# HVTN 505 Statistical Analysis Plan: Protocol V6 (Extended Follow-Up)

October 3, 2014

Study Statisticians
Peter Gilbert, Doug Grove, Holly Janes, Shannon Grant

Statistical Center for HIV/AIDS Research and Prevention
Vaccine Infectious Disease Division
Fred Hutchinson Cancer Research Center

Describes analysis methods to be used during extended follow-up of study participants as specified in Protocol V6

# Contents

# 1 Objectives of HVTN 505

HVTN 505 is a randomized, double-blind, placebo-controlled Phase IIb preventative HIV vaccine trial, conducted in Ad5 neutralizing antibody negative (titers $\leq 18$), circumcised, men and male-to-female (MTF) transgender persons who have sex with men, in the United States, ages 18-50 years, who satisfy criteria for being at high risk for acquiring HIV-1 infection.

Following the decision to halt study vaccinations after the April 22, 2013 DSMB review, the study transitioned into an extended follow-up phase. At the second interim analysis of the extended follow-up phase, no significant difference was found in the HIV-1 infection rate vaccine vs. placebo, and there was low conditional power to detect an increased rate of infection in the vaccine arm. Based on these results, the protocol was modified again (version 6) to reduce the intensity of study visits; only annual visits will be conducted after Month 24, and HIV-1 infected participants will only have three post-infection visits (at weeks 2, 4, and 24 post-infection diagnosis; PD). This document describes statistical methods that will be used to analyze data collected under protocol version 6.

## 1.1 Primary Objective

1. To evaluate the rate of study dropout in vaccine and placebo recipients
2. To evaluate the effect of the VRC DNA/rAd5 vaccine regimen on the rate of HIV-1 acquisition compared to placebo

## 1.2 Exploratory Objectives

1. To evaluate the effect of the VRC DNA/rAd5 vaccine regimen on viral load (VL) at the time of HIV-1 infection diagnosis
2. To evaluate the effect of the VRC DNA/rAd5 vaccine regimen on CD4+ T cell count and disease progression course
3. To evaluate HIV-1-specific and Ad5 vector-specific immune responses induced by the vaccine regimen
4. To evaluate the impact of viral genetic variation, host genetic factors, prophylactic ARV use, and other participant covariates including self-reported risk behavior on vaccine effects on study endpoints
5. To describe the prevalence of drug (TDF and FTC) resistance mutations after HIV-1 seroconversion among vaccine and placebo recipients reporting prophylactic ARV use

6. To evaluate immune correlates of risk of HIV-1 infection among vaccine recipients and (possibly) placebo recipients

# 2 Follow-Up Period

The goal is to follow all HIV-1 uninfected participants for 60 months post-enrollment. This will include 48 months of clinic visits plus a participant health contact at Month 60. Under version 4 of the protocol, participants were followed for 24 months post-enrollment. Under protocol version 5, participants who had previously completed 24 months of follow-up were brought back to the study sites to begin the extended follow-up.

Participants who are diagnosed with HIV-1 infection will be followed for 6 months post-infection diagnosis. Under version 4 of the protocol, HIV-1 infected participants were followed for 18 months post-enrollment. Participants diagnosed under version 5 of the protocol were followed only for 6 months post-infection diagnosis. Under version 6, infected participants will still be followed through 6 months post-infection, but the number of visits has been reduced to three: at weeks 2, 4, and 24 post-infection.

The analyses described in this document will use data from one of two follow-up periods for HIV-1 uninfected participants, as specified in detail below: 1) 0-24 months follow-up, or 2) including all follow-up through Month 48. Analyses of HIV-1 infected participants will use all available post-infection follow-up: 18 months for participants diagnosed under version 4 of the protocol (or data through the last post-infection visit), and 6 months for participants diagnosed under versions 5 or 6.

# 3 Study Populations

We define six study populations that are analyzed for addressing various study objectives. This terminology is used throughout the SAP.

**Safety Population:** Randomized subjects who receive at least one study injection of vaccine of placebo, i.e. all enrolled subjects

**Modified Intent-to-Treat (MITT) Population:** Subjects in the Safety Population who are HIV-1 negative on the date of first injection (Day 0)

**Week 28+ Population:** Subjects in the MITT Population who have an HIV-1 negative test result at or after Day 196, or their first positive HIV-1 test at or after Day 196

**Immunogenicity Population:** MITT subjects who have an HIV-1 negative test result at Month 7

**Immunogenicity Subcohort:** MITT subjects in the Immunogenicity Population for whom Month 7 immune responses are measured.

**MITT Infected Population:** MITT subjects who have a confirmed diagnosis of HIV-1 infection during the follow-up period through the Month 48 visit.

**Week 28+ Infected Population:** MITT-Infected subjects whose date of infection diagnosis is on or after Day 196 through their Month 48 visit.

**Pre-Week 28 Infected Population:** MITT-Infected subjects whose date of infection diagnosis is before Day 196.

The MITT Population and the Safety Population are very similar but not identical to a fully intention-to-treat population; the Safety Population differs by excluding randomized subjects who do not enroll; and the MITT population is the subset of the Safety Population that also excludes enrolled subjects discovered later to be HIV-positive by Day 0. Because of blinding and the brief length of time between randomization and enrollment– typically no more than 4 working days according to the HVTN Manual of Operations (HVTN MOP)– we expect almost all randomized subjects to be in the Safety Population. Given that eligibility for the study requires recent evidence of being HIV-1 uninfected (within 56 days prior to enrollment), we expect almost all enrolled subjects to also be in the MITT Population.

Since enrollment is concurrent with receiving the first vaccination, all participants will have received one vaccination and therefore are part of the Safety Population and will provide some safety data.

The primary analyses will consider all MITT infections. Week 28+ infections were the primary infection endpoints under version 4 of the protocol, because earlier "inter-current" infections would likely have occurred prior to the rAd5 vector vaccine/placebo boost and might have become established before the development of a complete vaccine-induced immune response. Under versions 5 and 6 of the protocol, primary interest lies in all HIV-1 infections, and this motivates the focus on all MITT infections. Supportive secondary analyses will restrict attention to Week 28+ infections.

Analyses of immunogenicity endpoints will be conducted in the Immunogenicity Population. The primary immunogenicity time-point is the Month 7 visit.

# 4 Diagnosis of HIV-1 Infections

The primary endpoint is diagnosis of HIV-1 infection during the follow-up period; additional endpoints are assessed in study participants diagnosed with HIV-1 infection. The occurrence of HIV-1 infection will be detected through HIV-1 ELISA tests administered at study visits at Months 6, 9, 12, 15, 18, 21, 24, 36, and 48 post-enrollment (see protocol Appendix B). Participants found to have evidence for HIV-1 infection will have additional testing using western blot and RNA PCR to confirm the diagnosis of HIV-1 infection.

The vaccine-induced immune responses may lead to false positive HIV-1 tests and difficulty in interpretation. Therefore the study utilizes a blinded Endpoint Adjudicator to review all data that support diagnoses of HIV-1 infection. Only HIV-1 infection cases confirmed by the Adjudicator will be counted as HIV-1 infections in the analysis.

## 4.1 Date of Infection Diagnosis

The Adjudicator will define the date of diagnosis based on his/her judgement of all of the available diagnostic data. This date is assigned based only on diagnostic results collected prospectively over time; it does not consider the results of HIV-specific PCR tests that may be performed on earlier samples that are tested later. The date of infection diagnosis will be used as the event-date for time-to-event analyses of HIV-1 infection diagnosis and as the time origin for the analysis of post-infection endpoints.

## 4.2 'Look-back' Procedure for HIV-Specific PCR Testing

For all subjects diagnosed with HIV-1 infection, the sample available at the nearest date before the diagnosis date will be tested using HIV-specific PCR. If it is positive, then the sample at the second nearest date before the diagnosis date will be tested using HIV-specific PCR. This procedure will be repeated until an HIV-specific PCR negative test result is obtained or until a test is done for a Day 0 sample. We define the 'earliest post-infection time-point' (EPIT) as the earliest date for which an HIV-specific PCR test is positive.

# 5 Timing of Final Analysis

The final analysis will take place soon after the last enrolled participant has reached the end of the Month 48 visit window; we refer to this as the final evaluation time (FET). As described in the monitoring plan in Section 8, the final analysis make take place earlier based on sequential monitoring of study dropout and the vaccine effect on HIV-1 acquisition.

# 6 Endpoints

## 6.1 Primary Endpoints

For each study endpoint defined below, we include the population in which it is assessed. Assessments of viral load (VL) and CD4+ T cell endpoints in HIV-1 infected subjects will exclude values measured after antiretroviral therapy (ART) initiation, because ART will typically have a strong effect on these biomarkers. Methods that assess VL endpoints that use actual values (not rank-based methods) will use the following approach to handle VL endpoints that are left- or right-censored by the quantification limits (40 to 10,000,000) of the primary VL assay. Left-censored values will be assigned value $\log_{10}(20)$ and right-censored values will be assigned value $\log_{10}(10,000,000) = 7$.

Study dropout is defined as termination from the study. Participants who terminate will be considered to have dropped out at the date of their last visit.

**Primary Endpoint 1 (MITT Population):** Study dropout through the Month 48 visit

**Primary Endpoint 2 (MITT Population):** HIV-1 infection diagnosed after Day 0 through the Month 24 visit

**Primary Endpoint 3 (MITT Population):** HIV-1 infection diagnosed after Day 0 through the Month 48 visit

## 6.2 Exploratory Endpoints

This section lists endpoints for several exploratory objectives of key interest. The full list of exploratory objectives is given in Section 1.2.

**Exploratory Endpoint 1 (MITT Infected Population)**: Diagnostic VL, defined as that obtained from the sample drawn at the visit at which a participant is diagnosed with HIV-1 infection

**Exploratory Endpoint 2 (MITT Infected Population)**: Measures of HIV-1 disease progression including: CD4+ T cell counts in HIV-1 infected volunteers; ART initiation; HIV-1 related clinical events and death

Under protocol version 6, CD4+ T cell counts are measured at Week 2 PD and ART initiation is assessed at Weeks 2, 4, and 24 PD (see protocol Appendix D). Since primary interest is in studying ART initiation as an indicator of disease progression, participants initiating ART before Week 10 PD will be censored at the time of ART initiation.

If and when HIV-1 infected participants experience an HIV-1 related clinical event, or die for any reason, will be recorded. HIV-1 related clinical events are defined as any documented AIDS-defining condition in the CDC 2008 definition (listed in Appendix A of Schneider

et al., 2008). The protocol chair and co-chairs will review each HIV-related clinical event and certify its acceptance as an endpoint for analysis. Other disease progression related events (such as cardiac, hepatic, renal, metabolic, and non-AIDS malignancies) will also be catalogued and analyzed.

**Exploratory Endpoint 3 (Immunogenicity Population)**: HIV-1 specific and rAd5 vector specific immune responses measured using

- HIV-1-specific and rAd5-specific unfractionated IFN-$\gamma$ ELISpot
- HIV-1-specific and rAd5-specific Intracellular cytokine staining (ICS)
- HIV-1-specific and rAd5-specific epitope mapping
- HIV-1 and rAd5 binding ELISA
- HIV-1 and rAd5 neutralization assay

The primary time-point for immunogenicity assessment is the Month 7 visit, 4 weeks after the rAd5 boost, and optionally at additional immunogenicity time points.

**Exploratory Endpoint 4 (MITT Population):** Host genetics

Host genetic factors including HLA alleles and KIR genes will be measured in all subjects in the MITT Infected Population and potentially also in the Immunogenicity Subcohort.

**Exploratory Endpoint 5 (MITT Population):** HIV-1 genetics

Full genome HIV-1 sequences will be measured from each subject in the MITT Infected Population.

**Exploratory Endpoint 7 (MITT Population):** Participant risk behavior and prophylactic ARV use

Participant risk behavior and prophylactic ARV use as measured by behavioral risk factor and prophylactic ARV use questionnaire as well as plasma ARV drug level testing. The plan for plasma ARV drug level testing is described in Section 7.9.

# 7 Statistical Analyses for Primary and Exploratory Endpoints

All vaccine activity and efficacy endpoints will be evaluated in the MITT population. Primary analyses will include all infections occurring in the MITT population, while supportive analyses will include only Week 28+ infections. Specifically, for analyses including only

Week 28+ infections, analyses of post-infection vaccine activity endpoints will only evaluate HIV-1 infections diagnosed at or following Day 196 and the analysis of HIV-1 infection will right-censor HIV-1 infection events diagnosed prior to Day 196.

Analyses of HIV-1 acquisition will be performed using either follow-up through Month 24, or all follow-up through Month 48. Correspondingly, these analyses will target estimation and inference of $\text{VE}^{\text{MITT}}(24)$, the multiplicative reduction in the hazard rate of HIV-1 infection (vaccine versus placebo) by the Month 24 visit, and estimation and inference of $\text{VE}^{\text{MITT}}(t)$, the multiplicative reduction by time $t$ post-enrollment for all times through to the Month 48 study visit. Analyses of $\text{VE}^{\text{MITT}}(24)$ will right-censor participants at the end of the Month 24 visit window, or, for participants who drop out prior to the Month 24 visit, at the date of the last available HIV test result. Analyses of $\text{VE}^{\text{MITT}}(t)$ will right-censor participants at time $t$, or, for participants who drop out prior to time $t$, at the date of the last available HIV test result.

Similarly, analyses of study dropout will be performed using either follow-up through Month 24, or all follow-up through Month 48. Analyses of dropout within 24 months will right-censor participants diagnosed with HIV-1 infection at the date of diagnosis and participants not dropping out or becoming infected at the end of the Month 24 visit window. Analyses of dropout including all available follow-up will right-censor participants diagnosed with HIV-1 infection at the date of diagnosis and participants not dropping out or becoming infected at the date of the last visit.

## 7.1 Behavioral risk score

Several analyses will make use of a baseline behavioral risk score for predicting HIV-1 infection risk. This score was developed using all data prior to study unblinding, using a multivariable Cox proportional hazards model and an all-subsets model selection procedure. The model includes two variables from the baseline behavioral risk score, which queried participants about risk behavior over the 3 months prior to enrollment: an indicator that the number of male sexual partners is greater than three and an indicator of unprotected receptive anal sex. The risk score is a weighted sum of these risk variables, each weighted by the estimated hazard ratio (HR). The risk score takes a value of 0 if a participant has neither risk factor, 1 if they have both, and an intermediate value (0.46 or 0.54) if they have one or the other.

Whereas all analyses that account for behavioral risk will use this *baseline* behavioral risk score, some analyses will also use the *post-baseline* behavioral risk score measured at post-infection visits. The behavioral risk score at a post-baseline visit is defined in the same way as for the baseline score, but uses the recorded values of the risk score variables at the given visit.

## 7.2 Primary Endpoint (Study Dropout)

### 7.2.1 Primary analysis

The timing of study dropout will be described for each treatment group using a cumulative incidence curve of the time between first injection and the date of dropout. This curve will be estimated by one minus the Kaplan-Meier estimator. The incidence of study dropout in each treatment group will be estimated by dividing the number of dropouts by the amount of person-time "at risk", and exact methods will be used to calculate 95% confidence intervals. To test for a difference in dropout rates between treatment groups, a score test will be used in the context of a Cox proportional hazards model. The proportional hazards assumption will be tested using the Grambsch and Therneau (1994) test based on Schoenfeld residuals; if the assumption is violated additional Cox regression modeling analyses will be performed that include time-dependent interactions between the natural logarithm of failure time and treatment arm.

To evaluate differences in dropout rates pre- vs. post-unblinding, the above analyses will be performed separately for the pre-unblinding period (censoring all subjects still "at risk" on April 22, 2013, the date of unblinding) and for the post-unblinding period ("starting the clock" for subjects on April 23, 2013). To test for a difference in dropout rates pre- vs. post-unblinding, a score test for the time-dependent indicator of post-unblinding follow-up time (an indicator as to whether the follow-up time is after April 22, 2013) will be reported. This test will be performed pooling treatment groups, and separately for each treatment group.

The instantaneous dropout rate will also be estimated as a function of follow-up time using the method of Gilbert et al. (2002). This analysis will be performed for both the pre-unblinding follow-up period and for the post-unblinding follow-up period, separately for each treatment group.

### 7.2.2 Secondary analysis

Secondary analyses of study dropout will restrict attention to Months 0-24 follow-up. These analyses will inform the validity of vaccine efficacy analyses restricted to Month 0-24 follow-up.

Other secondary analyses will assess baseline participant characteristics as predictors of study dropout. The baseline variables considered will be age, baseline behavioral risk score, race/ethnicity, HSV-2 serostatus, and BMI. Note that HSV-2 serostatus will not be considered if fewer than 95% of participants have measurements; this assay is specified as optional in the protocol. These covariates will be assessed as predictors in Cox proportional hazards models, both univariately and in the context of a multivariate model comprised of covariates predicting dropout with significance $p < 0.10$ based on two-sided score tests in univariate

Cox models.

## 7.3 Primary Endpoint (HIV-1 Infection Diagnosis)

### 7.3.1 Primary analysis

The timing of infections will be described for each treatment group using a cumulative incidence curve of the time between first injection and the date of HIV-1 infection diagnosis. This curve will be estimated by one minus the Kaplan-Meier estimator.

The vaccine effect on HIV-1 acquisition will be measured using the relative risk ($RR$) of HIV-1 diagnosis, defined as the hazard ratio (vaccine/placebo) in a continuous time Cox proportional hazards model. Conditional on a pre-specified method selection step described below, the method of Lu and Tsiatis (2008) will be used for estimating the $RR$ with a 95% Wald-based confidence interval, and for testing whether the $RR$ differs from 1 with a Wald statistic. If the pre-specified method selection step does not support use of the Lu and Tsiatis (2008) method, the standard maximum partial likelihood estimator will be used for estimating the $RR$ with a 95% Wald-based confidence interval, and a log-rank test will be used for testing whether the $RR$ differs from 1.

Given that the vaccine and placebo groups will be compared over a time period when participants are unblinded as to treatment assignment, the potential for confounding will be carefully considered. Specifically, in addition to estimating the marginal HR associated with vaccine assignment, we will also estimate the HR adjusted for potential confounding factors. First, we will assess the extent to which baseline participant characteristics and baseline behavioral risk score are predictive of HIV-1 infection risk. Variables that, when considered univariately and pooled over treatment arms, predict infection risk with significance $p < 0.10$ based on two-sided score tests in the Cox model will be added to the Cox model relating treatment assignment to HIV-1 infection risk. The baseline variables considered will be age, baseline behavioral risk score, race/ethnicity, HSV-2 serostatus, and BMI. Note that HSV-2 serostatus will not be considered if fewer than 95% of participants have measurements; this assay is specified as optional in the protocol. Both unadjusted HRs for vaccination (without adjustment for the baseline covariates) and adjusted HRs (with adjustment for baseline covariates) will be reported.

There is also the potential for confounding due to time-dependent factors such as risk behaviors and ARV use. However, adjustment for these factors is complicated by their differing collection schedules, missing data over time, and measurement error. Therefore, we view analyses that adjust for time-dependent confounders as secondary, rather than primary; these are described below in Section 7.3.2.

As mentioned above, under certain conditions instead of using standard Cox partial likelihood methods for estimation we will instead use the efficient estimation method of Lu and Tsiatis (2008). The version of the method that leverages baseline subject characteristics predictive

of HIV-1 infection diagnosis to improve efficiency will be used, which is implemented by the R package *speff2trial* evaluated in simulations by Xiaomin Lu, Michal Juraska, Rong Fu, and Holly Janes. Implementation of the Lu and Tsiatis method divides into three steps:

1. Assume a linear model for the function $f_0(X_1)$ of baseline covariates $X_1$ ($f_0$ is defined in Lu and Tsiatis, 2008, expression (13)) and use all-subsets model selection to pick the best model.
2. Assume a linear model for the function $g_0(X_1)$ of baseline covariates $X_1$ ($g_0$ is defined in Lu and Tsiatis, 2008) and use all-subsets model selection to pick the best model.
3. Solve the estimating equation on page 685 of Lu and Tsiatis (2008, middle panel), which uses fitted values from the best models selected in Steps 1 and 2.

In Steps 1 and 2 separately, BIC will be used as the criterion for choosing the best model. The key for the method to be objective is that the set of baseline covariates to consider for the predictive model and the algorithm for selecting among them is pre-specified. The following baseline covariates will be considered:

1. Age
2. Baseline behavioral risk score
3. Race/ethnicity
4. BMI
5. HSV-2 serostatus
6. All 2-way interactions of the above covariates
7. Square terms for age, baseline behavioral risk score, and BMI

Note that for modeling $f_0$ and $g_0$, as well as for modeling missingness in the primary endpoint viral load analysis, models with 2-way interaction terms always include the main effects.

The pre-specified method selection step is as follows. To use the Lu and Tsiatis method we require that the best treatment-arm-pooled model predicting HIV infection (based on BIC) has a c-statistic of at least 0.85. The purpose of this step is to only use the Lu and Tsiatis method in a scenario where the available covariates are good enough predictors of HIV infection to render the Lu and Tsiatis method meaingfully more efficient than the standard method that ignores baseline covariates. Simulation studies showed that scenarios with c-statistic below 0.85 tended to have less than 1-2% efficiency gain with the Lu and Tsiatis method. The reason to prefer the standard method in the case of no advantage for the Lu

and Tsiatis method, is to avoid potential controversy in the unexpected but possible event that the inferences are somewhat different via the Lu and Tsiatis and standard methods.

To assess the validity of the proportional hazards assumption of the Cox model, goodness-of-fit tests will be performed, including the Grambsch and Therneau (1994) test based on Schoenfeld residuals. If these diagnostics support failure of the proportional hazards assumption, additional Cox regression modeling analyses will be performed that include time-dependent interactions between the natural logarithm of failure time and treatment arm. In addition, the nonparametric smoothing method of Durham et al. (1999) based on Schoenfeld residuals may be used. These analyses will adjust for covariates in the same manner as the Cox model analysis that does not include the time-dependent interactions.

An additional analysis will be considered if marginal and conditional HRs associated with vaccination differ substantially, suggesting confounding, and if in addition the goodness-of-fit diagnostics support failure of the proportional hazards assumption. Under this analysis, cumulative probabilities of HIV-1 infection over time in each treatment arm, and additive differences and ratios of these cumulative probabilities (vaccine vs. placebo), will be estimated adjusted for potential confounders. In addition to the point estimates, 95% confidence intervals about the cumulative probabilities of HIV-1 infection over time for each treatment arm will be computed, as well as 95% confidence intervals about the additive difference and ratios over time. The collaborative targeted maximum likelihood estimation method of Stitelman and van der Laan (2010) will be used for estimation, which in addition to allowing confounding adjustment can correct for potential bias due to covariate-dependent censoring.

### 7.3.2 Secondary analyses

Secondary analyses of vaccine efficacy will consider four subgroups of the MITT population:

1. The Week 28+ population, those on-study on Day 196 and HIV-negative (i.e., not yet diagnosed as infected) prior to that;
2. The Week 28+ population who received rAd5/FFB;
3. The Week 28+ population who did not receive rAd5/FFB; and
4. (Per-protocol) the Week 28+ population who received all four immunizations with correct product administration and within visit windows.

The subgroups will help to address whether and how vaccine efficacy depends on the rAd5 vaccination. For each subgroup analysis, both VE(24) and VE over all available follow-up time will be evaluated. The methods described in Section 7.3.1 will be used for VE estimation for each subgroup and for each follow-up period.

For the analysis of per-protocol (PP) vaccine efficacy, the method of Gilbert, Shepherd, and Hudgens (2013) will additionally be used to assess causal per-protocol vaccine efficacy

among subjects per-protocol under both vaccine and placebo assignments. The method will yield estimated ignorance intervals, 95% estimated uncertainty intervals, and p-values for whether per-protocol vaccine efficacy differs from zero, which account for uncertainty due to partial non-identifiability of the causal vaccine efficacy parameters of interest as well as due to sampling variability.

Additional secondary analyses will assess the possibility of confounding by time-dependent variables. Specifically, if there is evidence of imbalance in the distribution of the two variables comprising the behavioral risk score between treatment arms over time, we will perform an analysis that adjusts for time-dependent predictors of HIV-1 infection risk. To evaluate evidence of imbalance, a logistic generalized estimating equations (GEE) model with exchangeable working correlation will be fit for the two risk behavior variables. We define evidence of imbalance between treatment arms as a two-sided p-value of less than 0.05 for the generalized Wald test of no difference between arms at any point in time, for either risk behavior variable. If this criterion is triggered, we will consider including the following time-dependent variables in the Cox model for estimating vaccine efficacy: a time-dependent version of the behavioral risk score, time-dependent indicators of STIs (gonorrhea, chlamydia, syphilis, and HSV if available), and time-dependent indicators of prophylactic ARV use as measured by questionnaire. The choice of time-dependent variables to include will be made based on data availability and evidence of association between the variables and HIV-1 infection risk.

## 7.4 Exploratory Endpoint 1 (Viral load at HIV-1 diagnosis)

The distribution of pre-ART VL at diagnosis will be compared between treatment arms using a Wilcoxon rank sum (WRS) test. It is anticipated that the rate of missing VL endpoints will be very small, and therefore a "complete case" analysis will be performed. If, however, the rate of missing endpoints is greater than 10% the robust likelihood-based method of Little and An (2004), described below for completeness, will be implemented. The version of the method that also allows for post-randomization selection bias may also be employed if there is evidence of a vaccine effect on the VL endpoint using the WRS test.

The robust likelihood-based method of Little and An (2004) was chosen because it is designed to minimize potential bias in the analysis that could occur due to missing VL endpoint data. Under plausible assumptions it will provide unbiased inferences if the missing endpoint data are missing at random (MAR).

It is possible that the Little and An method will provide unstable estimation. In the event of unstable estimation, then other methods for dealing with missing VL data would likely also be unreliable. In this case VL endpoint will be compared between treatment groups using a difference in sample averages among subjects with observed VL values, with 95% confidence interval and p-value based on the standard t-statistic.

We now describe how the Little and An method will be used in detail. The method is de-

scribed for a generic post-infection endpoint, $Y_p$. Only a subset of MITT infected subjects will contribute a value $Y_p$; subjects who dropped out or initiated ART prior to $Y_p$ measurement would not contribute values. Inferences about $Y_p$ apply to a population where ART is not prescribed before $Y_p$ measurement. Implementation of the method simplifies for $Y_p$ measured at the time of HIV-1 infection diagnosis; for these endpoints no other post-infection data are incorporated into the analysis.

### 7.4.1 Additional notation

Let $Z$ be vaccination assignment ($Z = 1$, vaccine; $Z = 0$, placebo) and let $X$ be a $q-$vector of baseline covariates and pre-infection covariates fully observed for everyone. Let $S$ be the indicator of whether a subject is diagnosed with HIV infection during the follow-up period. Note that for the analysis of the Week 28+ Infected Population, we define $S$ to be 0 if a subject's diagnosis date is before Day 196; whereas for the analysis of the MITT Infected Population, $S = 1$ for all MITT Infected subjects.

Subjects experiencing $S = 1$ are subsequently evaluated at $V$ visits, where variables $Y_1, \cdots, Y_{n_1}$ are collected at visit 1 (visit 1 is the diagnosis date), variables $Y_{n_1+1}, \cdots, Y_{n_1+n_2}$ are collected at visit 2, and so on, with variables $Y_{\sum_{i=1}^{V-1} n_i+1}, \cdots, Y_p$ collected at visit $V$, where $p = \sum_{i=1}^{V} n_i$. The entire collection of $p$ variables measured after $S = 1$ is $Y \equiv (Y_1, \cdots, Y_p)'$, where $Y_p$ is the outcome variable of interest. For $j = 1, \cdots, p$, let $M_j$ be the indicator of whether $Y_j$ is missing, and set $M = (M_1, \cdots, M_p)'$. The variables $Y$ are only meaningful if $S = 1$; therefore for uninfected subjects (i.e., $S = 0$) we set $Y \equiv *$ and $M \equiv *$.

Each participant has potential infection outcome $S(1)$ if assigned vaccine and $S(0)$ if assigned placebo. For $Z = 0, 1$, the potential outcomes $Y(Z) \equiv (Y_1(Z), Y_2(Z), \cdots, Y_p(Z))'$ and $M(Z) \equiv (M_1(Z), M_2(Z), \cdots, M_p(Z))'$ are defined if $S(Z) = 1$; otherwise $Y(Z) \equiv *$ and $M(Z) \equiv *$. With $\mu_z \equiv E(Y_p(z)|S(0) = S(1) = 1)$ for $z = 0, 1$, the *average causal effect* ($ACE$) estimand of interest is $ACE \equiv \mu_1 - \mu_0$. The goal of the primary analysis and the sensitivity analysis is to estimate the $ACE$ based on assumptions and the observed iid data $(Z_i, X_i, S_i, M_i, Y_i)$, $i = 1, \cdots, N$.

### 7.4.2 Assumptions for the primary analysis and for the sensitivity analysis to post-randomization selection bias

Throughout we make the following three identifiability assumptions.

**A1:** Stable Unit Treatment Values Assumption (SUTVA) (Rubin, 1978)

**A2:** The treatment assignment $Z$ is independent of $(X, S(0), S(1), M(0), M(1), Y(0), Y(1))$

**A3:** For infected subjects (with $S = 1$), the missing data mechanism for $Y$ is missing at random (MAR).

For subjects with $S = 1$, let $Y_{obs}$ denote the components of $Y$ that are observed, and $Y_{mis}$ denote the components of $Y$ that are missing. Let $f$ be the conditional cdf of $M$ given $Y$ and $S = 1$, $f(M|Y, S = 1, \nu)$, where $\nu$ denotes unknown parameters. MAR states that missingness depends only on the observed values $Y_{obs}$, i.e.,

$$f(M|Y, S = 1, \nu) = f(M|Y_{obs}, S = 1, \nu) \quad \text{for all } Y_{mis}, \nu.$$

### 7.4.3 Primary analysis and sensitivity analysis

The primary analysis will use the method of Little and An (2004), hereafter LA, to estimate the $ACE$, under A1-A3 and the additional assumption of no post-randomization selection bias. Because this method is a special case of the method used for the sensitivity analysis, we describe the sensitivity analysis method, and note the special case that corresponds to the primary analysis. The LA method was applied to data from the Step and VaxGen trials and validated in simulations by Gilbert and Jin (2010).

Allowing for possible post-randomization selection bias, the sensitivity analysis postulates additional assumptions that identify the $ACE$ and are indexed by fixed sensitivity parameters. Following Jemiai and Rotnitzky (2005), we suppose three models, which we refer to collectively as **A4:**

$$\begin{aligned} g_0(\Pr(S(1) = 1|S(0) = 1, Y_p(0) = y) &= \alpha_0 + \beta_0 y & (1)\\ g_1(\Pr(S(0) = 1|S(1) = 1, Y_p(1) = y) &= \alpha_1 + \beta_1 y & (2)\\ \Pr(S(0) = 1|S(1) = 1) &= \phi, & (3) \end{aligned}$$

where $g_0$ and $g_1$ are known invertible link functions whose inverses are continuous in $\alpha_0$ and $\alpha_1$; $\alpha_0$ and $\alpha_1$ are unknown parameters to be estimated; and $\beta_0$, $\beta_1$, and $\phi$ are known sensitivity parameters that are varied over plausible ranges. The sensitivity analysis will use logit links $g_0$ and $g_1$, in which case $\beta_0$ is interpreted as the difference in the log odds of infection in the vaccine group given infection in the placebo group with $y$ versus $y - 1$; and $\beta_1$ is interpreted similarly reversing the role of vaccine and placebo. The sensitivity analysis will vary $\beta_0$ and $\beta_1$ over the ranges $[-1.10, 1.10]$, which allows for odds ratios as much as 3-fold divergent from 1.0 (1.0 specifies no selection bias).

The parameter $\phi$ is interpreted as the probability that a subject infected in the vaccine group would also be infected in the placebo group. Setting $\phi = 1$ specifies the so-called 'monotonicity assumption,' that the vaccine does not increase the risk of infection for any subject. If the observed data on HIV infection rates support monotonocity (with an estimated hazard ratio (vaccine/placebo) of HIV infection less than or equal to 1), then the sensitivity analysis will be performed with $\phi = 1$. Otherwise, the sensitivity analysis will include values of $\phi < 1$. In particular, following the example described in Gilbert and Jin (2010), we will vary $\phi$ between $\phi_L$ and 1.0, where $\phi_L$ is set to two minus the upper 95% confidence limit for the hazard ratio (vaccine/placebo) of HIV infection. For example, if the upper 95% confidence limit for the

hazard ratio is 1.25, then setting $\phi = 2 - 1.25 = 0.75$ assumes that a vaccinated subject who becomes infected would have a 25% chance of avoiding infection had he/she been assigned placebo, which specifies a 25% plausible elevation of infection risk in the vaccine group.

The primary analysis is obtained by using logit links for $g_0$ and $g_1$ and by setting $\beta_0 = \beta_1 = 0$ and $\phi = 1$. If the data do not support monotonocity (with an estimated hazard ratio (vaccine/placebo) of HIV infection greater than 1), instead the assumption will be made that the vaccine does not decrease the risk of infection for any subject; this corresponds to assuming $\beta_0 = \beta_1 = 0$ and $\phi = p_0/p_1$ where $p_z \equiv \Pr(S(z) = 1)$ for $z = 0, 1$.

### 7.4.4 Estimation of the $ACE$ (given fixed values of $\beta_0, \beta_1, \phi$)

Let $\theta \equiv (p_0, p_1, \alpha_0, \alpha_1, \mu_0, \mu_1)'$. Then, as shown in Gilbert and Jin (2010), an unbiased estimating equation is given by

$$\sum_{i=1}^{N} U_i^M(\theta) = 0,$$

where

$$U_i^M(\theta) = (U_{1i}^M(\theta), U_{2i}^M(\theta), U_{3i}^M(\theta), U_{4i}^M(\theta), U_{5i}^M(\theta), U_{6i}^M(\theta))'$$

and

$$U_{1i}^M(\theta) = (1 - Z_i)(p_0 - S_i) \tag{4}$$

$$U_{2i}^M(\theta) = Z_i(p_1 - S_i) \tag{5}$$

$$U_{3i}^M(\theta) = (1 - Z_i)S_i \left\{ \left[ (1 - M_{pi}) g_0^{-1}(Y_{pi}; \alpha_0, \beta_0) + M_{pi} \widehat{E}_i[g_0^{-1}(Y_p; \alpha_0, \beta_0)] \right] - \phi \frac{p_1}{p_0} \right\} \tag{6}$$

$$U_{4i}^M(\theta) = Z_i S_i \left\{ \left[ (1 - M_{pi}) g_1^{-1}(Y_{pi}; \alpha_1, \beta_1) + M_{pi} \widehat{E}_i[g_1^{-1}(Y_p; \alpha_1, \beta_1)] \right] - \phi \right\} \tag{7}$$

$$U_{5i}^M(\theta) = (1 - Z_i)S_i \left\{ \mu_0 - \left[ (1 - M_{pi}) Y_{pi} g_0^{-1}(Y_{pi}; \alpha_0, \beta_0) + M_{pi} \widehat{E}_i[Y_p g_0^{-1}(Y_p; \alpha_0, \beta_0)] \right] \frac{p_0}{\phi p_1} \right\} \tag{8}$$

$$U_{6i}^M(\theta) = Z_i S_i \left\{ \mu_1 - \left[ (1 - M_{pi}) Y_{pi} g_1^{-1}(Y_{pi}; \alpha_1, \beta_1) + M_{pi} \widehat{E}_i[Y_p g_1^{-1}(Y_p; \alpha_1, \beta_1)] \right] \frac{1}{\phi} \right\}, \tag{9}$$

where

$$\widehat{E}_i[h(Y_p)] \equiv \widehat{E}[h(Y_p) | Y_{1i}^*, X_{2i}, \cdots, X_{qi}, Y_{1i}, \cdots, Y_{(p-1)i}, S_i = 1, Z_i]$$

for a function $h(\cdot)$, and

$$Y_{1i}^* = \text{logit} P(M_{pi} = 0 | X_{1i}, \ldots, X_{qi}, Y_{1i}, \ldots, Y_{(p-1)i}, S_i = 1, Z_i)$$

is the logit of the propensity score for $Y_{pi}$ to be observed for an infected subject.

We will use Gilbert and Jin's (2010) procedure for obtaining the predicted values $\widehat{E}_i[h(Y_p)]$ (described in more detail next), which fully specifies the equations and allows calculation of a solution.

### 7.4.5 Calculation of the fitted values for HVTN 505

Participants who drop out or initiate ART prior to $Y_p$ measurement, or who have no VL measurement at the visit(s) used to calculated $Y_p$, have $Y_p$ missing. We use Little and An's (2004) penalized spline propensity prediction method to predict $Y_p$ for subjects with missing data, and to build the model for whether $Y_p$ is observed for infected subjects. This implementation breaks down into three steps.

**Step 1:** We factor $P(M_p = 0|S = 1)$ as the probability of having $Y_p$ measured conditional on not starting ART and not dropping out before $Y_p$ measurement; multiplied by one minus the probability of starting ART before $Y_p$ measurement conditional on not dropping out before $Y_p$ measurement; multiplied by one minus the probability of dropping out before $Y_p$ measurement. Three logistic regression models will be fit for each of the three components, for the dichotomous outcomes having $Y_p$ measured, starting ART before $Y_p$ measurement, and dropping out before $Y_p$ measurement, respectively. Each model will relate the dichotomous outcome to covariates within the set $(X_1, \cdots, X_q, Y_1, \cdots, Y_{p-1})'$. Multiplying together the fitted values from the three models yields estimated propensities $\widehat{Y}_1^*$ for all subjects with $S = 1$. We will use an automated all-subsets model selection procedure to determine each of the three logistic regression models, with BIC as the criterion for the best model.

For each model, the baseline and pre-infection covariates $(X_1, \cdots, X_q)$ to consider are the baseline variables age, behavioral risk score, race/ethnicity, HSV-2 serostatus, number of priming vaccinations (DNA or DNA placebo) received prior to infection diagnosis, and whether the boost vaccination (rAd5 or rAd5 placebo) was received prior to infection diagnosis. In addition, the post-infection covariates $(Y_1, \cdots, Y_{p-1})$ to consider are the calendar time of infection diagnosis (tertiles are quartiles), whether a subject has a pre-seroconversion HIV PCR positive sample, the pre-ART CD4+ T cell counts and viral loads before $Y_p$ measurement, the squares of these VL and CD4 variables, any VL > 100,000 and any CD4 < 500 before $Y_p$ measurement, and their 2-way interaction terms. The model-building will be done using an automated procedure, to ensure objectivity of the analysis. The CD4 variables will be analyzed on the square-root scale.

For the analysis of endpoints at HIV-1 infection diagnosis, only baseline and pre-infection covariates will be considered.

**Step 2:** Secondly, a spline regression model of $Y_p$ on $\widehat{Y}_1^*$ is fit using subjects for whom $Y_p$ is observed. We use 10 equally spaced fixed knots and a truncated linear basis, although some adjustment in the number of knots may be needed to obtain a good fit. In addition, covariates among $(X_2, \cdots, X_q, Y_1, \cdots, Y_{p-1})$ that predict $Y_p$ will be entered into the regression model using a linear additive parametric model. An automated all-subsets linear regression model selection procedure will be used to select the best model (with criterion BIC for optimality), considering the same covariates as considered for Step 1. Based on this fitted model, the value $E[Y_p|Y_1^*, X_2, \cdots, X_q, Y_1, \cdots, Y_{p-1}, S = 1, Z = z]$ is predicted for each subject with $SM_p = 1$ using his or her estimated propensity $\widehat{Y}_1^*$ and other covariates.

In more detail, we replace one of the potential predictor variables, say $X_1$, by $Y_1^*$, and suppose

$$(X_2, \cdots, X_q, Y_1, \cdots, Y_{p-1} | Y_1^*, S = 1, Z = z) \sim N((s_{z2}^X(Y_1^*), \cdots, s_{zq}^X(Y_1^*), s_{z1}^Y(Y_1^*), \cdots s_{zp-1}^Y(Y_1^*)), \Sigma_z), \quad (10)$$
$$(Y_p | Y_1^*, X_2, \cdots, X_q, Y_1, \cdots, Y_{p-1}, S = 1, Z = z, \gamma_z) \sim N(s_{zp}^Y(Y_1^*) + r_z(X_1^*, \cdots, X_q^*, Y_1^*, \cdots, Y_{p-1}^*, \gamma_z), \sigma_z^2),$$

for $z = 0, 1$, where for $j = 2, \cdots, q$ $s_{zj}^X(Y_1^*) = E(X_j | Y_1^*, S = 1, Z = z)$ is a spline for the regression of $X_j$ on $Y_1^*$ where $X_j^* = X_j - s_{zj}^X(Y_1^*)$; and for $j = 1, \cdots, p-1$, $s_{zj}^Y(Y_1^*) = E(Y_j | Y_1^*, S = 1, Z = z)$ is a spline for the regression of $Y_j$ on $Y_1^*$, where $Y_j^* = Y_j - s_{zj}^Y(Y_1^*)$. Furthermore for $z = 0, 1$ $r_z$ is a parametric function with unknown parameter vector $\gamma_z$ that satisfies $r_z(Y_1^*, 0, \cdots, 0, \gamma_z) = 0$ for all $\gamma_z$. For subject $i$ in group $z$ with $S_i M_{pi} = 1$, the predicted value of $Y_p$ is obtained as $\widehat{E}_i[Y_p] = \widehat{s}_{zp}^Y(\widehat{y}_{i1}^*) + r_z(\widehat{x}_{i1}^*, \cdots, \widehat{x}_{iq}^*, \widehat{y}_{i1}^*, \cdots, \widehat{y}_{i(p-1)}^*; \widehat{\gamma}_z)$, where $\widehat{x}_{ij}^* = x_{ij} - \widehat{s}_{zj}^X(\widehat{x}_{i1}^*)$, $\widehat{y}_{ij}^* = y_{ij} - \widehat{s}_{zj}^Y(\widehat{y}_{i1}^*)$, $\widehat{s}_{zj}^X$ denotes the sample estimate of the spline $s_{zj}^X$, $\widehat{s}_{zj}^Y$ denotes the sample estimate of the spline $s_{zj}^Y$, $x_{ij}$ is the realization of $X_{ij}$, and $y_{ij}$ is the realization of $Y_{ij}$. For $z = 0, 1$ we will use the linear function

$$r_z(X^*, \gamma_z) = \gamma_z^T X^*,$$

where $X^*$ denotes the set of covariates $(X_2, \cdots, X_q, Y_1, \cdots, Y_{p-1})$ that are selected for the best-fitting linear regression model for predicting $Y_p$.

**Step 3:** Thirdly, fitted values $\widehat{E}_i[g_z^{-1}(Y_p; \alpha_z, \beta_z)]$ and $\widehat{E}_i[Y_p g_z^{-1}(Y_p; \alpha_z, \beta_z)]$ are computed using these regression fits and numerical integration. Specifically, using (10), we take

$$\widehat{E}_i[g_z^{-1}(Y_p; \alpha_z, \beta_z)] = \int g_z^{-1}(y; \alpha_z, \beta_z) \frac{1}{\widehat{\sigma}_z} d\Phi([y - \widehat{E}_i[Y_p]]/\widehat{\sigma}_z),$$

where $\Phi$ is the cumulative cdf of the standard normal distribution. The predicted value $\widehat{E}_i[Y_p g_z^{-1}(Y_p; \alpha_z, \beta_z)]$ is computed similarly using numerical integration.

The three steps are done for the vaccine and placebo groups separately.

### 7.4.6 Computational algorithm for estimating the $ACE$

As detailed in Gilbert and Jin (2010), the estimate $\widehat{ACE} = \widehat{\mu}_1 - \widehat{\mu}_0$ is computed with the following steps.

**Step 1:** Estimate $p_0$ by solving $\sum_{i=1}^N U_{1i}^M(p_0) = 0$ and estimate $p_1$ by solving $\sum_{i=1}^N U_{2i}^M(p_1) = 0$.

**Step 2:** Plug $\widehat{p}_0$ and the fitted values $\widehat{E}_i[g_0^{-1}(Y_p; \alpha_0, \beta_0)]$ into (6) and solve for $\alpha_0$ in $\sum_{i=1}^N U_{3i}^M(\alpha_0) = 0$ with a one-dimensional line search. Similarly plug $\widehat{p}_1$ and the fitted values $\widehat{E}_i[g_1^{-1}(Y_p; \alpha_1, \beta_1)]$ into (7) and solve for $\alpha_1$.

**Step 3:** Plug the estimates of $p_0$ and $\alpha_0$ and the fitted values $\widehat{E}_i[Y_p g_0^{-1}(Y_p;\alpha_0,\beta_0)]$ into (8) and solve $\sum_{i=1}^N U_{5i}^M(\mu_0)=0$ for $\mu_0$. Similarly solve $\sum_{i=1}^N U_{6i}^M(\mu_1)=0$ for $\mu_1$.

Under the monotonicity assumption (i.e., $\phi=1$) the computational algorithm is the same except that the second part of Step 2 is omitted ($\alpha_1$ is no longer relevant) and $U_{6i}^M(\theta)$ is replaced with

$$U_{6i}^M(\theta)=Z_iS_i\left\{\mu_1-\left[(1-M_{pi})Y_{pi}+M_{pi}\widehat{E}_i[Y_p]\right]\right\}.$$

### 7.4.7 Standard errors and confidence intervals

We use the bootstrap to obtain standard error estimators for $\widehat{ACE}$ and confidence intervals for $ACE$. Within each treatment group $Z=z$ separately, $B$ bootstrap data sets are constructed by sampling with replacement $N_z$ realizations of $(X_i,S_i,M_i,Y_i)|Z_i=z$, $z=0,1$. The $ACE$ is estimated as described above for each bootstrap data set. Then standard errors for $\widehat{\mu}_0$, $\widehat{\mu}_1$, and $\widehat{ACE}$ are estimated by the sample standard deviations of the bootstrap estimates, and $(1-\alpha)\times 100\%$ confidence intervals are obtained as the $\alpha/2$ and $1-\alpha/2$ percentiles of the bootstrap estimates.

### 7.4.8 Handling monotonely missing post-infection covariates

We have described the above method for the case that all infected subjects have complete data on the variables $X_1,\cdots,X_q,Y_1,\cdots,Y_{p-1}$ used to predict $Y_p$ or to predict whether $Y_p$ is observed. If more than 20% of infected subjects have an important predictor missing, then we will implement a version of the method that allows a monotone pattern of missing data (i.e., that $Y_{j+1},\cdots,Y_p$ are missing for all subjects for whom $Y_j$ is missing), using the approach of Little and An (2004). Specifically, the propensity spline model (10) can be applied sequentially to each block of missing values. By replacing missing values of covariates by the predicted values in sequential regression models, estimates $\widehat{E}_i[g_z^{-1}(Y_p;\alpha_z,\beta_z)]$ and $\widehat{E}_i[Y_p g_z^{-1}(Y_p;\alpha_z,\beta_z)]$ in equations (6)–(9) can be computed, and the resulting estimating equations $\sum_{i=1}^N U_i^M(\theta)=0$ can be solved for $\theta$.

### 7.4.9 Secondary biological activity analyses of post-infection endpoints

*Selection bias sensitivity analysis.* A sensitivity analysis will be used to produce point and 95% confidence interval estimates of the $ACE$ under a range of assumptions about potential selection bias resulting from the analyzed subgroups, as described above.

*Analysis using only Week 28+ infections.* A secondary analysis will evaluate the vaccine effect on post-infection endpoints among only Week 28+ infections, by excluding those infections that occur before Day 196. The analysis will use the method of Little and An (2004)

to estimate the $ACE$, under A1-A3 and the additional assumption of no post-randomization selection bias.

## 7.5 Exploratory Endpoint 2 (CD4+ T Cell Counts, ART Initiation, HIV-1 Related Clinical Events)

### 7.5.1 CD4+ T cell count at Week 2 post-diagnosis

The LA method will be applied to assess the vaccine effect on the pre-ART CD4+ cell count at Week 2 PD. The LA method is used so as to attempt to provide unbiased inferences, accounting for the loss of evaluable endpoints due to ART initiation or dropout. If the analysis suggests a vaccine effect, then the sensitivity analysis methods described above in Section 7.4.3 will be applied to evaluate the robustness of the vaccine effect. Participants using prophylactic ARVs near the time of HIV-1 infection will be evaluated separately.

### 7.5.2 Time to ART Initiation

The vaccine effect on the time between infection diagnosis and ART initiation will be evaluated with a Cox proportional hazards model. Participants initiating ART before Week 10 PD will be censored at time of ART initiation. If the analyses suggest there may be vaccine effects, then additional sensitivity analyses may be performed accounting for the possibility of post-randomization selection bias, using the method of Shepherd, Gilbert, and Lumley (2007).

### 7.5.3 Time to HIV-1 Related Clinical Events

The same method as used for the analysis of the time to ART initiation endpoint will be used, where the failure event is defined as the first event of any documented AIDS-defining condition in the CDC 2008 definition (listed in Appendix A of Schneider et al., 2008) or death (from any cause). Other disease progression related events (such as cardiac, hepatic, renal, metabolic, and non-AIDS malignancies) will be catalogued and analyzed in a similar fashion, should the number of these events be sufficient for analysis.

## 7.6 Exploratory Endpoint 3 (HIV-Specific and rAd5-Specific Immune Responses)

Data from the IFN-$\gamma$ ELISpot assay will be summarized for each treatment group using geometric means and percentages of subjects with a positive response (using the standard HVTN method for calling a positive response). Analogous summaries will be provided for

data from the ICS assay. If assay data are qualitative (i.e., positive or negative) then analyses will be performed by tabulating the frequency of positive response for each assay by group at each time point that an assessment is performed. For the vaccine group, crude and net binomial response rates will be presented with their corresponding exact 95% confidence interval estimates. Net response rates are calculated by subtracting the placebo response rate (i.e., false positive rate) from the vaccine response rate. For the placebo group, the crude binomial response rate and exact 95% confidence interval estimates will be presented. These immunogenicity results will be used to demonstrate that the vaccine is immunogenic in the study population.

## 7.7 Exploratory Endpoint 4: (Host Genetics)

The HLA alleles of all subjects in the MITT Infected Population, and possibly for subjects in the Immunogenicity Subcohort, will be measured. Subjects will be classified into one of three HLA groups defined by "protective", "neutral" and "unfavorable" alleles. This definition will be developed based on the literature on the relationship between HLA alleles and HIV progression prior to the final analysis. It is expected that the definition will be very similar to that used for the Step trial.

*Analysis of post-infection endpoints.* For each post-infection endpoint, we will use the analysis method specified above to assess whether the HLA group predicts the endpoint, and whether HLA group modifies the vaccine effect on the endpoint (interaction test). For the univariable endpoints such as pre-ART viral load, the interaction test will be based on a 2 degree of freedom generalized Wald test. For the time-to-event post-infection endpoints such as ART initiation, the interaction test will be based on a 2 degree of freedom partial likelihood ratio test in the Cox model.

## 7.8 Exploratory Endpoint 5 (HIV-1 Genetics)

The sieve analysis will be conducted using updates of the methods that were used for the Step (Rolland et al., 2011) and RV144 HIV vaccine efficacy trials. The sieve analysis plan will be finalized before conducting the sieve analysis.

## 7.9 Exploratory Endpoint 7 (Prophylactic ARV Use)

The sampling plan for measuring ARV drug use in plasma specimens will be determined based on the self-reported prophylactic ARV use data. Specifically, if the proportion of participants reporting use at any point in time during follow-up is less than 5%, and absent other data suggesting meaningful levels of use, no plasma specimens will be tested. If, however, the self-report data suggests greater than 5% of participants have used ARVs prophylactically, the ARV sampling plan specified in version 4 of the protocol will be applied. This plan measures

ARV drug levels in the plasma of all MITT infected participants, participants self-reporting use, and a random sample of participants not reporting use.

Using these data, an analysis will be performed to determine whether there is evidence that the vaccine effect on HIV-1 acquisition depends on prophylactic ARV use. A time-dependent "ARV use" variable will be defined as an indicator of detectable plasma drug levels as a function of time since enrollment. Here and henceforth, self-reported ARV use data will be used in the absence of plasma drug level data. Note that ARV use is only measured in a subset of participants and at selected time points; the subset of participants with ARV use measured constitute a stratified case-cohort sample from the trial population. We will use the weighted likelihood method of Li, Gilbert and Nan (2008) to accommodate the case-cohort sampling. The method uses a weighted likelihood approach to fit a Cox proportional hazards regression model to grouped survival data with stratified case-cohort sampling. We will use the version of the method that uses estimated weights, representing the stratum-specific (inverse) sampling fractions, since this improves efficiency. The covariates in the proportional hazards model will be vaccination assignment (vaccine vs. placebo), the time-dependent ARV use variable, and an interaction between ARV use and vaccination assignment. Since we will be missing full ARV use history on most subjects in the case-cohort sample, we will fill in the missing data in the following fashion. The goal is to generate complete ARV use data from the time of enrollment to HIV-1 infection or censoring. If a participant is selected for ARV assessment at a given visit and tests negative at the previous visit, all earlier visits until the previous ARV assessment will be considered negative. Similarly, going forward in time, once a participant tests negative at two consecutive visits, ARV use at all subsequent visits until the next ARV assessment, HIV-1 infection, or censoring will be set to negative. We may also explore a multiple-imputation approach to filling in the missing ARV use data; the Li et al. (2008) method was developed to handle missing covariate data using multiple imputation. A Wald test will be used to evaluate whether the vaccine effect on HIV acquisition differs according to ARV use.

We will also study the extent to which the vaccine effect on VL endpoints depends on prophylactic ARV use. Note that, because the VL analysis is conducted among HIV-1 seroconverters only, there is not a missing data problem as described above for the HIV acquisition endpoint. Specifically, we will have ARV use measured at the time of HIV diagnosis for all MITT infected participants. To study this question, we will reproduce the primary analysis of the VL endpoint, using the Little and An method, and in addition to vaccination assignment we will include as predictors an indicator of ARV use at the time of diagnosis and an interaction between vaccination assignment and ARV use. A Wald test will be used to test for an interaction between vaccination assignment and ARV use. A secondary analysis may be conducted in which we consider ARV use at the visit with earliest evidence of infection, rather than at the diagnosis visit.

# 8 Monitoring of Trial

The study will be monitored, potentially leading to modification or termination of the study. Interim analyses will occur every 6 months, with the first analysis scheduled for approximately September, 2014. The results of the interim analyses will be shared in a report to the Oversight Group that will keep the results confidential.

Interim analysis reports will include point estimates, 95% confidence intervals, and 2-sided p-values testing $H_0$: VE = 0% will be reported for $VE^{MITT}(24)$ and for $VE^{MITT}(t)$ for the latest available time point $t$. These parameters will also be estimated and tested for the four subgroups defined in Section 7.3.2. Point estimates and 95% confidence intervals will also be reported for dropout rates over study time and for pre- and post-unblinding follow-up periods. Results will be shown overall and separately by treatment arm. Two-sided p-values testing for a difference in dropout rates between treatment groups will also be reported, separately for pre- and post-unblinding periods.

## References


Borgan, O., Langholz, B., Samuelson, S. O., Goldstein, L., Pagoda, J. (2000). Exposure stratified case-cohort designs. *Lifetime Data Analysis* **6**, 39–58.

Breslow, N.E. (1970). Modification of the UMP test for binomial probabilities. *Journal of the American Statistical Association* **65**, 639–48.

Breslow, N.E., Lumley, T., Ballantyne, C.M., Chambless, L.E., Kulich, M. (2009). Using the whole cohort in the analysis of case-cohort data. *American Journal of Epidemiology* **169**, 1398–1405.

Buchbinder, S.P., Mehrotra, D.V., Duerr, A., et al. (2008). The Step study: The first test-of-concept efficacy trial of a cell-mediated immunity HIV vaccine. *Lancet* **372**, 1881–1893.

Chang, M.N., Guess, H.A. and Heyse, J.F. (1994). Reduction in the burden of illness: a new efficacy measure for prevention trials. *Statistics in Medicine* **13**, 1807–1814.

Durham L, Halloran M, Longini IMA (1999). Comparing two smoothing methods for exploring waning vaccine effects. *Applied Statistics* **48**:395-407.

Flynn, N.M., Forthal, D.N., Harro, C.D., Judson, F.N., Mayer, K.H., Para, M.F., and The rgp120 HIV Vaccine Study Group (2005). Placebo-controlled phase 3 trial of recombinant glycoprotein 120 vaccine to prevent HIV-1 infection. *Journal of Infectious Diseases* **191** 654–665.

Follmann, DA, Fay, M and Proschan, MA (2009). Chop-lump tests for vaccine trials. *Biometrics*, **65**:885-893.

Gilbert, PB, Wei, LJ, Kosorok, MR, and Clemens JD (2002). Simultaneous inferences on the contrast of two hazard functions with censored observations. *Biometrics* **58**:773-80.

Gilbert, PB, Bosch R, and Hudgens MG (2003). Sensitivity analysis for assessing vaccine effects on viral load in HIV-1 vaccine trials. *Biometrics* **59**:531-541.

Gilbert PB, Ackers M, Berman P, Francis D, Popovic V, Hu D, Heyward W, Sinangil F, Shepherd B, Gurwith M (2005). HIV-1 virologic and immunologic progression and antiretroviral therapy initiation among HIV-1 infected participants in an efficacy trial of a recombinant glycoprotein 120 vaccine. *Journal of Infectious Diseases*, **192**:974-983.

Gilbert PB and Jin Y (2010). Semiparametric estimation of the average causal effect of treatment on an outcome measured after a post-randomization event, with missing outcome data. *Biostatistics* **11**:34-47.

Gilbert PB, Sun Y (2005). Failure time analysis of HIV vaccine effects on viral load and antiretroviral therapy initiation. *Biostatistics* **6**:374-94.

Gilbert PB, Shepherd BE, Hudgens MG (2013). Sensitivity analysis of perprotocol time-to-event treatment efficacy in randomized trials. *Journal of the American Statistical Association*, in press.

Grambsch PM, Therneau TM (1994). Proportional Hazards Tests and Diagnostics Based on Weighted Residuals. *Biometrika* **81**:515-26.

Hammer, S.M., Efron, J.J., Reiss, P., et al. (2008). Antiretroviral treatment of adult HIV infection: 2006 recommendations of the International AIDS Society-USA panel. *Journal of the American Medical Association* **300** 555–570.

Heyse, JF, Kuter, BJ, Dallas, MJ, Heaton, P for the REST Study Team (2008). Evaluating the safety of a rotavirus vaccine: the REST of the story. *Clinical Trials*, **5**: 131-139.

Jemiai, Y. and Rotnitzky, A. (2005). Semiparametric methods for inferring treatment effects on outcomes defined only if a post-randomization event occurs. *Unpublished Ph.D. dissertation*, Harvard School of Public Health, Department of Biostatistics.

Li, Z., Gilbert, P., and Nan, B. (2008). Weighted likelihood method for grouped survival data in case-cohort studies with application to HIV vaccine trials. *Biometrics*, **64**: 1247-1255.

Little, R. and An, H. (2004). Robust likelihood-based analysis of multivariate data with missing values. *Statistica Sinica*, **14**: 949–968.

Little, R.J.A. and Rubin, D.B. (2002). *Statistical Analysis with Missing Data*. New York: Wiley.

Lu X, Tsiatis AA (2008). Improving the efficiency of the log-rank test using auxiliary covariates. *Biometrika*, **95**: 679–694.

Mehtrotra, DV and Heyse, J (2004). Use of the false discovery rate for evaluating clinical safety data. *Statistical Methods in Medical Research*, **13**: 227–238.

Mogg, R. and Mehrotra, D.V. (2007). Analysis of antiretroviral immunotherapy trials

with potentially non-normal and incomplete longitudinal data. *Statistics in Medicine* **26** 484–497.

Park Y, Tian L, Wei LJ (2006). One- and two-sample nonparametric inference procedures in the presence of a mixture of independent and dependent censoring. *Biostatistics*, **7**:252–267.

Patterson K, Prince H, Kraft E, et al. (2010). Exposure of extracellular and intracellular tenofovir and emtricitabine in mucosal tissues after a single dose of fixed-dose TDF/FTC: Implications for pre-exposure HIV prophylaxis. XIII International AIDS Conference, Vienna, Austria, July 18-23, 2010.

Pitisuttithum, P., Gilbert, P.B., Gurwith, M., et al. (2006). Randomized, double-blind, placebo-controlled efficacy trial of a bivalent recombinant glycoprotein 120 HIV-1 vaccine among injection drug users in Bangkok, Thailand. *Journal of Infectious Diseases* **194** 1661–1671.

Stitelman O, van der Laan M (2010). Collaborative targeted maximum likelihood for time to event data. *University of California, Berkeley, U.C.Berkeley Division of Biostatistics Working Papers*: paper 260.

Rolland MR, Tovanabutra S, deCamp AC et al. (2011). Genetic impact of vaccination on breakthrough HIV-1 sequences from the STEP trial. *Nature Medicine* **17** 366–371.

Robins JM, Blevins D, Ritter G, Wulfsohn M (1992). G-estimation of the effect of prophylaxis therapy for pneuocystis carinii pneumonia on the survival of AIDS patients. *Epidemiology*, **3**:319–336.

Robins, J.M. (1992). Estimation of the time-dependent accelerated failure time model in the presence of confounding factors. *Biometrika* **79** 321–334.

Robins JM, Rotnitzky A, Zhao LP (1994). Estimation of regression-coefficients when some regressors are not always observed. *Journal of the American Statistical Association*, **89**: 846–866.

Robins, J.M., Rotnitzky, A. and Zhao, L.P. (1995). Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *Journal of the American Statistical Association* **90** 106–121.

Robins JM, Hernán MA, Brumback B (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology*, **11**:550–560.

Rubin, D.B. (1978). Bayesian inference for causal effects. *Annals of Statistics* **6** 34–58.

Schacker TW, Hughes JP, Shea T, Coombs RW, Corey L (1998). Biological and virologic characteristics of primary HIV infection. *Annals of Internal Medicine*, **128**:613–20.

Schneider E, Whitmore S, Glynn MK, Domiguez K, Mitsch A, McKenna MT (2008). Revised Surveillance Case Definitions for HIV Infection Among Adults, Adolescents, and Children Aged <18 Months and for HIV Infection and AIDS Among Children Aged 18 Months to <13 Years– United States, 2008. *MMWR Recommendations and Reports* **57 (RR10)** 1–8.

Shepherd, B., Gilbert, P.B. and Lumley, T. (2007). Sensitivity analyses comparing time-to-event outcomes only existing in a subset selected post-randomization, conditional on covariates, with application to HIV vaccine trials. *Journal of the American Statistical Association* **102** 573–582.