ACTG A5391

Primary Statistical Analysis Plan

Version 3.0

Doravirine for Obese Persons on Integrase Inhibitors and Tenofovir Alafenamide (The Do IT Study)

Based on Protocol Version 2.0 (August 19, 2022)

Released for activation on 04FEB2025

ClinicalTrials.gov Identifier: NCT04636437

This is ACTG A5391 SAP Version 3.0 with names of authors, names of publication writing team members and analysis timeline redacted.

Table of Contents

Version History

Version	Changes Made	Date Finalized
1	Original Version	Aug 12, 2021
2	Update for protocol revision to version 2	Jan 11, 2023
3	Updates for final primary analysis	Feb 4, 2025

1 Introduction

1.1 Purpose

This Primary Statistical Analysis Plan (SAP) describes the primary estimands and other secondary outcome measures that will address specific study objectives and interim monitoring of the A5391 trial. The Primary SAP includes general analytic approaches for all primary estimands, secondary and other outcome measures in the primary manuscript(s) or submitted to ClinicalTrials.gov (regardless of the reporting timeline). The Primary SAP facilitates discussion of the statistical analysis components among the lead study investigators and statisticians, helping them agree on the statistical analyses to be performed and presented in the primary analysis report.

Analyses for the primary analysis report will be initiated once the last participant has completed the week 48 study visit; details on the analysis timeline will be finalized at a later time. In addition, all primary and secondary outcomes outlined in this SAP will be submitted to ClinicalTrials.gov within 1 year of the Primary Completion Date (PCD); PCD defined as 48 weeks after the last participant enrolls.

Additional specifications and details on other key elements of the primary analysis report, including tables, figures, and coding descriptions, are provided in the Analysis Implementation Plan (AIP).

Details for analyses of other objectives and outcome measures not included in the Primary SAP will be provided in a separate SAP.

1.2 Version History

1.2.1 Version 3

Major modification in v3 include the following:

- Update to handling of participants found ineligibile after randomization (note: those never starting treatment already excluded, so this revision applies to exclusion from efficacy analyses (but not safety) for participants found ineligibile who received study treatment
- Update to sample size considerations to account for change to accrual goal following final SMC interim review
- Updating analysis range for follow-up visits to be more inclusive of available data (i.e. wider intervals without gaps)
- Updating analysis set for insulin and insulin resistance (per HOMA-IR) outcomes to exclude persons taking exogenous insulin per clinical database reporting.

Minor updates:

- Updating ACTG network roster for members notified of final analysis completion
- Removing analysis timeline (as timeline maintained and updated in a separate document)
- Updated typo (vs rather than or) in the subgroup formulation for race and ethnicity

1.2.2 Version 2

Major modifications in v2 include the following:

- Update to align with updates in the trial protocol
  - participant eligibility criteria which defines the target population
  - added secondary outcome to incorporate reporting and analysis of BMD T scores from DXA evaluations
- Removal of exploratory objectives which are not covered by this SAP
- Updating of anchor time for calculation timing of evaluations
- Report contents section added
- Template analysis timeline added
- Writing team roster added

Minor modifications include the following:

- Update to language based on template updates (e.g. references to estimands in section 1.1.)

2 Trial Overview

2.1 Trial Design

A5391 is a phase IV, 48-week, 3-arm, open-label, randomized study to assess whether switching from an integrase strand transfer inhibitor (INSTI)-based regimen with tenofovir alafenamide/emtricitabine (TAF/FTC) or TAF/lamivudine (TAF/3TC) to a regimen containing doravirine (DOR) with TAF/FTC (or TAF/3TC) or DOR with tenofovir disoproxil fumarate (TDF)/FTC (or TDF/3TC) will result in reduced weight gain or weight loss compared with continuation of their current regimen.

The trial population consists of adults (≥ 18 years of age) living with HIV who have a body mass index (BMI) ≥30 kg/m^2, and who have maintained virologic suppression (HIV-1 RNA <50 copies/mL or below the lower limit of HIV-1 RNA detection available at the site) on an INSTI + TAF/FTC (or TAF/3TC), regimen for the 48 weeks prior to entry. [Note: in the previous version of the protocol, the target population were persons with BMI > 2.75 and with excessive weight gain under their INSTI-based regimen.]

The trial aims to enroll up to 150 (formerly 222) participants, targeting >50% participants assigned female sex at birth across all sites and >50% African American/Black participants at US sites. Randomization is stratified by sex assigned at birth and race (African American/Black vs non-African American/Black).

2.2 Primary Hypothesis

Compared to continued INSTI (bictegravir (BIC), dolutegravir (DTG), or raltegravir (RAL)) + TAF/FTC (or TAF/3TC), a switch to DOR+TAF/FTC (or TAF/3TC) will result in less weight gain or overall weight loss at 48 weeks among obese persons (i.e. BMI ≥30 kg/m^2). A switch from INSTI+TAF/FTC (or TAF/3TC) to DOR+TDF/FTC (or TDF/3TC) will result in further reduction in weight gain or overall weight loss compared to DOR+TAF/FTC (or TAF/3TC).

2.3 Trial Objectives

The following sections list the primary and secondary objectives from the trial protocol; corresponding protocol section numbering is shown in brackets.

2.3.1 Primary Objective

To assess if an antiretroviral therapy (ART) switch to DOR (either switch to DOR and remain on TAF/FTC [or TAF/3TC] or switch to DOR and TDF/FTC [or TDF/3TC]) results in differences in weight change over 48 weeks, compared to a strategy of remaining on current ART (INSTI+TAF/FTC [or TAF/3TC]) [Protocol Objective 1.2].

2.3.2 Secondary Objectives [key secondary objectives identified]

1) To assess if a switch to TDF/FTC (or TDF/3TC) from TAF/FTC (or TAF/3TC) results in differences in weight changes over 48 weeks by contrasting the 2 DOR-containing switch strategies (i.e., DOR+TAF/FTC [or TAF/3TC] versus DOR+TDF/FTC [or TDF/3TC]) [Protocol Objective 1.3.1]. [key]

2) To assess the effects of a switch from INSTI+TAF/FTC (or TAF/3TC) to DOR+TAF/FTC (or TAF/3TC), or DOR+TDF/FTC (or TDF/3TC) over 48 weeks on [Protocol Objective 1.3.2]:

- Safety and tolerability;
- Maintenance of virologic suppression;
- Minimum waist circumference;
- Fasting cardiometabolic parameters: glucose and insulin, homeostatic model assessment of insulin resistance (HOMA-IR) and lipids (triglycerides, low-density lipoprotein [LDL], high-density lipoprotein [HDL]);
- Dual-energy X-ray absorptiometry (DEXA) body composition measurements of total and regional fat and lean mass, and hip and lumbar spine bone mineral density.

2.4 Overview of Sample Size Considerations

The original targeted sample size is 222 participants **was updated to up to 150 participants following the final SMC review on June 5, 2023.**

The study is designed to evaluate whether an ART switch from BIC, DTG, or RAL with TAF/FTC or TAF/3TC to a regimen containing DOR with TAF/FTC (or TAF/3TC) or DOR with TDF/FTC (or TDF/3TC) results in reduced weight gain, or weight loss, as compared to continuation of current BIC-, DTG-, or RAL-containing ART. The primary comparison will examine the change (percent) in body weight (kg) from entry to week 48.

Original sample size justification (n=222)

The sample size is based on two-sided t-tests with 2.5% significance level and 80% statistical power to detect an absolute difference of 5% at week 48 between each switch arm and the control arm (continuing INSTI), assuming the standard deviation (SD) of within-participant changes is 9%.

Updated sample size justification (up to n=150)

While the trial was designed on the assumption that the standard deviation (sd) on the primary outcome (from external data) would be 9, an interim look of available sd from 37 participants (all accrued during v1) was much smaller at 5.3 (and similar sd between women (4.84) and men (5.05)). Using this observed standard deviation, using a smaller sample size (n=152) the trial would still have 80% power (using same assumptions of 2-sided t-test at 2.5% significance level of 2 pairwise comparisons), to detect a between arm difference in weight changes at 48 weeks of 3.58 percentage point (%pts), which was smaller than the originally specified 'clinical difference of interest' of 5 %pts. Under a slightly larger sd of 7, sample size (n=152) would provide 80% statistical power for a group difference of 4.73 %pts, similar to the original design. Finally, the total width of the 2-sided, 97.5% confidence interval about the treatment effect estimate was calculated with smaller sample size and was 5.2 %pts (for sd 5.3), and 6.9 %pts (for sd 7).

2.5 Overview of Formal Interim Monitoring

The study will be monitored by an independent ACTG-appointed Study Monitoring Committee (SMC). The SMC will review the study at least annually, with the first review occurring no more than 1 year after the first participant enrolls; interim reviews may also be convened if a concern is identified. The SMC will review ***unblinded, by-arm summaries of accrual, baseline characteristics, study and treatment status (including premature discontinuations), data availability/completeness, virologic failures, and AEs.*** Summaries of primary efficacy will not be provided during interim reviews.

There are no pre-specified stopping guidelines for this study, but the SMC should consider possible modifications or termination on bases of safety concerns or operational futility. With regards to safety**, if *10% or more of participants on the DOR-containing arms experience virologic failure or treatment-related AEs*,** this would be concerning to the study team. With respect to operational futility, the **SMC *would consider the overall and within arm LTFU rates, and treatment cross-over rates (particularly those switching from DOR-containing arm back to an INSTI-containing arm).*** As a benchmark, an ***overall LTFU rate of 20% or higher would be cause for concern, particularly if a higher rate is observed in the INSTI-containing arm where participants are not switched to DOR.***

3 Outcome Measures

Outcome measures are copied from the protocol, with slight changes shown with underline text. Only primary and secondary outcome measures, which are addressed in this SAP are included below.

3.1 Primary Outcome Measure

1) Change (percent) in body weight (kg) from entry to week 48.
   [For Primary Objective and Secondary Objective 1]

3.2 Secondary Outcome Measures

1) Change (percent) in body weight (kg) from entry to week 24.
   [For Primary Objective and Secondary Objective 1]

2) Change (absolute) in <u>minimum</u> waist circumference from entry to weeks 24 and 48.
   [For Secondary Objective 2]

   Minimum waist circumference is defined as the smallest horizontal circumference above the umbilicus and below the xiphoid process. The minimum waist circumference summarized in analysis is the average of three measurements taken in triplicate by the site investigators.

3) Change (absolute) in fasting cardiometabolic parameters (glucose, ***insulin***, ***HOMA-IR***, triglycerides, LDL, HDL) from entry to weeks 24 and 48. [***italicized*** based on batched, retrospective testing on stored plasma and serum at central labs}
   [For Secondary Objective 2]

4) Occurrence of confirmed plasma HIV-1 RNA >200 copies/mL following treatment initiation [For Secondary Objective 2]

   Confirmed plasma HIV-1 RNA > 200 copies/mL is defined as two consecutive results above 200 copies, with the testing of the second specimen within 2 weeks of site receipt of the first result, and the first result being from a specimen drawn after treatment initiation.

5) Occurrence of new Grade ≥3 AEs, through week 48. [For Secondary Objective 2]

   New Grade 3 or higher AEs are defined as Grade 3 or higher AEs that are new in onset or are aggravated in severity or frequency from the entry condition (e.g., Grade 1 or 2 at entry that escalates to Grade 3 or higher, or Grade 3 at entry that escalates to Grade 4 or higher), following treatment initiation

6) Occurrence of >10% reduction in CrCl as estimated by the CKD-EPI equation, relative to entry, through week 48. [For Secondary Objective 2]

7) Occurrence of premature discontinuation of study treatment through week 48. [For Secondary Objective 2]

   Premature discontinuation of study treatment is defined as stopping any component of the randomized study regimen, prior to the week 48 visit.

8) Change (percent) in DEXA body composition measures (total fat and lean mass, trunk fat, limb fat, and appendicular lean mass) from entry to week 48. [For Secondary Objective 2]

   Total fat, lean mass in kg as well as total fat % (i.e. %fat in tissue defined as [total fat mass / (total fat mass+ lean mass)]; will have %change outcome calculated

   Limb fat (in kg and %) ,where %fat in tissue, will be calculated from sum of arm/upper extremity & leg/lower extremity lean and fat masses reported

   Trunk fat (in kg and %) changes will be calculated

   Limb (appendicular) lean mass changes will be calculated

9) Change (percent) in DEXA hip and lumbar spine bone mineral density from entry to week 48. [For Secondary Objective 2]

   Hip BMD is defined via the following 4 outcome measures: left femoral neck, right femoral neck, left total hip (femur), and right total hip (femur) (each measured in g/cm3)

Lumbar spine BMD is defined as the arithmetic average of L1-L4 BMD in g/cm3

10) Change in categorization of bone mass (WHO classification) from baseline to week 48 based on minimum T score from left/right femoral neck, left/right total femur/hip, and lumbar spine(L1-L4).

4 Statistical Principles

4.1 General Considerations

The following analysis populations are defined for A5391:

- Screened Population: All participants who were screened for enrollment into the study.
- Randomized Population: All participants who were enrolled to the study (this is an intent-to-treat [ITT] population).
- Treated Population: All participants who received at least one dose of their randomized study treatment (this is a modified ITT population). For participants randomized to switch arms this includes those who took at least one dose of their randomized regimen; for participants randomized to remain on their current regimen this includes those who took at least one dose of their current regimen after study entry.
- **Eligible treated Population: randomized study treatment** **All participants who received at least one dose of and who were eligible as per eligibility criteria (this is an intent-to-treat [ITT] population).**

The following will be considered for timing of study visits:

Study Entry: **This is an interval defined by the first entry evaluations and finalization of study evaluations (including DEXA). If all entry evaluations completed on one day this is an interval of 1 day.** The start of study entry evaluation may be delayed following randomization for the requirement of being evaluated in a fasting state. From the protocol: Study treatment should be initiated as soon as possible after entry evaluations have been completed and within 14 days of entry.

- Treatment Initiation: Study treatment may be initiated following completion of study entry evaluations (except DEXA). (Note : this definition applies regardless of treatment assignment, and so control arm initiation will be the first dose taken following completion of entry evaluations.). Study treatment should be initiated as soon as possible after entry evaluations have been completed and within 14 days of entry
- Entry DEXA: Entry DEXA may occur up to 14 days post initiation of study treatment

Study week analysis windows are based on the Schedule of Evaluations (SOE) defined in the protocol and will be derived based on a given visit/specimen collection date, and the *entry start of evaluation date*

(i.e. the start date of the entry evaluation visit). **Therefore, the date of the first entry evaluations visit is day 0 (note: SVSDT from the entry visit).** In the event multiple results fall within the same analysis window, the one closest to the target time will be prioritized, or if equidistant, the earlier result will be prioritized.

SOE Visit	Protocol Range (days)	Analysis Range (days)
Screening	-45, -1	-45, -1
Entry	0 **(date of start of entry evaluations)**	[randomization, study entry visit *end* date]
Week 4	25, 35	21, **55**
Week 12	70, 98	**56, 124**
Week 24	154, 182	**125, 209**
Week 36	238, 266	**210, 287**
Week 48	322, 350	**288, 413**

Key study visits occur at entry, week 24, and week 48.

Entry: See above. These evaluations will also be referred to as baseline.

Week 24: The study visit at which key secondary outcomes will be evaluated.

Week 48: End of study treatment and last on-study visit.

There are three study arms under consideration in A5391, two investigational arms that involve switching from an INSTI-based regimen to DOR with or without a switch from TAF to TDF, and one control arm. Participants who are randomized to the control arm are expected to continue on their current ART regimen of INSTI+TAF/FTC (or TAF/3TC), where INSTI regimens may include BIC, DTC, or RAL.

Analysis of primary and secondary outcomes will be compared, separately, between each of the investigational switch arms and the control arm (remaining on current INSTI-based regimen). Each comparison will use a 2.5% type-I error rate to control for multiple comparisons. Secondary analysis will compare each of the primary and secondary outcomes between each of the switch arms using a 5% type-I error rate. No other adjustments for multiple comparisons will be made.

Continuous variables will be summarized using mean, standard deviation, median, interquartile range (Q1 and Q3), 10^{th} and 90^{th} percentile, and min and max; categorical variables will be summarized using frequency and percentage.

Analyses of occurrence outcomes (e.g. safety, tolerability) which are summarized across the trial sample as the proportion of participants who experienced the outcome of interest will include only those participants who were "at risk" for the outcome being observed during trial follow-up. In addition to excluding participants who never initiated study treatment (as those participants are taken off study follow-up), these also exclude participants who leave the study so early that the outcome of interest was not evaluated or measured. Details as to how these principles apply to specific outcomes are detailed below.

Analyses other than those related to screening and accrual will be utilizing the treated population analysis set **(safety) or eligible treated population set (efficacy).** This is due to protocol language that directs participants who never initiate treatment to be discontinued from trial follow-up. Any exceptions to this analysis approach are noted below.

5 Analysis Approaches

5.1 Analyses of the Primary Objective

The following table summarizes the primary objective and the associated estimand. Further details are provided after the table.

<table>
<tr><td colspan="3">Primary Objective: To assess if ART switch to DOR + TAF/FTC (or TAF/3TC) or to DOR + TDF/FTC (or TDF/3TC) results in differences in weight change over 48 weeks compared to a strategy of remaining on current ART (INSTI+TAF/FTC (or TAF/3TC)).</td></tr>
<tr><td>Estimand description</td><td colspan="2">Absolute difference (switch arm minus control arm) in the mean percent change in weight over 48 weeks, among adults living with HIV who have BMI ≥30 kg/m^2, while having maintained virologic suppression for the last year on an INSTI-based regimen.</td></tr>
<tr><td>Treatment</td><td colspan="2">DOR+TAF/FTC (or TAF/3TC), DOR+TDF/FTC (or TDF/3TC), or INSTI (BIC, DTG, or RAL)+TAF/FTC (or TAF/3TC)</td></tr>
<tr><td colspan="2">Target population</td><td>Analysis set (analysis population)</td></tr>
<tr><td colspan="2">Adults (≥ 18 years of age) living with HIV who have BMI ≥30 kg/m^2, on an INSTI-containing regimen + TAF/FTC (or TAF/3TC), and have maintained virologic suppression (HIV-1 RNA <50 copies/mL or below the lower limit of HIV-1 RNA detection available at the site) for at least 48 weeks prior to entry on this regimen.</td><td>Eligible Treated population</td></tr>
<tr><td colspan="2">Variable(s)</td><td>Outcome measure(s)</td></tr>
<tr><td colspan="2">Body weight (kg) at entry and week 48.</td><td>Change (percent) in body weight from entry to week 48.

The 48-week period will be based on the date of the week 48 visit, occurring between weeks 44 and 52 based on the analysis window.</td></tr>
<tr><td colspan="2">Handling of intercurrent events</td><td>Handling of missing data</td></tr>
<tr><td colspan="2">The following intercurrent events are relevant to the estimand:
- Death;
- Receipt of or change in medications known to cause changes in weight or interfere with DOR**;
- Change in randomized regimen;</td><td>Participants who discontinued follow-up before week 48 are considered ignorable will be excluded from analysis. Sensitivity analyses will consider participants who discontinue follow-up before week 48, using the last collected weight.</td></tr>
</table>

- Confirmed virology failure; - Pregnancy. Death is the only intercurrent event that will be accounted for in the primary analysis, other intercurrent events will be ignored according to a treatment policy strategy, but their impact will be assessed in supplementary analyses. For deaths, a *while alive strategy* is being taken to evaluate treatment effects. If deaths occur, the last observation taken before the death will be used to determine the variable.	Similarly, participants who having a missing baseline value will be missing for the change outcome to week 48. See sensitivity analysis below to address this contingency if it occurs.
Population-level summary measure	Analysis approach
Absolute differences (each switch arm minus control arm) in the mean percent change in weight from entry to week 48.	Absolute differences in the mean percent change in weight will be obtained using linear regression, with an indictor variable for treatment arm and adjustment for each stratification factor and weight at entry. This analysis will be done separately for each switch arm compared to the control arm.

** New Receipt of medications would include prohibited and precautionary medications outlined in the protocol and on the protocol specific webpage, changes to medications would include those outlined in the inclusion/exclusion criteria of the protocol. See AIP for list of medications.

Analysis Approach

The analysis of the primary outcome will compare the change (percent) in body weight (kg) from entry to week 48 between each switch arm and the control arm. The mean percent change in weight at week 48 will be compared separately between each of the investigational switch arms and the control arm using linear regression adjusting for stratification factors and weight at entry with a 2.5% type-I error rate; treatment arm will be included in each model as an indicator variable. The absolute difference in the mean percent change will be calculated, as well as a corresponding 97.5% confidence interval (CIs); p-values will also be provided. Participants who prematurely discontinue from the study, and therefore do not have a weight measurement at week 48 will be excluded from the primary analysis; the exception to this is if participants prematurely discontinue due to death, in which case their last weight measurement obtained prior to death will be used instead of the week 48 measurement. The primary analysis will be conducted according to an intent-to-treat principle, and therefore participants will be analyzed according to their randomized regimens, regardless if they change their ART regimen during follow-up (i.e. prematurely discontinue treatment).

Note: persons enrolled under the previous version of the protocol, with different target population, will be included in the primary analysis. Two supplementary analyses will be conducted (see below for more details):

1) Excludes persons enrolled who did not have screening BMI >= 30 kg/m2, and thus would not have qualified under version 2 of the protocol.
2) By subgroup defined by history of weight gain: those with documented weight gain as per version 1 requirements; those who did not have weight gain; those without historical weight gain information to meet the first 2 subgroups

Sensitivity Analyses – contingency

A. If any participants are missing their entry weight for change from baseline, a sensitivity analysis using, the screening weight used for trial eligibility will be conducted.

B. To evaluate the impact of missing data at week 48, controlled multiple imputation (MI) will be used only if more than 10% of week 48 weight values are missing overall or in one arm [1].

In this sensitivity analysis, missing data will be considered Missing at Random (MAR) and Missing Not at Random (MNAR). Under the MAR assumption, the conditional distribution of the unobserved data given the observed data (see below) will be same, regardless of whether the data were observed or not (i.e. missing week 48 weight values does not depend on the week 48 weight value); however, under the MNAR assumption, the probability a participant is missing their week 48 weight depends on what the value was of the unobserved week 48 weight (e.g. those who do not improve their weight trajectory may be more likely to drop out).

1. MAR Assumption

Standard MI will be conducted assuming data are MAR. The analysis model will be as defined for the primary analysis, and will adjust for treatment arm, weight at entry, and stratification factors. The imputation model used for MI will include all variables defined for the analysis model, and the addition of all weight measures observed prior to the participant being loss to follow up. Number of imputations will be determined by the percent of missing data, and will be assessed for precision. This analysis does not account for missing data pattern or reason for missingness.

2. MNAR Assumption

   To assess the impact of data being MNAR, we assume those missing week 48 weight outcomes have poorer responses than those with observed values at week 48 (i.e. those without an improved weight trajectory were more likely to drop out, or those on control arm more likely to drop out), and will utilize δ-based MI methods where δ corresponds to a fixed constant that will be added or subtracted to the MAR-imputed weight values at week 48, among those missing their week 48 weight. In this analysis, the choice of δ will be based on a percentage of the observed mean percent change in weight (entry to week 48), specifically will corresponds to 10%, 20%, 30%, 40% and 50% of this change in weight. This analysis will explore the impact of a range of fixed mean differences in the outcome for the unobserved data and evaluate the impact of the MNAR assumption on inference of the primary outcome. Note: Differences in observed effects are possible and expected in sensitivity analysis assuming data are MNAR.

The primary analysis will be considered robust to missing data, if overall interpretation of results are not impacted by the sensitivity analyses.

Supplementary Analyses

The following supplementary analyses are included to evaluate the impact of different assumptions on the inference of the primary outcome due to possible intercurrent events or other factors.

1. Evaluate impact of participants who, prior to week 48, used or changed medications known to cause changes in weight or that interfere with DOR as defined in the protocol (see AIP for list of medications), changed their randomized regimen (regardless of reason), had a confirmed virologic failure, or became pregnant.

   Approach: Repeat the primary analysis, but exclude participants who met the criteria above.

2. Evaluate impact of persons who were not obese (i.e. eligible as enrolled under version 1 but would not have been eligible per version 2 criteria)

   Approach: Repeat the primary analysis, but exclude those participants who enrolled under version 1 and whose screening BMI was < 30 kg/m2 (i.e. and thus would not have qualified under version 2.)

3. Evaluate impact of persons with historical weight gain in INSTI regimen.

Approach: See subgroup analysis below .

Supportive Analyses

Secondary outcome 1 (weight at wk 24) is included as supportive to the primary objective. Percent change in weight from entry to week 24 will be summarized in the same manner as the primary outcome. Sensitivity and supportive analyses noted above will also be examined.

In support of the primary analysis, comparing each switch arm to the control arm, an additional analysis will contrast the 48-week changes in weight between the switch arms (1 vs 2) to determine if there are additional benefits of a switch from TDF- to TAF-based treatment. The mean percent change in weight at week 48 will be compared between switch arms using the same methods outlined for the primary analysis, except a 5% type-I error rate will be used for this supportive comparison.

Subgroup Analyses

To evaluate the effects of switching to DOR from an INSTI-based ART regimen in specific populations, subgroup analyses will be conducted. Descriptive summaries of change (percent) in weight within each subgroups will be provided. The same approaches outlined for the primary analysis will be implemented for each subgroup. Within each **type of** subgroup, the difference between treatment arms will be determined, and compared between subgroups by a test of interaction and 97.5% CI and corresponding p-value. In the linear regression models, subgroup will be included as an indicator variable and an interaction term between treatment arm and subgroup will also be included; other model terms (weight at entry and stratification factors) will be included, as approproiate, as defined in the primary analysis. Pre-specified subgroups of interest include:

1. Sex (male sex at birth, female sex at birth) [stratification factor]
2. Gender (cis-female or trans-female on hormones at entry vs other)
3. Race (African American/Black vs non-African American/Black) [stratification factor]
4. Race/Ethnicity (African American/Black vs Hispanic vs other)
5. Age (<50, ≥ 50 years)
6. INSTI-regimen at entry (BIC, DTG, RAL)
7. History of weight gain in INSTI (yes, no, unknown/missing)

5.2 Analysis of Secondary Objectives

5.2.1 Analyses of the safety and tolerability objective

The outcomes for this objective includes participants who experience one of the following:

- Severity grade 3 or higher reportable AE
- > 10% reduction in calculated creatinine clearance (CRCL) compared to baseline
- Prematurely discontinue study treatment (as defined above)

Outcome calculation and descriptive analyses specifics:

The number and estimated proportion of participants experiencing each outcome will be summarized by arm. The denominator (i.e. at risk population set) will be the treated population analysis set with the following contingencies: for AEs -- the denominator will exclude any participants who left the study without any reportable AEs (including reportable AEs below grade 3), and before having any post-treatment initiation evaluations for AEs; for CRCL -- the denominator will exclude any participants who did not have any post-treatment initiation serum creatinine results available for CRCL calculation.

Only observed creatinine clearance values will be used to determine those participants meeting the CRCL reduction outcome. Screening values will be used as baseline if the entry value is missing; missing post-entry data will be ignored (and those assumed to have not met criteria for the outcome). All discontinuations of any component of ART regimen for any reason prior to the week 48 visit will be considered to have met the tolerability outcome. Timing and reasons for discontinuation will be summarized.

For each outcome, a summary of appropriate details about each outcome will also be included.

For example, the type, and severity of grade 3+ AEs will be summarized with MeDDRA categories, as well as summaries of which of these were related to study treatment, and the timing in relation to treatment initiation. For reductions in CRCL the distribution of % reduction among those meeting the outcome will also be given, as well as distribution of actual CRCL values meeting the outcome, and the distribution of timing of meeting the outcome post treatment initiation.

Hypothesis testing and Inference:

The proportion of participants will be compared between each switch (1 or 2) arm and the control (3) arm using mid-P Fisher's exact tests at the 2.5% significance level. As a supportive analysis, the two switch arms will be compared (1 vs. 2) using the same test, but at a 5% significance level.

2-sided, Wilson-based, exact score confidence intervals (coverage determined by significance testing in paragraph above) will be calculated for each estimated proportion.

5.2.2 Analyses of the maintenance of virologic suppression

The outcome for this objective is confirmed virologic failure defined as post-treatment and confirmed plasma HIV-1 RNA > 200 copies/mL. The denominator for the proportion summary statistic for this outcome will be based on the treated population analysis set minus any participants without any post-treatment initiation viral loads, and will use available HIV-1 RNA values for determining the numerator.

The proportion of participants meeting the outcome of virologic failure will be compared between each switch arm and the control arm using mid-p Fisher's exact tests at the 2.5% significance level. As a supportive analysis, the two switch arms will be compared (1 vs. 2) using the same test, but at a 5% significance level.

Corresponding 95% 2-sided Wilson-based exact score confidence intervals will be calculated for each estimated proportion with VF.

5.2.3 Analyses of minimum waist circumference and fasting cardiometabolic parameters

These outcomes will use methods parallel to those in the primary estimand, but with substituting for the appropriate variable measure. The following exceptions to this plan are as follows:

1) If entry/baseline lab measures are missing, then the change from baseline to outcome will also be missing as the fasting labs for these outcomes (glucose, insulin, HOMA-IR, triglycerides, LDL and HDL) were not scheduled to be collected at screening. Because it is assumed that missing baseline values (or samples to measure baseline values) will not be related to the outcomes (i.e MAR assumption), persons with missing data will be excluded, and contingencies for sensitivity analyses using MI as specified above will only be triggered if more than 20% of participants in a single arm or overall are missing outcomes at wk 48.

2) Supplementary analysis for triglycerides, LDL and HDL to adjust for lipid lower medication use during the study.

   This analysis will utilize subgroups based on reported lipid lower medication use during the study (yes/no), and use the subgroup analysis approach as outlined in the primary estimand section.

3) **The analysis set for insulin and HOMA-IR will exclude persons taking exogenous insulin (per concom meds) since these measurements are not interpretable among those taking insulin**.

5.2.4 Analyses of body composition by DEXA

Percent change in the following between baseline and week 48 will be calculated: total fat, lean mass, trunk fat, limb fat, appendicular lean mass.

The modeling and inference approach parallel to the primary estimand will be performed with the following considerations:

1) Persons missing DEXA at either baseline or week 48 will be excluded. Sensitivity analysis for missing data will not be performed.
2) A sensitivity analysis excluding persons whose DEXA measurements came from a different machine between baseline and week 48 will be performed.

5.2.5 Analyses of bone mass by DEXA

Percent change in the following bone mass measurements between baseline and week 48 will be calculated: total hip (left and right) and total lumbar spine (L1-L4)

The modeling and inference approach parallel to the primary estimand will be performed with the following considerations:

1) Persons missing DEXA at either baseline or week 48 will be excluded. Sensitivity analyses for missing data will not be performed.

2) A sensitivity analysis excluding persons whose DEXA measurements came from a different machine between baseline and week 48 will be performed.

Change in bone mass categorization from baseline to week 48 using minimum T-score at each time point among the following collection of Tscores: left &right femoral neck, left.&right total femur (hip), and lumbar spine (L1-L4). Missing subsets will be ignored (e.g. for example if DEXA only taken one side, then just available values will be used).

6 Analyses specific to interim monitoring

From section 2.5 above, there are a few key benchmarks for interim monitoring. To assure that observed rates (ie. proportions) reported do not include participants recently accrued but not yet with any post-randomization information, and thus not yet at risk for the monitoring outcomes such as virologic failure, treatment related AEs, loss to follow-up or treatment crossover; the following approach will be employed.

For virologic failure, participants are at risk for virologic failure once any post-treatment initiation HIV-1 viral load is available for analysis in the pushed clinical database (anticipated by ~ 8 weeks following randomization). For treatment related AEs, participants are at risk the earlier of a) first post-Rx clinic visit (initial visit scheduled at wk 4), or b) an AE is reported. For loss to follow-up, participants are at risk the earlier of a) 8 weeks lapsed since randomization, or b) person has prematurely discontinued from the study for any reason. For treatment crossover, participants are at risk once a post-treatment initiation clinic visit has occurred and is recorded in the clinical database pushed to the statistical center.

7 Report Contents

Report sections (outline)	**Supporting tables, figures and listings**
Background (study design, objectives, monitoring plan and history, trial milestones and updates); Statistical methods (sample size, analysis sets, any changes from specifications in SAP & AIP, hypothesis testing and inference summary)	CONSORT flow diagram for RCT
Screening (**interim** reports only) and Accrual	Summary of reasons for screen failure (interim only) Accrual tables by site, time, and stratification factors

Baseline characteristics of trial sample (overall and by assigned arm) *italics* included in clinicaltrial.gov record	Demographics: *age, sex, gender identity*, race and ethnicity *(DAIDS categories), and country* Body composition: *weight, BMI, waist circumference*, history of weight gain on INSTI Food security and Physical exercise questionnaires HIV disease status: ART regimen at screening, HIV-1R RNA , CD4 cell count Laboratory values (see AIP for list) Targeted concomitant medications (see protocol section 5.4) related to weight change Selected medical history (reportable diagnoses, signs/symptoms, other non-targeted medications)
Study Conduct (Participant and Study Treatment Disposition) *italics* for clinicaltrials.gov results record Participant self-report of Study treatment adherence Monitoring benchmarks (interim reports only)	Study status: *reasons,* timing *for premature study discontinuation* Study treatment disposition: timing of initiation, reasons for not initiating (if applicable), reasons and timing for premature discontinuation (*received study treatment, completed and not completed study treatment),* reasons and timing for changes Summary of 4-day recall of study treatment adherence by visit Monitoring benchmarks *(interim only)*
Targeted Concomitant Medications (section 5.4)	Summary of initiations or changes to pre-existing use during study follow-up
Adverse Events (standardized reports of *italicized* elements included in clinical trials.gov results record)	*Summary of New AEs by severity grade (grouped by Primary SOC and PT)* *Summary (*Listing of) *SAEs*, grade 4 AEs; AEs related to study treatment

	Listings of targeted events: deaths, pregnancies and virologic failures

Note: For each estimand in the Primary SAP, include summaries of the number and timing of each intercurrent event in each treatment group.

- *Analysis of primary outcome measures (reported in ClinicalTrials.gov)*
- *Analysis of secondary outcome measures (reported in ClinicalTrials.gov)*
- *Anything else relevant to the study (e.g., subgroup analyses, association analyses of primary, secondary or other outcome measures)*

8 Citations

1. Cro S, Morris TP, Kenward MG, Carpenter JR. Sensitivity analysis for clinical trials with missing continuous outcome data using controlled multiple imputation: a practical guide. Statistics in medicine. 2020 Sep 20;39(21):2815-42.

9 Associated Documents

Attachment 1: Writing Team Roster

Redacted.