Soft and hard commitment devices to increase HIV testing

Laura Derksen (University of Toronto)

Jason Kerwin (University of Minnesota)

Natalia Ordaz Reynoso (University of Minnesota)

Olivier Sterck (University of Oxford)

Contents

Preamble

This pre-analysis plan was finalized in September 2019, after the collection of baseline data, but before the digitization of the data that will be used to construct outcome variables. Some of the decisions made in this pre-analysis plan are based on a preliminary analysis of baseline data. Results from this preliminary analysis are provided in the appendix.

1 Research Strategy

In this study, we investigate whether a soft commitment device (a scheduled appointment) and a hard commitment device (a small gift of cell phone credit conditional on HIV testing) can increase HIV testing among men by reducing procrastination.

1.1 Interventions

We assess the impact of two interventions that aim at reducing procrastination and thereby overcome the problem of continuously putting off HIV testing.

(1) A soft commitment device: selected participants are offered scheduled appointments to get tested for HIV. Appointments are devices for mental and social commitment.

(2) A hard commitment device: selected participants are invited to invest in a mechanism that conditions a small transfer of cell phone credit on attending a clinic that offers HIV tests.

The scripts associated with these interventions are provided in the SurveyCTO questionnaire in the Appendix.

1.2 Experimental design

This study is a randomized controlled trial using a full factorial design. The randomization process was predefined in a .csv downloaded onto the survey tablets, and is hence reproducible. Randomization is at the individual level without stratification. Each study participant was automatically allocated to a study arm based on their Study ID which is composed of their enumerator ID, day of the study, and number of the survey. For example, the first enumerator's first respondent on the first day of the study has Study ID 010101.

Prior to applying the results of the randomization, we elicited incentivized preferences for a hard commitment device from each participant. Individuals were then randomly assigned to one of four study arms.

Arm 0: 25% of participants were assigned to the control group.

Arm 1: 25% of participants were offered the hard commitment device if they stated a preference for it. The original preference we elicited from all participants was honoured if subjects were selected for this treatment. If their preference was to refuse the commitment device, they received cell phone credit worth MK1000 immediately (1000 Malawian Kwacha, or approximately USD $1.36). If their preference was to accept, they forewent the immediate credit worth MK1000, but received the same amount of cell phone credit if they showed up at one of the HIV testing sites within two months of their interview. They do not have to agree to an HIV test to receive the cell phone credit.

Arm 2: 25% of participants were offered a scheduled appointment to get tested for HIV. They are allowed to refuse the appointment.

Arm 3: 25% of participants were offered both the hard commitment device and the appointment. Participants were allowed to refuse either one, or both.

This design is summarized in Table 1.

Table 1 – Experimental Design

Arm	Elicit preference for commitment device	Offer commitment hard device	Offer soft commitment device (appointment)
0 (control)	X		
1	X	X	
2	X		X
3	X	X	X

The intent-to-treat effect (ITT) of the soft commitment device (being offered an appointment) will be measured by comparing those randomly selected to get an appointment (groups 2 and 3) to those who were not (groups 0 and 1). As not everyone agreed to an appointment, we will also measure the Average Treatment Effect on the Treated (ATET) using an IV strategy.

The intent-to-treat effect of offering the hard commitment device will measured by comparing those who were assigned to groups 1 and 3 to those assigned to groups 0 and 2. We can also measure the effect on compliers using the same comparison, but restricting our sample to those who opted into the hard commitment device.

1.3 Sampling

The study site is Zomba Town in southern Malawi. We targeted a sample of approximately 1200 men who own a mobile phone, and who were recruited at bars and nightclubs in Zomba Town. By recruiting participants at bars, we aimed to target particularly high-risk men. We focus on men owning mobile phones as they are expected to be wealthier, which is associated with higher risk of HIV/AIDS in Malawi. We sampled exclusively men because they are tested for HIV less frequently than women are, principally because women who have children are subject to routine tests as part of antenatal care.

All participants received a MK1000 gift of cell phone credit at the end of the survey, and a MK500 cell phone credit voucher for redemption at any HIV testing site in Zomba Town; participants who were randomized into the hard commitment arm (Arm 1) and who opted for the commitment device had their initial MK1000 gift deferred until they appeared at one of the testing clinics. Participants do not have to agree to an HIV test to redeem the MK500 voucher (nor, if relevant, to receive the MK1000 credit).

1.4 Sample size

After excluding participants that do not satisfy inclusion criteria (see Section 2.1), our sample size is 1,232 participants. For the soft commitment device intervention (appointments), we have 636 participants in the treatment group and 596 participants in the control group for the intent-to-treat analysis.

For the hard commitment device intervention (option to receive money only if you appear for testing), we have 602 participants in the treatment group and 630 participants in the control group for the intent-to-treat analysis. Sample size by study arms is summarized in Table 2.

Table 2 – Sample size by study arm

Treatment	Respondents' choices and randomization	N
Hard CD	Refused the CD	607
	Opted into the CD, randomized out (receive credit worth MK1000 immediately)	332
	Opted into the CD, randomized in (receive credit worth MK1000 if shows-up at a clinic)	293
Soft CD	Randomized out (no appointment)	596
	Randomized in, but refuse the CD (no appointment)	220
	Randomized in, opted into the CD (appointment scheduled)	416

1.5 Data

We are collecting the following data on the subjects in this study:

(a) Baseline data collection.

During our initial interaction with study participants, we collected demographic information (gender, ethnicity, age, place of birth, educational achievement, marital status), as well as sets of questions aimed identifying mechanisms through which procrastination may operate. The SurveyCTO questionnaire is provided in the Appendix.

(b) Testing data

The project hired qualified HIV Diagnostic Assistants (HDAs) and integrated them into each of the HIV testing clinics in Zomba Town. Our study is run in coordination with the participating clinics and the district health office, to ensure that patients' privacy is respected and that we adhere to all legal and ethical requirements for protecting the data. We are adhering to strict confidentiality and data security protocols in order to protect the privacy of the men in our sample.

HDAs are in charge of collecting vouchers and performing HIV tests for study participants. Those who test negative are encouraged to seek a second confirmatory test after 3 months, in line with local protocol. HDAs are working with the clinics to integrate newly diagnosed individuals into ART initiation and care.

Our primary outcome variable will HIV testing. Participant voucher codes and phone numbers are recorded manually on a notebook for the purpose of identification. HDAs are also recording whether participants agree to be tested for HIV, as well as HIV test results by study arm, and individual ART initiation. The information in the notebooks will be digitized at the end of the experiment.

2 Empirical analysis

2.1 First-tier analysis: effects of treatments on testing

Our primary outcome variable is an indicator variable for HIV testing based on data collected by the HDAs at the clinics. The testing dummy will be equal to 1 for participants who tested

for HIV in person,[1] and equal to 0 otherwise. We will assess the intent-to-treat impact of the interventions on this testing indicator.

Our analysis will focus on the subsample of participants who satisfy the following conditions:

- They named one of the study clinics (within Zomba Town) when asked the following question: "We will give you a voucher for HIV testing. Which clinic would you like to redeem it at?"
- They do not report currently being on ART.

Men who do not meet those two criteria will be excluded from our analysis.

We will estimate the following equation using linear regressions:

$$Y_i = \alpha + \beta_1 H_i + \beta_2 S_i + \gamma' X_i + \mu_i + \tau_i + \lambda_i + \varepsilon_i \quad (1)$$

where Y_i is the outcome variable, H_i is a dummy variable that takes a value of one if the participant was randomly selected to get the hard commitment device, and S_i takes a value of one if the individual was randomly selected to get the soft commitment device. X_i is a list of baseline characteristics – listed in Table 3 – that are statistically significant in a regression on a past testing dummy, as well as the past testing dummy itself, the number of times respondents tested in the past, and a dummy equal to 1 if respondents opted into the hard commitment device. These control variables are included in the regression to increase precision. We will dummy out missing values of the baseline characteristics. We will control for date-of-survey fixed effects, μ_i, enumerator fixed effects τ_i, and "selected clinic" fixed effects λ_i.[2] The index i denotes an individual participant. As a robustness check, we will test whether our selection of controls matters for the results by estimating equation (1) with control variables selected using the double machine learning method of Chernuzhukov et al. (2017), using the Stata command pdslasso with an extended list of control variables described in Tables 3 and 4, as well as date of survey fixed effects, enumerator fixed effects and "selected clinic" fixed effects.

[1] The testing dummy is equal to 0 if a likely impostor tested on behalf of a study participant. Likely impostors are defined as follows: they either failed to provide consistent answers for at least two out of the three "security questions" (three security questions – labelled primary, districtbirth, and districtmother in the SurveyCTO - are asked both at baseline and at the clinics by HDAs to identify impostors), or have been directly identified as impostors by HDAs.

[2] The baseline survey asks respondents to name their preferred clinic, which is what we use as their "selected clinic". This does not have to be the same clinic where they actually get an HIV test. We cannot control for the clinic where the respondent actually gets tested because that is only defined for men who get an HIV test.

The coefficient β_1 measures the average intent-to-treat effect of the hard commitment device intervention. The coefficient β_2 measures the average intent-to-treat effect of the soft commitment device intervention (appointments). Effects should be interpreted in percentage-point terms.

We will measure the ATET for the soft commitment device using an IV strategy. The instrument is the treatment dummy S_i. The instrumented dummy is equal to 1 for respondents who were randomly selected to benefit from the soft commitment device and accepted to schedule an appointment with the treatment dummy S_i and equal to 0 otherwise.

For the hard commitment device, we know who the compliers are. We can therefore avoid the IV strategy (which reduces power), and estimate the ATET by estimating the following equation:

$$Y_i = \beta_0 N_i + \beta_1 S_i \times N_i + \beta_2 O_i + \beta_3 O_i \times H_i + \beta_4 O_i \times S_i + \gamma' X_i + \mu_i + \tau_i + \lambda_i + \varepsilon_i \quad (2)$$

where O_i is a dummy equal to one for participants who opted into the hard commitment device before learning whether or not they would receive it, and $N_i = 1 - O_i$ is a dummy equal to one for those who opted out. The coefficient β_3 in equation (2) measures the ATET of the hard commitment device (the effect of the hard commitment device on those who ex-ante wanted to receive it). The coefficients β_1 and β_4 will be analyzed in the second-tier analysis relating to heterogeneous treatment effects (Section 2.2).

We will present heteroskedasticity-robust standard errors for all estimates, with no clustering adjustment since our treatment is randomized at the individual level (Abadie et al., 2017).

Table 3 – Baseline characteristics used for the balance table, the list of control variables, and the attrition tests

Variable	Reference in the questionnaire (variable name)	Construction of indicator
Age	age	
Literacy dummy	read	
Married dummy	married	Equal to 1 if the respondent is married and 0 otherwise
Willingness to get tested (dummy)	testherenow planfuturetest	We will extract the first principal component of these two questions, following Filmer and Pritchett (2001)
Past testing dummy	evertested	The dummy evertested is 0 if timestested is equal to 0
Number of times tested since born	timestested	This variable is equal to 0 if evertested ==0.

Self-reported tendency to postpone testing (dummy)	cd_hivpostpone	The dummy is manually coded from the string variable cd_hivpostpone. The dummy is equal to 1 if the answer contains the words "afraid", "fear", "lazy", "busy", "laziness", "several" " i can't get tested", and equal to 0 otherwise.
Life expectancy with ART	lifegainarv	
Expected likelihood of being HIV+	owninfection	owninfection*10
Opted into the hard commitment device	cd_finaldecision	

Table 4 – Supplementary covariates considered when using the double machine learning method of Chernuzhukov et al. (2017)

Variable	Reference in the questionnaire (variable name)	Construction of indicator
Ethnic group dummies	ethnic	We will construct dummies for the five largest ethnic groups in our sample (Nyanja, Lomwe, Yao, Chewa, Ngoni) plus a residual category
Years of education	schooling	The indicator ranges from 0 to 13, where 13 is for 13 years of education or above.
Number of children	alivechildren	
Religion dummies	religion	Three dummies: Muslim, Christian, and Other.
An asset index	assets	We will extract the first principal component of the assets list, following Filmer and Pritchett (2001).
Boyfriend/girlfriend	chibwenzi	Equal to 1 if the respondent has a boyfriend/girlfriend and 0 otherwise
Sexual activity dummy	sexpast7days	Equal to 1 if the respondent had sex in the past 7 days and 0 otherwise
Number of partners in the last 12 months	totalpartners12	
Self-reported tendency to live for today	livefortoday	
Occupation dummies	occupation	We will construct dummies for the following categories of activities: "Occupation in military/police/security" "Occupation in skilled activity" "Occupation in transport sector" "Occupation in

		manual activity" "No occupation/student" "Other occupation (e.g. trade, agriculture)".
Household expenditures	foodexpend healthexpend schoolexpend businessexpend transportexpend entertainmentexpend accomodationexpend otherexpend	We will calculate the sum of the different categories of expenditures and transform the variable using the inverse hyperbolic sine function.
Opportunity cost of testing	wages	We will take the inverse hyperbolic sine of the variable wage We will also create a dummy equal to 1 for those with no wage.
Life-expectancy gain with ARV	liveHIVimmediately liveHIVnoARV	(liveHIVimmediately - liveHIVnoARV)
HIV/AIDS knowledge index	joycebanda knowARV costARV cureHIVherbs cureHIVpray mosquitosHIV	This variable is missing if the answer to the question "joycebanda" is "yes" (this question is an attention check). We will extract the first principal component of the five other questions, following Filmer and Pritchett (2001)
Serodiscordance expected likelihood	wifeposhusbneg	

2.2 Second-tier analysis: interactions and heterogeneous treatment effects

The second tier analysis can be split in three parts: (1) an analysis of interaction effects between the two treatments, (2) an analysis of heterogeneous treatment effects, and (3) an analysis of whether the treatment affects those least or most likely to test in the absence of intervention, using the endogenous stratification method of Abadie et al. (2018).

Interaction effects between the two treatments

We will estimate the following specification to test for interaction effects between the two treatments:

$$Y_i = \alpha + \beta_1 H_i + \beta_2 S_i + \beta_3 S_i \times H_i + \gamma' X_i + \mu_i + \tau_i + \lambda_i + \varepsilon_i \quad (3)$$

All variables are defined and handled in the same way as they were in equation (1). In terms of statistical inference, we will test whether the coefficient β_3 is statistically different from zero, to assess whether the hard and the soft commitment devices are substitutes ($\beta_3<0$) or complements ($\beta_3>0$).

We will also use a figure to represent the four marginal effects of interest as well as the total effect of both interventions:

- β_1 = marginal effect of the hard commitment device intervention in the absence of the soft commitment device intervention;
- β_2 = marginal effect of the soft commitment device intervention in the absence of the hard commitment device intervention;
- $\beta_1 + \beta_3$ = marginal effect of getting the hard commitment device intervention on top of the soft commitment device intervention;
- $\beta_2 + \beta_3$ = marginal effect of the soft commitment device intervention on top of the hard commitment device intervention;
- $\beta_1 + \beta_2 + \beta_3$ = total effect of both the hard and soft commitment devices.

Heterogeneous treatment effects

We will explore heterogeneous treatment effects with respect to two variables: (a) the expected likelihood of being HIV+ and (b) the preference for the hard commitment device.

(a) The functional form of the moderation relationship between the treatments and the expected likelihood of being HIV+ is unknown. On the one hand, those who believe they could be HIV positive could be more likely to react to the interventions because they have more learn/gain from the HIV test (excluding those who already know they are HIV+ and are on ART). On the other hand, those who believe they could be HIV positive often fear HIV tests because they fear learning they will die (Kaler and Watkins, 2010). The moderation relationship could therefore be positive, negative, or non-linear. To best capture a linear moderation relationship, we will estimate the following regression equation:

$$Y_i = \alpha + \beta_1 H_i + \beta_2 S_i + \beta_3 H_i \times Moderator_i + \beta_4 S_i \times Moderator_i + \beta_5 Moderator_i + {}'X_i + \mu_i + \tau_i + \lambda_i + \varepsilon_i. \quad (4)$$

To test for the presence of a non-linear moderation relationship, we will first create three dummies M_1, M_2 and M_3 based on the distribution of the moderator variable,[3] and then estimate the following regression equation:

[3] M1=1 if owninfection100==0, M2=1 if 0<wowninfection100<50, M3=1 if owninfection100>=50

$$Y_i = \alpha + \beta_1 H_i + \beta_2 S_i + \beta_3 H_i \times M_2 + \beta_4 H_i \times M_3 + \beta_5 S_i \times M_2 + \beta_6 S_i \times M_3 + \gamma' X_i + \beta_7 M_2 + \beta_8 M_3 + \mu_i + \tau_i + \lambda_i + \varepsilon_i. \quad (5)$$

In both equations (4) and (5), all variables are otherwise handled in the same way as in equation (1).

(b) The hard and soft commitment devices might be substitutes for one another. Let O_i be an indicator for opting into the hard commitment device, and $N_i = 1 - O_i$ be an indicator for opting out. Participants who opted into the hard commitment device ($O_i = 1$) expressed their desire for a commitment device. They might therefore be more likely to respond to the soft-commitment device intervention, especially if they were randomly selected to not receive the hard commitment device. We test this hypothesis by estimating two different equations.

First, we will estimate equation (2), which was introduced in section 2.1. In equation (2), we will estimate the coefficient β_1 to measure the effect of the soft commitment device on those who did not want a hard commitment device ($N_i=1$), and the coefficient β_4 to measure the effect of the soft commitment device on those who opted into the hard commitment device ($O_i=1$).

Second, we will estimate equation (7):

$$Y_i = \beta_0 N_i + \beta_1 N_i \times S_i + \beta_2 O_i + \beta_3 O_i \times H_i + \beta_4 O_i \times S_i \times (1 - H_i) + \beta_5 O_i \times S_i \times H_i + \gamma' X_i + \mu_i + \tau_i + \lambda_i + \varepsilon_i. \quad (7)$$

In this last equation (equation 7),

- β_3 = the effect of a hard commitment device on those who wanted to receive one, and who did not receive a soft commitment device.
- β_4 = the effect of a soft commitment device on those who wanted a hard commitment device but did not receive one.
- β_5 = the marginal effect of receiving a soft commitment device on those who wanted a hard commitment device and did receive one.

Endogenous stratification method of Abadie (2018)

We will implement the repeat split-sample (RSS) endogenous stratification procedure of Abadie et al. (2018) to explore whether the treatments affects those least or most likely to test in the absence of intervention. Intuitively, the method uses the control-group data to predict the

outcome variable using exogenous covariates, and then generates predicted values of the outcome for all observations including the treatment groups. The Abadie et al. RSS procedure explicitly address the problem of overfitting bias that occurs when the full sample of experimental controls is considered when estimating the predicted outcome in the absence of treatment. The approach relies on randomly splitting the sample in half, and using half of the data to predict the outcome variable (the first stage) and the other half to use the predicted outcomes for treatment effect heterogeneity analysis (the second stage). We will conduct 1000 random splits of our sample and do the two stages for each; our point estimates and standard errors will come from the mean and standard deviation of the estimates from stage 2 across the 1000 sample splits.

We will do this analysis with two different lists of covariates for the first stage of the procedure. First, we will use the control variables listed in Tables 3 as well as date of survey fixed effects, enumerator fixed effects, and "selected clinic" fixed effects. Second, we will use the double machine learning method of Chernuzhukov et al. (2017) to select covariates, starting with the extended list of covariates described in Tables 3 and 4 as well as date of survey fixed effects, enumerator fixed effects, and "selected clinic" fixed effects. In the second stage, we will consider 3 sub-groups constructed based on the predicted outcomes calculated in the first stage, following Abadie et al.

2.3 Third-tier analysis: exploratory work

The third-tier analysis is exploratory. This level of analysis is split in two parts: (1) an analysis of secondary outcomes and (2) an analysis of heterogeneous treatment effects whose direction, importance, or sign are theoretically ambiguous.

Secondary outcomes

The list of secondary-outcomes includes five dummy variables:

- Voucher dummy: a dummy variable equal to 1 for participants who redeemed their voucher and equal to 0 otherwise.
- Voucher/no testing dummy: a dummy variable equal to 1 for participants who redeemed their voucher but did not get tested for HIV, and equal to 0 otherwise.
- Tested HIV+ dummy: a dummy variable equal to 1 for participants who redeemed their voucher and tested HIV+, and equal to 0 otherwise.

- ART dummy: a dummy variable equal to 1 for participants who redeemed their voucher and agreed to initiate ARV treatment after testing HIV+, and equal to 0 otherwise.

The "Voucher dummy" and "Voucher/no testing dummy" aim at understanding why some participants might go to the clinic to redeem their voucher without agreeing to take the HIV test. As part of the hard commitment device, participants do not have to agree to an HIV test to get the money at the clinic. In a way, this hard commitment device intervention could increase voucher redemption but have no effect on testing because people still decide not to test when they are at the clinic (for example because they fear learning about the results, already know their serostatus, or do not have time to do the test). We do not know *ex ante* whether this type of behavior will be prevalent and worth analyzing in detail.

The "Tested HIV+" and "ART" dummies are considered as secondary not because they are uninteresting (quite the contrary) but rather because our statistical power will most likely be insufficient to detect any effects on these outcomes.

The analysis of secondary outcomes will follow the methodology described in sections 2.1 and 2.2. However, the HIV status and ART initiation outcomes are anonymous and are linked only to study arm, so we will not include covariates in these regressions.

Other heterogeneous treatment effects

We will explore heterogeneous treatment effects with respect to the following variables:

- Self-reported tendency to postpone testing (dummy)
- Self-reported tendency to live for today
- Life expectancy with ART

For these moderation relationships, we will estimate equation (4). All variables are otherwise handled in the same way as in equation (1).

3 Threats to internal validity and robustness tests

3.1 Balancing checks

We have two orthogonal treatment variables – cdassigned and apptassigned – whose pairwise coefficient of correlation is -0.0042. For each treatment variable, we will report an omnibus F-test of joint orthogonality following an OLS regression of the treatment indicator on all the controls listed in Tables 3 and 4 and the fixed effects. For each treatment variable, we also analyze the size of the normalized differences between the treatment and control group,

assuming that differences of 0.25 or less indicate good balance (Imbens and Rubin 2015). As we already have the baseline data at the time of writing this analysis plan, the results are presented in appendix.

3.2 Attrition

Attrition is not expected to be a problem for the main outcome of this study. Our main analysis, which relies on administrative data, will not suffer from attrition due to respondents not being found or not consenting to participate to a second survey.

3.3 Spillovers

In theory, vouchers and appointments could be passed to people in the control group. The HDAs will do rigorous identity checks to limit this possibility. Our main analysis focuses on average intent-to-treat effects.

3.4 Adjustment for multiple inference

In line with Lakens (2016), and Vanhove (2016), we adjust for multiple testing only when a single hypothesis is tested using multiple tests. We therefore do adjustments for multiple inference separately for each theoretical hypothesis tested in this research but not across hypotheses. We do not adjust for multiple inference across treatments (S or H) or across estimands (ITT or ATET).

The first-tier analysis proposes to test two different theoretical hypotheses (effects of the *H* and *S* interventions) with two different tests (β_1=0 and β_2=0 in equation 1). No adjustment is therefore needed for this level of analysis.

In the second-tier analysis, the study of the interaction between the two treatments is based on one test (β_3=0 in equation 3) that answers one question (are the hard and the soft commitment devices substitutes or complements?). No adjustment is therefore needed for this analysis.

In the second-tier analysis, the study of heterogeneous treatment effects and endogenous stratification includes different tests for each intervention. We therefore follow Anderson (2008) and Schaner (2018) and compute sharpened *q*-values that control the false discovery rate (FDR) following the two-step procedure proposed by Benjamini et al. (2006). The main tables will display sharpened *q*-values in brackets below traditional standard errors and p-values. These adjustments will be conducted separately for the heterogeneity tests related to each intervention (Table 4). We will explore other multiple hypotheses testing adjustments if new techniques become available.

We will not adjust for multiple hypotheses testing for the third level of analysis, which is exploratory in the sense that it aim at generating new hypotheses for future research rather than rejecting a pre-defined null-hypothesis (Hirschauer et al. 2019).

3.5 Heterogeneous treatment effect and omitted variable bias

Because we do not have exogenous variation in moderator variables, the analysis of heterogeneous treatment effects could be affected by omitted-variable bias. For each moderator, we will therefore also estimate a pooled specification that includes interaction terms between all control variables listed in Table 3 and the treatment dummies. We will not do this for the Abadie et al. (2018) approach, because it uses all the controls in Table 3 as predictors.

3.6 Testing without using the voucher

Participants might come to the clinic without their voucher, for example because they lost their phone, deleted the message, or rebooted the operating system of their mobile phone. In this case, participants are offered an HIV test but do not get the cell phone credit associated with redeeming the voucher. For these participants, the "testing dummy" – our main dependent variable – is equal to one if they agree to be tested. This situation is expected to be very infrequent as there is no financial incentive associated with testing without voucher. If more than 5% of all participants get tested for HIV without redeeming their voucher, we will assess whether our results are robust to considering a "voucher and testing" dummy equal to 1 for participants who redeemed their voucher and agreed to get tested for HIV and to 0 otherwise.

3.7 Impostors

Impostors are people who are not part of the experiment, but try to redeem the voucher of another individual to get money. Likely impostors are defined as follows: they either failed to provide consistent answers for at least two out of the three "security questions", or have been directly identified as impostors by HDAs.

In the main analysis, the testing dummy is equal to 1 for participants who agreed to be tested in person, and equal to 0 otherwise. The "testing dummy" is therefore equal to 0 if we suspect that an impostor tested for HIV on behalf of a study participant. As a robustness check, we will replicate the first-tier analysis without invalidating HIV tests made by likely impostors (i.e. we set their dummy variables to 1, since they appeared for testing).

4 Cost-effectiveness analysis

We aim to calculate the cost of the interventions per extra individual tested.

As we do not have a pure control group, we will focus on the marginal cost and marginal benefits of the two interventions on top of the voucher and gift, which are received by all participants. This means that the cost and benefits of the voucher will not be accounted for. Similarly, testing is the outcome variable and the cost of testing is not included in the cost-benefit calculus.

The cost of the interventions will be calculated ex-post as the sum of:

- The cost of research collaborators recruiting participants (wage + transport cost);
- The cost of sending reminders to people (wage + cost of text message).

5 Work plan

The timeline of the project is as follows

- May-June, 2019: Pilot
- June-July, 2019: Baseline survey and intervention
- July-September, 2019: Appointments and HIV testing at selected clinics
- September, 2019: Finalization and submission of the pre-analysis plan
- October, 2019: Digitization of administrative data (outcome variables)
- November, 2019 to December, 2020: Analysis and dissemination of results

The four researchers will actively contribute to each stage of the research. The lead roles are allocated as follows.

- Data collection setup – Laura Derksen
- Data collection monitoring – Natalia Ordaz Reynoso
- Data processing and analysis, including drafting pre-analysis plan – Olivier Sterck
- Paper drafting – Jason Kerwin

6 References

Abadie, A., Chingos, M. M., & West, M. R. (2018). Endogenous stratification in randomized experiments. *Review of Economics and Statistics*, *100*(4), 567-580.

Abadie, A., Athey, S., Imbens, G. W., & Wooldridge, J. (2017). *When should you adjust standard errors for clustering?* (No. w24003). National Bureau of Economic Research.

Anderson, M. L. (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.

Benjamini, Y., Krieger, A. M., & Yekutieli, D. (2006). Adaptive linear step-up procedures that control the false discovery rate. *Biometrika*, *93*(3), 491-507.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., & Newey, W. (2017). Double/debiased/neyman machine learning of treatment effects. *American Economic Review*, *107*(5), 261-65.

Filmer, D., & Pritchett, L. H. (2001). Estimating wealth effects without expenditure data—or tears: an application to educational enrollments in states of India. *Demography*, *38*(1), 115-132.

Haushofer, J., & Shapiro, J. (2016). The Short-term Impact of Unconditional Cash Transfers to the Poor: Experimental Evidence from Kenya. Quarterly Journal of Economics, 131(4), 1973–2042.

Hirschauer, N., Grüner, S., Mußhoff, O., & Becker, C. (2019). *Twenty steps towards an adequate inferential interpretation of p-values* (No. e3482v4). PeerJ Preprints.

Imbens, G. W., & Rubin, D. B. (2015). *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press.

Kaler, A., & Watkins, S. (2010). Asking God about the date you will die: HIV testing as a zone of uncertainty in rural Malawi. *Demographic research*, *23*(32), 905.

Lakens, D. (2016). Why you don't need to adjust your alpha level for all tests you'll do in your lifetime. Retrieved June 6, 2019, from http://daniellakens.blogspot.com/2016/02/why-you-dont-need-to-adjust-you-alpha.html.

Lee, D. S. (2009). Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *The Review of Economic Studies*, *76*(3), 1071-1102.

Samii, C. (2017). Notes on multiple comparisons and pre-specifying exploratory analyses. Retrieved June 6, 2019, from http://cyrussamii.com/?p=2500/.

Schaner, S. (2018). The persistent power of behavioral change: Long-run impacts of temporary savings subsidies for the poor. *American Economic Journal: Applied Economics*, *10*(3), 67-100.

Vanhove, Jan (2016), On correcting for multiple comparisons: Five scenarios. Retrieved June 11, 2019, from https://janhove.github.io/analysis/2016/04/01/multiple-comparisons-scenarios