HIV VACCINE TRIALS NETWORK

SCHARP at FRED HUTCH

STATISTICAL ANALYSIS PLAN

Protocol HVTN 120

Protocol Version 3.0

A phase 1/2a clinical trial to evaluate the safety and immunogenicity of ALVAC-HIV (vCP2438) and of MF59®- or AS01B-adjuvanted clade C Env protein, in healthy, HIV-uninfected adult participants

March 24, 2022

Version 1.3

Prepared by

Protocol Statisticians Zonglin He, PhD
Allan deCamp, PhD

FRED HUTCH CURES START HERE™

Approval Signature Page

HVTN 120

Statistical Analysis Plan

A phase 1/2a clinical trial to evaluate the safety and immunogenicity of ALVAC-HIV (vCP2438) and of MF59®- or AS01B-adjuvanted clade C Env protein, in healthy, HIV-uninfected adult participants

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

See email signature attached_____ See email signature attached

Allan deCamp, PhD Date

Staff Scientist

Fred Hutchinson Cancer Research Center

See email signature attached_____ See email signature attached

Zonglin He, PhD Date

Statistical Research Associate

Fred Hutchinson Cancer Research Center, SCHARP

Table of Contents

1 INTRODUCTION

The following describes the Statistical Analysis Plan (SAP) for the analysis of data from HVTN 120 for Safety Monitoring Board (SMB) reports, the Final Study Report (FSR) for Safety, Protocol Team (PT) reports for immunogenicity data, and the FSR for Immunogenicity. As detailed in SCHARP SOP-0013, Revision 6 (effective date: May 1, 2018), this SAP is required prior to the first analysis and must be approved by the lead protocol statisticians. SMB reporting begins shortly after enrollment opens, and subsequent revisions are expected to describe analysis of immunogenicity data. The SAP will be reviewed prior to the first SMB report and before the final analysis with all major revisions of the plan archived.

1.1. General Design Considerations

Protocol Title

A phase 1/2a clinical trial to evaluate the safety and immunogenicity of ALVAC-HIV (vCP2438) and of MF59®- or AS01B-adjuvanted clade C Env protein, in healthy, HIV-uninfected adult participants

Study products and routes of administration

- **ALVAC-HIV (vCP2438)** expresses the gene products 96ZM651 *gp120* (clade C strain) linked to the sequences encoding the HIV-1 transmembrane anchor (TM) sequence of *gp41* (28 amino acids clade B LAI strain) and *gag* and *pro* (clade B LAI strain). It is formulated as a lyophilized vaccine for injection at a viral titer $\geq 1 \times 10^6$ cell culture infectious dose $(CCID)_{50}$ and $< 1 \times 10^8$ $CCID_{50}$ (nominal dose of 10^7 $CCID_{50}$) and is reconstituted with 1 mL of sterile sodium chloride solution (NaCl 0.4%), administered IM as a single 1 mL dose.

- **Protein/MF59**: Bivalent Subtype C gp120/MF59: clade C TV1.C gp120 Env and clade C 1086.C gp120 Env, each at a dose of 100 mcg, mixed with MF59 adjuvant, administered IM as a single 0.5 mL dose.

- **Protein/$AS01_B$**: Bivalent Subtype C gp120/$AS01_B$: clade C TV1.C gp120 Env and clade C 1086.C gp120 Env, each at a dose of 20 mcg or 100 mcg, mixed with $AS01_B$ adjuvant, administered IM as a single 0.75 mL dose.

- **Placebo**: Sodium Chloride for Injection, 0.9%, administered IM.

Table 3-1 Schema

Group	N	Dose of each protein	Deltoid	Month 0 (Day 0)	Month 1 (Day 28)	Month 3 (Day 84)	Month 6 (Day 168)
1	50	100 mcg	Left	ALVAC-HIV	ALVAC-HIV	ALVAC-HIV	ALVAC-HIV
			Right	-	-	Protein/MF59 + Placebo*	Protein/MF59 + Placebo*
2	50	100 mcg	Left	ALVAC-HIV	ALVAC-HIV	ALVAC-HIV	ALVAC-HIV
			Right	-	-	Protein/AS01$_B$ + Placebo*	Protein/AS01$_B$ + Placebo*
3	50	20 mcg	Left	ALVAC-HIV	ALVAC-HIV	ALVAC-HIV	ALVAC-HIV
			Right	-	-	Protein/AS01$_B$ + Placebo*	Protein/AS01$_B$ + Placebo*
4	10	N/A	Left	Placebo	Placebo	Placebo	Placebo
			Right	-	-	Placebo + Placebo*	Placebo + Placebo*

* Two distinct placebo preparations for protein/adjuvant will be needed to maintain the blind since Protein/AS01$_B$ and Protein/MF59 consist of different injection volumes.

Design
Multicenter, randomized, controlled, double-blind trial

Duration per participant
12 months of scheduled clinic visits

Estimated total study duration
18 months (includes enrollment and follow-up)

Investigational New Drug (IND) study sponsor
DAIDS, NIAID, NIH, DHHS (Bethesda, Maryland, USA)

Study Locations
Southern Africa and United States

Study Chair
Z Mike Chirenje
UZ-UCSF Collaborative Research Program

1.2. Study Objectives and Endpoints

Primary objectives and endpoints

Primary objective 1

- To evaluate the safety and tolerability of ALVAC-HIV and bivalent gp120 protein/MF59 or bivalent gp120 protein/$AS01_B$.

Primary endpoints 1

- Severe local and systemic reactogenicity signs and symptoms (pain, tenderness, erythema, induration, fever, malaise/fatigue, myalgia, headache, nausea, vomiting, chills, arthralgia) up to 7 days after each vaccine dose

- AEs by body system, Medical Dictionary for Regulatory Activities (MedDRA) preferred term, severity, and assessed relationship to study products up to 30 days after each vaccine dose

- SAEs, AESIs, and new chronic conditions (requiring medical intervention for ≥ 30 days) throughout the study

- Laboratory measures: white blood cells (WBC), neutrophils, lymphocytes, hemoglobin, platelets, alanine aminotransferase (ALT), aspartate aminotransferase (AST), alkaline phosphate (ALP), and creatinine at baseline and following vaccinations

- AEs leading to early participant withdrawal or early discontinuation of study products administration throughout the study.

Primary objective 2

- To compare HIV-specific CD4+ T-cell response rates at the month 6.5 timepoint (2 weeks after the fourth vaccination) of ALVAC-HIV and bivalent gp120 protein/MF59 to each of the bivalent gp120 protein/$AS01_B$ vaccine regimens.

Primary endpoint 2

- HIV-specific CD4+ T-cell response rates as assessed by flow cytometry.

Primary objective 3

- To compare HIV-specific Env-gp120 binding antibody response magnitudes at the month 12 timepoint (6 months after the fourth vaccination) of ALVAC-HIV and bivalent gp120 protein/MF59 to each of the bivalent gp120 protein/$AS01_B$ vaccine regimens.

Primary endpoint 3

- HIV-specific Env-gp120 binding antibody response magnitude as assessed by multiplex assay.

Secondary objectives and endpoints

Secondary objective 1

- To further evaluate the systemic immune responses and the durability of immunogenicity of each vaccine regimen at the month 6.5 and month 12 timepoints.

Secondary endpoint 1

- HIV-specific total IgG binding antibody response breadth and magnitude as assessed by multiplex assay.
- Anti –V1/V2 scaffold IgG binding antibody responses as assessed by multiplex assay.
- HIV-specific CD4+ and CD8+ T-cell responses as assessed by flow cytometry.
- Additional immunogenicity assays may be performed on blood samples based on the HVTN Laboratory Assay Algorithm

Exploratory objectives

Exploratory objective 1

- To further evaluate immunogenicity of each vaccine regimen, additional immunogenicity assays may be performed on blood and optionally provided mucosal samples, including samples from other timepoints, based on the HVTN Laboratory Assay Algorithm.

Exploratory objective 2

- To assess whether the diversity of gut microbiome correlates with vaccine responses using optionally provided stool specimens.

Exploratory objective 3

- To conduct analyses related to furthering the understanding of HIV, immunology, vaccines, and clinical trial conduct.

1.3. Randomization

A participant's randomization assignment will be computer generated and provided to the HVTN CRS pharmacist through a Web-based randomization system. At each institution, the pharmacist with primary responsibility for dispensing study products is charged with maintaining security of the treatment assignments (except in emergency situations as specified in the HVTN MOP).

1.4. Blinding

Participants and site staff (except for site pharmacists) will be blinded as to participants' treatment group assignments. Study product assignments are accessible to those HVTN CRS

pharmacists, DAIDS protocol pharmacists and contract monitors, and SDMC staff who are required to know this information in order to ensure proper trial conduct. Any discussion of study product assignment between pharmacy staff and any other HVTN CRS staff is prohibited. The HVTN SMB members also are unblinded to treatment assignment in order to conduct review of trial safety.

When a participant leaves the trial prior to study completion, the participant will be told he or she must wait until all participants are unblinded to learn his or her treatment assignment.

In some cases, the CRS, PSRT, or study sponsor may believe unblinding of the site PI and participant would be appropriate to facilitate the clinical management of an AE or SAE. The HVTN Unblinding MOP specifies procedures for emergency unblinding, and for early unblinding for medical reasons.

1.5. Sample Size and Power

Safety

The goal of the safety evaluation for this study is to identify safety concerns associated with product administration. The ability of the study to detect SAEs can be expressed by the true event rate above which at least 1 SAE would likely be observed and the true event rate below which no events would likely be observed. Specifically, for each vaccine group of the study (n =100), there is a 90% chance of observing at least 1 event if the true rate of such an event is 2.3% or more; and there is a 90% chance of observing no events if the true rate is 0.1% or less.

Probabilities of observing 0, 1 or more, and 2 or more events among groups of size 100 are presented in Table 6-1 of the Protocol for a range of possible true adverse event rates. These calculations provide a more complete picture of the sensitivity of this study design to identify potential safety problems with a vaccine regimen.

An alternative way of describing the statistical properties of the study design is in terms of the 95% confidence interval (CI) for the true rate of an adverse event based on the observed data. Table 6-2 in the Protocol shows the 2-sided 95% CIs for the probability of an event based on a particular observed rate. Calculations are done using the score test method [1]. If none of the 100 participants receiving a vaccine regimen experience a safety event, the 95% 2-sided upper confidence bound for the true rate for such an event is 3.7%.

Immunogenicity

The primary immunogenicity evaluation is to compare Groups 2 and 3 (AS01B groups) separately to Group 1 (MF59 group) based on the following two criteria:

1. Establish superiority of the AS01B regimen in CD4+ T-cell response rate at month 6.5 for either TV1 or 1086; AND

2. Establish superiority of the AS01B regimen in magnitude of anti-gp120 antibody response at month 12 for 1086.

Power is calculated for each group comparison separately and assumes that:

1. The CD4+ T-cell response rate for the AS01B regimen is at least 20% higher for both antigens; AND,

2. The true geometric mean anti-gp120 binding antibody response is 1.5-fold higher for the AS01B regimen.

Power is based on data simulations using HVTN 100 CD4+ T-cell data from the month 6.5 visit (2 weeks after the fourth vaccination) and predicted anti-gp120 1086 antibody binding magnitude

at the month 12 visit. Simulated data sets were generated by sampling with replacement from the HVTN 100 data and power is reported as the percentage of the 10,000 simulated HVTN 120 trials that reject the composite null hypothesis defined by the primary immunogenicity evaluation ("Overall power") and separately for the null hypotheses defined by the CD4+ T-cell and the anti-gp120 antibody criteria. CD4+ T-cell response calls were simulated for the MF59 group (Group 1) by dichotomizing the ICS IL2/IFNγ CD4+ T-cell response magnitude using a threshold defined by the observed response rates of 49% to the TV1 antigen and 40% to the 1086 antigen. For the AS01B comparator group (Group 2 or 3), CD4+ T-cell response calls were simulated based on threshold corresponding to response rates of 69% for TV1 and 60% for 1086 antigens. The rationale for this approach is to maintain the between antigen correlation and allow simulation of a higher response rate in the AS01B group. Predicted HVTN 100 Env-gp120 binding data at month 12 were used for the second primary immunogenicity criterion. Env-gp120 binding magnitudes in the MF59 group (Group 1) at month 12 were predicted using HVTN 100 month 6.5 data and assuming the same antibody decline between 6.5 and 12 months that was obtained from modelling the decline of the antibody response to the A244 gp120 antigen in RV144 [76]. AS01B (Group 2 or 3) IgG response magnitudes were simulated on a logit scale using a fold-change equivalent to a 1.5-fold change between group geometric means on the MFI scale. Specifically, the logit transformation of MFI is defined by logit(MFI) = log((M-L)/L) – log(MFI/(M-MFI)) where $M=2^{15}$ and L=100, the upper and lower limits of detection of the binding assay multiplex assay (BAMA). The rationale for using the logit scale is to simulate MFI values within the range of the assay. Overall power, accounting for both primary immunogenicity criteria defined above, depends on inter-correlations between the CD4+ T-cell and anti-gp120 readouts. The correlation coefficient between month 6.5 TV1 and 1086 CD4+ Tcell magnitudes in HVTN 100 is high (0.86, Spearman rank correlation). Correlation coefficients between month 6.5 CD4+ T-cell magnitude (TV1 or 1086) and month 12 anti-gp120 1086 antibody magnitude are low (0.35 and 0.33, Spearman rank correlation). Although these correlations are difficult to interpret since they are based on predicted month 12 antibody responses, they are consistent with correlations between week 26 cellular and week 52 antibody responses from HVTN 096 and HVTN 205. Therefore, the results for overall power shown in Table 6-3 in the Protocol, are based on the correlation between month 6.5 CD4+ T-cell magnitude and predicted anti-gp120 1086 antibody magnitude. An alternative is to assume independence between the month 6.5 CD4+ T-cell responses and the month 12 anti-gp120 responses. Under this scenario the overall power is virtually the same (results not shown). Power calculations are based on 1-sided Fisher's exact tests for the TV1 and 1086 CD4+ T-cell response rate comparisons and a 1-sided t-test comparing log(MFI) for the anti-gp120 responses comparison. Comparisons are based on an alpha level of 0.025.

2 ANALYSIS COHORT DEFINITIONS

Since enrollment is concurrent with receiving the first study vaccination, all participants will provide some safety data. However, for immunogenicity analyses, it is possible that data may be missing for various reasons, such as participants terminating from the study early, problems in shipping specimens, low cell viability of processed peripheral blood mononuclear cells (PBMCs), or high assay background. Immunogenicity data from 17 phase 1 and 2 phase 2a HVTN vaccine trials, which began enrolling after June 2005 (data as of September 2014), indicate that 15% is a reasonable estimate for the rate of missing data at month 6.5. For this reason, sample size calculations account for 15 enrolled participants on each of the vaccine groups having missing data for the primary immunogenicity endpoints.

3 POTENTIAL CONFOUNDERS

Characterization of the safety of the vaccine is susceptible to confounding by adverse events not related to the vaccine that by chance occur more often in one arm of the trial than another. Therefore analyses involving adverse events will incorporate the reported relationship to product as assessed by the site clinician.

4 STATISTICAL ANALYSIS

This section describes the final study analysis, unblinded as to treatment group assignment. All data from enrolled participants will be analyzed according to the initial randomization assignment regardless of how many vaccinations they received. In the rare instance that a participant were to receive the wrong treatment at a vaccination time, a decision will be made as to how to report their safety data based on the nature of the wrong treatment and the number of correct and incorrect study injections the participant received; a revised SAP would document such a decision. Analyses are modified intent-to-treat in that individuals who are randomized but not enrolled do not contribute data and hence are excluded. Because of blinding and the brief length of time between randomization and enrollment—typically no more than 4 working days—very few such individuals are expected.

Analyses for primary endpoints will be performed using SAS and R. All other descriptive and inferential statistical analyses will be performed using SAS, StatXact, or R statistical software.

No formal multiple comparison adjustments will be employed for multiple safety endpoints, multiple primary immunogenicity endpoints, or secondary endpoints. However, multiplicity adjustments will be made for certain immunogenicity assays, as specified in the SAP, when the assay endpoint is viewed as a collection of hypotheses (e.g., testing multiple peptide pools to determine a positive response).

Immunogenicity data from this study may be combined with other phase 1/2a studies within the P5 partnership HIV vaccine program. Comparable eligibility criteria and validated assays for primary immunogenicity endpoints will be used to mitigate the potential bias introduced by combining data across studies conducted over an extended period of time.

4.1 Analysis variables

The analysis variables consist of baseline participant characteristics, safety, and immunogenicity for primary and secondary objective analyses.

4.2 Baseline comparability

Treatment groups will be compared for baseline participant characteristics using descriptive statistics.

4.3 Safety/tolerability analysis

Since enrollment is concurrent with receiving the first vaccination, all participants will have received at least 1 vaccination and therefore will provide some safety data.

4.3.1 Reactogenicity

The number and percentage of participants experiencing each type of reactogenicity sign or symptom will be tabulated by severity and treatment group and the percentages displayed graphically by group. For a given sign or symptom, each participant's reactogenicity will be counted once under the maximum severity for all injection visits. 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 groups.

4.3.2 AEs and SAEs

AEs will be summarized using MedDRA System Organ Class and preferred terms. 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 or 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 or the strongest recorded causal relationship to study product. Formal statistical testing comparing groups is not planned since interpretation of differences must rely heavily upon clinical judgment.

A listing of SAEs reported to the DAIDS Regulatory Support Center (RSC) Safety Office will provide details of the events including severity, relationship to study product, time between onset and last vaccination, and number of vaccinations received. A separate listing will do the same for AEs of special interest (AESI). AESI for this protocol include but are not limited to potential immune-mediated disorders; a sample list of AESI is provided in Appendix H of the Protocol. These listings will be submitted to the FDA in all annual reports and clinical trial reports.

4.3.3 Local laboratory values

Boxplots of local laboratory values will be generated for baseline values and for values measured during the course of the study by treatment group and visit. Each boxplot will show the first quartile, the median, and the third quartile. Outliers (values outside the boxplot) will also be plotted. If appropriate, horizontal lines representing boundaries for abnormal values will be plotted.

For each local laboratory measure, summary statistics will be presented by treatment group and timepoint, as well as changes from baseline for postenrollment values. In addition, the number (percentage) of participants with local laboratory values recorded as meeting Grade 1 AE criteria or above as specified in the DAIDS AE Grading Table will be tabulated by treatment group for each postvaccination timepoint. Reportable clinical laboratory abnormalities without an associated clinical diagnosis will also be included in the tabulation of AEs described above.

4.3.4 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 reason and treatment group.

4.4 Immunogenicity Analysis

4.4.1 General approach

For the statistical analysis of immunogenicity endpoints, data from enrolled participants will be used according to the initial randomization assignment regardless of how many injections they received. Additional analyses may be performed, limited to participants who received all scheduled injections per protocol. Assay results that are unreliable, from specimens collected outside of the visit window, or from HIV-infected participants after infection are excluded. Since the exact date of HIV infection is unknown, any assay data from blood draws 4 weeks prior to an infected participant's last seronegative sample and thereafter may be excluded. If an HIV-infected participant does not have a seronegative sample after enrollment, then all data from that participant may be excluded from the analysis.

Response rates will be analyzed by tabulating the frequency of positive response for each endpoint and treatment group at each timepoint for which an assessment is performed. For CD4+

and CD8+ T-cell response, response rates to the individual peptide pools will also be calculated. Crude response rates will be presented with their corresponding 95% confidence interval estimate calculated using the score test method [1]. For the primary endpoints of CD4+ T-cell response rates for TV1 and 1086, differences between groups will be tested with a 2-sided Barnard's exact test at an alpha level of 0.05. For secondary and exploratory assay endpoints, response rates and 95% confidence intervals will be calculated if appropriate for the endpoint. No adjustment will be made to the vaccine group estimates for the false positive rates in the placebo group.

The primary endpoint of anti-gp120 binding antibody response magnitude, and likely some of the secondary assays, will have quantitative assay data. Other quantitative measures include binding antibody magnitude from the multiplex assay, neutralizing antibody titers, area under the magnitude-breadth curve [AUC-MB] for the neutralizing antibody assay, and percentage of positive cells from the ICS assay. Quantitative data will be displayed as graphical and tabular summaries of the distributions by antigen, treatment group, and timepoint. For the primary and secondary immunogenicity endpoints, box plots and plots of estimated reverse cumulative distribution curves will be presented by group. For the primary endpoint, differences between groups will be tested with a nonparametric Wilcoxon rank sum test if the data are not normally distributed or with a 2-sample t-test if the data appear to be normally distributed, at an alpha level of 0.05.

Formal statistical comparisons between groups for secondary and exploratory immunogenicity endpoints are not objectives of the trial, although these are likely to be made to better understand the effect of each vaccine regimen. For comparisons in which the response rate for 1 of the groups is low (eg, $\leq 20\%$ for the class), statistical testing will use Barnard's exact test comparing the 2 response rates as most of the continuous data readouts would be left censored at the lower limit of detection. For comparisons in which the response rates for both groups are high (eg, $\geq 75\%$), the difference between groups will be tested using the continuous readouts with a nonparametric Wilcoxon rank sum test if the data are not normally distributed and with a 2-sample t-test if the data appear to be normally distributed.

Some immunologic assays have underlying continuous or count-type readout that are dichotomized into responder/nonresponder categories. If treatment group differences for these assays are best summarized by a mixture model, then either Lachenbruch's test statistic [2] will be used to evaluate the composite null hypothesis of equal response rates in the 2 groups and equal response distributions among responders in the 2 such groups. For estimation, differences in response rates between groups will be estimated using the methods described above, and in the subgroup of positive responders, differences in location parameters between groups will be estimated using the methods described above.

4.4.2 Missing data considerations

Based upon previous HVTN trials, missing 15% of immunogenicity results for a specific assay is common due to study participants terminating from the study early, problems in shipping specimens, or low cell viability of processed PBMCs. To achieve unbiased statistical estimation and inferences with standard methods applied in a complete-case manner (only including participants with observed data in the analysis), missing data need to be missing completely at random (MCAR). MCAR assumes that the probability of an observation being missing does not depend on any participant characteristics (observed or unobserved). When missing data are minimal (specifically, if no more than 20% of participants are missing values), then standard complete-case methods will be used, because violations of the MCAR assumption will have little impact on the estimates.

If a substantial amount of immunogenicity data are missing (at least 1 value missing from more than 20% of participants), then using the methods that require the MCAR assumption may give

misleading results. In this situation, analyses of the immunogenicity endpoints at a specific timepoint will be performed using parametric generalized linear models fit by maximum likelihood. These methods provide unbiased estimation and inferences under the parametric modeling assumptions and the assumption that the missing data are missing at random (MAR). MAR assumes that the probability of an observation being missing may depend upon the observed responses and upon observed covariates, but not upon any unobserved factors. Generalized linear models for response rates will use a binomial error distribution and for quantitative endpoints, a normal error distribution. For assessing repeated immunogenicity measurement, linear mixed effects models will be used. If the immunological outcomes are left- and/or right- censored, then the linear mixed effects models of Hughes [3] will be used, because they accommodate the censoring. In addition, secondary analyses of repeated immunogenicity measurements may be done using weighted GEE [4] methods, which are valid under MAR. All of the models described above will include as covariates all available baseline predictors of the missing outcomes.

4.5 Analyses prior to end of scheduled follow-up visits

Any analyses conducted prior to the end of the scheduled follow-up visits should not compromise the integrity of the trial in terms of participant retention or safety or immunogenicity endpoint assessments. In particular, early unblinded analyses by treatment assignment require careful consideration and should be made available on a need to know basis only. .

4.5.1 Safety

During the course of the trial, unblinded analyses of safety data will be prepared approximately every 4 months during the study for review by the HVTN SMB. Ad hoc safety reports may also be prepared for SMB review at the request of the HVTN 120 PSRT. Refer to the process described in the HVTN unblinding MOP any requests for unblinded safety data prior to the end of the scheduled follow-up visits.

4.5.2 Immunogenicity

The unblinded analysis of the primary endpoint of CD4+ T-cell response rates measured at month 6.5 will be conducted when all participants have completed the visit and all samples have been analyzed. Unblinded results will not be made publicly available until participants have completed their month 12 visit. Unblinded data analyses of secondary and exploratory endpoints (ie, additional immunogenicity assays) measured at month 6.5 will occur after the unblinded CD4+ T-cell immunogenicity analysis since selection of secondary assays is dependent upon the primary analysis. Analysis of secondary and exploratory objectives for month 6.5 will take place after the primary analysis and may be performed when assay data are available for analysis from at least 80% of participants.

Analysis of anti-gp120 binding antibody response magnitudes measured at month 12 will be conducted when all participants have completed the visit and all samples have been analyzed. Unblinded analysis of secondary and exploratory endpoints from the month 12 timepoint will take place after the primary analysis of anti-gp120 binding antibody and may be performed when assay data are available for analysis from at least 80% of participants.

The Laboratory Program will review analysis reports prior to distribution to the protocol chairs, DAIDS, vaccine developer, and other key HVTN members and investigators. Distribution of reports will be limited to those with a need to know for the purpose of informing decisions related to future trials. The HVTN leadership must approve any other requests for HVTN immunogenicity analyses prior to the end of the scheduled follow-up visits.

5 SAFETY TABLES AND FIGURES

5.1 List of Tables (see mock tables in Appendix A)

SMB reports and Safety FSRs include the following tables.

- Enrollment Report
- Demographics and Study Product Administration Frequencies
- Overall Protocol Status
- Maximum Local Reactogenicity Summary
- Maximum Systemic Reactogenicity Summary
- Adverse Experiences by Body System and Severity – By Decreasing Frequency
- Adverse Experiences by Preferred Term and Severity – By Decreasing Frequency – Includes Severe, Potentially Life-threatening or Fatal Events Only
- Adverse Experiences by Preferred Term and Severity – By Decreasing Frequency – Includes Events of All Severities
- Adverse Experiences by Preferred Term and Relationship to Study Product – By Decreasing Frequency – Includes Related Events Only
- Adverse Experiences by Preferred Term and Relationship to Study Product – By Decreasing Frequency – Includes Events of Any Relationship
- Expedited Adverse Events (EAEs) Reported to the Regulatory Support Center (RSC)
- Pregnancy Listing
- Adverse Events of Special Interest

Safety FSRs include the following additional tables.

- Social Impact Summary
- Local Lab Value Summary Statistics
- Local Laboratory Values Meeting Grade 1 AE Criteria or Above

5.2 Participant Listings

The following listings of participant-level data are included in the SMB reports.

- Discontinuations
- Pregnancies
- Severe or Life-Threatening Local and Systemic Reactogenicities
- Moderate or Severe Erythema and Induration
- Expedited Adverse Experiences (EAEs)
- Adverse Experiences of Special Interest (AESIs)
- Severe, Life-Threatening, or Fatal Adverse Experiences
- Study Product Administration Errors

- HIV Infection Results from Lab and Reported by Site

5.3 List of Graphs

- Maximum Local Reactogenicities
- Maximum Systemic Reactogenicities
- Boxplots of laboratory values for white blood cells (WBC), neutrophils, lymphocytes, hemoglobin, platelets, alanine aminotransferase (ALT), aspartate aminotransferase (AST), alkaline phosphate (ALP), and creatinine at baseline and following vaccinations

6 IMMUNOGENICITY TABLES AND FIGURES, BY ASSAY

6.1 Intracellular Cytokine Staining

Flow cytometry will be used to examine vaccine specific CD4+ and CD8+ T cell responses following stimulation of PBMCs with synthetic HIV peptides that span the proteins encoded by the vaccine. Vaccine specific T-cell response magnitudes and response rates will be assessed at the Month 6.5 timepoint/Visit 10 (2 weeks after the fourth vaccination) and the Month 12 timepoint/Visit 12 (6 months after the fourth vaccination). Visit 10 analyses will be conducted first. Afterwards Visit 12 analyses will be conducted and combined with the Visit 10 analyses, once Visit 12 specimens become available. Contingent on the results of the Visit 10 and Visit 12 analyses, additional reports may be generated for Visit 7. The following peptide pools will be used:

Peptide Pool	Pool ID	Lot Number
LAI-Gag	PP500025	11S227
LAI gp41 TM	PP502556	15H008AB
Env-1-ZM96	PP502442	14C093A
Env-2-ZM96	PP502446	14C097A
TV1 gp120	PP502557	15H020A
1086 gp120	PP502551	15H015A

The following functional markers will be measured from a 17-colormetric panel: AViD, CCR7, CD14, CD56, CD45RA, CXCR5, ICOS (CD278), PD-1 (CD279), CD3, CD4, CD8, CD154, IFN-γ, Granzyme B, IL-2, IL-4, IL-17a, and TNFα. The initial reports will present IL-2 and IFN-γ responses. Subsequent reports may be produced to include data from the other functional markers.

All analyses are performed on LAI-Gag, LAI-Gag gp41 TM, Env-1-ZM96, Env-2-ZM96, TV1 gp120 Env, and 1086 gp120 Env, individually and according to 2 levels of pooling:

1. Env-ZM96: Sum of Env-1-ZM96 and Env-2-ZM96
2. Any Env: the maximum of Env ZM96, 1086 gp120, and TV1 gp120, where Env ZM96 is the sum of Env-1-ZM96 and Env-2-ZM96; 3

3. Any HIV: the sum of Any Env, LAI-Gag gp 41 TM and LAI-Gag.

Positivity calls and subsequent analyses are performed separately for CD4+ and CD8+ T cell responses. Fisher's exact test criteria will be used for positivity calls before assays on all samples are complete. After, MIMOSA criteria will be used for positivity calls. If at least one peptide pool for a specific HIV-1 gene is positive, then the overall response to the gene is considered positive. If any peptide pool is positive, then the overall response to HIV-1 (Any HIV) for that T-cell subset is considered positive.

Positive response rates are compared between treatment groups using Barnard's exact test. Response magnitudes are compared between treatment groups using Wilcoxon rank sum test. Positive response rates are compared within treatment group, across timepoints using McNemar's test. Response magnitudes are compared within treatment group, across timepoints using the Wilcoxon signed rank test. Statistical tests with an unadjusted p-value ≤ 0.05 will be considered statistically significant.

The following comparisons of the immune response rate and magnitude will be performed:

- Pairwise comparisons of response rates and response magnitudes between all treatment groups.
- Pairwise comparisons of response rates and response magnitudes within treatment group, between the Visit 10 and Visit 12 timepoints.
- If additional assays are performed at Visit 7, then between time point comparisons, within treatment group, will be performed between Visit 7 and Visit 10 and between Visit 7 and Visit 12.

6.1.1 List of Tables

- Response rate table by lab, T cell subset, peptide pool, visit, day, and group/treatment arm for cells expressing either IL-2 and/or IFN-γ using Fisher's exact test criteria and/or MIMOSA criteria.
- Summary statistics (i.e., min, mean, median, max) of response magnitudes among responders for T-cell subset, peptide pool, visit, and group/treatment arm.
- Summary statistics (i.e., min, mean, median, max) of response magnitudes among all participants (positive and negative responders) for T-cell subset, peptide pool, visit, and group/treatment arm.
- Response rate and/or response magnitude comparisons of treatment arms as specified above.
- Response rate and/or response magnitude comparisons across time points as specified above.
- Listing of positive responders for cells expressing either IL-2 and/or IFN-γ based on the criteria used for the response rate table.

6.1.2 List of Graphs

- Boxplots of background-adjusted IFN-γ and/or IL-2 response magnitude by T cell subset, peptide pool, visit, and group/treatment arm using Fisher exact test and/or MIMOSA criteria. Graphs will be displayed with treatment groups side by side with one peptide pool, one T-cell subset and one visit per graph. Above the boxplots, concordant barplots of the response rates will be shown.
- Boxplots of background-adjusted IFN-γ and/or IL-2 response magnitude by T cell subset, HIV protein, visit, and group/treatment arm using Fisher exact test and/or MIMOSA criteria. Graphs will be displayed with treatment groups and visits side by side with one peptide pool, one T-cell subset per graph. Spaghetti plots that connect each participant response across visits will be superimposed on the boxplots. Above the boxplots, concordant barplots of the response rates will be shown.

6.2 Neutralizing Antibody

This section will be completed upon receipt of the lab study plan for the NAb assay.

6.3 Binding Antibody Multiplex Assay

Serum HIV-1-specific IgG antibody binding responses were measured at a 1:50 dilution against all antigens in the following table. Serum samples were tested at a starting 1:50 dilution followed by a 2-fold serial dilution for a total of six dilution points for antigens 00MSA 4076 gp140, 1086C_D7gp120.avi/293F, 96ZM651.D11gp120.avi, A1.con.env03 140 CF, Con 6 gp120/B, Con S gp140 CFI, and TV1c8_D11gp120.avi/293F at month 6.5. Serum IgG responses against specified antigens were measured on a Bio-Plex instrument (Bio-Rad) using a standardized custom HIV-1 Luminex assay (Tomaras et al. 2008; Tomaras et al. 2013; Haynes et al. 2012; Yates et al. 2014; Yates et al. 2018; Zolla-Pazner et al. 2014). The readout was background-subtracted mean fluorescence intensity (MFI), where background referred to a plate level control (i.e., a blank well run on each plate). Standard positive and negative controls were included in each assay to ensure specificity and for maintaining consistency and reproducibility between assays. The positive control includes purified polyclonal IgG from HIV subjects (HIVIG) using a 10-point standard curve (4PL fit) and CH58 V1V2 IgG (Liao et al. 2013). The negative controls were NHS (HIV-1 sero-negative human sera) and blank beads.

Several criteria were used to determine if data from an assay were acceptable and could be statistically analyzed. First, the blood draw date must have been within the allowable visit window as determined by the protocol. Post-infection samples from HIV-infected participants are excluded. Second, if the blank bead negative control exceeded 5,000 MFI, the sample was repeated. If the repeat value exceeded 5,000 MFI, the sample was excluded from analysis due to high background. The preset assay criteria for sample reporting were: coefficient of variation (CV) per duplicate values for each sample are $< 20\%$ and ≥ 50 beads counted per sample. To control for protein performance, the preset criteria include that the positive control titration in each assay must be within +/- 3 standard deviations of the mean for each antigen. The AUTC is calculated using the trapezoidal rule based on the raw MFI values truncated at zero across the log base 10 dilution per participant, isotype, antigen, and visit.

Binding Antibody Multiplex Assay (BAMA) testing will be performed on serum samples from each participant in the protocol at visit 2 (month 0, baseline), visit 10 (month 6.5, 2 weeks post the 4th vaccination), and visit 12 (month 12, 6 months post the 4th vaccination). Specimens from other timepoints as well as other HIV antigens and antibody isotypes may be assayed based on the results of the initial assay. If sufficient immunogenicity is observed for the primary analysis antigens, additional assays will be performed using the exploratory antigens.

The following table lists the antigens for the analysis.

Panel	Antigen	Clade	Endpoint
Vaccine gp120 panel	1086C_D7gp120.avi/293F	C	Primary objective 3
	96ZM651.D11gp120.avi	C	Primary objective 3
	TV1c8_D11gp120.avi/293F	C	Primary objective 3
Vaccine V1V2 panel	C.1086C_V1_V2 Tags	C	Secondary objective 1
	gp70-96ZM651.02 V1v2	C	Secondary objective 1
	gp70-TV1.21 V1V2	C	Secondary objective 1
	gp70-TV1.GSKvacV1V2/293F	C	Secondary objective 1
Other V1V2	gp70_B.CaseA_V1_V2	B	Exploratory objective
	gp70_B.CaseA2 V1/V2/169K	B	Exploratory objective
Concensus gp140/gp120	Con 6 gp120/B	M	Exploratory objective
	Con S gp140 CFI	M	Exploratory objective
Other gp140	00MSA 4076 gp140	A	Exploratory objective
	A1.con.env03 140 CF	A	Exploratory objective
gp41	gp41	B	Exploratory objective

MFI will be employed for this essay. The frequency of positive responders (% responders) will be tabulated to access HIV binding antibody responses between placebo and vaccine groups from MFI readouts.

Samples from post-enrollment visits are declared to have a positive response if they meet three conditions:

1. The Net MFI (MFI*, i.e., MFI minus Blank values, where 'Blank value' refers to a plate-specific background measure) were ≥ antigen-specific cutoff (based on the 95th percentile of the baseline samples).
2. The MFI* values were greater than 3 times the baseline MFI*.
3. The MFI values are greater than 3 times the baseline MFI values.

6.3.1 List of Tables

- The binding antibody positive response rates by isotype, antigen, visit and group
- Summary statistics of MFI* values (minimum, 25th percentile, median, 75th percentile, maximum, mean, standard deviation) among all participants, by isotype, antigen, visit number (visit month), and group
- Summary statistics of MFI* values (minimum, 25th percentile, median, 75th percentile, maximum, mean, standard deviation) among positive responders by isotype, antigen, visit number (visit month), and group.
- Summary statistics of binding antibody titers (AUTC) (minimum, 25th percentile, median, 75th percentile, maximum, mean, standard deviation) among all participants, by isotype, antigen, visit number (visit month), and group.
- Summary statistics of binding antibody titers (AUTC) (25th percentile, median, 75th percentile, maximum, mean, standard deviation) among positive responders, by isotype, antigen, visit number (visit month), and treatment group.

- Table of comparison of medians of AUTC between at the month 6.5 (2 weeks after the fourth vaccination) between T1 and each of T2 and T3 by isotype and the 3 antigens in Vaccine gp120 panel.
- Tables of comparison of medians of MFI* between at the month 12 (6 months after the fourth vaccination) between T1 and each of T2 and T3 by isotype and the 3 antigens in Vaccine gp120 panel.
- Listing of MFI* values among positive responders.

6.3.2 List of Figures

- Boxplots of MFI* values by isotype, antigen, visit, and group.
- Boxplots of binding antibody titers (AUTC) by isotype, antigen, visit, and group.

6.4 Antibody-Dependent Cell-mediated Cytotoxicity (Luciferase assay)

We utilized a modified version of a previously published ADCC luciferase procedure (Pollara et al. 2014). Briefly, CEM.NKRCCR5 cells (Trkola et al. 1999) were used as targets for ADCC luciferase assays after infection by one of the following HIV1 IMCs:

Complete IMC name	Accession Number	Abbreviated name	Vaccine Match
HIV Clade C TV1 gp120		HIV TV1	YES
HIV Clade C 96ZM651 gp120		HIV 96ZM651	YES
HIV 1086 B2 gp120		HIV 1086c	YES

Peripheral blood mononuclear cells (PBMCs) were obtained from a HIV-seronegative donor by leukapheresis and cryopreserved until the day of the assay. After thawing and overnight resting in RPMI 1640 supplemented with antibiotics, 10% fetal bovine serum (R10), and 10 ng/mL of IL-15, the PBMCs were used as effector cells at an effector-to-target ratio of 30:1.

Target and effector cells were plated in white 96-well half-area plates and co-cultured with 4-fold serial dilutions of trial participant serum starting at the 1:50 dilution. For each sample, percent specific killing was measured in duplicate at dilutions of 1:50, 1:200, 1:800, 1:3200, 1:12800, and 1:51200. Co-cultures were incubated for 6 hours at 37°C in 5% CO2. The final readout was the reduction of luminescence intensity generated by the presence of residual intact target cells that had not been lysed by the effector population in the presence of ADCC-mediating serum

antibodies. The percentage of killing was calculated using the formula: percent specific loss Luciferase activity = 100 * (RLU of target and effector well– RLU of test well)/(RLU of target and effector well).

In this analysis, the Relative Luminescence Units (RLU) of the target plus effector wells represents spontaneous lysis in the absence of any source of antibody and is used to calculate background activity. The monoclonal antibody (Synagis) and a cocktail of HIV1 monoclonal antibodies (A32, 2G12, CH44, and 7B2) were used as negative and positive controls, respectively.

6.4.1 List of Tables

- Response rate of peak baseline subtracted percent loss luciferase activity by treatment group and antigen
- Distribution of peak baseline subtracted percent loss luciferase activity among all participants by treatment group and antigen
- Distribution of peak baseline subtracted percent loss luciferase activity among positive responders by treatment group and antigen
- Distribution of pAUC baseline subtracted percent loss luciferase activity among all participants by treatment group and antigen
- Distribution of pAUC baseline subtracted percent loss luciferase activity among positive responders by treatment group and antigen
- Listing of positive responders
- Response rate comparison using Barnard's exact test
- Response magnitude comparison using Wilcoxon rank sum test
- Response AUC magnitude breadth comparison at Peak Time Point using Wilcoxon rank sum test
- Response AUC magnitude breadth comparison at Durability Time Point using Wilcoxon rank sum test

6.4.2 List of Figures

- ADCC Luciferase Response rate and magnitude by timepoint, antigen
- ADCC Luciferase Response AUC by timepoint, antigen
- ADCC Luciferase Magnitude-Breadth Response by timepoint

6.5 Antibody-Dependent Cell-mediated Cytotoxicity (GranToxiLux assay)

The qualified GranToxiLux Antibody-Dependent Cell-Mediated Cytotoxicity (GTL-ADCC) assay was performed as previously described (Pollara et al. 2011). Target cells were a clonal isolate of the CEM.NKRCCR5 CD4+ T-cell line (NIH AIDS Reagent Program, Division of AIDS, NIAID, NIH: from Dr. Alexandra Trkola (Trkola et al. 1999). These cells were coated with recombinant gp120s representing the HIV-1 envelopes of the subtype 1086_B2, 96ZM651, and TV1c8. Effector cells were PBMCs obtained from a HIV-seronegative donor with heterozygous for FcγR3A at position 158 (158F/V). PBMCs were obtained by leukapheresis to collect enough cells for completion of the study with a single donation, minimizing potential effector cell population variability effects on the study outcome. PBMCs were used at an effector cell to target cell ratio of 30:1. Serum samples were tested after five-fold serial

dilutions starting at 1:50. Each plate has one standardized positive control in duplicate and one standardized negative control in duplicate.

ADCC is quantified as net percent granzyme B activity, which is the percent of target cells positive for GTL (an indicator of granzyme B uptake) minus the percent of target cells positive for GTL when incubated with effector cells in the absence of a source of antibodies. Flow cytometry is used to quantify the frequency of granzyme B positive cells.

6.5.1 List of Tables

- peak activity response rates by antigen, timepoint, and treatment arm
- Distribution of peak activity, all participants
- Distribution of peak activity, positive responders
- Distribution of AUC, all participants
- Distribution of AUC, positive responders
- List of positive responders
- Response rate comparison using Barnard's exact test
- Response magnitude comparison using Wilcoxon rank sum test
- Response AUC magnitude breadth comparison at Peak Time Point using Wilcoxon rank sum test
- Response AUC magnitude breadth comparison at Durability Time Point using Wilcoxon rank sum test

6.5.2 List of Figures

- ADCC GranToxiLux Response rate and magnitude by timepoint, antigen
- ADCC GranToxiLux Response AUC by timepoint, antigen
- GranToxiLux Magnitude-Breadth Response by timpoint

7 REFERENCES

1. Agresti A, Coull BA. Approximate is better than "exact" for interval estimation of binomial proportions. Am Stat **1998**;52:119-26.
2. Lachenbruch PA. Comparisons of two-part models with competitors. Stat Med **2001**;20:1215-34.
3. Hughes JP. Mixed effects models with censored data with application to HIV RNA levels. Biometrics **1999**;55:625-9.
4. Rotnitzky A, Robins J. Analysis of semi-parametric regression models with non-ignorable non-response. Stat Med **1997**;16:81-102.

8 SAP Revision

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

SAP Version	Date	Modification
1.0	14 June 2018	Analysis plan for safety endpoints.
1.1	18 July 2019	Added analysis plan for ICS immunological endpoints.
1.2	26 April 2021	Added analysis plan for BAMA immunological endpoints.
1.3	24 March 2022	Added analysis plan for ADCC Luc and Flow endpoints.