

STATISTICAL ANALYSIS PLAN

**Feasibility and Acceptability of a Barbershop Based HIV Prevention Initiative Among
Heterosexual Men in Kalangala Islands, Uganda: A Cluster Randomized Trial**

HPTN 111 (Protocol Version 2.0)

Date SAP finalized for signature: 22 July 2025

Effective Date: Date of last signature

SAP Version: 4.0

STATISTICAL ANALYSIS PLAN

Protocol Name:	Feasibility and Acceptability of a Barbershop Based HIV Prevention Initiative Among Heterosexual Men in Kalangala Islands, Uganda: A Cluster Randomized Trial
Protocol Number:	<i>HPTN 111 (v2.0)</i>
SAP Author(s):	Emily Voldal (Staff Scientist); Devang Agravat (Statistical Research Associate II)
SAP Version:	4.0

Author(s):

Lead Unblinded Protocol Statistician	
Legal Name	Emily Voldal, PhD
Job Title	Staff Scientist
Signature & Date	See eTMF Signature Manifest

Lead Unblinded Statistical Research Associate	
Legal Name	Devang Agravat, MPH
Job Title	Statistical Research Associate II
Signature & Date	See eTMF Signature Manifest

Approval(s):



Protocol Chair	
Legal Name	Zubair Lukyamuzi, MBChB, MPH, Msc
Job Title	Protocol Chair
Signature & Date	DocuSigned by: <i>Zubair Lukyamuzi</i> 75B18C0DDAAF497... 8/1/2025

Unblinded Statistician	
Legal Name	Deborah Donnell, PhD
Job Title	Professor
Signature & Date	See eTMF Signature Manifest

The effective date of this document is the date of the latest signature.

TABLE OF CONTENTS

1.

INTRODUCTION

5

1.1

GENERAL DESIGN CONSIDERATIONS

6

1.2

STUDY OBJECTIVES AND ENDPOINTS

7

1.3

RANDOMIZATION

8

1.4

BLINDING

8

1.5

SAMPLE SIZE AND POWER

8

1.6

ANALYSIS SET(S)

9

1.7

STATISTICAL ANALYSIS ISSUES

10

2.

INTERIM ANALYSIS AND DATA MONITORING COMMITTEE

12

3.

GENERAL ANALYSIS METHODS

13

4.

TRIAL PARTICIPANT DISPOSITION

13

4.1

DISPOSITION OF PARTICIPANTS

13

4.2

PROTOCOL DEVIATIONS

14

5.

BASELINE DATA

14

6.

FEASIBILITY/ACCEPTABILITY ANALYSES (PRIMARY)

15

6.1

PARTICIPANT SURVEY RESPONSES

15

6.2

BARBER SURVEY RESPONSES

21

6.3

RECRUITMENT AND RETENTION

24

6.4

BARBER-PARTICIPANT INTERACTIONS AND INTERVENTION DELIVERY

26

7.

SECONDARY ENDPOINTS

29

8.

EXPLORATORY ENDPOINTS

35

9.

DEVIATIONS FROM PROTOCOL

35

10.

REFERENCES

35

11.

CHANGE HISTORY

35

LIST OF ABBREVIATIONS AND ACRONYMS

Term/Abbreviation	Definition
ACASI	Audio-Computer Assisted Self-Interview
CI	Confidence Interval
HIV	Human Immunodeficiency Virus
HPTN	HIV Prevention Trials Network
IQR	Interquartile range; for IQR, we will always present the 25 th and 75 th percentiles
ITT	Intention To Treat
Participant	Enrolled men who are clients of the participating barbershops; although barbers also participate in study procedures and surveys, they are referred to throughout as “barbers”
PrEP/PEP	Pre-Exposure Prophylaxis/Post-Exposure Prophylaxis
SAP	Statistical Analysis Plan
Shop	Barbershop
SMC	Study Monitoring Committee
SSP	Study Specific Procedures manual
STI	Sexually Transmitted Infection
Study Clinic/Study Site	Satellite study site co-located with an existing health facility in Kalangala (one site for the whole study)
Village	Town or village
VMMC	Voluntary Medical Male Circumcision

1. INTRODUCTION

This statistical analysis plan (SAP) details the statistical procedures to address the primary and secondary objectives specified in the Protocol Version 2.0 for HPTN 111. New versions of the SAP will be issued to document updates and changes in the plan. Analyses related to exploratory objectives and other secondary analyses not anticipated prior to study completion will be described in separate documents.

Formal interim monitoring will not be implemented in the study due to the short enrollment window and limited follow-up period and considering the minimal risk of the intervention to participants. The Study Monitoring Committee (SMC) will review summary information related to safety, study conduct, and preliminary endpoints, but those reports will not include any formal statistical analyses.

In addition to SMC reports, routine reports will be produced for the study operations team which include operational metrics including information about recruitment, retention, and delivery of the intervention (e.g. how often barbers are holding group sessions and counseling clients). Information from these reports will be given back to the site if any intervention to operations is necessary. No formal analyses will be included in these reports, and these reports are not described any further in this SAP.

1.1 General Design Considerations

Purpose: To test community-based approaches to engage heterosexual men at risk for HIV and specifically to assess the feasibility and acceptability of a barbershop based HIV prevention program.

Intervention: Participants recruited from intervention barbershops will receive the barbershop-based HIV prevention initiative. The trained barber will provide the following services during regular haircut services: general, status-neutral HIV education, HIV self-test kits, and information about where to receive HIV prevention services. The barber will also lead peer support group education every two months for clients enrolled in the study. Participants recruited from control barbershops will receive standard-of-care HIV prevention services which include facility-based HIV risk reduction counseling and testing and providing information about facility distributed HIV self-test kits.

Study duration: Approximately 24 months for study start-up, participant recruitment, and follow-up, with individual participants followed for 12 months.

Summary of design and considerations for analyses:

HPTN 111 is a cluster-randomized trial studying the feasibility and acceptability of a barbershop-based HIV prevention program as a community-based approach to engage heterosexual men at risk for HIV. 18 barbershops in the Kalangala district, Uganda were selected to participate, and randomized 2:1 to the barbershop-based HIV prevention initiative (intervention) or standard-of-care. Barbers in both arms recruit clients to join the study. Barbers in the intervention arm also deliver the HIV prevention program to their clients, including general, status-neutral HIV education and information about where to receive HIV prevention services, HIV self-test kits, and peer support groups. Standard-of-care HIV prevention services include facility-based HIV risk reduction counseling and testing, and HIV self-test kits distributed from existing facilities.

Up to 250 participants will be included in this study. Participants must be heterosexual men behaviorally vulnerable to HIV, at least 16 years old, and HIV-negative at enrollment. All participants will complete study clinic visits which include bio-behavioral data collection, HIV counseling and testing, and STI testing. All participants will have a study clinic visit at baseline and 6 months after enrollment (Week 26 visit). The original protocol specified that contingent on when participants were enrolled and funding for additional study time, some participants may have an additional study clinic visit at 12 months after enrollment (Week 52 visit). This has been updated via a letter of amendment so that all men may have a Week 52 visit.

Because this is a feasibility/acceptability study, the protocol allows for flexibility in case of issues with recruited shops or barbers. 18 shops have been selected and included in the randomization. Every effort will be made to retain those shops throughout the study. However, if a shop withdraws or is asked to withdraw by the study team (e.g. due to failure to enroll clients), the protocol allows for that shop to be replaced. Similarly, although we hope to retain individual barbers for the duration of the trial, there may be barbers who leave or join the study part-way through because their participation in the study is linked to their employment with a specific shop. It is technically possible that a barber could be employed at two different participating shops over the duration of the trial. We hope that selecting only one shop per town/village helps prevent this. Similarly, a participant could move from one shop to another. However, the participating shops are separated by a 5 km radius; therefore, we expect that it would be unlikely for a client to move to another participating shop in a different village. However, it may be more common for clients to move to a non-participating barbershop. In the absence of issues with switching shops, generally we believe that both barbers and clients are nested within barbershops. However, clients are not necessarily nested within barbers, as some shops may have multiple barbers and clients may not visit the same barber every time they get a haircut.

Each shop will be asked to recruit a minimum of 10 clients, stopping or pausing enrollment at specific shops as necessary to restrict the total sample size to 250 or fewer.

1.2 Study Objectives and Endpoints

Primary objective (feasibility/acceptability):

To evaluate the feasibility and acceptability of a barbershop-based HIV prevention initiative.

Primary endpoints (measured primarily in intervention arm):

- Participant responses from surveys at Week 26 and Week 52 about feasibility and acceptability of the intervention
- Barber responses from surveys about feasibility and acceptability of the intervention
- Recruitment rates during enrollment and retention in study and barbershop activities at Week 26 and Week 52
- Frequency of barber-participant interactions and proportion that include delivery of the intervention

Secondary objectives (effectiveness):

1. To compare completion of self-initiated HIV testing between intervention and control groups

Endpoint:

- Self-report of HIV test conducted between enrollment and study visit at Week 26, and between the Week 26 and Week 52 study visits

2. To evaluate the preliminary effectiveness of the intervention on change in behaviors associated with HIV acquisition

Endpoint:

- Self-reported risk behaviors at Week 26 and Week 52, including condom use and number of total and new sexual partners

3. To compare interest in or use of HIV prevention services between intervention and control groups

Endpoints:

- VMMC during the study
- Self-reported referral to PEP care or use of PEP
- Self-reported referral to PrEP care or use of any PrEP method during the study

Additional secondary objective:

To assess interest in long-acting PrEP among all participants by study arm

Endpoint:

Self-reported interest in long-acting PrEP at Week 52

Exploratory objective (effectiveness):

To evaluate the preliminary effectiveness of the intervention on incident STIs.

Endpoints:

- Incidence of HIV between screening and Week 52
- Incidence of gonorrhea and chlamydia between screening/enrollment and Week 52

1.3 Randomization

Barbershops were randomized 2:1 into the two arms, so that participants in 12 shops received the intervention and participants in 6 shops received standard-of-care. Covariate-constrained randomization was used so that shops in the two arms were balanced on village population, number of clients per day, and distance from the barbershop to the nearest health facility.

To accommodate a public randomization event (see SSP for details) which was transparent and understandable to the general public, the randomization was executed in two stages. In the first stage, covariate-constrained randomization was used to create three groups of six shops each. In the second stage, simple random assignment was used to assign each of the three groups to a treatment arm (two to intervention, one to control).

For the covariate restrictions in the first stage of randomization, we restricted the absolute difference in means between treatment arms. For each possible allocation of shops into three groups, there are three ways that the final 2:1 treatment arms could be assigned. We calculated the difference in means between treatment arms for each possible treatment assignment, and only allowed allocations to groups of three for which all possible treatment assignments met the restriction. Details on data, constraints, and numbers of allocations are in Section 4 of the SSP.

1.4 Blinding

There is no formal blinding in this trial. Barbers know which arm their shop has been assigned to before beginning recruitment. While they are recruiting clients for the trial, barbers are asked to not reveal which arm their shop belongs to. After enrollment, participants are notified of their barbershop's treatment assignment. Because of the public randomization event and unblinded barbers and participants, study sponsors, staff, and members of the community are likely to become aware of barbershop treatment assignments while the study is ongoing. To the extent possible, however, decisions regarding additions or modifications to planned analyses, determination of inclusion in or exclusion from analysis cohorts or datasets, and rules for handling missing or incomplete data will be made by statisticians, study leadership, or other study staff who do not have access to outcome data grouped by arm and/or without information about the treatment assignment of the case in question.

1.5 Sample Size and Power

Since the primary objective pertains only to the intervention arm, we decided to power the study based on the important secondary objective of self-initiated HIV testing, which is a binary participant-level variable measured between enrollment and Week 26 (or equivalently for power, between Week 26 and Week 52). Power calculations accounted for clustering of participants within shops and 2:1 randomization of shops. We assume there are 18 barbershops and allow the number of participants enrolled per shop to vary between 10 and 14, but assume that all shops enrolled the same number of participants. The assumed

between-cluster coefficient of variation of the true proportions are the same in each arm and equal to 0.25. If retention targets are met, loss to follow-up will be 10% for responses to questions about HIV testing on the 26- and 52-week questionnaires. However, we also consider a scenario with 20% loss to follow-up. We are interested in detecting an absolute difference in HIV testing rates between arms of 20%. We consider several possible baseline testing rates, and calculate power for 5% vs. 25%, 10% vs. 30%, and 15% vs. 35% of participants with HIV testing in the standard-of-care vs. intervention arms. Power calculations did not account for the covariate constraints used in the randomization.

To calculate power, we use the methods described in Hayes and Moulton (2017), with slight modification of Equation 7.6 to account for the trial design. To relate the power to the number of clusters, we use:

$$c_0 = 1 + (z_{\alpha/2} + z_{\beta})^2 \frac{\pi_0(1 - \pi_0) \frac{1}{m} + \pi_1(1 - \pi_1) \frac{1}{mn} + k_0^2 \pi_0^2 + k_1^2 \pi_1^2 \frac{1}{n}}{(\pi_0 - \pi_1)^2}$$

with the following definitions and assumptions:

- c_0 is the number of shops in the control arm, so the number of shops in the intervention arm is $2^* c_0$
- n is the allocation ratio, so $n=2$ for the 2:1 randomization
- m is the number of participants per shop
- k_0 and k_1 are the between-cluster coefficients of variation of the true proportions in the two arms. We assumed that $k_0=k_1=0.25$. According to Hayes and Moulton (2017), k in cluster-randomized trials is often less than 0.25, so the choice of $k=0.25$ may be conservative.
- π_0 and π_1 are the true proportion of participants who have self-initiated HIV testing in the two arms
- $z_{\alpha/2}$ and z_{β} are upper tail probabilities from a Normal distribution, for size $\alpha=0.05$ and power $100(1-\beta)$

See the protocol for a full table of power in all possible scenarios (Table 3 in protocol Version 1.0). If loss to follow-up is 10% and 10% or fewer of the participants in the control arm test for HIV, we have at least 80% power to detect an absolute difference of 20% if shops meet the minimum enrollment target (10 men per shop). If instead 15% of participants in the control arm test for HIV, power will be less than 80% if 10 or 12 participants are enrolled per shop, but 82% if 14 participants are enrolled per shop. If loss to follow-up is 20%, most considered scenarios still have close to 80% or higher power. The only considered scenario with less than 70% power is with 20% loss to follow-up, 10 enrolled participants per shop, and 15% of men in the control arm testing for HIV.

GENERAL DATA ANALYSIS CONSIDERATIONS

1.6 Analysis Set(s)

When summarizing information, there are four levels of information to consider: villages, shops, barbers, and participants. Villages are the highest level of clustering and the unit of randomization. At the time of randomization, each village contained one barbershop. Villages may have multiple barbershops in cases where a shop had to be replaced. If no replacement shop can be found within the original village, a shop may be recruited from a new area in the islands. In this case, the new shop will be considered part of the “village” of the shop it is replacing in analyses, even though they do not share the same physical village. Replacement shops will belong to the same treatment arm as the shop they are replacing. Barbers are associated with specific shops. At the time of randomization, all shops had only one enrolled barber. It is unlikely but possible that a barber could move from one study shop to another. When performing

analyses that account for clustering, a barber will be considered to belong to the shop at which they first enrolled with the study, even if they later switched shops. Participants may also switch shops. It is unlikely but possible that a participant could move from a shop to a shop in a different village. If a shop needs to be replaced, it is possible that participants may move from one shop to another within a village. For shop-level summaries, participants will belong to the shop at which they first enrolled. For village-level summaries, participants will belong to the village at which they first enrolled, which will account for cases where participants move to a replacement shop within the same village.

ITT cohorts: Throughout, primary analyses are intention to treat (ITT). For villages, ITT means as-randomized. For shops, ITT means the arm assignment of their village. For barbers and participants, ITT means the arm assignment of the shop they first enrolled with. For participants, the ITT population will exclude those who were subsequently found to be inappropriately enrolled due to violations of the inclusion/exclusion criteria that were deemed to warrant exclusion by the protocol committee. For barbers, the ITT population will include all enrolled barbers.

Participants not living with HIV: participants who are part of the ITT population but not identified as living with HIV prior to the time of the outcome measurement. HIV status is determined by site testing at study visits at baseline, Week 26, Week 52, and as needed at interim visits. The site reports the participant's HIV status after testing as "HIV-negative", "HIV-positive", or "additional testing needed". Cases marked "additional testing needed" will be resolved by re-testing at a later visit, or by adjudication if the site cannot resolve the HIV status. Participants who were HIV-positive at baseline would be excluded for assessments of the "not living with HIV" cohort at Week 26 and Week 52 (and would have been inappropriately enrolled). Participants who were HIV-positive at Week 26 (based on HIV testing at or prior to Week 26) would be excluded for assessments of the "not living with HIV" cohort at Week 52.

All participants in ITT population from shops randomized to the intervention arm: this cohort will be used for the analysis of feasibility/acceptability questions that apply only to participants who received the intervention. This will include all participants who enrolled at a shop that was randomized to the intervention arm, regardless of whether they continued attending the shop they enrolled at. This cohort may include some participants who never actually received all components of the intervention, which could have been due to participants switching barbershops or barbers not delivering the intervention services. Note that participants who enrolled at a shop that was randomized to the control arm will be excluded from this cohort, regardless of whether they later switched shops or reported receiving HIV prevention services from their barber.

Enrolled barbers from the intervention arm: this cohort will be used for the analysis of feasibility/acceptability questions that apply only to barbers who delivered the intervention. This will include all barbers who originally enrolled in the study at a shop that was randomized to the intervention arm, regardless of whether they continued working at that shop for the duration of the study. This cohort may include some barbers who never actually delivered all components of the intervention, which could have been due to barbers switching shops or a shop withdrawing early from the study. Note that barbers who enrolled at a shop that was randomized to the control arm will be excluded from this cohort, regardless of whether they later switched shops or reported delivering HIV prevention services.

VMMC cohort: participants in the "not living with HIV" population who reported that they were not circumcised at enrollment and have follow-up data on circumcision status. Participants who did not know their circumcision status at enrollment (or who did not answer the circumcision questions) will also be included, as long as they did not later report a pre-study circumcision.

1.7 Statistical Analysis Issues

Issues related to covariates

For any model-based comparison of treatment arms, we will adjust for the covariates used in the randomization. Note that the covariates used in randomization are measured at the shop level, although some are more related to village characteristics. For analysis methods that require aggregation to the village level, when there are multiple shops within a village we will use the covariate values from the shop that was followed the longest (i.e. the shop that contributed the most follow-up time from that village). Note that in a scenario where the randomized shop was replaced early in the study, the values of the covariates used in adjustment may not match the values used in the restricted randomization procedure.

Because of replacement shops and the recruitment of participants by barbers after randomization, it is possible that there could be some systematic imbalance between arms in terms of participant or barber or shop baseline characteristics. Baseline characteristics will be described by arm. If there are any baseline covariates that may affect the outcomes which are meaningfully imbalanced between arms, we will perform exploratory analyses that adjust the planned arm comparisons for the imbalanced covariates.

Issues related to clustered data

As described above, information in this trial is measured at four different levels: participants, barbers, barbershops, and villages. Because the number of participants per shop may vary, there may be substantial differences between participant-level summaries (weighting participants equally) and shop- or village-level summaries (weighting shops or villages equally, regardless of how many participants they enrolled). For example, a participant-level summary for the average of some measure y_{ij} which is an outcome from participant j in barbershop i would be $\bar{y} = \frac{1}{N} \sum_{ij} y_{ij}$ where N is the total number of participants across all barbershops. To calculate the average using a village-level summary, we would instead use $\bar{y} = \frac{1}{18} \sum_i \frac{1}{N_i} \sum_j y_{ij}$, where N_i is the number of participants in village i . Note that in addition to this clustering structure, some outcomes are measured at each visit, so individual participants and barbers may have repeated measures over time.

For data measured at the participant visit level (e.g. participant responses to surveys), we will take the following general approach. Answers from surveys will be summarized in detail at the visit level, or at the participant level for meaningful derived participant-level outcomes. For some details (e.g. missingness patterns and skip logic) it is most understandable to summarize data at the visit level, but we will stratify by visit (e.g. Week 26 and Week 52 for participants) whenever feasible for interpretability. We will identify simple (binary or continuous) outcomes that can be extracted from surveys, either by selecting particular questions or by deriving a scientifically meaningful outcome. For those simple outcomes only, we will summarize data at the village level, compute confidence intervals, and conduct any statistical analyses accounting for clustering at the village level. For interpretation of results and for planning future studies and interventions, it is also important to understand how much the outcome varies between villages; to accommodate this, we may accompany village-level analyses with descriptive tables or figures that show the outcome for each village and/or calculate statistics summarizing the patterns of clustering (e.g. the intra-cluster correlation coefficient) when applicable.

An analogous process will be considered for barber-level data, however because there is less potential variability in the number of barbers per village, the difference in weighting may be trivial. Formal analyses will still be done accounting for clustering at the village level but may not benefit from additional descriptive statistics on the clustering.

Issues related to missing data

Unless otherwise specified, all analyses are complete case. Formal statistical comparisons are made at the village level, but missing data at the participant or barber level could still bias results. If loss to follow-up is >10% different between arms at the level of data collection for any outcome with a formal comparison between arms, sensitivity analyses will be performed using multiple imputation. For outcomes that are measured in the intervention arm only, 95% confidence intervals will be re-calculated

using multiple imputation as a sensitivity analysis if the outcome is missing >25% at the level of data collection.

Other issues

We have not planned any correction for multiple comparisons. Because the primary outcome is feasibility/acceptability, the primary outcomes are mainly measured in the intervention arm only, and so there are no formal comparisons between arms that directly address the primary objective. The comparison between arms which is of most interest is the comparison of self-reported HIV testing rates, a secondary objective. This is the comparison which the study was powered for. To reflect this, we will present both 95% confidence intervals and p-values for the comparison of HIV testing rates. For all other comparisons, we will focus on descriptive statistics and compute 95% confidence intervals (but not report p-values).

All data will be tabulated per the evaluation visit as recorded on the form even if the assessment is outside of the visit window.

2. INTERIM ANALYSIS AND DATA MONITORING COMMITTEE

General strategy

No formal interim analyses or stopping rules are in place, due to the short enrollment window and limited follow-up period and considering the minimal risk of the intervention to participants. The Study Monitoring Committee will review summary information related to safety, study conduct, and preliminary endpoints, but those reports will not include any formal statistical analyses. Because assessments of feasibility/acceptability are complex and multi-faceted, we have not pre-specified any targets or stopping rules other than those defined in the protocol. Per the protocol, each shop is asked to recruit at least 10 clients to enroll in the study, and the target for retention is 90% at the Week 26 and Week 52 visits. Information presented to the SMC may include the following when applicable:

- Participant screening and reasons for ineligibility
- Summaries of enrollment, including number of participants enrolled per shop, progress of enrollment over time, and dates of activation and first and last enrollment
- Summaries of baseline characteristics of participants, barbers, and barbershops (see Section 5 for details), including presentations by arm to examine balance
- Visit completion for participants and barbers
- Study retention for participants, barbers, and barbershops
- Reported protocol deviations
- Barber engagement, including self-reported information about number of group sessions held, HIV self-test kits distributed, and delivery of information to clients
- Reported social impacts
- Endpoint monitoring, including information about recruitment and study activities conducted by barbers, survey questions on feasibility/acceptability, and participant-reported secondary endpoints

Because the treatment arms are different sizes, there will be no masked results reported to the SMC. Data will be presented either without information about the treatment assignment, or unblinded. Information will be presented by shop and/or by arm when it is necessary for additional insight into study conduct. Because of its impact on power, information about number of participants recruited, enrolled,

and retained will be presented by barbershop and/or village. Other than baseline characteristics, information presented by arm will include: recruitment and accrual tables, participant disposition and retention, and select endpoints in a closed report.

Recommendations from the SMC

In January 2025, the SMC reviewed study results, including participant outcomes through Week 26. Based on these results, the SMC recommended that 26-week results should be made public, as the results may impact the design of studies in development and would have no impact on the remaining follow-up of the ongoing HPTN 111 trial. After consulting with HPTN leadership, the study team decided to perform an interim analysis including Week 26 participant outcomes so that the results can be presented on an expedited timeline. This interim analysis included a sub-set of the analyses described below, including information about barber and participant recruitment and retention; analysis of primary outcomes such as feasibility/acceptability information from barber and participant surveys; and analysis of secondary outcomes such as HIV testing. For analyses we follow the principles outlined below, truncating outcomes to examine only the relevant time period or visits. This interim analysis is not described any further within this document, and all originally planned analyses will still be executed at the end of the study. Because our primary outcomes related to feasibility/acceptability are measured in the intervention arm only (with no comparison group), we will not make any multiple testing corrections to the final study analyses to account for this interim analysis. Note that there is no early stopping of follow-up, so results of this interim analysis will not affect the final results.

3. GENERAL ANALYSIS METHODS

All statistical tests and comparisons are two-sided. All confidence intervals are 95% confidence intervals, and all statistical tests have alpha level 0.05 unless otherwise specified. When appropriate, we will use exact methods for confidence intervals.

Summary statistics for categorical variables will include counts and percentages. For continuous variables, mean and standard deviation or median and inter-quartile range (25th percentile and 75th percentile) will be reported, along with the minimum, maximum, and number of missing observations. Whenever counts of events are presented by arm, care will be taken to emphasize that the treatment arms are different sizes and avoid directly contrasting counts.

4. TRIAL PARTICIPANT DISPOSITION

4.1 Disposition of Participants

We will describe how participants moved through study procedures using a table or diagram which includes the following measures within each study arm: number of participants screened, number screened out (and reasons for exclusion), number enrolled, number retained at Week 26, number retained at Week 52, and reasons for discontinuation for participants who terminated early from the study. If any participants switched shops or left their enrollment shop during the study, we will describe the frequency and patterns of movement.

We will describe how barbers moved through study procedures using a table or diagram which includes the following measures (by arm, where applicable): number of barbers trained, number of barbers randomized, number of barbers who joined the study after the randomization event, total number of barbers enrolled, and the number of barbers present at each quarterly visit with reasons for discontinuation for barbers who terminated early from the study. For control barbers, there is just one scheduled visit during participant enrollment, aligned with the first quarterly visit for intervention barbers;

for intervention barbers, quarterly visits continue for the duration of the study. If any barbers switched shops or left their enrollment shop during the study, we will describe the frequency and patterns of movement. We will also describe the enrollment of additional barbers throughout the study.

We will describe in a narrative or timeline how barbershops moved through study procedures, including information about when any shops were terminated early from the study and why, and when replacement shops joined the study.

4.2 Protocol Deviations

We will summarize any inappropriate enrollments, including whether any inappropriate enrollments resulted in exclusion from the ITT population. We will summarize any protocol deviations that are relevant to the assessment of feasibility/acceptability. We have not planned any per protocol analyses, so participants will be included in analyses based on their randomized treatment assignment regardless of protocol deviations. However, we will summarize self-reported information about what intervention components participants in both arms received. If there are protocol deviations that affect entire shops (e.g. a barber from the control arm was trained to deliver the intervention), we will consider addressing these in sensitivity analyses on a case-by-case basis.

5. BASELINE DATA

We are not planning to perform any formal statistical comparisons on baseline data. We will summarize baseline data for all levels of clustering. For villages and shops, we will summarize baseline data by arm. For all shops that were present at the randomization event, this data was collected a number of months before the study initiated. For replacement shops, data may have been collected at the time of their enrollment. We will also summarize barber information measured at the time of enrollment by arm. This will include information about their business that may affect delivery of the intervention. We will also summarize participant information measured at enrollment by arm. This will include information that may affect outcomes (e.g. history of HIV testing), and information that may affect the delivery of the intervention (e.g. frequency of haircuts). We may also describe answers to survey questions about the barber-client relationship, such as how strongly participants agree with the statement “I regularly discuss personal matters with my barber” and how strongly barbers agree with the statement “I regularly discuss personal matters with my clients”.

For participants, information collected at baseline which may be summarized by arm includes:

- Demographics, including age, relationship status, and education
- Information about barbershop attendance, including how frequently the participant gets a haircut and how long the participant has attended the barbershop they were recruited from
- Alcohol and drug use
- Sexual behavior, including HIV status of the participant's main partner (if applicable), information about engagement in different sex acts and transactional sex, and use of PrEP/PEP
- STI test results for gonorrhea, chlamydia, and syphilis
- HIV prevention services use, including when and where participants get tested for HIV and knowledge about VMMC and PrEP

To better understand differences between barbershops, some participant information may also be presented by barbershop. This could include demographics such as age and education, and other baseline characteristics that may play a role in the feasibility/acceptability or effectiveness of the intervention, such as frequency of haircuts or history of HIV testing.

6. FEASIBILITY/ACCEPTABILITY ANALYSES (PRIMARY)

Objective: To evaluate the feasibility and acceptability of a barbershop-based HIV prevention initiative

6.1 General feasibility/acceptability survey analysis methods

Estimands

Both barbers and participants answer a large number of questions relevant to feasibility and acceptability. In this section, we describe the estimand framework for these feasibility and acceptability surveys and then provide details about how surveys will be analyzed for both feasibility and acceptability, and both participants and barbers. For all these endpoints, the general objective is to evaluate the feasibility and acceptability of the barbershop-based HIV prevention intervention. The intervention has three main components: HIV education and counseling, HIV self-test kits, and barber-led peer group sessions. Participants can choose which components to engage with and how frequently they visit the barber. Although the intervention was designed with the idea that people will receive all three components, depending on their availability and preference participants randomized to the intervention arm may choose not to receive some or all components. In addition to choosing which components to engage with, participants also decide how frequently they want to receive education and counseling, self-test kits, and attend peer group sessions. Summarizing the feasibility/acceptability across all participants in the intervention arm will reflect an average intervention effect across people who have varying levels of engagement with the intervention.

The primary estimands for participants are the average feasibility/acceptability scores for a typical client from a typical shop delivering barbershop based HIV prevention services. To fully define these estimands, we also describe how intercurrent events will be handled. Intercurrent events are things that happen after randomization which could affect how we interpret an outcome (e.g. participant moved to a different city and stopped visiting the study barber) or whether we can measure an outcome (e.g. death of a participant). We apply below two common strategies for intercurrent events. First, the treatment policy strategy is used to include data in the analyses even though it may have been impacted by an intercurrent event, similar to an intent-to-treat analysis strategy. Second, the composite variable strategy incorporates the intercurrent event into the outcome definition explicitly so that the outcome is still well-defined if the intercurrent event affects whether we can measure something. The participant estimands are described by the following attributes:

1. **Population:** The target population is men who identify as a heterosexual male and are at least 16 years old, behaviorally vulnerable to HIV, and HIV-negative. They must also be a regular customer at a barbershop which has a customer base of at least 15 clients per week, has existed as a shop for at least 6 months, and has adequate space to deliver the intervention. The analysis set will be all participants enrolled at shops randomized to the intervention arm.
2. **Endpoint:** Survey responses about feasibility and acceptability (answered at Week 26 and Week 52), which are converted into three separate scores and averaged across visits. See below for details on how scores are calculated from surveys.
3. **Treatment of interest:** implementation of the barbershop-based HIV prevention intervention at the participant's usual barbershop. This is based on the randomized treatment assignment of the barbershop at which the participant enrolled, regardless of what intervention services participants report receiving.
4. **Handling of intercurrent events:**
 - a. **Discontinuation of study treatment for any reason:** this could include participants who decline components of the intervention, participants who move away from the island, and participants who can't access the intervention due to their original shop withdrawing from the study. We will use a treatment policy strategy to include these participants whenever

- data is available from any of their study visits. We will aim to continue to collect follow-up information at Week 26 and Week 52, regardless of discontinuation of treatment.
- b. No participation in group sessions: the feasibility/acceptability scores include questions about group sessions which are only answered by participants who attended a group session. In order to calculate scores for these participants to execute the treatment policy strategy described above, we will assign a neutral answer to group session questions for these participants when calculating an overall score (composite variable strategy).
 - c. Shop switching: this intervention is designed to leverage an existing barber-client relationship, so if a participant is accessing the intervention at a shop other than the one they enrolled at (e.g. in the case of a shop closure and replacement), the impact of the intervention may be different. We will use a treatment policy strategy to include these participants.
 - d. Death: we will use a treatment policy strategy to include these participants whenever data is available from any of their study visits.
 - e. Acquisition of HIV: although men can only enter the study if they test negative for HIV, there may be participants living with HIV at the time of follow-up. We will use a treatment policy strategy to include these participants.
5. Summary measure: Average feasibility and acceptability scores, weighting villages equally and weighting participants equally within villages. See below for details on how scores will be summarized.

The primary estimands for barbers are the average feasibility/acceptability scores for a typical barber from a typical shop delivering barbershop based HIV prevention services. These barber estimands are described by the following attributes:

1. Population: The target population is barbers in Kalangala Island with a primary level of education who work at a shop that has adequate space to deliver the intervention, has existed for at least six months, and has a customer base of at least 15 clients per week. The analysis set will be all barbers enrolled at shops randomized to the intervention arm.
2. Endpoint: Survey responses about feasibility and acceptability (answered at quarterly visits), which are converted into two separate scores and averaged across visits. See below for details on how scores are calculated from surveys.
3. Treatment of interest: implementation of the barbershop-based HIV prevention intervention at the barber's barbershop. This is based on the randomized treatment assignment of the barbershop at which the barber enrolled, regardless of what intervention services barbers report delivering.
4. Handling of intercurrent events:
 - a. Discontinuation of study intervention delivery for any reason: this could include barbers who lose their job or move away, and barbers who can't deliver the intervention due to issues out of their control (e.g. unavailability of self-test kits, or participants declining the intervention). We will use a treatment policy strategy to include these barbers whenever data is available from any of their study visits. We will aim to continue to collect follow-up information at quarterly visits as long as the barber might have new insights on experience with the intervention. If a barber is leaving their shop or moving away from the island, we will aim to have at least one quarterly visit after they have finished providing HIV-related services to clients.
 - b. No participation in group sessions: the feasibility/acceptability scores above include questions about group sessions which are only answered by barbers who led a group session. In order to calculate scores for these barbers to execute the treatment policy strategy described above, we will assign a neutral answer to group session questions for these barbers (composite variable strategy).
 - c. Death: we will use a treatment policy strategy to include these barbers whenever data is available from any of their study visits.

5. Summary measure: Average feasibility and acceptability scores, weighting villages equally and weighting barbers equally within villages. See below for details on how scores will be summarized.

Note that for secondary endpoints a similar estimand approach will be used, with specific information about endpoints and summary measures (including comparisons between groups when relevant) described in their respective analysis sections below.

Calculation of scores

Multiple endpoints described below involve scores to combine a group of related questions about feasibility or acceptability. Questions that inform these scores comprise of a statement or question, with a five-level Likert scale response. For each question answers will be scored from -2 to 2 so that the answer that is least supportive of the intervention will be given a score of -2, and the answer that is most supportive of the intervention will be given a score of 2. To combine multiple questions into a score for a particular person at a particular time point, we will average the responses to all the questions. As a result, all scores will have a possible range from -2 (least supportive) to 2 (most supportive). All computed scores will be based on at least three questions; for domains of interest with only one or two questions, the questions will be analyzed separately. Note that the scores are computed for each individual (barber or participant) at each visit. To get an overall score for each individual, we will average scores across all visits. To get an overall score for each village or shop, we will average scores across all individuals within that village or shop. For example, suppose y_{ijk} is a feasibility or acceptability score from participant j in village i at time k . We are interested in the average score $\bar{y} = \frac{1}{12} \sum_i \frac{1}{N_i} \sum_j \frac{1}{N_j} \sum_k y_{ijk}$, which first takes the average over the $k = 1, \dots, N_j$ visits for participant j , then takes the average over the $j = 1, \dots, N_i$ participants within village i , then takes the average over the $i = 1, \dots, 12$ villages randomized to the intervention. This is a village-level summary of scores across the duration of the trial. Note that the number of visits may differ between individuals, but this weights individuals equally within village regardless. Similarly, the number of individuals may differ between villages, but this weights villages equally regardless. If there were one shop per village, this would be equivalent to equally weighting shops. However in the case of replacement shops within villages, this method weights shops within villages by the number of participants, so that shops which enrolled or serviced more customers within a single village have more influence on the village-level average.

General analysis approach

To report results from a score, we will do the following. We will present summaries of the score (mean, median, IQR, range, etc.), which has possible values between -2 (least supportive) to 2 (most supportive). We may also describe what proportion of observations fall above or below certain benchmarks for descriptive purposes only. We will also summarize answers to individual questions that make up the score, for example using a centered bar plot. Because there are no comparison groups for feasibility or acceptability, our primary analysis will be reporting a mean score with a 95% CI.

Note that although we are carefully weighting the mean score to align with our estimand of interest, it is informative to calculate simple descriptive statistics at the level of the data collection (i.e. visits). There may be discrepancies between summaries of visit-level data and weighted village-level means. We plan to report information about scores at the visit level, possibly including minimum and maximum scores, information about missing data, and detailed information about answers to individual questions. However, whenever possible the information accompanying the primary analysis will focus on village-level averages, (e.g. the median, IQR, and range of the village-level average scores), so that it aligns with the point estimate of interest and shows how average scores vary between villages. We may also stratify

visit- or individual-level summary statistics by village, which aligns more closely with the point estimate of interest.

Throughout this analysis report, to analyze the scores, we will present the village-level average as the point estimate of interest (the average score, as described above) and provide summaries of the variability in village-level observations such as IQR and range. For our main analysis of the primary five feasibility/acceptability scores defined below, we will also calculate the 95% confidence interval for the mean based on bootstrap sampling. For bootstrap confidence intervals, we will use simple random sampling with replacement of the villages, and no sampling within villages (Field and Welsh 2007). For each re-sampled set of 12 villages, we will calculate the mean score. We will repeat this at least 10,000 times and take the 2.5th and 97.5th percentile of all mean scores as the 95% confidence interval.

Additional analyses

Although primary interest is in examining the feasibility and acceptability of the entire intervention package, we may also explore feasibility and acceptability of each of the three intervention components separately (HIV education and counseling, HIV self-test kits, and peer group sessions). To do this, we would identify which of the questions within each of the primary scores are related to each specific intervention component. For each intervention component, if there are at least three questions, we would calculate a score based on those questions, using the same scoring method as described above. If there are only one or two questions that fall into that domain, we will not calculate a formal score; instead, we will rely on summary statistics for individual questions to provide insight.

This analysis report focuses on pre-specified scoring systems which summarize feasibility or acceptability over the entire study duration. We may also use exploratory analyses to investigate whether feasibility and acceptability change over time. This may be particularly important for barbers, who may not have experience delivering all components of the intervention at their Week 13 visit.

Our primary estimand is to understand feasibility and acceptability across all participants, regardless of their engagement in the intervention. However if there are a substantial proportion of participants who report not receiving a particular component of the intervention (e.g. did not attend group sessions), we may perform supplementary analyses stratified by participation status, which examine feasibility/acceptability among participants who experienced the relevant intervention, and among participants who did not engage in the relevant intervention. For group sessions, information about feasibility/acceptability is collected only among participants who attended at least one group session in the relevant time period, so supplementary analyses would highlight only the responses among participants who attended group sessions.

6.2 Participant survey responses

Question: How do participants feel about the feasibility of the intervention?

Population: All participants in ITT population from shops randomized to the intervention arm

Outcomes: Using questions from the “Barbershop Services Acceptability (Follow-up)” form, we will compute two feasibility scores: one that focuses on the participants’ perspective on barber qualities, and the other that focuses on the participant’s own experience. Our primary interest is in the overall intervention feasibility, but we may also examine questions for each of the three intervention components (HIV education and counseling, HIV self-test kits, and peer group sessions).

Participant feasibility score of barber qualities

- 1) The barber was knowledgeable about HIV and prevention. Strongly disagree, Disagree, Neutral, Agree, Strongly agree (*topic: HIV education and counseling*)
- 2) The barber was knowledgeable about HIV self-testing. Strongly disagree, Disagree, Neutral, Agree, Strongly agree (*topic: HIV self-test kits*)
- 3) The barber was non-judgmental and listened to my thoughts. Strongly disagree, Disagree, Neutral, Agree, Strongly agree
- 4) I felt that the barber-led group sessions provided a safe and comfortable space to discuss HIV-related topics. Strongly disagree, Disagree, Neutral, Agree, Strongly agree (*topic: peer group sessions; note – participants who did not attend a group session will be given a score of zero for this question*)

Participant feasibility score of participant experience

- 1) I feel that I have more knowledge about general HIV prevention and treatment services after talking with my barber. Strongly disagree, Disagree, Neutral, Agree, Strongly agree (*topic: HIV education and counseling*)
- 2) I feel that I have more knowledge about HIV testing after talking with my barber. Strongly disagree, Disagree, Neutral, Agree, Strongly agree (*topic: HIV self-test kits*)
- 3) I feel that I have more knowledge about PrEP after talking with my barber. Strongly disagree, Disagree, Neutral, Agree, Strongly agree
- 4) I feel that I have more knowledge about PEP after talking with my barber. Strongly disagree, Disagree, Neutral, Agree, Strongly agree
- 5) I feel that I have more knowledge about safe male circumcision after talking with my barber. Strongly disagree, Disagree, Neutral, Agree, Strongly agree

Participants answer these questions twice, at Week 26 and Week 52. At each visit, surveys will be scored and then aggregated over time using the strategy described in Section 6.1.

Descriptive analysis: To check the quality of the data and ensure all significant negative experiences have been captured, we will produce summaries which show the answers for each survey question (possibly including the ones loosely related to feasibility/acceptability which did not fit into any primary score) at the visit level. For each score, we will use summaries and/or visualizations showing the responses to individual questions that make up the score (e.g. using stacked bar plots). Because there are no specific intervention components with more than two questions within either feasibility score, we will not calculate any scores for the feasibility of the specific intervention components. Instead, we will rely on the summaries of individual questions to provide insight on feasibility of different aspects of the intervention.

Summaries of scores will include visit-level data to show the variability of responses at the visit level (e.g. minimum and maximum, IQR). However, our point estimates of interest are the average scores across villages as described in Section 6.1. For the average scores, we will describe the variability by using plots and/or summaries of the n=12 village-level scores (e.g. minimum and maximum, IQR). We will also compute a 95% CI for the village-level average score using bootstrap resampling of villages as described in Section 6.1.

Methods for clustered data: Surveys are collected at each visit. Basic descriptive statistics will be based on visit-level data to describe missingness patterns and incorporate skip pattern logic, and

summarize additional questions about feasibility/acceptability. For the derived scores, survey scores are averaged first within participants and then within villages, so that summaries focus on village-level means (weighting participants equally within villages regardless of follow-up time, and weighting villages regardless of size).

Question: How do participants feel about the acceptability of the intervention?

Population: All participants in ITT population from shops randomized to the intervention arm

Outcomes: Using questions from the “Barbershop Services Acceptability (Follow-up)” form, we will compute an acceptability score. Our primary interest is in the overall intervention acceptability, but we may also examine questions for each of the three intervention components (HIV education and counseling, HIV self-test kits, and peer group sessions).

Participant acceptability score

- 1) How satisfied are you with the HIV information provided by your barber? Very satisfied, Satisfied, Neutral, Dissatisfied, Very dissatisfied (*topic: HIV education and counseling*)
- 2) How comfortable were you discussing HIV and other health related topics with your barber? Very comfortable, Somewhat comfortable, Neutral, Somewhat uncomfortable, Very uncomfortable (*topic: HIV education and counseling*)
- 3) I am more comfortable seeking HIV prevention (i.e., testing) and treatment services after talking with my barber. Strongly disagree, Disagree, Neutral, Agree, Strongly agree
- 4) I felt that talking to my barber about HIV topics and receiving a haircut was too time consuming. Strongly disagree, Disagree, Neutral, Agree, Strongly agree (*topic: HIV education and counseling*)
- 5) I felt uncomfortable with the idea of taking an HIV self-test kit from my barber. Strongly disagree, Disagree, Neutral, Agree, Strongly agree (*topic: HIV self-test kits*)
- 6) I liked being able to get an HIV self-test kit from my barber. Strongly disagree, Disagree, Neutral, Agree, Strongly agree (*topic: HIV self-test kits*)
- 7) I worried about other people in the shop hearing my barber talking to me about HIV. Strongly disagree, Disagree, Neutral, Agree, Strongly agree (*topic: HIV education and counseling*)
- 8) I worried that I might experience stigma going to the barber who talks about HIV. Strongly disagree, Disagree, Neutral, Agree, Strongly agree
- 9) Talking with my barber has allowed me to think positively about seeking and/or using HIV prevention (i.e., PrEP and PEP) and treatment (i.e., ART) medications. Strongly disagree, Disagree, Neutral, Agree, Strongly agree
- 10) I felt that the barber-led group sessions were useful in improving my knowledge about HIV prevention and treatment services. Strongly disagree, Disagree, Neutral, Agree, Strongly agree (*topic: peer group sessions; note – participants who did not attend a group session will be given a score of zero for this question*)

Participants answer these questions twice, at Week 26 and Week 52. At each visit, surveys will be scored and then aggregated over time using the strategy described in Section 6.1.

Descriptive analysis: To check the quality of the data and ensure all significant negative experiences have been captured, we will produce summaries which show the answers for each survey question

(possibly including the ones loosely related to feasibility/acceptability which did not fit into any primary score) at the visit level. We will use summaries and/or visualizations showing the responses to individual questions that make up the participant acceptability score (e.g. using stacked bar plots). For the HIV education and counseling intervention component, we may also calculate an average score for the questions related to this specific component (as identified in the list above). For the other two intervention components, because there are too few questions, we will not calculate scores specific to these components; instead, we will rely on the summaries of individual questions to provide insight on acceptability.

Summaries of the acceptability score will include visit-level data to show the variability of responses at the visit level (e.g. minimum and maximum, IQR). However, our point estimate of interest is the average score across villages as described in Section 6.1. For the average score, we will describe the variability by using plots and/or summaries of the $n=12$ village-level scores (e.g. minimum and maximum, IQR). We will also compute a 95% CI for the village-level average score using bootstrap resampling of villages as described in Section 6.1.

Methods for clustered data: Surveys are collected at each visit. Basic descriptive statistics will be based on visit-level data to describe missingness patterns and incorporate skip pattern logic, and summarize additional questions about feasibility/acceptability. For the derived scores, survey scores are averaged first within participants and then within villages, so that summaries focus on village-level means (weighting participants equally within villages regardless of follow-up time, and weighting villages regardless of size).

6.3 Barber survey responses

Question: How do study barbers feel about the feasibility of delivering the intervention?

Population: Enrolled barbers from the intervention arm

Outcomes: Using questions from the “Barbershop Intervention Acceptability (Follow-up)” form, we will compute a feasibility score. Our primary interest is in the overall intervention feasibility, but we may also examine questions for each of the three intervention components (HIV education and counseling, HIV self-test kits, and peer group sessions).

Barber feasibility score

- 1) How easy was it to provide HIV prevention information to your clients while giving a haircut? Very easy, Somewhat easy, It was not easy or hard, Somewhat difficult, Very difficult (*topic: HIV education and counseling*)
- 2) How comfortable were you discussing HIV and other health related topics with your clients? Very comfortable, Somewhat comfortable, Neutral, Somewhat uncomfortable, Very uncomfortable (*topic: HIV education and counseling*)
- 3) I felt comfortable providing HIV prevention information to my clients. Strongly disagree, Disagree, Neutral, Agree, Strongly agree (*topic: HIV education and counseling*)
- 4) I felt uncomfortable with the idea of giving HIV self-test kits to my clients. Strongly disagree, Disagree, Neutral, Agree, Strongly agree (*topic: HIV self-test kits*)
- 5) I was uncomfortable talking about HIV with a client when there were other clients waiting in the shop. Strongly disagree, Disagree, Neutral, Agree, Strongly agree (*topic: HIV education and counseling*)

- 6) I felt comfortable providing information about HIV prevention and treatment services to my clients. Strongly disagree, Disagree, Neutral, Agree, Strongly agree (*topic: HIV education and counseling*)
- 7) I felt comfortable providing information about HIV testing to my clients. Strongly disagree, Disagree, Neutral, Agree, Strongly agree (*topic: HIV self-test kits*)
- 8) I felt comfortable providing information about PrEP to my clients. Strongly disagree, Disagree, Neutral, Agree, Strongly agree
- 9) I felt comfortable providing information about PEP to my clients. Strongly disagree, Disagree, Neutral, Agree, Strongly agree
- 10) I felt comfortable providing information about safe male circumcision to my clients. Strongly disagree, Disagree, Neutral, Agree, Strongly agree
- 11) I felt that I was able to answer the questions my clients had about HIV. Strongly disagree, Disagree, Neutral, Agree, Strongly agree (*topic: HIV education and counseling*)
- 12) I felt comfortable leading the group sessions with my clients. Strongly disagree, Disagree, Neutral, Agree, Strongly agree (*topic: peer group sessions; note – barbers who did not lead a group session will be given a score of zero for this question*)
- 13) I felt knowledgeable about the topics we discussed during the group session. Strongly disagree, Disagree, Neutral, Agree, Strongly agree (*topic: peer group sessions; note – barbers who did not lead a group session will be given a score of zero for this question*)
- 14) I felt confident leading a group of my clients in discussing HIV information. Strongly disagree, Disagree, Neutral, Agree, Strongly agree (*topic: peer group sessions; note – barbers who did not lead a group session will be given a score of zero for this question*)

Barbers answer these questions quarterly for the duration of the study. At each visit, surveys will be scored and then aggregated over time using the strategy described in Section 6.1.

Depending on rates of enrollment, barbers may not have engaged in all components of the intervention before their Week 13 visit. For example, participants are not expected to request self-test kits until 3 months after the participants enroll in the study, so barbers answering feasibility/acceptability questions about HIV self-test kits at Week 13 may not have experience distributing self-test kits yet.

Descriptive analysis: To check the quality of the data and ensure all significant negative experiences have been captured, we will produce summaries which show the answers for each survey question (possibly including the ones loosely related to feasibility/acceptability which did not fit into any primary score) at the visit level. We will use summaries and/or visualizations showing the responses to individual questions that make up the barber feasibility score (e.g. using stacked bar plots). For the HIV education and counseling intervention component and the peer group sessions intervention component, we may also calculate an average score for the questions related to each specific component (as identified in the list above). For the other intervention component of HIV self-test kits, because there are too few questions, we will not calculate scores specific to these components; instead, we will rely on the summaries of individual questions to provide insight on feasibility.

Summaries of the feasibility score will include visit-level data to show the variability of responses at the visit level (e.g. minimum and maximum, IQR). However, our point estimate of interest is the average score across villages as described in Section 6.1. For the average score, we will describe the variability by using plots and/or summaries of the n=12 village-level scores (e.g. minimum and maximum, IQR). We will also compute a 95% CI for the village-level average score using bootstrap resampling of villages as described in Section 6.1.

Methods for clustered data: Surveys are collected at each visit. Basic descriptive statistics will be based on visit-level data to describe missingness patterns and incorporate skip pattern logic, and summarize additional questions about feasibility/acceptability. For the derived scores, survey scores are averaged first within barbers and then within villages, so that summaries focus on village-level means (weighting barbers equally within villages regardless of follow-up time, and weighting villages regardless of size).

Question: How do study barbers feel about the acceptability of delivering the intervention?

Population: Enrolled barbers from the intervention arm

Outcomes: Using questions from the “Barbershop Intervention Acceptability (Follow-up)” form, we will compute an acceptability score. Our primary interest is in the overall intervention acceptability, but we may also examine questions for each of the three intervention components (HIV education and counseling, HIV self-test kits, and peer group sessions).

Barber acceptability score

- 1) I felt that my clients appreciated discussing HIV prevention information with me. Strongly disagree, Disagree, Neutral, Agree, Strongly agree (*topic: HIV education and counseling*)
- 2) I like being able to give HIV self-test kits to my clients. Strongly disagree, Disagree, Neutral, Agree, Strongly agree (*topic: HIV self-test kits*)
- 3) I worried about people not wanting to come to my shop because they don’t want to hear about HIV. Strongly disagree, Disagree, Neutral, Agree, Strongly agree (*topic: HIV education and counseling*)
- 4) I enjoyed providing HIV information and discussing other health related topics with my clients. Strongly disagree, Disagree, Neutral, Agree, Strongly agree (*topic: HIV education and counseling*)
- 5) I felt that the barber-led group sessions were useful for my clients to learn more about HIV. Strongly disagree, Disagree, Neutral, Agree, Strongly agree (*topic: peer group sessions; note – barbers who did not lead a group session will be given a score of zero for this question*)
- 6) I felt that my clients appreciated attending the barber-led group sessions. Strongly disagree, Disagree, Neutral, Agree, Strongly agree (*topic: peer group sessions; note – barbers who did not lead a group session will be given a score of zero for this question*)

Barbers answer these questions quarterly for the duration of the study. At each visit, surveys will be scored and then aggregated over time using the strategy described in Section 6.1.

Depending on rates of enrollment, barbers may not have engaged in all components of the intervention before their Week 13 visit. For example, participants are not expected to request self-test kits until 3 months after the participants enroll in the study, so barbers answering feasibility/acceptability questions about HIV self-test kits at Week 13 may not have experience distributing self-test kits yet.

Descriptive analysis: To check the quality of the data and ensure all significant negative experiences have been captured, we will produce summaries which show the answers for each survey question (possibly including the ones loosely related to feasibility/acceptability which did not fit into any primary score) at the visit level. We will use summaries and/or visualizations showing the responses to individual

questions that make up the barber acceptability score (e.g. using stacked bar plots). For the HIV education and counseling intervention component, we may also calculate an average score for the questions related to the specific component (as identified in the list above). For the other intervention components, because there are too few questions, we will not calculate scores specific to these components; instead, we will rely on the summaries of individual questions to provide insight on acceptability.

Summaries of the acceptability score will include visit-level data to show the variability of responses at the visit level (e.g. minimum and maximum, IQR). However, our point estimate of interest is the average score across villages as described in Section 6.1. For the average score, we will describe the variability by using plots and/or summaries of the $n=12$ village-level scores (e.g. minimum and maximum, IQR). We will also compute a 95% CI for the village-level average score using bootstrap resampling of villages as described in Section 6.1.

Methods for clustered data: Surveys are collected at each visit. Basic descriptive statistics will be based on visit-level data to describe missingness patterns and incorporate skip pattern logic, and summarize additional questions about feasibility/acceptability. For the derived scores, survey scores are averaged first within barbers and then within villages, so that summaries focus on village-level means (weighting barbers equally within villages regardless of follow-up time, and weighting villages regardless of size).

6.4 Recruitment and retention

Question: How quickly were barbers able to recruit participants?

Population: All enrolled participants (from both arms)

Outcomes: Participant enrollment in the study

Duration of recruitment at each shop will begin on the first day the shop was allowed to recruit clients, and end when the shop is asked to stop recruitment. These dates are specific to each shop. If a shop is asked to temporarily pause enrollment (e.g. due to limits on the total number of participants and uneven recruitment between shops), this time will be removed from the recruitment time.

Barbers used a recruitment log to record how many clients they talked to about the study on each day, and how many clients were referred to study staff each day. By adding these daily reports up over the duration of enrollment, we can infer how many times a barber talked to clients about the study, but not the total number of clients contacted. From these daily logs, we will derive the following shop-level variables:

- 1) Total number of times clients were contacted about the study (sum of all daily client contact counts)
- 2) Total number of client referrals (sum of all daily client referral counts)

Descriptive analysis: Total number of participants enrolled and recruitment time will be summarized overall and by shop. Summaries will also be made by arm. Recruitment over time may be plotted to show any variation in participant enrollment over time since shop recruitment began. We will calculate the number of participants enrolled per village-week (total number of participants enrolled in a village divided by the total weeks of recruitment from all shops in that village).

We will summarize the reports from the barbers about how many clients they contacted and how many clients were referred to study staff each day. These summaries may include information about how complete the log form reporting is (e.g. proportion of days between shop activation and the end of participant enrollment that have an associated recruitment log form), and whether it aligns with the

recruitment time for each shop. We will calculate the number of participants enrolled per client visit (total number of participants enrolled in a village divided by the total number of times a barber reported talking to clients about the study).

Methods for clustered data: Information about enrollment and recruitment will be summarized primarily overall, pooling all shops and villages (equally counting each enrollment, regardless of which shop recruited the participant). For summaries of the derived shop-level variables from daily logs, data will be presented at the shop level (equally weighting shops). Note that this distinguishes between shops within a single village.

Question: How well were participants retained in the study?

Population: Participants in ITT population (from both arms)

Outcomes: Participant visit completion at Week 26 and Week 52

All participants are expected to have two study visits (Week 26 and Week 52). For the purposes of measuring participant retention, any early termination due to reasons other than death will be counted as missed visits.

Descriptive analysis: Visit completion (percent of participants attending visits within the allowable window) will be summarized overall and by arm for each scheduled visit (Week 26 and Week 52, equally weighting participants). For shops in the intervention arm, visit completion by shop may also be summarized.

Methods for clustered data: Information about visit completion will be stratified by visit (Week 26 or Week 52); for each specific visit, information will be summarized at the level of data collection, equally weighting participants within each time point.

Question: How well were intervention barbers retained in the study?

Population: Enrolled barbers from shops randomized to deliver the intervention

Outcomes: Visit completion at quarterly visits

Visit completion summaries will be only among barbers expected to have the relevant visit. For the purposes of measuring barber retention, any early termination due to reasons other than death will be counted as missed visits.

Descriptive analysis: Visit completion will be summarized by quarterly visit (equally weighting barbers). We will also summarize what percent of barbers were terminated before the end of the study, and reasons for termination.

Methods for clustered data: Information about visit completion will be stratified by visit (Week 13, Week 26, etc.); for each specific visit, information will be summarized at the level of data collection, equally weighting barbers within each time point. Information about termination is collected at the barber level, and will be presented at the barber level (equally weighting barbers).

6.5 Barber-participant interactions and intervention delivery

Question: Based on participant's retrospective reports, how frequently did participants interact with their barber and what interventions did they receive?

Population: Participants in ITT population

Outcomes: On the Barbershop Services Received form, participants report what services they received from barber visits, and what services they received via group sessions. This form is filled out at the Week 26 and Week 52 visits.

The first set of questions ask about experiences during haircuts/barbershop services (excluding group sessions). Information collected includes:

- 1) Did the participant attend the barbershop they were recruited from since the last study visit? Yes, No
- 2) How many times did the participant attend their recruiting barbershop for a haircut since the last study visit?
- 3) The last time they went to the recruiting barbershop for a haircut, approximately how long did the participant spend at the shop? Include waiting time.
- 4) Mark all HIV-related services they received from a barber during their most recent haircut at the recruiting shop: HIV education, HIV prevention referral information, HIV treatment referral information, HIV self-test kit information, HIV self-test kits to take home, None, Other (specify)*
- 5) Mark all HIV-related services they received from a barber from the recruiting shop at any of their haircuts since the last study visit: HIV education, HIV prevention referral information, HIV treatment referral information, HIV self-test kit information, HIV self-test kits to take home, None, Other (specify)*

For participants who attended a barbershop other than the one they were recruited from during the time frame of interest, questions 2-5 are repeated for services they received at other participating shops, so that answers can be combined to determine the participant's treatment exposure from all participating shops. The form also asks how many times a participant attended barbershops outside the study.

The second set of questions asks about group sessions. Information collected includes:

- 1) Did the participant attend a barber-led group session since the last study visit? Yes, No*
- 2) How many group sessions did they attend since the last study visit?*
- 3) Which barber/shop led the group session? (checkboxes for shops, or unknown)*
- 4) Mark all topics discussed during the last group session they attended: What is HIV, How HIV is spread, Condom use, HIV PrEP, HIV PEP, HIV testing, HIV self-testing, STI management, Circumcision, Referral/community resources, Other (specify), None*

Note that questions marked by asterisks above are specifically related to intervention components that we would expect to be received in the intervention arm only. We are primarily interested in summarizing these in the intervention arm but will also examine responses in the control arm to detect contamination. The other questions are related to how participants interact with their barbers, which are applicable to both arms but have special importance in the intervention arm, since it indicates how often participants could have been receiving the intervention from their barbers. We are interested in whether these metrics (namely number of visits and duration of haircuts) differ between arms, which may give some insight on how the intervention impacted the barbers' business and participants' behavior.

From this visit-level data, the following participant-level outcomes will be derived (combining Week 26 and Week 52 responses):

- 1) Whether participants received any intervention services from any barber at haircuts between enrollment and Week 52 (Yes/No)
- 2) Whether participants attended any barber-led group session between enrollment and Week 52 (Yes/No)
- 3) Average number of visits to the recruiting barbershop (average of counts from Week 26 and Week 52)
- 4) Average time spent at the shop for the participant's last haircut (average of times from Week 26 and Week 52)

Descriptive analysis: Participant responses will be summarized both overall and separately at Week 26 and Week 52. Visit-level information described above will be summarized (equally weighting visits), divided by arm to understand whether there was any contamination in the control group and examine barbershop visit frequency in the control group. Participants who report not attending any participating barbershops during the period in question will be excluded from summaries and analyses of number of visits, duration of haircuts, services received, etc. We will describe how frequently participants received services at barbershops other than their enrollment shop. For participants who received HIV-related services at another shop (e.g. participants who transferred from one shop to another), we will combine their answers about all participating shops to reflect all services received whenever possible.

Additional summaries will be done for the intervention arm only. We will derive two binary variables for each participant describing receipt of barber services, as described above. These binary variables will be summarized at the village level (equally weighting villages) to examine what percent of enrolled clients received services and attended group sessions. We will also describe how these outcomes vary between villages, e.g. possibly by plotting the percent of participants who received services per village.

Statistical analysis: We will compare the number of visits to the recruiting shop and the time spent at the shop for the last haircut between arms. To combine participants' responses at Week 26 and Week 52, we will use each participant's average response over the two study visits, as described above. If a participant is missing the relevant data at only one visit, the data from the other visit will be used. If a participant is missing the relevant data at both visits, they will be excluded. Linear regression of aggregated data will be used to account for clustering at the village level. This analysis equally weights participants within barbershops regardless of follow-up duration, so long as the participants have at least one non-missing follow-up measurement. Villages are given equal weight, regardless of the number of participants.

We wish to test whether the mean number of visits to the recruiting shop differs between arms, accounting for clustering within villages and adjusting for covariates used in the randomization (village population, number of clients per day, and distance from the barbershop to the nearest health facility). Throughout this analysis plan, adjustment for the randomization variables will mean adjusting for a linear term for each variable. To do this, we will aggregate data to the village level and use linear regression to estimate a treatment effect, adjusted for randomization covariates (Hayes and Moulton 2017, Section 10.5). We will report the difference in means between arms with 95% confidence intervals for the difference. The time spent at the shop for the most recent haircut will be analyzed using the same methods. The steps below detail how this analysis will be conducted.

1. Outcome data will be aggregated to the village level. Let Y_{ij} be the outcome (e.g. number of visits) for participant j from village i ; note that here, Y_{ij} is an average over all the study visits for participant j . Then the aggregated outcome for village i is $Y_i = \frac{1}{N_i} \sum_j Y_{ij}$, where N_i is the number of participants in village i .

2. A model will be fit to the village-level aggregated data. We will use linear regression to estimate an adjusted difference in means between the treatment arms. The model will include the three randomization covariates, and an indicator for the treatment arm:

$$Y_i = \alpha + \beta_t t_i + \beta_{z1} z_{i1} + \beta_{z2} z_{i2} + \beta_{z3} z_{i3} + \varepsilon_i$$

where t_i is 1 if the village was randomized to the intervention arm and 0 if the village was randomized to the control arm, z_{i1} is the village population, z_{i2} is the number of clients per day, and z_{i3} is the distance from the barbershop to the nearest health facility. We assume ε_i are Normally distributed and independent, which is appropriate since we expect correlation within villages, but not between villages. Note that these covariates are measured at the barbershop level; see Section 1.7 for details on covariate values for villages with multiple barbershops.

3. Model results will be summarized. The intervention effect of interest is β_t , which is the average difference in the aggregated outcome (e.g. average number of visits per village) between villages randomized to intervention vs. control, adjusted for village population, number of clients per day, and distance to nearest health facility. We will also present a 95% confidence interval for β_t .

Methods for clustered data: Information about barbershop attendance and services received is collected at each participant visit. Basic descriptive statistics will be based on visit-level data to describe missingness patterns and incorporate skip pattern logic, and summarize additional details. For the key participant-level derived variables listed above, survey responses are aggregated for each participant, and descriptive statistics about the outcome are based on participant-level data (equally weighting all participants). Formal analyses of the participant-level variables weights participants equally within villages, and weights villages equally regardless of size.

Question: Based on barber logs, how frequently did barbers deliver the intervention during client visits, and what components of the intervention were delivered?

Population: Enrolled barbers in the intervention arm

Outcomes: Each shop keeps a form to record daily study activity (Services Provided Log/intervention tracking log). Information in this log includes the following for each barbershop visit from a study participant:

- 1) HIV education provided? Yes, No, Client declined
- 2) HIV self-test information provided? Yes, No, Client declined
- 3) Number of test kits provided
- 4) HIV prevention/care referral provided? Yes, No
- 5) Number of other clients in attendance

The following village-level variables will be derived:

- 1) Total number of visits from participating clients
- 2) Proportion of visits that included delivery of the intervention, including any of the following: HIV education, HIV self-test information, HIV self-test kits, or HIV prevention/care referral.
- 3) Total number of HIV self-test kits distributed

Descriptive analysis: Information from the services provided log will be summarized primarily at the village level using the derived variables above, equally weighting daily reports across the duration of the study (not accounting for the number of participants or barbers in each shop). More detailed information may be summarized at the level of data collection, equally weighting client visits. We may also examine whether characteristics changed over time by stratifying summaries of intervention delivery into 26-week periods. Since each enrolled client is supposed to receive one self-test kit approximately 3 months after

each study visit, we could compare this schedule to the barber's reported distribution of self-test kits to see if they distributed enough HIV tests to allow all enrolled clients to test on the appropriate schedule. However, barbers were allowed to distribute multiple test kits to the same person and to clients not enrolled in the study, so distributing a large number of test kits does not imply that all enrolled clients received a self-test kit.

Methods for clustered data: Information about client visits is collected every time a participating client visits the barbershop; this information will be summarized at the level of data collection, equally weighting each client visit. For the derived village-level variables listed above, descriptive statistics are based on village-level data (equally weighting villages).

Question: How frequently were group sessions held?

Population: Enrolled barbers in the intervention arm

Outcomes: Study staff attend each group session and complete the Group Session log form. The village-level variable of total number of group sessions held will be derived based on the group session logs.

Descriptive analysis: We will summarize the number of group sessions delivered in each village. We may also examine whether the frequency of group sessions changed over time by stratifying summaries into 26-week periods. In addition to summarizing average or overall metrics, we will also summarize how much delivery of group sessions varied between villages, including the village-level range.

Methods for clustered data: Information about group sessions is collected at each group session on a log form. Note that per the protocol, each shop should have one group session every two months; under this scenario, equally weighting group session observations will correspond to equally weighting villages. For the derived village-level variable described above, descriptive statistics are based on village-level data (equally weighting villages).

7. SECONDARY ENDPOINTS

7.1 Self-initiated HIV testing

Objective: To compare completion of self-initiated HIV testing between intervention and control groups

Question: Is there a difference in the percent of participants who got HIV testing between study visits between arms?

Population: Participants not living with HIV

Outcomes: We will use the number of tests since last study visit from the HIV Prevention Services Follow-up form to derive binary variables. Participants who tested 0 times during the relevant period will be "no", participants who tested one or more times will be "yes". The following participant-level variables will be derived:

- Yes/no did the participant test for HIV between enrollment and Week 26

- Yes/no did the participant test for HIV between Week 26 and Week 52
- Yes/no did the participant test for HIV both between enrollment and Week 26, and between Week 26 and Week 52

Descriptive analysis: Information about HIV testing (e.g. number of times tested, whether self-test kit was used, where testing was done) will be summarized by arm at the visit level (equally weighting participant visits). This will include the average number of times participants tested for HIV between visits in each arm. Only participants who have a non-missing outcome at Week 26 will be included in the primary Week 52 summaries, since all questions refer to the time since last study visit. We may also describe the correlation between the survey responses at Week 26 and 52 with any information gathered between visits on the Interim HIV Testing log form. We will summarize the three main participant-level outcomes bulleted above, and describe how the percent of men testing varies between villages.

Statistical analysis: We will compare the proportion of participants reporting HIV testing between arms for the first portion of the study (between enrollment and the Week 26 visit) and the second portion of the study (between the Week 26 and Week 52 visits) separately. In addition, the proportion who reported HIV testing during both periods will be compared between arms. To estimate the difference in proportion of men testing, we will use a risk difference from a village-level analysis. Inference will be based on linear regression of aggregated village-level outcomes. For each outcome, we will report the difference in percentages between arms with a 95% confidence interval. The regression analysis will be analogous to the one described above under the primary outcomes, where the village-level average $Y_i = \frac{1}{N_i} \sum_j Y_{ij}$ is equivalent to the proportion of participants within each village who tested for HIV. Linear regression of these village-level averages estimates the difference in percent of men testing for HIV between arms, adjusted for the covariates used in the randomization. We will also report a p-value for the difference between arms for each of the three HIV testing outcomes from the regression models. We will use a two-sided test with $\alpha=0.05$, and no adjustment is planned for multiple comparisons. As a sensitivity analysis, we will also calculate the p-values described above using the Wilcoxon rank-sum test on the village-level averages (but note that this does not incorporate adjustment of covariates).

Methods for clustered data: Information about HIV testing is collected at each participant visit. Basic descriptive statistics will be based on visit-level data to describe missingness patterns and incorporate skip pattern logic, and summarize additional information about testing patterns. For the three derived variables listed above, survey responses are aggregated for each participant, and descriptive statistics about the outcome are based on participant-level data (equally weighting all participants). Formal analyses of the participant-level variables weights participants equally within villages, and weights villages equally regardless of size.

7.2 Risk behaviors

Objective: To evaluate the preliminary effectiveness of the intervention on change in behaviors associated with HIV acquisition

Question: Is there a difference between arms in self-reported risk behaviors?

Population: Participants in ITT population

Outcomes: Self-reported risk behaviors at Week 26 and Week 52 on the ACASI Sexual Behavior form. Topics covered include condom use and HIV status of sexual partners. The following are key factors collected at each visit:

- 1) Whether participants used a condom the last time they had vaginal sex with a partner. For participants with a main partner and other sexual partners, they will be counted as 'yes' if they answered 'yes' to both "Did you use a condom the last time you had sex with someone other than your main partner" and "Did you use a condom the last time you had vaginal sex with your main partner".
- 2) Total number of sexual partners in the last three months (number of non-main partners, plus one if the participant has a main partner)
- 3) Number of new sex partners in the last three months (partners are considered "new" if participants have not had sex with them in the past year)
- 4) Number of sexual partners with unknown HIV status in the last three months
- 5) Among participants with a main partner, HIV status of the main partner (HIV positive, HIV negative, or unknown)

From this visit-level data, the following participant-level outcomes will be derived (combining Week 26 and Week 52 responses) and compared between arms:

- 1) Whether participants ever reported condomless sex at their last sexual encounter from Week 26 or Week 52 (Yes/No)
- 2) Average number of total sexual partners from the three months prior to Week 26 and three months prior to Week 52 (average of counts from Week 26 and Week 52)
- 3) Among participants who ever had a main partner, whether participants ever had a main partner with unknown HIV status at Week 26 or Week 52 (Yes/No)

Descriptive analysis: Information about self-reported risk behaviors will be summarized by arm in detail at the visit level (equally weighting participant visits, pooling across Week 26 and Week 52). We will summarize the three main participant-level risk behavior outcomes (listed above), and describe how the percentages and means vary between villages.

Statistical analysis: We will compare the three main participant-level risk behavior outcomes between arms. For binary outcomes, we will estimate a risk difference and 95% confidence intervals. For count outcomes, we will estimate a difference in means and 95% confidence intervals. We will perform village-level analyses, using linear regression of aggregated data as described above, adjusted for the covariates used in randomization.

Methods for clustered data: Surveys are collected at each participant visit (e.g. number of sex partners in past three months). Basic descriptive statistics will be based on visit-level data to describe missingness patterns and incorporate skip pattern logic, including important context such as whether participants have a main partner. For key variables, survey responses are aggregated for each participant using the rules above (e.g. average number of sex partners across visits), and descriptive statistics about the outcome are based on participant-level data (equally weighting all participants). Formal analyses of the key participant-level variables weights participants equally within villages, and weights villages equally regardless of size.

7.3 HIV prevention services

Objective: To compare interest in or use of HIV prevention services between intervention and control groups

Question: Is there a difference between arms in uptake of VMMC?

Population: VMMC cohort (participants not living with HIV who did not report that they were circumcised at enrollment, including participants who did not know their circumcision status)

Outcome: Participants self-report new circumcisions at each study visit (“Have you been newly circumcised since your last study visit?”, asked at Week 26 and Week 52). From this visit-level data, we will derive the following participant-level outcome: whether participants were circumcised between enrollment and Week 52 (Yes/No).

Descriptive analysis: The proportion of men who reported new circumcisions will be summarized by arm. Participants will be included in the analysis even if they were missing the question about circumcision at one of the follow-up visits, by assuming that there was no VMMC in the missing time period. Participants missing both follow-up measurements will be excluded. We will summarize the percent of men with new circumcisions, and describe how the percent varies between villages.

Statistical analysis: We will estimate the difference in percentages between arms using linear regression of aggregated data which accounts for clustering by village and adjusts for covariates used in randomization. We will report the difference in percent of men who are newly circumcised between arms and 95% confidence intervals.

Methods for clustered data: Surveys on VMMC status are collected at each participant visit. Basic descriptive statistics will be based on visit-level data to describe missingness patterns. Survey responses are aggregated for each participant using the rules above (i.e. any new circumcision over the course of the study), and summarized at the participant level (equally weighting all participants). Formal analyses of the outcome weights participants equally within villages, and weights villages equally regardless of size.

Question: Is there a difference between arms in self-reported referral to PEP care or use of PEP?

Population: Participants not living with HIV

Outcomes: Participants report whether they have gone to a health facility to get PEP since the last study visit (Yes/No), whether they were provided with a PEP prescription (Yes/No), and whether they have used PEP since the last study visit (Yes/No). On the ACASI, participants also report whether they have taken PEP in the past three months to prevent getting HIV. From this visit-level data, the following participant-level outcomes will be derived (combining Week 26 and Week 52 responses):

- 1) Whether participants ever went to a health facility to get PEP between enrollment and Week 52 (Yes/No)
- 2) Whether participants ever got a PEP prescription between enrollment and Week 52 (Yes/No)
- 3) Whether participants ever used PEP between enrollment and Week 52 (Yes/No)
- 4) Whether participants ever reported using PEP in the past three months on the ACASI (Yes/No)

Descriptive analysis: Participant responses about PEP use will be summarized by arm at the participant level (equally weighting participants). Participant responses from the surveys at Week 26 and Week 52 will be combined as described above. Participants will be included in summaries as long as they have the relevant measurement at either the Week 26 or Week 52 time point by assuming that the missing time period did not include any PEP access or use. Participants missing both follow-up measurements will be excluded.

Note that the questions about accessing a health facility and obtaining a prescription are only relevant to participants who are not already using PEP. At the time of enrollment, on the ‘HIV Prevention Services’ CRF, participants are asked whether they have ever used PEP (Yes/No). However, due to the nature of PEP it is difficult to identify which participants already had a prescription or supply of PEP pills that they

were utilizing at the time of enrollment. As a result, we will include all participants in summaries of PEP access and use, even though some of them may have already had access to PEP at the time of enrollment.

Statistical analysis: We will compare PEP use in the past three months (combining Week 26 and Week 52 by summarizing whether participants reported PEP use either at Week 26 or at Week 52), and estimate the difference in percent using PEP between arms. We will also compare the proportion of participants who have ever gone to a health facility to get PEP, and estimate the difference in percentage between arms. We will perform village-level analyses, using linear regression of aggregated data and adjusting for covariates used in randomization. We will report the difference in percentages and 95% confidence interval. If there are differences in PEP use between arms, we may perform exploratory analyses to see whether differences in sexual behavior between arms might explain PEP use.

Methods for clustered data: Surveys are collected at each participant visit (e.g. PEP use since the last study visit). Basic descriptive statistics will be based on visit-level data to describe missingness patterns and incorporate skip pattern logic. Survey responses are aggregated for each participant using the rules above (e.g. ever using PEP), and descriptive statistics about the outcome are based on participant-level data (equally weighting all participants). Formal analyses of select participant-level variables weights participants equally within villages, and weights villages equally regardless of size.

Question: Is there a difference between arms in self-reported referral to PrEP care or use of PrEP?

Population: Participants not living with HIV

Outcomes: Participants report whether they have gone to a health facility to get PrEP since the last study visit (Yes/No), whether they were provided with a PrEP prescription (Yes/No), and whether they have used PrEP since the last study visit (Yes/No). On the ACASI, participants also report whether they have taken PrEP in the past three months to prevent getting HIV. From this visit-level data, the following participant-level outcomes will be derived (combining Week 26 and Week 52 responses):

- 1) Whether participants ever went to a health facility to get PrEP between enrollment and Week 52 (Yes/No)
- 2) Whether participants ever got a PrEP prescription between enrollment and Week 52 (Yes/No)
- 3) Whether participants ever used PrEP between enrollment and Week 52 (Yes/No)
- 4) Whether participants ever reported using PrEP in the past three months on the ACASI (Yes/No)

Descriptive analysis: Participant responses about PrEP use will be summarized by arm at the participant level (equally weighting participants). Participant responses from the surveys at Week 26 and Week 52 will be combined as described above. Participants will be included in summaries as long as they have the relevant measurement at either the Week 26 or Week 52 time point by assuming that the missing time period did not include any PrEP access or use. Participants missing both follow-up measurements will be excluded.

Note that the questions about accessing a health facility and obtaining a prescription are only relevant to participants who are not already using PrEP. At the time of enrollment, on the 'HIV Prevention Services' CRF, participants are asked whether they are currently using PrEP (Yes/No). Participants who are using PrEP at the time of enrollment will be excluded from summaries of PrEP access (including going to health facilities and getting prescriptions), but included in summaries of PrEP use.

Statistical analysis: We will compare PrEP use in the past three months (combining Week 26 and Week 52 by summarizing whether participants reported PrEP use either at Week 26 or at Week 52), and estimate the difference in percent using PrEP between arms. We will also compare the proportion of

participants who have ever gone to a health facility to get PrEP (excluding participants who already have a prescription for PrEP), and estimate the difference in percentage between arms. We will perform village-level analyses, using linear regression of aggregated data and adjusting for covariates used in randomization. We will report the difference in percentages and 95% confidence interval.

Methods for clustered data: Surveys are collected at each participant visit (e.g. PrEP use since the last study visit). Basic descriptive statistics will be based on visit-level data to describe missingness patterns and incorporate skip pattern logic. Survey responses are aggregated for each participant using the rules above (e.g. ever using PrEP), and descriptive statistics about the outcome are based on participant-level data (equally weighting all participants). Formal analyses of select participant-level variables weights participants equally within villages, and weights villages equally regardless of size.

7.4 Interest in long-acting PrEP

Objective: To assess interest in long-acting PrEP among all participants by study arm

Question: How do men in this population feel about long-acting PrEP, and is there a difference in interest in long-acting PrEP between arms?

Population: Participants in ITT population

Outcomes: Self-reported interest in long-acting PrEP at Week 52. Topics covered include interest in long-acting PrEP and preferences for product and delivery method, with a total of seven questions. The key question of interest is: "How interested would you be in taking long-acting injectable PrEP?", which participants answered with very uninterested, somewhat uninterested, neutral, somewhat interested, or very interested. For formal analysis, we will convert this to a binary outcome indicating whether participants are interested (somewhat interested or very interested) in long-acting PrEP vs. not (neutral, somewhat uninterested, or very uninterested).

Note that the questions about long-acting PrEP are asked only at Week 52, so there are no repeated measures within participants.

Descriptive analysis: Information about self-reported interest in long-acting PrEP will be summarized overall and by arm at the participant level (equally weighting participants).

Statistical analysis: We will compare the binary outcome of interest in long-acting PrEP between arms. We will estimate a risk difference and 95% confidence intervals. We will perform village-level analyses, using linear regression of aggregated data as described above, adjusted for the covariates used in randomization.

Methods for clustered data: Surveys are collected at each participant's Week 52 visit. Basic descriptive statistics will be based on participant-level data (equally weighting all participants). Formal analyses of the key interest outcome weights participants equally within villages, and weights villages equally regardless of size.

8. EXPLORATORY ENDPOINTS

Exploratory endpoints will be described in a separate analysis plan, in collaboration with the Laboratory Center.

9. DEVIATIONS FROM PROTOCOL

In the protocol Version 1.0 Section 8.6.2, it states that “the proportion of participants who report receiving a referral for PrEP and/or PEP or attending a clinic to be screening for PrEP/PEP eligibility, will be compared between arms” to address the secondary outcome of comparing interest in or use of HIV prevention services. In the forms used in the study, participants provide information about whether they have gone to a health facility to get PrEP/PEP, whether they were given a PrEP/PEP prescription, and their use of PrEP/PEP, but we do not directly record whether participants received a referral. As a result, this SAP focuses only on the clinic attendance element in Section 7.3, and not the referral itself.

In the protocol Version 1.0, it implies that some outcomes will be compared separately at Week 26 and Week 52. In this SAP, we instead focus on comparing outcomes over the whole study period, and leave time-specific comparisons to future work

10. REFERENCES

Field, C.A., & Welsh, A.H. (2007). Bootstrapping clustered data. Journal of the Royal Statistical Society: Series B (Statistical Methodology)., 69(3), 369-390. <https://doi.org/10.1111/j.1467-9868.2007.00593.x>

Hayes, R. J., & Moulton, L. H. (2017). Cluster randomised trials, second edition. CRC Press. <https://doi.org/10.4324/9781315370286>

11. CHANGE HISTORY

Version		Affected Section(s)	Activity Description
Number	Effective Date		
1.0	August 14, 2024	N/A	New document
2.0	January 27, 2025	Sections 2.1, 2.2, 2.5, 3.1, 3.2, 4, 5, 6.1, 7, 8, 9, 11	- Updated background to reflect the development that all participants may be followed through Week 52, and correspondingly revised analyses to focus on the overall study period instead of separately analyzing Week 26 and Week 52 outcomes, (Sections 2.1, 8, 9, 11). - Removed analysis of feasibility/acceptability opinions before being exposed to the intervention,

			and moved information about confidence intervals to Section 5 (Sections 5, 8.1, 8.2) - Added a supplementary analysis examining feasibility/acceptability only among treated participants (Section 8.1) - Added a sensitivity analysis for the two-stage analysis of HIV testing (Section 9.1) - Clarified how different questions from the participant's survey on interactions with barbers will be used, and how we will handle any cases where participants received services at multiple participating barbershops (Section 8.4) - Clarified how to handle missing data when defining the VMMC cohort (Section 3.1, 9.3) - Added descriptive information about shop switching and additional enrollment of barbers (Section 6.1) - Clarified in primary endpoints list that participant surveys will examine feasibility as well as acceptability (Section 2.2) - Clarified that shop characteristics with respect to imbalance between arms are specifically baseline characteristics (Section 3.2) - Made minor grammatical corrections (Sections 2.5, 3.2, 4, 7)
3.0	March 21, 2025	Sections 1.2, 1.6, 1.7, 2, 3, 5, 6, 7	- Added an explicit discussion of estimands for feasibility/acceptability survey responses (Section 6.1) - Revised the analysis of feasibility/acceptability surveys to focus on separate derived feasibility and acceptability scores, including new exploratory analyses specifically related to these scores (Sections 3, 6). - Based on simulation results, the two-stage analysis procedure was replaced with linear regression of aggregated data (Sections 1.7, 6, 7) - Added detail about how data will be summarized at different levels of clustering, including a paragraph on methods for handling clustered data for each question of interest and more explicit identification

			of how derived higher-level outcomes are defined (Sections 1.7, 6, 7) - Added a description of the recommendations from the SMC and early release of results (Section 2) - Modified the outcome related to peer group session attendance to include all reported attendance regardless of topics discussed (Section 6.5) - Specified that 95% confidence intervals for engagement in intervention components will be calculated via bootstrapping (Section 6.5) - Revised the analysis of feasibility/acceptability of client recruitment to summarize individual questions and stratify by visit (Section 6.3) - Added clarification about including or excluding participants who report PrEP and PEP use at baseline (Section 7.3) - Added note about an exploratory analysis to contextualize PEP use (Section 7.3) - Deleted an outdated paragraph about only following a subset of participants to Week 52 (Section 1.7) - Deleted an outdated sentence about analyzing baseline feasibility and acceptability (Section 5) - Deleted an outdated sentence about only following a subset of participants to Week 52 (Section 6.4) - Made minor grammatical corrections (Sections 1.2, 1.6, 6, 7)
4.0	Date of last signature	Sections 1, 5, 6, 7	- Updated the SAP to align with V2.0 of the study protocol, including adding a new secondary objective (Sections 1, 7.4) - Added some detail and clarification about the estimands in the primary analysis (Section 6.1) - Refined prioritization of analyses so that the SAP focuses on analyses necessary to address the objectives, with potentially important supportive analyses described as optional analyses that may be performed if resources allow (Sections 1.7, 5, 6, 7). Removed some analyses which were beyond the scope of the SAP, including: mention of an exploratory factor analysis (Section 6.1); summaries of questions about how barbers felt about study

			recruitment (Section 6.3); 95% confidence intervals for summaries of participant engagement (Section 6.5); detailed summaries about the content of group sessions and expected schedule of group sessions (Section 6.5); detailed information about risk behaviors (Section 7.2).
--	--	--	---