# Using a Discounted Hour at Clinics to Extend Healthcare Access to the Underserved

## *Pre-Analysis Plan*

Isabelle Cohen, Rebecca Dizon-Ross, Mika Inoue,
Seema Jayachandran, and Fiker Negash

First version: February 14, 2026
This version: March 13, 2026

# 1 Introduction

This document pre-specifies analyses for a randomized trial in rural northern Uganda that tests whether second-degree price discrimination — specifically in the form of an hour each morning when care is offered at a discounted price — can enable healthcare clinics to increase their reach to underserved individuals.

We collaborate with OneDay Health (ODH), a nonprofit that operates health centers in remote communities located far from a government health clinic. In the absence of an ODH center, care at a government health care is *de facto* expensive or out of reach because of the travel, and while "drug shops" sell medicine such as antimalarials, the advice given is often inappropriate. ODH aims to fill this void by setting up clinics that provide basic care. ODH hires a qualified nurse to run the center; rents a room in a village to house the health center, which serves households from several nearby villages; uses solar power for lighting; and offers care for a range of conditions such as malaria, upper respiratory infections, pneumonia, sexually transmitted infections, and wound care.

ODH uses donor money to set up a new health center (e.g., site selection, nurse recruitment, furniture) but aims to cover most of the subsequent operating costs via patient fees. The fees for a visit average 8800 UGX ($2.51 using 3500 UGX = $1). Many people in the community can and do pay these fees. However, the fees are likely prohibitive for poorer households and for individuals in less poor households for whom the household has a low willingness to pay for healthcare. We use the term underserved (US) to encompass both types of people who are priced out of care.

This project will test whether second-degree price discrimination can enable ODH to expand the care it provides to the US without cannibalizing too much of its revenue from full-fee-paying patients. Through a randomized trial, we introduce a new, lower-priced service tier that consists of a 90-minute window every morning when a visit costs a flat fee of 1000 UGX (about $0.30). Second-degree price discrimination relies on the lower-priced service tier being lower value, and one way this is achieved is by making the discount valid only during a specific, short window of time. In addition, the discounted visits are done as group visits when feasible and appropriate, and full-price paying customers have priority in the queue during the window hour, except in cases of medical emergencies. The goal is for the new service tier to be lower value than the more expensive standard

offering, but without comprising the quality of the medical care itself.

# 2 Sample

## Health centers and villages

The study sample consists of 239 villages in the catchment areas of 44 ODH centers in northern Uganda. The sample was determined using the following criteria. Within each center, we include villages (a) that are not the village where the center is located[1], and (b) from which at least 0.7 patients per week visit the center on average during the pre-period. To avoid overcrowding at centers,[2] we (c) cap the number of study villages per center at 10, dropping the largest village when the criterion binds (1 village); (d) exclude villages whose average visits per week in the pre-period exceeds 18 (2 villages); (e) and exclude the largest village in four centers where our study villages, in aggregate, contribute a particularly high weekly volume of visits (4 villages). In addition, because our field plan allocates at most two days to conducting surveys in a village, we (f) exclude villages with more than 750 households (based on preliminary information from village leaders), as surveying all interested households might not be possible (4 villages).

We include ODH centers (g) that are located in northern Uganda and (h) have at least 3 villages in their catchment area that meet our other criteria. From this set, we exclude one center close to ODH's headquarters to use for piloting.

## Households

The unit of randomization is the village. To make the within-center randomization design simpler operationally for ODH nurses, eligibility for discounts will be based on showing a physical voucher with a unique ID at the center (rather than being based on self-reported village of residence). We will hand out vouchers, color-coded to indicate the treatment arm, to households in all study villages.

To enroll households in a cost-effective way, we will hold an enrollment and voucher

[1]ODH patient records track the patient's village. ODH asked to exclude the village where the center is located because adhering to the randomized assignment would be harder for the nurse in the village where she resides.

[2]If the intervention were scaled up, ODH would hire additional nurses.

distribution event in each village. Prior to the event, we will use trained village heath teams (VHTs) to conduct treatment-blinded advertising. The VHTs will tell households in the village that there will be information and the possibility of a discount at ODH for those who attend the event in the village center the next day. The event will be divided across one or two "enrollment days" when baseline surveys take place and one "lottery explanation day."

During enrollment day(s), we will conduct short (five-minute) baseline surveys and hand out vouchers with a unique ID number that are valid as long as they have been stamped by the surveyor during the enrollment day. VHTs and LC1s (local elected officials) will assist the field team to ensure that only those from the specified village can pick up vouchers. In villages with 250 households or fewer, there will be a single enrollment day; in villages with more than 250 households, we will conduct two enrollment days to increase the share of households included. During the enrollment day(s), we will also share information on typical prices at ODH.

The lottery explanation day takes place after all enrollment days are completed. Voucher recipients return to the village center, and the specifics of the discount are announced and explained. We distribute a simple pictograph to act as a mnemonic. This approach of separating surveys from treatment announcement ensures that the full sample of households for a village is enrolled and surveyed before treatment status is revealed. At the same time, those who only attend on the first day are still in possession of a valid voucher.

We conduct this activity in the control group as well as the treatment groups, giving the control group vouchers that entitle them to a small discount at ODH. We use this procedure, rather than a pure control group, for several reasons. First, this enables comparable baseline data collection in the treatment and control groups. Collecting the baseline data in a central location is low-cost, and the small discount provides the control group compensation for their time. Second, voucher distribution likely has a marketing or salience effect, which we want to hold constant across arms. Third, offering a small discount to the control group gives them an incentive to bring their voucher to the ODH center when they come for care, which enables us to link visits to the baseline data.

We aim to enroll every household in the village, with one member of each household attending the enrollment day. Households that agree to participate in the study and

complete the baseline survey will be enrolled in the household-survey sample. A complementary sample are patients at the ODH center, where we will use de-identified patient records provided to us by ODH.

Enrollment and the start of treatment will be staggered over a roughly 10-week period.

# 3 Treatment arms and randomization

Villages will be assigned to the control group, the discount-window treatment group, or a low-price treatment group:

- **Control group (C):** Eligible for 200 UGX ($0.06) discount off the normal fees for any visit during operating hours.
- **Discount window treatment (PD, for price discrimination):** Eligible for a 1000 UGX flat fee from 8 am to 9 am *de jure* and 8 am to 9:30 am *de facto*; if shown at other times when the center is open, the voucher entitles the user to a 200 UGX discount off the regular price.[3]
- **Low-price group (LP):** Eligible for 1000 UGX flat fee any time during operating hours. The purposes of LP are to test mechanisms and to provide a benchmark for assessing PD's performance at increasing access relative to revenue loss, as explained in the Hypotheses section.

Roughly 1/3 of the villages will be assigned to Control, 1/2 of villages to the PD treatment, and 1/6 to the LP treatment. Randomization is stratified by center. There are 3 to 10 villages per center, and we can only exactly match our treatment probabilities in the centers with 6 villages, so standard randomization results in many "misfits." Instead, we use different randomization probabilities by stratum size; in simulations, doing so improves statistical power. In addition, the randomization procedure we use reduces the maximum number of LP and PD villages per center, which guards against over-crowding.[4]

[3]The 200 UGX discount the rest of the day, first, provides an incentive to present the voucher, which enables data merging, and, second, ensures that PD draws only from the price points used in C and LP, allowing us to test whether PD outcomes simply reflect a convex combination of those two price regimes, or whether sorting via PD expands the frontier of achievable outcomes.

[4]Specifically, we multiply the target treatment probabilities of 1/3, 1/2, and 1/6 by the stratum size and then round to a multiple of 0.5. If the rounded value is an integer, that gives the number of villages for the arm for that stratum size. If not, the integers above and below the rounded value map to two

The intervention will begin in the first villages on February 15, 2026. Given the tentative schedule of the staggered roll-out, we expect the intervention to end in in the last villages in August 2026.

# 4 Data

## Patient register data

The main outcome data source are ODH patient records. ODH nurses record details of patient visits in printed registers. The research team worked with ODH to modify the registers to make them machine readable and to include a few variables relevant for the study such as the voucher ID number and proxies for household poverty, namely household (HH) head's education, whether the HH owns a phone, and whether the HH has a toilet. As an additional key measure of being underserved (which might vary within household), the registers include information on what care the patient got the last time they were sick (no care, ODH, government health clinic, drug shop, other), as well as whether this is the patient's first visit to ODH.[5]

The patient register data suffice to test most of the hypotheses, such as whether there are more patient visits from treatment villages relative to control villages and whether the increase is relatively larger for the US.

## Baseline household survey

We can also test the hypotheses among our enrolled sample of households using data from patient registers linked to baseline survey data, with the linking based on voucher IDs.

During the voucher distribution, a short baseline survey will collect information from one

possible configurations for that stratum size, and we randomize among them. Because the LP arm offers the most generous subsidy, to avoid overwhelming centers, we cap the number of LP villages per center at 1—which binds only in the largest centers (8-10 villages) — and assign the excess villages to Control. To compensate for this lower probability of assignment to LP in large strata, we round up to the nearest 0.5 for LP in smaller strata (3 to 5 villages), rounding down for Control. (The PD arm's value is always exactly a multiple of 0.5, so no rounding is needed.) The expected treatment probabilities, based on the number of strata in our sample by stratum size, are 32.4% Control, 50.0% PD, and 17.6% LP. As the treatment probabilities vary by stratum size, we include stratum-size fixed effects in the analysis.

[5]These variables are predetermined at the start of treatment but become endogenous to treatment thereafter. The measures collected in the baseline survey do not have this limitation.

representative from each household. During advertising, the instructions will be for the female head of HH or the wife of the male head of HH to attend. The survey will collect basic demographics about household members and household wealth proxies, specifically the HH head's education; HH head's spouse's education; whether the HH owns a phone, any beds, a toilet, and any livestock; and where the HH head, the wife of the HH head, and the youngest child received care the last time they were sick or injured. In addition, we will collect information on the total number of children in the household (see below), whether anyone in the HH has ever been to ODH and whether the respondent or their children have foregone care in the last year due to cost.

The household representative will be given separate vouchers for each individual age 16 and older in the household.[6] Each voucher has a unique ID, which is recorded in the survey along with the age and gender of the voucher holder and the number of boys and girls under sixteen for which the voucher holder is a primary parent. Because the nurse records the voucher ID number in patient registers for voucher-using visits, we will be able to merge the outcome data (patient register data) with baseline data.

The advantages of the baseline data are four-fold. First, with the patient data, we observe only those who show up at the ODH center, whereas incorporating the baseline data allows us to assess the composition of who takes up relative to who is offered treatment. Second, we can collect more household wealth proxies than is feasible on the patient register. Third, we collect truly predetermined measures of place of care the last time it was needed and ever having been to ODH before, whereas in the patient registers, these variables eventually become endogenous to treatment. Fourth, we can assess whether there is baseline balance across arms among households who experience the salience effect of voucher distribution.

## Baseline LC1 survey

At baseline we will also conduct a short survey of the LC1 (village elected official) to collect village-level proxies for poverty, mostly for baseline balance tests.

[6] Children can access discounted care if accompanied by an adult from their household who has a voucher.

# 5 Hypotheses

The hypotheses below pertain to log patients visits. We suppress "log" for simplicity.

## Primary hypotheses

H1: More patient visits in PD than C. *(Adding the discount window increases demand.)*

H2: More patient visits in PD than C for US patients compared to non-US patients. *(Increased demand caused by the discount window is larger for the underserved.)*

H3: *US patients place less value on quality relative to price compared to non-US patients.*

H3 is the hypothesis that PD (relative to C) draws in more US than non-US patients not just because of more price sensitivity among the US but also because of less sensitivity to quality — i.e., that the single crossing property holds.[7] The LP arm, relative to C, gives an estimate of pure price sensitivity, and a key reason for including the arm is that it enables testing H3, or the role that quality degradation plays in PD drawing in relatively more US patients.

Intuitively, if the relative treatment effect of LP on US visits (relative to non-US visits) is the same as for PD, that is suggestive that PD only drew in more US because of their greater price sensitivity, with quality degradation playing no additional role; in contrast, if LP has less of an advantage than PD for US visits (relative to non-US), that suggests that PD's advantage for US stems also from quality degradation. Thus our primary test compares the ratio of the treatment effect of PD among US relative to non-US with the ratio of the treatment effect of LP among US relative to non-US, testing whether PD draws in relatively more US than LP does. This is numerically equivalent to testing whether the ratio of "quality sensitivity" (proportional increase in patient visits between PD and LP) to "price sensitivity" (proportional increase in patient visits between C and LP) is smaller for US than non-US patients — a reduced form test of the single crossing

[7] Strictly speaking, heterogeneity in price sensitivity alone would also be consistent with the single-crossing property (SCP), since SCP requires only that the ratio of quality to price sensitivity is lower for US than non-US patients, which could be satisfied even if quality sensitivity is the same across groups. However, the more substantive question for our setting is whether quality degradation is playing an active role in the targeting — i.e., whether PD achieves better targeting than a pure price cut (LP) would. H3 is therefore best understood as a test of whether quality differentiation contributes to targeting, not just whether SCP holds in a technical sense.

property.

An alternative route to establishing H3 is via two supplementary tests jointly. The first is a quality responsiveness test: since LP offers the same low price as the PD window but at higher quality, comparing LP to PD isolates the role of quality, and we can test whether US patients are less responsive to this quality improvement than non-US patients, i.e., whether the proportional increase from PD to LP is smaller for US than non-US (SH4, also listed as a standalone secondary hypothesis below). SH4 can be combined with one of two other tests to be sufficient for the single-crossing property, each under mild conditions. 1) PD heterogeneity by US status (H2): The intuition is that H2 establishes that PD draws in relatively more US patients, but this could reflect price sensitivity alone; SH4 rules that out by showing the gap is larger in PD than LP, implying that differential quality sensitivity must also be at work. 2) Heterogeneity in price sensitivity by US status (SH3): whether the proportional increase from C to LP is larger for US than non-US patients. Combined with SH4, SH3 is also jointly sufficient for showing the ratio of quality relative to price sensitivity is lower for the US (i.e., the single-crossing property), as SH4 shows the numerator is lower for US and SH3 shows the denominator is higher.

In addition to these reduced-form tests, we also plan to estimate a structural demand model to compare the quality-relative-to-price-sensitivity of US and non-US patients.

**US Measures** We will use three main measures of being US, one based on type of care received when sick in the past, the second based on household poverty indicators, and the third, a data-driven measure of price sensitivity. The first measure is an individual-level measure: whether the person's last care was no care (rather than care at ODH, a drug shop, a government clinic, or other.) We will also report results using last care was not at ODH or the individual or household had never been to ODH before, which we expect to be higher-powered tests.[8] We note, however, that the last-care measures show some inconsistencies across centers and between the baseline survey and register data; we will explore these inconsistencies further to decide on the exact measure to be used.

The second main measure of being US will be a HH-level poverty index that combines

[8] As an alternative specification, we will use last care was no care or at a drug shop, grouping these two categories of lacking adequate care; drug shops dispense medicines but are not staffed by a qualified health care professional.

male HH head education, female HH head education*, has a toilet, owns a phone, owns a bed*, owns livestock*, and anyone in the HH having foregone care in the last year due to cost*, where the variables with asterisks are only available when conducting analyses that link patient records to household surveys.[9] We will use a continuous measure (discretizing the variable to fit within our negative binomial estimation framework), but because results are often simpler to interpret with an indicator variable, we will also show results using an indicator for being above-median in terms of poverty.

Our final measure will be a data-driven predicted price sensitivity score: specifically, a prediction of heterogeneous treatment effects of LP relative to C, constructed from baseline characteristics using a causal forest.[10] We can use this measure in heterogeneity analyses involving PD to test whether price-sensitive patients are also less deterred by quality degradation — i.e., whether quality degradation can help screen out those with high WTP from taking up the low-price product, as in classic price discrimination models. This measure will vary at the individual level (though we will also explore allowing the prediction to vary by visit characteristics such as diagnosis).

## Secondary hypotheses

SH1: More patient visits in LP than C. *(Demand is decreasing in price.)*

SH2: More patient visits in LP than PD. *(Demand is increasing in quality.)*

SH3: More patient visits in LP than C for US patients compared to non-US patients. *(The underserved are especially price elastic.)*

SH4: More patient visits in LP than PD for non-US patients compared to US patients.

[9] Male HH head's education or female HH head's education will be missing in single-headed HHs. The asterisk applies to only one of the two HH head education variables because the register records a single HH head education variable, which may reflect either a male or female HH head depending on household composition.

[10] With infinite data, we could use a split-sample approach to fit the prediction, but this would come with too much of a power sacrifice. To help mitigate overfitting concerns, we can also compare LP villages before and after treatment begins to create alternative predictions that are not subject to concerns of overfitting the control group data. This fully addresses the concern when testing H2, which compares PD to C and does not use the LP arm. For H3, overfitting the LP data leads to a conservative test of the hypothesis: if certain types of patients have spuriously high outcomes in LP (and hence spuriously high predicted price sensitivity), they would also have a spuriously high LP vs PD response (in expectation), making them look more quality-sensitive than they are. This would bias us against finding the negative relationship between price and quality sensitivity that H3 predicts.

*(The underserved are less sensitive to quality degradation at a given price.)*

SH5: Fewer full-fee patient visits in PD than C; more full-fee patient visits in PD than LP. *(The discount window causes some patients who would have paid the full fee to shift to the lower-price option, but still retains some full-fee patients (and more than LP).)*

SH6: Fewer discounted patient visits in PD than LP. *(Lowering quality / restricting the time of discounts reduces their usage.)*

SH7: Ratio of mild malaria to severe malaria visits is higher in PD and LP (pooled) than C. *(Lower prices encourage earlier care and reduce more serious illness.)*

## Heterogeneity analyses

We will test whether PD (both relative to C and relative to LP) has heterogeneous effects along the following dimensions:

- Gender
- Age group (under age 5, age 5 to 18, age 18+)
- Village-level poverty index (based on LC1 survey and aggregated HH survey data)
- Distance to ODH center

We do not have signed hypotheses except that PD will lead to larger increases in visits among poorer villages, and to larger increases in visits among adolescent girls, who may be underserved at baseline. The heterogeneity analyses are mostly exploratory/will be reported for completeness.

## Exploratory analyses

Some additional exploratory analyses we intend to conduct are to

- Estimate heterogeneity by age-gender combinations (e.g., girls age 5 to 18 compared to boys age 5 to 18).
- Simulate the impacts of third degree price discrimination based on observables (e.g.,

a strategy of targeting a discount only to women could be simulated as the combination of the control arm for men and the LP arm for women).

- Examine whether the ratio of access gains to revenue losses is more favorable in PD than LP (i.e., does PD retain much of the access gains of LP without as much loss in revenue). Specifically, we will compare across LP and PD the ratio of the additional visits generated relative to the revenue per patient foregone. We will use both total visits and visits by US patients as the measure of access.
- Test if H2 is stronger in centers with more or larger villages assigned to PD. The hypothesis is that congestion enhances the screening out of non-US during the discount window.
- Estimate treatment effects for the period after the intervention has ended, to assess intertemporal substitution and/or persistence.

# 6 Estimation

## Based on a sample of village-weeks

Our main estimation approach will use visits recorded in the patient registers. The outcome will be count of patient visits to the ODH center, aggregated at the village-week or village-week-demographic group level. The counts can use either all visits or all visits with a recorded voucher ID. The first approach over-includes relative to the enrolled households and means we cannot use variables from the baseline survey. The second approach under-includes visits, as some patients may not bring their vouchers. In particular, patients who only would receive a 200 UGX discount (control group and PD outside of the discount window) have less of an incentive to bring their voucher. Thus, the main trade-off is that using all visits is more comprehensive, with less chance of under-counts that vary by treatment arm, while restricting to vouchers users enables a richer or more accurate measure of US constructed from baseline survey data.

We will estimate treatment effects using negative binomial models. This accommodates count data and an expectation that treatment will have a proportional effect on visits. In simulations with the pre-period data, we find that the negative binomial models are higher powered than Poisson models. However, they sometimes fail to converge and they

produce biased coefficients when proportional treatment effects are larger than about 5x, so we will also use Poisson models if these issues materialize. Similarly, simulations suggest higher power aggregating to the week-level rather than day-level, but we will aggregate to the day level as a robustness check.

We will estimate a difference-in-differences specification using negative binomial pseudo maximum likelihood. Let $Y_{vgt}$ denote total visits for village $v$ for time period $t$ for group $g \in \{NonUS, US\}$, which refer to the HH or individual poverty proxy for being underserved. $Y_{vt}$ without the subscript $g$ denotes total visits for village $v$ for time period $t$. Standard errors will adjust for non-independence within villages.

We will use pre-intervention data combined with post-intervention data for the period when the vouchers are valid. $Post_{vt} = 1$ from the day that the treatment assignment is revealed to the village (which is also the first day the vouchers are valid) to the end of the period when the vouchers are valid. We will use pre-period data from roughly October 2025 onward. Alternatively, because the beginning of treatment is staggered over several weeks, we can use the same number of pre-period weeks per center.

### Main effects of treatment

To estimate overall treatment effects, we will estimate the equation below:[11]

$$\mathbb{E}[Y_{vt}] = \exp\Big(\alpha + \beta_{LP} LP_v + \beta_{PD} PD_v + \beta_{Post} Post_t + \theta_{LP} LP_v \times Post_t + \theta_{PD} PD_v \times Post_t + \text{FE} + \varepsilon_{vt}\Big). \quad (1)$$

H1 maps to $\theta_{PD} > 0$ in equation (1). SH1 maps to $\theta_{LP} > 0$, and SH2 maps to $\theta_{LP} > \theta_{PD}$.

[11] FE denotes fixed effects. Besides stratum size-by-*Post* fixed effects, we plan to include center-week FEs because, based on simulations, they improve power.

**Heterogeneous treatment effects by US status**

To test the relative effects for US and non-US, we will estimate a triple difference model with two observations per village-week, one for US and one for non-US:[12]

$$\begin{aligned}\mathbb{E}[Y_{vgt}] = \exp\Big(&\alpha + \beta_{LP}LP_v + \beta_{PD}PD_v + \beta_{Post}Post_t + \gamma NonUS_g \\ &+ \theta_{LP}LP_v \times Post_t + \theta_{PD}PD_v \times Post_t + \delta_{Post}NonUS_g \times Post_t \\ &+ \delta_{LP}LP_v \times NonUS_g + \delta_{PD}PD_v \times NonUS_g + \lambda_{LP}LP_v \times Post_t \times NonUS_g \\ &+ \lambda_{PD}PD_v \times Post_t \times NonUS_g + \text{FE} + \varepsilon_{vgt}\Big). \end{aligned} \tag{2}$$

H2 maps to $\lambda_{PD} < 0$ in equation (2), or PD attracts relatively more US patients than non-US patients. SH3 maps to $\lambda_{LP} < 0$. SH4 maps to $\lambda_{PD} < \lambda_{LP}$. The primary test of H3 — that the ratio of the LP treatment effect for US relative to non-US is less than the ratio of the PD treatment effect for US relative to non-US — maps to $\frac{\theta_{LP}}{\theta_{LP}+\lambda_{LP}} < \frac{\theta_{PD}}{\theta_{PD}+\lambda_{PD}}$ in the case of a binary measure of US, which simplifies to $\theta_{LP}\lambda_{PD} < \theta_{PD}\lambda_{LP}$. The latter is the condition for the case of a continuous measure of US as well.[13] Alternately, we can establish H3 by jointly testing SH4 with either H2 or SH3.

A second way to test the relative effects of our treatments by US status is to estimate the following OLS regression at the visit level, which tests how the characteristics of patients who visit the clinic vary across treatment arms.[14] Defining $NonUS_{iv(i)t}$ as the non-US

[12]For continuous measures, we will discretize them and have more than two observations per village-week, e.g., into quintiles with five observations per village-week.

[13]In the continuous case, where one cannot calculate the treatment effects of PD and LP for two distinct groups (i.e., where $\frac{\theta_{LP}}{\theta_{LP}+\lambda_{LP}} < \frac{\theta_{PD}}{\theta_{PD}+\lambda_{PD}}$ does not compare treatment effects across two distinct groups), the single crossing condition is that the ratio of the treatment effect of PD relative to the treatment effect of LP is decreasing in $NonUS_g$. From equation (2), the treatment effect of PD for a given group $g$ is $\theta_{PD} + \lambda_{PD} \cdot NonUS_g$ and the treatment effect of LP is $\theta_{LP} + \lambda_{LP} \cdot NonUS_g$. By the quotient rule, the condition that $\frac{\theta_{PD}+\lambda_{PD}\cdot NonUS_g}{\theta_{LP}+\lambda_{LP}\cdot NonUS_g}$ is decreasing in $NonUS_g$ requires: $\lambda_{PD}(\theta_{LP} + \lambda_{LP} \cdot NonUS_g) - \lambda_{LP}(\theta_{PD} + \lambda_{PD} \cdot NonUS_g) < 0$, which simplifies to $\theta_{LP}\lambda_{PD} < \theta_{PD}\lambda_{LP}$ — the same condition as in the binary case.

[14]This "characteristics regression" tests the same underlying hypotheses as the equation (2) "volumes regression," but does so by asking directly whether visitor composition differs across arms rather than whether visit counts respond differentially by wealth group. The two approaches are therefore complementary. However, they are not equivalent: the characteristics regression implicitly tests a condition on the ratio of visit *levels* across groups, while the visit-count regression tests a condition on *log* changes. The latter maps more naturally to constant elasticity demand models standard in this literature, but power calculations suggest that the characteristics regression is often higher powered for testing heterogeneity (H2, SH3, SH4), perhaps because it is estimating fewer parameters.

status of visitor $i$ who visited the clinic from village $v$ on day $t$. We estimate:

$$\begin{aligned} NonUS_{iv(i)t} =& \alpha + \gamma_{LP} LP_{v(i)} + \gamma_{PD} PD_{v(i)} + \beta_{Post} Post_t \\ &+ \theta^e_{LP} LP_{v(i)} \times Post_t + \theta^e_{PD} PD_{v(i)} \times Post_t + \text{FE} + u_{ivt}, \end{aligned} \tag{3}$$

where the superscript $e$ distinguishes these coefficients from the coefficients in equation (2). H2 maps to $\theta^e_{PD} < 0$ in equation (3), SH3 maps to $\theta^e_{LP} < 0$, and SH4 maps to $\theta^e_{LP} > \theta^e_{PD}$. (Note that there is not an analog of the direct test of H3 using equation (3); the only way to establish H3 using equation (3) is by combining SH4 with either H2 or SH3.)

A final way to test the relative effects of our treatments by US status is to use within-arm timing variation. We can aggregate the data at the village-week-demographic level separately for two types of visits — those that occur between 8 and 9:30 am ("window visits") and those that occur at other times ("non-window visits"). Using $w$ to index whether a visit falls in a window or non-window time, we can then estimate the following regression using data from the Control and PD groups:[15]

$$\begin{aligned} \mathbb{E}[Y_{vgtw}] = \exp\Big(&\alpha + \beta_{PD} PD_v + \beta_{Post} Post_t + \gamma NonUS_g + \xi_1 Window_w \\ &+ \theta_{PD} PD_v \times Post_t + \delta_{PD} PD_v \times NonUS_g \\ &+ \lambda_{PD} PD_v \times Post_t \times NonUS_g + \xi_2 PD_v \times Window_w + \xi_3 Post_t \times Window_w \\ &+ \xi_4 NonUS_g \times Window_w + \xi_5 PD_v \times Post_t \times Window_w \\ &+ \xi_6 Post_t \times NonUS_g \times Window_w + \xi_7 Post_t \times NonUS_g \\ &+ \pi_{PD} PD_v \times Post_t \times NonUS_g \times Window_w + \text{FE} + \varepsilon_{vgtw}\Big). \end{aligned} \tag{4}$$

H2 maps to $\pi_{PD} < 0$.

**Treatment effects on specific types of visits**

Some of our hypotheses entail restricting the sample to a subset of visits.

We will estimate effects for full-fee visits only to test for cannibalization.[16] Specifically,

[15] LP data can also be pooled with Control group data for power.

[16] We will code visits receiving the 200 UGX discount as "full-fee"; "discounted visits" here means those where the 1000 UGX flat fee was applied.

we will estimate a variant of equation (1), but restricting visits to full-fee visits; we denote the count of full-fee visits as $Y^{FF}$.

$$\mathbb{E}[Y_{vt}^{FF}] = \exp\Big(\alpha + \beta_{LP}LP_v + \beta_{PD}PD_v + \beta_{Post}Post_t + \theta_{LP}LP_v \times Post_t + \theta_{PD}PD \times Post_t + \text{FE} + \varepsilon_{vt}\Big). \quad (5)$$

SH5 maps to testing $\theta_{PD} < 0$ and $\theta_{PD} > \theta_{LP}$ in equation (5). Note that $\theta_{PD}$ compares PD with the control group; because of the salience effect, the intervention could cause full-fee visits to increase in PD villages, i.e., it is possible that $\beta_{Post} + \theta_{PD} > 0$. Because the dependent variable may equal 0 for many observations in the LP group,[17] which our simulations suggest can hurt power and/or prevent convergence, we can also estimate a variant of equation (5) that omits the LP group to test the $\theta_{PD} < 0$ part of the hypothesis.

We will estimate effects for discounted visits only to see if degrading the quality of the discounted product through PD deters take-up of the discount, relative to offering the discount for all visits in LP. Specifically, we will estimate a variant of equation (1), but restricting visits to discounted visits; we denote the count of discounted visits as $Y^{Disc}$. We will exclude the control group (who should not receive any discounted visits) from the regression, as the hypothesis is not relevant for them. PD will be the omitted group:

$$\mathbb{E}[Y_{vt}^{Disc}] = \exp\Big(\alpha + \beta_{LP}LP_v + \beta_{Post}Post_t + \theta_{LP}LP_v \times Post_t + \text{FE} + \varepsilon_{vt}\Big). \quad (6)$$

SH6 maps to testing $\theta_{LP} > 0$ in equation (6).

To test if lower prices induce earlier visits among those with malaria, thereby reducing severe malaria visits, we will pool PD and LP villages, denoting them by $Treat$, and estimate heterogeneous effects by type of malaria visit. The outcome uses mild or severe malaria visits, with other visits excluded. Here, $g \in \{Mild, Severe\}$.

$$\mathbb{E}[Y_{vgt}^{Mal}] = \exp\Big(\alpha + \beta_T Treat_v + \beta_{Post}Post_t + \gamma Mild_g + \theta_T Treat_v \times Post_t + \delta_{Post}Mild_g \times Post_t + \delta_T Treat_v \times Mild_g + \lambda_T Treat_v \times Post_t \times Mild_g + \text{FE} + \varepsilon_{vgt}\Big). \quad (7)$$

[17] Full-fee visits will not mechanically be 0 in LP since people may come out of hours or the nurse may mistakenly not give them a discount.

SH7 maps to $\lambda_T > 0$, or the lower-priced options increase mild malaria visits more than severe malaria visits. Because such an effect would take some time to materialize, we will exclude the first week of post-period data for this test.[18]

As with the heterogeneity analyses by US status (equations 2 through 4), SH7 can also be tested using regressions of other forms (e.g., a characteristics regression with the share of mild malaria visits as the outcome — the analog of equation 3).

## Based on a sample of person-weeks (or household-weeks)

An alternative approach to using the sample of patient visits is to test the hypotheses using the sample of voucher recipients in households enrolled at baseline. In this case, the sample also includes people who never visit the clinic. In this analysis, we do not have pre-period visit data.

Here, we will estimate a linear probability model where the outcome is whether the person visited the clinic in a given week. For example, the analog of equation (2) is:

$$\begin{aligned} AnyVisit_{ivgt} =& \alpha + \beta_{LP} LP + \beta_{PD} PD + \gamma NonUS \\ &+ \delta_{LP} LP \times NonUS + \delta_{PD} PD \times NonUS + \text{FE} + \varepsilon_{ivgt}. \end{aligned} \tag{8}$$

H2 maps to $\delta_{PD} < 0$ here, for example.

Because matching children to the baseline data might be challenging, we will also estimate these regressions aggregated to the household-week rather than person-week. We can also collapse the data to one observation per person and estimate a linear probability model where the outcome is $AnyVisit_{ivgt}$ over the intervention period or a count model where the outcome is total visits $Y_{ivg}$. The advantage of maintaining the panel structure is that it allows us to account for the different start time of the post-period by village.

[18]For patients whose counterfactual care would have been other paid care (e.g., government clinics), an offsetting factor is that savings at ODH are larger for severe malaria care than mild care, potentially inducing greater substitution from other paid care to ODH for severe than mild malaria. A sharper test would therefore restrict to those whose counterfactual care (proxied by pre-period care) would be no care, ODH, or drug shops.