Statistical Analysis Plan (SAP)

Study Title: Supporting Oral Pre-exposure Prophylaxis Decision Making Among Pregnant Women in Lilongwe, Malawi

NCT number NCT06394323
Document Date 08/08/2025

Supporting oral pre-exposure prophylaxis decision making among pregnant women in Lilongwe, Malawi: a pilot study

Statistical Analysis Plan

Lauren Hill, PhD MSPH

Principal Investigator, University of North Carolina at Chapel Hill

Friday Saidi, MBBS, MMed

Site Lead Investigator, University of North Carolina (UNC) Project–Malawi

Jessica Keys, PhD, MSPH

Biostatistician, University of North Carolina at Chapel Hill

Carlie L. Sulpizio, MPH

Graduate Research Assistant, University of North Carolina at Chapel Hill

Ivy Kaliati

Data Research Associate

UNC Project Malawi

Version 1.1

08 August 2025

Table of Contents

1. Amendments

Current Version	Current Version Date	Summary of Changes
Version 1.0	July 22, 2025	Finalized Version 1.0
Version 1.1	August 8, 20205	Updated formatting, chance imbalance, and primary outcomes analysis

2. Acronyms

AIM	Acceptability of Intervention Measure
AIPW	Augmented Inverse Probability Weighting
CI	Confidence Interval
DCS	Decisional Conflict Scale
FIM	Feasibility of Intervention Measure
HIV	Human Immunodeficiency Virus
IAM	Intervention Appropriateness Measure
ITT	Intention-To-Treat
IPTW	Inverse Probability of Treatment Weights
MI	Multiple Imputation
MyChoice	MyChoice for HIV Prevention Intervention
PrEP	Pre-exposure Prophylaxis
SAP	Statistical Analysis Plan
SD	Standard Deviation
SDM	Shared Decision Making
SOC	Standard of Care

3. Introduction

This statistical analysis plan (SAP) details the statistical procedures that address the study objectives specified in the September 2024 **protocol** of the investigator-initiated NIH-funded pilot study titled: "**Supporting oral pre-exposure prophylaxis decision making among pregnant women in Lilongwe, Malawi: a pilot study**," also known as MyChoice Study 1.

The purpose of this pilot study is to evaluate the **feasibility**, **acceptability**, and **appropriateness** of a shared decision-making (SDM) intervention, **MyChoice,** designed to support HIV-negative

pregnant women in their decisions about initiating oral pre-exposure prophylaxis (PrEP). This SAP details the planned analyses for both the primary implementation outcomes and secondary cognitive outcome (decisional conflict), providing a framework for interpreting findings that will inform the design of a future efficacy trial.

4. Study Objectives and Summary

4.1 Protocol Title:

Supporting oral pre-exposure prophylaxis decision making among pregnant women in Lilongwe, Malawi: a pilot study

4.2 Study Design:

This study is a single-blind, pilot feasibility trial of the MyChoice shared decision making (SDM) intervention compared to standard of care (SOC) for pregnant women considering daily oral PrEP. We will evaluate the feasibility, acceptability, and appropriateness of the MyChoice intervention.

Population: The primary population for this study is pregnant women, aged 18 years or older, at risk of HIV infection that have not initiated PrEP. We focus on pregnant women because of the elevated HIV risk faced by women in the perinatal period.

Sample size: 100 participants in total, comprised of

- N=50 HIV negative pregnant women randomized to the MyChoice intervention arm
- N=50 HIV negative pregnant women randomized to the Standard of Care (SOC) arm

Follow-up: Participants will be followed for approximately three months from enrollment, with study visits at months 0, 1, 2, and 3 (see **Appendix A** for-visit windows).

Study site: Lilongwe, Malawi (Bwaila Hospital)

Study duration: The length of participant follow-up is approximately three months from enrollment (month 0).

Study groups: Participants randomly assigned to the *control group* will receive SOC HIV prevention counseling in accordance with the usual care in Malawi per national guidelines. Women randomly assigned to the MyChoice *intervention group* will receive the MyChoice study intervention, My Choice for HIV Prevention (MyChoice) counseling, a counselor-delivered shared decision-making approach for pregnant women considering PrEP.

Clinical Trial ID: NCT06394323, clinicaltrials.gov

4.3 Study Aims

Specific aims which will be completed through the pilot study are as follows:

Aim 1: Evaluate the acceptability, appropriateness, and feasibility of the MyChoice PrEP shared decision-making intervention.

Aim 2: Assess the plausibility of intervention effects on a proximal cognitive endpoint (decisional conflict) in preparation for a future efficacy trial.

4.4 Study Outcomes

We will assess the following primary and secondary outcomes in this feasibility pilot study.

4.4.1 Primary Outcomes

Primary outcomes will include participant-reported perceptions of the MyChoice intervention (assessed through questionnaire self-report using validated scales)[1]:

- *Intervention acceptability*: the extent to which participants perceive the intervention to be agreeable, palatable, or satisfactory will be assessed using the Acceptability of the Intervention Measure (AIM) scale. [1]
- *Intervention appropriateness*: the perceived relevance and usefulness to support decision making will be assessed using the Intervention Appropriateness Measure (IAM) scale.[1]
- *Intervention feasibility*: the extent to which the intervention is feasible or practical will be assessed using the Feasibility of Intervention Measure (FIM) scale.[1]

Primary study outcomes will be assessed through participant questionnaires at the enrollment visit among MyChoice arm participants. Acceptability and Appropriateness outcomes will also be assessed at the month 2 follow-up visit among MyChoice arm participants to understand how perceptions of the MyChoice intervention change over time.

4.4.2 Secondary Outcomes

The secondary outcome of decisional conflict will be assessed through participant self-report using a validated scale to assess the extent to which the counseling received by participants in either arm helped them make decisions about HIV prevention during pregnancy and breastfeeding.

Decisional conflict or perceptions regarding decision uncertainty, satisfaction, clarity of personal values, and support for decision-making will be assessed with the Decisional Conflict Scale.[2–5]

The secondary outcome will be assessed in the study questionnaire at the enrollment visit among all participants after receipt of study counseling.

5. Randomization and Masking Procedures

Once enrolled, at their first visit (month 0, expected to be the same date as screening), participants will be randomly assigned to one of two study arms (MyChoice or SOC) at a 1:1 study arm ratio. The randomization assignments will be generated in advance using a permuted block design. All

randomization plans, including a back-up randomization plan for replacement randomization ID (RID), will be generated using SAS version 9.4 (SAS Institute, Cary, NC). The treatment arm assignment, as linked to a RID, will be placed in opaque sealed envelopes and sequentially numbered with these RID.[6] Once RID is assigned, the corresponding intervention (MyChoice or SOC) is administered as outlined in the protocol.

6. General Statistical Considerations

In this pilot study, the primary objective is to understand **acceptability**, **feasibility**, and **appropriateness** of the MyChoice intervention. Given the pilot nature of the study, emphasis will be placed on **estimation and precision** rather than hypothesis testing.

6.1 Descriptive Statistics

Initial analysis will include descriptive analyses to characterize the population on features such as demographics (e.g., age, marital status, income, education) and each outcome of interest. We will examine whether key sociodemographic features (including but not limited to: age, income, education, marital status, gestational age, perceived HIV risk, and experience of IPV) differ by study arm (n (%) for categorical data and median (25th, 75th percentile), mean (SD), median (IQR) and min-max for continuous measures). Testing and confidence intervals (CIs) will be two-sided and at a 95% confidence level (i.e., alpha 0.05) with no adjustment for multiplicity, unless stated otherwise. Any characteristics differing significantly by arm may be included as covariates in all analyses comparing study arms.

6.2 Estimation and Precision

Given the pilot nature of the study, emphasis will be put on estimation and precision of estimates, rather than null hypothesis testing. We will use large-sample methods (e.g., Wald CIs) when the nominal CI coverage level is tenable. Primary endpoints, which are evaluated only among the participants that received the MyChoice intervention, will be analyzed strictly descriptively with no formal hypothesis testing.

6.3 Multiplicity and Confidence Interval Coverage

An alpha of 0.05 will be used throughout analysis to compute 95% CIs, with no adjustment for multiplicity. In the case of analyses of continuous data with $n<30$, 95% CI coverage properties of large-sample methods will be evaluated in simulation studies or existing exact statistical methods (e.g., exact CI for a risk difference) will be used.

6.4 Analysis Population and Approach

Acceptability, **appropriateness**, and **feasibility** endpoints will be restricted to women who were randomized to and participated in the MyChoice intervention. For the secondary outcome of decisional conflict, the SOC group will serve as the referent. Analyses will be conducted using an intention-to-treat (ITT) approach, analyzing participants according to their original randomization assignment.

6.5 Small Data

Appropriate descriptive statistics will follow standard formats (e.g., n (%), median (IQR), mean (SD) if subgroup size are small (e.g., <5 participants per category for categorical data or <10 participants in comparison group for continuous analyses), an exact 95% CI will be calculated as a sensitivity analysis. (e.g., exact CI for a risk difference).

6.6 Chance Imbalance

We anticipate that our measured baseline covariates will be balanced between the arms by randomization. However, if there is evidence of considerable chance imbalance in key baseline covariates, a **post hoc sensitivity analysis** may be conducted for the secondary outcome using **Augmented Inverse Probability Weighting (AIPW)** to adjust for these imbalances and can be used to produce double robust effect estimates compared to Inverse Probability of Treatment Weighting (IPTW). This analysis will be exploratory in nature and will not replace the unadjusted primary analysis. A doubly robust approach would also be suitable, if feasible, given our sample size (n=50 per arm). The protocol does not address chance imbalance, and such an analysis would be a sensitivity approach.

6.7 Missing Data

Missing (unevaluable) data are anticipated to be uncommon for the primary endpoints. Therefore, we will conduct a complete case analysis of the primary study endpoints, excluding participants that are missing all items pertaining to a given scale measure. We may also provide best-case, worst-case upper and lower bounds around our estimates in the case of >10% unevaluable data, assuming either all missing endpoints fall into the favorable (best-case) or unfavorable (worst-case) categories. Women who did not complete the 2-month visit will be excluded from the denominator of those primary endpoints assessed at that visit (AIM, IAM).

Among a given study arm for the secondary outcome (DCS), if >10% of participants in either randomization group are unevaluable, multiple imputation (MI) using chained equations (aka fully conditional specification [FCS] method) may be applied under a missing at random assumption (White 2011). Thirty or more MI datasets may be created, and results will be combined using Rubin's rule. Each imputation model will be specified prior to conducting outcome analyses by randomization group, and missingness patterns will be reviewed prior to finalizing imputation models. The imputation models will consider interaction terms between randomization group and covariates and will include at least all the covariates in the outcome analysis model; covariate values will also be imputed for the chained equations as needed. Interactions will be evaluated for inclusion only if sample size allows. Factors that are related to both the endpoint and to missingness will be important for inclusion in the imputation models.

6.8 Analysis Cohorts

- **Intention-To-Treat (ITT)**: all participants who are enrolled and randomly assigned to and received the MyChoice or SOC intervention.

7. Primary Outcomes

Acceptability, Appropriateness, and Feasibility Outcomes

7.1 Acceptability Endpoint (a)

(i) AIM acceptability scale mean score for the MyChoice intervention at **months 0 and 2,** respectively.

7.2 Appropriateness Endpoint (b)

(i) IAM appropriateness scale mean score for the MyChoice intervention at **months 0 and 2**, respectively.

7.3 Feasibility Endpoints (c)

(i) FIM feasibility scale mean score for the MyChoice intervention at **month 0**

For **all** primary outcomes, the **month 0** visit will be the primary result for each endpoint. For **acceptability** and **appropriateness** measures, **month 2** result will serve as the supplemental result.

7.4 Cohort and Study Arm

These outcomes will be assessed only among participants randomized to the MyChoice intervention arm using ITT.

7.5 Measures

All primary outcomes will be assessed through participant self-report in the SDM Intervention Acceptability and Appropriateness instrument using the following standardized instruments (Appendix B):

(a) Acceptability will be measured using the **Acceptability of Intervention Measure (AIM) scale**.
(b) Appropriateness will be measured using the **Intervention Appropriateness Measure (IAM) scale**.
(c) Feasibility will be measured using the **Feasibility of Intervention Measure (FIM) scale**.

Each scale consists of four items rated on a 5-point Likert scale coded from 1 to 5 with 1 representing “completely disagree” and 5 representing “completely agree”, corresponding to increasing levels of (a) acceptability, (b) appropriateness, or (c) feasibility, respectively. During implementation coding was reversed, such that 1 corresponds to “completely agree” and 5 to “completely disagree”.[1] To maintain consistency with the original published scales, items will be reversed coded prior to calculating a mean scale score. Each of these questions includes a non-response option, coded as 999. This will be set to missing for calculating summary scores. Each

participant's summary score for a given scale will be calculated as the **mean** of the non-missing item responses. Summary scores will range from 1 to 5, with higher scores indicating more favorable perceptions of the intervention.

(a) The items used to assess MyChoice intervention **acceptability** at month 0 and month 2 consists of the four-item AIM scale rated on a 5-point scale[1]. After reverse coding (see above), each item will range from 1= "completely disagree" to 5= "completely agree," so that higher scores reflect increasing acceptability. At least one item among the four must be completed to calculate an acceptability score for a participant.

(b) The items used to assess MyChoice intervention **appropriateness** at month 0 and month 2 consist of the four item IAM rated on a 5-point scale[1]. After reverse coding, each item will range from 1= "completely disagree" to 5= "completely agree" so that higher scores reflect increasing appropriateness. At least one item among the four must be completed to calculate an appropriateness score for a participant.

(c) The items used to assess MyChoice intervention **feasibility** at month 0 consists of the four-item FIM scale rated on a 5-point scale[1]. After reverse coding, each item will range from 1= "completely disagree" to 5= "completely agree," so that higher scores reflect increasing feasibility. At least one item among the four must be completed to calculate a feasibility score for a participant. score for a participant.

7.6 Analysis

(i) **Item-Level Summary:** Among participants that received the MyChoice intervention, we will describe the distribution of responses for each individual item within the **acceptability** (a) and **appropriateness** (b) and **feasibility** (c) scales. For **acceptability** and **appropriateness,** we will summarize responses at **month 0** and **month 2**, separately. For **feasibility** (c), we will describe the distribution of responses to each of the question items at **month 0**, as this construct was not assessed at Month 2. Descriptive statistics will include counts, percentages, and graphical summaries were appropriate.

(ii) **Scale-Level Summary:** We will then calculate a mean score for each participant that received the MyChoice intervention at each applicable time point Across participants at each time point, we will estimate the mean of these individual means the common standard error using the sum of squared differences between the individual and population means, and the 95% CI around the sample mean for each construct (acceptability, appropriateness, feasibility). These calculations will be performed separately for **month 0** and **month 2**. Participants with at least one or more completed item on a given scale will be included in these analyses. We will calculate the weighted group mean where we weight the mean for each participant by the inverse of their individual variance and get a pooled SE (square root of the sum of individual variance estimates divided by the N^2)

The formula is as follows:

Weighted mean of means ($\bar{x}_W$) where each participant $i's$ variance ($Var(\bar{x}_i)$) from their individual mean $\bar{x}_i$ is multiplied by their individual mean ($\bar{x}_i$) and summed across all participants N and then divided by the sum of the variance for participants $i = 1$ to N.

$$\bar{x}_W = \frac{\sum_i^N w_i \bar{x}_i}{\sum_i^N w_i}, \quad W_i = \frac{1}{Var(\bar{x}_i)}$$

Pooled standard error

$$(SE_{pool}): \sqrt{\frac{1}{N^2}\sum_i^N Var(\bar{x}_i)}$$

(iii) **Descriptive Comparisons Over Time:** Although the study is not powered to detect statistically significant within-subject changes, we will describe observed changes in mean scale scores for acceptability and appropriateness from **Month 0** to **Month 2** to provide insight into potential shifts in participant perceptions over time.

8. Secondary Outcomes

Decisional Conflict Scale

8.1 Endpoints

(i) Decisional Conflict scale mean score at **month 0**

8.2 Hypotheses

Participants in the MyChoice intervention group will report, on average, decisional conflict scores lower than participants in the control group.

H_0: The mean DCSs of the MyChoice and SOC arms are equal.

H_1: The mean DCSs of the MyChoice and SOC arms are not equal (2-tailed). For 1-tailed test, you would use an H_1 of the mean DCS in the MyChoice arm is lower than the mean DSC in the SOC arm.

8.3 Cohort and Study Arm

Secondary outcomes will be assessed only among participants randomized to both the MyChoice Intervention and SOC arm using ITT.

8.4 Measures

(a) The DCS scale contains 16 items rated on a 5-point scale. The scale questions may be found on the Decisional Conflict Scale REDCap instrument (**Appendix C**). Each scale item is rated from 1 to 5 (“strongly agree” to “strongly disagree”). Each of the questions include a non-response option (999), that will be set to missing for analysis. DCS scores will be computed based on women’s responses to statements, where higher scores

indicate greater decisional conflict, while lower scores indicate diminished conflict and heightened certainty in decision-making. DCS scores are computed as the mean rating across items multiplied by 25 (theoretical range: 0 to 100). The scale will be administered at the month 0 visit following completion of study counselling.

8.5 Analysis

(i) **Item-Level Summary:** We will describe the responses for each of the 16-items on the Decisional Conflict Scale (DCS) at **month 0.** Descriptive statistics will include counts, percentages, and graphical summaries, as appropriate.

(ii) **Scale-Level Summary:** For each participant, in each study arm, we will calculate the mean of their completed DCS at **Month 0** and multiply that result by 25 to produce a standardized summary score ranging from 0 to 100, in accordance with the established scoring procedures. Only participants with at least one completed DCS item will be included in the analysis. For each study arm, we will calculate the mean of participants' individual summary scores (mean of means), the SD, and the 95% Ci using the common variance.

(iii) **Between-Group Comparisons:** We will compare the mean DCS scores between the intervention (MyChoice) and SOC arms using the Welch's t-test (assuming unequal variance). We will also report the absolute difference in mean scores between arms (SOC as the referent), along with the associated 95% CI using the pooled variance assumption.

(iv) **Missing Data:** If more than >10% of participants in either randomization group are unevaluable, multiple imputation (MI) may be applied as described in section 6.7.

(v) **Adjustment for Imbalances:** If key demographic variables or baseline characteristics related to decisional conflict differ substantively by study arm, we will use AIPW to adjust for these differences, as described in section 6.6

Appendix A: Visits and Visit Windows

Study visit schedule

Procedure	Screening	Enrollment Visit	Follow-up visits		
			1 mo. (-/+ 15 days)	2mo. (+/- 15 days)	3 mo. (+/- 15 days)
Study introduction	X				
Eligibility assessment	X				
Informed consent		X			
SDM counseling (intervention)		X			
Questionnaires		X	X	X	X
PrEP assisted referral		X			
Qualitative interviews			X		X
Adherence measurement (PrEP adopters only)					
Pill count/self-report			X	X	X
DBS				X	

Appendix B: Primary Outcome Questionnaires (AIM, IAM, FIM)

The Acceptability of Intervention Measure (AIM), Intervention Appropriateness Measure (IAM), and Feasibility of Intervention Measure (FIM) was adapted from the Weiner Psychometric assessment of three newly developed implementation outcome measures. (2017) [1] Response options were switched to from Completely Disagree to Completely Agree to Completely Agree to Completely Disagree to mimic the response options from other scale that were used throughout the study.

Instructions: I'd like to know what you think about the study you've been participating in to support your use of an HIV prevention method of your choice. Remember, this program consists of PrEP, female, and male condoms. Please indicate how you feel about each of these statements by indicating a number from 1 (completely agree) to 5 (completely disagree).

Acceptability of Intervention

	Completely Agree	Agree	Neither Agree nor disagree	Disagree	Completely Disagree
The shared-decision counseling study meets my approval					
The shared-decision counseling study is appealing to me.					
I like the shared-decision counseling I received.					
I welcome the shared-decision counseling.					

Instructions: I'd like to know what you think about the study you've been participating in to support your use of an HIV prevention method of your choice. Remember, this program consists of PrEP, female, and male condoms. Please indicate how you feel about each of these statements by indicating a number from 1 (completely agree) to 5 (completely disagree).

Intervention Appropriateness Measure

	Completely Agree	Agree	Neither Agree nor disagree	Disagree	Completely Disagree
The shared-decision counseling seems fitting.					
The shared-decision counseling seems suitable.					
The shared-decision counseling seems applicable.					
The shared-decision counseling seems like a good match.					

Instructions: Now I'd like to ask you about how much you agree with the following 4 statements regarding the decision counseling you took part in. Please indicate how you feel about each of these statements by indicating a number from 1 (Completely agree) to 5 (completely disagree).

Feasibility of Intervention Measure

	Completely Agree	Agree	Neither Agree nor Disagree	Disagree	Completely Disagree
The shared-decision counseling seems implementable.					
The shared-decision counseling					

seems possible.	
The shared-decision counseling seems doable.	
The shared-decision counseling seems easy to use.	

Appendix C: Secondary Outcome Questionnaire (DCS)

Instructions: Now, please think about the decision you made to use a HIV prevention method as a part of this study. Please indicate how you feel about each of these statements by indicating a number from 1 (strongly agree) to 5 (strongly disagree).

	Strongly Agree	Agree	Neither Agree nor Disagree	Disagree	Strongly Disagree	No Response
I know which options are available to me.						
I know the benefits of each option.						
I know the risks and side effects of each option.						
I am clear about which benefits matter most to me.						
I am clear about which risk and side effects matter most to me.						

I am clear about which is more important to me (the benefits or the risks and side effects).
I have enough support from others to make a choice.
I am choosing without pressure from others.
I have enough advice to make a choice.
I am clear about the best choice for me.
I feel sure about what to choose.
This decision is easy for me to make.
I feel I have made an informed choice.
My decision shows what is important to me.
I expect to stick with my decision.
I am satisfied with my decision.

References

1. Weiner BJ, Lewis CC, Stanick C, et al. Psychometric assessment of three newly developed implementation outcome measures. *Implement Sci*. 2017;12(1):108. doi:10.1186/s13012-017-0635-3

2. O'Connor AM. Validation of a decisional conflict scale. *Med Decis Making*. 1995;15(1):25-30. doi:10.1177/0272989X9501500105

3. Garvelink MM, Boland L, Klein K, et al. Decisional Conflict Scale Use over 20 Years: The Anniversary Review. *Med Decis Making*. 2019;39(4):301-314. doi:10.1177/0272989X19851345

4. Knops AM, Goossens A, Ubbink DT, Legemate DA, Stalpers LJ, Bossuyt PM. Interpreting patient decisional conflict scores: behavior and emotions in decisions about treatment. *Med Decis Making*. 2013;33(1):78-84. doi:10.1177/0272989X12453500

5. Sun Q. *Predicting Downstream Effects of High Decisional Conflict: Meta-Analyses of the Decisional Conflict Scale*. 2005.

6. Doig GS, Simpson F. Randomization and allocation concealment: a practical guide for researchers. *J Crit Care*. 2005;20(2):187-191; discussion 191. doi:10.1016/j.jcrc.2005.04.005