Statistical Analysis Plan (The Akiba Trial)

Authors: Lizzie Bair, Mary Putt, Harsha Thirumurthy, Dean Yang

Date: April 7, 2025 (Version 1)

Modifications: --

This document describes the analysis plan for the Akiba study, a two-arm parallel group randomized controlled trial testing the impact of a prize-linked savings intervention on composite HIV/STI incidence among men in western Kenya. This plan was drafted and finalized after baseline data collection was completed and prior to unblinding of the investigators to treatment assignment.

For a complete overview of the study background, setting, experimental design, interventions, outcomes, and data collection process please see the following:

- The protocol paper
    - Egbe, T.I., Omollo, O.D., Wesonga, J.O., Bair, E.F., Chakrabarti, A., Putt, M.E., Celum, C.L., Camlin, C.S., Napierala, S., Agot, K. and Thirumurthy, H., 2022. A savings intervention to reduce men's engagement in HIV risk behaviors: study protocol for a randomized controlled trial. *Trials* 23(1), p.1018.
- Clinical Trials.gov preregistration
    - https://clinicaltrials.gov/study/NCT05385484
- American Economic Association registration
    - https://www.socialscienceregistry.org/trials/9484
- Our IRB approved study protocol is available upon request

I. OUTCOMES

We tested participants for HIV, *Chlamydia trachomatis* (CT), and *Neisseria gonorrhoeae* (NG) at baseline, 12 months, and 24 months. We also tested participants for herpes simplex virus type 2 (HSV-2) at baseline and 24 months. Participants who tested HIV-positive at their baseline or 12 months (or had documentation of being HIV-positive) were not tested at subsequent visits. Participants who tested positive for HSV-2 at their baseline were not tested at 24 months. Participants who tested positive for the curable STIs (CT and NG) were provided treatment by the study, while those tested HIV-positive were referred to government health facilities for treatment.

We measured secondary outcomes in surveys that were conducted every 6 months beginning at the study baseline.

Primary outcome:

- Composite incidence of one or more of HIV and three other STIs (HSV-2, CT, NG) over 24 months. This is a binary variable that will indicate whether a participant acquired HIV, HSV-2, CT, or NG between the baseline visit and the 24-month visit. For participants who are HIV-positive and/or HSV-2-positive at baseline, the outcome will be defined over all STIs other than HIV and/or HSV-2. For the curable STIs (CT and NG), we provided treatment for all participants who tested positive, thus making it possible to define incidence for CT and NG among these participants.

Secondary outcomes:

- Total Savings
    - In total, as well as separately as follows:
        - Formal savings: the sum of savings balances in all formal bank accounts
        - Informal savings: savings held outside of formal institutions, including savings groups and money saved at home or elsewhere
- Physical assets: value of all assets (personal as well as productive)
- Investments
    - Any housing improvements (yes/no)
    - Human capital investments: household spending on education (will adjust for number of children in school)
- Expenditures
    - Food
    - Non-food household expenditures (including personal items; clothing and shoes; household items; utilities; and transport)
    - Alcohol
        - On self only
        - On self and others
    - Transactional sex
    - Gambling

- Sexual risk behavior and alcohol use
    - Number of transactional sex partners in the past 3 months
    - Engagement in any transactional sex in the past 3 months
    - Total number of sexual partners in the past 3 months
    - Condom use at last sex in the past 3 months
    - Number of transactional encounters in the past 3 months
    - Hazardous alcohol use (AUDIT-C ≥4)
- Self-care Behaviors
    - Self-reported engagement in HIV care among participants who are HIV-positive at baseline
- Intimate partner violence
    - Perpetuation of physical, psychological, or sexual IPV
    - Experienced physical, psychological, or sexual IPV
- Future orientation:
    - Assessed by a commonly-used scale that asks participants to make hypothetical choices between money received at two different times (e.g., 1,000 Shillings in 1 month vs. 1,100 Shillings in 2 months)
    - Cantril scale which measures participant's well-being
    - 11-point scale that assesses participant's expectations about their longevity
- Stress (Cohen's Perceived Stress Scale, PSS)
    - Overall PSS score
    - Perceived helplessness score (subscale of PSS)
    - Perceived self-efficacy score (subscale of PSS)
- Perceived financial security
- Food security

II. ANALYTIC APPROACH

We will conduct intention-to-treat analyses using the full sample of participants. Except as noted, hypothesis tests will be two-sided, with a Type I error rate of 0.05 and confidence intervals (CIs) will be 95%.

We will begin by summarizing participants' baseline characteristics, including their demographic characteristics, weekly earnings, savings, and sexual behaviors. These characteristics will be summarized by study group (intervention and control). For the intervention group, we will also report process measures related to the intervention, including proportion of participants who completed the financial training and who opened an account with our partner bank, savings behavior in the bank account, number of lottery prizes awarded, and value of prizes received.

We will report composite HIV/STI incidence per person-year of follow up. For secondary outcomes, we will report the mean (SD), proportion, or median (IQR) for each of outcomes at baseline and each follow-up time point.

III. Analysis of primary outcome

Our primary analyses will determine the effect of the intervention on compositive HIV/STI incidence. The primary outcome will be analyzed using a Poisson model with robust standard errors (Zou, 2004). Controlling for baseline HIV and HSV-2 status, age, marital status and primary occupation, as well as exposure time, we will estimate the following model:

$$E(Y_i) = \exp(\alpha + \beta_1 I(Z_i = 1) + \beta_2 HIV + \boldsymbol{\delta X}_i + \log(Exposure)).$$

Y_i is the primary outcome for participant *i*, $Z_i = 1$ indicates the participant was randomized to the treatment arm, and $\exp(\beta_1)$ is the rate ratio of the primary outcome for the treatment arm relative to the control conditional on baseline HIV status (as noted in our study protocol, randomization to the intervention was stratified on baseline HIV status), covariates $\boldsymbol{X}_i$, and exposure time ($Exposure_i$). The covariates, $\boldsymbol{X}_i$, include baseline HSV-2 status and the following pre-specified baseline characteristics: age, marital status and primary occupation. $Exposure_i$ is the number of days that participant *i* is at risk of contracting all eligible HIV/STIs, and we control for this since there is some variation in the number of the days that elapsed between baseline and the 24-month tests for HIV/STIs. For participants who were positive for CT and/or NG at baseline, exposure time is the number of days between treatment for the baseline infection and the 24-month follow-up. The primary outcome is considered non-missing if participants were tested for HIV/STIs at 24 months, or if they tested positive for CT, NG, or HIV at 12 months.

We will also conduct a secondary analysis of the primary outcome and estimate the effect of the intervention using an ordinary least squares (OLS) model (see section VII). The OLS model will estimate the risk difference in the primary outcome due to the savings intervention.

The primary hypothesis test considers whether the log of the rate ratio of composite HIV/STI incidence at 24 months differs from 0.

$$H_0: \beta_1 = 0$$

$$H_a: \beta_1 \neq 0$$

With only one pre-specified primary hypothesis, we will not adjust for multiple comparisons. The Type I error rate will be 0.05.

IV. Analyses of secondary outcomes

We will conduct additional analyses to estimate the effect of the intervention on secondary outcomes. These outcomes will be analyzed longitudinally using generalized random effects models with a log-link for binary and count outcomes, and an identity link for continuous outcomes.

Conditional on the mean-zero, normally-distributed random effect, (γ_i), the expected value for the outcome at the t^{th} time (t=12, 18, 24) for the interaction model can be written:

$$g\big(E(Y_{it}|\gamma_i)\big) = \alpha + \beta_1 I(Z_i = 1) + \beta_2 HIV_i + \beta_t M_{it} + \beta_{tz} M_{it} I(Z_i = 1) + \boldsymbol{\delta X}_i$$

where $g()$ represents the link function. The time of the outcome (12, 18 or 24 months; month 6 is the reference) is indicated by M_{it}. The three coefficients for the time effect among the control are β_t with treatment by time interaction terms β_{tz} The model includes the stratification variable (baseline HIV status) and $\delta \boldsymbol{X}_i$ represents the product of the adjustment terms and the vector of baseline covariates. Covariates described in the primary outcome analysis as well as the baseline measure of the outcome will be included. When looking at human capital investments, we will also adjust for number of children in school.

The main effects model is:

$$g\big(E(Y_{it}|\gamma_i)\big) = \alpha + \beta_1 I(Z_i = 1) + \beta_2 HIV_i + \beta_t M_{it} + \boldsymbol{\delta X}_i$$

Using a likelihood ratio test, we will test whether

$$H_0: \beta_{tz} = 0 \text{ for all t=\{12, 18, 24\}}$$

$$H_a: \beta_{tz} \neq 0 \text{ for at least one of t =\{12, 18, 24\}}$$

If the global test of the interaction is significant at a Type I error rate of 0.10, we will estimate the treatment effect for each timepoint using a Type I error rate of 0.05 with a Holm-Bonferroni adjustment. If the treatment by time interaction is not significant at the 0.10 level, we will report the overall treatment effect.

The Type I error rate of 0.10 is used for the global test because while we are interested in temporal variation in the treatment effect, the study likely has limited power to detect the interaction compared to the main effect.

Model fit for continuous outcomes will be examined using residuals and the outcome transformed as needed. Extreme values for count outcomes (e.g. number of sexual partners in the past 3 months) will be Winsorized. For continuous outcomes, coefficients represent the mean change in the outcome for intervention participants vs. control, while for binary and count outcomes coefficients will represent log risk ratios.

V. Subgroup analyses

These exploratory analyses will assess the intervention effect on the primary outcome and secondary outcomes including total savings, physical assets, alcohol expenditures, transactional sex expenditures, and each of the sexual risk behavior and alcohol use outcomes for the following subgroups:

1. Risk-seeking preferences at baseline (low vs. medium/high)
2. Positivity for HIV or other STIs at baseline (None versus any as proxy for risk level)
3. Age (Above/below median age at baseline)
4. Socioeconomic status (above/below median total assets at baseline)
5. Marital status (married or living as married vs. not at baseline)

Subgroup analyses will follow the methodology described in Wang et. al. (2007) using a test of the interaction between the treatment effect and the subgroup to guide inference. Forest plots will be created with estimates and confidence intervals for each subgroup. Multiple comparisons will not be formally addressed in order to retain power to detect effects of interest for future studies. However, with five unadjusted hypothesis tests, we acknowledge that the family-wise Type I error rate, the probability of detecting at least one false positive given no effect in any subgroup, is 0.23.

VI. Attrition, non-response, and missing values

Dropout patterns will be described by study group, and baseline characteristics of individuals with missing primary outcome data will be described by study group and compared to those who remained in the study. If missing data rates exceed 10% we will perform multiple imputation using baseline data.

VII. Sensitivity analyses

We will conduct a variety of sensitivity analyses to assess the robustness of our treatment effect estimates on the primary outcome. If the amount of missing data for the primary outcome exceeds 10% in either arm we will perform multiple imputation with fully conditional specification. We note that the maximum likelihood approaches proposed here are valid as long as covariates explaining the missingness are included in the model.

We will first undertake a bounding exercise for the treatment effect by including all participants enrolled in the study in the analytic sample, and in one analysis setting the outcome for those who did not have a primary outcome measurement (those who were not tested at 12 or 24 months, or who tested negative for CT, NG, and HIV at 12 months) to all negative for intervention participants and positive for comparison participants, and in a second analysis setting all intervention participants to positive and comparison participants to negative. analyses.

We will also estimate treatment effects on the primary outcome with an OLS model, which is more commonly used by economists.

$$Y_i = \alpha + \beta_1 I(Z_i = 1) + \beta_2 HIV_i + \boldsymbol{\delta X}_i + \beta_3 Exposure_i + \varepsilon_i$$

Finally, we will do a sensitivity analysis including self-reported STIs that participants indicated were treated in the 6-monthly surveys as incident cases to assess the robustness of our results to potential undermeasurement.

References

Wang, R., Lagakos, S. W., Ware, J. H., Hunter, D. J., & Drazen, J. M. (2007). Statistics in medicine—reporting of subgroup analyses in clinical trials. *New England Journal of Medicine*, *357*(21), 2189-2194.

Zou, G. (2004). A Modified Poisson Regression Approach to Prospective Studies with Binary Data. *American Journal of Epidemiology*, *159*(7), 702–706. https://doi.org/10.1093/aje/kwh090