**IMPAACT 2010**

**Phase III Study of the Virologic Efficacy and Safety of Dolutegravir-Containing versus Efavirenz-Containing Antiretroviral Therapy Regimens in HIV-1-Infected Pregnant Women and their Infants**

**"VESTED"**
**Virologic Efficacy and Safety of ART Combinations with TAF/TDF, EFV, and DTG**

**DAIDS Study ID #30129**
**IND # 133,438 Held by DAIDS**
**ClinicalTrials.gov ID: NCT03048422**

# Statistical Analysis Plan

Version 5.0

October 26, 2020

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

This is the IMPAACT 2010 SAP Version 5.0 with names of authors, names of publication writing team members, and analysis timeline redacted.

## Table of Contents

# 1. Introduction

This document describes the proposed content for the primary statistical analysis of IMPAACT 2010, focusing on analyses that address the major randomized comparisons for key safety, tolerability and efficacy outcome measures. This includes the key analyses that will form the core of any presentation or publication used to disseminate the primary conclusions of the study. It is recognized that this statistical analysis plan (SAP) may be modified by the Study Team as new information becomes available or to reflect recommendations made by the DSMB.

The primary outcomes can be measured at two different study times: at pregnancy outcome and through visit week 50 postpartum. In order for the study results to have maximum impact, the study data results will be released once all women have had a pregnancy outcome instead of waiting until all participants complete follow-up through visit week 50. The final primary analysis results will be described in two separate reports after all pregnancy outcomes and after all follow-up is completed; analyses for both timepoints will be described throughout this document.

Major changes to the previous SAP are noted in **bold** throughout the document.

# 2. Core Manuscript Writing Team

Core writing team names were redacted from this version of the SAP.

# 3. Study Overview

## 3.1 Study Design

IMPAACT 2010 is a Phase III, three-arm, randomized, open-label study to compare the virologic efficacy and safety of three antiretroviral regimens — 1) dolutegravir(DTG)+ emtricitabine(FTC)/ tenofovir alafenamide(TAF)(DTG+FTC/TAF), 2) dolutegravir(DTG)+ emtricitabine(FTC)/tenofovir disoproxil fumarate (TDF)(DTG+FTC/TDF), and 3) efavirenz (EFV)/emtricitabine(FTC)/tenofovir disoproxil fumarate (TDF) (EFV/FTC/TDF) — for HIV-1-infected pregnant women and their infants. Mother-infant pairs will be randomized at 14-28 weeks gestation in a 1:1:1 ratio to receive either DTG+FTC/TAF (Arm 1), DTG+FTC/TDF (Arm 2), or EFV/FTC/TDF (Arm 3). The primary objectives are

to compare the DTG-containing arms (Arms 1 and 2) to the EFV-containing arm (Arm 3) with respect to maternal viral suppression at delivery; to compute all pairwise regimen comparisons for a composite adverse pregnancy outcome (spontaneous abortion, fetal death, preterm delivery, or small for gestational age); and to compute all pairwise regimen comparisons for maternal and infant grade 3 or higher adverse events through 50 weeks postpartum. All mothers and infants are planned to be followed through 50 weeks postpartum. The total sample size is 639 mother-infant pairs (approximately 213 per arm) and accrual is expected to be completed within 12 months after the first pair is enrolled. The total study duration will be approximately 31 months. The sample size will be re-evaluated with the possibility of adjustment prior to the closure of participant accrual. Mother-infant pairs will be stratified at randomization by gestational age (14-18 weeks, 19-23 weeks, 24-28 weeks) and by country.

The primary virologic efficacy analysis will combine the two DTG-containing randomized arms for comparison to EFV/FTC/TDF. The DTG-containing arms will be combined because viral load at delivery is not expected to vary between these arms. This combined analysis also weighs feasibility against the need to provide information on the possibility of effect modification of DTG in the presence of TDF or TAF. Powering the study for all three pairwise comparisons would be ideal scientifically; however, this would require an infeasible sample size.

Viral suppression at delivery was chosen as the primary efficacy endpoint because of the importance of maintaining high viral suppression to minimize perinatal HIV transmission. A conclusion for superiority in viral suppression of a DTG-containing regimen relative to EFV/FTC/TDF might hasten the adoption of a DTG-containing regimen for use during pregnancy and postpartum. However, a conclusion of non-inferiority might allow for an increase in the options available to pregnant and postpartum women. Because of these two important possible conclusions at the end of the study, the primary analysis will be an analysis of non-inferiority while a superiority comparison is specified as an important secondary analysis. Testing both superiority and non-inferiority does not constitute multiple testing; however, non-inferiority must be established before a conclusion of superiority can be made.

A conclusion of non-inferiority will be predicated on the following considerations:

- The 95% confidence interval for the difference excludes the non-inferiority margin.
- There is a high degree of consistency with the protocol plans. For example, the levels of protocol deviations and loss to follow-up will need to be similar to those in previous studies in which superiority has been shown (e.g., PROMISE, PROMOTE).
- The EFV-containing arm shows its usual efficacy.
- The ITT and the per-protocol analyses show similar results.

All three pairwise regimen comparisons will be conducted for the adverse pregnancy and maternal adverse event primary objectives. The DTG-containing arms (Arm 1 and Arm 2) will not be combined for these primary analyses because little is known about safety in pregnancy through 1 year postpartum between these two arms.

An open-label study design was selected because the number of pills per day and instructions for administration vary between treatment arms. Potential sources of bias with an open-label study include increased attribution and reporting of specific toxicities in a treatment arm. Site investigators may be more likely to change antiretroviral therapy if a treatment arm is perceived as suboptimal. Efforts to minimize these biases are determined in the study protocol by establishing

standard criteria and guidelines for toxicity and ARV regimen management, safety-related data recording and reporting, and inadequate virologic response.

Missing data will be handled using three strategies: complete case analysis, multiple imputation, and sensitivity analyses. The complete case analysis will remove records with missing data. Since the amount of missing data is expected to be low in IMPAACT 2010, the complete case analysis will be the principal analysis. Multiple imputation and sensitivity analyses will be used to measure the robustness of the study findings. Congruence, as measured by the differences in the estimated point estimates and 95% CIs, between the complete case analyses and the missing data methods will be used to measure robustness of the analysis to the incomplete records. Multiple imputation and sensitivity analyses will be used for all of the four primary analyses. Multiple imputation and sensitivity analyses will be used for the secondary analyses if the percentage of missing data is unacceptably high (e.g. >10%). The details on which missing data methods will be used for each outcome is described in the analysis sections below.

## 3.2 Hypotheses

Among HIV-1-infected pregnant women initiating ART between 14 and 28 weeks gestation, and their infants:

- A DTG-containing regimen will be non-inferior to EFV/FTC/TDF with regard to virologic efficacy at delivery
- Rates of adverse pregnancy outcomes will not significantly differ between the three study ART regimens
- Rates of maternal and infant grade 3 or higher adverse events will not significantly differ between the three study ART regimens

## 3.3 Study Objectives and Outcome Measures

### 3.3.1 Primary Objectives and Outcome Measures

The primary and secondary objectives are per protocol version 2.0, **LOA #3**. Outcome measures presented in the final postpartum analysis report will be summarized as defined in protocol version 2.0, **LOA #3**. Primary outcome measures that will be presented in the final primary pregnancy outcome analysis report are identified below with the following superscripts based on definition of delivery:

[†]Includes data up to 14 days postpartum (post-birth) in primary pregnancy analysis report

[‡]Includes data through 28 days postpartum (post-birth) in primary pregnancy analysis report

[††]Includes data reported on study through 50 weeks postpartum (post-birth) as of data freeze date in primary pregnancy analysis report

[‡‡]Includes measurements up to but not including post-delivery.

Primary Objectives and Outcome Measures

- Whether treatment initiated during pregnancy with a DTG-containing regimen is non-inferior to EFV/FTC/TDF with regard to virologic efficacy (HIV-1 RNA <200 copies/mL) at delivery
  **Outcome Measure:** HIV-1 RNA <200 copies/mL at delivery (up to 14 days postpartum), using real-time test results obtained at site laboratories[†]

- Whether rates of the following safety outcomes differ for any pairwise regimen comparison:

  - Adverse pregnancy outcomes (spontaneous abortion, fetal death, preterm delivery, or small for gestational age)
    **Outcome Measure:** Composite outcome of spontaneous abortion (occurring at <20 weeks gestation), fetal death (occurring at ≥20 weeks gestation), preterm delivery[1] (<37 completed weeks), or small for gestational age[2] (<10th percentile)[†]

  - Maternal grade 3 or higher adverse events through 50 weeks postpartum
    **Outcome Measure:** Maternal grade 3 or higher adverse events, including events resulting in death due to any cause, through 50 weeks postpartum (refer to **Protocol** Section 7.3.3 for severity grading; for this outcome measure, grade 3 or higher creatinine levels and creatinine clearance rates will be defined based on absolute values (>1.8 x ULN for creatinine, <60 mL/min for creatinine clearance rate)[†]

  - Infant grade 3 or higher adverse events through 50 weeks postpartum
    **Outcome Measure:** Infant grade 3 or higher adverse events, including events resulting in death due to any cause, through 50 weeks postpartum[‡]

[1] The estimation of gestational age at delivery and study entry will follow the guidance from the American College of Obstetricians and Gynecologists (ACOG). The ACOG algorithm determines an estimated due date (EDD) by combining information from the estimated due date from ultrasound measurements and the self-reported last menstrual period ("Methods for Estimating Due Date." The American College of Obstetricians and Gynecologists, no. 700, May 2017).

To account for missing data, the following hierarchy will be used to estimate gestational age:

1) ACOG algorithm using the ultrasound biometry measurements calculated by the INTERGROWTH-21 formulas
2) ACOG algorithm using ultrasound EDD (Missing biometry measurements)
3) Ultrasound biometry measurements calculated by the INTERGROWTH-21 formulas (Missing LMP)
4) Ultrasound EDD (Missing LMP and the biometry measurements)
5) Last menstrual period (Missing ultrasound measurement)
6) Site-determined gestational age at study entry (Applicable to interim analyses in which missing data have yet to be queried, missing LMP and ultrasound data)

[2]Infants will be classified as small for gestational age based on methods from the INTERGROWTH-21st Project. Birthweight, sex, and gestational age at delivery will be used to calculate the infant percentile of size for gestational age at delivery. Villar, et.al. "International Standards for Newborn Weight, Length, and Head Circumference by Gestational Age and Sex: the Newborn Cross-Sectional Study of the INTERGROWTH-21st Project." The Lancet, vol. 384, 6 Sept. 2014. More information on the INTERGROWTH-21st Project can also be found at intergrowth21.tghn.org.

### 3.3.2 Secondary Objectives and Outcome Measures

The secondary objectives of this study are to evaluate the following among HIV-1-infected pregnant women and their infants:

- Whether treatment initiated during pregnancy with a DTG-containing regimen is superior to EFV/FTC/TDF with regard to virologic efficacy (HIV-1 RNA <200 copies/mL) at delivery
  **Outcome Measure:** HIV-1 RNA <200 copies/mL at delivery using real-time test results obtained at site laboratories (superiority analysis) †

- Whether the following differ when comparing a DTG-containing regimen initiated during pregnancy to EFV/FTC/TDF:
  - Proportion of mothers with HIV-1 RNA <50 copies/mL at delivery
    **Outcome Measure:** HIV-1 RNA <50 copies/mL at delivery (up to 14 days postpartum) using batched test results obtained from central laboratory
  - Proportion of mothers with HIV-1 RNA <200 copies/mL at 50 weeks postpartum
    **Outcome Measure:** HIV-1 RNA <200 copies/mL at 50 weeks postpartum using real-time test results obtained from site laboratories
  - Time to maternal HIV-1 RNA <200 copies/mL
    **Outcome Measure:** Time to first HIV-1 RNA <200 copies/mL through delivery (up to 14 days postpartum) using real-time results obtained from site laboratories†

- Whether the following differs for any pairwise regimen comparison:
  - Proportion of mothers with HIV-1 RNA <200 copies/mL using the standardized FDA snapshot algorithm at delivery and at 50 weeks postpartum
    **Outcome Measures:**
    - HIV-1 RNA <200 copies/mL at delivery (up to 14 days postpartum) using real-time test results obtained from site laboratories and FDA[3] snapshot algorithm†
    - HIV-1 RNA <200 copies/mL at 50 weeks postpartum using real-time results obtained from site laboratories and FDA snapshot algorithm

- Whether rates of the following differ when comparing a DTG-containing regimen to EFV/FTC/TDF:
  - Adverse pregnancy outcomes (spontaneous abortion, fetal death, preterm delivery, or small for gestational age)
    **Outcome Measure:** Composite outcome of spontaneous abortion (occurring at <20 weeks gestation), fetal death (occurring at ≥20 weeks gestation), preterm delivery (<37 completed weeks), or small for gestational age (<10$^{th}$ percentile) †
  - Maternal grade 3 or higher adverse events through 50 weeks postpartum
    **Outcome Measure:** Maternal grade 3 or higher adverse events, including events resulting in death due to any cause, through 50 weeks postpartum†
  - Infant grade 3 or higher adverse events through 50 weeks postpartum
    **Outcome Measure:** Infant grade 3 or higher adverse events, including events resulting in death due to any cause, through 50 weeks postpartum‡

- Whether rates of the following differ for any pairwise regimen comparison:
  - A composite outcome of spontaneous abortion, fetal death, preterm delivery, small for gestational age, or major congenital anomaly
    **Outcome Measure:** Composite outcome of spontaneous abortion (occurring at <20 weeks gestation), fetal death (occurring at ≥20 weeks gestation), preterm delivery (<37 completed weeks), or small for gestational age (<10$^{th}$ percentile) or major congenital anomaly[4] ††

- A ranked composite infant safety outcome measure through 50 weeks postpartum
  **Outcome Measure:** Ranked composite infant safety outcome through 50 weeks postpartum
- Infant HIV infection through 50 weeks postpartum
  **Outcome Measure:** Infant HIV infection at delivery and through 50 weeks postpartum[†]
- Infant mortality through 50 weeks postpartum
  **Outcome Measure:** Infant death due to any cause through 50 weeks postpartum[‡]
- Infant bone toxicity at 26 weeks postpartum
  **Other Outcome Measure:** Infant whole body and lumbar spine BMC **values** based on DXA scan at 26 weeks postpartum
- Maternal bone toxicity at 50 weeks postpartum
  **Other Outcome Measure:** Maternal lumbar spine and hip BMD Z-scores based on DXA scan at 50 weeks postpartum
- Markers of maternal and infant renal toxicity through 50 weeks postpartum
  **Outcome Measures:**
  - Maternal serum creatinine and creatinine clearance rate (Cockcroft-Gault formula)[†]
  - Infant serum creatinine and creatinine clearance rate (Schwartz formula)

  **Other Outcome Measure:** Maternal serum creatinine and creatinine clearance rate (Cockcroft-Gault formula), urine protein creatinine ratio, beta 2 microglobulin, and retinol binding protein
- Antiretroviral drug resistance observed with each maternal ART regimen:
  - Among mothers who experience virologic failure (at baseline and time of virologic failure)
    **Secondary Outcome Measure:** HIV-1 antiretroviral drug resistance mutations at the time of maternal virologic failure (and at screening for mothers with resistance detected at the time of virologic failure to determine if the resistance was present prior to enrollment or occurred post-enrollment) using the Stanford algorithm
  - Among HIV-infected infants (at time of HIV diagnosis)
    **Secondary Outcome Measure:** HIV-1 antiretroviral drug resistance mutations at the time HIV diagnosis for HIV-infected infants

- Whether treatment initiated during pregnancy with each regimen is non-inferior with regard to preterm delivery and, separately, small for gestational age for any pairwise regimen comparison
  **Outcome Measures:**
  - Preterm delivery (<37 completed weeks)[†]
  - Small for gestational age (<10$^{th}$ percentile)[†]

- Whether changes in maternal weight differ for any pairwise comparison or when comparing a DTG-containing regimen to EFV/FTC/TDF
  **Other Outcome Measures:**
  - Change in maternal weight antepartum (from study entry to delivery) [‡‡]
  - Change in maternal weight postpartum (from delivery to 50 weeks postpartum)
  - Change in maternal weight overall (from study entry to 50 weeks postpartum)

**Other outcome measures are not submitted to ClinicalTrials.gov.**

[3]FDA's Guidance for Industry, "The Human Immunodeficiency Virus-1 Infection: Developing Antiretroviral Drugs for Treatment", Appendix A, November 2015.

[4]Major congenital anomaly will be defined for this study consistent with the definition of malformation provided by Holmes LB, Westgate MN. Inclusion and exclusion criteria for malformations in newborn infants exposed to potential teratogens. Birth defects research Part A, Clinical and molecular teratology. 2011;91(9):807-12, i.e., a structural abnormality with surgical, medical, or cosmetic importance. The following will not be considered major congenital anomalies: genetic disorders, chromosome abnormalities, minor anomalies, birth marks, positional deformities, prematurity related findings, prenatal ultrasound-only findings (i.e., findings only identified by ultrasound and not by the examining pediatrician), and polydactyly (postaxial, type B). Findings consistent with the definition of malformation will be included in analyses as major congenital anomalies. All suspected congenital anomalies will be entered into eCRFs and a small group of study investigators (blinded to randomized study arm), including an expert on birth defects, will review all suspected anomalies in near real time to determine whether they meet the study-specific definition of major congenital anomaly.

## 3.4 Visit and Evaluation Schedule

Throughout this document, analysis timeframes described as "at week X", includes the study visit window. Antepartum clinic visits are scheduled every 4 weeks (± 2 weeks) prior to delivery. Participants will complete a delivery clinic visit as soon as possible after delivery, within a targeted window of 14 days after the pregnancy outcome. If the visit cannot be conducted within the targeted window, it may be conducted within an allowable window of 27 days after delivery. Viral load results will only be considered as delivery measurements if tested within 14 days after the pregnancy outcome. Both the mother and infant are scheduled to return at 6 weeks postpartum (± 2 weeks) and then continue follow-up visits every 12 weeks (±6 weeks) until 50 weeks postpartum.

If a mother's study drug regimen is modified by switching DTG or EFV to another ARV, an additional study visit should be scheduled approximately 4 weeks (± 1 week) after the switch. If a maternal participant has an HIV viral load ≥ 200 copies/mL at or after at least 24 weeks on the study or an infant receives a positive HIV NAT result, additional HIV-1 RNA testing should be conducted within 28 days of the date of specimen collection for the initial test.

If a mother is pregnant at the visit week 50 postpartum visit, the pregnancy outcome and ARV changes during the subsequent pregnancy must be recorded. If confirmation of a maternal virologic failure or infant HIV infection is pending after the visit week 50 postpartum visit, the confirmatory test and result must be recorded as well.

**Baseline for Mothers and Birth Data for Infants**
Maternal baseline values are the measurements closest to or on study entry date. Ultrasound measurements used to estimate gestational age are prioritized to the earliest values available during the current pregnancy. Ultrasounds are permitted up to 14 days after randomization.

Infant birth values are the measurements closest to or on their birthdate. Values will be considered as birth data up to 14 days after delivery.

# 4. Monitoring and Re-Evaluation of Sample Size

## 4.1 DSMB

The study will be monitored by a NIAID-convened Data and Safety Monitoring Board (DSMB), which will perform its monitoring functions consistent with the NIAID Policy on Data and Safety Board Operations.

Interim analyses of study progress and conduct as well as participant safety will be reviewed at least biannually starting within 12 months after the first mother-infant pair is randomized. Two interim efficacy analyses will be performed when approximately 33% and 66% of the anticipated information on viral suppression at delivery have been observed, unless otherwise requested by the DSMB. Assuming participant accrual as described in Section 3.8 of the protocol and that the average gestational age at entry is 24 weeks, it is projected that 33% of the total information will be available 10 months after the first mother-infant pair is enrolled in the study.

The DSMB will be provided with masked by-arm study data reports. Based on any of its reviews, the DSMB may recommend that the study proceed as currently designed, proceed with design modifications, or be discontinued. The DSMB may also provide operational recommendations to address any study implementation challenges that may be identified.

**Statistical Stopping Guidelines for Evaluating Efficacy**
Efficacy decision boundary guidelines will be based on the group sequential Lan-DeMets O'Brien-Fleming like error spending approach. The Z test statistic for the difference in proportions when comparing the DTG-containing arms to the EFV-containing arm will be compared to the z-value decision boundary guideline computed from the Lan-DeMets O'Brien-Fleming like error spending approach. The z-value decision boundary guideline is approximately ±3.7 at 33% of information and ±2.5 at 66% of information. As the decision boundary guidelines change depending on the amount of information at the interim analysis, these boundaries are only given for illustrative purposes. If the Z test statistic is larger in magnitude than the z-value decision boundary guideline, in either direction, then early termination of the study may be considered. A conclusion for non-inferiority of the combined DTG-containing arms relative to the EFV-containing arm at an interim analysis in the absence of superiority would not warrant stopping for efficacy. Because a conclusion for superiority will be predicated on a decision for non-inferiority, the non-inferiority analyses should also be considered by the DSMB.

**Assessment of Futility for the Efficacy Comparisons**
IMPAACT 2010 should not be stopped at an interim analysis for futility if it appears unlikely that a conclusion of superiority or non-inferiority can be made at the maximal sample size. There are two reasons why that futility of the primary efficacy comparisons should not be considered at interim analyses: 1) the study will be quite advanced in its conduct by the time a reasonable number of participants have outcome information for the futility analyses, and 2) it will be important to collect as much safety data as possible for the 3 primary safety objectives.

## 4.2 Re-Evaluation of Sample Size

The main purpose of the sample size re-evaluation is to use the observed data to inform the following parameters used in the sample size calculations: the percent of participants with evaluable data and the percent of mothers with HIV-1 RNA <200 copies/mL at delivery. This is to evaluate whether a sample size increase is needed to achieve the desired statistical power for the primary efficacy objective of the study. The sample size re-evaluation will take place at one interim analysis, once sufficient information is available to estimate the proportion of mothers with HIV-1 RNA <200 copies/mL at delivery. To maintain power for the primary safety analyses, the sample size should not be decreased.

The re-evaluation will be undertaken by a statistician who is blinded to the randomized treatment arms. The statistician will be given study datasets without access to the randomized treatment arms. The re-evaluated sample size will be computed for the DSMB at the interim analysis in which at least 192 women have delivery HIV-1 RNA data available (approximately 33% of the expected information) or at the scheduled DSMB review before accrual is completed, whichever comes first. If accrual of 639 mother-infant pairs is expected to be completed before the next scheduled DSMB meeting, accrual may be paused when 639 pairs have been enrolled but accrual would not formally close until the DSMB has reviewed the sample size adjustment. The DSMB will review the sample size re-evaluation based on a pre-specified algorithm and make recommendations concerning sample size.

Further guidelines for the DSMB in considering sample size re-evaluation include the following:

- The sample size should not increase because the percent of non-evaluable women at delivery is higher than 15%. The primary rationale for this guideline is that the percent could potentially be so high relative to the anticipated percent suppressed that the results of the study may be difficult to interpret. Thus, the main point of considering the percent evaluable is because a rate lower than 15% may be relevant in offsetting any increase in sample size needed due a lower than expected suppression percentage at delivery.

- The DSMB will be provided with estimated primary outcome rates by arm at each interim efficacy analysis. A concern, however, with sample size re-evaluation is that it may affect the Type I error rate and introduce bias in estimates of the difference in outcomes between arms (and associated confidence intervals). This is generally avoided if the interim information used to re-evaluate the sample size is pooled over randomized treatment arms.

- A sample size re-evaluation algorithm will be followed. To re-power the study, the amount of information will be fixed to the amount used to calculate the sample size of 639 pairs. The maximal amount of information for this study is:

$$\mathrm{I} = \left[ .8 * (1 - .8)/(.9 * 639 * \frac{2}{3}) + .8 * (1 - .8)/(.9 * 639 * \frac{1}{3}) \right]^{-1} \approx 797.2$$

Maximal information was computed from the inverse of the variance of the estimate of the difference in binomial proportions calculated under the alternative hypothesis. At the time of the sample size re-evaluation, the observed pooled binomial proportion ($\hat{P}$) will be used to compute the re-evaluated sample size. If we let ε be the proportion evaluable, after solving the information formula for N, the formula to re-compute the total sample size is:

$$N = \frac{\mathrm{I}}{\varepsilon} * \left[\frac{9}{2}\hat{P}(1-\hat{P})\right].$$

Following this procedure will aid in maintaining 80% power for a non-inferiority margin of 10% assuming that the comparison groups have the same but arbitrary suppression probability.

The sample size was re-evaluated and presented to the DSMB in January 2019. The DSMB recommended on February 4, 2019 that the study continue as planned with no alteration in study size.

## 5. Analyses

All of the primary and secondary analyses will be performed with the principle of ITT. For the analyses of superiority, the ITT analysis is the primary analysis. Per-protocol analyses of non-inferiority will also be conducted.

There are two secondary non-inferiority objectives and one non-inferiority analysis for the primary efficacy objective. Non-inferiority analyses typically require additional considerations relative to an analysis of superiority, including the congruence between an ITT analysis and a per-protocol analysis. Thus, the non-inferiority analyses, ITT and per-protocol, should be given equal weight in the non-inferiority analyses.

Per-protocol analyses will censor or exclude participants' data who do not remain on the randomized regimen up to the time of the measured outcome. Any change, defined as the first switch, first addition or first removal of any of any of the ARVs in the randomized regimen will be considered a change from the randomized regimen, with the exception of switches made due to a requirement for other concomitant medications.

The unit of analysis for infant and pregnancy outcomes will be the mother-infant pair. Analysis summaries will use the most extreme (worst) infant or pregnancy outcome for each pregnancy. For example, if one infant of a multiple birth becomes HIV-infected but the other does not, this will count as one event in the analyses. If both infants become infected, this will count as one event for the mother-infant pair in the analysis, not two. This is a conservative approach and addresses the fact that data on infants from multiple births may be related. Analyses of live born infants will be conducted on the unique participant level (twins will be treated as independent).

Summaries and regimen comparisons will not be adjusted for the two stratification factors. By arm regimen comparisons for each stratum of gestational age at study entry (14-18 weeks, 19-23 weeks, or 24-28 weeks) will be included in the final study report. P-values for these treatment comparisons will be provided if there is a statistically significant difference (p-value <0.05) between the

treatment arm and gestational age at study entry.

There are ten primary comparisons to be made, including one efficacy comparison and nine safety comparisons. There will be no adjustments for multiple comparisons. Type I error rates (α) will be controlled at the comparison level where a statistical difference will be declared if the p-value is less than 0.05. The p-value for the primary efficacy comparison will be adjusted for the repeated interim analyses using the error spending approach mentioned above, and the safety analyses will use an unadjusted nominal p-value. The safety analyses will not be adjusted for multiple comparisons or interim analyses to minimize the chance of a Type II error (β). The trial-wise Type I error will be larger than 5%.

Per NIH policy for Phase III and pivotal Phase II and IV studies, NIH requires primary analyses of treatment comparisons to be summarized by sex and by race and treatment interactions with sex and race to be tested. These analyses are required so do not represent multiple comparisons and will be presented in the primary study analysis regardless of power issues. Analyses restricted to one sex or race will present the appropriate data (analyses including only mothers need to be presented by race, not by sex).

Women with no post-baseline follow-up will be excluded from post-baseline analyses. Visits after the date of randomization will be considered post-baseline visits.

## 6. Primary Outcome Measures

### Efficacy: Viral Suppression

**Purpose:** to address whether the DTG-containing regimen is non-inferior to EFV/FTC/TDF with regard to virologic efficacy at delivery when initiated during pregnancy

Primary Efficacy Outcome Measure

- HIV-1 RNA <200 copies/mL at delivery (up to 14 days postpartum), using real-time test results obtained at site laboratories

**Statistical Methods**

Analyses will compare the DTG-containing arms to the EFV-containing arm. Viral suppression will be summarized using a difference in binomial proportions with a success defined as HIV-1 RNA <200 copies/mL at delivery. The denominator for this analysis will include women who have an HIV-1 RNA measurement within the delivery visit window (+14 days). A two-sided, two-sample test in proportions will be used for this analysis with normal approximation. Confidence intervals and p-values for the difference in proportions will be adjusted for the interim analyses to maintain a two-sided Type 1 error rate of 0.05 using the time ordered calculations. Ninety-five percent repeated confidence intervals (RCI) will be computed at interim analyses. A 10% non-inferiority margin will be used for the primary analysis. A p-value less than 0.05 testing the equality of proportions will be considered statistically significant, and a p-value testing the equality of the proportions will be provided if non-inferiority is established. Similarly, a 95% CI (or RCI at interim analyses) that excludes a difference of 10% in favor of the EFV-containing regimen will be considered to be statistically significant for the non-inferiority analysis.

After collecting baseline data, the observed percent of women with a baseline viral load <200

copies/mL was higher than expected. A descriptive analysis to describe the impact that this observation has on the analysis conclusion will be conducted by re-computing the primary delivery viral load analysis among women with a baseline viral load of <200 copies/mL and women with ≥ 200 copies/mL separately. An interaction between the study arm and baseline viral will be tested using logistic regression.

Sensitivity analyses will be carried out using multiple imputation for women missing the HIV-1 RNA measurement within the delivery visit window. A linear mixed effect model will be used to impute missing delivery viral loads. The predictor will be time on study with the outcome of the log viral load before or on the delivery visit. Viral load measurements below the limit of quantification will be set to the limit of quantification. This model will include all participants that have at least one viral load measurement before or on delivery. Viral load will not be imputed for mothers with no viral load measurements. Natural splines will be used to model the viral load trajectories over time. Bayesian information criteria (BIC) will be used to select the degrees of freedom of the spline model. Furthermore, a random intercept and random slope(s) (multiple slopes if BIC selects a non-linear term) will be fit at the participant level. These participant level random effects will be used to impute the missing VL data. Because the VL trajectory might vary by arm, three imputation models will be used, one for each arm.

For interim analyses, the efficacy data will be presented to the DSMB when approximately 33% and 66% of participants have viral suppression data at delivery. A masked by-arm analysis will be given. See section 4.1 for stopping guidelines based on efficacy decision boundary guidelines.

## Safety

**Purpose:** to address whether safety outcome rates differ between the three regimens initiated during pregnancy.

Primary Safety Outcome Measures

- Composite outcome of spontaneous abortion (occurring at <20 weeks gestation), fetal death (occurring at ≥20 weeks gestation), preterm delivery (<37 completed weeks), or small for gestational age (<10th percentile)
- Maternal grade 3 or higher adverse events, including events resulting in death due to any cause, through 50 weeks postpartum (refer to Protocol Section 7.3.3 for severity grading; for this outcome measure, grade 3 or higher creatinine levels and creatinine clearance rates will be defined based on absolute values (>1.8 x ULN for creatinine, <60 mL/min for creatinine clearance rate)
- Infant grade 3 or higher adverse events, including events resulting in death due to any cause, through 50 weeks postpartum

The protocol requires grading of events according to Version 2.1 of the Division of AIDS Table for Grading the Severity of Adult and Pediatric Adverse Events (DAIDS AE Grading Table).

**Statistical Method**

Safety data will be presented to the DSMB every six months. All primary safety analyses will compare DTG+FTC/TDF (Arm 2) to DTG+FTC/TAF (Arm 1), DTG+FTC/TDF (Arm 2) to EFV/FTC/TDF (Arm 3), and DTG+FTC/TAF (Arm 1) to EFV/FTC/TDF (Arm 3).

**Spontaneous Abortion, Fetal Death, Preterm Delivery, and Small for Gestational Age**

The analyses of the composite adverse pregnancy outcomes will be summarized as differences in proportions. A two-sided, two-sample test for the equality of the difference in proportions will be used for the by arm comparisons. A nominal (not adjusted for interim analyses) p-value less than 0.05 will be considered statistically significant. The Type I error rate for each comparison will be 5%. Two important separate sensitivity analyses will be conducted: 1) by including mothers who withdrew (i.e. lost to follow-up) before experiencing any one of these pregnancy outcomes as failures to account for the possibility of informative missingness, and 2) by stratifying by the presence of a major congenital anomaly. The analysis stratifying by the presence of a major congenital anomaly will be conducted to investigate effect modification if there is a by-arm imbalance in major congenital anomalies.

Conclusions based on the analysis stratified by the presence of a major congenital anomaly should be interpreted with care as some major congenital anomalies may have occurred after randomization. At interim analyses, a sensitivity analysis will be conducted by restricting the analysis to women that with an expected due date four weeks before the date the data were frozen. This sensitivity analysis will be conducted because early events are likely to be selected at interim analyses, and without this restriction the prevalence of these events might be overestimated.

**Maternal Grade 3 or Higher Adverse Events**

The primary maternal safety analysis will use a Kaplan-Meier estimate of the event probability through 50 weeks postpartum. At the final analysis through 50 weeks postpartum**,** the proportion experiencing an event for each group will be computed from the Kaplan-Meier survival curve and Greenwood's formula will be used to compute the variance of the difference. A Z score will be computed by taking the difference in the estimated probabilities and dividing this difference by the square root of the sum of variances as estimated by Greenwood's formula. Since the time from randomization to a pregnancy outcome will vary among the mothers, the visit week 50 event Kaplan-Meier probability estimate will be taken from the time at the average time from randomization to the mother's pregnancy outcome plus 50 weeks. The Z score will then be compared to the standard normal curve to compute a p-value. For the primary pregnancy analysis report, a log rank test of the equality of survival curves will be presented. A nominal (not adjusted for interim analyses) p-value less than 0.05 will be considered statistically significant. The Type 1 error rate for each comparison will be 0.05. Important sensitivity analyses will be conducted by including mothers who drop out of the study early as failures to account for the possibility of informative missingness. At interim analyses the log-rank of the equality of survival curves will be presented to the DSMB.

Grade 3 and 4 creatinine clearance adverse events reported due to a percent decrease from baseline or creatinine adverse events reported due to an increase from baseline will be excluded from the primary maternal safety analysis. Only women with a calculated creatinine clearance value of <60 mL/min or mL/min/1.73m2 (grade 3 or higher creatinine clearance) or >1.8 x ULN creatinine (grade 3 or higher creatinine) will be counted as failures in the primary maternal safety analysis. See Section 9 for further details regarding creatinine clearance and creatinine grading.

**Infant Grade 3 or Higher Adverse Events**

The primary infant safety analysis will use a Kaplan-Meier estimate of the event probability from birth through visit week 50. The proportion experiencing an event for each group will be computed from the Kaplan-Meier survival curve and Greenwood's formula will be used to compute the variance of the difference. A Z-score will be computed by taking the difference in the estimated

probabilities and dividing this difference by the square root of the sum of variances as estimated by Greenwood's formula. The Z score will then be compared to the standard normal curve to compute a p-value. For the primary pregnancy analysis report, a log rank test of the equality of survival curves will be presented through the first 28 days after birth. A nominal (not adjusted for interim analyses) p-value less than 0.05 will be considered statistically significant. The Type 1 error rate for each comparison will be 0.05. Important sensitivity analyses will be conducted by including infants who drop out of the study early as failures to account for the possibility of informative missingness. At interim analyses the log-rank test testing the equality of survival curves will be presented to the DSMB.

## 7. Secondary Outcome Measures

**Secondary Outcome Measures**

- HIV-1 RNA <200 copies/mL at delivery (up to 14 days postpartum) using real-time test results obtained at site laboratories
- HIV-1 RNA <50 copies/mL at delivery (up to 14 days postpartum) using batched test results obtained from central laboratory
- Time to first HIV-1 RNA <200 copies/mL through delivery (up to 14 days postpartum) using real-time results obtained from site laboratories
- HIV-1 RNA <200 copies/mL at 50 weeks postpartum using real-time test results obtained from site laboratories
- HIV-1 RNA <200 copies/mL at delivery using real-time test results obtained from site laboratories and FDA snapshot algorithm
- HIV-1 RNA <200 copies/mL at 50 weeks postpartum using real-time results obtained from site laboratories and FDA snapshot algorithm
- Composite outcome of spontaneous abortion (occurring at <20 weeks gestation), fetal death (occurring at ≥20 weeks gestation), preterm delivery (<37 completed weeks), small for gestational age (<10th percentile) or major congenital anomaly
- Ranked composite infant safety outcome through 50 weeks postpartum
- Infant HIV infection at delivery and through 50 weeks postpartum
- Infant death due to any cause through 50 weeks postpartum
- Maternal serum creatinine and creatinine clearance rate (Cockcroft-Gault formula)
- Infant serum creatinine and creatinine clearance rate (Schwartz formula)
- HIV-1 antiretroviral drug resistance mutations at the time of maternal virologic failure (and at study entry for mothers with resistance detected at the time of virologic failure to determine if the resistance was present at enrollment or occurred post-enrollment) using the Stanford algorithm
- HIV-1 antiretroviral drug resistance mutations at the time HIV diagnosis for HIV-infected infants
- Preterm delivery (<37 completed weeks)
- Small for gestational age (<10th percentile)
- Change in maternal weight antepartum (from study entry to delivery)
- Change in maternal weight postpartum (from delivery to 50 weeks postpartum)
- Change in maternal weight overall (from study entry to 50 weeks postpartum)

Analyses to test the equality of differences in event probabilities at delivery will be performed using the same approach as for the primary analyses. This includes estimates of the probability of mothers with HIV-1 RNA <50 copies/mL at delivery, the probability of HIV-1 RNA <200 copies/mL at delivery

and through 50 weeks postpartum as defined by the FDA snapshot algorithm the probability of mothers with adverse pregnancy outcomes, HIV drug resistance mutations acquired post-randomization as defined by the Stanford algorithm, and the probability of infant HIV infection at delivery. A two-sided, two-sample test for the equality of the difference in proportions will be used for the by-arm comparisons.

A non-inferiority analysis of preterm delivery and small for gestational age will be conducted as a secondary analysis. The non-inferiority analysis will be conducted because the protocol team believes that is important to set a predefined decision rule for a conclusion of non-inferiority before the data are observed. A non-inferiority margin of 10% will be used. This margin reflects the judgment of the protocol team. The non-inferiority analysis will be conducted to test if DTG+FTC/TDF is non-inferior to EFV/FTC/TDF, if DTG+FTC/TAF is non-inferior to EFV/FTC/TDF, and if DTG+FTC/TAF is non-inferior to DTG+FTC/TDF. For the non-inferiority analyses, a p-value will not be computed; instead, statistical significance for non-inferiority will be determined by the appropriate limit of a 95% confidence interval rejecting the non-inferiority margin. This procedure for the non-inferiority analyses controls a one-sided Type I error rate of 0.025.

Visit week 50 postpartum cumulative event probabilities will be estimated using the Kaplan-Meier method. This analysis estimates confirmed infant HIV infection through 50 weeks postpartum, and infant mortality through 50 weeks postpartum. Greenwood's formula will be used to compute the standard error of the difference between the estimated probabilities. For the analysis of neonatal deaths, a log rank test for the equality of survival curves will be used.

Time until viral suppression (first plasma HIV RNA viral load <200 copies/mL) from study entry through delivery will be compared using a log-rank test, and will be displayed graphically using the Kaplan-Meier estimate of the survival curve.

Continuous outcomes, which include markers of maternal and infant renal toxicity, will be analyzed using two sample t-tests. A linear mixed effects model with a robust variance estimator will also be used to estimate the rate and change between arms of creatinine clearance over time.

Infants and pregnancy outcomes will be classified into the worst group using the following hierarchy:

1. Infant death
2. Spontaneous abortion (<20 weeks gestation) or fetal death (≥20 weeks gestation)
3. Infant HIV infection
4. Extremely and very early preterm (<32 completed weeks)
5. Major congenital anomaly
6. Preterm delivery (<37 completed weeks)
7. Small for gestational age (<10th percentile)
8. Hospitalization
9. Grade 3 or 4 adverse event
10. None of the above

Proportions of participants in each classified group will be summarized and all three pairwise regimen comparisons will be compared using ordinal logistic regression. In addition to the standard ordinal logistic regression analysis, a weighted ordinal logistic regression will be employed to give each outcome type the same weight, using the following equation,

$$\frac{\#\ of\ participants}{(\#\ of\ participants\ with\ a\ given\ outcome)\ X\ (\#of\ hierarchy\ levels)}$$

At each level in the hierarch we will set a binary event threshold (i.e., events at or above and below a given level). For each threshold, we will fit a logistic regression model.

Longitudinal maternal pregnancy weight, postpartum weight, and weight since randomization will be compared by arm with a time and arm interaction using generalized estimating equations. Analyses of weight will be compared for all pairwise comparisons and by comparing the EFV-containing arm to the DTG-containing arms. A sensitivity analyses will be conducted using inverse probability weighting on gestational age at delivery. A two-sided, two-sample test for the equality of the difference in proportions of women defined as being obese or overweight at postpartum week 50 will be applied for the by-treatment arm comparisons.

**Missing Data Methods for Secondary Analyses**

Missing data methods will be used there is a large amount of missing data (e.g. >10% missing) or if requested by the study team. This section lays out the planned methods to account for missing data. Missing data methods will only be used to investigate the robustness of the complete case analyses, as previously discussed.

Multiple imputation, as described in the primary efficacy analysis section, will be used for the following outcome measures:

- HIV-1 RNA <200 copies/mL at delivery (up to 14 days postpartum) using real-time test results obtained at site laboratories
- HIV-1 RNA <50 copies/mL at delivery (up to 14 days postpartum) using batched test results obtained from central laboratory
- HIV-1 RNA <200 copies/mL at 50 weeks postpartum using real-time test results obtained from site laboratories
- Infant serum creatinine and creatinine clearance rate (Schwartz formula)

The FDA snapshot algorithm accounts for missing data by including a categorization of missed visits and study discontinuations; therefore, no special missing data methods will be used for the following outcome measures:

- HIV-1 RNA <200 copies/mL at delivery using real-time test results obtained from site laboratories and FDA snapshot algorithm
- HIV-1 RNA <200 copies/mL at 50 weeks postpartum using real-time results obtained from site laboratories and FDA snapshot algorithm

Women or infants, as appropriate, who discontinue study early before experiencing an outcome listed below will be counted as failures (as an additional component of the outcome) for the following outcomes:

- Composite outcome of spontaneous abortion (occurring at <20 weeks gestation), fetal death (occurring at ≥20 weeks gestation), preterm delivery (<37 completed weeks), or small for gestational age (<10th percentile) or major congenital anomaly

- Ranked composite infant safety outcome through 50 weeks postpartum
- Infant HIV infection at delivery and through 50 weeks postpartum
- Infant death due to any cause through 50 weeks postpartum
- HIV-1 antiretroviral drug resistance mutations at the time of maternal virologic failure (and at study entry for mothers with resistance detected at the time of virologic failure to determine if the resistance was present at enrollment or occurred post-enrollment) using the Stanford algorithm
- HIV-1 antiretroviral drug resistance mutations at the time HIV diagnosis for HIV-infected infants
- Preterm delivery (<37 completed weeks)
- Small for gestational age (<10th percentile)

In addition to counting mother-infant pairs as failures if they discontinue the study early before experiencing an outcome, a second sensitivity analysis will count missing data as failures for the primary adverse pregnancy outcome measure (composite outcome of experiencing spontaneous abortion, fetal death, preterm delivery, or small for gestational age).

## 8. Other Outcome Measures

**DXA Scans**

**The following other outcome measures (stated in protocol version 2.0 LoA #3) address secondary objectives of bone mineral toxicity:**

- **Maternal lumbar spine and hip BMD Z-scores based on DXA scan at 50 weeks postpartum**
- **Infant whole body and lumbar spine BMC values based on DXA scan at 26 weeks postpartum**

**Differences between treatment arms in maternal BMD Z-scores (lumbar spine and hip) and infant BMC values (whole body and lumbar spine) will be analyzed using two sample t-tests assuming unequal variance. Infant whole body values will be analyzed with and without the head. For maternal DXAs a sensitivity analysis testing for the interaction of treatment arm and duration exposed to Medroxyprogesterone acetate will be included. Differences between pooled groups comparing DTG to EFV and TDF to TAF will also be summarized descriptively.**

**Pregnancy Dating**

Accuracy of the estimated date of delivery (EDD) is important for clinical care and identifying adverse pregnancy outcomes. The ACOG algorithm is used to estimate the EDD on IMPAACT 2010 and determines the estimated rates of adverse pregnancy outcome measures such as preterm delivery and infants small for gestational age. The estimate of gestational age for this study uses a centrally calculated value based upon the ultrasound fetal parameters in conjunction with the ACOG algorithm. This strategy has not been used in other IMPAACT studies. This approach was taken to minimize two sources of variability: 1) site-level variability due to different ultrasound machines and formulas, and 2) individual level ultrasound and LMP uncertainty due to varying fetal growth patterns, timing of ovulation, and recall in memory. The first source of variability is accounted for by applying a single formula to the fetal parameters. The second source of variability is accounted for using the ACOG algorithm that redates the gestational age based on the LMP and ultrasound. It is unknown how much this strategy will reduce variability in the estimate of gestational age. To help plan for future studies we will compare the variability using ultrasound machine alone, LMP alone, the ACOG with the ultrasound machine estimate, and IMPAACT 2010 estimate. Rates of EDD

redating using the ACOG algorithm will also be described by gestational age at ultrasound.

## 9. Summaries to Investigate the Impact of SARS-COV-2

**The SARS-COV-2 pandemic occurred during the IMPAACT 2010 follow-up. The majority of women and infants completed study follow-up before the SARS-COV-2 pandemic, however, there were a handful of women and infants who still needed to complete week 38 and week 50 study visits. On March 31st, 2020 the study team sent a clarification memo to the sites that widened the week 50 window and clarified how to conduct at home visits (see SAP Section 11). Summaries that describe the impact of SARS-CoV-2 on IMPAACT 2010 will focus on the week 38 and 50 visits, and measurements of viral load, infant growth, and maternal weight. The pre-pandemic period will be defined as before March 31, 2020, and the post-pandemic period will be defined on and after March 31, 2020.**

**Data Quality Summaries**

- **Frequency of COVID-related missingness in the post-pandemic period overall and over time**
- **Summaries comparing pre-pandemic period to post-pandemic period:**
  - **Frequency of the week 50 viral load, infant growth, and weight data availability**
  - **Per-person per-month rate of reported adverse events**
  - **Per-person per-month rate of virologic failures**

**Retention**

- **Percentage of women and infants discontinue the study early comparing the pre-pandemic versus the post-pandemic period**

**Visit Completion**

- **Percent of missed week 38 and week 50 visits overall and by site comparing the pre-pandemic period to the post-pandemic period**
- **Percent of post-pandemic week 50 visits that were conducted within the pre-pandemic week 50 window (through week 56) compared to the extended visit window (week 74)**

**Outcome**

- **Multiple imputation of the missing week 50 viral loads, maternal weight, and infant growth**
- **Multiple imputation of the week 50 viral loads, maternal weight, and infant growth considering participants who were missing these outcomes or with measurements outside of the pre-pandemic window treated as missing**

## 10. Outline of Report Contents

The following summaries are planned content of the final analysis report:

1. Accrual
2. Eligibility Violations
3. Selected Characteristics at Baseline

4. Study Status and Loss to Follow-up
5. Study Treatment Status and Duration of Treatment
6. Safety
7. Pregnancy Outcomes
8. Repeat Pregnancy Frequency and Outcomes
9. Primary Outcome Measures
10. Secondary Outcome Measures (those required to be submitted to ClinicalTrials.gov)

## 11. Protocol History

IMPAACT 2010 protocol version 1.0 was finalized in December 2016. Clarification Memoranda (CM) #1 and #2 were also finalized in July 2017 before study opening. Protocol version 2.0 was implemented on December 8, 2017 before study opening. Major changes to protocol version 2.0 were primarily due to requests from the FDA and DSMB.

CM #1, July 21, 2017:

- IMPAACT P1026s TAF PK results met the criteria for opening IMPAACT 2010 to accrual.

CM #2, July 21, 2017:

- The Division of AIDS Table for Grading the Severity of Adult and Pediatric Adverse Events (DAIDS AE Grading Table) used was updated to Version 2.1, dated March 2017.
- Fetal ultrasound scans performed during the current pregnancy prior to the study screening period may also be used. The earliest results should be used for estimating gestational age.
- Clarifications were added about recording requirements for maternal and infant physical examinations, concomitant medications, and ultrasound scans.

At the suggestion of the DSMB and FDA, the following major changes were implemented to the protocol:

Protocol Version 2.0, December 8, 2017:

- The primary objective testing whether treatment initiated during pregnancy with a DTG-containing regimen is superior to the EFV-containing regimen with regard to virologic efficacy at delivery was changed to a non-inferiority test. The secondary objective tests superiority.
- The total sample size increased from 549 mother-infant pairs to 639 mother-infant pairs to increase the power for the non-inferiority analysis.
- The accrual cap was lifted for more than 30 mother-infant pairs to enroll each month.
- Comparing the time to maternal HIV-1 RNA <200 copies/mL between a DTG-containing regimen to EFV/FTC/TDF was added as a secondary objective.
- Comparing adverse pregnancy outcomes, maternal adverse events, and infant adverse events between a DTG-containing regimen to EFV/FTC/TDF was added as a secondary objective.
- The allowable window for conducting a Delivery Visit was expanded to 27 days after the pregnancy outcome. The targeted window for Delivery Visits remains up to 14 days and the primary and secondary delivery outcome measures will only be used if measured within the targeted 14 day window.
- An exclusion criterion was added to exclude mothers with a history of antiretroviral drug

resistance mutations that would impact selection of ART regimen.
- If study-supplied study drugs of DTG, FTC/TAF, FTC/TDF, or EFV/FTC/TDF are not available, non-study supplies of these ARVs were approved to be prescribed and dispensed through study site pharmacies, following review and approval by the CMC. Small for gestational age will be calculated from the INTERGROWTH-21$^{st}$ Project, which is a complement to WHO published international growth standards by developing international growth standards for fetuses and infants.
- EPDS, PSQI, and GAD-7 will also be administered at maternal early discontinuation visits.
- When visit windows overlap by a period of one day; the day of overlap was determined to be prioritized for the completion of the earlier of the two visits. Missed evaluations were also approved to be taken at the next scheduled visit.
- Exceptions to antiretroviral receipts within six months prior to study entry were contracted to TDF or FTC/TDF for pre-exposure prophylaxis instead of all ARVs.

On July 27, 2018, LoA #1 for protocol version 2.0 was distributed to sites. LoA #1 was in response to results from an ongoing observational study in Botswana reporting the potential increased risk of neural tube defects (NTDs) associated with exposure to DTG at the time of conception. These findings were not of immediate concern for mother-infant pairs on IMPAACT 2010 because DTG is initiated at or after 14 weeks gestation. However, women may become pregnant on study prior to index pregnancy and might be exposed to DTG. Because of this, accrual was paused on May 18, 2018 until LoA #1 could be implemented. The major changes implemented by LoA #1 for protocol version 2.0 were as follows:

- At study entry, mother reports that she does not wish to become pregnant again for at least 50 weeks after her current pregnancy and that she is willing to use effective contraception during this period.
- Pregnancy testing is required at each scheduled postpartum study visit.
- Maternal fertility intentions , willingness to use effective contraceptive postpartum in relation to study requirements, and contraception history will be collected
- Glucose, hematocrit, and stored whole blood for folate and HbA1c will be collected for testing
- Contraception counseling and a documentation plan for initiating effective contraception postpartum will be provided
- EAE reporting categories for women are changed in Arm 3 from SUSARs to SAEs. EAE reporting across arms will be consistent.
- Women who conceive while taking DTG postpartum will be switched from DTG to an alternate ARV if still in the first trimester when the pregnancy is identified.
- Mothers who become pregnant on study should be offered an ultrasound scan between 14 and 22 weeks gestation to confirm the dating of pregnancy and evaluate fetal anatomy.
- Exploratory outcomes were added to Section 9.2.3:
    - Maternal folate, glucose, and HbA1c
    - Infant folate and glucose

Clarification Memorandum (CM) #1 for protocol Version 2.0 was distributed on September 25, 2018. This was due to a study team concern with the definition of grade 3 or 4 creatinine clearance: <60 mL/min or mL/min/1.73m$^2$ or ≥30% decrease from baseline (similarly, an increase of at least 1.5 x ULN creatinine above baseline). The team considered a percent decrease in creatinine clearance from baseline as unimportant for IMPAACT 2010 because large decreases in creatinine clearance are expected after delivery. This memo specified that grades should be based on absolute values of

creatinine and creatinine clearance only, and not on grades that are based on changes from baseline. The CM also noted that primary maternal safety analysis would only include creatinine and creatinine clearance events based upon absolute values rather than changes in value from baseline.

LoA # 2 for protocol Version 2.0 was finalized on August 23, 2019 for the following reasons:

- Allow plasma stored at the Screening Visit for HIV-1 RNA viral load testing to describe the virologic profiles of enrolled mothers in response to the higher than expected suppression rates among women at entry
- Clarify that the primary outcome measure of maternal grade 3 or 4 adverse events would only include creatinine or creatinine clearance adverse events that met the DAIDs grading table criteria for a grade 3 or higher event based on values alone and not change in levels.
- Clarify infant bone outcome measures should be based on absolute bone mineral content (BMC) values instead of Z-scores, because international normative BMC values have not been established for infants.
- Add a secondary study objective and outcome measures and statistical analyses to describe the effects of study drug regimens on changes in maternal weight

**LoA #3 for protocol Version 2.0 was finalized on June 10, 2020 to incorporate the contents of protocol Clarification Memorandum (CM) #2 (issued on March 31, 2020) that implemented safeguards for study participants against SARS-CoV-2. Sites were instructed to follow applicable government, health authority, and institutional policies with respect to conduct of study visits and procedures, with utmost importance placed on the health and well-being of study participants and study staff. Week 50 visits were broadened to include up to 12 weeks before and up to 24 weeks after the visit target date. Alternate laboratories using alternate assays were permitted if labs could not be performed within a site's Protocol Analyst List. Sites with limited capacity to conduct in-person study clinic visits were advised to prioritize maternal HIV-1 RNA viral load testing and infant diagnostic HIV nucleic acid testing. Subsequently, maternal serum creatinine, followed by other maternal chemistries and hematology were the next highest priority for data collection. The following were allowed to be administered or obtained remotely: medical and medication histories, adherence assessments, counseling, and support, contraception counseling, and questionnaires. DXA scans could be missed or skipped. Sites were instructed to document when missed, out-of-window, off-site, remote contact, or impartial visits occurred, alternate laboratories or assays were used, or alternate study drugs were provided.**

# 12. Appendix

## SAP Version History

| Version | Changes Made | Effective Date |
|---|---|---|
| 1.0 | Original Version | 12/08/17 |
| 2.0 | • Added strategies to handle missing data in study design section<br>• Further defined baseline and study visit weeks<br>• Added missing data methods for secondary analyses<br>• Clarified definition for gestational age and added hierarchy of estimation methods to account for missing data<br>• Updated dates for version 2.0 of the protocol in Protocol History<br>• Added information on study randomization<br>• Added outline of report contents<br>• Added protocol history for LoA #1 for protocol version 2.0<br>• Based on DSMB recommendation, principle results will be released after all women have pregnancy outcome results, while follow-up to 50 weeks postpartum is continuing<br>• Defined gestational age strata in Section 6 | 10/4/18 |
| 3.0 | • Added analysis of primary efficacy outcome measure by maternal suppression status at baseline<br>• Added inclusion criteria for creatinine and creatinine clearance events in the primary maternal safety analysis<br>• Added further analyses for creatinine clearance<br>• Added protocol history for CM #1 for protocol version 2.0<br>• Added signature page | 03/13/19 |
| 4.0 | • Describe final pregnancy analysis report and analyses included<br>• Updated formatting of outcome measures<br>• Provided further clarification for creatinine and creatinine clearance events for primary maternal adverse safety outcome measure<br>• Added secondary objective and outcome measure on maternal weight change<br>• Added analysis for maternal weight change<br>• Added analysis of pregnancy dating methods<br>• Added major changes that were included in LoA #2 for protocol version 2.0 | 10/17/19 |
| 5.0 | • Added analysis of DXA scan results<br>• Added logistic regression and weighted ordinal logistic models to hierarchy analysis<br>• Added analyses that describe effect of SARS-CoV-2 on study | 10/26/2020 |