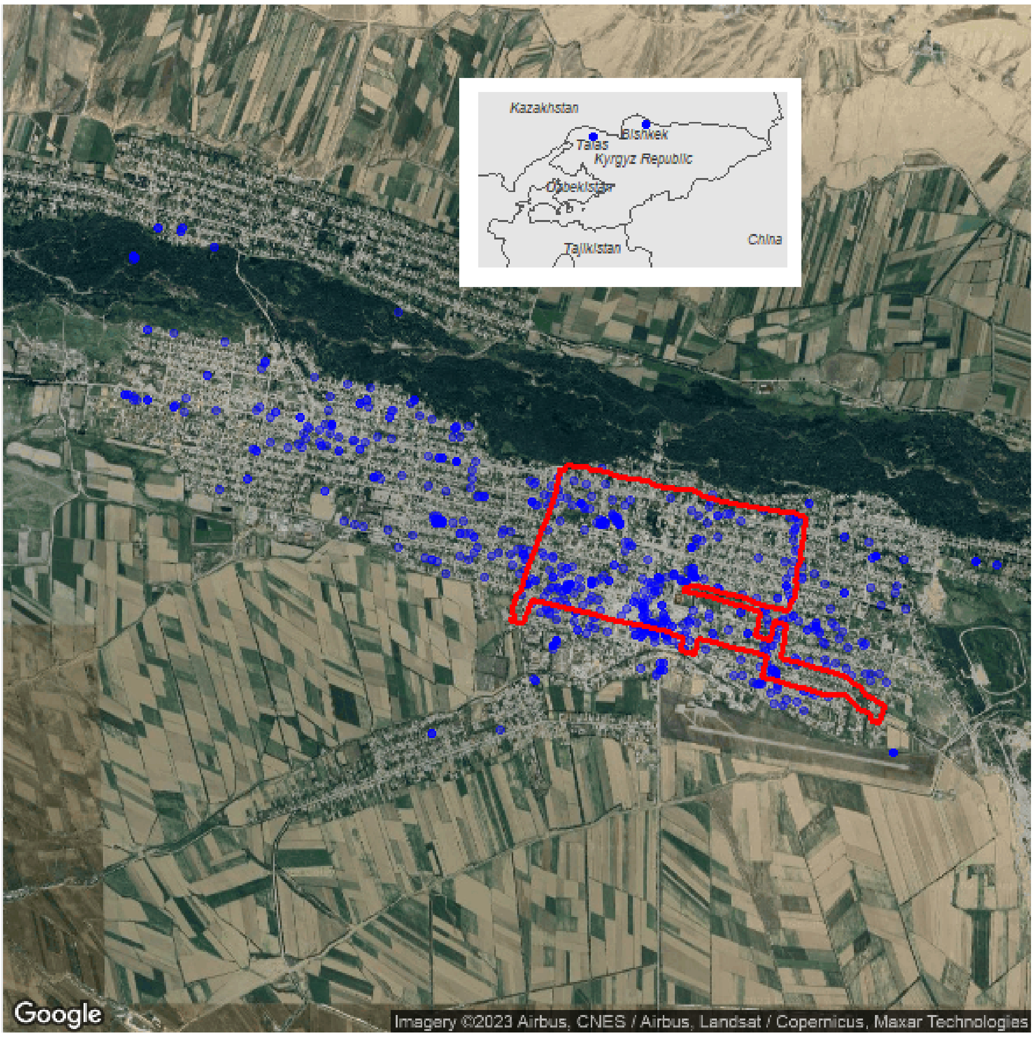
[27] Those who already had connections at baseline may have slightly decreased using a house tap as the main source, but the result is not robust. Any drop in use of municipal water is likely due to the promise of installing household water meters, which was part of the community outreach before the project began. Metered billing can either decrease or increase costs. This can have a negative effect on households with low or uncertain income, or those that do not like the risk of a high bill.

of projects in peri-urban environments. While improved municipal infrastructure can improve the efficiency of household chores, we find no differences in the time that treated and untreated households spend on water-related chores. This is presumably because well water is available less than 100 meters from the home, on average.

Despite the positive treatment effects, overall house tap adoption in Talas remains below 50%. With well water as a viable alternative, policy makers need to continue to maintain the network to attract and maintain new customers. This will lower the per person cost of providing the water and better allow the utility companies to cover costs with water tariffs. Water network extensions and service quality must therefore be well managed to avoid a diminishing willingness to pay bills into the future. Focusing on service attributes that matter most to customers, such as water pressure, will be important. A negative trend in perception could undermine confidence in the new water infrastructure, reduce uptake, and stall further progress.

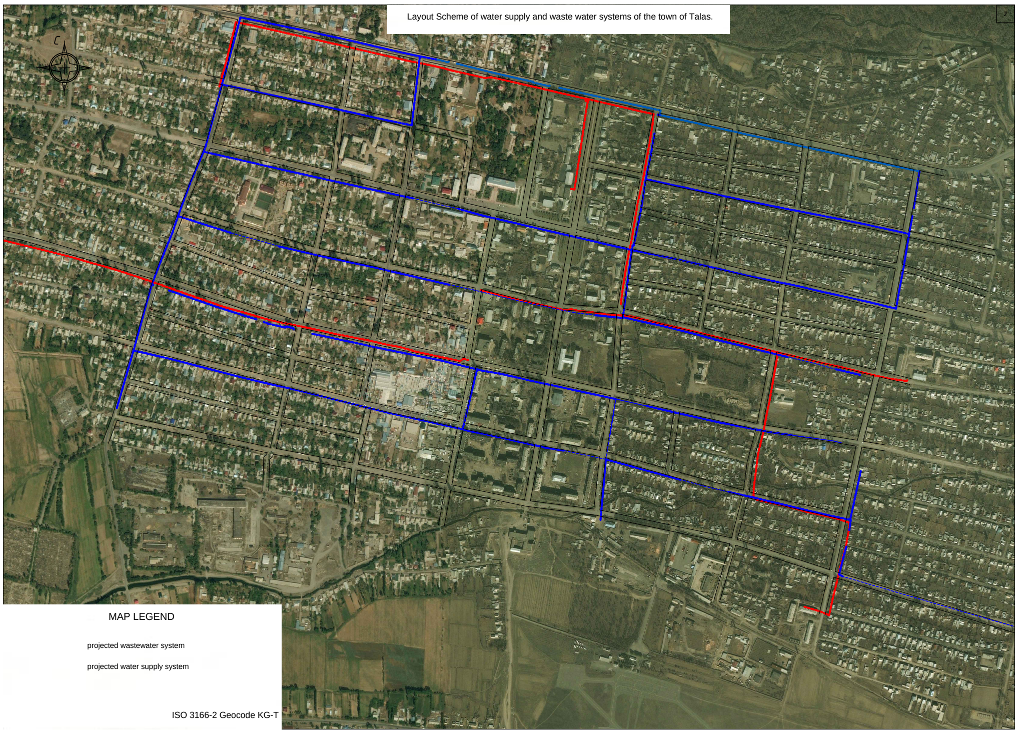
9 Charts and tables

Figure 1: Map of Talas, treatment area and panel sample



Note: All households inside the red shape are situated within 2 city blocks of a repaired pipe, as per data from the water utility company in Talas. Blue geo-location dots indicate approximate location of households in the sample (a wobble was added to protect household anonymity).

Figure 2: Map of the treatment area



Note: This map shows the location of water and wastewater network repairs under the EBRD program in Talas. Black double lines indicate streets. Blue lines indicate municipal water mains that received repairs and red lines indicate wastewater pipes that received repairs.

Table 1: Baseline sample counts and distribution by strata, panel and non-panel

Strata	**Sample did not receive improvements**		**Sample that received improvements**		**Total**	
A: Panel sample						
W/o water network connection	156	0.75	70	0.49	226	0.65
Connected w/ yard tap	13	0.06	10	0.07	23	0.07
Connected w/ house tap, no sewerage	24	0.12	17	0.12	41	0.12
Connected w/ house tap & sewerage	14	0.07	45	0.32	59	0.17
Total	207		142		349	
B: Non-panel sample						
W/o water network connection	271	0.72	101	0.47	372	0.63
Connected w/ yard tap	14	0.04	8	0.04	22	0.04
Connected w/ house tap, no sewerage	40	0.11	30	0.14	70	0.12
Connected w/ house tap & sewerage	53	0.14	74	0.35	127	0.21
Total	378		213		591	

Table 2: Baseline sample compared with KIHS

Variable	Talas, this study		KIHS	
	panel sample	entire sample	national	Talas region
House (vs flat)	0.76	0.76	0.76	0.93
	(0.43)	(0.42)	(0.42)	(0.25)
Household size (no. people)	3.52	3.54	3.91	4.39
	(1.50)	(1.61)	(1.79)	(1.73)
House tap (main water source)	0.33	0.34	0.31	0.09
	(0.47)	(0.47)	(0.46)	(0.29)
Well (main water source)	0.60	0.62	0.14	0.51
	(0.49)	(0.49)	(0.34)	(0.50)
Worked in the last 7 days?	0.28	0.27	0.44	0.26
	(0.45)	(0.45)	(0.50)	(0.44)
Subjective health	4.15	4.13	3.92	4.09
	(0.54)	(0.55)	(0.58)	(0.49)

Note: The table reports mean values, with standard deviation in parentheses. Means from this study data include stratification weights. The panel sample means also includes attrition weights. KIHS refers to the Kyrgyz Integrated Household Survey, obtained from the National Statistics Committee of Kyrgyzstan. KIHS means include stratification weights. Subjective health is average of self-reported health for all household members, scale of 1 to 5.

Table 3: Activities included in the time use diary module

Category	Code	Activity description
Personal care	1	Sleeping
	2	Resting (doing nothing, sitting thinking etc.)
	3	Dressing, undressing, doing make-up, other personal care
	4	Washing, bathing, showering
Eating, drinking	5	Eating or drinking/ having a meal (at home)
Housework	6	Preparing food and drinks, cooking
	7	Sweeping/dusting/tidying house, ironing or mending clothes
	8	Washing floors, dishes, clothes
	9	Fetching/collecting, storing/managing water
	10	Caring for domestic animals (not pets)
	11	Tending crops/vegetables in plot or farm
	12	Home manufacture, selling homemade goods
	13	Other household tasks
Travel	14	Travelling (meaning moving from point A to point B either by walking/jogging, cycling, taking a bus, train, car or any other means of transport)
Employment	15	Work for waged job (include paid, unpaid, overtime and work brought home)
	16	Self-employed work, including work for family business
	17	Buying and selling, acting as a go-between, trade activities
	18	Looking for work
Education & courses	19	Formal education
	20	Recreational courses and study
Voluntary work	21	Voluntary work for or on behalf of an organisation, charity or sports club
Caring for children & adults	22	Physical care of own children
	23	Playing with, reading to, helping with homework (own children)
	24	Caring for adults who live with you
	25	Caring for adults or children who don't live with you (not as voluntary or paid work)
Shopping & appointments	26	Shopping (incl. internet shopping), banking (incl. internet banking), post-office, appointments with doctor, dentist, hairdresser, plumber etc.
Leisure	27	Watching TV and DVDs, listening to radio, reading, at-home hobbies
	28	Working in the garden for recreation
	29	Playing sports, exercising, hunting, fishing
	30	Spending time w/ friends, family, neighbours, or contacting them by phone, text, e-mail, letter etc.
	31	Visits to cinema, club/bar, restaurant, cinema, sporting events, concert/theatre, museum/monuments, library etc.
	32	Attending mosque, church, temple, synagogue, or other religious meetings, praying alone, attending political or other meetings
	33	Other leisure activities

Table 4: Service attributes and levels used in the Discrete Choice Experiment

Service attribute	**Description**	**Possible levels:** one level randomly assigned for each DCE choice
Main source of water		- In-house connection provided by utility company; - Yard tap, connection provided by utility company; - Communal standpipe provided by utility company
Quality of water		- Clean and drinkable - Clean and drinkable with some shortcomings (smell and taste) - Somewhat dirty but sometimes used for drinking
Water pressure	The time it takes to fill a 3 liter kettle (respondents were shown a photo of a typical kettle)	- Strong pressure: less than 10 seconds - Adequate pressure: 10-20 seconds - Low pressure: more than 20 seconds
Frequency of intermittence	How often piped water is interrupted	- Once in 5 years - A few times yearly - Once a month - Several times weekly
Duration of intermittence	Average duration of interruptions	- 1 hour - 2 hours - 5 hours - 12 hours

Table 5: Summary of water connection types available for each strata of households

	Sample Strata			
	Not a utility company customer	**In-house water from utility company, no sewerage**	**Yard tap only from utility company**	**In-house water from utility company, with sewerage**
Option 1	Status quo (“as now”)	Status quo (“as now”)	Status quo (“as now”)	Status quo (“as now”)
Option 2	Water from utility company (connection type is specified and randomly chosen)	Water from utility company, connection is “in-house”	Water from utility company, and never communal	Water from utility company, connection is “in-house”
Option 3	Water from utility company, connection is “in-house”	Water from utility company, connection is “in-house”	Water from utility company, connection is “in-house”	Water from utility company, connection is “in-house”
Option 4	Water services as in option 3 + mains sewerage connection	Water services as in option 3 + mains sewerage connection	Water services as in option 3 + mains sewerage connection	No alternative 4

Note: Each strata has a different set of possible connection types that can be used in the choice experiment. For each DCE choice task, the connection type and levels for each service attribute (Table 4) are randomly chosen. “Communal”, in Option 2 refers to communal standpipes that the utility company provides.

Figure 3: DCE percentage show card

YOUR WATER BILL	100 KGS	150 KGS	200 KGS	250 KGS	300 KGS
AN INCREASE IN WATER BILL					
10%	+ 10 KGS	+ 15 KGS	+ 20 KGS	+ 25 KGS	+ 30 KGS
20%	+ 20 KGS	+ 30 KGS	+ 40 KGS	+ 50 KGS	+ 60 KGS
30%	+ 30 KGS	+ 45 KGS	+ 60 KGS	+ 75 KGS	+ 90 KGS
50%	+ 50 KGS	+ 75 KGS	+ 100 KGS	+ 125 KGS	+ 150 KGS
70%	+ 70 KGS	+ 105 KGS	+ 140 KGS	+ 175 KGS	+ 210 KGS
100%	+ 100 KGS	+ 150 KGS	+ 200 KGS	+ 250 KGS	+ 300 KGS

Table 6: Means of respondent traits at baseline, balanced panel only

Variable	**N**	**Sample did not receive improvements** (1)	**N**	**Sample that received improvements** (2)	**Difference T-test** (1)-(2)
Distance to Talas center	195	4.953 (0.215)	124	2.365 (0.157)	2.588***
House (vs flat)	195	0.861 (0.030)	124	0.624 (0.049)	0.237***
House size (no. rooms)	195	3.617 (0.109)	124	2.972 (0.129)	0.646***
Plot of land	195	0.831 (0.030)	124	0.574 (0.049)	0.257***
Economic condition	195	2.954 (0.079)	124	2.769 (0.085)	0.185
Worked in last 7 days	195	0.313 (0.037)	124	0.250 (0.044)	0.063
Receives remittances	195	0.180 (0.029)	124	0.301 (0.045)	-0.122**
Household size (people)	195	3.580 (0.122)	124	3.406 (0.131)	0.174
Subjective health	195	4.146 (0.043)	124	4.119 (0.050)	0.028
Respondent age	195	42.751 (1.123)	124	43.497 (1.366)	-0.746
Power cuts per year	195	5.234 (0.456)	124	5.350 (0.760)	-0.117
Collects water outside	195	0.358 (0.035)	124	0.371 (0.046)	-0.012
Wastewater connection	195	0.099 (0.026)	124	0.385 (0.049)	-0.286***
Subjective assessment of water attributes					
Quality	195	1.959 (0.017)	124	1.908 (0.033)	0.051
Pressure	195	1.205 (0.043)	124	1.009 (0.048)	0.196***
Min. cuts per year	195	1.842 (0.702)	124	1.857 (0.251)	-0.014
Avg. cut duration	62	3.629 (0.585)	59	3.703 (0.446)	-0.075

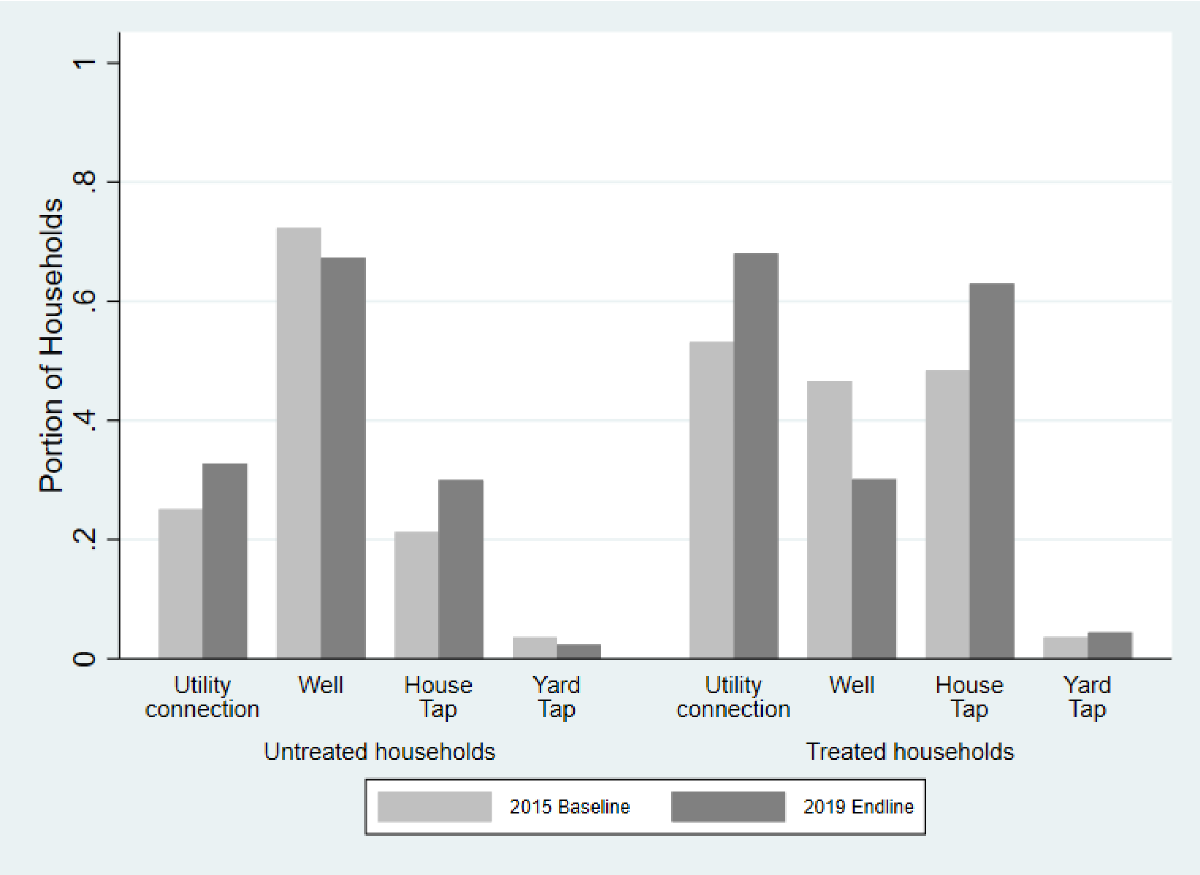
Notes: Weighted means (attrition and design weights) with standard deviations in parentheses. Subjective health is average of self-reported health for all household members, scale of 1 to 5. Economic condition: scale of 1-5, 1 being “Hardly make ends meet, don’t have enough money to buy the most necessary products”, 3 is “Enough for living, but it is difficult to buy some durables, such as furniture, refrigerator, TV” and 5 is “We can buy almost everything we want”. Water quality: 1 means "Somewhat dirty but sometimes used for drinking" , 2 means "Clean and drinkable with some shortcomings (smell and taste)", 3 means "Clean and drinkable". Pressure: based on estimated time to fill a standard kettle. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

Table 7: Means of outcome variables at baseline, balanced panel only

Variable	**N**	**Sample did not receive improvements** (1)	**N**	**Sample that received improvements** (2)	**Difference T-test** (1)-(2)
Main source of water					
Utility company	195	0.251 (0.034)	124	0.532 (0.047)	-0.280***
Well	195	0.723 (0.035)	124	0.466 (0.047)	0.257***
House tap	195	0.215 (0.033)	124	0.495 (0.048)	-0.280***
Yard tap	195	0.036 (0.012)	124	0.036 (0.014)	-0.000
WTP for water service attributes					
Better quality	195	2.274 (0.007)	124	2.270 (0.011)	0.004
Reduced cut frequency	195	2.826 (0.014)	124	2.772 (0.018)	0.054**
Reduced cut duration	195	3.642 (0.067)	124	3.860 (0.112)	-0.217*
Pressure	195	1.793 (0.025)	124	1.811 (0.028)	-0.018
Time use (minutes per day)					
Bathing	195	25.935 (4.665)	124	10.695 (2.536)	15.240***
Water chores	195	114.967 (6.675)	124	113.793 (7.466)	1.174
Non-water chores	195	224.075 (13.680)	124	194.500 (16.996)	29.574
Childcare	195	102.791 (10.383)	124	101.939 (11.534)	0.852
Paid work	195	150.920 (16.368)	124	165.005 (20.913)	-14.085
Leisure	195	167.573 (8.057)	124	154.754 (10.323)	12.819

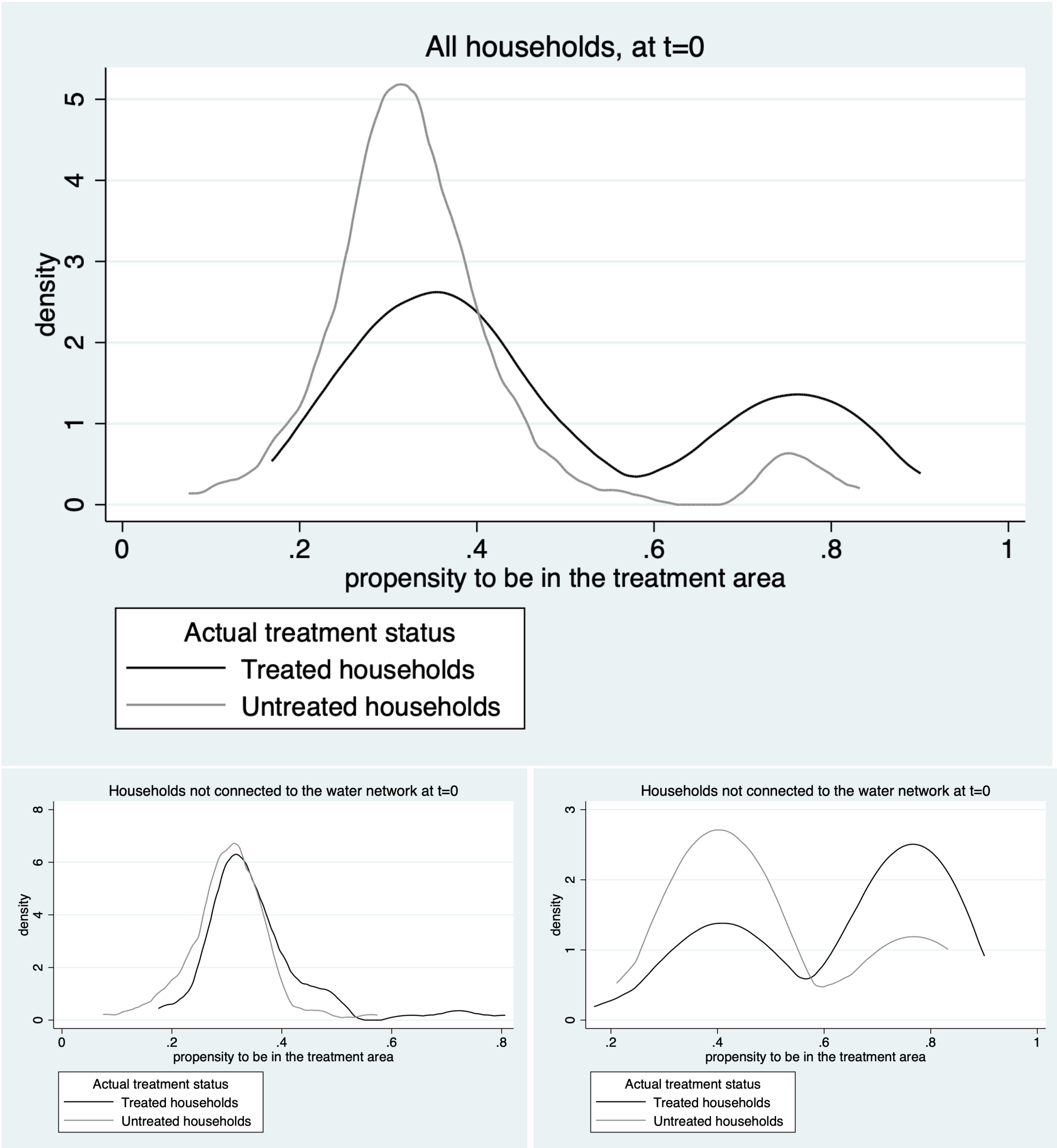
Notes: Weighted means (attrition and design weights) with standard deviations in parentheses. Main source of water: house and yard taps (rows 3 and 5) are subsets of utility company (row 1). Definition of WTP outcomes appear in Table 4. Water chores include "Washing floors, washing dishes, washing clothes" , "Fetching/collecting or storing/managing water". ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

Figure 4: Main source of water in each wave, by treatment group



Note: 95% of households use either a utility connection or well as the main water source. 1% report using both and 3% report using neither. House tap and yard tap are subcategories of "Utility connection". Three households, 0.9%, that use utility company water access it through a street tap (not pictured here). All means include weights.

Figure 5: Propensity score densities, by treatment group and baseline connection status



Note: Kernel densities of the propensity to be in the treated area, based on observable traits. Scores are calculated as predicted values from a logistic regression. The overlapping distributions demonstrate common support between those actually situated in the treatment area and those situated outside. These propensity scores are used for the doubly robust difference in differences estimation, and are different from the attrition propensity scores.

Table 8: Uptake of utility company services and taps, with and without propensity score weights

	Generalized DiD			DRDiD		
Outcome	**Full sample**	**Unconnected at $t = 0$**	**Connected at $t = 0$**	**Full sample**	**Unconnected at $t = 0$**	**Connected at $t = 0$**
Any water network connection	0.06	0.29***	-0.02	0.09	0.28***	-0.06*
	(0.172)	(0.000)	(0.706)	(0.022)	(0.000)	(0.087)
House tap, Main water source	0.06	0.21***	-0.04	0.08**	0.20***	-0.03
	(0.333)	(0.001)	(0.595)	(0.040)	(0.002)	(0.536)
Yard tap, Main water source	0.04	0.09***	0.02	0.03	0.09***	-0.03
	(0.118)	(0.008)	(0.690)	(0.157)	(0.008)	(0.403)
Well, Main water source	-0.10**	-0.31***	0.01	-0.12***	-0.31***	0.05
	(0.030)	(0.000)	(0.795)	(0.003)	(0.000)	(0.138)
N	672	454	218	672	454	218

Note: Results are from panel, difference-in-difference and doubly-robust difference-in-difference estimations, with two time periods, household fixed effects and robust error terms clustered at the household level. Includes stratification and attrition weights. Dependent variables are listed in Column 1. $t = 0$ refers to baseline connection status. Estimations for households connected to the municipal water network at $t = 0$ condition on whether the household is connected via a house or yard tap. p-values are in parentheses. For the reader's convenience, Significance is additionally indicated with * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 9: WTP and Time Use impacts, with and without propensity score weights

		Generalized DiD		**DRDiD**	
	Outcome	**Unconnected at** $t = 0$	**Connected at** $t = 0$	**Unconnected at** $t = 0$	**Connected at** $t = 0$
WTP for:	Quality	-0.02	0.05	-0.02	0.07
		(0.405)	(0.249)	(0.572)	(0.334)
	Pressure	0.14***	-0.02	0.12**	-0.01
		(0.013)	(0.707)	(0.048)	(0.868)
	Reduced cut duration	-0.54***	-0.50	-0.49**	-0.43
		(0.011)	(0.150)	(0.025)	(0.217)
	Reduced cut frequency	-0.07*	-0.01	-0.08**	-0.04
		(0.075)	(0.890)	(0.033)	(0.657)
Time spent:	Water chores	1.53	25.69	4.04	22.44
		(0.909)	(0.168)	(0.781)	(0.369)
	Bathing	27.00***	11.74	24.86***	-0.77
		(0.000)	(0.194)	(0.000)	(0.918)
	labor	-5.55	56.71	3.44	-15.07
		(0.887)	(0.371)	(0.932)	-15.07
	Child rearing	-16.24	35.78	-8.93	29.15
		(0.515)	(0.199)	(0.715)	(0.401)
	Leisure	-10.17	26.06	-13.16	60.76
		(0.693)	(0.433)	(0.647)	(0.153)
	N	454	218	454	218

Note: Results are from panel, doubly-robust difference-in-difference estimations with two time periods, household fixed effects and robust error terms clustered at the household level. Includes stratification and attrition weights. Dependent variables are listed in Column 1. $t = 0$ refers to baseline connection status. Estimations for households not connected to the network at $t = 0$ capture intention to treat. Estimations for households connected to the municipal water network at $t = 0$ condition on whether the household is connected via a house or yard tap. p-values are in parentheses. For the reader's convenience, Significance is additionally indicated with * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

References

Abadie, Alberto, 2005, "Semiparametric difference-in-differences estimators." *The Review of Economic Studies*, *72* (1), 1–19.

Agesa, Richard U and Jacqueline Agesa, 2019, "Time spent on household chores (fetching water) and the alternatives forgone for women in Sub-Saharan Africa: Evidence from Kenya." *The Journal of Developing Areas*, *53* (2).

Amoah, Anthony and Peter G Moffatt, 2021, "Willingness to pay for reliable piped water services: evidence from urban Ghana." *Environmental Economics and Policy Studies*, *23* (4), 805–829.

Aslam, Hina, Jian Liu, Abeer Mazher, Dagne Mojo, Muhammad Imran, and Chao Fu, 2018, "Willingness to pay for improved water services in mining regions of developing economies: case study of a coal mining project in Thar Coalfield, Pakistan." *Water*, *10* (4), 481.

Beach, Brian, 2022, "Water infrastructure and health in US cities." *Regional Science and Urban Economics*, *94*, 103674.

Brock, J Michelle and Leon Bost, 2021, "Impact of water infrastructure improvements on household time-use." *EGAP Registry, published by OSF.* osf.io/n9ewh.

Chen, Yvonne Jie, Namrata Chindarkar, and Jane Zhao, 2019, "Water and time use: evidence from Kathmandu, Nepal." *Water Policy*, *21* (S1), 76–100.

Devoto, Florencia, Esther Duflo, Pascaline Dupas, William Parienté, and Vincent Pons, 2012, "Happiness on tap: Piped water adoption in urban Morocco." *American Economic Journal: Economic Policy*, *4* (4), 68–99.

Dinkelman, Taryn and L Rachel Ngai, 2022, "Time use and gender in Africa in times of structural transformation." *Journal of Economic Perspectives*, *36* (1), 57–80.

Ferraro, Paul J and Juan José Miranda, 2017, "Panel data designs and estimators as substitutes for randomized controlled trials in the evaluation of public programs." *Journal of the Association of Environmental and Resource Economists*, *4* (1), 281–317.

Foster, Tim, Cindy Priadi, Krishna Kumar Kotra, Mitsunori Odagiri, Emily Christensen Rand, and Juliet Willetts, 2021, "Self-supplied drinking water in low-and middle-income countries in the Asia-Pacific." *npj Clean Water*, *4* (1), 37.

Fujita, Yasuo, Ayumi Fujii, Shigeki Furukawa, and Takehiko Ogawa, 2005, "Estimation of willingness-to-pay (WTP) for water and sanitation services through contingent valuation method (CVM): A case study in Iquitos City, The Republic of Peru." *JBICI Rev*, *11*, 59–87.

Garrido, Melissa M, Amy S Kelley, Julia Paris, Katherine Roza, Diane E Meier, R Sean Morrison, and Melissa D Aldridge, 2014, "Methods for constructing and assessing propensity scores." *Health services research*, *49* (5), 1701–1720.

Gidey, Kidu Mezgebo and Ewnetu Zeleke, 2015, "Households' willingness to pay for improved water services in urban areas: A case study from Nebelet town, Ethiopia." *Journal of Development and Agricultural Economics*, *7* (1), 12–19.

Goett, Andrew A, Kathleen Hudson, and Kenneth E Train, 2000, "Customers' choice among retail energy suppliers: The willingness-to-pay for service attributes." *The Energy Journal*, *21* (4).

Greene, William H, *Econometric analysis*, Pearson Education India, 2020.

Gross, Elena, Isabel Günther, and Youdi Schipper, 2018, "Women are walking and waiting for water: The time value of public water supply." *Economic Development and Cultural Change*, *66* (3), 489–517.

Hensher, David, Nina Shore, and Kenneth Train, 2005, "Households' willingness to pay for water service attributes." *Environmental and Resource Economics*, *32* (4), 509–531.

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

Imbens, Guido W and Jeffrey M Wooldridge, 2009, "Recent developments in the econometrics of program evaluation." *Journal of Economic Literature*, *47* (1), 5–86.

IOM, 2021, "International Migration in Kyrgyzstan."

Kahneman, Daniel and Jack L Knetsch, 1992, "Valuing public goods: the purchase of moral satisfaction." *Journal of Environmental Economics and Management*, *22* (1), 57–70.

Kim, Taehyeon, Jihoon Shin, Jinseok Hyung, Kibum Kim, Jayong Koo, and YoonKyung Cha, 2021, "Willingness to pay for improved water supply services based on asset management: a contingent valuation study in South Korea." *Water*, *13* (15), 2040.

Koolwal, Gayatri and Dominique Van de Walle, 2013, "Access to water, women's work, and child outcomes." *Economic Development and Cultural Change*, *61* (2), 369–405.

Luce, R Duncan, "Decision making; an experimental approach." 1959.

Macours, Karen and Teresa Molina Millan, 2017, "Attrition in Randomized Control Trials: Using tracking information to correct bias."

Maliva Islam, Md, Ali Akber, and Md Atikul Islam, 2018, "Willingness to pay for improved drinking water in the southwest coastal Bangladesh." *International Journal of Sustainable Development & World Ecology*, *10* (2), 119–132.

McFadden, Daniel, 1974, "The measurement of urban travel demand." *Journal of Public Economics*, *3* (4), 303–328.

Meeks, Robyn C, 2017, "Water works the economic impact of water infrastructure." *Journal of Human Resources*, *52* (4), 1119–1153.

Meunier, Simon, Dale T Manning, Loic Queval, Judith A Cherni, Philippe Dessante, and Daniel Zimmerle, 2019, "Determinants of the marginal willingness to pay for improved domestic water and irrigation in partially electrified Rwandan villages." *International Journal of Sustainable Development & World Ecology*, *26* (6), 547–559.

Rubin, Donald B, 2001, "Using propensity scores to help design observational studies: application to the tobacco litigation." *Health Services and Outcomes Research Methodology*, *2* (3), 169–188.

Sant'Anna, Pedro HC and Jun Zhao, 2020, "Doubly robust difference-in-differences estimators." *Journal of Econometrics*, *219* (1), 101–122.

Törnqvist, Rebecka, Jerker Jarsjö, and Bakhtiyor Karimov, 2011, "Health risks from large-scale water pollution: Trends in Central Asia." *Environment International*, *37* (2), 435–442.

Winter, James C, Gary L Darmstadt, and Jennifer Davis, 2021, "The role of piped water supplies in advancing health, economic development, and gender equality in rural communities." *Social Science & Medicine*, *270*, 113599.

Wooldridge, Jeffrey M, *Econometric Analysis of Cross Section and Panel Data*, MIT press, 2010.

World Bank, 2022, "World Development Indicators." https://data.worldbank.org/indicator/SM.POP.TOTL?locations=KG (Accessed 23 November 2022).

World Bank, "Central Asia: Towards Water-Secure Sustainable Economies." http://documents.worldbank.org/curated/en/923041591902082151/Central-Asia-Towards-Water-Secure-Sustainable-Economies (Accessed 4 October 2022).

Appendix

A Sampling

To build sampling frames, we used two sets of data that were available at the local authority level for Talas: a) the customer database from the local water company and b) a database of addresses from the local territorial authority. Database (a) contains addresses of all water and sewage service customers. The data are broken down to distinguish between houses and flats. Database (b) provides a full list of addresses for the city. It is from the department that is responsible for organizing collection of local taxes. Combining the lists allowed us to stratify the baseline sample by whether or not a house (or flat) was connected to the municipal water network.[28] We sampled twice as many addresses as the target sample size. If any household refused to answer the survey, it was replaced with the next address on the list, within the same strata.

The aim for the endline was to collect a household panel sample. We executed up to four visits to the same household within one month to obtain the second interview. If the person responsible for or knowledgeable about household finance refused to be interviewed, or was unavailable, the interview was conducted with another knowledgeable household member, so long as they met certain criteria. The replacement respondent should have lived in the household for the past three years and have been responsible for or knowledgeable about the household finances.[29] In cases where the residents had moved, we attempted to interview the same person at a different address, but finding movers (based on phone number collected five years before) was rare. It happened eight times in total.

B Panel attrition

In this section, we analyze whether the imbalance in our sample across waves, caused by attrition, poses a threat to the consistency of our estimates. According to Wooldridge (2010), it is possible to obtain consistent estimates from an unbalanced panel if the attrition is uncorrelated with the (time varying) error term. If attrition is correlated with the error term, results will be biased.[30] Attrition is a threat to internal validity when it results in unrepresentative samples. It threatens external validity when certain groups are completely absent from the sample but would otherwise drive results (i.e. under heterogeneous treatment effects) (Macours and Molina Millan, 2017). All things considered, while attrition in our panel is unconventionally high, it appears to be correlated with observable characteristics in a way we can account for using attrition weights.

We also assess whether the implied level of migration is realistic for Talas, to evaluate the external validity of the sample. In this data, the attrition is not only due to refusal, but due to sampled respondents moving away between waves (Table A.3). If take-up of municipal water and house taps is lower for people

[28] The final population list was cross-referenced with the 2009 national census.

[29] 37.53% of the panel observations fall into the category of having different baseline and endline respondents. Results are robust to controlling for gender differences between members of the household that respond in each wave.

[30] Consistent estimation also requires that the rank condition is satisfied in the fixed effects model run on the panel with attrition. The standard OLS rank condition requires that there are no exact linear relationships between independent variables in the model (i.e. the matrix of independent variables is full rank) (Greene, 2020). Attrition can result in violations of the rank condition if the observations that are gone were pivotal in making sure the condition was satisfied in the first place. Without full rank the matrices cannot be inverted and the estimation fails. We do not encounter this problem with any of our estimations.

that migrate prior to or during the EBRD program, omitting migrants can lead to overestimation of the treatment effects (Macours and Molina Millan, 2017). But if take-up is not likely to differ and migration is not correlated with the treatment, it may be ignorable. This could be the case if migration is normal for this population, such that a large variety of households are equally likely to migrate at any period. Moreover, if the portion of the sample lost to migration is consistent with official statistics, it provides evidence that our attrition from migration may not be an anomaly. Therefore the second part of this appendix describes how that amount of the baseline sample lost to migration may indeed be in line with migration statistics.

B.1 Attrition patterns and weights

Researchers typically use weights, based on baseline correlates of attrition, to help correct for time varying respondent traits that cannot themselves be observed. We do the same; attrition weights are included in the results in Table 8 and Table 9. We explain the construction of the weights here. We also show that results are robust to excluding the weights, and to over-weighing panel observations that are most similar to attriting households.

The first step to designing attrition weights is to assess whether attriters are "missing at random (MAR)". MAR, somewhat counter-intuitively, means that there is a measurable systematic difference between attriters and non-attriters that can be accounted for using other variables in the data set. This is in contrast to "missing completely at random", where there are no correlates (observable or unobservable) of attrition. Researchers use multivariate regression to assess if the attriters are MAR across waves. The attrition dummy is the dependent variable and baseline observables are the independent variables. Determining correlates of attrition allows for construction of weights that make the final estimates more representative. The idea is that households sharing baseline traits also share future dynamic patterns, on average.

Results from such a regression are in Table A.4. Independent variables are listed in the left hand column. We test a large variety of baseline traits, including: tap ownership, willingness to pay, time use, treatment assignment and any potentially theoretically important variables (that might be correlated with treatment response, such as economic condition and household size). Due to our very rich data set, we can include dummies for whether the household receives remittances, ownership of key assets and respondent risk aversion. Risk aversion is important because risk preference may dictate adoption of taps (and trusting water provision to the utility company) more so than other household traits. It is commonly unobservable.

We find evidence that attriters are MAR. Column (1) of Table A.4 shows multivariate correlates of attrition in the full sample. Only household size and respondent age are significantly correlated with attrition (both with $p = 0.10$). Larger households are less likely to attrit, presumably because it is easier to find a respondent in the household at any given time. Meanwhile, older respondents are more likely to be retired, giving them a lower opportunity cost of time spent with the interviewer. Patterns diverge for those who were and were not connected to the water pipe network at baseline. Those connected at baseline who have higher self-reported economic condition and those who collect water outside are more likely to attrit (Column 1 of Table A.6). Meanwhile, among those unconnected at baseline, there is larger attrition in the treatment group (Column 1 of Table A.5).

Attrition that is correlated with treatment assignment is a particular problem if the treatment causes the attrition. We therefore look at attrition correlates for the treatment and control groups, separately (Columns 2 and 3 of each table). In our data, the control group has no observable correlates of attrition. We do, however, identify correlates of attrition for the treatment group. For households that were unconnected at baseline (Table A.5, Column 2), attrition in the treatment group is negatively correlated with owning computers, having children in the house and having higher self-reported health. Meanwhile, people who recently worked, households with more adults and households with more rooms are more likely to attrit. Those that treat water before drinking it and those that have a higher WTP for pressure are also more likely to attrit.

Households in the treated area with connections at baseline demonstrate different attrition patterns. For this group, attrition is more likely among: those having worked recently, those that spend more time on non-water chores, people who obtain water inside the house, people who spend less time on water chores and those who are younger. We also find that attrition in this group is more likely among households further from the city center, but this should not be over interpreted because the treated zone is not very large to begin with.

Overall, no dominant pattern emerges that would suggest the treatment caused attrition. Instead, it seems to be a combination of endogenous selection of households into connected and unconnected properties and that households with more working adults, and fewer children, are harder to keep in the sample of treated households. We propose that these traits are consistent with attrition being caused by opportunity costs of time, rather than the treatment itself.

In summary, we show that attrition, while very high, is seemingly missing at random (MAR). Also, it does not appear that treatment assignment itself will have caused attrition. We identify several correlates of attrition, some of which are important in the treatment group but not in the control group. Using inverse probability weighting can correct for these imbalances, and we turn to this next.

Inverse Probability Weights

Once we have identified correlates of attrition, we can use inverse probability weights to correct for it (Macours and Molina Millan (2017)). We describe the construction of the weights in detail here. Results from Robustness checks appear in Figure A.2 and Figure A.3.

As we discuss in the main text, we produce two different weights: a representative weight and an attriter weight. The aim of the former is to make the results more representative of the original random sample. The aim of the latter is to check if over-weighting attriters (the majority of the original sample) changes the results.

For the representative weight, we generate a binary variable equal to one for all baseline observations and zero for all endline/panel observations. We regress this variable on a collection of covariates using a linear probability model. Weights are the inverse of the predicted values from that regression. Respondents that have a higher predicted probability of attrition, and are thus less represented in the balanced panel, will get a boost while those with a low predicted probability are weighted down. The key assumption is that people with the same set of observables would have responded to the treatment the same way.

The covariates used in the model include those that could be correlated with being in the treatment area and with the outcomes of interest. We start with a large model, including all variables from Table A.4 and their interaction with treatment assignment. We use t-tests to check for balance between the weighted panel sample and the reference group (e.g. the entire baseline sample). If satisfactory balance is not achieved, we drop interactions or variables with very small/insignificant coefficients. We run the regression again and again test for balance. We do this until we achieve balance, as described in Garrido et al. (2014). We measure balance using Rubin's B and Rubin's R. Rubin's B is the absolute standardized difference of linear propensity score means in the two groups. Rubin's R is the ratio of variances of the propensity scores, in the two groups.[31]

The included variables are those listed in Table A.4, minus three variables. The excluded variables are: whether the household stores water, whether they treat water before drinking it and the time

[31] Ideally, for sufficient balance, B should be less than 25% and R should be less than 0.5 (Rubin, 2001).

spent on leisure activities. These are highly correlated with other variables of the model and none of them contribute meaningfully to attrition.[32] Both Rubin's B and Rubin's R from this model are within acceptable ranges, at 18.8% and 1.01, respectively. The black line in Figure A.1 shows the distribution of the representative weights. It ranges from 1.5 to 4.8, and is less than 2.9 for three quarters of the sample.

We generate the attriter weight using the same procedures as the representative weight, but with a different dependent variable. The dependent variable is a binary variable equal to one for all attriting observations and zero for all endline/panel observations. The full interaction model also performs poorly for this dependent variable. The model without interactions performs well, but we gain additional improvements if we allow one interaction, between the treatment dummy and a dummy indicating if there is a connection at baseline. This model achieves Rubin's B equal to 12.5 and Rubin's R equal to 0.88. The grey distribution in Figure A.1 shows the attriter biased weights.

Figure A.2 and Figure A.3 graph treatment effects under no attrition weight, the representative weight and the attriter weight, with 95% confidence intervals. The charts show that the results are robust to all three scenarios.

B.2 Migration

In this section we address the high rate of migration in our sample. Attrition from migration is a common concern for panel surveys in developing economies because it is a common occupational choice (Macours and Molina Millan, 2017). First, it is important to observe that migrants leave from the treatment and control areas in equal proportion. This means that the migration should not impact the internal validity of the estimates.

Second, we ascertain that the level of migration indicated in our data is likely to be standard for Talas. The Kyrgyz Republic has high rates of international migration, most of which is labor migration. In 2015, international migrant stock was just over 200,000 people, or 3.4% of the population of the country (World Bank, 2022). In the five-year period from 2007-2012, the Kyrgyz Republic lost approximately 20 percent of its population to international migration, the 29th highest globally, and similar in proportion to Sudan (in the same period) (World Bank, 2022). Most of this is labor migration – over 30% of Kyrgyz GDP comes from remittances.

Rural to urban migration is also prevalent in the Kyrgyz Republic. According to the IOM, Talas Province lost 5% of its population in the 10 years prior to 2020 to internal migration (IOM, 2021). Other provinces fared similarly, with rates ranging from 5-11%. Moreover, the proximity of Talas to the capital city, Bishkek, facilitates migration over daily commuting. According to Google Maps, the drive would take around five hours.

Thus, outside data sources indicate that the level of migration implied by our data may be an upper bound but is not extreme. In any given year there may be considerable numbers of people who cannot be reached, either due to permanent or temporary migration. Assuming that correlates of migration are not changing over time, the results may then be considered externally valid, generalizable to other towns with similar labor mobility.

Taken together, while the sample suffers from severe attrition, there is evidence to suggest that the attrition is seemingly random. Our very rich data set reduces the risk that a theoretically important group is unaccounted for by the attrition weights. Lastly, the implied rate of migration over the five-year period of our study seems to be consistent with outside data sources. We therefore think that the

[32] Each of these three variables has severely decreased balance in the matched sample. Rubin's B from a linear model that includes all variables listed in Table A.4 is 28%. The desirable threshold of Rubin's B is 25%.

evidence points in the direction of attrition not being a major threat to the validity of the results.

Table A.1: Baseline descriptive statistics, all observations (panel + non-panel)

Variable	**N**	**Sample did not receive improvements** (1)	**N**	**Sample that received improvements** (2)	**Difference T-test** (1)-(2)
Distance to Talas center	591	4.916	303	1.674	3.242***
		(0.102)		(0.048)	
House (vs. flat)	597	0.807	303	0.515	0.292***
		(0.019)		(0.030)	
House size (no. rooms)	597	3.381	303	2.871	0.509***
		(0.061)		(0.077)	
Owns a plot of land	597	0.751	303	0.499	0.252***
		(0.020)		(0.030)	
Economic condition	597	2.789	303	2.671	0.118
		(0.043)		(0.057)	
Worked in the last 7 days	597	0.319	303	0.291	0.028
		(0.020)		(0.028)	
Household size (people)	597	3.358	303	3.091	0.268**
		(0.064)		(0.085)	
Subjective health	590	4.151	303	4.160	-0.009
		(0.025)		(0.038)	
Respondent age	597	42.303	303	40.944	1.359
		(0.613)		(0.874)	
Power cuts per year	597	5.105	303	5.676	-0.571
		(0.262)		(0.544)	
Collects water outside	590	0.361	303	0.283	0.077**
		(0.020)		(0.026)	
Wastewater connection	597	0.167	303	0.467	-0.300***
		(0.019)		(0.030)	
Subjective assessment of water attributes					
Quality	597	1.935	303	1.918	0.017
		(0.013)		(0.018)	
Pressure	597	1.184	303	1.080	0.103**
		(0.024)		(0.037)	
Min. cuts per year	597	2.085	303	2.604	-0.519
		(0.511)		(0.542)	

Notes: Weighted means (attrition and design weights) with standard deviations in parentheses. Subjective health is average of self-reported health for all household members, scale of 1 to 5. Economic condition: scale of 1-5, 1 being “Hardly make ends meet, don’t have enough money to buy the most necessary products”, 3 is “Enough for living, but it is difficult to buy some durables, such as furniture, refrigerator, TV” and 5 is “We can buy almost everything we want”. Water quality: 1 means "Somewhat dirty but sometimes used for drinking" , 2 means "Clean and drinkable with some shortcomings (smell and taste)", 3 means "Clean and drinkable". Pressure: based on estimated time to fill a standard kettle. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

Table A.2: Baseline descriptive statistics, all observations (panel + non-panel)

Variable	**N**	**Sample did not receive improvements** (1)	**N**	**Sample that received improvements** (2)	**Difference T-test** (1)-(2)
Main source of water					
Utility company	597	0.296	303	0.612	-0.316***
		(0.021)		(0.028)	
Well	597	0.687	303	0.392	0.296***
		(0.021)		(0.028)	
House tap	597	0.260	303	0.585	-0.325***
		(0.020)		(0.029)	
Yard tap	597	0.035	303	0.027	0.009
		(0.007)		(0.009)	
WTP for water service attributes					
Better quality	597	2.268	303	2.285	-0.017**
		(0.005)		(0.007)	
Reduced cut frequency	597	2.809	303	2.781	0.028*
		(0.009)		(0.013)	
Reduced cut duration	597	3.621	303	3.960	-0.339***
		(0.040)		(0.059)	
Pressure	597	1.784	303	1.859	-0.076***
		(0.013)		(0.018)	
Time use (minutes per day)					
Bathing	597	20.739	303	17.850	2.889
		(1.790)		(2.757)	
Water chores	597	103.151	303	106.570	-3.419
		(3.531)		(4.561)	
Non-water chores	597	201.012	303	175.004	26.008**
		(7.537)		(10.028)	
Child rearing	597	81.240	303	100.092	-18.851*
		(5.312)		(8.423)	
Paid work	597	149.824	303	172.988	-23.164
		(8.760)		(13.580)	
Leisure	597	166.942	303	165.530	1.412
		(5.203)		(7.455)	

Notes: Weighted means (attrition and design weights) with standard deviations in parentheses. Main source of water: house and yard taps (rows 3 and 5) are subsets of utility company (row 1). Definition of WTP outcomes appear in Table 4. Water chores include "Washing floors, washing dishes, washing clothes" , "Fetching/collecting or storing/managing water". ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

Table A.3: Baseline sample attrition composition

Attrition reason	**Frequency**	**Percent of total**	**Percent of attriters**
Panel household	349	38.78	
Respondent refusal	281	31.22	51.00
Respondent moved/migrated	196	21.78	35.57
Other attrition reason	74	8.22	13.43
Total	900	100.00	100.00

Notes: We classify as "Other attrition reasons" all households with missing data for this variable. We cross-checked to confirm that these households were indeed non-panel – all panel and non-panel households were verified using addresses and household roster information from each wave.

Table A.4: Correlates of attrition for Full sample

Dep var: Attrition dummy	Full Sample (1)		Treatment (2)		Control (3)	
In treatment area	0.274	(0.164)				
Water utility connection	-0.194	(0.704)	-0.881	(0.363)	-0.166	(0.789)
House tap, main water source	0.484	(0.325)	0.263	(0.787)	0.474	(0.429)
Well, main water source	0.329	(0.369)	0.070	(0.938)	0.410	(0.354)
Wastewater connection	0.230	(0.550)	-0.484	(0.440)	1.145*	(0.042)
Treats water (e.g. boiling)	-0.011	(0.964)	0.143	(0.736)	0.052	(0.870)
Stores water	-0.107	(0.734)	-0.494	(0.577)	-0.139	(0.696)
Collects water outside	-0.178	(0.341)	-0.625	(0.070)	0.043	(0.855)
Distance to Talas center	7.183	(0.366)	43.736	(0.263)	6.859	(0.422)
House (vs. flat)	-0.039	(0.923)	-0.850	(0.266)	0.157	(0.747)
Household size (people)	-0.180*	(0.031)	0.170	(0.385)	-0.263**	(0.008)
No. children	0.065	(0.534)	-0.493*	(0.041)	0.187	(0.134)
House size (no. rooms)	-0.011	(0.875)	0.096	(0.510)	-0.014	(0.866)
Worked in last 7 days	0.299	(0.090)	0.863*	(0.016)	0.111	(0.607)
Respondent age	-0.012*	(0.038)	-0.015	(0.173)	-0.009	(0.227)
Subjective health	-0.126	(0.420)	-0.236	(0.424)	-0.137	(0.486)
Risk tolerance	0.056	(0.175)	-0.024	(0.771)	0.083	(0.105)
Receives remittances	-0.214	(0.257)	-0.449	(0.201)	-0.091	(0.699)
Economic condition	-0.115	(0.280)	-0.448*	(0.024)	0.046	(0.732)
Power cuts per year	-0.093	(0.451)	-0.103	(0.653)	-0.088	(0.597)
Owns a plot of land	0.055	(0.864)	-0.076	(0.911)	0.052	(0.886)
Owns a PC	-0.260	(0.057)	-0.442	(0.193)	-0.165	(0.241)
Has a car	0.061	(0.686)	0.410	(0.186)	-0.072	(0.694)
Has bank account	0.184	(0.251)	0.091	(0.772)	0.225	(0.254)
Bathing (min.)	0.001	(0.768)	0.001	(0.876)	-0.000	(0.986)
Water chores (min.)	-0.001	(0.147)	-0.005*	(0.017)	-0.001	(0.532)
Non-water chores (min.)	0.000	(0.523)	0.002	(0.062)	-0.000	(0.521)
Paid labor (min.)	-0.000	(0.700)	0.001	(0.455)	-0.000	(0.454)
Leisure (min.)	-0.000	(0.631)	0.000	(0.875)	-0.001	(0.314)
Child rearing (min.)	-0.001	(0.375)	0.001	(0.464)	-0.001	(0.201)
WTP quality	-0.209	(0.776)	0.117	(0.933)	-0.407	(0.658)
WTP reduced cut frequency	0.492	(0.272)	0.491	(0.554)	0.091	(0.886)
WTP reduced cut duration	0.050	(0.606)	0.136	(0.432)	0.094	(0.456)
WTP pressure	-0.056	(0.851)	0.983	(0.107)	-0.522	(0.153)
Observations	857		291		566	
prob(f) > F	0.247		0.134		0.283	

Notes: Weighted means (attrition and design weights) with standard deviations in parentheses. Subjective health is average of self-reported health for all household members, scale of 1 to 5. Economic condition: scale of 1-5, 1 being "Hardly make ends meet, don't have enough money to buy the most necessary products", 3 is "Enough for living, but it is difficult to buy some durables, such as furniture, refrigerator, TV" and 5 is "We can buy almost everything we want". Water quality: 1 means "Somewhat dirty but sometimes used for drinking" , 2 means "Clean and drinkable with some shortcomings (smell and taste)", 3 means "Clean and drinkable". Pressure: based on estimated time to fill a standard kettle. Risk tolerance is self-evaluated on a scale of 1-10. *p*-values in parentheses ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

Table A.5: Correlates of attrition for households **Unconnected at** $t = 0$

Dep var: Attrition dummy	Full Sample (1)		Treatment (2)		Control (3)	
In treatment area	0.638*	(0.016)				
Well, main water source	0.354	(0.452)	1.665	(0.279)	0.194	(0.713)
Treats water (e.g. boiling)	0.095	(0.769)	2.051*	(0.038)	0.108	(0.764)
Stores water	-0.117	(0.753)	-0.375	(0.813)	-0.275	(0.502)
Collects water outside	0.049	(0.832)	1.028	(0.119)	-0.102	(0.701)
Distance to Talas center	9.464	(0.314)	-57.812	(0.342)	11.600	(0.257)
House (vs. flat)	0.058	(0.922)	0.150	(0.885)	0.277	(0.687)
Household size (people)	-0.174	(0.075)	0.844*	(0.031)	-0.273*	(0.012)
No. children	0.040	(0.748)	-1.152**	(0.003)	0.164	(0.244)
House size (no. rooms)	0.039	(0.627)	0.657**	(0.009)	-0.034	(0.700)
Worked in last 7 days	0.168	(0.427)	1.439*	(0.012)	0.062	(0.800)
Respondent age	-0.011	(0.132)	-0.010	(0.573)	-0.015	(0.085)
Subjective health	-0.121	(0.528)	-1.098*	(0.034)	-0.098	(0.657)
Risk tolerance	0.033	(0.484)	-0.207	(0.061)	0.060	(0.290)
Receives remittances	-0.383	(0.108)	-0.194	(0.786)	-0.319	(0.237)
Economic condition	0.072	(0.597)	-0.332	(0.340)	0.107	(0.498)
Power cuts per year	-0.122	(0.419)	-0.294	(0.356)	-0.036	(0.844)
Owns a plot of land	0.136	(0.688)	0.151	(0.877)	0.197	(0.597)
Owns a PC	-0.305	(0.064)	-1.984**	(0.002)	-0.202	(0.192)
Has a car	-0.003	(0.989)	0.045	(0.933)	0.004	(0.985)
Has a bank account	0.267	(0.181)	-0.202	(0.719)	0.326	(0.149)
Bathing (min.)	-0.000	(0.905)	-0.005	(0.087)	0.000	(0.989)
Water chores (min.)	-0.001	(0.297)	-0.005	(0.119)	-0.001	(0.621)
Non-water chores (min.)	-0.000	(0.893)	0.001	(0.542)	-0.000	(0.728)
Paid labor (min.)	-0.000	(0.648)	0.002	(0.154)	-0.001	(0.356)
Leisure (min.)	-0.000	(0.634)	0.002	(0.362)	-0.001	(0.389)
Child rearing (min.)	-0.001	(0.214)	-0.004	(0.091)	-0.001	(0.292)
WTP quality	-0.769	(0.419)	-0.659	(0.750)	-1.049	(0.347)
WTP reduced cut frequency	-0.750	(0.366)	0.008	(0.998)	-0.694	(0.466)
WTP reduced duration cuts	0.219	(0.277)	0.165	(0.814)	0.171	(0.437)
WTP pressure	-0.367	(0.346)	3.208**	(0.007)	-0.925*	(0.038)
Constant	5.115	(0.199)	-1.162	(0.920)	6.818	(0.146)
Observations	562		134		428	
p	0.674		0.107		0.786	

Notes: Weighted means (attrition and design weights) with standard deviations in parentheses. Subjective health is average of self-reported health for all household members, scale of 1 to 5. Economic condition: scale of 1-5, 1 being “Hardly make ends meet, don’t have enough money to buy the most necessary products”, 3 is “Enough for living, but it is difficult to buy some durables, such as furniture, refrigerator, TV” and 5 is “We can buy almost everything we want”. Water quality: 1 means "Somewhat dirty but sometimes used for drinking" , 2 means "Clean and drinkable with some shortcomings (smell and taste)", 3 means "Clean and drinkable". Pressure: based on estimated time to fill a standard kettle. Risk tolerance is self-evaluated on a scale of 1-10. p-values in parentheses ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

Table A.6: Correlates of attrition for households **Connected at** $t = 0$

Dep var: Attrition dummy	Full Sample (1)		Treatment (2)		Control (3)	
In treatment area	-0.271	(0.448)				
House tap, main water source	0.154	(0.818)	0.457	(0.747)	1.013	(0.251)
Well, main water source	0.362	(0.614)	0.839	(0.488)	1.335	(0.272)
Wastewater connection	0.018	(0.970)	-1.771	(0.096)	1.382	(0.132)
Treats water (e.g. boiling)	-0.543	(0.274)	-1.854	(0.075)	-0.362	(0.727)
Stores water	0.582	(0.514)	-1.847	(0.325)	1.122	(0.456)
Collects water outside	-0.835*	(0.048)	-2.143**	(0.002)	0.790	(0.308)
Distance to Talas center	4.871	(0.799)	259.191**	(0.009)	10.855	(0.660)
House (vs. flat)	0.096	(0.914)	-3.033	(0.103)	-0.671	(0.686)
Household size (people)	-0.320	(0.090)	-0.436	(0.192)	-0.411	(0.163)
No. children	0.195	(0.388)	0.148	(0.699)	0.462	(0.185)
House size (no. rooms)	-0.106	(0.486)	-0.163	(0.491)	0.131	(0.640)
Worked in the last 7 days	0.738	(0.052)	1.394*	(0.029)	0.327	(0.595)
Respondent age	-0.019	(0.099)	-0.046*	(0.010)	0.018	(0.388)
Subjective health	-0.203	(0.481)	-0.357	(0.423)	-0.262	(0.627)
Risk tolerance	0.160	(0.094)	0.141	(0.436)	0.315	(0.082)
Received remittances	0.008	(0.983)	-0.361	(0.503)	0.575	(0.389)
Economic condition	-0.414*	(0.033)	-0.471	(0.201)	-0.197	(0.586)
Power cuts per year	-0.244	(0.356)	0.023	(0.952)	-0.536	(0.284)
Owns a plot of land	-0.324	(0.740)	-0.023	(0.990)	0.842	(0.653)
Owns a PC	-0.009	(0.976)	0.010	(0.985)	0.019	(0.972)
Has a car	0.259	(0.409)	0.798	(0.111)	-0.636	(0.250)
Has a bank account	0.012	(0.969)	0.215	(0.662)	-0.231	(0.656)
Bathing (min.)	-0.001	(0.861)	0.019	(0.153)	0.000	(0.971)
Water chores (min.)	-0.003	(0.152)	-0.013**	(0.002)	-0.001	(0.809)
Non-water chores (min.)	0.002	(0.134)	0.004*	(0.046)	-0.001	(0.568)
Paid labor (min.)	-0.000	(0.955)	-0.002	(0.311)	0.001	(0.347)
Leisure (min.)	-0.001	(0.523)	-0.003	(0.363)	-0.001	(0.801)
Child rearing (min.)	0.001	(0.499)	0.002	(0.259)	-0.001	(0.651)
WTP quality	-0.385	(0.813)	0.982	(0.698)	0.591	(0.855)
WTP reduced cut frequency	1.103	(0.110)	1.330	(0.289)	0.959	(0.463)
WTP reduced cut duration	0.106	(0.428)	0.389	(0.119)	0.072	(0.729)
WTP pressure	0.486	(0.515)	1.395	(0.241)	0.557	(0.663)
Constant	1.093	(0.817)	-3.025	(0.717)	-4.473	(0.555)
Observations	295		157		138	
p	0.703		0.364		0.566	

Notes: Weighted means (attrition and design weights) with standard deviations in parentheses. Subjective health is average of self-reported health for all household members, scale of 1 to 5. Economic condition: scale of 1-5, 1 being “Hardly make ends meet, don’t have enough money to buy the most necessary products”, 3 is “Enough for living, but it is difficult to buy some durables, such as furniture, refrigerator, TV” and 5 is “We can buy almost everything we want”. Water quality: 1 means "Somewhat dirty but sometimes used for drinking" , 2 means "Clean and drinkable with some shortcomings (smell and taste)", 3 means "Clean and drinkable". Pressure: based on estimated time to fill a standard kettle. Risk tolerance is self-evaluated on a scale of 1-10. p-values in parentheses ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

Figure A.1: Representative attrition weights

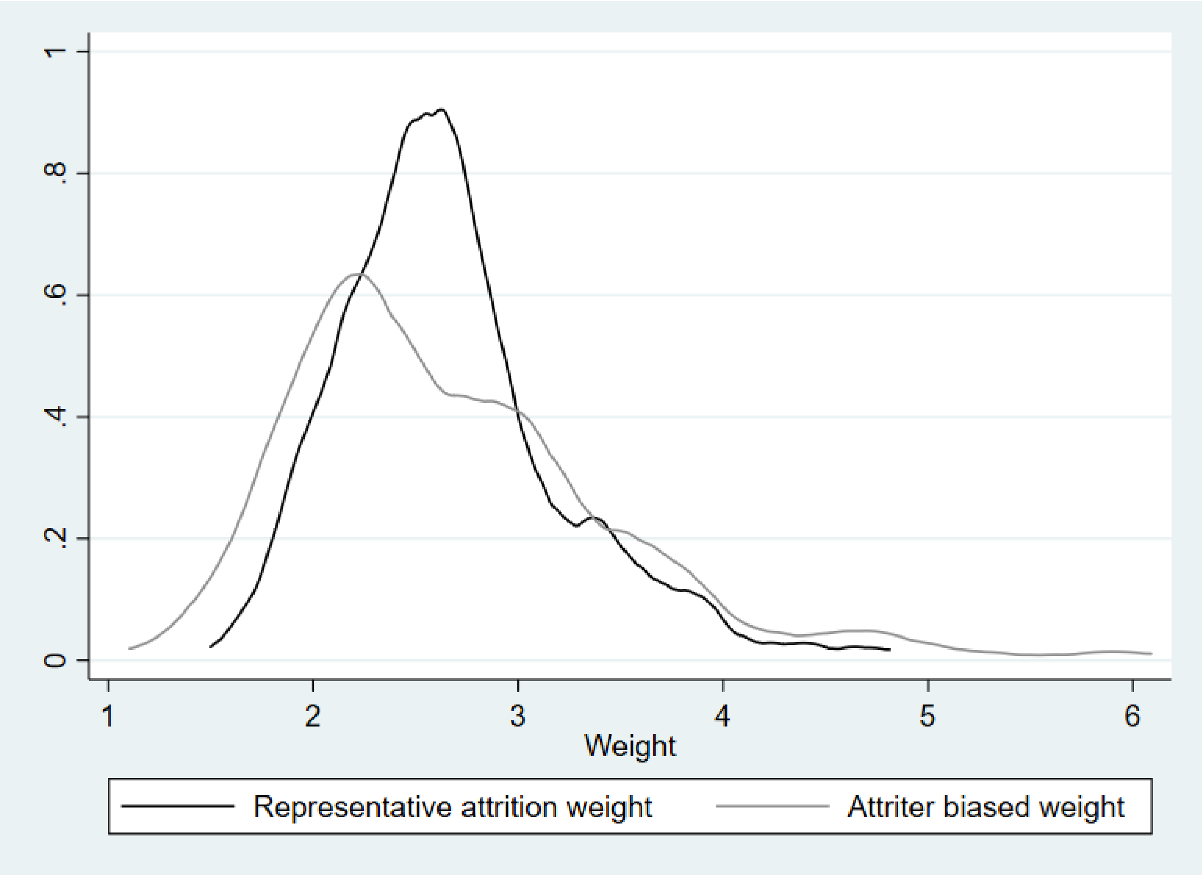
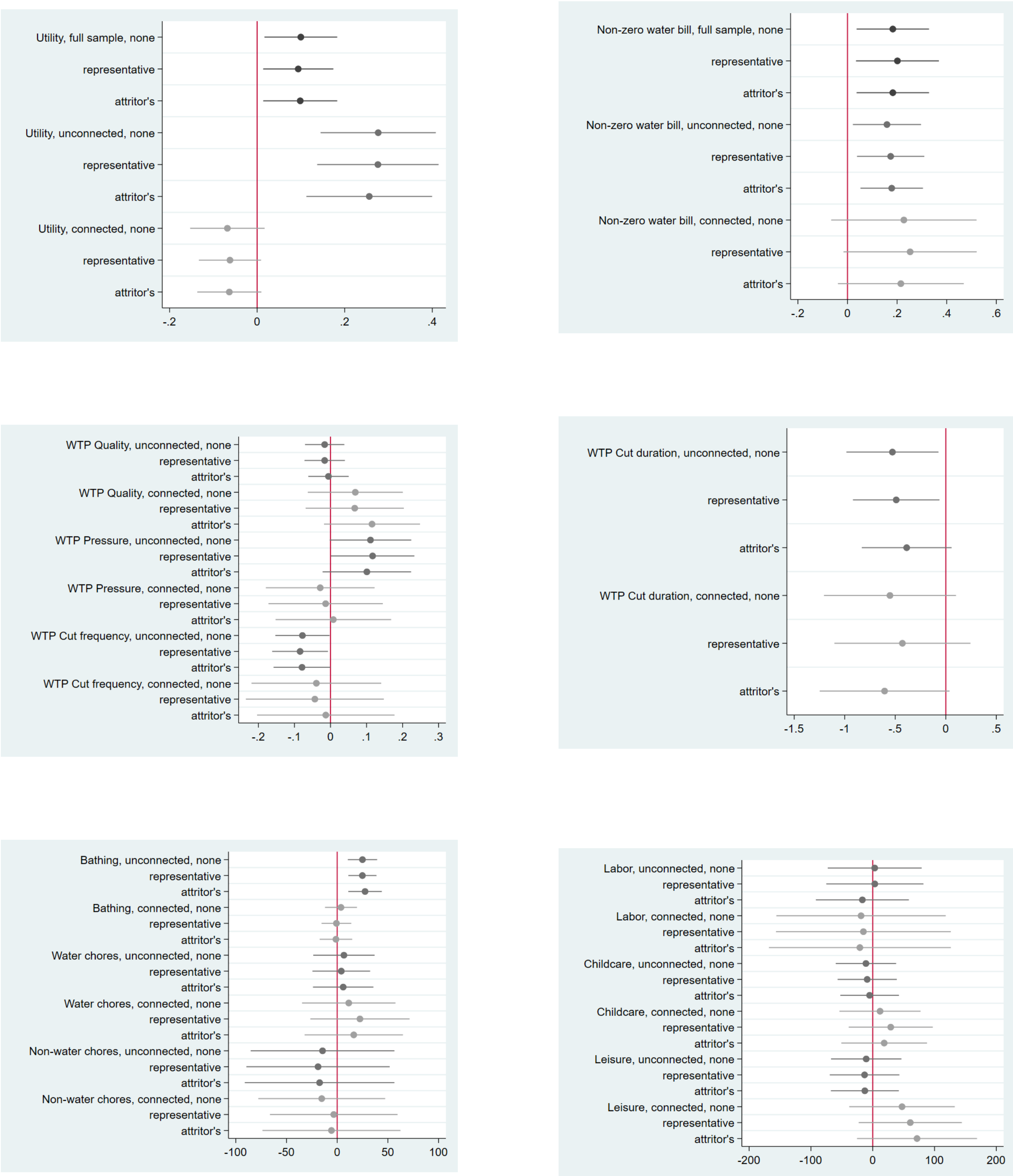
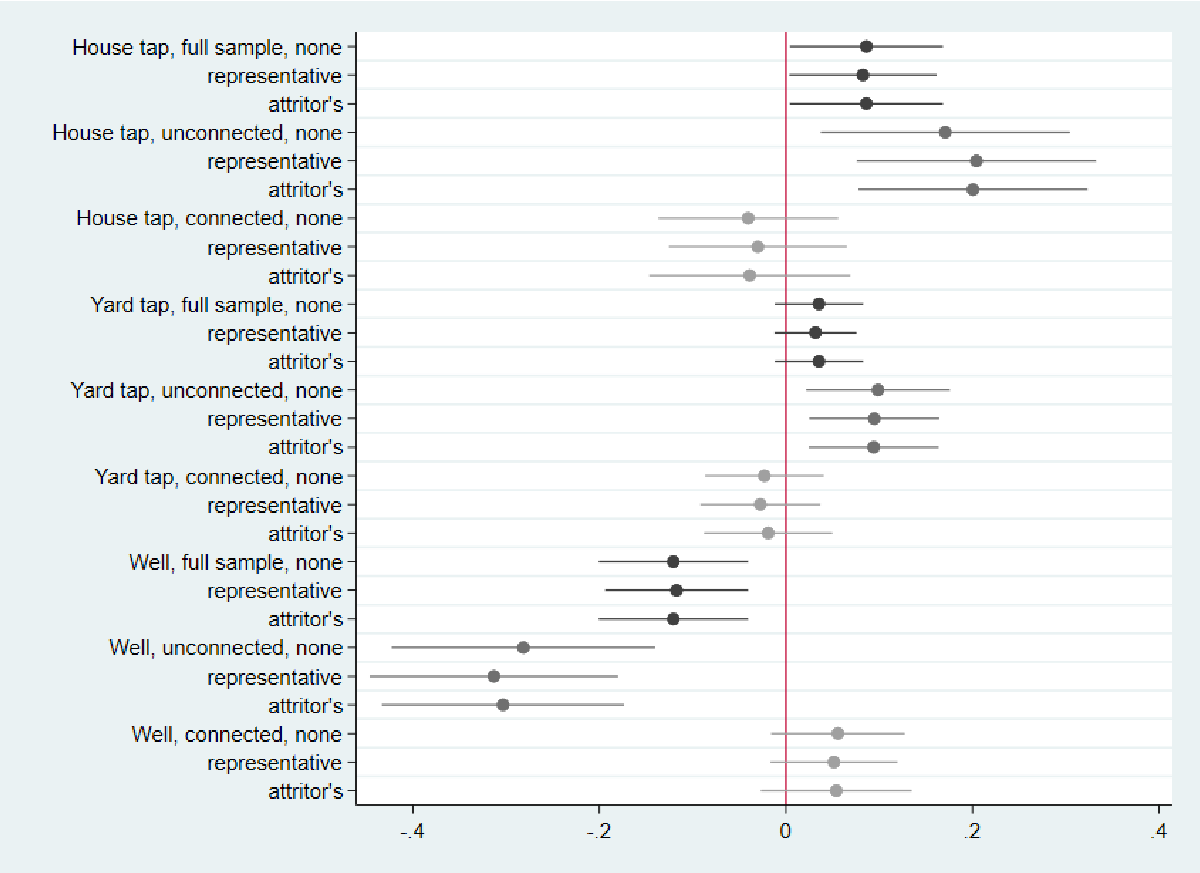


Figure A.2: Robustness checks with different weight regimes



Note: The charts report coefficient estimates for the impact of water network improvements from panel, doubly-robust difference-in-difference estimations with two time periods, household fixed effects and robust error terms clustered at the household level. Includes stratification weights. Dependent variables, baseline connection status and type of attrition weight are listed on the vertical axis. Estimations for households not connected to the network at $t = 0$ capture intention to treat. Estimations for households connected to the municipal water network at $t = 0$ condition on whether the household is connected via a house or yard tap. Spikes depict 95% confidence intervals.

Figure A.3: Robustness checks with different weight regimes, household infrastructure



Note: The charts report coefficient estimates for the impact of water network improvements from panel, doubly-robust difference-in-difference estimations with two time periods, household fixed effects and robust error terms clustered at the household level. Includes stratification weights. Dependent variables, baseline connection status and type of attrition weight are listed on the vertical axis. Estimations for households not connected to the network at $t = 0$ capture intention to treat. Estimations for households connected to the municipal water network at $t = 0$ condition on whether the household is connected via a house or yard tap. Spikes depict 95% confidence intervals.

C Robustness checks

The tables below present results from a more narrowly defined treatment group compared to our main results. We define the narrower treatment group as houses located in the one row parallel to a repaired pipe segment, along both sides of the pipe. This definition increases the probability of a contaminated control group. Nonetheless, the pattern of results for uptake of municipal water and household water infrastructure is the same. Coefficient and standard error estimates are also similar. For WTP and time use, results are less stable for example bathing does not increase as much and the coefficient for WTP Pressure is no longer significant. Meanwhile, WTP for reducing duration of cuts becomes positive and significant.

Table A.7: Uptake of utility company services and taps, narrow treatment band (DRDiD)

Outcome	**Full sample**	**Unconnected at** $t = 0$	**Connected at** $t = 0$
Any water network connection	0.10**	0.28***	-0.04
	(0.050)	(0.001)	(0.341)
House tap, Main water source	0.12**	0.25***	0.01
	(0.019)	(0.002)	(0.886)
Yard tap, Main water source	0.02	0.07*	-0.04
	(0.408)	(0.096)	(0.250)
Well, Main water source	-0.13***	-0.34***	0.04
	(0.008)	(0.000)	(0.310)
N	672	454	218

Note: Households on a street under which a repaired water main is situated are considered in the treatment area (i.e. w/in 1 house of a repaired pipe). All other households are untreated. Results are from panel, doubly-robust difference-in-difference estimations with two time periods, household fixed effects, sampling weights and attrition weights. Dependent variables are listed in Column 1. $t = 0$ refers to baseline connection status. Estimations for households connected to the municipal water network at $t = 0$ condition on whether the household is connected via a house or yard tap. p-values are in parentheses. For the reader's convenience, Significance is additionally indicated with * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.8: WTP and time use impacts, narrow treatment band (DRDiD)

	Unconnected at $t = 0$		**Connected at $t = 0$**	
Outcome	**Coefficent**	**p-value**	**Coefficent**	**p-value**
WTP for water attributes				
Pressure	0.09	(0.214)	-0.10	(0.228)
Quality	-0.02	(0.600)	0.05	(0.334)
Reduced cut duration	0.78***	(0.003)	0.27	(0.367)
Reduced cut frequency	-0.08	(0.114)	0.08	(0.331)
Time spent (minutes)				
Water chores	-4.72	(0.764)	-21.07	(0.352)
Bathing	16.10**	(0.023)	1.39	(0.869)
Paid work	11.19	(0.801)	-46.75	(0.464)
Child rearing	9.42	(0.734)	30.90	(0.355)
Leisure	-15.17	(0.626)	22.39	(0.603)
N	672	454	218	

Note: Households on a street under which a repaired water main is situated are considered in the treatment area (i.e. w/in 1 house of a repaired pipe). Results are from panel, doubly-robust difference-in-difference estimations with two time periods, household fixed effects, sampling weights and attrition weights. Dependent variables are listed in Column 1. Estimations for households connected to the municipal water network at $t = 0$ condition on whether the household is connected via a house or yard tap. p-values are in parentheses. For the reader's convenience, Significance is additionally indicated with * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

D Propensity score diagnostics

Table A.9 shows the model diagnostics and results from logistic regressions for various models we tested for the propensity score component of the DRDiD. The objective is to establish a set of variables that allows us to predict where any given household might be located, relative to the treatment boundary. We tried a large variety of variables and combinations, to improve fit (not all models that we tested appear here). In all models the pseudo R^2 is low, but the joint significance of variables is below 0.001. The best fitting model is also the one with the best propensity score performance (column (4)). Due to the richness of the data set, it is unlikely that there is a key set of unobservables we could obtain to improve the model fit. It is more likely that the location of the treatment boundary is only weakly correlated with household traits. We nonetheless use the DRDiD approach to alleviate concerns over common trends.

Table A.9: Models to construct propensity score, logit

		Alternative models		Preferred model
Diagnostics	(1)	(2)	(3)	(4)
Rubin's B	16.8	25.1	34.4	22.8
Rubin's R	1.51	1.98	1.81	0.65
Pseudo R^2	0.085	0.089	0.096	0.099
Prod(χ^2) > X^2	0.000	0.000	0.000	0.000
Variables				
Wastewater	2.343***	2.183***	2.091**	1.906**
	(0.000)	(0.001)	(0.001)	(0.018)
House (vs. flat)	0.696	0.822	1.396*	1.365*
	(0.260)	(0.197)	(0.076)	(0.092)
House size (no. rooms)	-0.142*	-0.141	-0.143	-0.157
	(0.099)	(0.106)	(0.129)	(0.105)
House tap, main water source		0.084	0.194	0.232
		(0.889)	(0.751)	(0.707)
Well, main water source		0.196	0.194	0.203
		(0.728)	(0.730)	(0.719)
Any water network connection		0.466	0.507	0.499
		(0.488)	(0.450)	(0.456)
Economic condition			-0.112	0.036
			(0.418)	(0.917)
Household size (no. people)			0.038	0.175
			(0.617)	(0.476)
Owns backyard ag. plot			-0.569	-0.549
			(0.249)	(0.268)
Econ. cond. x people				-0.046
				(0.588)
Receives remittances x Owns PC				0.262
				(0.407)
Wastewater connection x Owns PC				0.163
				(0.741)
Constant	-0.842	-1.227	-1.123	-1.529
	(0.159)	(0.120)	(0.182)	(0.220)
Observations	345	345	340	340

Notes: Data are the baseline survey observations for all panel households. Rubin's B and R are measures of how balanced the samples are on observables, after applying the propensity score model in the same column. The variable *Economic condition* take values 1-5, 1 being "Hardly make ends meet, don't have enough money to buy the most necessary products", 3 is "Enough for living, but it is difficult to buy some durables, such as furniture, refrigerator, TV" and 5 is "We can buy almost everything we want". *Owns PC* is a binary variable equal to 1 if the household owns at least 1 personal computer. p-values in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$