Liquidity constraints: Allocating clean water to the rural poor

Pre-analysis plan

Fiona Burlig, Amir Jina, and Anant Sudarshan*

December 5, 2022

Abstract

How should scarce resources be allocated in developing countries? Weitzman (1977) highlights a trade-off between prices, which generate allocative efficiency, and quotas, which might have desirable distributional consequences. In partnership with a private company supplying clean water to rural Odisha, India, we plan to run an experiment to measure the relative effectiveness of different allocation mechanisms. We will measure the price elasticity of demand for clean water, health effects from consuming clean water, and the extent to which liquidity constraints and intra-household inefficiencies reduce consumption. To do so, we implement a cluster-randomized trial, where 160 villages are randomized into a pure control group and multiple treatment arms: (i) discounts; (ii) a monthly quota; and (iii) an exchangeable quota, where unused allocation can be exchanged for cash. We subsequently randomize which households within each treatment village will receive treatment. We plan to measure effects of treatment on water consumption and health outcomes using a combination of survey and administrative data.

Key words: Water; resource allocation; quotas
JEL Codes: D45; O13; Q25; Q53

*Burlig: Harris School of Public Policy and Energy Policy Institute (EPIC), University of Chicago, and NBER. Email: burlig@uchicago.edu. Mailing address: Keller Center, 1307 E 60th St., Chicago, IL 60637. Jina: Harris School of Public Policy and EPIC, University of Chicago, and NBER. Email: amirjina@uchicago.edu. Mailing address: Keller Center, 1307 E 60th St., Chicago, IL 60637. Sudarshan: Department of Economics, University of Warwick. Email: anant.sudarshan@warwick.ac.uk. Mailing address: S2.92 Department of Economics, Social Science, University of Warwick, Coventry, CV4 7AL, UK. We thank Kishan Nanavati and Krupa Paltasingh of Spring Health Water India Pvt. Ltd. for their partnership on this project. We thank Kashif Ahmed, Zoya Khan, Ambalika Khanna, Nisha Koppa, Yashaswi Mohanty, and Chinmaya Sahoo for excellent research assistance. Nadia Lucas provided key research design development assistance. We gratefully acknowledge financial support from the Oak Foundation, the Templeton World Charity Foundation, and the Becker Friedman Institute's Development Economics Research Fund. All remaining errors are our own. This project received IRB approval from the University of Chicago (Protocol No. IRB22-0036), and is registered on the AEA RCT registry (Identification No. 10545). This PAP describes our experimental design, data collection, and outlines our planned analysis, including definitions of the key outcome variables and main regression specifications. We will update this PAP with additional details before our endline survey is complete. We leave open the possibility of additional analysis informed by what we learn through implementation.

Contents

1 Introduction **1**

2 Experimental design **1**
2.1 Sample Selection . . . 2
2.2 Timeline . . . 4

3 Data collection and outcome variables **4**
3.1 Record of scratch-card distribution . . . 4
3.2 Administrative data . . . 4
3.3 Survey . . . 5
3.3.1 Baseline . . . 6
3.3.2 High-frequency survey . . . 6
3.3.3 Endline . . . 6
3.4 Water quality measurements . . . 6

4 Analysis **7**
4.1 Main effect 1: Household-level treatment effects on water consumption . . . 7
4.1.1 Intent to treat effects . . . 7
4.2 Main effect 2: Price elasticity of demand . . . 8
4.2.1 Testing for liquidity constraints . . . 9
4.3 Main effect 3: Household-level treatment effects on health . . . 10
4.3.1 Short-run ITT effects . . . 10
4.3.2 Short-run LATEs . . . 10
4.3.3 Medium-run effects . . . 11
4.4 Secondary effects: Non-health benefits . . . 11
4.5 Secondary effect: Spillovers . . . 11
4.6 Heterogeneity . . . 12
4.6.1 Water quantity effects . . . 13
4.6.2 Demand elasticity . . . 13
4.6.3 Health effects . . . 13
4.7 Multiple hypothesis testing . . . 14

1 Introduction

In 2022, we began a cluster-randomized experiment in Odisha, India, in partnership with Spring Health India Pvt. Ltd — a private company that sells clean water to rural households in Odisha, India.

The clean water sold by Spring Health comes originally from local sources and is then treated using electrochlorination, in a treatment plant powered using decentralized solar electricity. For the majority of Spring Health water treatment plants, the input source of water is groundwater. In most cases there is one treatment facility per village, although a small share of villages share a treatment plant. Spring Health trains a local entrepreneur who operates and maintains the treatment facility and is normally also the owner of the well providing the input water.

The electro-chlorination process is intended to remove coliform and other organic contaminants, not contamination due to arsenic compounds, cynanide etc. That said we are aware of no evidence of arsenic contamination of groundwater in the areas of Odisha where SH functions. In this note, unless explicitly specified, we use 'treated water' or 'clean water' to indicate water that has been treated to remove coliform, or that has passed a coliform water test.

In the status-quo business model, any household in a village served by Spring Health can enroll for clean water deliveries. Payments are made against orders fulfilled and households may purchase varying amounts of water in the month and may discontinue purchases at any point. Spring Health water is packaged in 20 litre bottles each priced at INR 25 (about 30 cents). Bottles are delivered directly to households. This model differs from many prior studies that have studied the economics of clean water in developing countries (e.g. Kremer et al. (2011), Berry, Fischer, and Guiteras (2020); Dupas et al. (forthcoming)), because households do not have to spend time to obtain clean water and there is little risk of contamination between the source of clean water (the treatment plant) and reaching the household.

2 Experimental design

In partnership with Spring Health we conduct a field experiment designed to test different ways of offering water to households, inspired by Weitzman (1977). These include offering an unlimited water for sale at different prices, providing a water quota for free, or an entitlement model where households can choose not to consume upto the entitlement limit in exchange for cash rebates. Figure 1 displays the design graphically and we describe the experiment in more detail below.

2.1 Sample Selection

Working with Spring Health, we selected 160 villages as the site of the experiment. These villages were selected because they have a robust existing Spring Health presence. More precisely, all villages in the experiment had been served by Spring Health for at least 24 months prior to the beginning of the study, and Spring Health did not expect operational challenges in carrying out the experiment. These villages were then randomly assigned to four treatment arms to set up a cluster-randomized design as described below.

Pure control (40 villages) Villages in this arm are completely "business-as-usual." Throughout the experiment, all households in these villages were able to purchase Spring Health water at their prevailing market price. No surveys were conducted in these villages. Thus our only source of data on these villages comes from Spring Health's administrative data, which is generated through the normal course of business.

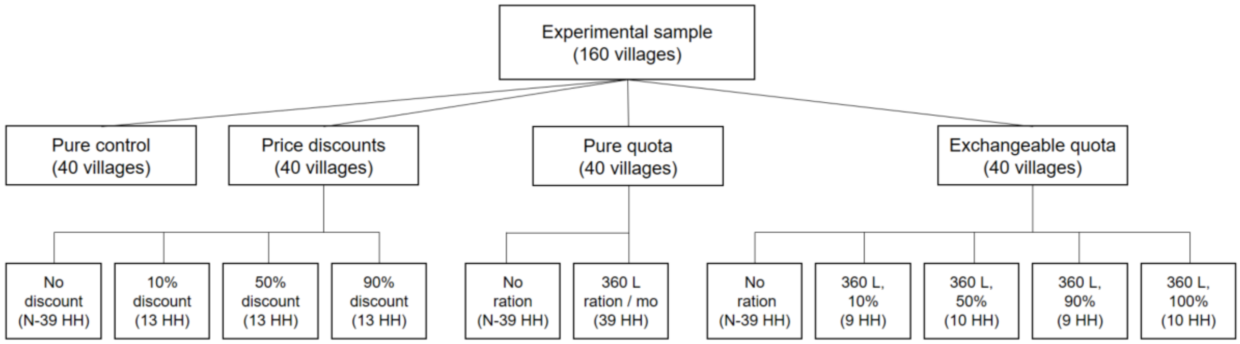
Free quotas (40 villages) Each household in villages assigned to the 'free quota' condition is randomly assigned to receive either an offer of 400 litres of water per month per month for the duration of the experiment or nothing (equivalent to a free quota of 0 litres per month)[1]. Households who do not receive a free quota continue to have the option of purchasing clean water from Spring Health at full price. Likewise households who receive a free quota could purchase additional water at full price if they exhaust their quota. Households could also opt not to use some or all of their quota.

Household level randomization was conducted by distributing scratch cards drawn from a shuffled deck, to every home in the village. The deck for every village included 39 treatment cards (a non-zero quota). However since the total number of distributed cards varied by the number of households, the probability of being treated is not constant. Figure 2 shows an example scratch card. Note that all scratch cards look identical so the field operators distributing cards do not know whether they contain an offer until the household uses them. Each card has a unique ID number but a mapping from this number to the offer on the card was only available to the research team. This mapping was used by the research team to check the number of offers in each deck prior to delivering them to field operators.

Price discounts (40 villages) In price discount villages, 13 households were randomly assigned to receive an offer of a 10% discount for the duration of the experiment, 13 were assigned to receive an offer of a 50% discount, and 13 were offered a 90% discount. The remaining households received no discount. Households who did not receive a discount could continue to purchase Spring Health water at the market price. These discounts were also implemented through the distribution of a shuffled deck of scratchcards.

1. This was benchmarked to be slightly above the mean consumption for households in our pilot of 360 litres per month

Figure 1: Experimental design diagram



Notes: This figure displays the details of our experimental design: we randomly assign 160 villages to a pure control group, price discount group, pure quota group, or exchangeable quota group. Within the three treatment groups, we randomly assign households to sub-treatments. In price discount villages, every household is offered either no discount, a 10% discount, a 50% discount, or a 90% discount. In pure quota villages, every household is offered either no quota or a quota of 400 litres per month. In exchangeable quota villages, every household is offered no quota or a 400 litre per month quota with a cash-back value of 10%, 50%, 90%, or 100% of the market price.

Exchangeable quota (40 villages) In exchangeable quota villages, 38 households in each village were given the offer of a 400 litre quota just as in the free quota condition. The difference in this treatment arm was that households could redeem unused water below this quota for cash. Specifically, for every unclaimed bottle of water households were entitled to receive a payment equal to 10% of the market price (9/38 households), 50% of the market price (10/38), 90% of the market price (9/38 households), or 100% of the market price (10/38 households). As in the other arms, households who did not receive a non-zero exchangeable quota remained eligible to buy Spring Health water at full price.

Free one-time 100L (all treatment villages) In addition to the treatments described above, 5 households in each treatment village were randomized to receive a one-time offer of 100 free litres of water.Households who chose not to take this offer received nothing in return. These households could otherwise purchase water at the market price. We include this arm to understand the extent to which Spring Health water is an experience good, by comparing the number of bottles consumed by these households to the control households — any difference greater than five will suggest that exposure to Spring Health water drives future purchases — but plan to otherwise exclude these households from our main analyses.

Figure 2: Sample scratch card

Notes: This figure shows an example promotional scratch card used to randomize households into treatment groups. The left shows the front of the card, which is common among all offers. The right shows the back of the card, which differs across offers.

2.2 Timeline

For logistical reasons, we will run the experiment in phases. We randomly divide villages across phases, preserving equal treatment and control group sizes in each phase. Figure 3 shows the implementation timeline, including the scratch card distribution, water distribution, and survey data collection (described in more detail in Section 3 below).

3 Data collection and outcome variables

We will bring together several sources of data to conduct our analysis. The PI team has not had access to these data when this PAP was uploaded to the AEA registry.

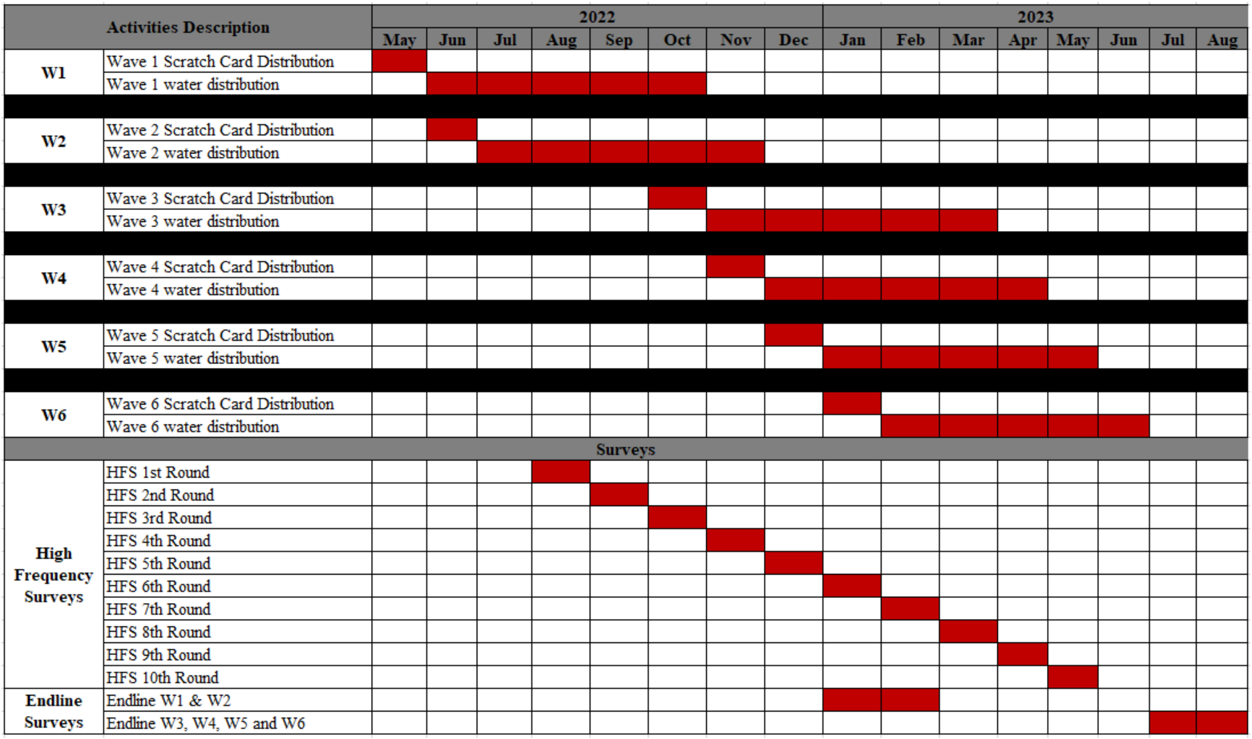
3.1 Record of scratch-card distribution

While distributing scratch cards to each household, we record the village name, the name of the household head, whether the household is an existing Spring Health customer, contact and address details, the offer type, and the scratch card identification number, which allows us to link a household to its treatment offer. We use this data to estimate intent-to-treat effects, as described in Section 4 below. We can also use this dataset to conduct a set of randomization checks.

3.2 Administrative data

One of our main outcomes of interest is consumption of treated water. To measure this, we collect administrative water sales records from Spring Health. For every household that purchases water from Spring Health – including both households who received a treatment offer and those who did not – we collect daily information on the number of bottles of water

Figure 3: Experimental timeline



Notes: This figure shows the experimental timeline in the form of a Gantt chart.

purchased from Spring Health, and at what price. For exchangeable quota households, we use these data to calculate how much money they are owed at the end of each reporting period.

Because Spring Health's administrative data are complete, and our treatment offer dataset enumerates every household in each village, we can also infer that households who do not appear in the Spring Health administrative dataset must consume 0 litres. This yields a household-by-date panel of Spring Health water purchases.

3.3 Survey

In addition to collecting administrative data from Spring Health, we are administering a series of surveys to measure household responses to our interventions. We do not conduct surveys in pure control villages. In each treatment village, we sample 15 randomly-selected households, stratified by treatment arm including the control group.[2]

2. In price discount villages, we survey 4 control households and 3 households in each discount level; in pure quota villages we survey 6 control households and 7 quota households; and in exchangeable quota villages, we survey 3 control households, 2 10% exchange households, 3 50% exchange households, 2 90% exchange households, and 3 100% exchange households.

Each surveyed household will be visited five times: once for the baseline survey (one month after scratch card distribution), three times for "high-frequency" check-ins (two,three, and four months after scratch card distribution), and once for the endline survey (five months after scratch card distribution). During the second or third high-frequency check-in, we also conduct water testing in survey households. Figure 3 presents a rough experimental timeline. We describe each survey in more detail below. Through each round of the survey, households who dropped out of the survey were replaced by selecting a similar backup at random from the corresponding treatment arm.

3.3.1 Baseline

We conduct a brief baseline survey, consisting of basic household demographic information, including total number of household members, number of children, and number of infants; main source of water for cooking and handwashing, laundry and cleaning, gardening and farming, and drinking; drinking water treatment behavior; water purchase behavior; information on earnings in the past month; information on missed school in the past week; and data on household illness, including symptoms, clinic visits, and expenses.

3.3.2 High-frequency survey

In the high-frequency surveys, we ask a subset of the questions in the baseline questionnaire: we ask for information on earnings in the past month; information on missed school in the past week; and data on household illness, including symptoms, clinic visits, and expenses. This allows us to measure any treatment effects on work and school attendance and health, which may be relatively short-lived and therefore difficult to capture in the endline survey.

3.3.3 Endline

We will update this PAP with detailed information on the endline survey prior to putting it into the field. We will also update Section 4 below to include analysis that relies on the endline survey.

3.4 Water quality measurements

During the final high-frequency survey, we will also visit survey households to take measurements of water quality. During this period, we will also take village-wide water quality measurements, including of Spring Health's original water sources and their treated water. We will update this PAP with detailed information about water quality testing prior to fielding this part of the design.

4 Analysis

4.1 Main effect 1: Household-level treatment effects on water consumption

4.1.1 Intent to treat effects

We begin by estimating the intent-to-treat effects of each of our treatments on clean water consumption.

Spring Health water: Extensive margin We first estimate a cross-sectional extensive-margin regression to understand whether our treatments changed households' use of Spring Health water:

$$\begin{aligned}\mathbf{1}[Q_i > 0]_i = {} & \eta_1 \cdot 10\% \text{ discount}_i + \eta_2 \cdot 50\% \text{ discount}_i + \eta_3 \cdot 90\% \text{ discount}_i \\ & + \eta_4 \cdot \text{Pure quota}_i + \eta_5 \cdot 10\ \% \text{ exchange}_i + \eta_6 \cdot 50\ \% \text{ exchange}_i + \eta_7 \cdot 90\ \% \text{ exchange}_i \\ & + \eta_8 \cdot 100\ \% \text{ exchange}_i + \gamma_v + \varepsilon_i \end{aligned} \tag{1}$$

where $\mathbf{1}[Q_i > 0]_i$ is an indicator equal to 1 if and only if household i purchased any Spring Health water during the course of the experiment, the treatment indicators are each listed, γ_v is a village fixed effect, and ε_i is an error term. As a baseline specification, we compare treated consumers to only untreated customers within treated villages. As robustness, we also run a version of this specification which includes pure control customers. For this, as well as all other household-level regressions in the Analysis section, we will estimate a version that includes a vector of pre-determined controls, $\mathbf{X}_i$, selected using LASSO.

We also estimate a pooled version, where we combine treatments of different levels:

$$\mathbf{1}[Q_i > 0]_i = \eta_1 \cdot \text{Any discount}_i + \eta_2 \cdot \text{Pure quota}_i + \eta_3 \cdot \text{Any exchange}_i + \gamma_v + \varepsilon_i \tag{2}$$

Spring Health water: Quantity We next estimate the effect of our treatments on the quantity of Spring Health water purchased by households. We first use a cross-sectional specification:

$$\begin{aligned}Q_i = {} & \eta_1 \cdot 10\% \text{ discount}_i + \eta_2 \cdot 50\% \text{ discount}_i + \eta_3 \cdot 90\% \text{ discount}_i \\ & + \eta_4 \cdot \text{Pure quota}_i + \eta_5 \cdot 10\ \% \text{ exchange}_i + \eta_6 \cdot 50\ \% \text{ exchange}_i + \eta_7 \cdot 90\ \% \text{ exchange}_i \\ & + \eta_8 \cdot 100\ \% \text{ exchange}_i + \gamma_v + \varepsilon_i \end{aligned} \tag{3}$$

where Q_i is the total amount of Spring Health water purchased by household i during the experimental time period, and all other variables are the same as in Equation (1) above. We also estimate a pooled version identical to Equation (2) with Q_i on the left-hand side.

In addition, we estimate a panel model that includes time fixed effects to soak up additional identifying variation and serve as a basis for our event study specification.

$$\begin{aligned} Q_{it} = {} & \beta_1 \cdot 10\% \text{ discount}_i + \beta_2 \cdot 50\% \text{ discount}_i + \beta_3 \cdot 90\% \text{ discount}_i \\ & + \beta_4 \cdot \text{Pure quota}_i + \beta_5 \cdot 10 \text{ \% exchange}_i + \beta_6 \cdot 50 \text{ \% exchange}_i + \beta_7 \cdot 90 \text{ \% exchange}_i \\ & + \beta_8 \cdot 100 \text{ \% exchange}_i + \gamma_v + \delta_t + \varepsilon_{it} \end{aligned} \tag{4}$$

where Q_{it} is the quantity of Spring Health water purchased by household i in village v on date t, we include an indicator variable for households who were *offered* each treatment, γ_v is a village fixed effect, δ_t is a month fixed effect, and ε_{it} is an error term, clustered at the individual level. We will estimate Equation (4) both with and without the pure control villages. We will also estimate a version where we pool all treatments of the same type.

Dynamic effects Next, we estimate the extent to which our treatment effects change over time. To do this, we estimate an event study version of our panel model. In partcular, we regress:

$$\begin{aligned} Q_{it} = {} & \sum_{m=0}^{6} \beta_1^m \cdot [\text{months from 10\% discount start} = m]_i + \beta_2^m \cdot [\text{months from 50\% discount start} = m]_i \\ & + \beta_3^m \cdot [\text{months from 90\% discount start} = m]_i + \beta_4^m \cdot [\text{months from pure quota start} = m]_i \\ & + \beta_5^m \cdot [\text{months from 10\% exchange start} = m]_i + \beta_6^m \cdot [\text{months from 50\% exchange start} = m]_i \\ & + \beta_7^m \cdot [\text{months from 90\% exchange start} = m]_i + \beta_8^m \cdot [\text{months from 100\% exchange start} = m]_i \\ & + \gamma_v + \delta_t + \varepsilon_{it} \end{aligned} \tag{5}$$

where we now estimate m treatment effects of each type, one for each month after treatment begins. We will also estimate a version where we pool all treatments of the same type.

4.2 Main effect 2: Price elasticity of demand

Price discount villages A key parameter is the price elasticity of demand for clean water. To estimate this, we first run the following specification, using price discount villages only:

$$\log(Q)_i = \beta \cdot \log(P)_i + \gamma_v + \varepsilon_i \tag{6}$$

where P_i is the price of Spring Health water, and identifying variation comes from random assignment of households to discount levels. In addition, we will construct a figure that plots average consumption at each price level.

Exchangeable quota villages Households in our exchangeable quota villages face the same marginal incentive as households in the price discount villages: the opportunity cost of water for a household with a 10% exchangeable quota is 10% of the market price. As a result, we can use these villages to generate a second estimate of the price elasticity of demand. To estimate this, we estimate Equation (6) instead using exchangeable quota villages only, where we define P_i to be the opportunity cost of using a quota bottle. In this case, the identifying variation comes from random assignment of households to exchangeable quota levels.

4.2.1 Testing for liquidity constraints

Next, we compare households facing the same marginal incentive in the price discount villages and exchangeable quota villages, to test for the presence of liquidity constraints. While both sets of households face the same marginal incentive, exchangeable quota households begin with a water allocation, so any differences between these two demand curves should reflect liquidity constraints. To test for this, we estimate, using (i) the price discount villages, the exchangeable quota villages, and the control villages; and (ii) the price discount villages and exchangeable quota villages only:

$$\log(Q)_i = \beta_1 \cdot \log(P)_i + \beta_2 \cdot \log(P)_i \times \mathbf{1}[\text{Exchangeable quota}]_v + \gamma_v + \varepsilon_i \tag{7}$$

where the main coefficient of interest is β_2, which estimates the differential effect of being in an exchangeable quota village (denoted by the indicator function) on the price elasticity of demand.

We also estimate the difference in quantities at each price point for discount vs. exchangeable quota households with the following specification:

$$\begin{aligned} Q_i = {} & \beta_1 \cdot [10\% \text{ discount price}]_i + \beta_2 \cdot [10\% \text{ discount price}]_i \times \mathbf{1}[\text{Exchangeable quota}]_v \\ & + \beta_3 \cdot [50\% \text{ discount price}]_i + \beta_4 \cdot [50\% \text{ discount price}]_i \times \mathbf{1}[\text{Exchangeable quota}]_v \\ & + \beta_5 \cdot [90\% \text{ discount price}]_i + \beta_6 \cdot [90\% \text{ discount price}]_i \times \mathbf{1}[\text{Exchangeable quota}]_v \\ & + \gamma_v + \varepsilon_{it} \end{aligned} \tag{8}$$

where the difference between β_1 and β_2, β_3 and β_4, and β_5 and β_6 is the difference in demand at each point between the discount arm and the exchangeable quota arm.

4.3 Main effect 3: Household-level treatment effects on health

4.3.1 Short-run ITT effects

Next, we measure intent-to-treat effects of each of our treatments on health. To do this, we estimate the following specification:

$$\begin{aligned}\bar{H}_i = {} & \beta_1 \cdot 10\% \text{ discount}_i + \beta_2 \cdot 50\% \text{ discount}_i + \beta_3 \cdot 90\% \text{ discount}_i \\ & + \beta_4 \cdot \text{Pure quota}_i + \beta_5 \cdot 10 \% \text{ exchange}_i + \beta_6 \cdot 50 \% \text{ exchange}_i + \beta_7 \cdot 90 \% \text{ exchange}_i \\ & + \beta_8 \cdot 100 \% \text{ exchange}_i + \gamma_v + \varepsilon_i \end{aligned} \tag{9}$$

where $\bar{H}_i$ is the average answer to whether any household member was sick in the past week by household i across all high-frequency survey waves. We will additionally run this specification using the other health questions in our survey, including whether any household members had diarrhea, vomiting, or abdominal pain. We will also estimate this using childrens' outcomes on the left-hand side. Because we do not conduct high-frequency surveys in pure control villages, for these treatment effects, we exclude the pure control villages from the sample. We will also estimate a version where we pool all treatments of the same type.

4.3.2 Short-run LATEs

Our key objective with our health measures is to quantify the impacts of clean water consumption on health. In order to do this, we move from reduced-form estimates of intent-to-treat effects to 2SLS estimates of LATEs, instrumenting for water consumption with indicators of treatment assignment. We use the following system of equations:

First stage:

$$\begin{aligned} Q_i = {} & \beta_1 \cdot 10\% \text{ discount}_i + \beta_2 \cdot 50\% \text{ discount}_i + \beta_3 \cdot 90\% \text{ discount}_i \\ & + \beta_4 \cdot \text{Pure quota}_i + \beta_5 \cdot 10 \% \text{ exchange}_i + \beta_6 \cdot 50 \% \text{ exchange}_i + \beta_7 \cdot 90 \% \text{ exchange}_i \\ & + \beta_8 \cdot 100 \% \text{ exchange}_i + \gamma_v + \varepsilon_i \end{aligned} \tag{10}$$

Second stage:

$$\bar{H}_i = \pi \cdot \widehat{Q_i} + \gamma_v + \varepsilon_i \tag{11}$$

where π is the coefficient of interest, mapping out the relationship between consumption of clean water and health. We will also estimate a version where we pool all treatments of the same type.

4.3.3 Medium-run effects

We also measure medium-run ITT and LATE effects of our treatments on health using data from the endline survey. We will update this PAP to include these specifications prior to fielding the endline survey.

4.4 Secondary effects: Non-health benefits

In addition to measuring treatment effects on health, we expect that providing clean water at lower cost may lead households to (A) spend less time collecting water, (B) spend less money treating water, (C) running out of water less often, and (D) switching primary water sources. We test for these by estimating the following specification:

$$\begin{aligned} \bar{NH}_i = {} & \beta_1 \cdot 10\% \text{ discount}_i + \beta_2 \cdot 50\% \text{ discount}_i + \beta_3 \cdot 90\% \text{ discount}_i \\ & + \beta_4 \cdot \text{Pure quota}_i + \beta_5 \cdot 10\ \% \text{ exchange}_i + \beta_6 \cdot 50\ \% \text{ exchange}_i + \beta_7 \cdot 90\ \% \text{ exchange}_i \\ & + \beta_8 \cdot 100\ \% \text{ exchange}_i + \gamma_v + \varepsilon_i \end{aligned} \tag{12}$$

where $\bar{NH}_i$ is the average answer by household i to our high-frequency survey questions on the amount of time spent collecting water, amount of money spent purchasing water, amount of money spent treating water, chance of running out of water, and the chance that the household listed a particular source as its main source of water. Because we do not conduct high-frequency surveys in pure control villages, for these treatment effects, we exclude the pure control villages from the sample. We will also estimate a version where we pool all treatments of the same type.

4.5 Secondary effect: Spillovers

We test for the possibility that our treatments impact the demand for Spring Health water among control households by comparing control households in treatment villages against those in pure control villages.

Spillovers on Spring Health customer status First, we test whether treatments induced control households to sign up for Spring Health deliveries. Using a sample of only control and pure control households, we estimate:

$$\mathbf{1}[Q > 0]_v = \beta_1 \cdot \text{price discount}_v + \beta_2 \cdot \text{pure quota}_v + \beta_3 \cdot \text{exchangeable quota}_v + \varepsilon_v \tag{13}$$

where $\mathbf{1}[Q > 0]_v$ is the count of households purchasing at least 1 bottle of Spring Health water in each village, v. We include the entire sample of villages in this regression. For pure control villages, we include all households in the count; for treated villages, we include only control households in the count. In treated villages, to account for the fact that some households are treated, we scale the count variable by $\frac{1}{1-\text{treated household share}_v}$.

Spillovers on water purchases To test for spillover effects of treatment on water purchases, we estimate the following equation, on a sample that excludes all households who received a treatment offer:

$$Q_v = \beta_1 \cdot \text{price discount}_v + \beta_2 \cdot \text{pure quota}_v + \beta_3 \cdot \text{exchangeable quota}_v + \varepsilon_v \tag{14}$$

where Q_v is the sum of Spring Health water purchased over the course of the experiment in each village v. As above, we include the entire sample of villages in this regression, and scale the quantity for treated villages to account for the existence of treated households.

We do not estimate spillover effects on health, because we restricted our household surveys to treatment villages for budgetary and power reasons. However, if there are any spillover effects between treatment and control households, and if there is a positive effect of clean water on health, any health effects we estimate in Equations (9) and (11) can be viewed as lower bounds.

4.6 Heterogeneity

In addition to our average treatment effects (ITTs and LATEs), we are interested in heterogeneity along several dimensions. Here, we highlight specific dimensions of heterogeneity that we can explore in the absence of the endline survey, and describe our approach for estimating these effects. We will update this section to describe further plans for heterogeneous treatment effect analysis using endline data prior to fielding the endline, where we plan to use similar methods.[3] We expect different dimensions of heterogeneity to matter for different outcome variables, so we segment our discussion by outcome. In this section, we discuss our main heterogeneous effects specifications, though we may estimate heterogeneity along additional dimensions in the final paper.

3. For implementation timing reasons, we plan to collect a variety of characteristics that are unlikely to change in the endline survey rather than in the baseline survey. We plan to test for heterogeneity using these exogenous characteristics. We will not test for heterogeneity along dimensions we expect to vary with treatment.

4.6.1 Water quantity effects

We will estimate heterogeneous treatment effects on clean water quantities using specifications of the following form:

$$Q_i = \sum_{k \in \text{treatments}} \pi_k \cdot \text{Treatment}_{ik} \cdot \text{Covariate}_i + \eta_k \cdot \text{Treatment}_{ik} + \beta \cdot \text{Covariate}_i + \gamma_v + \varepsilon_i \tag{15}$$

where the covariates of interest are: household size (and separately number of children); household income; and the quality of existing drinking water options. We will estimate versions of this specification where we pool and separate individual sub-treatments.

4.6.2 Demand elasticity

We will estimate heterogeneous elasticities of demand using the following specification:

$$\log(Q)_i = \pi \cdot \log(P)_i \cdot \text{Covariate}_i + \beta \cdot \log(P)_i + \eta \cdot \text{Covariate}_i + \gamma_v + \varepsilon_i \tag{16}$$

where the covariates of interest are: household size (and separately number of children); household income; and the quality of existing drinking water options. We will estimate this separately for price discount villages only and for exchangeable quota villages only.

We will also test for heterogeneous effects on liquidity constraints by comparing the price discount and exchangeable quota villages. Here, the main covariate of interest from the baseline survey is income. We plan to supplement this with other covariates we collect in the endline.

4.6.3 Health effects

We also plan to estimate heterogeneity in health effects, using the following ITT specification:

$$\bar{H}_i = \sum_{k \in \text{treatments}} \pi_k \cdot \text{Treatment}_{ik} \cdot \text{Covariate}_i + \eta_k \cdot \text{Treatment}_{ik} + \beta \cdot \text{Covariate}_i + \gamma_v + \varepsilon_i \tag{17}$$

For health, we plan to examine heterogeneity by the number of children in the household, household income, quality of household drinking water measured at baseline, and whether the household treats their drinking water at baseline.

We will also test for heterogeneous health LATEs, where the first stage is Equation (15), and the second stage is:

$$\bar{H}_i = \pi \cdot \widehat{Q_i} \cdot \text{Covariate}_i + \eta \cdot \widehat{Q_i} + \beta \cdot \text{Covariate}_i + \gamma_v + \varepsilon_i$$

4.7 Multiple hypothesis testing

This study has a limited number of key outcome variables. We are primarily concerned with (i) consumption of clean water; and (ii) health. (i) is measured with one outcome variable. For (ii), we will (a) focus on a primary outcome, whether any household member was sick in the past week, and (b) generate an index over the following health variables from our high-frequency survey: whether any household member was sick in the past week, whether any household members had diarrhea, vomiting, or abdominal pain, and outcomes for children. We will also report False Discovery Rate $q-$values following Anderson (2008) for the health outcomes.

References

Anderson, Michael. 2008. "Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects." *Journal of the American Statistical Association* 103 (484): 1481–1495.

Berry, James, Greg Fischer, and Raymond Guiteras. 2020. "Eliciting and utilizing willingness to pay: Evidence from field trials in northern Ghana." *Journal of Political Economy* 128 (4): 1436–1473.

Dupas, Pascaline, Basimenye Nhlema, Zachary Wagner, Aaron Wolf, and Emily Wroe. forthcoming. "Expanding access to clean water for the rural poor: Experimental evidence from Malawi." *American Economic Journal: Economic Policy.*

Kremer, Michael, Jessica Leino, Edward Miguel, and Alix Peterson Zwane. 2011. "Spring cleaning: Rural water impacts, valuation, and property rights institutions." *Quarterly Journal of Economics* 126 (1): 145–205.

Weitzman, Martin. 1977. "Is the price system or rationing more effective in getting a commodity to those who need it most?" *Bell Journal of Economics* 8 (2): 517–524.