This is ACTG A5403 SAP Version 1.0 with names of authors, names of publication writing team members and analysis timeline redacted.

Approvals

ACTG A5403

Primary Statistical Analysis Plan

Version 1.0

October 11, 2023

Giving Standardized Estradiol Therapy In Transgender Women to Research Interactions with HIV Therapy: the GET IT RIgHT Study

ClinicalTrials.gov Identifier: NCT06005610

Protocol Version 2.0 (June 30, 2023)

Created by:

[redacted]

Statistical and Data Analysis Center
Harvard T.H. Chan School of Public Health

Table of Contents

Version History

Version	Changes Made	Date Finalized
1	Original Version for version 2.0 of the protocol document	October 11, 2023

1 Introduction

1.1 Purpose

This Primary Statistical Analysis Plan (SAP) describes the primary and key secondary estimands and other secondary outcome measures that will address specific study objectives of the A5403 trial. The Primary SAP includes general analytic approaches for all primary estimands, key secondary estimands, 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.

The Primary Analysis Implementation Plan (AIP) provides detailed outlines of tables, figures, and coding descriptions.

A separate SAP will provide outlines of analyses for other objectives and outcome measures not included in the Primary SAP.

1.2 Version History

Not applicable; original version.

2 Study Overview

2.1 Overview of Study Design

A5403 is a 48-week, open-label, non-randomized, multicenter trial to evaluate both ART and 17-β estradiol exposure in adult transgender women (TW) and other individuals identifying as female or transfeminine but with male sex assigned at birth (henceforth referred to in aggregate as TW) living with HIV-1 (PWH). The trial aims to enroll 90 participants, or 30 per ART group, who are not currently on feminizing hormone therapy (FHT) and virologically suppressed on one of 3 targeted ART regimens:

Group 1: BIC-treated group: Participants taking bictegravir (BIC) + tenofovir alafenamide (TAF) + emtricitabine (FTC); n=30

Group 2: DTG-treated group: Participants taking dolutegravir (DTG) once daily + tenofovir disoproxil fumarate (TDF) + (FTC or lamivudine [3TC]); n=30

Group 3: Boosted DRV-treated group: Participants taking any regimen containing darunavir plus cobicistat (DRV/c); n=30

All participants will continue on ART and receive study-supplied 17-β estradiol for FHT for study weeks 0-48. Oral 17-β estradiol 2 mg once daily will be initiated at study entry. At weeks 4, 12, 24, and 36, study clinicians may titrate 17-β estradiol in 2 mg increments to achieve the desired participant goals and target hormone concentrations, as measured locally at each visit. Participants will not be required to have their estradiol dose titrated. Titration above the recommended therapeutic range will not be permitted for safety reasons. De-escalation for tolerability or toxicity is allowed.

Steady-state antiretroviral and steady-state 17-β estradiol PK will be assessed throughout the study using intensive and sparse PK sampling. Sparse PK sampling will occur at entry and weeks 4, 12, 24, 36, and 48 within 22-26 hours after last dose and before daily ART and 17-β estradiol administration. Fifteen participants from each ART group (who opt in via informed consent) will undergo 8 hour intensive PK sampling at weeks 0, 24 and 48 (or premature discontinuation); samples will be collected 30 minutes prior to ART and 17-β estradiol dosing, and then 1, 2, 3, 4, 6, and 8 hours after dosing.

All trial outcomes are defined for the 48-week period following oral 17-β estradiol initiation. The primary completion date (PCD) for the study is when the final participant enrolled is evaluated at week 48 for the primary outcomes.

2.2 Hypotheses

2.2.1 Primary Hypotheses

1. [Hypothesis 1.1.1] Across oral 17-β estradiol doses from 2-10 mg daily, there will not be clinically significant reductions in ART exposure.

2. [Hypothesis 1.1.2] Over 48 weeks, participants receiving DRV-containing ART will have lower 17-β estradiol exposure than participants receiving DTG- or BIC-containing regimens.

2.2.2 Secondary Hypotheses

1. [Hypothesis 1.2.1] Adverse events related to 17-β estradiol will not differ by ART regimen but may differ by estradiol dose.

2. [Hypothesis 1.2.2] Participants receiving DRV-containing ART will require higher doses of 17-β estradiol to achieve similar feminizing hormone therapy (FHT) satisfaction as participants receiving BIC- or DTG-containing ART.

3. [Hypothesis 1.2.3] Virologic effectiveness will be maintained in each ART regimen plus FHT over 48 weeks.

4. [Hypothesis 1.2.4] Higher 17-β estradiol concentrations will be associated with greater weight gain and greater perturbations of lipids and insulin/glucose.

5. [Hypothesis 1.2.5] The relationship between 17-β estradiol dose and serum concentrations will reflect a dose proportional relationship.

2.2.3 Exploratory Hypotheses (to be addressed in Primary Analysis)

1. [Hypothesis 1.3.3] 17-β estradiol administration will result in changes in sex hormone profiles (e.g., lower total and free testosterone, higher sex hormone binding globulin (SHBG) and free estradiol).
   Note: Only testosterone will be addressed in the primary analysis.

2.3 Study Objectives

This Primary SAP addresses the following primary and secondary objectives in the study protocol. Other study objectives in the protocol will be addressed in subsequent analysis plans.

Analysis of the study objectives below will be analyzed under either superiority or equivalence framework, depending on the hypothesis for the objective. However, inference will focus on estimation of relevant group/effect differences and confidence interval about these estimates, rather than hypothesis testing (i.e. p-values).

While the plan is to conduct the primary (final) analysis once the last participant has completed the week 48 study visit, all queries have been resolved, and the routine process for final analysis preparation of the database has completed, the following is an alternate plan under the contingency that one (or more) ART cohort completes accrual and follow-up far in advance of the others. In this case, objectives which do not require inclusion of the ART cohort still under accrual or in follow-up (e.g. objectives which include analyses within ART cohort) may undergo final analysis and public presentation once the data to support those analyses have completed quality control and assurance processes and are considered final. In this contingency, the final analysis will occur in 2 stages.

2.3.1 Primary Objectives

1. [Objective 1.4.1] To estimate trough concentrations of received ARVs before versus with each dose of oral 17-β estradiol (2-10 mg), both in presence and in absence of concomitant anti-androgen use.

2. [Objective 1.4.2] To compare the trough serum estradiol concentrations achieved, at each 17-β estradiol doses, between groups receiving boosted versus non-boosted ART.

2.3.2 Secondary Objectives

Only those objectives as part of the primary analysis are included here.

1. [Objective 1.5.1] To compare occurrence of FHT-associated adverse effects across ART groups and estradiol doses.

2. [Objective 1.5.2] To summarize intervention satisfaction across ART groups.

3. [Objective 1.5.3] To compare virologic suppression within ART groups, at each received 17-β estradiol dose over 48 weeks.

4. [Objective 1.5.4] To compare the PK parameters (AUC, Ctrough, Cmax, Tmax) of 17-β estradiol over 48 weeks in a subset of TW undergoing intensive PK between groups receiving boosted versus non-boosted ART.

5. [Objective 1.5.5] To assess changes in weight, anthropometrics, lipids, glucose, and insulin sensitivity in relationship to estradiol concentrations.

6. [Objective 1.5.6] To assess the changes in estradiol concentrations across each 17-β estradiol dosage and ART group.

2.3.3 Exploratory Objectives (to be addressed in Primary Analysis)

1. [Objective 1.6.3] To describe changes in sex hormones following 17-β estradiol administration. Note: Only testosterone will be address in the primary analysis.

2.4 Overview of Sample Size Considerations

The total sample size accrual goal is 90 participants, 30 per ART group. However, as analyses are specified at each dose of oral 17-β estradiol, and titration of estradiol dose over time is anticipated, at doses higher than the initial 2mg QD, the effective sample size will likely be smaller than the enrolled sample.

Summary of ART Primary PK Objective Sample Size Calculations

For the primary objective of evaluating changes in trough ART concentrations, a fixed sequence design and an equivalence hypothesis will be used. Therefore, two-one sided tests (TOST) on the t-test and pre-specified no effect bounds for the geometric mean ratio (GMR) can be used to calculate statistical power.

The sample size was derived using assumptions on the standard deviation of differences in natural log-transformed trough concentrations (i.e., transformed geometric ratios). Assuming a range of inter-person coefficient of variation (CV) of Ctrough of ART of 0.352 to 0.47, and a range of hypothesized correlation between Ctrough within the same participant of 0 to 0.5, the calculated standard deviation of intra-person difference in log Ctrough ranges from 0.34 to 0.63.

Assuming the true (alternative) GMR of 1 (i.e. equivalence), a significance level (alpha) of 0.05, and an effective sample size of 20 participants per ART group, there is >80% power for equivalence under no-effect bounds of (0.7, 1.43) for standard deviation of <0.50.

See protocol section 10.4.1 for more details regarding sample size and power calculations.

Summary of Estradiol Primary PK Objective Sample Size Calculations

For the primary objective of comparing serum estradiol between groups defined by ART regimen at each dose of oral 17-β estradiol, a parallel groups PK design and superiority hypothesis comparing the boosted DRV ART group against the other 2 ART groups will be used.

Assuming a between participant CV of Ctrough of serum estradiol = 0.60 (i.e., highly variable), a significance level of 0.05, and a similar effective sample size in ART groups across increasing estradiol doses, as the effective sample size decreased from 30 in boosted ART group vs 60 non-boosted ART group to 15 vs 30, the alternative GMR for serum estradiol that can be detected with adequate power (≥80%) decreases from 0.70 to 0.60.

See protocol section 10.4.2 for more details regarding sample size and power calculations.

Sample Size for Intensive PK Subsample

Fifteen participants from each ART group (who opt in via informed consent) will undergo intensive PK sampling. Participants schedule for intensive PK sampling who are lost to follow up can be replaced to ensure a minimum of 15 participants per ART group.

The approach for sample size calculation was similar to the primary objective for PK, but utilized intensive sampling for estimation of estradiol PK parameters of Cmax, Tmax, and AUC. Assuming a CV of 0.27, a sample size of 15 has 80% power that a 2-sided, 95% CI on the estradiol PK parameter has a half-width of 0.168.

See protocol section 10.4.3 for more details regarding sample size and power calculations.

Accrual goals for race/ethnicity

The trial aims to enroll at least 50% of participants identifying as non-white or Latine. Participants who do not provide race or ethnicity will be grouped as white/non-Latine. Participants who identify as multiple races/ethnicities and endorse either non-White or Latine will be grouped as non-white/Latine.

2.5 Overview of Formal Interim Monitoring

The study will undergo interim feasibility and safety review at least annually by an ACTG-appointed Study Monitoring Committee (SMC) (as per appointment via the ARTS TSG).

Timing: The first interim review will occur no more than 12 months after enrollment of the first participant or 25 participants complete 24 weeks of follow-up (in order to have conduct and safety data on at least two possible estradiol dose titrations), whichever occurs earlier. Subsequent reviews will occur approximately annually, unless recommend otherwise by the SMC.

Triggers: Additional SMC review may be triggered by the following feasibility triggers:

- if accrual at 18 months after first enrollment is less than 50% of the total expected accrual (e.g., fewer than 45 of anticipated 90 participants across all groups)
- if greater than 25% of currently-accrued participants (if at least 30 participants have been enrolled) are lost or withdrawn from study follow-up.

An expedited safety review by the SMC for potential trial modification will also be triggered by any of the following:

- **Any death that is not clearly attributed to a cause other than the study product (e.g., accident, trauma)**
- **2 or more Grade 3 or higher, treatment-related adverse events**
- **2 or more Grade 2 or higher allergic/hypersensitivity reactions related to study treatment**
- **2 or more of the same Grade 3 or higher, unexpected, treatment-related adverse events (and screening and accrual will be held until the SMC expedited safety review has occurred).**

An interim review may also be convened if a concern is identified by the DAIDS clinical representative, the study chairs, or study statistician in consultation with the team.

Scope: Any/all SMC reviews will include administrative/trial conduct data (accrual, study treatment disposition, adherence to ART and study treatment, realization of estradiol titration, and study disposition) as well as safety information (reported AEs and SAEs), including safety-related outcomes (protocol defined virologic failure and targeted events). See SPDSMP for details regarding SMC reports.

Monitoring guidelines:

Safety: It is recommended that the trial be stopped in the case of any death that is caused by study treatment, and that trial modification (including stopping) should be considered in the case of unexpected serious illness (i.e., SAE) caused by the study treatment. Protocol Section 7.4 also lists AE triggers for expedited safety review, which may also result in trial modification following SMC review.

Efficacy: Because there are no plans to initiate batched centralized PK testing prior to study completion, there are no plans for inclusion of early information on either of the primary outcomes (ART or estradiol). Locally measured estradiol concentrations will be available to sites for clinical management of FHT, but will not be summarized for SMC review due to inter-assay variability between laboratories. The realization of titration will provide information on how these local hormone concentrations were used in trial conduct.

3 Outcome Measures

All primary and secondary outcome measures listed below will have results submitted to ClinicalTrials.gov. All outcomes are defined for the 48-week period following oral 17-β estradiol initiation; therefore, all results will be reported in a single results submission, one year after PCD. See AIP for details on ClinicalTrails.gov submission.

3.1 Primary Outcome Measure(s)

1. **[Outcome 10.2.1.1]** Trough concentration of the analytes BIC, DTG, and DRV in plasma at each received dose of oral 17-β estradiol (0 (pre-FHT or baseline), 2, 4, 6, 8, 10 mg) summarized at the participant level as the following two outcomes:
    - **a.** Intra-person geometric ratios of trough ART concentration at each received estradiol dose relative to pre-FHT (i.e., baseline).
    - **b.** Indicator of concentration being above drug-specific externally defined threshold, based on published estimates of each drug's protein bound (or adjusted) effective concentration 90/95 (i.e. $PAEC_{90}$ or $PBEC_{95}$).

    Log-transformed trough concentrations (Ctrough) in ng/mL from plasma samples collected within 22-26 hours after the last 17-β estradiol dose, measured over 48 weeks. Analyte concentrations < LLoQ at entry and on study imply adherence issues and may be used to inform the PK analysis set.

    If multiple observations are available at the same estradiol dose, the first sampled qualifying steady-state trough concentration (based on calendar time) taken will be used. Other methods will be outlined as sensitivity analyses in the analysis approach below.

    Intra-person geometric ratios will be calculated as the log of Ctrough at week x – log of Ctrough at baseline.

    Drug-specific externally defined thresholds will be defined from the published estimates of either protein bound, or protein adjusted effective concentrations (at either 90 or 95 boundary), also known as $PAEC_{90}$ or $PBEC_{95}$, as modeled from invitro studies.

2. **[Outcome 10.2.1.2]** Trough serum total estradiol assessed at each received dose of oral 17-β estradiol (i.e., 2 mg, 4 mg, 6 mg, 8 mg, 10 mg), as quantified via batch testing at central lab.

    Log-transformed total 17-β estradiol Ctrough in ng/mL from serum samples collected within 22-26 hours after the last dose of ART and 17-β estradiol, measured over 48 weeks. Results < LLoQ at entry will be imputed as 0 ng/mL.

    If multiple observations are available at the same estradiol dose, the first qualifying trough concentration (based on time since last dose) taken will be used. Other methods will be outlined as sensitivity analysis in the analysis approach below.

3.2 Secondary Outcome Measures

1. **[Outcome 10.2.2.1]** Trough concentration of the analytes TFV-DP, FTC-TP, and 3TC-TP in non-viable PBMCs at each received dose of oral 17-β estradiol (0 (pre-FHT or baseline), 2, 4, 6, 8, 10 mg) summarized at the participant level as the following two outcomes:
    - **a.** Intra-person geometric ratios of trough ART concentration at each received estradiol dose relative to pre-FHT (i.e., baseline).
    - **b.** Indicator of concentration being above drug-specific externally defined threshold.

    Analyte trough concentration from non-viable PBMC samples will use parallel methods to those outlined for Primary Outcome #1.

2. **[Outcome 10.2.2.2]** Occurrence of any reportable adverse event deemed **related** to 17-β estradiol over 48 weeks.

    The following language defines reportable adverse event per Section 7.2 of the protocol.

    Post-entry and while on study treatment, all signs (including an abnormal laboratory finding), symptoms, or diagnoses new in onset or aggravated in severity or frequency from the baseline condition, indicated as related to 17-β estradiol by the site (or the CMC if adjudication differs from that of the site investigator), and meeting any of the following criteria:

    - ≥ Grade 3
    - AE leading to estradiol interruption, dose reduction or permanent discontinuation , regardless of severity grade
    - ≥ Grade 2 cholecystitis, elevated liver enzymes, or hypertension
    - ≥ Grade 1 lipid or glucose abnormalities and worsening by at least 1 grade since baseline
    - Coronary heart disease or other cardiovascular disease, cancer (exclusive of basal/squamous cell skin cancer), diabetes/pre-diabetes, or any vascular event (see below for listing), regardless of grade
    - Met SAE definition (see below) or expedited AE reporting requirements

    The Division of AIDS Table for Grading the Severity of Adult and Pediatric Adverse Events (DAIDS AE Grading Table), corrected Version 2.1, July 2017, and the DAIDS EAE Manual Version 2.0 will be used.

3. **[Outcome 10.2.2.3]** Occurrence of SAEs and new events of coronary artery disease (CAD)/Cardiovascular disease (CVD), cancer (exclusive of basal/squamous cell skin cancer), diabetes mellitus (DM)/pre-DM, and vascular event, defined in protocol section 7.2, over 48 weeks.

    Post-entry and while on study treatment, all events listed in outcome measure and defined below.

    As per ICH guidelines, a SAE is defined as any untoward medical occurrence that results in death, is life-threatening, requires inpatient hospitalization or prolongation of existing hospitalization, results in persistent or significant disability/incapacity, is a congenital anomaly/birth defect, or is an important medical event that may not be immediately life-threatening or result in death or hospitalization but may jeopardize the patient or may require intervention to prevent one of the other outcomes listed.

Vascular events include arterial events (strokes and myocardial infarctions) or deep venous thrombotic events (venous thromboembolism, deep venous thrombosis, retinal vein thrombosis, and pulmonary embolism).

4. **[Outcome 10.2.2.4]** Indicators of testosterone suppression at each received dose of oral 17-β estradiol, defined as any serum total testosterone <50 ng/dL.

   Testosterone will be collected and stored at study entry and weeks 4, 12, 24, 36, and 48 prior to changes in estradiol dose. Stored samples will be tested for total testosterone batched and retrospectively.

   Results < LLoQ will be considered to be <50 ng/dL.

   If multiple observations are available at the same estradiol dose, the last testosterone concentration (based on calendar time) taken will be used. See analysis plan for sensitivity analyses using repeated measures or other approaches.

5. **[Outcome 10.2.2.5]** Indicators of virologic suppression of HIV, defined as plasma HIV-1 viral load <50 copies/ml, measured over 48 weeks.

   HIV-1 viral load will be collect at study entry and weeks 12, 24, 36, and 48.

6. **[Outcome 10.2.2.6]** Changes in overall transgender congruence score from baseline, measured over 48 weeks.

   The FHT Satisfaction Survey will be administered at entry and weeks 24 and 48.

   The overall transgender congruence score is calculated from participant response to the 12-question Transgender Congruence Scale (TCS). Participants rate each item on a 5-point Likert-type scale (i.e., 1 = strongly disagree, 2 = somewhat disagree, 3 = neither agree nor disagree, 4 = somewhat agree, 5 = strongly agree). Questions '*The way my body currently looks does not represent my gender identity.'*, *'I do not feel that my appearance reflects my gender identity.'*, and *'I am not proud of my gender identity.'* are reversed scored. The overall score is the average of the response to the 12 questions, with higher scores indicating a higher level of congruence [1].

   Positive changes from baseline indicate improvement in transgender congruence.

   Absolute change from baseline will be summarized.

7. **[Outcome 10.2.2.7]** Area under the curve over 8 hours (AUC 0-8h) of estradiol based on intensive sampling among a subset of participants.

   Whole blood samples will be collected pre-dose (within 30 minutes prior to ART and 17-β estradiol dosing), and then 1, 2, 3, 4, 6, and 8 hours after dosing at entry and weeks 24 and 48.

Standard noncompartmental techniques will be used to determine these PK parameters using the software package Phoenix WinNonLin (Certara®). The AUC will use the linear up/log down version of the trapezoidal rule. This version of the trapezoidal rule uses linear interpolation between untransformed data up to Cmax, and between log-transformed data from Cmax through Clast.

8. **[Outcome 10.2.2.8]** Absolute and percent changes in weight and BMI from baseline, measured over 48 weeks.

   Weight and height will be measured at study entry, and weight will be measured at weeks 4, 12, 24, 36, and 48. First available height will be used to calculate BMI for all time points.

9. **[Outcome 10.2.2.9]** Absolute and percent changes from baseline in minimum waist and maximum hip circumference and waist-hip ratio measured over 48 weeks.

   Waist and hip circumference, using standardized measuring procedures, will be collected at study entry and weeks 4, 12, 24, 36, and 48.

10. **[Outcome 10.2.2.10]** Absolute changes in fasting lipids from baseline (e.g., direct LDL, total cholesterol, triglycerides, HDL), measured over 48 weeks.

    Fasting lipid panel will be performed at study entry and weeks 12, 24, and 48.

11. **[Outcome 10.2.2.11]** Absolute changes in glucose and insulin sensitivity (HOMA-IR) from baseline, measured over 48 weeks.

    Glucose will be measured at study entry and weeks 4, 12, 24, 36, and 48. Insulin will be measured from stored samples, which will be collected at study entry and weeks 4, 12, 24, 36, and 48. Insulin sensitivity will be calculated as:

    HOMA-IR = (fasting glucose in mmol/L * fasting insulin in μU/L) /22.5 [2]

3.3 Exploratory Outcome Measures

1. **[Outcome 10.2.3.5]** Other PK parameters (e.g., Cmax, Tmax) of estradiol based on intensive sampling, among a subset of participants.

   Whole blood samples will be collected pre-dose (within 30 minutes prior to ART and 17-β estradiol dosing), and then 1, 2, 3, 4, 6, and 8 hours after dosing at entry and weeks 24 and 48.

   The PK parameters for 17-β estradiol from intensive sampling include:

   - Maximum concentration (Cmax)
   - Time to maximum concentration (Tmax)
   - Minimum concentration (Cmin)
   - Time to minimum concentration (Tmin)
   - Trough concentration (Ctrough)
   - Oral clearance (CL/F)

- Volume of distribution (Vd)
- Elimination half-life (T1/2)

Standard noncompartmental techniques will be used to determine these PK parameters using the software package Phoenix WinNonLin (Certara®). Cmax will be taken as the maximum observed concentration. Tmax is the time at which Cmax occurs. Ctrough will be taken as the pre-dose measurement. Apparent oral clearance will be calculated as CL/F. The T1/2 will be determined using regression analysis when possible along with other estimates of exposure as deemed appropriate based on review of the data.

4 General Considerations

4.1 Analysis Populations

Safety Analysis Population: Participants who initiate 2 mg of oral 17-β estradiol.

PK Analysis Population (As-Treated Analysis): Participants who initiate 2 mg of oral 17-β estradiol with no evidence of non-adherence or misuse of either ART regimen or oral 17-β estradiol from any source (e.g., including but not limited to self-report or objectives measures of adherence, such as undetectable drug levels, or estradiol levels indicative of extra-trial dosing such as PK results) around the time of PK samples drawn. Participants will be included in any week/dose analysis for which they were adherent proximal to sampling. Analyses will be as-treated, based on the received dose of estradiol, rather than the prescribed dose. Analyses will also be as-treated with respect to changes in ART regimen.

4.2 Definition of Visit Windows for Analysis

Clinic visits are scheduled to take place at screening, pre-entry (if needed), entry/day 0, and at weeks 4, 12, 24, 36, and 48. Schedule visit windows are ±7 days for week 4, and ±14 days for all subsequent visits.

Analysis visit windows will be formed around each study visit using the midpoints between adjacent weeks as cutoffs. See details in Primary AIP.

Baseline: Participants are directed to initiate 17-β estradiol within 3 days of study entry. Therefore, baseline is defined as the last evaluation prior to the first dose of 17-β estradiol.

4.3 Analysis Approaches

All analyses will use log-transformed concentrations of drug or hormones or PK parameters calculated from those concentrations of drug or hormones in order to utilize normality assumptions in the analysis. PK concentrations and parameters will be summarized by geometric means and CVs. Other descriptive statistics will include key percentiles.

<u>No-Effect Bounds</u>

For the primary PK analyses for ART levels, the 90% confidence interval about each GMR will be calculated and compared to no-effect reference interval of [0.7, 1.43]. The location and coverage of the confidence interval will be the focus of inference, rather than the p-value from a hypothesis test. A reference interval wider than the conservative (0.8,1.25) bioequivalence interval was chosen due to the higher variability of Ctrough parameter (versus other parameters like AUC that combine multiple measurements). This approach has been used in other trials when the PK parameter has exhibited high variability.

Inference will focus on estimation of relevant group/effect differences and confidence interval about these estimates, rather than hypothesis testing (i.e. p-values).

<u>Sources of Bias</u>

For the primary ART PK analysis, participants serve as their own control, and thus this analysis is not subject to bias due to absence of randomization.

For the primary estradiol PK analysis, comparisons between analysis groups are subject to bias based on both of the following:

1) confounding by prognostic factors for serum total estradiol which may be imbalanced between analysis groups not allocated via randomization in this trial, and

2) confounding by indication via the estradiol titration scheme.

While the former source of bias may be mitigated via calculating adjusted GMR for known and measured prognostics factors which have large imbalances between analysis groups (supplementary analysis #1), the later bias is more difficult to tackle. Importantly, confounding by indication is not anticipated to occur or impact the analysis and interpretation at the same rate throughout the trial. For instance, because everyone is started at the 2 mg dose and evaluated at week 4 at this dose before implementing the titration scheme, confounding by indication does not apply for this estradiol dose. It is anticipated that this bias may be minimal at the lower doses of estradiol, which represent the earlier phases of the titration scheme (e.g., 4 mg/day and potentially 6 mg/day). Therefore, this issue is of more concern at the higher doses of estradiol (6-10 mg/day), where it could substantively impact the results. Therefore, the interpretation of comparisons at higher doses, as well as interpretation of results at lower versus higher doses, will require caution and acknowledgment of this issue.

5 Estimands and Estimation

5.1 Primary Estimands

5.1.1 Primary ART PK Estimand: Intra-person GMR

<table>
<tr><td colspan="2">Primary Objective 1: To estimate trough concentrations of received ARVs before versus with each dose of oral 17-β estradiol (2-10 mg), both in presence and in absence of concomitant anti-androgen use.</td></tr>
<tr><td>Estimand description</td><td>Among adult TW-PWH who are virally suppressed on a targeted ARV regimen and who start or restart oral 17-β estradiol at 2 mg/day and whose dose may be titrated (and anti-androgen may be added after 24 weeks) as informed by safety, hormone levels and participant wishes; the intra-person geometric mean ratios of trough concentrations of each applicable ARV analyte (BIC, DTG, DRV) measured from plasma at each received dose of 17-β estradiol compared to without estradiol.</td></tr>
<tr><td>Treatment</td><td>Oral 17-β estradiol for 48 weeks, initiated at 2 mg/day and titrated as informed by safety evaluations, target hormone levels and participant wishes 4, 12, 24 and 36 weeks later, with allowed increases at each titration of 2 mg. After 24 weeks, anti-androgen may be added.</td></tr>
<tr><td>Target population</td><td>Analysis set</td></tr>
<tr><td>Adult persons with HIV who were assigned male sex at birth, who are virologically suppressed on a targeted ART regimen, and who wish to take feminizing hormone treatment for the purposes of developing or enhancing female secondary sex characteristics.</td><td>PK Analysis Population (see Section 4.1)</td></tr>
<tr><td>Variable(s)</td><td>Outcome measure(s)</td></tr>
<tr><td>Trough plasma concentrations of applicable target ARV analyte (BIC, DTG, DRV) measured from plasma samples before FHT and at each dose of 17-β estradiol</td><td>Intra-person geometric ratios of trough ART analyte concentration at each received 17-β estradiol dose relative to pre-FHT. See Outcome 10.2.1.1a for details.</td></tr>
<tr><td>Handling of intercurrent events</td><td>Handling of missing data</td></tr>
<tr><td>1) Premature stopping of estradiol due to AE: while on treatment
2) Premature stopping of estradiol due to lack of efficacy: while on treatment
3) Use of other/alternative FHT (e.g. street hormones): while on treatment
4) Use of estradiol at dose other than prescribed by titration: treatment policy
5) Change in background ART regimen: while on treatment</td><td>Based on target population, participants taking estradiol, there are no missing data due to lost to follow up or study discontinuation.

Missing data due to lab errors is ignorable because it is non-informative, and rare.</td></tr>
<tr><td>Population-level summary measure</td><td>Analysis approach</td></tr>
<tr><td>At each dose of 17-β estradiol, geometric mean ratios (GMR) of trough concentrations of each received target ART (BIC, DTG, DRV) with versus without feminizing hormone treatment, bounded by 90% confidence intervals.</td><td>Fixed sequence design where participants are their own control (i.e. geometric ratios calculated for each person, and mean and CI calculated across participants). Each ART group is analyzed separately.

The 90% Confidence interval about each GMR will be calculated and compared to no effect reference interval of [0.7, 1.43]

If multiple observations are available at the same estradiol dose, the first sampled qualifying steady-state trough concentration (based on calendar time) will be used. Choice of sensitivity analysis 1 or 2 will be based on prevalence of replicates.</td></tr>
</table>

	Sensitivity Analysis: If multiple observations are available at the same estradiol dose, the minimum of qualifying trough concentrations sampled at steady-state will be used.

Supplementary Analyses

1. Proportion of persons whose trough concentration of ARV (BIC, DTG, DRV) is below a specified reference level at a particular dose of oral 17-β estradiol
   Outcome Measure [10.2.1.1b]: Indicator of concentration being above drug-specific threshold, defined from the drug specific $PAEC_{90}$ or $PBEC_{95}$ from published invitro studies.
   Population-level Summary Measure: Proportion of people with ARV trough concentration below the drug specific threshold as defined above, at a particular dose of oral 17-β estradiol, bounded by exact 95% confidence interval calculated using Wilson-score method.

2. By subgroups, where subgroups defined by
   a. anti-androgen use at/around the time of PK sampling.
   b. Naïve vs. restarting FHT

Trough Concentrations of TFV-DP, FTC-TP, and 3TC-TP [Secondary Outcome 10.2.2.1]

Analysis of TFV-DP, FTC-TP, and 3TC-TP trough concentrations will use parallel methods to those described immediately above.

5.1.2 Primary 17-β Estradiol PK Estimand

Primary Objective 2: To compare the trough serum estradiol concentrations achieved, at each 17-β estradiol doses, between groups receiving boosted versus non-boosted ART.	
Estimand description	Among adult TW-PWH who are virally suppressed on a targeted ARV regimen and who start or restart oral 17-β estradiol at 2 mg/day and whose dose may be titrated (and anti-androgen may be added after 24 weeks) as informed by safety, hormone levels and participant wishes; the geometric mean ratios of trough concentrations of total serum estradiol, at each dose of 17-β estradiol, compared between those whose ART regimen is boosted versus not boosted.
Treatment	Oral 17-β estradiol for 48 weeks, initiated at 2 mg/day and titrated as informed by safety evaluations, target hormone levels and participant wishes 4, 12, 24 and 36 weeks later, with allowed increases at each titration of 2 mg. After 24 weeks, anti-androgen may be added.

Target population	Analysis set
Adult persons with HIV who were assigned male sex at birth, who are virologically suppressed on a targeted ART regimen, and who wish to take feminizing hormone treatment for the purposes of developing or enhancing female secondary sex characteristics.	PK Analysis Population (see Section 4.1)
Variable(s)	**Outcome measure(s)**
Trough concentration (Ctrough) of centrally- and batch-tested total serum estradiol measured at end of daily dosing interval at each received dose of oral 17-β estradiol (range 2, 4, 6, 8, 10 mg/day).	Outcome measure as defined by the variable. See Outcome 10.2.1.2 for details.
Handling of intercurrent events	**Handling of missing data**
1. Premature stopping of estradiol due to AE: while on treatment 2. Premature stopping of estradiol due to lack of efficacy: while on treatment 3. Use of other/alternative FHT (e.g. street hormones): while on treatment 4. Use of estradiol at dose other than prescribed by titration: treatment policy 5. Change in background ART regimen: while on treatment	Based on target population, participants taking estradiol, there are no missing data due to lost to follow up or study discontinuation. Missing data due to lab errors are ignorable because it is non-informative, and rare.
Population-level summary measure	**Analysis approach**
At each dose of 17-β estradiol (2, 4, 6, 8 and 10 mg), geometric mean ratio (GMR) of total serum Ctrough estradiol between those whose ART regimen is boosted versus not boosted, with 90% CI.	Parallel groups approach between non-randomized groups, using pooled estimate of variance for CI calculation. Normality assumptions on natural log transformation of Ctrough outcomes. Sensitivity Analysis 1: Using Satterthwaite version of standard error calculation to explore the influence of assumption of equal variance between groups on calculation of the GMR and CI. The 90% Confidence interval about each GMR will be calculated and compared to no effect reference interval of [0.7, 1.43] The geometric means and CVs and other descriptive statistics (key percentiles) within each analysis group will also be calculated.

	If multiple observations are available at the same estradiol dose, the first sampled qualifying steady-state trough concentration (based on calendar time) will be used. One of the following sensitivity analyses (depending on the rate of replicates) will be used: Sensitivity Analysis 1: If multiple observations are available at the same estradiol dose, the minimum of qualifying trough concentrations sampled at steady-state will be used. Sensitivity Analysis 2: If multiple observations are available at the same estradiol dose, the geometric mean of the qualifying trough concentrations sampled at steady-state will be used.

Supplementary Analysis

1. Adjusted GMR where adjustment via normal regression modeling for prognostic factors (e.g. geographical location) which may be imbalanced between comparisons groups despite accrual limits to mitigate confounding. Note: this can also be used to adjust for anti-androgen use, and for restarting vs. naïve to FHT.

2. P-value from Wilcoxon-Rank Sum test between analysis groups (based on untransformed data).

5.2 Secondary Estimand(s)

5.2.1 Occurrence of related Adverse Events Secondary Estimand

Note: This analysis will be repeated for Outcome 10.2.2.3 Occurrence of SAEs and new events of CAD/CVD, cancer, DM/pre-DM, and vascular event, defined in protocol section 7.2, over 48 weeks.

<table>
<tr><th colspan="3">Secondary Objective 1: To compare occurrence of FHT-associated adverse effects across ART groups and estradiol doses.</th></tr>
<tr><td>Estimand description</td><td colspan="2">Among adult TW-PWH who are virally suppressed on a targeted ARV regimen and who start or restart oral 17-β estradiol at 2 mg/day and whose dose may be titrated (and anti-androgen may be added after 24 weeks) as informed by safety, hormone levels and participant wishes; probability of having at least one adverse event deemed related to 17-β estradiol over 48 weeks while receiving 17-β estradiol.</td></tr>
<tr><td>Treatment</td><td colspan="2">Oral 17-β estradiol for 48 weeks, initiated at 2 mg/day and titrated as informed by safety evaluations, target hormone levels and participant wishes 4, 12, 24 and 36 weeks later, with allowed increases at each titration of 2 mg. After 24 weeks, anti-androgen may be added.</td></tr>
<tr><th>Target population</th><th colspan="2">Analysis set</th></tr>
<tr><td>Adult persons with HIV who were assigned male sex at birth, who are virologically suppressed on a targeted ART regimen, and who wish to take feminizing hormone treatment for the purposes of developing or enhancing female secondary sex characteristics.</td><td colspan="2">Safety Analysis Population (see Section 4.1)</td></tr>
<tr><th>Variable(s)</th><th colspan="2">Outcome measure(s)</th></tr>
<tr><td>Occurrence of any reported adverse event deemed related to 17-β estradiol over 48 weeks while on study treatment.</td><td colspan="2">Outcome measure as defined by the variable. See Outcome 10.2.2.2 for details.</td></tr>
<tr><th>Handling of intercurrent events</th><th colspan="2">Handling of missing data</th></tr>
<tr><td>1. Premature stopping of estradiol due to AE: while on treatment/Composite
2. Premature stopping of estradiol due to lack of efficacy: while on treatment
3. Use of other/alternative FHT (e.g. street hormones): while on treatment
4. Use of estradiol at dose other than prescribed by titration: treatment policy
5. Change in background ART regimen: while on treatment</td><td colspan="2">Based on target population, participants taking estradiol, there are no missing data due to lost to follow up or study discontinuation.</td></tr>
<tr><th>Population-level summary measure</th><th colspan="2">Analysis approach</th></tr>
<tr><td>Proportion of people having at least one adverse event deemed related to 17-β estradiol over 48 weeks, overall and by ART regimen.</td><td colspan="2">[If Indicated] Two-degree of freedom tests will explore if the number of participants with qualifying safety outcomes differs across the three ART groups, with pairwise comparisons informed by the results of the omnibus testing.</td></tr>
</table>

Supplementary Analysis:

Association of 17-β estradiol dose with safety outcomes will utilize estimation of each safety outcome incidence in order to adjust for time of exposure to various estradiol doses.

6 Analysis of Other Objectives

6.1 To describe changes in sex hormones following 17-β estradiol administration. [Exploratory objective 1.6.3]

Testosterone Suppression [Outcome 10.2.2.4]

Proportion of participants with suppressed testosterone at each dose of 17-β estradiol (2, 4, 6, 8, and 10 mg) will be summarized by ART group and anti-androgen use (use versus no use). Analysis will be conducted in the subset of the PK Analysis Population who has not undergone orchiectomy.

Missing testosterone concentrations will be excluded from this analysis.

Supplementary analysis: By subgroups defined by anti-androgen use.

6.2 To compare virologic suppression within ART groups, at each received 17-β estradiol dose over 48 weeks [Secondary objective 1.5.3]

Virologic Suppression [Outcome 10.2.2.5]

Proportion of participants with virologic suppression of HIV will be summarized by ART group. Analysis will include the PK Analysis Population.

Missing viral loads will be excluded from this analysis.

6.3 To summarize intervention satisfaction across ART groups. [Secondary objective 1.5.2]

Transgender Congruence Score [Outcome 10.2.2.6]

Overall transgender congruence scores and changes from baseline at weeks 24 and 48 will be summarized by ART group and by 17-β estradiol dose at time of measurement. Analysis will include the Safety Analysis Population.

Change in transgender congruence score at each week will be compared between ART groups using Kruskal-Wallis Test, by 17-β estradiol dose at time of measure. Change in transgender congruence score at each week will be modeled by ART group and 17-β estradiol dose.

To test secondary hypothesis (Participants receiving DRV-containing ART will require higher doses of 17-β estradiol to achieve similar FHT satisfaction as participants receiving BIC- or DTG-containing ART):

Transgender Congruence Score at each week will be modeled by ART group (boosted vs. not-boosted ART) and 17-β estradiol dose. The interaction between ART and estradiol dose will be tested to see if the effect of the dosage on the score differs in the two groups.

Supplemental Analysis: By FHT history (re-started FHT versus naïve to FHT)

6.4 To compare the PK parameters (AUC, C_{trough}, C_{max}, T_{max}) of 17-β estradiol over 48 weeks in a subset of TW undergoing intensive PK between groups receiving boosted versus non-boosted ART [Secondary objective 1.5.4]

PK parameters (AUC 0-8h, C_{trough}, C_{max}, T_{max}) based on intensive sampling [Outcome 10.2.2.7, 10.2.3.5]

Participants included in the PK Analysis population who also underwent intensive PK sampling will be included in intensive PK parameter summaries and analyses.

PK parameters for 17-β estradiol from intensive sampling on a subset of participants at weeks 24 and 48 by will be summarized by ART group and by 17-β estradiol dose at time of measurement.

Analyses of intensive 17-β estradiol PK parameters will use parallel methods to those described in Section 5.1.2 Third Primary Estimand. The no effect reference interval may vary for parameters.

The distribution of 17-β estradiol PK parameters $T_{max,}$ T_{min}, and T1/2 will be compared between analysis groups at each estradiol dose using Wilcoxon-Rank Sum tests.

6.5 To assess changes in weight, anthropometrics, lipids, glucose, and insulin sensitivity in relationship to estradiol concentrations. [Secondary Objective 1.5.5]

6.5.1 Anthropometric Outcome Changes [Outcome 10.2.2.8, 10.2.2.9]

Anthropometric outcomes (weight, BMI, waist, hip, and waist-to-hip ratio) and their absolute and percent changes from baseline in will be summarized by ART group. Analysis will include the Safety Analysis Population.

Maximum change (change from baseline to highest value) and dose at maximum change, regardless of time point, will be summarized. [Time of maximum change will be summarized.]

Within-person changes from 0 to 2 mg estradiol, 0 to 4 mg estradiol, etc., will be summarized to examine a stepwise increase as estradiol dose increases.

The hypothesized association of estradiol dose with increased changes in anthropometric outcomes will be assessed.

Supplemental analyses:

- Frequency of meeting the AHA metabolic syndrome criteria (and estradiol dose at time of meeting criteria):
    - Central or abdominal obesity: >40 inches waist circumference

6.5.2 Lab Outcome Changes [Outcome 10.2.2.10, 10.2.2.11]

Lab results (fasting lipids, glucose, and insulin sensitivity) and their absolute changes from baseline will be summarized by ART group. Analysis will include the Safety Analysis Population.

The relationship with estradiol dose may be explored in a parallel fashion to anthropometric measurements.

Supplemental analyses:

- Fasting lipid results divided into subgroups by statin use.
- Frequency meeting the AHA metabolic syndrome criteria (and estradiol dose at time of meeting criteria):
    - High triglycerides: ≥150 mg/dL, or taking medicine for high triglycerides
    - Low HDL cholesterol: <40 mg/dL, or taking medicine for low HDL cholesterol
    - High blood pressure: ≥130/85 mm Hg, or taking medicine for high blood pressure
    - High fasting glucose: ≥100 mg/dL, or taking medicine for high blood glucose
- Frequency of developing prediabetes and diabetes (as defined by ADA), while taking estradiol, and estradiol dose at time of meeting criteria
- Incidence rate of developing HOMA-IR > 2 while taking estradiol, and estradiol dose at time of meeting criteria
- Frequency of use of lipid lowering drugs, anti-hypertensive medication, or glucose lowering medications

6.6 To assess the changes in estradiol concentrations across each 17-β estradiol dosage and ART group. [Secondary Objective 1.5.6]

Estradiol Concentrations [Outcome 10.2.1.2, 10.2.2.7]

The outcome to address this 'dose proportionality' objective are the estradiol concentrations. [This will include all participants who started estradiol.] Secondarily, the outcomes is the estradiol AUCs (as measured within intensive PK sample). [This will include all participants in the intensive PK sample who started estradiol.]

The primary covariate of interest is the estradiol dose at the time that the concentration or AUC is measured (and reported dose rather than prescribed dose). Other covariates of interest include both ART cohort group and use of anti-androgen.

Intercurrent event (ICE) handling: All analyses are 'while on treatment', which includes both estradiol and ART.

Using regression models, the analysis will explore the relationship between the outcomes and estradiol dose, while controlling for ART group and anti-androgen use (e.g. as additional covariates or by stratification), and incorporating repeated measures by participants. The focus will be on estimation, rather than testing. For example, a linear model would imply a constant rate of change in concentration across the estradiol dose interval from 2 – 10 mg. Generalized linear models, including linear splines, quadratic ,etc. will be explored for goodness of fit before parameter interpretation.

7 Report Contents

Additional details will be provided in an Analysis Implementation Plan (AIP).

1. CONSORT Diagram
2. Study Entry
    a. Screening
    b. Accrual
    c. Eligibility Violations
3. Baseline Characteristics
4. Study Retention
5. Study Treatment (17-β estradiol) titration and adherence
6. Other FHT initiation
7. ART Treatment adherence
8. Safety
    a. Adverse events and deaths
    b. Virologic failure
    c. Targeted events
9. Analysis sets and analysis exclusions
10. Analysis of primary outcome measures: ART PK analyses
11. Analysis of primary outcome measures: Estradiol PK analyses
12. Analysis of secondary outcomes measures:
    a. Safety outcomes
    b. Testosterone Suppression
    c. Virologic Suppression of HIV
    d. Transgender Congruence Score
    e. Intensive Estradiol PK parameters
    f. Anthropometric Outcomes
    g. Lab Outcomes

8 References

1. Kozee HB, Tylka TL, Bauerband LA. Measuring Transgender Individuals' Comfort With Gender Identity and Appearance: Development and Validation of the Transgender Congruence Scale. Psychology of Women Quarterly 2012; 36(2): 179-196.

2. Sarafidis P, Lasaridis A, Nilsson P, et al. Validity and reproducibility of HOMA-IR, 1/HOMA-IR, QUICKI and McAuley's indices in patients with hypertension and type II diabetes. J Hum Hypertens 2007; 21: 709–716. https://doi.org/10.1038/sj.jhh.1002201

9 Associated Documents

Attachment 1: Draft Writing Team Roster

[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]

[redacted]

[redacted]

[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]

Attachment 2: Timetable for Primary Analysis and Manuscript Preparation

[redacted]

[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]		[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]