NCT03060629

HVTN 705/VAC89220HPX2008 Statistical Analysis Plan for Safety, Trial Monitoring, and Vaccine Efficacy Analysis

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

and

Janssen Vaccines & Prevention B. V.

Protocol version 4.0 (June 12, 2020)
SAP version 3.4
June 30, 2022

HIV VACCINE
TRIALS NETWORK

SAP Approval Signature Page

HVTN 705/VAC89220HPX2008
A multicenter, randomized, double-blind, placebo-controlled phase 2b efficacy study of a heterologous prime/boost vaccine regimen of Ad26.Mos4.HIV and aluminum phosphate-adjuvanted Clade C gp140 in preventing HIV-1 infection in women in sub-Saharan Africa

I have read this Statistical Analysis Plan and approve its contents.

See appended email approval.

PPD

Date

PPD

HIV Vaccine Trials Network
Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center

See appended email approval.

PPD

Date

PPD

HIV Vaccine Trials Network
Department of Statistics, University of Washington
Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center

PPD

See appended email approval.

Date

PPD

HIV Vaccine Trials Network
Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center
Department of Biostatistics, University of Washington

See appended email approval.

PPD

Date

PPD

Janssen Vaccines & Prevention B.V.
On behalf of PPD, Janssen Vaccines Prevention B.V.,
as PPD designated backup PPD

SAP Modification History

The version history of, and modifications to, this statistical analysis plan are described below.

Date: October 12, 2017
Protocol version: 1.0 (February 17, 2017)
SAP version: 1.0

Date: June 4, 2019
Protocol version: 3.0 (July 26, 2018)
SAP version: 1.1

Modifications:

1) Section 10.1.2: Updated timing of non-efficacy interim analyses to line up with DSMB meetings: "Thereafter, non-efficacy interim analyses will proceed at every scheduled DSMB meeting (anticipated every six months) until the end of Stage 1."

2) Section 7.1.2: Corrected days of reporting from 28 to 30: "AEs occuring during the first 30 days..."

3) Section 10.3: Corrected timing of the first operational futility analysis report from November to October 2018: "A report provided to the DSMB will be included in 6-monthly closed DSMB reports, starting in October 2018..."

4) Section 10.3: Clarified content and timing of treatment-blinded operational futility analysis reports to the Oversight Group: "Furthermore, a treatment-blinded report will be generated for distribution to the OG before each DSMB meeting takes place and will report estimates listed in (a)–(c) and (e)–(g) above calculated based on treatment-blinded data in scenarios (i)–(iii). The reported results pertaining to estimates (a)–(c) and (e)–(g) under scenarios (i)–(iii) will be identical to those in the DSMB report."

5) Section 8.1: Added the additive-difference vaccine efficacy against HIV-1 infection over time as a parameter of interest in support of the primary analysis of VE given its attributable risk interpretation and public health impact: "In addition, to assess potential time-effects of vaccine efficacy, the Kaplan-Meier method will be used to plot the estimated cumulative incidence rates over time for the vaccine and placebo groups. This method will be used to estimate (i) cumulative vaccine efficacy over time, defined as (1 minus the ratio (vaccine/placebo) of cumulative incidence by time t) $\times$ 100%, and (ii) additive-difference vaccine efficacy over time, defined as the difference (placebo minus vaccine) in cumulative incidence by time t, with the method of Parzen, Wei, and Ying (1997) applied to obtain pointwise and simultaneous 95% CIs. In addition, the longitudinal targeted minimum loss-based estimation method as implemented in the R

package `survtmle` will be used for estimation, which in addition to allowing confounding adjustment can correct for potential bias due to covariate-dependent censoring."

6) Section 10.3.1: Deleted the PP incidence rate scenario: "pooled infection rate [PP $VE(7–24) = 50\%$ scenario]: $0.5 \times 0.042 + 0.5 \times 0.5 \times 0.042 = 0.0315$ infections/person-year at-risk" because the futility analysis is first conducted in the MITT cohort, and then the PP cohort is only extracted as a subset of the MITT cohort after a completed simulation of MITT cohort time-to-event data, which is clarified by the statement "Conditioning on interim data, a complete time-to-event data set is simulated for the MITT cohort, and the PP cohort is extracted at the end."

Date: February 12, 2020
Protocol version: 3.0 (July 26, 2018)
SAP version: 1.2

Modifications:

1) Updated the non-efficacy monitoring boundary presented to and approved by the DSMB on November 26, 2019. This update entailed the following revisions:

   – In Section 10: "Vaccine efficacy is monitored by the independent DSMB at each DSMB meeting, with monitoring triggers based on numbers of primary HIV-1 infection endpoints (and, for non- and high efficacy, the length of completed follow-up by all on-study participants) for when the DSMB meetings begin to include formal evaluation of the potential-harm, non-efficacy, and high-efficacy stopping boundaries (these triggers are described below)."

   – Table 5, last column: "6-monthly starting once all participants reach 13 months, and 60 MITT infections are observed, then through the end of Stage 1"

   – In Section 10.1.1: "...(with the exception that the first non-efficacy analysis would be done and reported to the DSMB by secure means at the time when all participants in follow-up reach 13 months of follow-up, and 60 total MITT primary endpoint events are observed)."

   – In Section 10.1.1: "The tests start at the 10^{th} total infection and are performed continuously until the non-efficacy monitoring commences."

   – in Section 10.1.2: "Such analyses will start when two conditions are met: (1) all participants in follow-up have reached 13 months of follow-up since enrollment, and (2) at least 60 MITT infections have been observed."

   – In Section 10.1.2: "By checking confidence intervals for both VE(0–24) and VE(7–24), and requiring completed 13 months of follow-up, the monitoring plan is designed to protect against stopping prematurely based on ramping vaccine efficacy over the intercurrent period of 0–13 months."

2) Section 10.1.2: Revised the definition of the non-efficacy interim analysis timepoint: "In this event, cumulative vaccine efficacy through time t, $VE(0\text{–}t)$ and $VE(7\text{–}t)$, will be estimated with fixed time point t chosen to be the latest possible time point where stable estimation of both VE parameters can be achieved; this is operationalized by defining t as the maximum time point when at least 150 participants in the per-protocol cohort are observed to be at risk for the primary efficacy endpoint in each treatment arm."

Date: January 12, 2021
Protocol version: 4.0 (June 12, 2020)
SAP version: 2.0

Modifications:

- Added Section 9 on the scope and timing of the primary analysis to be conducted at the end of Stage 1
- In Section 2.2, secondary objectives and endpoints 3 and 4, revised "month 12 through..." as "month 13 through...", i.e., 4 weeks after the 4^{th} vaccination visit
- In Section 8.2.3, revised the statement

  "In particular, we perform these analyses among: per-protocol subjects who are HIV-1 uninfected at Month 12, studying endpoints occurring between Month 12 and 24; per-protocol subjects who are HIV-1 uninfected at Month 12, studying endpoints occurring between Month 12 and 36."

  as

  "In particular, we perform these analyses in the FIS cohort, studying endpoints occurring between Month 13 and 24; and in the FIS cohort, studying endpoints occurring between Month 13 and 36."
- In Section 8.2.3, revised the statement

  "In particular, we perform these analyses among: MITT subjects who are HIV-1 uninfected at Month 12, studying endpoints occurring between Months 12 and 24; MITT subjects who are HIV-1 uninfected at Month 12, studying endpoints occurring between Months 12 and 36"

  as

  "In particular, we perform these analyses among: MITT subjects who are HIV-1 uninfected at Month 13, studying endpoints occurring between Month 13 and 24; MITT subjects who are HIV-1 uninfected at Month 13, studying endpoints occurring between Month 13 and 36..."

Date: May 17, 2021
Protocol version: 4.0 (June 12, 2020)
SAP version: 3.0

Modifications:

- In Section 8.1, added more details on the operational definition of the primary analysis time point τ and the construction of the CI for VE. To this end, the following text was added: "More specifically, the primary VE parameter, $VE_{7-24}(\tau)$, will be estimated at a fixed time point τ chosen to be the latest possible time point when stable estimation using follow-up data through the Month 24 visit can be achieved; this is operationalized by defining τ as the maximum time point with 150 participants in the per-protocol cohort observed to be at risk for the primary endpoint in each of the placebo and vaccine groups. All times-to-event will be right-censored at the month 24 visit. Each of the two cumulative incidence parameters in $VE_{7-24}(\tau)$ will be estimated using the transformed Nelson-Aalen estimator for the cumulative hazard function evaluated at time τ defined above. We will use the delta method to obtain the asymptotic 95% CI for the log cumulative incidence ratio (vaccine/control) and then back-transform these confidence bounds to the VE scale. Both the point estimate and the 95% CI for VE will be reported."
- In Section 8.2.1, the following statement was added: "The VE parameter VE(0–24) will be estimated at the same time point τ as defined in Section 8.1."
- In Sections 8.1 and 8.2.1, the notation in the primary and secondary hypothesis tests was updated to emphasize that the hypothesis tests are performed at time point τ.
- In Section 9, items 1(c) and 1(d) were updated to emphasize that estimation of the primary VE parameter and the primary hypothesis test are performed at time point τ.

Date: August 13, 2021
Protocol version: 4.0 (June 12, 2020)
SAP version: 3.1

Modifications:

- In Section 8.5, Table 3 was added, specifying a set of baseline covariates used as input features by the superlearner for estimation of the behavioral risk score. Also, Table 4 summarizing the library of prediction and screening algorithms used by the superlearner was revised. Finally, the cross-validation procedure was revised as follows:

  "If the number of primary endpoints in each treatment group is > 30, we will use superlearner with 5-fold cross-validation, separately for the vaccine and placebo groups; otherwise leave-one-out cross-validation will be used. For the cross-validated superlearner, 5-fold cross-validation will be used for the outer cross-validation irrespective

of the number of primary endpoints. These cross-validation rules align with those made for the baseline behavioral risk score analysis in Moderna's COVE trial of the mRNA-1273 vaccine."

Date: April 5, 2022
Protocol version: 4.0 (June 12, 2020)
SAP version: 3.2

Modifications:

- In Section 8.2.2, the method used to estimate the covariate-adjusted vaccine efficacy was modified. The CFSurvival R package will now be used. In contrast to the method in the previous version of this SAP, CFSurvival does not require the discretization of time. As such, the Super Learner library in Table 2 has been modified to include survival regression methods.

Date: May 16, 2022
Protocol version: 4.0 (June 12, 2020)
SAP version: 3.3

Modifications: After the final analysis was implemented on May 2, 2022, the protocol statisticians became aware of the fact that Section 8.1 on the Primary Analysis of Vaccine Efficacy did not specify a statistical analysis of the primary parameter of per-protocol vaccine efficacy through 24 months that could be adequately justified, and therefore an update to the SAP is needed. This ascertainment is based purely on methodological grounds and is not influenced by the results.

The relevant text in the SAP v3.2 that defined the incorrect approach is as follows: "More specifically, the primary VE parameter, $VE_{7-24}(\tau)$, will be estimated at a fixed time point τ chosen to be the latest possible time point when stable estimation using follow-up data through the Month 24 visit can be achieved; this is operationalized by defining τ as the maximum time point with 150 participants in the per-protocol cohort observed to be at risk for the primary endpoint in each of the placebo and vaccine groups. All times-to-event will be right-censored at the month 24 visit."

The issue is that this specification inadvertently discarded more than 80% of the study participants in the risk-set at the month 24 visit, even though these participants were in fact still at risk at the month 24 visit (defined as HIV negative at all visits with HIV testing results including an HIV negative test result at the month 24 visit). More specifically, for the final analysis data set, 969 per-protocol vaccine recipients and 1008 per-protocol placebo recipients were at-risk at 24 months, yet only 166 per-protocol vaccine recipients and 159 per-protocol placebo recipients were included in the risk-set for the estimation of the cumulative incidence probabilities through the final time point of $\tau = 24.13$ months. An appropriate survival analysis method should include all of the information in the data, which would include the data from the 969 and 1008 per-protocol participants noted above. The

combination of defining the final time point τ to be after the month 24 visit date for most participants ($\tau = 24.13$ months), and the right-censoring of follow-up times of participants testing HIV negative at month 24 to their month 24 visit dates, caused this problem. The origin of the error was that in group sequential monitoring interim analyses, conducted when only partial follow-up data were available, it was appropriate and useful to define the final time point τ of the cumulative incidence and vaccine efficacy parameters based on the latest time point with at least 150 participants at-risk at the latest time point; this ensured stable inference as noted in the v3.2 SAP. However, this issue no longer exists for the final analysis, because for the final analysis all enrolled participants have follow-up well beyond the month 24 visit, such that the condition serves no purpose, and needlessly and inadvertently removes statistical information. Therefore, the carrying over of the statistical method designed for the interim analyses to the final analysis caused the problem.

To remedy this problem, Section 8.1 was revised as follows:

- The v3.2 sentence "The failure times of participants without HIV-1 infection by the month 24 visit are right-censored at the date of the last HIV-negative test or at the month 24 visit, whichever occurs earlier."

  was changed to

  "The failure times of participants with last HIV-negative test at or after the month 24 visit are right-censored at the right edge of the month 24 allowable visit window [defined as 60 days after the target day (728 days) for the month 24 visit, equal to 25.91 months post-enrollment]. For participants with last HIV-negative test prior to the month 24 visit and without diagnosis of the HIV-1 infection primary endpoint at or before the month 24 visit, their failure times are right-censored at the date of the last HIV-negative test."

- The following sentence in v3.2 was deleted, given that this issue is now resolved based on the above sentence: "Dropouts will be censored at the time of their last HIV-1 negative test before or at the month 24 visit."

- The two v3.2 sentences "More specifically, the primary VE parameter, $VE_{7-24}(\tau)$, will be estimated at a fixed time point τ chosen to be the latest possible time point when stable estimation using follow-up data through the Month 24 visit can be achieved; this is operationalized by defining τ as the maximum time point with 150 participants in the per-protocol cohort observed to be at risk for the primary endpoint in each of the placebo and vaccine groups. All times-to-event will be right-censored at the month 24 visit."

  were changed to the single sentence

  "More specifically, the primary VE parameter, $VE_{7-24}(\tau)$, will be estimated at the fixed time point $\tau = 25.91$ months post-enrollment with the specification of the value of τ determined as noted above."

- In Section 8.2.1, the sentence "As in the per-protocol analysis, the failure times of MITT participants without HIV-1 infection are right-censored at the date of the last HIV-negative test or at the month 24 visit, whichever occurs earlier."

  was changed to

  "The failure times of MITT participants for the analysis of VE(0–24) are right-censored following the same approach as used for the primary analysis of vaccine efficacy."

- Similar updates were needed to clarify the data analysis of vaccine efficacy through to Month 36. In particular, the following text was added to Section 8.2.3:

  "For all analyses evaluating VE through Month 24, the failure times will be right-censored following the same approach as used for the primary analysis of vaccine efficacy. For analyses that aim to evaluate VE through Month 36, the right-censoring approach used will depend on the number of participants who attend their month 36 visit at the end of follow-up. The failure time convention right-censoring approach used and definition of the estimand of interest will depend on how large this number is in each of the two arms. In particular:

Case 1: If at least 150 participants in each arm of the per-protocol cohort have attended their Month 36 visit and had a negative HIV-1 test at that visit, then VE will be evaluated through time $\tau = 36.5$ months, where 36.5 denotes the upper allowable visit window for the month 36 visit. The failure times of participants with last HIV-1 negative test at or after the month 36 visit are right-censored at the right edge of the month 36 allowable visit window (month 36.5). For participants whose last HIV-1 negative test occurs prior to the month 36 visit and without diagnosis of a primary HIV-1 infection endpoint at or before the month 36 visit, their failure times are right-censored at the date of their last HIV-1 negative test.

Case 2: If, in at least one of the two arms of the per-protocol cohort, there are not 150 participants that have attended their Month 36 visit and had a negative HIV-1 test at that visit, then, in order to ensure stable estimation, VE will be evaluated through a time τ that falls before 36.5 months. The time τ will be defined as the maximum time point at or after which 150 participants in both the vaccine and placebo groups of the per-protocol cohort have HIV-1 negative tests. For participants whose last HIV-1 negative test occurs prior to time τ and without diagnosis of a primary HIV-1 infection endpoint at or before time τ, their failure times are right-censored at the date of their last HIV-1 negative test. The failure times of all other participants will be right censored at time τ."

- Similarly, the following sentence was added to Section 9:

  "The analyses of vaccine efficacy through 30 months will define the final time point for analysis (τ) and the right-censoring process in the same way as described for the analyses of vaccine efficacy through 24 months, except now indexed off of the month 30 visit instead of the month 24 visit."

Date: July 18, 2022
Protocol version: 4.0 (June 12, 2020)
SAP version: 3.4

Modifications: Clarified text in Section 10.4: "Monitoring for Performance Standards of Quality of Trial Conduct." Specifically:

- Changed 'whereas others are specific to the AMP trials' to 'whereas others are specific to the HVTN 705 trial.'
- Changed 'adherence to receipt of infusions (target 90% of infusions received, with minimally acceptable level of 70% of infusions received).'

  to

  'adherence to study interventions (target 95% adherence for receipt of first three doses, with minimally acceptable level of 80%; and target 90% adherence for receipt of first four doses, with minimally acceptable level of 80%)'

Contents

List of Figures

List of Tables

1 Overview of HVTN 705/VAC89220HPX2008

HVTN 705/VAC89220HPX2008 is a multicenter, randomized, double-blind, placebo-controlled phase 2b efficacy study and plans to enroll 2600 HIV-uninfected women between the age of 18 and 35. The objective of the study is to evaluate the preventive vaccine efficacy (VE) of a heterologous prime/boost regimen utilizing Ad26.Mos4.HIV and aluminum-phosphate adjuvanted Clade C gp140 for the prevention of HIV infection (from confirmed HIV-1 infections diagnosed between the Month 7 and Month 24 visits).

2 Objectives and Endpoints

HVTN 705/VAC89220HPX2008 evaluates the preventive VE of a heterologous prime/boost regimen utilizing Ad26.Mos4.HIV and aluminum-phosphate adjuvanted Clade C gp140 for the prevention of HIV infection in HIV-seronegative women residing in sub-Saharan Africa (from confirmed HIV-1 infections diagnosed between the Month 7 and Month 24 visits).

2.1 Primary Objectives and Endpoints

Primary objective 1:
To evaluate the preventive vaccine efficacy (VE) of a heterologous prime/boost regimen utilizing Ad26.Mos4.HIV and aluminum-phosphate adjuvanted Clade C gp140 for the prevention of HIV infection in HIV-seronegative women residing in sub-Saharan Africa from confirmed HIV-1 infections diagnosed between the Month 7 and Month 24 visits

Primary endpoint 1:
Vaccine efficacy as derived from confirmed HIV-1 infections diagnosed between the Month 7 and Month 24 visits

Primary objective 2:
To evaluate the safety and tolerability of a heterologous prime/boost regimen utilizing Ad26.Mos4.HIV and aluminum-phosphate adjuvanted Clade C gp140 for the prevention of HIV infection in HIV-seronegative women residing in sub-Saharan Africa

Primary endpoint 2:
Local and systemic reactogenicity signs and symptoms for 3 days after each vaccination, adverse events for 30 days after each vaccination, and serious adverse events, AESIs, and adverse events leading to early participant withdrawal or early discontinuation of study product(s) administration for the entire duration of the study

2.2 Secondary Objectives and Endpoints

Secondary objective 1:
To evaluate vaccine efficacy from enrollment through 24 months

Secondary endpoint 1:
HIV-1 infection diagnosed after enrollment through 24 months post enrollment

Secondary objective 2:
To evaluate vaccine efficacy from enrollment through 36 months if Stage 2 occurs

Secondary endpoint 2:
HIV-1 infection diagnosed after enrollment through 36 months post enrollment

Secondary objective 3:
To evaluate vaccine efficacy from month 13 through month 24

Secondary endpoint 3:
HIV-1 infection diagnosed after month 13 through 24 months post enrollment

Secondary objective 4:
To evaluate vaccine efficacy from month 13 through month 36 if Stage 2 occurs

Secondary endpoint 4:
HIV-1 infection diagnosed after month 13 through 36 months post enrollment

Secondary objective 5:
To evaluate the immunogenicity of the vaccine regimen

Secondary endpoint 5:
Immune responses at the study visits following the third and fourth vaccinations from assays based on the HVTN Laboratory Assay Algorithm such as vaccine-specific binding antibodies and T-cell responses.

Secondary objective 6:
To evaluate immunogenicity and immune response biomarkers among vaccine recipients after the third vaccination as correlates of risk of subsequent HIV acquisition and correlates of vaccine efficacy, if deemed applicable.

Secondary endpoint 6:
Immune responses from assays based on the HVTN Laboratory Assay Algorithm (available at https://atlas.scharp.org/) and/or more assays down-selected from a larger pool of pilot studies, in HIV-1infected vaccine cases and HIV-1uninfected vaccine controls

Secondary objective 7:
To evaluate VE adjusting for various demographic and other baseline characteristics

Secondary endpoint 7:
HIV-1 infection diagnosed after the third vaccination by demographic and other baseline characteristics

Secondary objective 8:

If significant positive evidence of vaccine efficacy from month 7 through 24 months is seen, to assess if and how vaccine efficacy depends on genotypic characteristics of HIV such as signature mutations

Secondary endpoint 8:
HIV-1 infection diagnosed after month 7 through Month 24 and genotypic characteristics of viral sequences from HIV-1infected participants at HIV-1 diagnosis, such as signature site mutations

Secondary objective 9:
To evaluate and compare genomic sequences of viral isolates from HIV-1infected vaccine and placebo recipients, and use sieve analysis methods to assess whether VE differs by genotypic or phenotypic characteristics of exposing HIVs and whether there is evidence of vaccine-induced immune pressure on the viral sequences

Secondary endpoint 9:
Viral sequences from HIV-1infected participants at the earliest available postinfection time-point and possible subsequent visits

2.3 Exploratory Objectives

Exploratory objective 1:
To evaluate vaccine effects (vaccine activity) on virologic and immunologic outcomes (eg, HIV-1 viral load (VL) and postdiagnosis CD4+ T-cell count) among HIV-1infected participants for 6 months post diagnosis accounting for ARV use

Exploratory objective 2:
To explore the association between the vaginal microbiome as well as genital inflammation, and HIV infection risk

Exploratory objective 3:
To evaluate early and innate immune responses (eg, whole blood transcriptomics, serum cytokines) one day after the third vaccination (ie, the first protein boost) as correlates of risk of subsequent HIV acquisition

Exploratory objective 4:
To evaluate local and systemic reactogenicity signs and symptoms that arise from day 4 to day 7 at a subset of clinical research sites

Exploratory objective 5:
To further evaluate the immunogenicity of the vaccine regimen, additional immunogenicity assays may be performed, and assays may be performed on samples from other timepoints, based on the HVTN Laboratory Assay Algorithm

Exploratory objective 6:
To assess use of biomedical interventions and biological and behavioral factors in the study

cohort and how they modify vaccine efficacy

Exploratory objective 7:
To evaluate the role of host genetic factors in the immune response to the vaccine regimen and in vaccine effects on study endpoints

Exploratory objective 8:
To perform comparative analyses of correlates of risk identified in HVTN 705 and those identified in other HIV vaccine efficacy studies

Exploratory objective 9:
To conduct analyses related to furthering the understanding of HIV, immunology, vaccines, and clinical trial conduct

3 Follow-Up Period

All participants will be followed for at least 24 months post-enrollment. Participants will receive vaccinations at Months 0, 3, 6, and 12 and be followed for HIV infection for a period of at least 2 years (stage 1) after enrollment until the primary analysis at the end of stage 1 is performed (when the last subject reaches the month 24 visit). Participants who become HIV-1 infected during the study will be followed for approximately 6 months after confirmation of diagnosis.

4 Study Populations

The following study populations or analysis sets are used for addressing the study objectives.

- **Full Analysis Set (FAS)**:all randomized participants who receive at least 1 vaccine administration
- **Modified Intent-to-Treat (MITT) Population**: participants in the FAS who are HIV-1 uninfected on the date of first vaccination.
- **Per-Protocol (PP) Population**: participants in the FAS who are HIV-1 uninfected 4 weeks after the 3rd vaccination visit, who received all planned vaccinations at the first 3 vaccination visits within the respective visit windows and have no other major protocol deviations that were judged to possibly impact the efficacy of the vaccine.
- **Full Immunization Set (FIS)**: participants in the FAS who are HIV-1 uninfected 4 weeks after the 4th vaccination visit and who receive all planned vaccinations within the respective visit windows.

- **At risk Immunogenicity Cohort (IC-at risk)**: participants in the FAS who are selected for measurement of immune response endpoints at the primary immunogenicity timepoints and who are HIV-1 uninfected 4 weeks after the 3rd vaccination visit, who have no other major protocol deviations that were judged to possibly impact the efficacy of the vaccine.
- **Per Protocol Immunogenicity Cohort (IC-PP)**: participants in the IC-at risk who received all planned vaccinations at the first 3 vaccination visits within the respective visit windows.

The MITT population and the FAS are very similar but not identical to a full Intention-to-Treat Cohort (ie, all randomized participants); the FAS differs by excluding randomized volunteers who do not enroll (ie, dont receive any vaccinations); and the MITT population is the subset of the FAS that also excludes randomized participants discovered later to be HIV-1 positive by day 0. Because of blinding and the brief length of time between randomization and enrollment (typically no more than 4 working days) we expect almost all randomized volunteers to be in the FAS. Given that eligibility for the study requires recent evidence of being HIV-1 uninfected (within 45 days prior to enrollment), we expect almost all enrolled participants to also be in the MITT Cohort.

The analyses of safety will be performed on the FAS. The primary analysis of vaccine efficacy will be based on the PP population. Secondary analyses of vaccine efficacy will be based on the MITT population and the FIS. Analyses of vaccine immunogenicity and immune correlates of risk will be based on IC-at risk, IC-PP and the FIS (for those with immunogenicity outcomes).

In addition, 4 cohorts of participants who are diagnosed with HIV-1 infection during the trial are analyzed for addressing various study objectives. Terminology for these cohorts is defined in Table 1, which will be used throughout the SAP.

Since this is a proof-of-concept trial, all efficacy analyses will be done according to the as treated principle (ie, actually received treatments), except for analyses using the MITT population.

In the unexpected event of a duplicate enrollment, the interim safety data will be reported for each enrollment, considering these as separate participants, while noting in the report that a duplicate enrollment occurred. All final analyses will only include unique participants. For duplicate enrollments, the data collected under each enrollment will be combined and the participant will be identified using the participant ID of the first enrollment. For 'as treated' analyses, the treatments received across both enrollments will be considered when determining the treatment group. MITT analyses will use the treatment group assigned at randomization from the first enrollment. These doubly-enrolled participants are excluded from the PP cohort.

Table 1: Cohorts of HIV-1 infected study participants.

Cohort Name	Definition of Cohort
MITT infected by 24 Months cohort	Participants in the MITT population who are diagnosed with HIV-1 infection during the follow-up period after enrollment through the Month 24 visit
MITT infected by 36 Months cohort	Participants in the MITT population who are diagnosed with HIV-1 infection during the follow-up period after enrollment through the Month 36 visit
Per-Protocol infected by 24 Months cohort	Participants in the PP population who are diagnosed with HIV-1 infection during the follow-up period on or after the Month 7 visit through the Month 24 visit
Per-Protocol infected by 36 Months cohort	Participants in the PP population who are diagnosed with HIV-1 infection during the follow-up period on or after the Month 7 visit through the Month 36 visit

5 Definition of the Primary Efficacy Endpoint (Documented HIV-1 Infection)

5.1 HIV Testing Postvaccination

The vaccine efficacy endpoint is diagnosis of HIV-1 infection during the follow-up period. Following enrollment, HIV testing will take place at scheduled clinic visits defined in Appendix E and Appendix G of the CTP.

In-study HIV testing will be performed according to the HVTN HIV diagnostic testing algorithms. Routinely, specimens are initially assayed with an HVTN Lab Program approved HIV 1/2 enzyme immunoassay (EIA) or chemiluminescent microparticle immunoassay (CMIA). If the EIA/CMIA is reactive, nucleic acid polymerase chain reaction (PCR) test to detect HIV-1 RNA will be performed as indicated in the algorithm. The algorithm is repeated on a second specimen to confirm a diagnosis of HIV-1 infection. The second specimen for confirmatory testing may be collected at an interim visit (designated as visit #.X, where # is the visit at which the first reactive test was obtained and X designates the interim visit; specified in Appendix F and Appendix H of the CTP). Samples to be stored for future immunogenicity or virology studies will also be collected at this time (Appendix F of the CTP).

A case will be defined as a participant with confirmed detectable HIV-1 nucleic acid PCR on

2 different specimen collection dates. The nucleic acid test will most commonly be the HIV-1 RNA PCR viral load test. Confirmation of HIV-1 infection will be determined as dictated by the HVTN HIV testing algorithm (available on the HVTN 705 protocol-specific website). Before issuing an HIV-1 infection report for a participant diagnosed with HIV-1 infection prior to study unblinding, all testing results will be reviewed by a blinded, independent Endpoint Adjudicator(s) and/or designee(s).

If a participant is confirmed to have become HIV-1 infected prior to study unblinding, plasma HIV-1 viral RNA will be measured on archived samples collected according to Appendix E. In addition to plasma HIV-1 viral RNA testing, participants may also have measurements of immunogenicity assessments and a clinical assessment performed at all postinfection visits (see Appendix F of the CTP).

5.2 Endpoint Adjudication

The diagnostic criteria for HIV-1 infection outside the setting of a vaccine trial are well accepted. However, definitive diagnosis of HIV-1 infection in the context of having received an HIV vaccine that is even partially effective may be more difficult. Specifically, if the immune responses elicited by vaccination are capable of completely suppressing viral replication, or if vaccination alters the normal serological response upon exposure to HIV-1, standard diagnostic tests may be more difficult to assess. Therefore, the HVTN will have an endpoint adjudication process to assess all serological and virological testing, in a blinded manner, on each participant in the trial who, prior to study unblinding, tests positive per the HVTN 705 HIV-1 diagnostic testing algorithm. The assessment of the Endpoint Adjudicator(s) or designee(s) will be reported to the SDMC and to the HIV diagnostics laboratory.

The Endpoint Adjudicator(s) and/or designee(s) must notify the SDMC within 1 working day of any confirmed HIV-1 infection. The HIV diagnostics lab will inform the clinic of the outcome of the HIV testing algorithm (ie, HIV infected, HIV uninfected, or redraw required).

The Endpoint Adjudicator(s) and/or designee(s) will be an expert in the fields of infectious diseases or laboratory medicine independent of the clinical investigators participating in this trial.

5.3 Date of HIV-1 Infection for vaccine efficacy

The primary analysis will be done in the PP population where participants becoming HIV infected or dropping out before the visit (4 weeks post vaccination 3 visit) after the third vaccination or not having received all of the first 3 vaccinations within the specified time window or having a major specified protocol deviation, will be excluded from the analysis. The date of HIV-1 diagnosis will be the draw date of the first sample that leads to a positive test result by the diagnostic algorithm described above. Dropouts will be censored at the

time of their last HIV-1 negative test.

6 Interactions of Study Statisticians with the Data and Safety Monitoring Board

At each 6-monthly Data and Safety Monitoring Board (DSMB) meeting, the study statisticians will present Open, Chair/Medical Officer, and Closed Reports; all tables and figures included in the Closed Report are specified in Appendix A with statistical analyses of safety further described in Section 7. For a subset of tables and figures in the Closed report, the Open Report includes tables and figures with the same information except with trial information presented pooled across the two treatment groups to preserve blinding to treatment assignment. In addition, the following interim monitoring reports will be presented at each DSMB meeting:

- monitoring report of vaccine efficacy for potential harm, non-efficacy, and high efficacy per Section 10.1, where
  - potential harm monitoring starts at the 10^{th} pooled MITT infection endpoint and is continually performed with each additional endpoint until the time non-efficacy/high efficacy monitoring is triggered,
  - non-efficacy monitoring starts at the 60^{th} pooled MITT infection and proceeds at every scheduled DSMB meeting thereafter (anticipated every six months) until the end of Stage 1, and
  - high efficacy monitoring is harmonized with non-efficacy monitoring and starts when 150 MITT participants reach their Month 36 visit (end of Stage 2); if Stage 2 occurs, one additional high efficacy interim analysis will be performed at 6 months after the end of Stage 1;
- monitoring of the use of Truvada as pre-exposure prophylaxis (PrEP) first reported for the April 2019 DSMB meeting per Section 10.2;
- monitoring report for futility to assess vaccine efficacy per Section 10.3 starting at the latest 12 months after the first participant is enrolled and continuing every 6 months thereafter; and
- monitoring report for performance standards of quality of trial conduct per Section 10.4.

7 Statistical Analysis of Safety

The analyses of safety will be performed on the FAS, all randomized participants who receive at least 1 vaccine administration. All safety analyses will be tabulated by treatment group

(active vaccine, placebo) according to the as-treated principle.

7.1 Baseline Comparability

Treatment groups will be compared for baseline characteristics including demographics and laboratory measurements, using descriptive statistics (percentages, means, ranges).

7.1.1 Reactogenicity

The number and percentage of participants experiencing each type of reactogenicity sign or symptom will be tabulated by severity,treatment group and the length of reactogenicity follow-up (3 days versus 7 days) and the percentages will be displayed graphically by arm and the length of reactogenicity follow-up (3 days versus 7 days). For a given sign or symptom, each participants reactogenicity will be counted once under the maximum severity for each injection visit. In addition, to the individual types of events, the maximum severity of local pain or tenderness, induration or erythema, and of systemic symptoms will be calculated. Kruskal-Wallis tests will be used to test for differences in severity between arms.

7.1.2 AEs and SAEs

AEs occuring during the first 30 days after vaccination will be summarized using MedDRA System Organ Class and preferred terms. SAEs will be shown for the whole study period. Tables will show by treatment group the number and percentage of participants experiencing an AE within a System Organ Class or within preferred term category by severity and by relationship to study product. For the calculations in these tables, a participant with multiple AEs within a category will be counted once under the maximum severity and by causal relationship to study product. Formal statistical testing comparing arms is not planned since interpretation of differences must rely heavily upon clinical judgment. Parallel analyses will include all AEs and AEs leading to participant withdrawal or early discontinuation of study product(s). A listing of SAEs reported to the Janssen Global Medical Safety Group will provide details of the events including severity, relationship to study product, time between onset and last vaccination, and number of vaccinations received.

7.1.3 Reasons for vaccination discontinuation and early study termination

The number and percentage of participants who discontinue vaccination and who terminate the study early will be tabulated by treatment arm and including the reason for discontinuation.

8 Statistical Analysis of Vaccine Efficacy

Except where specified, all vaccine efficacy endpoint analyses are performed in the PP cohort. All analyses only use samples and data collected prior to study unblinding. The final vaccine efficacy analysis will occur when the last enrolled participant has reached the month 36 visit.

8.1 Primary Analysis of Vaccine Efficacy

The time between enrollment and the date of HIV-1 infection diagnosis determined in Section 5 is evaluated for all subjects. Subjects becoming HIV-infected or dropping out before their month 7 visit or not having received all of the first 3 vaccinations within the specified time window or having a major protocol deviation will be excluded from the primary PP analysis. The failure times of participants with last HIV-negative test at or after the month 24 visit are right-censored at the right edge of the month 24 allowable visit window [defined as 60 days after the target day (728 days) for the month 24 visit, equal to 25.91 months post-enrollment]. For participants with last HIV-negative test prior to the month 24 visit and without diagnosis of the HIV-1 infection primary endpoint at or before the month 24 visit, their failure times are right-censored at the date of the last HIV-negative test.

We define the primary vaccine efficacy parameter, VE(7–24), as one minus the probability of the primary efficacy endpoint between the month 7 and month 24 visit for the vaccine group divided by the probability of the primary efficacy endpoint between the month 7 and month 24 visit for the placebo group times 100 percent. More specifically, the primary VE parameter, $VE_{7-24}(\tau)$, will be estimated at the fixed time point $\tau = 25.91$ months post-enrollment with the specification of the value of τ determined as noted above. Each of the two cumulative incidence parameters in $VE_{7-24}(\tau)$ will be estimated using the transformed Nelson-Aalen estimator for the cumulative hazard function evaluated at time τ defined above. We will use the delta method to obtain the asymptotic 95% CI for the log cumulative incidence ratio (vaccine/control) and then back-transform these confidence bounds to the VE scale. Both the point estimate and the 95% CI for VE will be reported.

$VE(\tau)$ is the target parameter for the primary analysis of overall VE. The primary analysis tests

$$\begin{aligned} &\text{the null hypothesis } \mathrm{H}_0\text{: } VE_{7-24}(\tau) = 0\% \text{ versus} \\ &\text{the alternative hypothesis } \mathrm{H}_1\text{: } VE_{7-24}(\tau) \neq 0\% \end{aligned} \tag{1}$$

using a 2-sided $\alpha = 0.05$ level Wald test of the equality of log cumulative hazard functions at τ for the vaccine group and the control group.

Cox proportional hazards model will also be used for estimating VE(7-24), measured by 1 minus the hazard ratio (vaccine vs. placebo) and for a score test of whether the VE(7-24) differs from 0%.

As a sensitivity analysis to the primary analysis of vaccine efficacy in the PP population, targeted minimum loss-based estimation (TMLE) is used to estimate this vaccine efficacy parameter. TMLE is used to estimate the cumulative incidences of HIV-1 infection over time for each of the vaccine and placebo groups, through to Month 24. These estimates will then be contrasted to estimate the primary VE parameter VE(7-24), which adjusts for covariates. This analysis can correct for bias due to measured participant covariates that predict both per-protocol status and HIV-1 infection. As part of the implementation of the TMLE the Super Learner is used to generate initial estimates of the conditional censoring distribution and the iterated conditional means.

In addition, to assess potential time-effects of vaccine efficacy, the Kaplan-Meier method will be used to plot the estimated cumulative incidence rates over time for the vaccine and placebo groups. This method will be used to estimate (i) cumulative vaccine efficacy over time, defined as (1 minus the ratio (vaccine/placebo) of cumulative incidence by time t) $\times$ 100%, and (ii) additive-difference vaccine efficacy over time, defined as the difference (placebo minus vaccine) in cumulative incidence by time t, with the method of Parzen, Wei, and Ying (1997) applied to obtain pointwise and simultaneous 95% CIs. In addition, the longitudinal targeted minimum loss-based estimation method as implemented in the R package `survtmle` will be used for estimation, which in addition to allowing confounding adjustment can correct for potential bias due to covariate-dependent censoring.

8.2 Secondary Analyses of Vaccine Efficacy

8.2.1 Modified Intention to Treat Vaccine Efficacy Estimate

As a key secondary analysis, we report the MITT vaccine efficacy from Months 0-24, i.e. VE(0–24). This analysis is conducted within the MITT cohort of participants who are HIV-1 uninfected on the date of first vaccination. The failure times of MITT participants for the analysis of VE(0–24) are right-censored following the same approach as used for the primary analysis of vaccine efficacy. We define the MITT vaccine efficacy as one minus the probability of the MITT efficacy endpoint between the month 0 and month 24 visit for the vaccine group divided by the probability of the MITT efficacy endpoint between the month 0 and month 24 visit for the placebo group times 100 percent. The VE parameter VE(0–24) will be estimated at the same time point τ as defined in Section 8.1. Each of the two cumulative incidence parameters in $VE_{0-24}(\tau)$ will be estimated using the transformed Nelson-Aalen estimator for the cumulative hazard function evaluated at time τ defined in Section 8.1.

This key secondary analysis tests

$$\begin{aligned} &\text{the null hypothesis } \mathrm{H}_0\text{: } VE_{0-24}(\tau) = 0 \text{ versus} \\ &\text{the alternative hypothesis } \mathrm{H}_1\text{: } VE_{0-24}(\tau) \neq 0 \end{aligned} \tag{2}$$

using a 2-sided $\alpha = 0.05$ level Wald test of the equality of log cumulative hazard functions at τ for the vaccine group and the control group within the MITT cohort.

Cox proportional hazards model will also be used for estimating VE(0-24), measured by 1 minus the hazard ratio (vaccine vs. placebo) and for a score test of whether the VE(0-24) differs from 0%.

8.2.2 Vaccine Efficacy Estimates that Adjusts for Covariates

As a supportive analysis of the hypotheses in (1), a doubly robust estimator will be used to estimate cumulative incidences of the primary efficacy endpoint over time in the per protocol population, accounting for both per-protocol and right censoring status. The estimator used for this analysis is described in Westling et al. (2021). Super Learner (van der Laan, Polley, and Hubbard, 2007; Westling, Luedtke, Gilbert, and Carone, 2021) is used to generate initial estimates of the conditional protocol violation and censoring distribution and the iterated conditional means. The Super Learner library is specified in Table 2. If any of the algorithms included in the Super Learner library return an error when run on the trial data (e.g., due to the rarity of the event), then they will be removed. Each method considers adjustment for vaccine/placebo assignment, baseline demographic covariates, and the baseline behavioral risk score built via supervised learning as described in Section 8.5. The particular demographic covariates considered are: site, country, and age at enrollment.

VE parameters will be estimated by 1 minus the ratio (vaccine group/placebo group) of the estimates of the HIV incidence probability between months 7 and 24. Influence-curve based variance estimators of each cumulative incidence are used, and the delta method is applied to obtain the variance estimator of the log cumulative incidence ratio. Point estimates and 95% pointwise and simultaneous Wald CIs for cumulative incidence curves and $VE(t)$ curves will be plotted. For the final time point of 24 months, 2-sided Wald p-values will be reported.

8.2.3 Vaccine Efficacy over Time Periods Other Than Months 0–24, Months 7–24

We will repeat the methods from Section 8.1 in the per-protocol cohort in windows other than Months 7-24. In particular, we perform these analyses in the FIS cohort, studying endpoints occurring between Month 13 and 24; and in the FIS cohort, studying endpoints occurring between Month 13 and 36. We will repeat the methods from Section 8.2.1 in the MITT cohort in windows other than Months 0–24. In particular, we perform these analyses among: MITT subjects who are HIV-1 uninfected at Month 13, studying endpoints occurring between Month 13 and 24; MITT subjects who are HIV-1 uninfected at Month 13, studying endpoints occurring between Month 13 and 36; and MITT subjects, studying endpoints occurring between Months 0 and 36.

For all analyses evaluating VE through Month 24, the failure times will be right-censored following the same approach as used for the primary analysis of vaccine efficacy. For analyses that aim to evaluate VE through Month 36, the right-censoring approach used will depend

Table 2: Models included in the Super Learner library. Z denotes vaccine/placebo assignment, B baseline behavioral risk score as described in Section 8.5, and W a vector of baseline demographic covariates. The columns indicate what type of candidate estimator was used (GLM = generalized linear model, step = stepwise GLM using both AIC and BIC as selection criteria, GAM = generalized additive model [additive Cox regression in the case of survival endpoint], RF = random survival forests, KM = Kaplan Meier) and what covariates were included ($x * y$ indicates a cross product between covariates x and y).

Model type	Covariates
Propensity score estimate	
GLM	Z
GLM	$Z + B + W$
GLM	$Z * B$
step	$Z * B + W$
Event and censoring estimates	
KM	$\emptyset$
Cox	Z
Cox	$Z * B$
Cox	$Z * B + W$
GAM	$Z + B$
GAM	$Z + B + W$
RF	$Z + B + W$

on the number of participants who attend their month 36 visit at the end of follow-up. The failure time convention right-censoring approach used and definition of the estimand of interest will depend on how large this number is in each of the two arms. In particular:

Case 1: If at least 150 participants in each arm of the per-protocol cohort have attended their Month 36 visit and had a negative HIV-1 test at that visit, then VE will be evaluated through time $\tau = 36.5$ months, where 36.5 denotes the upper allowable visit window for the month 36 visit. The failure times of participants with last HIV-1 negative test at or after the month 36 visit are right-censored at the right edge of the month 36 allowable visit window (month 36.5). For participants whose last HIV-1 negative test occurs prior to the month 36 visit and without diagnosis of a primary HIV-1 infection endpoint at or before the month 36 visit, their failure times are right-censored at the date of their last HIV-1 negative test.

Case 2: If, in at least one of the two arms of the per-protocol cohort, there are not 150 participants that have attended their Month 36 visit and had a negative HIV-1 test at that visit, then, in order to ensure stable estimation, VE will be evaluated through a time τ that falls before 36.5 months. The time τ will be defined as the maximum

time point at or after which 150 participants in both the vaccine and placebo groups of the per-protocol cohort have HIV-1 negative tests. For participants whose last HIV-1 negative test occurs prior to time τ and without diagnosis of a primary HIV-1 infection endpoint at or before time τ, their failure times are right-censored at the date of their last HIV-1 negative test. The failure times of all other participants will be right censored at time τ.

8.2.4 Vaccine Efficacy Accounting for Number of Founding Viruses

Another secondary analysis may be conducted to assess the vaccine efficacy on HIV-1 acquisition over 24 months using the method of Follmann and Huang (2015) that incorporates information on the number of HIV-1 founder viruses in HIV-1-infected participants. The method has increased efficiency relative to Cox proportional hazards regression if the vaccine reduces the number of founders.

8.3 Analysis of Exploratory Objective 1

The use of oral FTC/TDF as PrEP will be assessed as described in Section 10.2.

8.4 Analysis of Exploratory Objective 2

If there is substantial PrEP use detection at baseline, then the Cox model will be used to assess whether vaccine efficacy significantly differs in subgroups with detectable versus undetectable baseline PrEP use, and to make inferences on vaccine efficacy separately in each subgroup. In addition, if there is substantial PrEP use detected over the course of the study, then a Cox model with PrEP detectability as a time-varying covariate will be used to assess different vaccine efficacy by time-varying subgroup and for each subgroup separately.

More specifically, the primary analysis will be repeated where only MITT infection endpoints with no evidence of PrEP use at the time of HIV-1 diagnosis and at the time of the earliest evidence of infection will be included in the analysis. Intracellular tenofovir (TFV) concentration (see Section 8.3) will be used to determine eligibility for this analysis. A participant is eligible if the TFV concentration is below the lower limit of detection as defined above at the diagnosis visit and at the visit with earliest evidence of HIV-1 infection (if different from the diagnosis visit). Since the ARVs are only detectable in plasma for roughly 14 days (Patterson et al. 2010) and some participants may have become infected before the 14-day window, with this approach we are not assured that all those included in the analysis were not using prophylactic ARVs at the time of infection. Therefore an additional analysis may be conducted that addresses this issue by also excluding participants from the HIV-1 acquisition analysis if they self-reported ARV use in the last 30 days at either the diagnosis

visit or the last visit prior to diagnosis. These analyses will evaluate uninfected participants without accounting for data on their TFV concentrations.

8.5 Analysis of Exploratory Objective 3

Several analyses will make use of a baseline behavioral risk score variable. This section describes how this variable will be constructed.

Let Y denote the primary efficacy endpoint indicator (i.e., $Y = 1$ if the endpoint was observed and $Y = 0$ otherwise), W a vector of demographic and behavioral risk factors collected at baseline and specified in Table 3, and Z denote a vaccine indicator ($Z = 0, 1$ denotes placebo, vaccine). For the i^{th} participant, we define the baseline behavioral risk score as a bivariate vector $\left(\widehat{P}(Y_i = 1 \mid Z_i = 0, W_i), \widehat{P}(Y_i = 1 \mid Z_i = 1, W_i)\right)$, where we use loss-based super-learning to estimate the best model for $P(Y = 1 \mid Z = z, W) = E_0[Y \mid Z = z, W]$, $z = 0, 1$, where E_0 denotes expectation under the true data generating distribution. This is a standard prediction problem. We estimate $E_0(Y \mid Z = z, W)$ with a minimizer of the risk of a loss: $\psi_0 = \arg\min_\psi P_0 L(\psi)$, with $Pf \equiv \int f(o) dP(o)$. We select binary log-likelihood loss $L(\psi)(O) = -\{Y \log \psi(W) + (1 - Y) \log(1 - \psi(W))\}$, given its good performance for a rare event outcome, and we stratify by Z when fitting the risk minimization problem to ensure that we are estimating $P(Y = 1 \mid Z, W)$ rather than $P(Y = 1 \mid W)$. To construct an optimal estimator among any given class of candidate estimators, we use loss-based super-learning. The oracle inequality for the cross-validation selector guarantees that the estimator is asymptotically at least as good as any candidate in the set of candidate estimators. We refer to van der Laan, Polley, and Hubbard (2007) and Rose and van der Laan (2011) for details.

Let $\hat{\Psi}_j : \mathcal{M}_{NP} \to \Psi(\mathcal{M})$ be a candidate estimator that maps an empirical distribution of $(O_1, \ldots, O_n)$ (i.e., an element of the nonparametric model $\mathcal{M}_{NP}$ of probability distributions) into the parameter space $\Psi(\mathcal{M}) = \{\Psi(P) : P \in \mathcal{M}\}$, $j = 1, \ldots, J$. This library of candidate estimators could include a variety of parametric model based estimators as well as a variety of machine learning algorithms, possibly coupled with different dimension-reduction strategies, and possibly indexed by a variety of tuning parameters.

Let $B_n \in \{0, 1\}^n$ be a random split of the sample into a training sample $\{i : B_n(i) = 0\}$ and validation sample $\{i : B_n(i) = 1\}$. For example, if we use V-fold cross-validation defined by a partitioning of the sample in V equal size groups, then B_n has V possible realizations, each occurring with probability $1/V$, and each split corresponds with setting the components of B_n in one of the V-folds equal to 1 and setting the other components equal to 0. Let P^0_{n,B_n} and P^1_{n,B_n} be the empirical distributions of the training and validation sample corresponding with split-vector B_n, respectively. The cross-validated risk of the j-th candidate estimator is then defined as $E_{B_n} P^1_{n,B_n} L(\hat{\Psi}_j(P^0_{n,B_n}))$.

Define $\hat{\Psi}_\alpha = \sum_{j=1}^J \alpha_j \hat{\Psi}_j$ as a weighted linear combination of the candidate estimators, where

Table 3: Baseline covariates used as input features W by the prediction/screening algorithms included in the construction of the superlearner model

Variable	**Definition**
site	Site Name
countryl	Country Name
aage	Age at Enrollment in Years (continuous)
bmibl	BMI at Enrollment (continuous)
arace	Race (Black, Colored/Mixed, Indian, Multiple, White)
syph	Syphilis Diagnosis at Enrollment (Positive, Negative, Not done/indeterminate)
ngonor	Gonorrhea Diagnosis at Enrollment (Positive, Negative, Not done/indeterminate)
ctrach	Chlamydia Trachomatis Diagnosis at Enrollment(Positive,Negative,Not done/indeterminate)
trich	Trichomonas Diagnosis at Enrollment (Positive, Negative, Not done/indeterminate)
agefsex	Age at First Sexual Intercourse in Years by the Time of Enrollment (continuous)
agesexp*	Age in Years of Oldest Sex Partner (continuous)
analsex*	Anal Sex (N = No, Y = Yes, Not Asked)
conduse	Condom Use by the Time of Enrollment (N = No, Y = Yes, Not Asked)
diagsti*	Diagnosed with STI (DK = Don't Know, N = No, Y = Yes)
exchsex*	Exchange Services for Sex (N = No, Y = Yes)
gendisch*	Genital Discharge (N = No, Y = Yes)
gensores*	Genital Sores or Ulcers (N = No, Y = Yes)
homearea	Type of Area Living In by the Time of Enrollment (Rural/Countryside, Urban/City/Town)
homemat	Main Materials of Home by the Time of Enrollment (Formal, Informal, Traditional, Other)
homserv	Building Has at Least Three Services by the Time of Enrollment (N = No, Y= Yes)
invag*	Insert Item (paper, cloth, etc.) Into Vagina to Make It More Dry for Sex (N = No, Y= Yes)
livbrths	How Many Babies Alive at Birth by the Time of Enrollment (continuous)
livwpart	Living with Main Sex Partner by the Time of Enrollment (N=No, Y=Yes, NA)
mainpart	Married or Have Main Sex Partner by the Time of Enrollment (N = No, Y= Yes)
nsexact*	Number of Sex Acts (number of times) (continuous)
nsexp*	Number of Sex Partners (number of people) (continuous)
othpart	Main Sex Partner Has Other Partners by the Time of Enrollment (N=No, Y=Yes, NA)
sexhivp*	Sex with HIV+ Partner (DK = Don't Know, N = No, Y = Yes)
usexalc*	Unprotected Sex with Alcohol Use (Never, 1-2 times, 3-5 times, 6 or more times)
usexhivp*	Unprotected Sex with HIV+ Partner (DK = Don't Know, N = No, Y = Yes)

* The reference time-period is the previous month before enrollment

the weights α_j are restricted to be non-negative and sum to 1. The cross-validation selector for the continuous family $\{\hat{\Psi}_\alpha : \alpha\}$ of candidate estimators is defined as:

$$\alpha_n = \arg\min_\alpha E_{B_n} P^1_{n,B_n} L(\hat{\Psi}_\alpha(P^0_{n,B_n})),$$

and the super-learner is then defined as $\hat{\Psi}(P_n) = \hat{\Psi}_{\alpha_n}(P_n)$.

By the oracle inequality for the cross-validation selector we have that, if the expectation of the loss-based dissimilarity $\min_\alpha E_{B_n} P_0\{L(\hat{\Psi}_\alpha(P^0_{n,B_n})) - L(\psi_0)\}$ between the oracle selected estimator and ψ_0 converges to zero at a slower rate than $1/n$, then

$$\frac{E_0 E_{B_n} P_0\{L(\hat{\Psi}_{\alpha_n}(P^0_{n,B_n})) - L(\psi_0)\}}{E_0 \min_\alpha E_{B_n} P_0\{L(\hat{\Psi}_\alpha(P^0_{n,B_n})) - L(\psi_0)\}} \to 1, \text{ as } n \to \infty.$$

Table 4: Prediction and screening algorithms used by the superlearner for building the baseline behavioral risk score in the HVTN 705 trial

Learner	**Screen***
SL.mean	all
SL.glm	all
	glmnet
	univar_logistic_pval
	highcor_random
SL.glm.interaction	glmnet
	univar_logistic_pval
	highcor_random
SL.glmnet	all
SL.gam	glmnet
	univar_logistic_pval
	highcor_random
SL.xgboost	all
SL.ranger.imp	all

* Screen details:
all: includes all variables
glmnet: includes variables with non-zero coefficients
in the standard implementation of SL.glmnet that optimizes
the lasso tuning parameter via cross-validation
univar_logistic_pval: Wald test 2-sided p-value
in a logistic regression model < 0.10
highcor_random: if pairs of quantitative variables
with Spearman rank correlation > 0.90, select one of
the variables at random

In other words, excluding the unrealistic situation in which one of our candidate estimators is a correctly specified parametric model, the super-learner is asymptotically equivalent with the oracle selected estimator.

In addition, we can evaluate the super-learner by its cross-validated risk, using a cross-validation scheme S_n (e.g., using V-fold cross-validation again as in the super-learner):

$$\text{CV-RISK} = E_{S_n} P^1_{n,S_n} L(\hat{\Psi}(P^0_{n,S_n})),$$

which involves rerunning the super-learner on learning samples $\{i : S_n(i) = 0\}$ and evaluating it on test samples $\{i : S_n(i) = 1\}$, and averaging the performance across the different splits. When we do this for the trial, we will let S_n denote a random variable representing a mixture of 20 randomly selected 10-fold cross-validation schemes, so that the cross-validated risk can be evaluated by averaging the cross-validated risk from 20 different 10-fold cross-validation splits.

This represents an estimator of the true conditional risk

$$E_{S_n} R(\hat{\Psi}(P^0_{n,S_n}) \mid P_0) \equiv E_{S_n} P_0 L(\hat{\Psi}(P^0_{n,S_n})),$$

and one can also construct a Wald-type 95% confidence interval for the latter true conditional risk parameter $E_{S_n} R(\hat{\Psi}(P^0_{n,S_n}) \mid P_0)$ given by CV-RISK $\pm 1.96\sigma_n/\sqrt{n}$, where $\sigma_n^2 = E_{S_n} P^1_{n,S_n} \left\{ L(\hat{\Psi}(P^0_{n,S_n})) - E_{S_n} P^1_{n,S_n} L(\hat{\Psi}(P^0_{n,S_n})) \right\}^2$. The theory behind the asymptotic correctness of this data adaptive confidence interval is given in Hubbard, Kherad-Pajouh, and van der Laan (2016). We build the super-learner using the R package `SuperLearner` available on `CRAN`.

If the number of primary endpoints in each treatment group is > 30, we will use superlearner with 5-fold cross-validation, separately for the vaccine and placebo groups; otherwise leave-one-out cross-validation will be used. For the cross-validated superlearner, 5-fold cross-validation will be used for the outer cross-validation irrespective of the number of primary endpoints. These cross-validation rules align with those made for the baseline behavioral risk score analysis in Moderna's COVE trial of the mRNA-1273 vaccine.

Table 4 shows the combinations of prediction and screening algorithms used by the super-learner. We will plot point and 95% CI estimates of the cross-validated area-under-the-ROC curves (AUCs) (Hubbard, Kherad-Pajouh, and van der Laan, 2016) for each individual learning approach as well as for discrete super-learner and super-learner, and, for each $z = 0, 1$, we will separately select the learner with the lowest cross-validated AUC for finalizing the risk scores $\left(\widehat{P}(Y_i = 1 \mid Z_i = 0, W_i), \widehat{P}(Y_i = 1 \mid Z_i = 1, W_i)\right)$. For each $z = 0, 1$, we will report the cross-validated AUC of the best model as a summary of the quality of the risk score for the given study group.

9 Primary Analysis to Be Conducted at the End of Stage 1

The target Month 24 visit date for the last enrolled participant is May 25, 2021, with the upper allowable limit of the visit window on June 22, 2021. Unless the trial is stopped early, the target data cut date for the end-of-Stage-1 primary analysis will be May 28, 2021.

If the trial reaches the end of Stage 1, the primary analysis report will include the following SAP-specified analyses:

1. Primary objective 1 (parts of Section 8.1)
    (a) Estimation of month 7–24 incidence rate, with 95% CI, by treatment arm in the PP cohort

(b) Estimation of month 7–24 cumulative incidence over time by treatment arm in the PP cohort

(c) Estimation of month 7–24 cumulative VE over time, with 95% pointwise and simultaneous CI, in the PP cohort, with the point estimate and 95% CI for the primary VE parameter reported at time point τ

(d) Two-sided Wald test of H_0: month 7–24 cumulative $VE(\tau) = 0\%$ at τ in the PP cohort

(e) Estimation of month 7–24 HR VE using the Cox model, with 95% CI, in the PP cohort

(f) Two-sided score test of H_0: month 7–24 HR $VE = 0\%$ in the PP cohort

2. Secondary objective 2 (parts of Section 8.2.1)

   (a) Repeat 1(a)–(f) except for the month 0–24 period and the MITT cohort

3. Secondary objective 3 (parts of Section 8.2.1)

   (a) Repeat 1(a)–(c) and (e) except for the month 13–24 period and the FIS cohort

   (b) Repeat 1(a)–(c) and (e) except for the month 13–24 period and the cohort of MITT participants uninfected at month 13

4. Additional efficacy analyses requested by Janssen

   (a) Repeat 1(a)–(c) and (e) except for the cohort of MITT participants uninfected at month 7

   (b) Repeat 1(a) except for the month 7–30 and 7–36 periods and the PP cohort

   (c) Repeat 1(b)–(c) except for the month 7–36 period and the PP cohort

   (d) Repeat 1(a) except for the month 0–30 and 0–36 periods and the MITT cohort

   (e) Repeat 1(b)–(c) except for the month 0–36 period and the MITT cohort

   (f) Repeat 1(a) except for the month 13–30 and 13–36 periods and the FIS cohort

   (g) Repeat 1(b)–(c) except for the month 13–36 period and the FIS cohort

5. Primary objective 2

   (a) Maximum local reactogenicity by treatment assignment and the length of reactogenicity follow-up (3 days versus 7 days) (table and figure)

   (b) Maximum systemic reactogenicity by treatment assignment and the length of reactogenicity follow-up (3 days versus 7 days) (table and figure)

   (c) Adverse events by system organ class, severity, and treatment assignment (table)

   (d) Adverse events by preferred term, severity, and treatment assignment (table)

(e) Adverse events related to study product by preferred term, severity, and treatment assignment (table)

(f) Listing of adverse events of special interest (AESIs), by preferred term and severity (listing)

(g) Serious adverse events (SAEs) reported to the global safety database ordered by treatment assignment, participant, and decreasing severity (listing)

(h) Adverse events leading to early participant withdrawal or early discontinuation of study administrations (listing)

6. Limited demographics and baseline disease characteristics, disposition and adherence tables, by treatment arm

Janssen will propose to the OG a list of recipients at Janssen of the analysis report. The OG's approval will be required before report distribution.

The analyses of vaccine efficacy through 30 months will define the final time point for analysis (τ) and the right-censoring process in the same way as described for the analyses of vaccine efficacy through 24 months, except now indexed off of the month 30 visit instead of the month 24 visit.

10 Trial Monitoring

The trial is monitored in four ways: (1) interim monitoring of VE for potential harm, non-efficacy, and high efficacy; (2) monitoring of the use of Truvada as PrEP; (3) monitoring for futility to assess VE; and (4) operational monitoring for other performance standards of quality of trial conduct. Data on these monitoring activities are collated into interim reports and are presented to the independent DSMB every 6 months. Following each DSMB meeting, the DSMB reports to the Oversight Committee (OC) a summary of the trial review, which may include recommendations to modify or terminate the trial.

10.1 Interim Monitoring of Vaccine Efficacy for Potential Harm, Non-Efficacy, and High Efficacy

Vaccine efficacy is monitored by the independent DSMB at each DSMB meeting, with monitoring triggers based on numbers of primary HIV-1 infection endpoints (and, for non- and high efficacy, the length of completed follow-up by all on-study participants) for when the DSMB meetings begin to include formal evaluation of the potential-harm, non-efficacy, and high-efficacy stopping boundaries (these triggers are described below). The approach to sequential monitoring of VE is similar to that described in Gilbert, Grove et al. (2011). The monitoring of VE is also similar to that used for the HVTN 505 Phase 2b HIV-1 vaccine

efficacy trial (Hammer et al., 2013). The trial uses interim monitoring of VE to stop early for:

- potential harm [establish that VE(0–24) $< 0\%$ based on a 2-sided monitoring-adjusted 90% confidence interval lying below 0%]
- non-efficacy [establish that both VE(0–24) and VE(7–24) $< 40\%$ and $H_0 : VE \leq 0\%$ is not rejected, based on a 2-sided 95% nominal confidence interval for VE lying below 40% and covering 0%]
- high efficacy [establish that VE(0–36) $> 70\%$ based on a 2-sided 95% nominal confidence interval lying above 70%].

The interim monitoring for non-efficacy and high-efficacy is based on 2-sided 95% confidence intervals in order that the result of the study that would be reported in the abstract of a journal article would convey convincing evidence supporting the conclusion of non-efficacy or high efficacy. Freidlin, Korn, and Gray (2010) discuss a rationale for this approach to non-efficacy monitoring. The interim monitoring for potential harm is based on 2-sided 90% confidence intervals for prudence to protect the safety of study participants, i.e., less precision is required to meet a guideline for potential harm than to meet guidelines about non-efficacy or high efficacy. The interim monitoring plan is summarized in Table 5.

The potential harm monitoring is done after every primary endpoint event starting at the 10^{th} total pooled over the treatment groups, through to the time at which the non-efficacy monitoring commences. Once the non-efficacy monitoring begins, the non-efficacy monitoring serves the purpose of detecting a harmful effect of the vaccine regimen to elevate the endpoint rate compared to placebo. Details of the procedures used for the three monitoring outcomes are described next.

10.1.1 Potential harm monitoring

Heyse et al. (2008) and the HVTN 505 Phase 2b HIV-1 vaccine efficacy trial (Hammer et al., 2013) are examples of randomized, placebo-controlled efficacy trials that used continuous monitoring for an elevation in the endpoint rate in the active versus control treatment arm. Continuous monitoring means that an unblinded statistician has visibility to the treatment assignment of each diagnosed MITT HIV-1 infection as they are determined in real-time, and, after each confirmed HIV-1 infection diagnosis, this statistician notes whether a stopping boundary is reached that indicates that the relative cumulative rate of HIV-1 infection (RR, vaccine/placebo) exceeds one. If the stopping boundary is met, then the unblinded statistician immediately informs the Chair of the DSMB and the Executive Secretary of the DSMB through secure communication procedures. As such, the potential harm monitoring is in real-time, with a result possible at any time, whereas in contrast the non-efficacy and high efficacy monitoring is conducted only at the 6-monthly scheduled DSMB meetings

Table 5: Summary of interim monitoring of VE

Monitoring Outcome	Hypotheses	Testing Approach	Size	Monitoring Plan	Timing of Interim Analyses
Potential Harm	$H_0 : VE(0\text{–}24) \geq 0\%$ vs. $H_1 : VE(0\text{–}24) < 0\%$	Exact 1-sided binomial test of the proportion of infections assigned to vaccine group	1-sided $\alpha = 0.05$	Near-constant* 1-sided p-value cut-off controlling the FWER at $\alpha = 0.05$	After every MITT infection from 10^{th} total until first non-efficacy analysis
Non-Efficacy	$H_0 : VE(0\text{–}24) \geq 40\%$ vs. $H_1 : VE(0\text{–}24) < 40\%$ and $H_0 : VE(7\text{–}24) \geq 40\%$ vs. $H_1 : VE(7\text{–}24) < 40\%$	Wald test	1-sided $\alpha = 0.025$	Unadjusted 95% CIs for VE(0–24) and VE(7–24): lower bounds $< 0\%$ and upper bounds $< 40\%$	6-monthly starting once all participants reach 13 months, and 60 MITT infections are observed, then through the end of Stage 1
High Efficacy	$H_0 : VE(0\text{–}36) \leq 70\%$ vs. $H_1 : VE(0\text{–}36) > 70\%$	Wald test	1-sided $\alpha = 0.025$	Unadjusted 95% CI for VE(0–36): lower bound $> 70\%$	Harmonized with non-efficacy monitoring, starting when 150 MITT participants reach 36 months of follow-up; if Stage 2 occurs, 1 additional analysis halfway through Stage 2

*An increasing per-test alpha until a constant level is reached that, if applied to all subsequent tests, maintains the specified FWER (see Table 6)

through the end of Stage 1 or at any extra DSMB meetings requested by the DSMB (with the exception that the first non-efficacy analysis would be done and reported to the DSMB by secure means at the time when all participants in follow-up reach 13 months of follow-up, and 60 total MITT primary endpoint events are observed).

The potential harm monitoring is done using an exact one-sided binomial test of the null hypothesis $H_0 : p \leq 1/2$ versus the alternative hypothesis $H_1 : p > 1/2$, where p is the probability that an HIV-1-infected participant was assigned to the vaccine group (as compared to being assigned to the placebo group). The tests start at the 10^{th} total infection and are performed continuously until the non-efficacy monitoring commences. Each test is performed at a prespecified nominal/unadjusted alpha-level, which may vary over the multiple tests. The alpha-level used for each test is determined indirectly as follows: first, we choose the overall type I error rate we are willing to accept over the course of the monitoring (the overall probability that we reach a stopping boundary during the trial when the vaccine regimen is actually safe, i.e., true $RR = 1$, equivalently $p = 1/2$); and second, we choose whether/how to vary the alpha-level from test-to-test. Using these pieces of information we determine the exact alpha-levels to be used for each test.

Table 6: Potential harm monitoring stopping boundaries: One-sided p-value cut-offs for rejecting $H_0 : p \leq 1/2$ in favor of $H_1 : p > 1/2$ with an exact 1-sided binomial test where p is the probability that an HIV-1-infected participant was assigned to the vaccine group

No. of Infections in Vaccine Group	No. of Infections in Placebo Group	Per-Test Type I Error Rate (1-Sided P-value Cut-off)
10	0	0.011
10	1	0.014
11	2	0.014
13	3	0.014
15	4	0.014
16	5	0.014
18	6	0.014
20	7	0.014
21	8	0.014
23	9	0.014
24	10	0.014
26	11	0.014
27	12	0.014
28	13	0.014
30	14	0.014
31	15	0.014
33	16	0.014
34	17	0.014
35	18	0.014
37	19	0.014
38	20	0.014
39	21	0.014

An overall 1-sided type I error rate of 0.05 is chosen for the family-wise error rate of the multiple hypothesis tests starting at the 10^{th} HIV-1 infection endpoint through to the 60^{th}. This type I error rate is chosen to balance the competing goals of participant safety and preventing false positive results. To prevent stopping too early, perhaps due to spurious results caused by wide sampling variability, stopping prior to the accumulation of 10 total infections was ruled out. Table 6 shows the potential harm stopping boundaries in terms of the one-sided p-value cut-offs for a selected set of potential harm interim analyses starting at the 10^{th} HIV-1 infection event.

10.1.2 Non-efficacy monitoring

The DSMB will monitor the vaccine for non-efficacy defined as evidence that it is highly unlikely that the vaccine has a beneficial effect on acquisition of VE(0–24) or of VE(7–24) of 40% or more. Such analyses will start when two conditions are met: (1) all participants in follow-up have reached 13 months of follow-up since enrollment, and (2) at least 60 MITT infections have been observed. Condition (2) is chosen as the minimum MITT infection total at which a point estimate of zero VE would just correspond to an unadjusted upper 95% confidence bound for VE(0–24) equal to 0.40 based on a Cox proportional hazards model; the approach, stated in condition (2), to starting non-efficacy monitoring was suggested by Freidlin, Korn, and Gray (2010). Thereafter, non-efficacy interim analyses will proceed at every scheduled DSMB meeting (anticipated every six months) until the end of Stage 1.

The criterion for non-efficacy is that, for both VE(0–24) and VE(7–24), the lower 95% confidence bound lies below 0% and the upper 95% confidence bound lies below 40%. By checking confidence intervals for both VE(0–24) and VE(7–24), and requiring completed 13 months of follow-up, the monitoring plan is designed to protect against stopping prematurely based on ramping vaccine efficacy over the intercurrent period of 0–13 months.

At each non-efficacy interim analysis, VE(0–24) and VE(7–24) are estimated with 2-sided 95% unadjusted confidence intervals using the same method as for the final analysis of VE described in Section 8.1. Based on projected accrual and HIV-1 incidence in the placebo group, it is expected that some participants will have reached the Month 24 study visit by the time of the first non-efficacy interim analysis, such that VE(0–24) and VE(7–24), which are both defined based on cumulative incidence of HIV-1 infection through to the Month 24 visit, can be estimated. At the early interim analyses, however, it is possible that only very few participants will have reached the Month 24 study visit by the time of the analysis, precluding the ability to estimate VE(0–24) and VE(7–24) with adequate precision. In this event, cumulative vaccine efficacy through time τ, $VE_{0-24}(\tau)$ and $VE_{7-24}(\tau)$, will be estimated with fixed time point τ chosen to be the latest possible time point where stable estimation of both VE parameters can be achieved; this is operationalized by defining τ as the maximum time point when at least 150 participants in the per-protocol cohort are observed to be at risk for the primary efficacy endpoint in each treatment arm.

In all interim analyses of VE, participants who have not experienced the primary endpoint will be right-censored at the time of their last visit.

10.1.3 High efficacy monitoring

Monitoring of high efficacy allows early detection of a highly protective vaccine if there is evidence that VE(0–36) $> 70\%$. Stage 1 high efficacy analyses will be harmonized with those for non-efficacy monitoring, with the exception that the high efficacy analyses will only start once at least 150 participants in the MITT cohort have reached the terminal Month 36 visit.

This condition ensures that sufficient follow-up has accumulated to estimate VE(0–36). If Stage 2 occurs, there will also be one final high efficacy interim analysis at the midpoint of Stage 2, defined as 6 months after the end of Stage 1.

The criterion for high efficacy is that the unadjusted 95% confidence interval for VE(0–36) lies above 70%. Note that, while the potential harm and non-efficacy monitoring is restricted to infections diagnosed up to Month 24, the monitoring for high efficacy counts all infections up to Month 36 because early stopping for high efficacy would only be warranted under evidence for durability of vaccine efficacy.

10.2 Monitoring of the Use of PrEP

The use of oral FTC/TDF as PrEP (either off-study or provided in the study) may impact study outcomes (e.g., by lowering the HIV-1 incidence rate rendering a loss of statistical power). Dried blood spot (DBS) samples will be used for assessment of quantitative concentrations of intracellular tenofovir diphosphate (TFV-DP).

The prevalence of oral FTC/TDF use will be estimated and reported both as any detectable use and as effective use. More specifically, estimated percentages of person-years at-risk (PYR) during any detectable FTC/TDF use and during inferred effective FTC/TDF use will be reported.

Next we summarize how inferred effective use is measured. Current knowledge about PrEP in women indicates that consistent use of 6–7 doses a week is required to achieve protection. The lower quartile of simulated TFV-DP levels in DBS at 6 doses per week is 1064 fmol/punch (Castillo-Mancilla et al., 2013). In this trial, we will use 1,000 fmol/punch as the cut-off to define effective PrEP use based on the lower quartile cited above. Ongoing work with calibration of DBS from directly observed dosing studies may refine these thresholds. PrEP use measures will be reported by arm to the DSMB; in addition, both the OC and the protocol team leadership will see pooled estimates of FTC/TDF use.

DBS samples will be collected and stored for prospective monitoring of PrEP use at all study sites at pre-specified fixed sample collection days each month (see Section 10.2.1 for details). Furthermore, in order to increase the precision of PrEP use estimation, the frequency of sample collection may be increased.

At a given calendar time T (e.g., a fixed date prior to a scheduled DSMB meeting), we are interested in the population-level parameter, the percentage of person-years at-risk for HIV infection on effective PrEP use between initiation of DBS sample storage and time T. The definition of this parameter assumes that we have an assay readout from stored samples that accurately measures effective PrEP use as a binary outcome at the time the sample was drawn; importantly, it does not require an accurate measurement of effective PrEP use the day before or for any period of time earlier than the sampling day. In addition to estimating the percentage of person-years at risk on effective PrEP use, we define a similar parameter,

the percentage of person-years at risk exposed to detectable PrEP using the lower limit of quantitation (LLOQ) of the DBS assay.

Define the target parameter of interest as

$$\Phi(T) = \frac{\int_{T_0}^{T} p(t)E[Y(t)]dt}{\int_{T_0}^{T} E[Y(t)]dt}$$

where T_0 is the time since the first person enrolled after DBS storage commenced, $p(t)$ is the percent of participants on effective PrEP use at time $t \in [T_0, T]$, and $E[Y(t)]$ is the expected number of participants with DBS storage at-risk for HIV at time t.

We estimate $\Phi(T)$ based on the binary PrEP use readout from the DBS assay and the DBS sampling plan. Let i, ranging from 1 to N, index study participants and let j, ranging from 1 to M_i, index participant DBS collection dates that are sampled for assaying. For each sample collected, we have an indicator x_{ij} of effective PrEP use which is only measured if the DBS sampling indicator, Δ_{ij}, is equal to 1. The estimated percentage of person-years at risk on effective PrEP use from the initiation of PrEP monitoring time T_0 until time T is defined as

$$\widehat{\Phi}(T) = \frac{\Sigma_{i=1}^{N}\Sigma_{j=1}^{M_i}\Delta_{ij}x_{ij}\pi_{ij}^{-1}P_{ij}}{\Sigma_{i=1}^{N}\Sigma_{j=1}^{M_i}P_{ij}}$$

where the sampling probability, π_{ij}, and person-years, P_{ij}, are defined below.

Let k, ranging from 1 to K, index the DBS sampling plan collection intervals $[T_0, T_1]$, $(T_1, T_2]$, ..., $(T_{K-1}, T_K]$ where the right endpoint of each interval is a sample collection date (with $T_K \equiv T$) and define $\mathcal{T}_k$ as the k^{th} interval. Let $t_{i1} < t_{i2} < \ldots < t_{iM_i}$ be the sampling times in $[T_0, T]$ for the i^{th} participant and define t_{i0} as the maximum of T_0 and the i^{th} participants enrollment time. The DBS sampling plan determines which samples, collectively across participants, will be assayed for PrEP use. Define the set of samples, S_{ij}, as all samples collected during the same collection interval $\mathcal{T}_k$ as the sample i, j. That is, $S_{ij} \equiv \{i', j' | t_{i'j'} \in \mathcal{T}_k \text{ for } k \text{ s.t. } t_{ij} \in \mathcal{T}_k\}$. Define the sampling probability as

$$\pi_{ij} \equiv \frac{\Sigma_{i',j' \in S_{ij}}\Delta_{i'j'}}{|S_{ij}|}$$

where $|S_{ij}|$ is the number of samples in set S_{ij}. Define person-years, P_{ij}, as $t_{ij} - t_{i(j-1)}$. Note that our parameter of interest can be defined and estimated for the entire study cohort as well as for subregions (e.g., South African sites). We will report bootstrap 95% confidence intervals for $\Phi(T)$.

The same approach is used for point and confidence interval estimation of the percentage of person-years at risk exposed to detectable PrEP use.

An example of PrEP use report statistics are shown in Table 7.

Table 7: Summary of detectable and effective PrEP use

Number of DBS specimens collected over the 1st batch period	N_1
Number of DBS specimens assayed over the 1st batch period	N_2
Proportion of assayed specimens with TFV-DP above LLOQ (95% CI)	N_3/N_2 (x.xx, x.xx)
Proportion of assayed specimens with TFV-DP above effective use threshold* (95% CI)	N_4/N_2 (x.xx, x.xx)
Percent person-years on detectable PrEP through date T=xx (95% CI)#	$\widehat{\Phi}(T)_d$ (x.xx, x.xx)
Percent person-years on effective PrEP through date T=xx (95% CI)#	$\widehat{\Phi}(T)_e$ (x.xx, x.xx)
Repeat through Nth batch period	

*1000 fmol/punch

#$\widehat{\Phi}(T)_d$ is an estimate of the target parameter of interest based on measured TFV-DP above the LLOQ. Similarly, $\widehat{\Phi}(T)_e$ is based on TFV-DP above the 1000 fmol/punch threshold for effective PrEP use.

10.2.1 Simulation Study for PrEP Monitoring

A simulation study was conducted to help determine the DBS sampling schedule and, subsequently, the number of DBS samples that should be collected and assayed to achieve sufficient precision of estimation of the target parameters of interest. For a simulated trial, the PrEP monitoring plan was implemented based on the following assumptions: 1) enrollment falls randomly between Monday and Friday; 2) follow-up visits are scheduled on a randomly selected weekday during the week of the target visit date; 3) missed visits are distributed uniformly at a rate of 10%; 4) the dropout rate is 0.1/PYR in both treatment arms based on an exponential distribution; 5) the HIV infection rate in the placebo arm is 0.042/PYR based on an exponential distribution; and 6) initiation of DBS sample collection begins as of November 1, 2017. PrEP use was simulated at various constant rates among trial participants, and, for each rate, a random sample of participants is assumed to be continually on PrEP during the entire follow-up period. Simulation results are presented for the number of samples collected spaced at 6 month intervals. The intervals are spaced such that the last collection time point is approximately 3.5 months prior to the next DSMB meeting to allow the intracellular TFV-DP assay to be run and a report generated. In this simulation, the following three DBS sampling schedule scenarios are considered:

S1. all samples are collected and assayed from visits held on a business day closest to the 15th of each month,

S2. all samples are collected and assayed from visits held on business days closest to the 1st and the 15th of each month,

S3. all samples are collected and assayed from visits held on business days closest to the 1st, the 10th, and the 20th of each month.

Simulation results for DBS sampling schedules S1–S3 are shown in Figures 1–3, respectively. For 6-monthly intervals shown on the x-axis, all samples assayed cumulatively through the

end of the month are used to estimate PrEP use prevalence from the beginning of DBS sample collection until the given timepoint using the estimator described in Section 10.2.

The results indicate that approximately 500 assayed samples are required to estimate PrEP use prevalence with adequate precision. This number of samples is projected to accumulate by the end of May 2019 under S1, by the end of December 2018 under S2, and by the end of October 2018 under S3. The gain in precision becomes limited after the number of assayed samples grows beyond 700 under S1 or beyond 1,000 under S2 or S3; however, continuing to assay samples over the course of the trial will allow estimation of temporal trends in PrEP use.

Based on these simulation results, the trial will follow the DBS sampling schedule S2 as defined above.

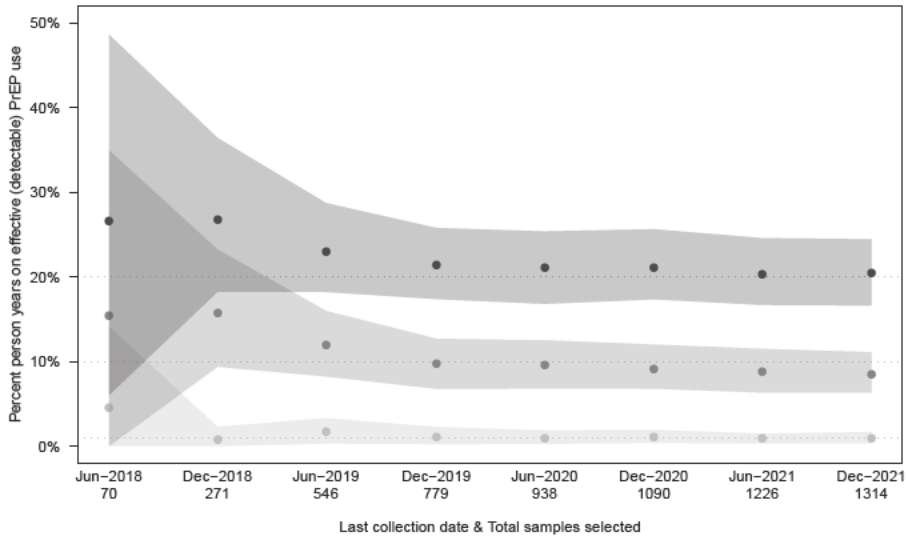

Figure 1: Estimates of PrEP use prevalence under DBS sampling schedule S1 considering three PrEP use scenarios (20%, 10% and 0% of PYRs). The total number of samples that would be assayed according to the simulation study is shown below each date on the x-axis. Estimates are shown as dots of varying color intensity corresponding to the three PrEP use scenarios. The color bands show 95% pointwise bootstrap confidence intervals.

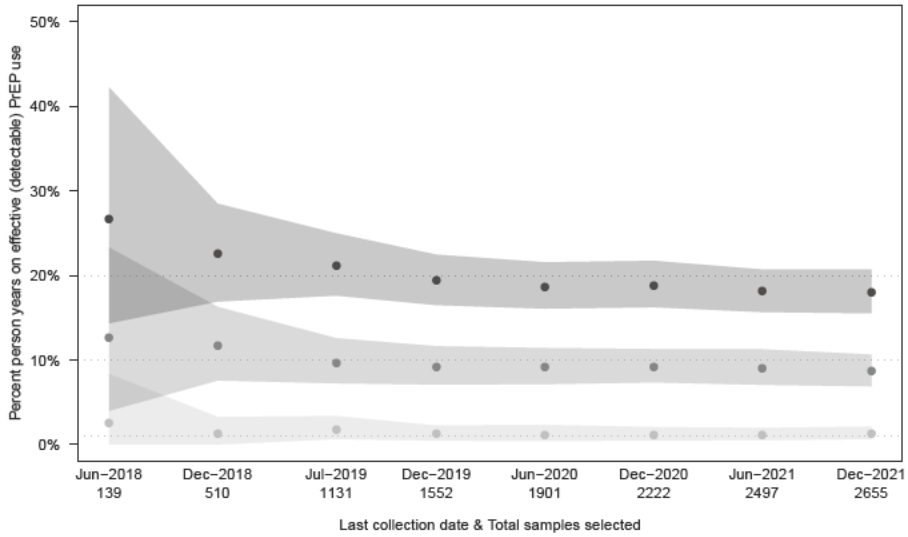

Figure 2: Estimates of PrEP use prevalence under DBS sampling schedule S2 considering three PrEP use scenarios (20%, 10% and 0% of PYRs). The total number of samples that would be assayed according to the simulation study is shown below each date on the x-axis. Estimates are shown as dots of varying color intensity corresponding to the three PrEP use scenarios. The color bands show 95% pointwise bootstrap confidence intervals.

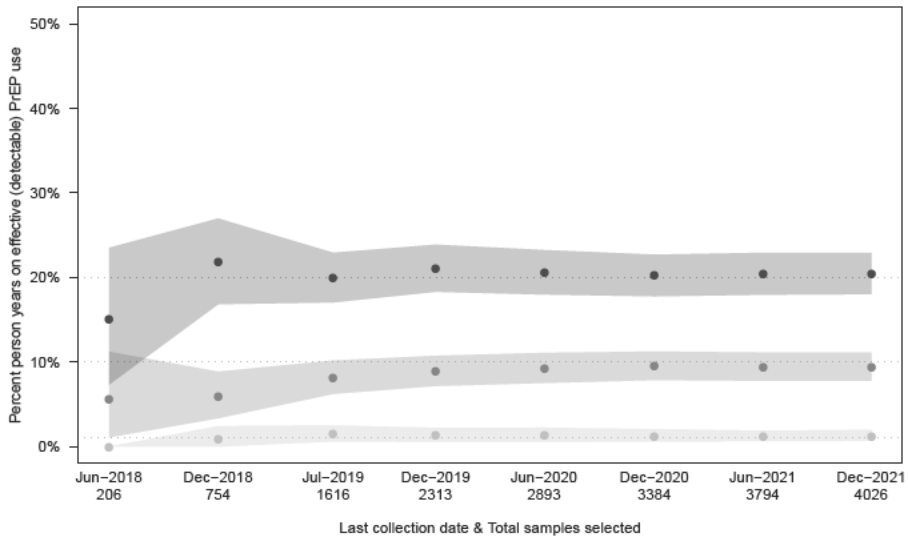

Figure 3: Estimates of PrEP use prevalence under DBS sampling schedule S3 considering three PrEP use scenarios (20%, 10% and 0% of PYRs). The total number of samples that would be assayed according to the simulation study is shown below each date on the x-axis. Estimates are shown as dots of varying color intensity corresponding to the three PrEP use scenarios. The color bands show 95% pointwise bootstrap confidence intervals.

10.3 Monitoring for Futility to Assess Vaccine Efficacy

The objective of monitoring the trial for futility to assess VE is to monitor progress toward the minimal needed target number of treatment arm-pooled HIV-1 primary endpoint infections by the end of Stage 1 in the MITT and PP cohorts. Four targets are monitored for:

1. the total number of MITT HIV-1 infections needed to achieve 80% power to detect VE(0–24) = 43% (i.e., VE(0–7) = 25%, VE(7–24) = 50% and assuming a constant placebo incidence rate over time),
2. the total number of MITT HIV-1 infections needed to achieve 60% power to detect VE(0–24) = 43% (i.e., VE(0–7) = 25%, VE(7–24) = 50% and assuming a constant placebo incidence rate over time),
3. the total number of PP HIV-1 infections needed to achieve 80% power to detect VE(7–24) = 50%,
4. the total number of PP HIV-1 infections needed to achieve 60% power to detect VE(7–24) = 50%,

each calculated based on trial simulations using the R `seqDesign` package. The target numbers in 1.–4. are 108, 65, 79, and 49, respectively. The rationale for considering the four targets is as follows: (i) if the trial cannot achieve 80% power to detect VE(7–24) = 50% (targets 1. and 3.), considerations about enrollment modification or expansion are warranted, (ii) if the trial cannot achieve even 60% power to detect VE(7–24) = 50% (targets 2. and 4.), considerations about completing the trial early for futility to assess VE are warranted, and (iii) the MITT cohort allows for greater precision of estimation of the number of infections by the end of Stage 1 (targets 1. and 2.), while the PP cohort is used in the primary VE analysis (targets 3. and 4.).

Two versions of the futility monitoring report will be generated. A report provided to the DSMB will be included in 6-monthly closed DSMB reports, starting in October 2018, and will report:

(a) the estimated distribution of the total (i.e., treatment arm-pooled) number of HIV-1 infections in the MITT cohort by the end of Stage 1, with corresponding power to reject $H_0 : VE(0\text{–}24) \leq 0\%$ using a 1-sided 0.025-level Wald test under the alternative hypothesis that $VE(0\text{–}24) = 43\%$,

(b) the estimated probability that the total number of HIV-1 infections in the MITT cohort by the end of Stage 1 is ≥ 108 (target 1) with 95% credible intervals,

(c) the estimated probability that the total number of HIV-1 infections in the MITT cohort by the end of Stage 1 is ≥ 65 (target 2) with 95% credible intervals,

(d) the estimated distribution of the number of HIV-1 infections in the MITT cohort by the end of Stage 1 in each treatment arm,

(e) the estimated distribution of the total (i.e., treatment arm-pooled) number of HIV-1 infections in the PP cohort by the end of Stage 1, with corresponding power to reject $H_0 : VE(7\text{–}24) \leq 0\%$ using a 1-sided 0.025-level Wald test under the alternative hypothesis that $VE(7\text{–}24) = 50\%$,

(f) the estimated probability that the total number of HIV-1 infections in the PP cohort by the end of Stage 1 is ≥ 79 (target 3) with 95% credible intervals,

(g) the estimated probability that the total number of HIV-1 infections in the PP cohort by the end of Stage 1 is ≥ 49 (target 4) with 95% credible intervals, and

(h) the estimated distribution of the number of HIV-1 infections in the PP cohort by the end of Stage 1 in each treatment arm.

The distributions in (a), (d), (e), and (h) will also be summarized by the mean number of HIV-1 infections with a Wald 95% confidence interval. The estimation procedures for (a)–(h) will be conducted under each of the following three scenarios:

(i) the treatment arm-pooled infection rates in (a)–(c) and (e)–(g), and the two treatment arm-specific infection rates in (d) and (h) used for generating future data are based on a Bayesian model and the prior assumptions that $VE(0\text{–}7) = 25\%$, $VE(7\text{–}24) = 50\%$ (the design alternative) and the placebo incidence rate is constant over time,

(ii) the treatment arm-pooled infection rate in (a)–(c) and (e)–(g), and the two treatment arm-specific infection rates in (d) and (h) used for generating future data are based on a Bayesian model and the prior assumption that $VE = 0\%$ (the null hypothesis), and

(iii) the treatment arm-pooled infection rate in (a)–(c) and (e)–(g) used for generating future data is based on a Bayesian model and the prior assumption that the infection rate equals the observed to-date infection rate.

The reason for conducting the estimation procedure under (i)–(iii) is that the purpose of the results in (b) is to trigger considerations about enrollment modifications, whereas the purpose of the results in (c) is to trigger considerations about early trial completion due to futility, where it is desired to reach a guideline based on (b) more easily/readily than a guideline based on (c). Accordingly, the results for (b) are interpreted focusing on the prior of VE(0–7) = 25% and VE(7–24) = 50% (i.e., scenario (i)), which makes it more likely to reach a guideline than the prior of VE(0–24) = 0%, and the results for (c) are interpreted focusing on the prior of VE(0–24) = 0% (i.e., scenario (ii)), which makes it less likely to reach a guideline than the prior of VE(0–7) = 25% and VE(7–24) = 50%. Results for (b) and (c) based on carrying forward the observed to-date infection rate in scenario (iii) as well as results for (f) and (g) in scenarios (i)–(iii) based on the PP cohort provide additional

guidance to the DSMB regarding considerations about enrollment modifications or early trial completion.

Furthermore, a treatment-blinded report will be generated for distribution to the OG before each DSMB meeting takes place and will report estimates listed in (a)–(c) and (e)–(g) above calculated based on treatment-blinded data in scenarios (i)–(iii). The reported results pertaining to estimates (a)–(c) and (e)–(g) under scenarios (i)–(iii) will be identical to those in the DSMB report.

In addition, a special DSMB and OG report may be generated approximately 2 months before the projected completion of enrollment in order to provide timely information for a potential decision to modify enrollment before the enrollment apparatus is closed down.

While it is the primary responsibility of the OG to make decisions regarding trial operations and modifications based on the monitoring of treatment-blinded primary endpoints, given the resource issues involved, DSMB review is also needed because issues of scientific integrity are also involved. More specifically, the DSMB can evaluate the progress toward primary endpoint targets in the context of the treatment-unblinded data, and based on this review may recommend to the OG to complete the trial early due to reaching a guideline for futility to assess VE (specified below).

The monitoring for futility to assess VE includes the following guidelines for trial modifications:

- **Guideline for enrollment modifications.** If, in the VE(0–24) = 43% scenario for the prior distribution in (i) using the robust prior defined in Section 10.3.2.1, the estimated probability of reaching 108 total infections in the MITT cohort by the end of Stage 1 is less than 25%, the OG may consider enrollment modifications with the intention to be able to conduct the primary VE analysis with sufficiently high power.

- **Guideline for futility.** If, in the VE(0–24) = 0% scenario for the prior distribution in (ii) using the standard $\mathsf{Ga}(\alpha, \beta)$ prior as in Section 10.3.2, the estimated probability of reaching 65 total infections in the MITT cohort by the end of Stage 1 is less than 25%, the DSMB may recommend completing the trial early based on the inability to conduct the primary VE analysis with sufficiently high power. However, since this is a proof-of-concept trial, a high bar is desired for completing the trial early for futility, and therefore if this event occurs yet the non-efficacy monitoring has not started or the non-efficacy boundary has not been reached, then this guideline for futility also requires that the estimated VE(0–24) is $< 30\%$.

If enrollment is incomplete at the time of an interim futility analysis, then the outlined estimation procedures will use the average observed enrollment rate in approximately the last 6 months for generating future enrollment data. A Bayesian approach will be used for generating future HIV-1 incidence data, conditional on observed data to-date. More specifically, the estimates in (a)–(c) and (e)–(g) will condition on the observed to-date treatment

arm-pooled HIV-1 incidence rate in the respective cohort, whereas the estimates in (d) and (h) will condition on the observed to-date treatment arm-specific HIV-1 incidence rates in the respective cohorts. All estimates in (a)–(d) and (e)–(g) will also use the observed to-date treatment arm-pooled dropout rate in the respective cohort for generating future dropout data. Further details of these calculations, including the prior distributions, are described in Section 10.3.1.

If, at any time, these guidelines for futility to assess VE are met and yet it appears that value exists in continuing the trial, the statisticians will provide the DSMB and the Leadership Group with additional information, as appropriate, for use in their consideration of whether to recommend early trial completion.

10.3.1 Estimation of the number of HIV-1 infection endpoints at an interim analysis

The method for estimating the probability distribution of the number of HIV-1 infection endpoints by the end of Stage 1 is based on the following approach to simulating this trial. The trial is modeled as a combination of three processes—enrollment, dropout, and HIV-1 infection—and a large number of trials is simulated. The three processes are assumed independent and their distributions are taken to be Poisson, exponential, and exponential, respectively. Data are generated at the level of the individual participant, such that, for each participant, we obtain an enrollment time, an (underlying true) infection time, and a dropout time. Only the minimum of the infection and dropout times is observable, and the average value for this minimum is beyond the duration of the trial, such that neither event will be observed for most participants.

In the absence of observed trial data, the treatment arm-pooled as well as the treatment arm-specific parameters for the infection and dropout processes are chosen to match our pre-trial assumptions about these rates. In addition, the infection rate considers both the design alternative of $VE = 50\%$ and the null hypothesis of $VE = 0\%$ in the calculation of the total and treatment arm-specific numbers of endpoints. More specifically, treatment arm-pooled calculations in (a)–(c) and (e)–(g) assume

- pooled infection rate [MITT $VE(0\text{–}7) = 25\%$ and $VE(7\text{–}24) = 50\%$ scenario]: $0.5 \times 0.042 + 0.5 \times \left(\frac{7}{24} \times 0.75 + \frac{17}{24} \times 0.5\right) \times 0.042 = 0.0330$ infections/person-year at-risk, and
- pooled infection rate [$VE(0\text{–}24) = 0\%$ scenario]: 0.042 infections/person-year at-risk.

Conditioning on interim data, a complete time-to-event data set is simulated for the MITT cohort, and the PP cohort is extracted at the end.

We also assume a treatment arm-pooled dropout rate of 0.10 dropouts/person-year at-risk.

Treatment arm-specific calculations in (d) and (h) assume

- infection rate in the control arm: 0.042 infections/person-year at-risk,
- infection rate in the vaccine arm [MITT $VE(0\text{–}7) = 25\%$ and $VE(7\text{–}24) = 50\%$ scenario]: $\left(\frac{7}{24} \times 0.75 + \frac{17}{24} \times 0.5\right) \times 0.042 = 0.0241$ infections/person-year at-risk, and
- infection rate in the vaccine arm [PP $VE(7\text{–}24) = 50\%$ scenario]: $0.5 \times 0.042 = 0.021$ infections/person-year at-risk.

In each treatment arm, the dropout rate is assumed to be 0.10 dropouts/person-year at-risk.

The first step in simulating each trial is to enroll a certain number of participants per week according to a random draw from a Poisson distribution with rate parameter as listed above. Enrollment continues week-by-week until a total of 2,600 participants is reached. Second, each participant is assigned an exact enrollment day, uniformly distributed within their enrollment week. Following enrollment, the infection and dropout times are drawn from their respective exponential distributions, and the lesser of the two is recorded as occurring at the given time (possibly outside the time-window of the trial). We consider dropout events to have occurred at the dropout time (in days) that was generated (assuming it was less than the infection time). For participants who become HIV-1 infected, we record their time of diagnosis as the time of the first study visit following the true infection time. It is this time of diagnosis that we observe for infected participants.

A modification of the above procedure for simulating an efficacy trial is used for estimating metrics of futility to assess VE at a given interim analysis. The modification entails using the observed trial data to estimate parameters of the processes, rather than relying entirely on pre-trial assumptions. In particular:

- enrollment rate: if enrollment is incomplete, estimated based on the rate observed in approximately the last 6 months in the study,
- infection rate: drawn from a posterior distribution of the infection rate formed by combining the observed data with our prior specification about the infection rate based on the pre-trial assumptions, and
- dropout rate: estimated based on the treatment arm-pooled rate observed to date.

The rationale for a Bayesian approach for the infection rate (see Section 10.3.2 for details) is to help stabilize the infection rate early in the trial when insufficient time will have passed to accrue many infections. If we were to rely solely on the observed infections, we might by chance obtain very low rates, which would lead to an unrealistic prediction of the number of endpoints.

We consider various different gamma prior distributions for the infection rate in each of scenarios (i)–(iii) reflecting different weights assigned to the prior distribution (see Section 10.3.2 for details). Gamma distributions are considered because they are conjugate to the exponential distribution used for generating future infection data.

At a given interim analysis, 10^4 trials are simulated using the above procedure and treatment arm-pooled infection and dropout rates for estimates in (a)–(c) and (e)–(g). Separately, another set of 10^4 trials is simulated using the above procedure, treatment arm-specific infection rates, and the treatment arm-pooled dropout rate for estimates in (d) and (h). Each of these trials yields a projected number of infections by the end of Stage 1. These projected numbers of infections from each trial will be used to estimate the entire distribution of the number of infections by the end of Stage 1. The probability of reaching the target number of infections will be estimated as the proportion of trials with the projected number of infections greater than or equal to the target.

Figures on enrollment, HIV-1 incidence and dropout over time will also be included to aid interpretation of the results.

10.3.2 A Bayesian model for the HIV-1 incidence rate in estimation of the number of HIV-1 infection endpoints at an interim analysis

Let n_k and T_k denote, respectively, the infection count and the observed total person-time at risk at the time of the k-th futility analysis, pooling over all treatment arms. Additionally, let T^* denote the estimated total person-time at risk for the primary efficacy analysis at the end of Stage 1. Let the prior distribution of the pooled HIV-1 incidence rate p be $\mathsf{Ga}(\alpha, \beta)$ parametrized such that the prior mean $E\,p = \alpha/\beta$ (the same Bayesian method applies to the treatment arm-specific HIV-1 incidence rate). In scenario (i) for the treatment arm-pooled incidence rate, we additionally consider a robust prior distribution described in Section 10.3.2.1 used in the calculation evaluating the enrollment modifications guideline.

Generally, assuming that, conditional on p, the times to infection follow $\mathsf{Exp}(p)$, the posterior mean of p at the time of the k-th analysis equals

$$\begin{aligned} E[p \,|\, \text{data}] &= \frac{\alpha + n_k}{\beta + T_k} \\ &= \frac{\alpha}{\beta}\,\frac{\beta}{\beta + T_k} + \frac{n_k}{T_k}\,\frac{T_k}{\beta + T_k}, \end{aligned} \tag{3}$$

i.e., the posterior mean can be interpreted as a convex combination of the prior mean and the observed incidence rate. For a given $\beta > 0$, the weight on the prior mean at the first analysis depends on the accumulated person-time at risk (T_1), and the weight will decrease in subsequent analyses because $\beta/(\beta+T_k)$ is a decreasing function of T_k, which is a desirable Bayesian property.

In order to identify α and β, it is desirable that the prior mean equals the pre-trial assumed treatment arm-pooled incidence rate p^* (e.g., under PP VE(7–24)=50%, $p^* = 0.5 \times 0.042 + 0.5 \times 0.5 \times 0.042 = 0.0315$), i.e.,

$$\frac{\alpha}{\beta} = p^*. \tag{4}$$

Furthermore, we propose to consider three values of β that correspond to the weights $w = \frac{1}{2}$, $\frac{1}{3}$ and $\frac{1}{4}$ on the prior mean at the time when 50% of the estimated total person-time at risk has been accumulated, i.e., for each value of w, β is defined as the solution to the equation

$$\frac{\beta}{\beta + T^*/2} = w.$$

It follows that

$$\beta = \beta(w, T^*) = \frac{wT^*}{2(1-w)}, \tag{5}$$

and the estimation of T^* is described in Section 10.3.2.2. For $w = \frac{1}{2}$, $\frac{1}{3}$ and $\frac{1}{4}$, we obtain $\beta = \frac{T^*}{2}$, $\frac{T^*}{4}$, and $\frac{T^*}{6}$, respectively.

At the k-th futility analysis and for each of the three values of β, we will sample the treatment arm-pooled HIV-1 incidence rate from the posterior $\mathsf{Ga}(\alpha + n_k, \beta + T_k)$ for generating future data and report the weight $\frac{\beta}{\beta+T_k}$ on the prior mean in the convex combination (3).

10.3.2.1 A robust prior model for the HIV-1 incidence rate in the calculation evaluating the guideline for enrollment modifications

The robust prior model (Schmidli et al., 2014) is implemented for the guideline to trigger enrollment modifications since it is designed to maximize the probability of meeting the guideline for large downward deviations from the protocol-assumed incidence rates, while minimizing a false trigger for protocol-assumed incidence rates.

The prior distribution of p is defined as a weighted mixture of two gamma distributions,

$$(1 - w_R)\mathsf{Ga}(\alpha_I, \beta_I) + w_R\mathsf{Ga}(\alpha_V, \beta_V),$$

where we set $w_R = 0.2$, and $\mathsf{Ga}(\alpha_V, \beta_V)$ and $\mathsf{Ga}(\alpha_I, \beta_I)$ represent the weakly informative and informative component of the mixture prior, respectively. The parameters β_V and β_I are calculated following (5) with $w = 1/1000$ and $w = 1/3$, respectively (and T^* per Section 10.3.2.2). Subsequently, α_V and α_I are calculated following (4) with p^* set to the pre-trial assumed treatment arm-pooled MITT incidence rate of 0.0330 infections/person-year at risk for both components of the mixture (i.e., assuming VE(0–7) = 25%, VE(7–24) = 50%, and a constant placebo incidence rate over time).

The posterior distribution at the time of the k-th analysis is derived following the conjugacy principle, as in (3), which results in a mixture of conjugate posteriors with updated weights

$$(1 - \widetilde{w}_{R,k})\mathsf{Ga}(\alpha_I + n_k, \beta_I + T_k) + \widetilde{w}_{R,k}\mathsf{Ga}(\alpha_V + n_k, \beta_V + T_k),$$

where

$$\widetilde{w}_{R,k} \propto w_{R,k} f_V / \left\{ w_{R,k} f_V + (1 - w_{R,k}) f_I \right\}$$

with $f_\cdot$ equal to

$$f_\cdot = \frac{\Gamma(\alpha_\cdot + n_k)/(\beta_\cdot + T_k)^{\alpha_\cdot + n_k}}{\Gamma(\alpha_\cdot)/\beta_\cdot^{\alpha_\cdot}}$$

(see, e.g., Bernardo and Smith, 2000, Section 5.2.3, pages 279–282).

10.3.2.2 Estimation of the total person-years at risk by the end of Stage 1

The total target sample size is $N = 2600$, the duration of Stage 1 follow-up per participant is $\tau = 2$ years, the pre-trial assumed dropout rate is $d^* = 0.1$ dropouts per person-year at risk (PYR), and, in the MITT $VE(0\text{–}7) = 25\%$ and $VE(7\text{–}24) = 50\%$ scenario, the pre-trial assumed treatment arm-pooled HIV-1 incidence rate is $p^* = 0.5 \times 0.042 + 0.5 \times \left(\frac{7}{24} \times 0.75 + \frac{17}{24} \times 0.5\right) \times 0.042 = 0.0330$ cases per PYR.

We consider the standard right-censored failure time analysis framework. Denoting the failure and censoring times as T and C, respectively, we assume that T is independent of C, $T \sim \mathsf{Exp}(p^*)$, and $C \sim \mathsf{Exp}(d^*)$. It follows that $X := \min(T, C) \sim \mathsf{Exp}(p^* + d^*)$ and

$$\begin{aligned} T^* &= N \times E[\min(X, \tau)] \\ &= N \times \left\{ E[X \mid X \le \tau]\, P(X \le \tau) + \tau\, P(X > \tau) \right\} \\ &= N \times \left\{ (p^* + d^*) \int_0^\tau x \exp^{-(p^*+d^*)x} \dot{x} + \tau \exp^{-(p^*+d^*)\tau} \right\} \\ &= N \times \frac{1 - \exp^{-(p^*+d^*)\tau}}{p^* + d^*}\,. \end{aligned}$$

This results in $T^* = 4565.85$ PYRs. For comparison, if all N participants were followed for τ years, the total PYRs would be $N\tau = 5200$ years.

Subsequently, for $T^* = 4565.85$ PYRs, if $T_1 = 0.2\,T^*$, the weights $\frac{\beta}{\beta + T_1}$ on the prior mean at the first futility analysis in the MITT cohort corresponding to $w = \frac{1}{2}$, $\frac{1}{3}$, and $\frac{1}{4}$ are 0.71, 0.56, 0.45, respectively. If $T_1 = 0.3\,T^*$, the respective weights on the prior mean are 0.63, 0.45, and 0.36.

10.4 Monitoring for Performance Standards of Quality of Trial Conduct

The protocol team and study investigators will have performance standards regarding the quality of trial conduct in addition to the study endpoint rate. Some of these use standard metrics detailed in the Network Evaluation Metrics and Standards document, whereas others are specific to the HVTN 705 trial. Some of these standards will relate to achievement of targeted levels of:

1. participant enrollment into the trial (targets based on protocol assumptions).
2. retention of participants (target 5% annual dropout or less, with minimally acceptable level of no more than 10% annual dropout; also target 90% visit attendance among participants under follow-up [NEC standard]).

3. adherence to study interventions (target 95% adherence for receipt of first three doses, with minimally acceptable level of 80%; and target 90% adherence for receipt of first four doses, with minimally acceptable level of 80%).

4. quality and timeliness of HIV-1 diagnostic testing.

5. quality and timeliness of data collected on case report forms.

10.4.1 Expanded details for reporting on item 4: quality and timeliness of HIV-1 diagnostic testing

1. **Timeliness:** Turnaround time from blood collection to diagnostic reporting is summarized. This process monitors the site, the site-processing lab, the shipping company and the actual diagnostics lab. In addition the turnaround time from arrival in the lab to reporting is monitored, which isolates the turnaround time to the lab.

2. **Quality:** No additional monitoring for HIV-1 diagnostic quality is done beyond the fact that the labs participate in CAP and VQA and they are all audited annually by DAIDS.

10.4.2 Expanded details for reporting on item 5: quality and timeliness of data collected on case report forms

This reporting will use NEC standard metrics, as detailed in the Network Evaluation Metrics and Standards document (pages 12-14). In summary, the reporting outputs are as follows:

1. **For Quality:** For **QC Rate**, the standard metric for satisfactory quality is < 10 QCs per 100 pages. The denominator for this metric includes the total number of pages entered in the database during the time period. For refaxes, only the most recent page faxed is included. All CRF pages faxed to SCHARP are included in this calculation. Total CRF pages are labeled (Total Pages1) in the DMQ table.

   For **QC Resolution**, the standard metric is $> 80\%$ of QCs resolved in < 7 calendar days.

   For **Percent EDCd**, note that quality tends to be higher when pages are submitted via EDC (real time validation, faster submissions, etc), and this metric incentivizes sites to use EDC.

2. **Timeliness:** The standard metric for CRF Submission Rate (% pages faxed/EDCd on time) is $> 90\%$ of pages submitted within < 4 calendar days. The denominator for this metric is not total pages, as it excludes pages that, according to study operations, may not be completed and/or faxed immediately following a study visit (e.g., screening visit forms). Log-based forms are also not included. For refaxes, only the initial page faxed

is included. Only time-critical CRF forms are calculated. Screening and log based forms are excluded. CRFs included in this calculation are labeled (Total Pages2). HIV test results are excluded.

The DSMB and the leadership of the HVTN 705 trial will monitor whether the trials are achieving at least minimally acceptable levels of key performance standards. The DSMB will make recommendations to improve areas that are deficient. Termination of a trial would be considered if it appears unlikely that minimally acceptable performance will be achieved.

11 Statistical Software

All analyses described in this SAP will be conducted in validated instances of *R* and *SAS*.

12 Roles of Study Statisticians

HVTN SDMC statisticians will be blinded or unblinded to treatment group. During protocol development and after primary follow-up is completed, there will be no distinction between the roles; both types of statisticians will be responsible for designing and analyzing the study. During the primary follow-up period, however, only the treatment-unblinded statistician(s) will see interim data broken down by treatment group. Their role will be to conduct the interim monitoring and to produce and present reports on accruing data to the study DSMB. During the primary follow-up period, treatment-blinded statisticians will see only the interim data pooled across study groups. This way, treatment-blinded statisticians can assist protocol leadership in making decisions about modifications to the protocol without being influenced by interim efficacy results.

Appendix A Mock Tables and Figures for DSMB Closed Report

Mock tables and figures for the DSMB Closed Report are included below. The Closed Report summarizes trial information pooled across the treatment groups (labeled Total) and by masked treatment assignment (labeled A and B). A subset of these tables and figures will form the Open Report, in which trial information is reported pooled across the treatment groups.

DRAFT

HVTN 705/VAC89220HPX2008

Data and Safety Monitoring Board (DSMB)
CLOSED Report
All Tables and Figures

Tables and Figures

Table ENROLL1. Disposition of Study Participants, by Treatment Assignment
Population: All Screened Subjects (N =)

	Total	A	B
Number Randomized	xxxx	xxxx	xxxx
Enrolled*	xxxx (xx.x%)	xxxx (xx.x%)	xxxx (xx.x%)
Not Enrolled	xxxx (xx.x%)	xxxx (xx.x%)	xxxx (xx.x%)
Pending Enrollment[+]	xxxx (xx.x%)	xxxx (xx.x%)	xxxx (xx.x%)
Number Enrolled*	xxxx	xxxx	xxxx
On Study[1]	xxxx (xx.x%)	xxxx (xx.x%)	xxxx (xx.x%)
Completed Study	xxxx (xx.x%)	xxxx (xx.x%)	xxxx (xx.x%)
Off Study Early	xxxx (xx.x%)	xxxx (xx.x%)	xxxx (xx.x%)

*Enrollment implies receipt of the first vaccination.
[+]This category reports participants randomized prior to the cutoff date, but not vaccinated prior to it.
[1]Study completion requires at least 24 months of follow-up for uninfected participants and 6 months post-diagnosis for infected participants.

Table BL1. Baseline Participant Characteristics, by Treatment Assignment
Population: Full Analysis Set (N = xxxx)

	Total	A	B
Total Enrolled	xxx	xxx	xxx
Age (Years)			
18 – 20	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
21 – 30	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
31 – 35	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Median (Min, Max)	xx (xx, xx)	xx (xx, xx)	xx (xx, xx)
BMI			
<18.5	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
18.5-25	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
≥25	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Race[+]			
Black	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Colored/Mixed	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
White	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Indian	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Asian	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Other	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Multiple	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)

[+]Participants my report more than one racial category – those who do so are categorized as 'Multiple'.
Some counts may not total to the number of participants, due to non-response. Missingness is not included explicitly as it is minimal.

Table BL2. Baseline Risk Behaviors, by Treatment Assignment
Population: Full Analysis Set (N = xxxx)

The reference time-period for these questions is the previous month, except for Condom Use

	Total	A	B
Total Enrolled	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Number of Sex Partners in the Last Month			
0	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
1	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
2	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
3-4	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
>=5	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Median (Min, Max)	xx (xx, xx)	xx (xx, xx)	xx (xx, xx)
Condom use, general frequency			
Always	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Sometimes	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Never	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Anal Sex			
Yes	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Had an HIV+ Partner			
Yes	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Unprotected Sex with HIV+ Partner			
Yes	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Exchange of Sex for Money/Gifts			
Yes	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Diagnosed with or treated for STI			
Yes	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Alcohol Use and Unprotected Sex			
Never	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
1-2 times	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
3-5 times	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
6 or more times	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)

Participants are not required to answer these questions, so counts may not match the number of participants.
Missingness is not included explicitly as it is minimal.
Percentages are relative to the total number of participants enrolled.

Table STATUS1. Study Status and Reasons for Early Study Termination, by Treatment Assignment
Population: Full Analysis Set (N = xxxx)

	Total		A		B	
Total Enrolled	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)
Status						
On Study, In Trt Phase	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)
On Study, Completed Trt Phase	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)
On Study, Discontinued Trt	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)
Completed Study, Completed Trt Phas	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)
Completed Study, Discontinued Trt	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)
Off Study Early, Completed Trt Phase	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)
Off Study Early, Discontinued Trt	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)
Reasons for Early Study Termination	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)
Death	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)
Participant refused further participation	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)
Unable to adhere to visit schedule	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)
Participant relocated	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)
Unable to contact	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)
Investigator decision	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)
Inappropriate enrollment	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)
Duplicate screening/enrollment	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)
Early study closure	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)
Participant incarcerated	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)
Other	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)
Early Study Termination Due to an AE	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)

Table STATUS2. Vaccination Status and Reasons for Discontinuation, by Treatment Assignment
Population: Full Analysis Set (N = xxxx)

	Total		A		B	
Total Enrolled	xxx		xxx		xxx	
Treatment Status						
In Trt Phase	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)
Completed Trt Phase	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)
Discontinued Trt	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)
Reasons for Discontinuation of Vacc.	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)
Death	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)
Adverse Experience	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)
Reactogenicity Symptom	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)
Other Clinical Event	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)
Unable to Contact / Out of Window	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)
Participant Refused Vaccination	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)
Other	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)
Missing	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)

Table RETEN1. Visit Retention, by Treatment Assignment
Population: Full Analysis Set (N = xxxx)

	Total	A	B
Total Enrolled	xxx	xxx	xxx
Month 3 / Vaccination 2			
Expected* for visit or terminated[+]	xxx	xxx	xxx
Completed	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Missed	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Terminated[+]	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Month 6 / Vaccination 3			
Expected* for visit or terminated[+]	xxx	xxx	xxx
Completed	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Missed	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Terminated[+]	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Month 6.5 / Post-Vaccination			
Expected* for visit or terminated[+]	xxx	xxx	xxx
Completed	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Missed	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Terminated[+]	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Month 12 / Vaccination 4			
Expected* for visit or terminated[+]	xxx	xxx	xxx
Completed	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Missed	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Terminated[+]	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)

*Participants are considered expected for a visit when they reach the end of their visit window.
[+]Terminated prior to completion of visit as defined by submission of a specimen collection form.
Participants that terminate from the study continue to be marked as terminated and counted at subsequent visits.

Table TRTADH1. Treatment Adherence by Vaccination Visit and Treatment Population: Full Analysis Set (N = xxxx)

	Total	A	B
Month 3 / Vaccination 2			
Expected* or terminated[+]	xxx	xxx	xxx
Received treatment	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Missed visit and trt, still on-trt	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Completed Visit, Missed trt, still on-trt	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Discontinued treatment[+], still on-study	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Terminated study prior to vaccination	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Month 6 / Vaccination 3			
Expected* or terminated[+]	xxx	xxx	xxx
Received treatment	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Missed visit and trt, still on-trt	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Completed Visit, Missed trt, still on-trt	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Discontinued treatment[+], still on-study	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Terminated study prior to vaccination	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Month 12 / Vaccination 4			
Expected* or terminated[+]	xxx	xxx	xxx
Received treatment	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Missed visit and trt, still on-trt	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Completed Visit, Missed trt, still on-trt	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Discontinued treatment[+], still on-study	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Terminated study prior to vaccination	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)

*Participants are considered expected for a visit when they reach the end of their visit window.
[+]Terminated prior to completion of visit as defined by submission of a specimen collection form.
Participants that terminate from the study continue to be marked as terminated and counted at subsequent visits.

Table TRTADH2. Overall Cumulative Treatment Adherence
Population: Full Analysis Set (N = xxxx)

	Total	A	B
Month 3 / Vaccination 2			
Expected or terminated	xxx	xxx	xxx
Received all treatments	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Missed 1 treatment, on-trt	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Discontinued treatment*, on-study	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Terminated prior to vaccination	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Month 6 / Vaccination 3			
Expected or terminated	xxx	xxx	xxx
Received all treatments	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Missed 1 treatment, on-trt	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Missed 2 treatments, on-trt	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Discontinued treatment*, on-study	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Terminated prior to vaccination	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Month 12 / Vaccination 4			
Expected or terminated	xxx	xxx	xxx
Received all treatments	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Missed 1 treatment, on-trt	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Missed 2 or more treatments, on-trt	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Discontinued treatment*, on-study	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Terminated prior to vaccination	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)

*Discontinued treatment at or before the indicated visit..

Table PREG1. Pregnancy Listing
Population: Full Analysis Set (N = xxxx)

Trt	Publication ID	Pregnancy Outcome	LMP Onset Date	Pregnancy Outcome date	Date of Last Vacc. Prior to Outcome	Time from Prior Vacc. To LMP	# Vacc. Prior to LMP	# Vacc. Prior to Outcome	Total # Vacc.
A	XXX-XXXX	text	ddMMMyyyy	ddMMMyyyy	ddMMMyyyy	text	text	X	XX
A	XXX-XXXX	text	ddMMMyyyy	ddMMMyyyy	ddMMMyyyy	text	text	X	XX
A	XXX-XXXX	text	ddMMMyyyy	ddMMMyyyy	ddMMMyyyy	text	text	X	XX
	.								
	.								
	.								
B	XXX-XXXX	text	ddMMMyyyy	ddMMMyyyy	ddMMMyyyy	text	text	X	XX
B	XXX-XXXX	text	ddMMMyyyy	ddMMMyyyy	ddMMMyyyy	text	text	X	XX
B	XXX-XXXX	text	ddMMMyyyy	ddMMMyyyy	ddMMMyyyy	text	text	X	XX

Table RE1. Maximum Local Reactogenicity by Treatment Assignment
Population: Full Analysis Set (N = xxxx)

	Total (N=xxx)	A 3 Day* (N=xxx)	B 3 Day* (N=xxx)	A 7 Day* (N=xxx)	B 7 Day* (N=xxx)
Pain					
None	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Mild	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Moderate	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Severe	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Potentially Life-Threatening	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Tenderness					
None	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Mild	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Moderate	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Severe	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Potentially Life-Threatening	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Pain and/or Tenderness					
None	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Mild	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Moderate	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Severe	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Potentially Life-Threatening	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Erythema					
None / Not Gradable	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Gr 1: 2.5 - <5 cm / 6.25 - < 25cm2	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Gr 2: 5 - <10cm / 25 - <100cm2	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Gr 3: >=10cm / >=100cm2 / Complications	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Gr 4: Complications	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
.					
.					
.					

Participants are counted once per reactogenicity sign/symptom according the maximum severity level experienced across all vaccinations.
*Particpants for which 7-day reactogenicity data are collected are only shown in the respective columns (7 Day). Their reactogenicity in the first 3 days is not reflected together with the rest of 3-Day reactogenicity data.

Table RE2. Maximum Systemic Reactogenicity by Treatment Assignment
Population: Full Analysis Set (N = xxxx)

	Total (N=xxx)		A 3 Day* (N=xxx)		B 3 Day* (N=xxx)		A 7 Day* (N=xxx)		B 7 Day* (N=xxx)	
Malaise and/or Fatigue										
None	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)
Mild	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)
Moderate	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)
Severe	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)
Potentially Life-Threatening	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)
Myalgia										
None	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)
Mild	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)
Moderate	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)
Severe	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)
Potentially Life-Threatening	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)
Headache										
None	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)
Mild	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)
Moderate	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)
Severe	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)
Potentially Life-Threatening	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)
Nausea										
None	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)
Mild	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)
Moderate	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)
Severe	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)
Potentially Life-Threatening	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)
.										
.										
.										

Participants are counted once per reactogenicity sign/symptom according the maximum severity level experienced across all vaccinations.
*Particpants for which 7-day reactogenicity data are collected are only shown in the respective columns (7 Day). Their reactogenicity in the first 3 days is not reflected together with the rest of 3-Day reactogenicity data.

Table EAE1. Expedited Adverse Events (EAEs) Reported to the Regulatory Support Center
Ordered by Treatment Assignment, Participant, and Decreasing Severity
Population: Full Analysis Set (N = xxxx)

Trt.	Publ. ID	Severity	EAE No.	Adverse Experience	Onset Date	Relation to Vaccine – Stdy Site	Relation to Vaccine – Med Off.	Num. Prev. Vacs	Days Since Last Vacc.
A	XXX-XXXX	text	XXXXXX	text	ddMMMyyyy	text	text	X	XX
	XXX-XXXX	text	XXXXXX	text	ddMMMyyyy	text	text	X	XX
	XXX-XXXX	text	XXXXXX	text	ddMMMyyyy	text	text	X	XX
	.								
	.								
	.								
	.								
B	XXX-XXXX	text	XXXXXX	text	ddMMMyyyy	text	text	X	XX
	XXX-XXXX	text	XXXXXX	text	ddMMMyyyy	text	text	X	XX
	XXX-XXXX	text	XXXXXX	text	ddMMMyyyy	text	text	X	XX

Table AE1. Grade 2-5 Adverse Events by System Organ Class, Severity, and Treatment Assignment
Ordered by Decreasing Frequency
Population: Full Analysis Set (N = xxxx)

System Organ Class / Severity	Total (N=xxx)		A (N=xxx)		B (N=xxx)	
Participants with one or more AEs						
Moderate and Greater	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)
Severe and Greater	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)
Life Threatening	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)
Infections and infestations						
Moderate and Greater	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)
Severe and Greater	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)
Life Threatening	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)
Gastrointestinal disorders						
Moderate	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)
General disorders and administration Country conditions						
Moderate and Greater	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)
Severe	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)
Musculoskeletal and connective tissue disorders						
Moderate	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)
Injury, poisoning and procedural complications						
Moderate and Greater	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)
Severe and Greater	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)
Life Threatening	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)
Respiratory, thoracic and mediastinal disorders						
Moderate and Greater	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)
Severe	xxx	(xx.x%)	xxx	(xx.x%)	xxx	(xx.x%)
.						
.						
.						

n's are the number of participants reporting one or more AEs within a specific system organ class.
Percentages are calculated as n divided by the number of enrolled x 100.
AE records included in the table have been coded into MedDRA codes by SCHARP clinical staff.

Table AE2. Grade 2-5 Adverse Events by High Level Term, Severity, and Treatment Assignment
Ordered by Decreasing Frequency
Population: Full Analysis Set (N = xxxx)

	Total (N=xxx)	A (N=xxx)	B (N=xxx)
Participants with one or more AEs			
Moderate and Greater	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Severe and Greater	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Life Threatening	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Upper respiratory tract infections			
Moderate	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Streptococcal Infections			
Moderate	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Allergies to foods, food additives, drugs and other chemicals			
Moderate and Greater	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Severe and Greater	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Abdominal and gastrointestinal infections			
Moderate and Greater	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Severe and Greater	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Life Threatening	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Diarrhea (excl infective)			
Moderate	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
.			
.			
.			

n's are the number of participants reporting one or more AEs within a specific system organ class.
Percentages are calculated as n divided by the number of enrolled x 100.
AE records included in the table have been coded into MedDRA codes by SCHARP clinical staff.

Table AE3. Grade 2-5 Adverse Events Related to Study Product by Preffered Term, Severity, and Treatment Assignment; Ordered by Decreasing Frequency
Population: Full Analysis Set (N = xxxx)

	Total (N=xxx)	A (N=xxx)	B (N=xxx)
Participants with one or more related AEs			
Moderate and Greater	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Severe and Greater	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Life Threatening	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Upper respiratory tract infections			
Moderate	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Streptococcal Infections			
Moderate	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Allergies to foods, food additives, drugs and other chemicals			
Moderate and Greater	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Severe and Greater	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Abdominal and gastrointestinal infections			
Moderate and Greater	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Severe and Greater	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Life Threatening	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Diarrhea (excl infective)			
Moderate	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
.			
.			
	.		

n's are the number of participants reporting one or more related AEs within a specific system organ class.
Percentages are calculated as n divided by the number of enrolled x 100.
AE records included in the table have been coded into MedDRA codes by SCHARP clinical staff

Table AE4. Adverse Events of Special Interest (AESIs), by Preferred Term and Severity
Population: Full Analysis Set (N = xxxx)

Trt	Publ. ID	Severity	Adverse Experience	Onset Date	Relation to Vaccine	Num. Prev. Vacs	Days Since Last Vacc.
A	XXX-XXXX	text	text	ddMMMyyyy	text	X	XX
A	XXX-XXXX	text	text	ddMMMyyyy	text	X	XX
A	XXX-XXXX	text	text	ddMMMyyyy	text	X	XX
B	XXX-XXXX	text	text	ddMMMyyyy	text	X	XX
B	XXX-XXXX	text	text	ddMMMyyyy	text	X	XX
B	XXX-XXXX	text	text	ddMMMyyyy	text	X	XX

Table HIV1. HIV-1 Infections Adjudication Timeline
Population: Infected Cohort (N = xxxx)

Trt	Initial Positive Sample Draw Date	Adjudication Completion Date	Total Process Length (Days)	Initial Positive Draw Date to Initial Positive Test Date (Days)	Initial Positive Test Date to Confirmation Draw Date (Days)	Confirmation Draw Date to Confirmation Test Date (Days)	Confirmation Test Date to Adjudication Posting (Days)	Adjudication Posting to Completion (Days)
A	ddMMMyyyy	ddMMMyyyy	xxx	xxx	xxx	xxx	xxx	xxx
A	ddMMMyyyy	ddMMMyyyy	xxx	xxx	xxx	xxx	xxx	xxx
.	.	.						
.	.	.						
.	.	.						
B	ddMMMyyyy	ddMMMyyyy	xxx	xxx	xxx	xxx	xxx	xxx
		Median (Min, Max):	**xx (xx, xxx)**	**xx (xx, xx)**	**xx (xx, xx)**	**xx (xx, xx)**	**xx (xx, xxx)**	**xx (xx, xx)**

Records are listed in the order of the first positive sample draw date within each treatment group.

Table HIV2.1 HIV-1 Infection Listing
Population: Infected Cohort (N = xxxx)

Cohort[1]	Trt Group	Publ. ID	Number of Vaccinations Received	Initial Positive Sample Draw Date	Initial Pos. Visit	Initial Pos. Study Week	Adjudication Date	Retrospective Testing Results Required	Retrospective Pos. at Enrollment
Non-MITT	A	xxx-xxxx	xxx	ddMMMyyyy	xxx	xxx	ddMMMyyyy	Yes	Yes
		xxx-xxxx	xxx	ddMMMyyyy	xxx	xxx	ddMMMyyyy	Yes	Yes
.				.			.		
	B	xxx-xxxx	xxx	ddMMMyyyy	xxx	xxx	ddMMMyyyy	No	No
		xxx-xxxx	xxx	ddMMMyyyy	xxx	xxx	ddMMMyyyy	No	No
.									
.									
MITT	A	xxx-xxxx	xxx	ddMMMyyyy	xxx	xxx	ddMMMyyyy	No	No
		xxx-xxxx	xxx	ddMMMyyyy	xxx	xxx	ddMMMyyyy	No	No
.				.			.		
	B	xxx-xxxx	xxx	ddMMMyyyy	xxx	xxx	ddMMMyyyy	No	No
		xxx-xxxx	xxx	ddMMMyyyy	xxx	xxx	ddMMMyyyy	No	No
.									
.									
.									

Records are listed in the order of the first positive sample draw date within each cohort and treatment combination.

Table HIV2.2 HIV-1 Infection Listing
Population: Per Protocol Infected Cohort (N = xxxx)

Trt Group	Publ. ID	Number of Vaccinations Received	Initial Positive Sample Draw Date	Initial Pos. Visit	Initial Pos. Study Week	Adjudication Date	Retrospective Testing Results Required	Retrospective Pos. at Enrollment
A	xxx-xxxx	xxx	ddMMMyyyy	xxx	xxx	ddMMMyyyy	Yes	Yes
	xxx-xxxx	xxx	ddMMMyyyy	xxx	xxx	ddMMMyyyy	Yes	Yes
			.			.		
B	xxx-xxxx	xxx	ddMMMyyyy	xxx	xxx	ddMMMyyyy	No	No
	xxx-xxxx	xxx	ddMMMyyyy	xxx	xxx	ddMMMyyyy	No	No
A	xxx-xxxx	xxx	ddMMMyyyy	xxx	xxx	ddMMMyyyy	No	No
	xxx-xxxx	xxx	ddMMMyyyy	xxx	xxx	ddMMMyyyy	No	No
			.			.		
B	xxx-xxxx	xxx	ddMMMyyyy	xxx	xxx	ddMMMyyyy	No	No
	xxx-xxxx	xxx	ddMMMyyyy	xxx	xxx	ddMMMyyyy	No	No

Records are listed in the order of the first positive sample draw date within each cohort and treatment combination.

Figure EFF1.1. Cumulative HIV Incidence Among Participants by Treatment Arm
Population: Per Protocol (N = xxxx)

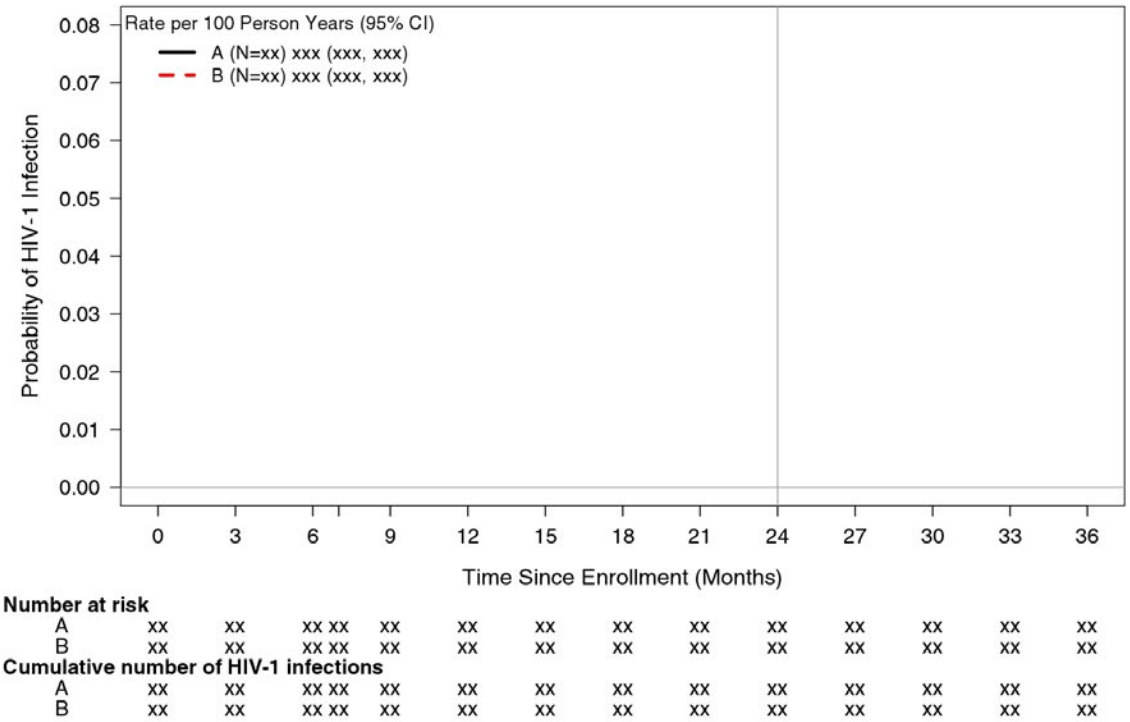

Figure EFF1.2. Cumulative HIV Incidence Among Participants by Treatment Arm
Population: MITT (N = xxxx)

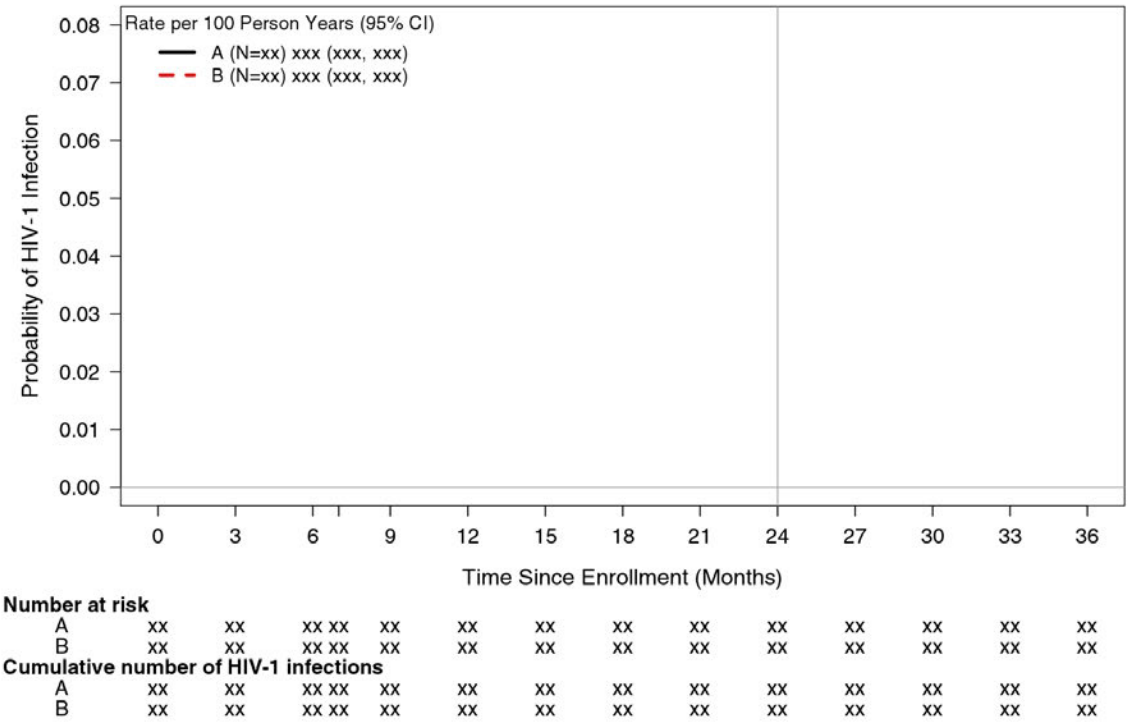

Figure EFF2.1. Vaccine Efficacy Over Time
Population: Per Protocol (N = xxxx)

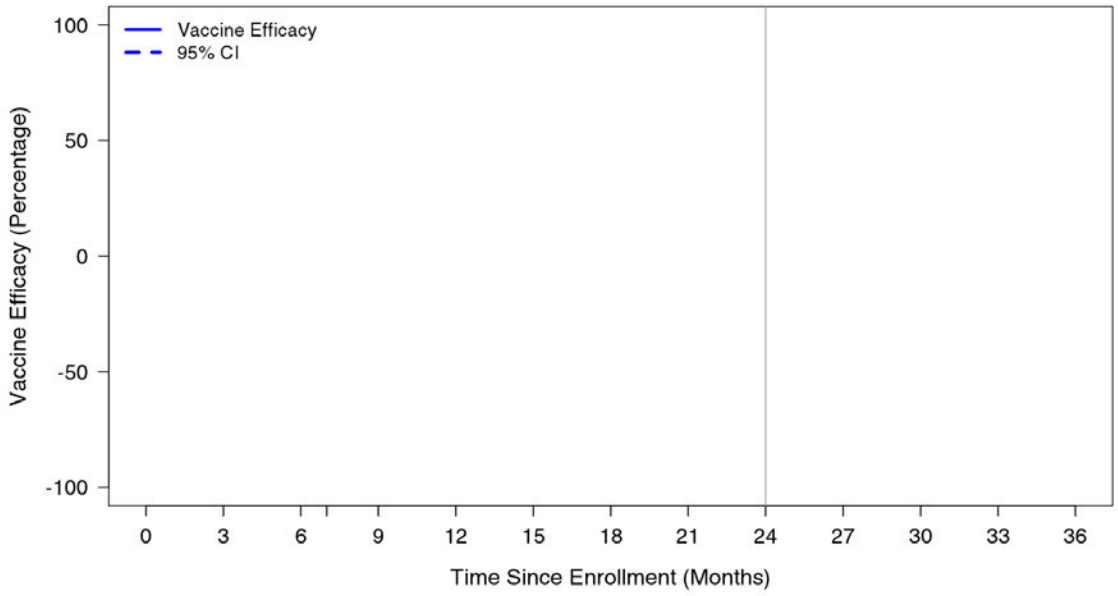

Figure EFF2.2. Vaccine Efficacy Over Time
Population: MITT (N = xxxx)

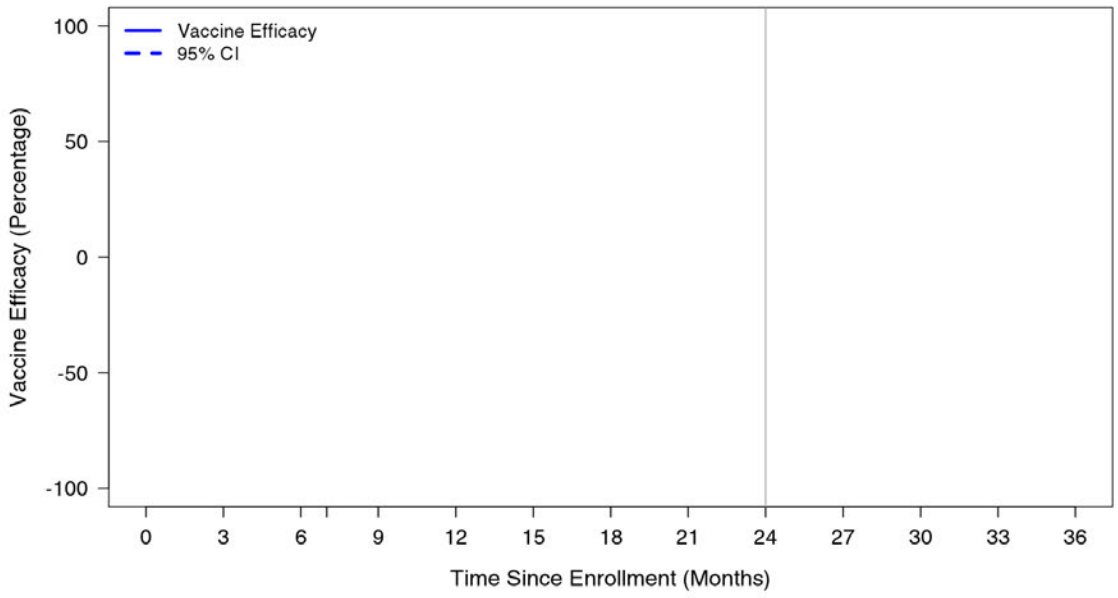

Table EFF1.1. Rate of HIV Infection and Vaccine Efficacy, According to Age and BMI Groups
Population: Per Protocol (N = xxxx)

	A				**B**				**Vaccine Efficacy**
	# Evaluated	# Inf.	# PY	Rate/PY	# Evaluated	# Inf.	# PY	Rate/PY	% (95% CI)
All subjects	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx (xxx, xxx)
Age Group									
18 – 20	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx (xxx, xxx)
21 – 25	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx (xxx, xxx)
26 – 35	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx (xxx, xxx)
BMI Group									
<18.5	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx (xxx, xxx)
18.5-25	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx (xxx, xxx)
≥25	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx (xxx, xxx)

Table EFF1.2. Rate of HIV Infection and Vaccine Efficacy, According to Age and BMI Groups
Population: MITT (N=xxxx)

	A				**B**				**Vaccine Efficacy**
	# Evaluated	# Inf.	# PY	Rate/PY	# Evaluated	# Inf.	# PY	Rate/PY	% (95% CI)
All subjects	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx (xxx, xxx)
Age Group									
18 – 20	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx (xxx, xxx)
21 – 25	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx (xxx, xxx)
26 – 35	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx (xxx, xxx)
BMI Group									
<18.5	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx (xxx, xxx)
18.5-25	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx (xxx, xxx)
≥25	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx (xxx, xxx)

References

Bernardo, J. M. and Smith, A. F. M. (2000), *Bayesian Theory*, Wiley Series in Probability and Statistics.

Castillo-Mancilla, J. R., Zheng, J.-H., Rower, J. E., Meditz, A., Gardner, E. M., Predhomme, J., Fernandez, C., Langness, J., Kiser, J. J., Bushman, L. R., and Anderson, P. L. (2013), "Tenofovir, Emtricitabine, and Tenofovir Diphosphate in Dried Blood Spots for Determining Recent and Cumulative Drug Exposure," *AIDS Research and Human Retroviruses*, 29, 384–390.

Follmann, D. and Huang, C.-Y. (2015), "Incorporating founder virus information in vaccine field trials," *Biometrics*, 71, 386–396.

Freidlin, B., Korn, E. L., and Gray, R. (2010), "A general inefficacy interim monitoring rule for randomized clinical trials," *Clinical Trials*, 7, 197–208.

Gilbert, P., Grove, D., Gabriel, E., Huang, Y., Gray, G., Hammer, S., et al. (2011), "A sequential Phase 2b trial design for evaluating vaccine efficacy and immune correlates for multiple HIV vaccine regimens," *Statistical Communications in Infectious Diseases*, 3.

Hammer, S. M., Sobieszczyk, M. E., Janes, H., Karuna, S. T., Mulligan, M. J., Grove, D., Koblin, B. A., Buchbinder, S. P., Keefer, M. C., Tomaras, G. D., Frahm, N., Hural, J., Anude, C., Graham, B. S., Enama, M. E., Adams, E., DeJesus, E., Novak, R. M., Frank, I., Bentley, C., Ramirez, S., Fu, R., Koup, R. A., Mascola, J. R., Nabel, G. J., Montefiori, D. C., Kublin, J., McElrath, M. J., Corey, L., Gilbert, P. B., and Team, H. . S. (2013), "Efficacy Trial of a DNA/rAd5 HIV-1 Preventive Vaccine," *New England Journal of Medicine*, 369, 2083–2092.

Heyse, J. F., Kuter, B. J., Dallas, M. J., Heaton, P., and Team, R. S. (2008), "Evaluating the safety of a rotavirus vaccine: the REST of the story," *Clinical Trials*, 5, 131–139.

Hubbard, A. E., Kherad-Pajouh, S., and van der Laan, M. J. (2016), "Statistical Inference for Data Adaptive Target Parameters," *International Journal of Biostatistics*, 12, 3–19.

Parzen, M., Wei, L., and Ying, Z. (1997), "Simultaneous confidence intervals for the difference of two survival functions," *Scandinavian Journal of Statistics*, 24, 309–314.

Rose, S. and van der Laan, M. J. (2011), *Targeted Learning: Causal Inference for Observational and Experimental Data*, Springer Series in Statistics.

Schmidli, H., Gsteiger, S., Roychoudhury, S., O'Hagan, A., Spiegelhalter, D., and Neuenschwander, B. (2014), "Robust Meta-Analytic-Predictive Priors in Clinical Trials with Historical Control Information," *Biometrics*, 70, 1023–1032.

van der Laan, M. J., Polley, E. C., and Hubbard, A. E. (2007), "Super learner." *Statistical Applications in Genetics and Molecular Biology*, 6, 1–23.

Westling, T., Luedtke, A., Gilbert, P., and Carone, M. (2021), "Inference for treatment-specific survival curves using machine learning," *arXiv preprint arXiv:2106.06602; under revision at J Am Stat Assoc.*