Published in final edited form as:

Stat Commun Infect Dis. 2017 January ; 9(1): . doi:10.1515/scid-2016-0001.

Basis and Statistical Design of the Passive HIV-1 Antibody Mediated Prevention (AMP) Test-of-Concept Efficacy Trials

Peter B. Gilbert^{1,2,*}, Michal Juraska¹, Allan C. deCamp¹, Shelly Karuna¹, Srilatha Edupuganti³, Nyaradzo Mgodhi⁴, Deborah J. Donnell¹, Carter Bentley¹, Nirupama Sista⁵, Philip Andrew⁵, Abby Isaacs¹, Yunda Huang¹, Lily Zhang¹, Edmund Capparelli⁶, Nidhi Kochar¹, Jing Wang¹, Susan H. Eshleman⁷, Kenneth H. Mayer^{8,9,10}, Craig A. Magaret¹, John Hural¹, James G. Kublin¹, Glenda Gray^{11,1}, David C. Montefiori¹², Margarita M. Gomez¹³, David N. Burns¹⁴, Julie McElrath¹, Julie Ledgerwood¹⁵, Barney S. Graham¹⁵, John R. Mascola¹⁵, Myron Cohen^{16,¶}, and Lawrence Corey^{1,¶}

¹Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA

²Department of Biostatistics, University of Washington, Seattle, Washington, USA

³Department of Medicine, Emory University, Atlanta, Georgia, USA

⁴University of Zimbabwe – University of California San Francisco Research Program, Harare, Zimbabwe

⁵FHI 360 Durham, North Carolina, USA

⁶Department of Pediatrics, University of California, San Diego, California, USA

⁷Department of Pathology, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA

⁸The Fenway Institute, Boston, Massachusetts, USA

⁹Department of Psychiatry, Harvard Medical School, Boston, Massachusetts, USA

¹⁰Department of Medicine, Beth Israel Deaconess Medical Center, Boston, Massachusetts, USA

¹¹South African Medical Research Council, Cape Town, South Africa; Perinatal HIV Research Unit, University of the Witwatersrand, Braamfontein, Johannesburg, South Africa

¹²Duke University Medical Center, Durham, North Carolina, USA

¹³Vaccine Research Program, Division of AIDS, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland, USA

*Corresponding author: pgilbert@scharp.org (PBG).

¶These authors contributed equally to this work.

Conflict of interest statement: All authors have no potential conflicts of interest to declare.

Meetings where this research was presented: P.B.G. presented an early version of this research at the *Passive Immunization Trials to Inform Vaccine Design Workshop* at the Global HIV Vaccine Enterprise in August of 2014 in New York City, with talk title “Statistical issues in assessing treatment efficacy and correlates of protection in monoclonal antibody efficacy trials.”

¹⁴Prevention Sciences Program, Division of AIDS, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland, USA

¹⁵Vaccine Research Center, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland, USA

¹⁶Department of Medicine, University of North Carolina, Chapel Hill, North Carolina, USA

Abstract

Background—Anti-HIV-1 broadly neutralizing antibodies (bnAbs) have been developed as potential agents for prevention of HIV-1 infection. The HIV Vaccine Trials Network and the HIV Prevention Trials Network are conducting the Antibody Mediated Prevention (AMP) trials to assess whether, and how, intravenous infusion of the anti-CD4 binding site bnAb, VRC01, prevents HIV-1 infection. These are the first test-of-concept studies to assess HIV-1 bnAb prevention efficacy in humans.

Methods—The AMP trials are two parallel phase 2b HIV-1 prevention efficacy trials conducted in two cohorts: 2700 HIV-uninfected men and transgender persons who have sex with men in the United States, Peru, Brazil, and Switzerland; and 1500 HIV-uninfected sexually active women in seven countries in sub-Saharan Africa. Participants are randomized 1:1:1 to receive an intravenous infusion of 10 mg/kg VRC01, 30 mg/kg VRC01, or a control preparation every 8 weeks for a total of 10 infusions. Each trial is designed (1) to assess overall prevention efficacy (PE) pooled over the two VRC01 dose groups vs. control and (2) to assess VRC01 dose and laboratory markers as correlates of protection (CoPs) against overall and genotype- and phenotype-specific infection.

Results—Each AMP trial is designed to have 90% power to detect PE > 0% if PE is 60%. The AMP trials are also designed to identify VRC01 properties (i.e., concentration and effector functions) that correlate with protection and to provide insight into mechanistic CoPs. CoPs are assessed using data from breakthrough HIV-1 infections, including genetic sequences and sensitivities to VRC01-mediated neutralization and Fc effector functions.

Conclusions—The AMP trials test whether VRC01 can prevent HIV-1 infection in two study populations. If affirmative, they will provide information for estimating the optimal dosage of VRC01 (or subsequent derivatives) and identify threshold levels of neutralization and Fc effector functions associated with high-level protection, setting a benchmark for future vaccine evaluation and constituting a bridge to other bnAb approaches for HIV-1 prevention.

Keywords

Correlates of protection; Clinical trial; HIV prevention; Monoclonal antibody; Sieve analysis; Surrogate endpoint; VRC01

INTRODUCTION

Rationale for testing VRC01 for HIV-1 prevention efficacy

Thirty-two years after publication of the first report of AIDS (Centers for Disease Control 1981), the global HIV-1 epidemic continues and effective biomedical interventions are still needed to reduce the acquisition of HIV-1. While many countries have made inroads on

leveling HIV-1 prevalence over the last few years, microepidemics of infection are occurring in nearly all regions, even in countries possessing the full toolkit of proven prevention approaches (UNAIDS 2013, WHO in partnership with UNICEF and UNAIDS 2013).

Antiretroviral drugs (ARVs) have been shown to be effective for HIV prevention (Cohen et al. 2011, Grant et al. 2010, McCormack et al. 2016, Molina et al. 2015, Baeten et al. 2012, Cohen et al. 2016). However, efficacy trial results have been heterogeneous in the level of adherence and in the level of efficacy, with possible differences by gender and route of exposure (Van Damme et al. 2012, Marrazzo et al. 2015, Patterson et al. 2011, Cottrell et al. 2016). In addition, widespread use of ARVs for HIV prevention on a population basis is beset by challenges of poor adherence, side effects, and resource limitations (van der Straten et al. 2012, AIDSinfo 2014). For these reasons, a biomedical HIV prevention approach with sustained activity, an acceptable safety profile, and less dependence upon adherence is still needed. In this regard, development of a safe and effective vaccine is needed for a rapid and sustained decline in HIV incidence (Fauci, Folkers, and Marston 2014).

Over the past several years, there has been a concerted and successful effort to isolate broadly neutralizing antibodies (bnAbs) to HIV-1 from chronically-infected individuals (Li et al. 2007, Simek et al. 2009, Wu et al. 2010, Gray et al. 2009, Gray et al. 2011a, Corti et al. 2010, Walker et al. 2011, Scheid et al. 2011, Binley et al. 2008, Sather et al. 2009, Falkowska et al. 2012, Walker et al. 2009, Walker et al. 2010, McLellan et al. 2011, Kong et al. 2013, Huang et al. 2012). Research on the epitopes these antibodies target on the HIV-1 Env glycoprotein and on lineages and maturation pathways of these antibodies (Stamatatos et al. 2009, Walker et al. 2010, Wu et al. 2010, Gray et al. 2009) has informed efforts to design recombinant protein immunogens that might elicit such antibodies (Walker and Burton 2010, Falkowska et al. 2012, Burton, Stanfield, and Wilson 2005, Kwong and Mascola 2012, Georgiev et al. 2013). Several of these antibodies have shown considerable *in vitro* neutralization breadth, inhibiting 80–90% of isolates in *in vitro* neutralization assays (Wu et al. 2010, Zhou et al. 2010). Two antibodies to the CD4 binding site of the HIV-1 envelope, VRC01 and 3BNC117, have entered phase 1 clinical trials (e.g., VRC01 in VRC601, VRC602, HVTN 104). These antibodies have the potential to be used as passive immunization when administered as long-acting pre-exposure prophylaxis (PrEP) or post-exposure prophylaxis (PEP).

VRC01 monoclonal antibody

VRC01 is a human IgG1 monoclonal antibody (mAb), directed against the CD4-binding site of the HIV-1 envelope glycoprotein. VRC01 was developed by the National Institute of Allergy and Infectious Diseases' Vaccine Research Center (VRC). VRC01 has been tested in several non-human primate (NHP) challenge experiments (Shingai et al. 2014, Pegu et al. 2014). *In vitro* assays demonstrated that 90% of 190 HIV-1 isolates across all clades tested showed sensitivity to neutralization (Pegu et al. 2014). VRC01 has a 50% inhibitory concentration (IC₅₀) of < 50 mcg/mL against 91% of HIV-1 isolates and an IC₅₀ < 1 mcg/mL against 72% of HIV-1 isolates (Zhou et al. 2010, Wu et al. 2010). In NHP studies, both virus sensitivity and a range of serum antibody concentrations appear to influence protection from mucosal challenge.

Trials in rhesus macaques demonstrated that VRC01 delivered at different doses with a range of serum concentrations could prevent simian HIV acquisition (Shibata et al. 1999, Baba et al. 2000, Hofmann-Lehmann et al. 2002, Nishimura et al. 2002, Hessel et al. 2009, Moldt et al. 2012, Pegu et al. 2014, Ko et al. 2014). Phase 1 trials of these human-derived antibodies have shown them to be well tolerated at intravenous (IV) doses up to 40 mg/kg every 4–8 weeks and to have broad *in vitro* neutralizing activity (Ledgerwood et al. 2015, Mayer et al. 2016).

An important requisite for designing vaccines that elicit bnAbs is understanding the level of neutralizing activity required to achieve protection in humans. It will be important to know whether such levels vary by infection route or neutralizing sensitivity of the infecting isolate. Since VRC01 contains the Fc portion of IgG1 and displays Fc receptor-mediated effector functions (Ko et al. 2014), knowledge of the role of antibody-mediated effector functions (e.g., ADCC, ADCP, ADCVI) in protection will also be critical (Florese et al. 2009, Forthal et al. 2005, Asmal et al. 2011, Tudor et al. 2009).

Passive administration of antibodies to prevent the acquisition of and/or treat infectious diseases has been utilized in medicine for over 100 years (Graham and Ambrosino 2015). Most notably, passive administration of palivizumab (Null et al. 1998) has informed RSV vaccine development by defining neutralizing antibody levels that correlate with reduced disease severity in RSV-infected infants. This precedent has informed the conceptual framework for the AMP trials. As antibodies are usually more expensive to manufacture than small molecules, defining the optimal concentrations that correlate with protection from acquisition is an important prerequisite for future product design.

Overview of the AMP trial primary objectives

The Antibody Mediated Protection (AMP) trials consist of HVTN 704/HPTN 085 (ClinicalTrials.gov #NCT02716675, Evaluating the Safety and Efficacy of the VRC01 Antibody in Reducing Acquisition of HIV-1 Infection Among Men and Transgender Persons Who Have Sex With Men) and HVTN 703/HPTN 081 (ClinicalTrials.gov #NCT02568215, Evaluating the Safety and Efficacy of the VRC01 Antibody in Reducing Acquisition of HIV-1 Infection in Women). These harmonized protocols evaluate the safety and efficacy of VRC01, administered in 10 intravenous (IV) infusions every 8 weeks, in reducing acquisition of HIV-1 infection. HVTN 704/HPTN 085 enrolls 2700 HIV-uninfected men and transgender persons (TG) in North and South America (the United States [US], Peru, Brazil) and Switzerland (Switzerland) who have anal sex with men or TG; HVTN 703/HPTN 081 enrolls 1500 HIV-uninfected sexually active women in sub-Saharan Africa (Botswana, Kenya, Malawi, Mozambique, South Africa, Tanzania, Zimbabwe). The first primary objective of these trials is to evaluate safety and tolerability of VRC01; the second primary objective is to evaluate efficacy of VRC01 for prevention of HIV-1 infection. Study participants are randomized to receive 10 mg/kg or 30 mg/kg IV doses of VRC01 or a control infusion, every 8 weeks for a total of ten infusions.

For each trial, the primary efficacy analysis assesses prevention efficacy (PE) against the primary endpoint of documented HIV-1 infection by the Week 80 visit in the two bnAb groups pooled versus the control infusions group; throughout, PE denotes prevention

efficacy of the pooled bnAb groups versus control. (A key secondary analysis assesses PE against VRC01-sensitive HIV-1s, as described under Statistical Analyses below.) The sample sizes are selected to provide 90% power to detect PE = 60% (rejecting the null hypothesis of PE = 0%) for each trial.

Overview of the AMP secondary objective to assess correlates of protection (CoPs)

A secondary objective of the AMP trials is to develop a marker(s) of the VRC01 bnAb that correlates with the level and antigenic specificity of protection against HIV-1 infection. This objective is assessed using the following secondary endpoints:

1. Serum concentration of VRC01 in participants assigned to receive the bnAb (measured by ELISA and a neutralizing antibody assay)
2. Serum bnAb-mediated neutralization and Fc effector functions to HIV-1 Env proteins representing the variability of the VRC01 antibody footprint
3. Sequences of breakthrough HIV-1 infections from the earliest available HIV-1-positive plasma samples
4. Neutralization and Fc effector function sensitivity to VRC01 (purified and in VRC01 recipient sera) of HIV-1 strains from HIV-1 infected trial participants, based on analysis of the earliest available post-HIV-1-infection serum samples

This article summarizes the study design of each AMP trial and the statistical analysis plan for assessing the primary efficacy objective and the secondary CoPs objective; it does not address the design and analysis for the primary safety and tolerability objective.

AMP TRIAL DESIGN

Participants are tested for HIV-1 infection at every study visit (4-weekly) and at any time following participant report of possible HIV-1 exposure during the study. Participants are followed 12 weeks beyond the Week 80 visit for extended assessment of safety and tolerability, as well as for tracking the rate of waning of VRC01 serum concentration. Therefore, the planned study duration for each participant is 92 weeks.

Rationale for the AMP trial design

Several considerations inform the design of the AMP trials. First, each trial is powered to determine whether passive administration of bnAbs can block HIV-1 acquisition. Based on data from *in vitro* neutralization studies, NHP challenge studies, and pharmacokinetic (PK) data from the phase 1 clinical trials VRC 602 and HVTN 104, the doses selected are designed to evaluate the activity of VRC01 across a range of serum concentrations in a diverse population of at-risk persons in multiple geographic regions. The AMP trials are designed to inform the optimal dosage for widespread use of the antibody or its subsequent derivatives, as well as to benchmark the types of effector functions associated with efficacy, thus constituting an important bridge to other bnAbs and other bnAb delivery systems.

Control infusions are used in each AMP trial based on the important scientific and public health questions each trial seeks to answer and on an acceptable risk/benefit balance. In

particular, the control infusion arm creates a double-blinded trial that ensures an unbiased assessment of PE and product safety. Hence, the AMP trials meet the standards commonly required for ethical use of control groups in clinical trials (World Medical Association 2002, Emanuel and Miller 2001, Emanuel, Wendler, and Grady 2000).

Two VRC01 doses are included in the AMP trials for a number of reasons. First, the serum VRC01 level required to provide protection against HIV-1 infection in humans is unknown. Defining the lowest efficacious VRC01 dose is essential for defining a target for this antibody and its derivatives that can be developed commercially. Typically, higher doses of drug products are associated with lower tolerability due to side effect profiles or operational features such as length of drug administration time (e.g., IV infusion time). At present the 30 mg/kg dose cannot be administered subcutaneously (SC) Due to challenges in adequately scaling IV administration to populations most in need of protection against HIV-1 infection, and the complexity and cost of manufacturing mAbs, the feasibility of long-term 30 mg/kg dosage is limited. Therefore, it is important to evaluate the level of efficacy at lower antibody concentrations.

The appropriateness of including the 10 mg/kg dosing group is supported by the high percentage of isolates that are susceptible to antibody concentrations as low as 1 mcg/mL, one-fourth the median trough level anticipated with 10 mg/kg dosing every 8 weeks based on HVTN 104 and the large degree of overlap in the VRC01 concentrations between the 10 and 30 mg/kg dose groups for a considerable period of time after each infusion (Figure 1). Figure 2 shows predicted median VRC01 serum concentrations over time for the 10 and 30 mg/kg dose groups and Figure 3 shows measured VRC01 serum concentrations for participants in the HVTN 104 trial who received VRC01 at 10 mg/kg (N=12) or 30 mg/kg (N=12) (Mayer et al. 2016). Based on *in vitro* neutralization studies and NHP challenge studies, the expected trough levels of VRC01 are projected to be an order of magnitude higher than the *in vitro* IC50 levels required to protect against a majority of circulating HIV-1 strains. In particular, at a trough level of 16 mcg/mL observed in the 30 mg/kg group, the breadth of neutralization based on IC50/IC80 values is 94%/93% against a panel of 56 clade B isolates, and 80%/73% against a panel of 200 southern African clade C isolate. At a trough level of 4 mcg/mL observed in the 10 mg/Kg group, the corresponding breadth of neutralization based on IC50/IC80 values is 93%/82% against clade B, and 75%/58% against clade C.

Figure 4 illustrates a model of how PE could manifest itself in the AMP trials. For the scenario that PE is 40% in the 10 mg/kg dose group and 80% in the 30 mg/kg dose group, it shows expected cumulative incidences of HIV-1 infection since the most recent infusion by treatment group, under a PK model for VRC01 concentrations and a model linking higher current concentration to a lower instantaneous hazard of HIV-1. In the VRC01 group HIV-1 incidence starts low post-infusion and steadily increases as VRC01 concentration decreases, whereas in the control infusions group HIV-1 incidence is constant over time post-infusion. A planned analysis assesses PE via such estimated cumulative incidence curves illustrated in Figure 4 (see Prevention Efficacy Parameters and Associated Hypothesis Tests and Study Questions).

The variability of VRC01 concentration across individuals and time also facilitates evaluation of antibody effector functions (eg, ADCC, ADCP, ADCVI), in addition to neutralization, for their association with the level of protection. This information will aid the development of other bnAbs and other biomedical HIV-1 prevention modalities. Hence, in addition to providing a test of concept for passive immunoprophylaxis for HIV-1 prevention, the data generated by this trial will help guide development of functionally enhanced mAbs and the development of long-term delivery strategies (e.g., through a vaccine immunogen or vectored immunoprophylaxis). This trial affords a unique opportunity to correlate serum antibody levels and the potency of effector functions with protective efficacy, in a trial in which these parameters can be measured close to the time of HIV-1 acquisition, providing benchmarks for vaccine development and helping define targets of antibody effector mechanisms for protein- or viral-vector-based vaccines. The data will also provide critical information by which to evaluate which NHP models (eg, SHIV challenge stocks) are most predictive of bnAb efficacy in humans.

HIV-1 prevention package provided to study participants

All study participants are provided with standard-of-care HIV-1 prevention (WHO 2015) and will have access to oral PrEP free of charge when it has been approved by the relevant ethics committee/institutional review board and national regulatory bodies, accounting for community consultation and advice from persons and groups with bioethics and human subjects protection expertise. It is anticipated that standards of HIV-1 prevention will continue to evolve in the coming years, but it is also expected that diversity across geographic regions and risk groups will persist with regards to recommendations for HIV prevention standards. Neither the study design nor the concept of the trial are influenced by this diversity in PrEP guidelines, given that the primary analysis of PE assesses overall efficacy of VRC01 averaging over participant subgroups with different risk factors and HIV prevention practices. The contribution of these prevention efforts to reduce HIV-1 incidence is monitored and is taken into account in the statistical power calculations (see Summary of HIV-1 Incidence in the Two Study Populations).

Overview of trial monitoring

Each AMP trial is monitored for safety of VRC01 and in the following four ways: (1) feasibility monitoring for retention and adherence to infusions; (2) sequential monitoring of PE including for potential harm; (3) monitoring for futility to assess PE, defined as an inability to answer the primary efficacy objective in a timely manner; and (4) operational monitoring for other performance standards of quality of trial conduct. Data on these monitoring activities are reported to the independent Data Safety and Monitoring Board (DSMB) every 6 months. While guidelines for trial modifications based on the monitoring operate separately for each of the AMP trials, the same multinational DSMB monitors the two trials, facilitating its ability to account for all information when making recommendations (Ellenberg, Fleming, and DeMets 2002).

As these are the first large-scale studies with intravenous administration of a biomedical intervention for prevention of sexual HIV-1 transmission, the feasibility monitoring (1) includes an early feasibility check after at least 120 participants in each trial have reached

the Week 32 visit. A guideline is specified for an adequately high follow-up retention rate and not permanently discontinuing infusions by the Week 32 visit. The monitoring of PE (2) is summarized below in the Statistical Plan for Sequential Monitoring of PE. Monitoring of futility to assess PE (3) includes projections at each DSMB meeting of how long it will take until there are enough total HIV-1 infection events such that each trial has approximately 70% power to reject $H_0: PE = 0\%$ if $PE = 60\%$ in the primary analysis (this number is approximately 29 total HIV-1 infections diagnosed by the Week 80 visit). If the projected time to the 29th total infection in a given trial is beyond a specified duration with high probability, then a guideline would be met for recommending completing the trial based on the inability to answer the primary PE question in a timely manner. The time projections are also used as input into potential trial modifications, such as extending accrual or increasing enrollment at certain study sites. The projections of the time to the 29th total infection are computed as posterior distributions in a Bayesian approach that models each trial as a combination of three processes—enrollment, dropout, and HIV-1 infection. To reduce variability of the projections the parameters for the infection process are chosen based on a combination of pre-trial assumptions and the data observed thus far in the study. For operational monitoring (4), standard measures on enrollment, dropout, expected and observed visit attendance, permanent discontinuation of infusions, infusion adherence patterns, and accuracy and timeliness of data entry and resolution of data queries is summarized in the DSMB reports.

STATISTICAL DESIGN

Prevention Efficacy Parameters and Associated Hypothesis Tests and Study Questions

The sample sizes are selected to provide approximately the same amount of statistical information for each trial, operationalized as the same expected number of HIV-1 infections in the control infusions group. For a given trial, let PE denote the overall prevention efficacy of the 10 mg/kg and 30 mg/kg VRC01 groups pooled together, compared to the control group:

$$PE = [1 - F_A(t=80) / F_C(t=80)] \times 100\%,$$

where $F_A(t)$ [$F_C(t)$] is the cumulative probability of documented HIV-1 infection by t weeks after the first infusion for participants assigned to receive VRC01 [control infusion]. The primary analysis tests the null hypothesis $H_0: PE = 0\%$ vs. $H_1: PE > 0\%$ using a 2-sided $\alpha = 0.05$ level Wald test that accounts for sequential monitoring of PE for potential harm, non-efficacy, and high efficacy, summarized below. An important secondary analysis assesses PE based on a proportional hazards model that compares the time from the most recent infusion to HIV-1 infection between the VRC01 and control groups, which will have greater statistical power than the primary analysis if infections tend to occur late in infusion intervals (Figure 4).

Secondary analyses assess PE parameters PE10 and PE30 for the individual VRC01 dose groups versus the control infusions group. Secondary hypothesis tests [with questions of interest being addressed] are:

H0: PE10 = 0% vs. H1: PE10 > 0% [Does 10 mg/kg dosing of VRC01 protect?]

H0: PE30 = 0% vs. H1: PE30 > 0% [Does 30 mg/kg dosing of VRC01 protect?]

H0: PE10 = 0% and PE30 = 0% vs. H1: PE10 > 0% or PE30 > 0% with at least 1 strict inequality [Is there dose-response protection?]

H0: PE10 = PE30 vs. H1: PE10 ≠ PE30 [Does dose affect the level of protection?]

After the two trials are completed, the above hypothesis tests are repeated pooling the data sets, with interaction tests applied to assess whether PE differs by trial cohort.

Rationale for the PE parameter that measures prevention efficacy averaging over participant subgroups

The primary PE parameter measures the efficacy of VRC01 averaging over the 10 mg/kg and 30 mg/kg subgroups, in the population-based context of study participants with different risk characteristics that encompass various transmission routes, host genetics, HIV risk taking behaviors, and participant decisions to use different elements of the HIV prevention package. Given that prevention choices and practices may change during the course of each trial, the PE parameter is interpreted as averaging over this temporal as well as among-participant heterogeneity [see (Janes et al. 2013) for further discussion of these issues].

Primary Study Endpoint (Documented HIV-1 Infection)

The HVTN laboratory algorithm for diagnosis of HIV-1 infection is based on a combination of results from the Abbott ARCHITECT HIV-1 antigen/antibody Combo test, an HIV-1 RNA assay, and the Biorad HIV-1 Geenius HIV-1/HIV-2 rapid test. All testing is done at a central laboratory. An independent adjudication committee reviews all diagnostic data and determines each HIV-1 infection primary endpoint and the first study visit at which the HIV-1 infection is documented, defined as the date of HIV-1 infection diagnosis. This procedure is the same as that used in previous HVTN trials except that earlier generation kits were used [e.g., in the Step trial, the Uni-Gold Recombigen HIV test (Trinity Biotech) and the Multispot HIV-1/HIV-2 Rapid Test (Bio-Rad) were used (Buchbinder et al., 2008)].

While the HVTN algorithm is highly sensitive and specific, its operating characteristics for individuals receiving infusions of VRC01 are still being studied, for potential modification for AMP. The two main theoretical concerns for the AMP trials being investigated are: (1) a VRC01 recipient could have a false positive HIV-1 infection result if he or she is tested only with a reactive anti-VRC01 antibody-based screening test; and (2) VRC01 could suppress HIV-1 viral load early in HIV-1 infection such that detection of HIV-1 infection could be delayed for participants in the VRC01 group versus the control infusions group (Lynch et al. 2015, Caskey et al. 2015). The analysis of each AMP trial will assess the impact of VRC01 on plasma HIV-1 viral load during the first 6 months after HIV-1 infection diagnosis.

Issue (1) is not a concern because the HVTN algorithm can almost perfectly discriminate antibodies to VRC01 versus HIV-1 infection due to the small (35 residue) footprint of VRC01. Moreover, seven commercially available diagnostic kits have been tested at VRC01 serum concentrations ranging from 0.78 to 1600 micrograms/mL, without yielding any false positive results. Issue (2) poses open questions given limited data at this time on operating

characteristics of HIV-1 diagnostic assays early in HIV-1 infection for individuals who acquire HIV-1 while receiving VRC01. It is theoretically possible that some HIV-1-infected participants would not be diagnosed as HIV-1 positive until the VRC01 infusions are discontinued. To ensure that all HIV-1 infections during the trial are diagnosed, all participants will be followed up to the Week 92 visit (20 weeks post last infusion), by which time it is expected that VRC01 will be undetectable in most participants. A more subtle effect from issue (2) would be that HIV-1 infection endpoints of VRC01 recipients tend to register a few weeks after those of control infusion recipients, even if HIV-1 acquisition events happen on the same time-scale for the two groups. This diagnostic delay would cause a small and expected negligible amount of bias in the analysis of PE (Meier and Gilbert 2005), but its potential for biasing the correlates analysis is greater given the need to estimate VRC01 concentrations near HIV-1 acquisition times.

The studies of diagnostic assay operational characteristics are expected to yield a highly specific and sensitive finalized algorithm for defining HIV-1 infection. In particular, the Biorad Geenius rapid test allows Fiebig staging and quantitation of band intensity, which may be useful in interpreting seroconversion patterns in the context of VRC01. Moreover, a surprising result such as profound HIV-1 suppression by VRC01 post-acquisition would itself constitute a landmark result for the HIV-1 prevention field.

Statistical Plan for Monitoring of Prevention Efficacy

For each trial, PE is sequentially monitored 6-monthly by the independent DSMB. The AMP trial statisticians provide the DSMB a guideline for whether to recommend stopping the study early for potential harm, non-efficacy, or high efficacy, based on interim estimates and 95% confidence intervals for PE. Gilbert et al (Gilbert et al. 2011) describe the conceptual approach to this sequential monitoring of PE, which is also similar to that used for the HVTN 505 phase 2b HIV-1 vaccine efficacy trial (Hammer et al. 2013).

Table 1 summarizes the statistical approach to monitoring PE. Each of the three events of potential harm, non-efficacy, and high efficacy is monitored by comparing HIV-1 incidence between the VRC01 groups pooled versus control infusions, to harmonize with the primary analysis. The potential harm monitoring tests for negative PE after each diagnosed HIV-1 infection starting at the twentieth, and provides the result to the DSMB immediately if the pre-specified monitoring boundary is crossed. This monitoring controls the family-wise error rate of falsely hitting the potential harm boundary at level 0.05, accounting for the multiple hypothesis tests.

Non-efficacy and high-efficacy monitoring commence once a minimal number of total HIV-1 infection events occur that allows for an adequately small nominal 95% confidence interval about PE. These two monitoring types are done at each 6-monthly DSMB meeting. The non-efficacy monitoring recommends stopping if low PE at best is established with high confidence, defined as the nominal 95% confidence interval for PE lying below 40%. This type of non-efficacy stopping boundary was recommended by Freidlin, Korn, and Gray (Freidlin, Korn, and Gray 2010), with rationale that the primary efficacy result that would be reported in the abstract of the journal article would clearly communicate a result of low efficacy at best with precision. Once non-efficacy monitoring commences, the potential

harm monitoring after each infection ceases, as the non-efficacy monitoring now provides the function of detecting an elevated rate of HIV-1 infection.

The high efficacy monitoring recommends early stopping only if PE is high with high confidence, implemented as the nominal 95% confidence interval for PE lying above 70%. A stringent benchmark for high efficacy is used because there is not a strong ethical rationale to expeditiously cross control infusion recipients over to the VRC01 group (given that the infusions cease by week 72 for all study participants), and it is important to complete the trial follow-up in the case of substantial PE in order to answer the PE and CoP objectives with maximal precision. The non-efficacy and high efficacy monitoring are based on nominal confidence intervals instead of monitoring-adjusted confidence intervals for simplicity, and because the adjustment would have a very minor effect.

Trial Sample Size

Because each trial is designed to have 90% power to reject $H_0: PE = 0\%$ if PE is at least 60%, the design may be classified as an intermediate-sized phase 2b screening efficacy trial (Fleming and Richardson 2004, Gilbert 2010, Rida et al. 1997). Two parallel trials are conducted each with this level of power because (1) PE may vary by route of transmission/population and it is of interest to learn the answer in each distinct setting; (2) tolerability, feasibility/adherence, and other factors that affect PE may vary in the two trials, and measuring these factors in two distinct settings may help in interpreting the PE results; (3) pooling the data from the two trials increases the power for assessing CoPs; and (4) it is ethically acceptable to learn about PE concurrently based on parallel comparisons with inert control groups, whereas it would likely not be ethically warranted to conduct sequential controlled trials if the first trial showed a substantial level of protection (i.e., there is a time window of opportunity).

Sample size calculations—The sample size calculations are based on the power of a 1-sided 0.025-level Wald test for comparing log-transformed cumulative incidences of HIV-1 infection by the Week 80 visit between the pooled VRC01 groups versus the control group, in the presence of the sequential monitoring of PE summarized above. Power is computed using the open source R package seqDesign available at the Comprehensive R Archive Network (CRAN).

Figure 5 shows the sample size per group required to have 90% power to reject $H_0: PE = 0\%$ in favor of $H_1: PE > 0\%$ for different values of PE set to be the same for the 10 and 30 mg/kg dose groups ($PE = PE_{10} = PE_{30}$), for each trial. The results show that about $N = 2300$ and 1250 total participants are needed to achieve 90% power to detect $PE = 60\%$ in the Americas/Switzerland and Africa trial, respectively. Figure 5 assumes 3% and 5.5% annual HIV-1 incidence in the control group for the Americas/Switzerland and Africa trial, respectively, with rationale summarized below. The selected $N=2700$ and $N=1500$ sample sizes build in conservative margins, where the smallest control group incidence under which there is at least 90% power to detect $PE = 60\%$ is 2.3% and 4.2%, respectively.

Assumptions for the sample size calculations—The following assumptions undergird the sample size calculations, common for the two trials unless otherwise noted:

- 10% annual dropout incidence in each of the 3 study groups;
- 32-month (30-month) uniform accrual in the Americas/Switzerland (Africa) trial halved during the first 3 months
- Study visits every 4 weeks for HIV-1 diagnostic tests through Week 80
- 3% (5.5%) annual HIV-1 incidence in the Americas/Switzerland (Africa) trial control infusions group [with a conservative margin down to 2.3% (4.2%)]
- PE10 and PE30 within an 8-week infusion interval are the same for each of the 10 infusion intervals
- Sequential monitoring of PE is conducted as summarized in Table 1.

Summary of HIV-1 Incidence in the Two Study Populations

Over the past 25 years, MSM and TG have been the only risk groups in the US for which estimated HIV-1 incidence has increased, with more recent increases in incidence noted particularly among young MSM of color (Hall et al. 2008, Prejean et al. 2011), (Centers for Disease Control and Prevention 2015). Among gay and bisexual men, Black MSM—especially those between the ages of 13 and 24—are the group most affected by HIV-1 (Centers for Disease Control and Prevention 2012), (Koblin et al. 2013). HIV-1 prevalence in sub-Saharan Africa is the highest worldwide, with acquisition rates among heterosexual women as high as 6–8% annually in some locations (UNAIDS 2014). In recent clinical trials in southern Africa, annual HIV-1 incidence rates of 1.9–9.1% have been observed among sexually active women (Padian et al. 2007, Celum et al. 2008, Gray et al. 2011b, Gray et al. 2014, Abdool et al. 2010, Van Damme et al. 2012, NIAID 2013).

Rationale for the background HIV-1 incidence assumptions—The assumption of 3% annual HIV-1 incidence in the Americas/Switzerland trial control group accommodates background PrEP use by assuming incidence is decreased by two parameters: (1) the fraction of person-years at-risk (PYRs) during which PrEP is used; and (2) the level of efficacy of PrEP during PrEP use. Figure 6 shows the annual HIV-1 incidence in the control group as a function of these 2 factors. Based on HVTN 505 incidence data (US participants only), we consider an assumption of 1.6% annual control group incidence in the US (which is expected if 27% of person-years at risk are during PrEP use and PrEP efficacy is 90% during use), and the same 1.6% annual control incidence in the planned n=70 enrollees in Switzerland. Then if we assume a 5% annual control group incidence in South American participants (Grant et al. 2010) (with 5% of person-years at risk during PrEP use and 90% PrEP efficacy during use), 48% of participants enrolled at South American sites would yield the assumed 3% annual control group incidence rate.

The assumption of 5.5% annual HIV-1 incidence for the control group of sub-Saharan African women is based on recent data directly measuring HIV-1 incidence in women aged 18–40 at several African sites (Gray et al. 2014, Abdool et al. 2010, Van Damme et al. 2012, NIAID 2013, Marrazzo et al. 2015), (Rees et al. 2015). It is difficult to predict the degree of PrEP use by sub-Saharan African women during the trial. Given that PrEP is not currently available in the public sector for sub-Saharan African women and is currently seldom used

outside of controlled clinical trials, we assume gradual scale up and estimate that no more than 15% of the PYRs would be during PrEP use periods. Assuming 90% PrEP efficacy during PrEP use would mean that the 5.5% annual incidence assumption corresponds to assuming 6.4% annual incidence in the absence of any PrEP use.

Robustness analysis—Each trial can accommodate up to about 50% PrEP use and still plausibly achieve approximately 80% power to detect $PE = 60\%$. In particular, the Americas/Switzerland trial has 79% power with as low as 1.65% control group annual incidence, a scenario attained if 50% of PYRs are without PrEP use and a 3% control group annual incidence and the other 50% are with PrEP use at 90% PrEP efficacy. The Africa trial has 79% power with as low as 3.02% control group annual incidence, a scenario attained if 50% of PYRs are without PrEP use and a 5.5% control group annual incidence and the other 50% are with PrEP use at 90% PrEP efficacy.

Power for prevention efficacy of the individual VRC01 dose groups—Under the design parameters listed above including 3.0% and 5.5% control group annual incidences, each trial has about 80% power to detect PE_{10} or PE_{30} equal to 55%.

Statistical Operating Characteristics of the Trials

For the selected total sample size of the Americas/Switzerland trial ($N = 2700$), Figure 7 shows probabilities of early stopping for potential harm, non-efficacy, and high efficacy, in addition to the probability that the final result will be significant $PE > 0\%$ (i.e., unconditional power). Supplementary Table 1 provides the same information in tabular form. The corresponding figure and table for the Africa trial are almost identical and are not shown.

Figure 7 shows greater than 90% power to detect $PE = 60\%$ for each trial, and greater than 95% probability that the trial would stop early for non-efficacy if there is no PE. There is low risk (less than 5%) of a potential harm result if $PE = 0\%$, and stopping for high efficacy would only occur if PE is very high, above 90%.

Table 2 shows, for each trial, the numbers of HIV-1 infection endpoints during the trial expected under the null hypothesis scenario $PE = PE_{10} = PE_{30} = 0\%$ and under the design alternative hypothesis scenario $PE = PE_{10} = PE_{30} = 60\%$. The calculations show that 37 HIV-1 infection endpoints are expected in the control group of each trial. Under the design alternatives $PE_{10} = PE_{30} = 60\%$, 30 total HIV-1 infection primary endpoints are expected pooled over the two VRC01 groups in the Americas/Switzerland trial and 31 total in the Africa trial. These numbers of breakthrough cases for the $PE_{10} = PE_{30} = 60\%$ scenario are specified for illustrative power calculations for the CoP analysis described below.

STATISTICAL ANALYSES

Overall Prevention Efficacy Primary Objective

The primary analysis of PE analyzes the modified intention-to-treat (MITT) cohort, comprising all randomized participants who receive the first infusion and are not later determined to have been HIV-1 infected at entry. The MITT analysis analyzes participants

according to the initial randomization assignment regardless of how many infusions they receive after the first.

Cumulative incidences of HIV-1 infection over time for individual and pooled study groups are estimated by a transformation of Nelson-Aalen estimators (Kalbfleisch and Prentice 2002) for the respective cumulative hazard functions of HIV-1 infection. PE parameters are estimated by one minus the ratio (VRC01/control group) of these cumulative incidence estimators. For the final primary analysis time point (Week 80 visit), 2-sided Wald p-values are reported.

Supportive analyses of cumulative incidence curves and PE(t) over time are conducted using targeted minimum loss based estimation (tMLE) (Moore and van der Laan 2009, van der Laan and Rose 2011). The tMLE method can decrease bias by allowing dropout to depend on participant covariates and can improve statistical precision by leveraging information in baseline participant characteristics predictive of HIV-1 infection and/or dropout.

Complementary secondary analyses assess PE using Cox proportional hazards models and score tests for whether PE differs from zero. This analysis uses as the failure time the time from the most recent infusion to the estimated date of HIV infection, which can have increased power compared to the primary analysis if instantaneous PE is higher when VRC01 concentrations are higher.

The Cox models are selected for the secondary analysis so that the interpretability and validity of the primary analysis does not depend on the proportional hazards assumption. The Cox model will also be used for a secondary analysis of PE in subgroups defined by the level of cumulative infusion adherence over time included in the model as a time-varying covariate. An advantage of using the Cox model is that some of the CoP analyses described below use a Cox model, thereby harmonizing the overall and marker-subgroup analyses of PE.

Statistical Analyses and Power Calculations for the CoP Secondary Objective

VRC01 markers as individual-level correlates of protection (CoPs)—The secondary objective of the AMP trials is addressed by measuring VRC01 markers over time in VRC01 group breakthrough HIV-1-infected cases, and in a random sample of HIV-1 uninfected VRC01 group participants (i.e., the ‘marker subset’), integrated with genotypic and phenotypic sieve analyses of breakthrough HIV-1 viruses accounting for knowledge of the VRC01 antibody footprint and HIV-1 Env genetic signatures associated with VRC01 susceptibility. The main analyses of the secondary objective pool over the two VRC01 dose groups to maximize statistical power, and include interaction tests for whether associations differ in the two dose groups. The CoP analyses are conducted for each trial individually, as well as for the two trials pooled, with interaction tests applied to assess whether associations differ by trial.

The design and data requirements for successful identification of CoPs are as follows:

1. Sufficient sample size to observe enough VRC01 recipients diagnosed with incident HIV-1 infection during the primary follow-up period through to the Week 80 visit
2. A result of significant beneficial overall VRC01 PE > 0%
3. Sufficiently frequent HIV-1 testing schedule and accurate HIV-1 diagnostics
4. Wide variability in the concentration and effector functions of VRC01 over time and individuals, and the ability to measure these characteristics such that their timing relative to HIV-1 acquisition is estimated with adequately low measurement error
5. VRC01 marker sampling design that allows sufficiently accurate modeling of mAb characteristics over time

CoP analyses assess whether and how PE varies over subgroups defined by marker readouts. Note that the serum concentration VRC01 markers are defined without reference to a particular HIV-1 Env sequence. In contrast, the effector function markers are defined relative to specific HIV-1 Envs or panels of Envs, such that they may be linked to immunogenicity endpoints that would be used in future trials of candidate HIV-1 vaccines.

Sampling of VRC01 markers—Specimens are collected and stored from every study participant to preserve blinding, with serum collected at every study visit (4-weekly through Week 92 and at Day 61) and PBMC collected at the Week 0, 40, 72, 92 study visits. Characteristics of VRC01 or immune responses to VRC01 are measured predominantly in participants assigned to the VRC01 groups. Markers are measured from participants in the ‘marker subset,’ consisting of all primary endpoint HIV-1-infected cases in the VRC01 groups, and a random sample of participants from each VRC01 group who reach the Week 80 study visit HIV-1 negative (together with a small random sample of control group participants for blinding and quality control purposes for laboratory conduct of assays). The random sample is chosen after the final primary analysis to create a common control:case ratio across key strata defined by factors such as dose group. The markers are measured at each 4-weekly study visit through Week 92 (to see the tail of decline during the 20-week period post last infusion at Week 72), as well as at the visit scheduled 5 days after the second infusion, where for VRC01 group cases measurements are made through to the first HIV-1 positive sampling date and at time points post infection diagnosis (Week 0, 2, 4, 6, 8, 12, 24). The Day 5 post-second infusion visit is included because it occurs shortly after the end of the distributive phase of VRC01, which together with the samples drawn just before infusions (Week 0, 8, ..., 72), the samples drawn 4 weeks post-infusions (Week 4, 12, ..., 76), and the Week 80, 84, 88, 92 samples 8–20 weeks post last infusion constitute the basis for modeling marker curves over time.

Overview of the analysis of a VRC01 marker as a CoP—For a given VRC01 recipient, let $S(t)$ be a marker characteristic at time t post enrollment (first infusion date) in days, for example the time-concentration curve. The entire curve $S(t)$ cannot be known exactly given that the marker is measured at a grid of time-points. However, a PK model can be used to express the whole curve $S(t)$ as a function of the measured marker readouts and

error terms. One approach to the CoP analysis estimates the “prevention efficacy curve” $PE(S(t) = s)$ defined as

$$PE(S(t) = s) = [1 - (\text{hazard- mAb}(S(t) = s)) / (\text{hazard- control}(S(t) = s))] \times 100\%,$$

where $\text{hazard-mAb}(S(t) = s)$ is the hazard rate of HIV-1 infection at time t for a participant assigned to either VRC01 group with $S(t) = s$ and $\text{hazard-control}(S(t) = s)$ is the hazard rate of HIV-1 infection at time t for a participant assigned to the control group who would have had $S(t) = s$ had s/he been assigned to one of the VRC01 groups (thus $S(t)$ is a counterfactual random variable for control infusion recipients). The $PE(s)$ curve for each fixed marker level $S(t)=s$ is interpreted as the percent multiplicative reduction in the instantaneous incidence of HIV-1 infection for the subgroup of VRC01 recipients with current value of the marker equal to s versus this same subgroup had they, counter to fact, not been assigned to receive VRC01. As such, $PE(s)$ approximately measures the causal effect of VRC01 to reduce the per-exposure probability of HIV-1 infection for the subgroup with marker value s at the time of exposure. By measuring PE approximately as per-exposure PE, the analysis aims to generate clues about the mechanisms of protection, in addition to generating knowledge about the statistical correlates of protection (Plotkin and Gilbert 2012).

A second set of CoP analyses assess a ‘fixed-time’ version of the PE curve that defines subgroups by a marker readout at an early fixed time point such as the Week 4 or 5 days post second infusion visit. The fixed-time PE curve is defined as:

$$PE(s) = [1 - F_A(t=80|s) / F_C(t=80|s)] \times 100\%,$$

where $F_A(t|s)$ is the cumulative probability of documented HIV-1 infection by t weeks after the first infusion for participants assigned to receive VRC01 with marker level s at the fixed time point, and $F_C(t|s)$ is the cumulative probability of documented HIV-1 infection by t weeks after the first infusion for participants assigned to receive the control infusions with counterfactual marker level s at the fixed time point (had they been assigned to receive VRC01). This latter curve is a natural extension of the overall PE curve, defined identically except it restricts to the subgroup defined by the marker readout under assignment to VRC01. The same type of curve has been used for assessing fixed-time CoPs in vaccine efficacy trials (eg, (Follmann 2006, Gilbert and Hudgens 2008, Qin et al. 2008, Huang, Gilbert, and Wolfson 2013, Gabriel and Gilbert 2014, Gilbert et al. 2014)). The fixed-time curve is probably less important for generating clues about mechanisms of protection and more important for identifying an early marker that is prognostic for the level of protection conferred by VRC01. The RV144 immune correlates study (Haynes et al. 2012) and follow-up analyses assessed fixed-time (Week 26) correlates of risk and CoPs.

In addition to estimation, hypothesis testing is performed to assess whether a VRC01 marker modifies PE. Within the framework used for estimation described above, this entails testing time-varying and fixed-time PE curves, respectively:

$$H0:PE(S(t)=s)=PE \text{ vs. } H1:PE(S(t)=s) \text{ varies in } s, \quad (1)$$

$$H0:PE(S=s)=PE \quad \text{vs. } H1:PE(S=s) \quad \text{varies in } s. \quad (2)$$

Rejecting either null hypothesis $H0$ in favor of $H1$ supports the role of the marker in protection. The CoP analysis pools over the 2 VRC01 groups to improve power through a larger sample size and through greater inter-individual variability of the VRC01 markers. Given that the exact dates of infusions have a major impact on VRC01 markers $S(t)$ such as concentration over time (e.g., notice the profound influence of missed infusions on $S(t)$ in Figure 3), the analysis accounts for the exact dates of infusions that are expected to show considerable heterogeneity across trial participants.

A time-window approach to assessing VRC01 markers as CoPs—To explain the concept of this approach, first suppose all participants received all scheduled infusions and rigidly followed the visit schedule for infusions and HIV-1 diagnostic tests. Then this approach compares HIV-1 incidence during the first 4 weeks of the infusion intervals (i.e., weeks 1–4, 9–12, 17–20, ..., 73–76) to HIV-1 incidence during the second 4 weeks of the infusion intervals (i.e., weeks 5–8, 13–16, 21–24, ..., 77–80), aggregating over the ten 8-week infusion intervals. Because VRC01 has much higher concentration during the first halves of the infusion intervals compared to the second halves (e.g., demonstrated in Figure 3), detection of higher HIV-1 incidence in the second halves of visit windows would imply that higher concentrations are associated with a lower rate of HIV-1 infection.

However, because of missed infusions and HIV-1 diagnostic testing visits, the heterogeneity among participants of infusion dates (which is allowable as specified by broad visit windows), a better approach accounts for the set of exact infusion dates of each participant. As such, our approach to implementing the time-window approach compares the mean time between the last infusion and infection for VRC01 group cases, with the mean of the participant-specific weighted average time between infusions of VRC01 group control participants, where the weights allow for changing HIV-1 incidence over time. The analysis stratifies by the 10 mg/kg and 30 mg/kg dose groups and aggregates estimates over these two dose groups. In addition, an important element of the analysis uses predictions of the HIV-1 infection times for all VRC01 group cases. Models for these predictions are built using information on HIV-1 diagnostic tests, intra-host diversity of HIV-1 sequences, HIV-1 viral loads, and perhaps other factors. Initial research to build predictors of HIV-1 infection time based on HIV-1 sequence information has been conducted using validation data sets of acutely HIV-1-infected individuals (Edlefsen et al. 2016).

A related approach to assessing CoPs integrates a PK model of the serum concentration of VRC01 over time, comparing the mean VRC01 concentration at the time of HIV-1 infection for VRC01 group cases with the mean of participant-specific weighted average VRC01 concentrations over follow-up time. The participant-specific average VRC01 concentration

over follow-up time is intended to reflect the expected concentration at the time of a random HIV exposure, with weighting used as above to allow for changing HIV-1 incidence over time.

Comparing cumulative HIV-1 incidence between the 10 mg/kg and 30 mg/kg dose VRC01 groups for assessing group-level CoPs—Table 3 reports power calculations for comparing HIV-1 incidence between the 10 mg/kg and 30 mg/kg VRC01 groups, for each trial separately and pooling across the two trials. This comparison assesses a ‘group-level CoP’ in that detection of higher PE in the higher VRC01 dose group would establish, based on a direct causal analysis stemming from the randomization, greater PE caused by a higher dose. The results show that each trial is well-powered to detect moderate differences pooling over the two trials and to detect only large differences for each individual trial (right-most column) (e.g., to detect $PE_{30} = 70\%$ greater than $PE_{10} = 30\%$, there is 93% power pooling over the trials and 68% for each individual trial). In contrast, there is high power to correctly select the best VRC01 dose group within each trial given moderate efficacy differences. However, the objective of comparing the two particular dose levels is less important than other correlates objectives that focus on estimating the concentration needed for high PE.

VRC01 markers—Markers of VRC01 effector functions to each HIV-1 Env protein in a panel of HIV-1 Envs are measured by assaying VRC01 recipient samples. A particularly relevant HIV-1 Env panel represents the VRC01 antibody footprint diversity of HIV-1s exposing trial participants. For example, relevant for the Americas/Switzerland trial, Supplementary Figure 1(A) shows a logo plot of the VRC01 antibody footprint motifs comprising 35 amino acid sites (Wu et al. 2010) based on 276 subtype B U.S. and South American sequences from the Los Alamos National Laboratory (LANL) sequence database with submission dates between 2006 and 2014 (Crooks et al. 2004, Schneider and Stephens 1990). There are 274 unique VRC01 footprints, of which 272 occur in only one sequence. In the Africa trial subtype C predominates, although subtypes A and C also circulate at some of the trial sites. Supplementary Figure 1(B) shows a parallel logo plot for 444 sub-Saharan African subtype C sequences from the LANL database with submission dates between 2006 and 2014. There are 441 unique VRC01 footprints, of which 438 occur in only one sequence.

Neutralization, virion capture, ADCC, and ADCP exemplify assays used for developing effector function markers of interest. Research is needed to develop the most useful readouts/variables to study as CoPs. An example marker based on a neutralization assay is the weighted area under the magnitude-breadth neutralization curve of the VRC01 footprint Env panel, where the weighting is by the estimated prevalence of VRC01 footprints in HIV-1s circulating at the trial sites during the trial (Huang et al. 2009). Such a marker aims to summarize the level of neutralization of HIV-1s to which a VRC01 recipient is exposed.

For VRC01 group participants with an HIV-1 infection primary endpoint, an important marker is the level at which his/her sample at the time of HIV-1 acquisition has an effector function response against his/her breakthrough founder virus or viruses. (This analysis requires modeling the marker at the time of HIV-1 acquisition based on the available grid of

marker measurements.) A result where pre-infection serum samples from infected participants generally fail to neutralize their founder viruses- yet can neutralize breakthrough founder viruses of control infusion recipients- would support neutralization as a CoP and mechanistic CoP. This analysis is repeated for other mAb effector functions (i.e., Fc) as a way to discriminate which functions are or are not CoPs, and to rank the effector functions by their contribution to explaining the level of protection. This ‘phenotypic sieve analysis’ framework, previously applied to HIV-1 vaccine efficacy trials, is summarized by Montefiori (Montefiori 2014).

Statistical power for assessing a VRC01 marker as a CoP—The marker $S(t)$ must vary within VRC01 recipients over time and/or among VRC01 recipients in order to study and identify a CoP, with more variability in one or both components providing greater statistical power to detect a CoP. In essence, to detect CoPs it is necessary that a large fraction of person-years at risk of HIV-1 infection are in high protection zones and a large fraction of person-years at risk are in low protection zones. (A “high protection zone” is a period of time during which VRC01 recipients have high levels of a marker and high levels of efficacy and a “low protection zone” is a period of time during which VRC01 recipients have low levels of a marker and low or absent levels of efficacy.) Data from VRC 601, VRC 602, and HVTN 104 show wide variability in the VRC01 concentration over time, and more limited variability among VRC01 recipients. The limited among-VRC01 recipient variability motivates the 3-group study design that randomly assigns participants to two different VRC01 doses in order to increase this type of variability.

Achieving adequate power to detect CoPs requires that the HIV-1 diagnostic testing is more frequent than the infusion schedule, and constitutes the rationale for the chosen HIV-1 testing schedule every 4 weeks. The data from all of the employed diagnostic testing kits is incorporated into the modeling of $S(t)$ and of HIV-1 infection times, and is integrated into the statistical analysis for estimation and testing of the PE curves. Lastly, achieving adequate power to detect CoPs requires studying markers that can be measured with a high ratio of potentially protection relevant variability versus protection irrelevant variability, where the latter variability could stem from technical measurement error of the assay or to error in estimation of the timing of infection. Because of this fact the trial design relies on parallel assay studies or pilot assay studies within AMP to qualify marker variables for the CoP analysis, and on development of models for $S(t)$ over continuous time t . Detecting CoPs requires a large amount of the right kind of variability in $S(t)$, not merely a large amount of variability due to any reason.

We provide one power calculation for assessing a trichotomous marker $S(t)$, the concentration of the VRC01 at time t post enrolment, as a CoP, via testing the null hypothesis of a flat PE curve [i.e., testing H_0 in equation (1) above]. This illustrative calculation divides the pooled VRC01 group at a given time t post-enrolment into three groups: $S(t) = 0$ indicates Low, $S(t) = 1$ indicates Medium, and $S(t) = 2$ indicates High for some specified definitions of Low, Medium, High. Therefore at any given time t after enrollment, $S(t)$ divides the pooled VRC01 group into 3 subgroups. In our illustration we take “0” and “2” to be a VRC01 serum concentration < 10 mcg/mL and > 50 mcg/mL, respectively, and “1” an intermediate concentration. The actual data analysis would vary the

thresholds used in dichotomous and trichotomous markers and explore how the PE curve depends on the different thresholds (Fong, Di, and Permar 2015).

Based on the model of the VRC01 time-concentration pattern depicted in Figure 1, we conduct a power calculation indexed by the percentage of PYRs of VRC01 recipients in the (Low, Medium, High) zones, and indexed by scenarios of levels of PE within each zone. We consider power calculations pooling data across both VRC01 groups and both trials, assuming 61 VRC01 group HIV-1-infected cases (the expected number, see Table 2) with VRC01 marker data and 305 VRC01 group HIV-1 uninfected controls with VRC01 marker data (5:1 control:case ratio).

Figure 8 shows power curves for the scenario that (30%, 40%, 30%) of PYRs at risk are in the (Low, Medium, High) zones (which we approximately expect to occur based on VRC 601, VRC 602, and HVTN 104, as depicted in Figure 1) and PE = (30%, 60%, 80%) for the subgroups in these zones (averaging over the 3 subgroups yields overall PE = 60%, the design alternative). The power curves also account for noise in the measurement of $S(t)$ via a parameter ρ that equals the proportion of the inter-individual marker variability that is potentially protection relevant (Gilbert, Janes, and Huang 2016); see (Gilbert, Janes, and Huang 2016) for details on how the power calculations are conducted. The results show that the noise level majorly affects power, underscoring that only assays meeting qualification criteria will be included in CoP studies. The results also show that for a reasonably tight assay ($\rho = 0.9$), there is about 90% power to detect a CoP (with effect size 30%, 60%, 80% efficacy) at the selected trial sample size. The parameter ρ also reflects noise due to uncertainty in the exact times t of HIV-1 acquisition at which values of $S(t)$ are needed. For the scenario that this extra variability reduces ρ to 0.7, there is about 75% power to detect a CoP with the indicated effect size (Figure 8).

The TZM-bl assay that is planned for measuring VRC01 concentrations is validated and known to have low technical measurement error, and the ELISA assay that is also planned for measuring VRC01 concentrations is now being validated. For these assays the major source of biologically-irrelevant variability is expected to be due to interval censoring of HIV-1 infection times and the imperfect ability to predict these times.

Figure 9 shows the same power analysis as Figure 8, fixing $\rho = 0.9$ and varying the percentage of PYRs at risk that are in the (Low, Medium, High) zones. The results show that power increases sharply with the percentage of VRC01 recipient follow-up that is in the Low and High regions, demonstrating the principle that potentially protection relevant variability in the marker is a strong determinant of power to detect a CoP. Supplementary Figure 2 shows the same calculations for the Americas/Switzerland MSM + TG trial, except using an effect size of PE = 15%, 60%, 85% across the Low, Medium, and High $S(t)$ categories (results are almost identical for the Africa trial). The figure shows that an individual trial has 80% power for this effect size if 25%, 50%, 25% of PYRs at risk are in the (Low, Medium, High) zones.

Genotypic and phenotypic sieve analysis—Genotypic sieve analysis analyzes sequences of breakthrough founder HIV-1s from the earliest available post-infection sample

and seeks to identify amino acid “signatures” that differentiate breakthrough sequences in the VRC01 groups versus the control group, in terms of differential PE against HIV-1s according to some genotypic characteristics. For example, the genotypic sieve analysis assesses differential PE against different genotypes of HIV-1 with genotype defined by (1) Number of mismatches to the subtype B (or C) consensus sequence of the VRC01 footprint (subtype selected to match the circulating subtype), (2) Number of known VRC01 neutralization escape mutations in the VRC01 footprint, (3) Number of known mAb effector function escape mutations in the VRC01 footprint for other effector functions such as virion capture, ADCC, or ADCP, (4) Number of potential N-linked glycosylation sites within a given radius of the mAb footprint, and (5) Length of variable loops in proximity to the VRC01 footprint. To illustrate the utility of genotypic sieve analysis, we consider a hypothetical result that would support the idea that VRC01 protects via specific neutralization targeting of VRC01’s epitope. In this illustration, the HIV-1 genotype is defined by four ordered categories according to whether there are 0, 1, 2, or >2 neutralization escape mutations in the VRC01 antibody footprint. The hypothetical result would be an estimated PE of 95% against HIV-1s with zero escape mutations and a PE of 0% against HIV-1s with >2 neutralization escape mutations, and intermediate and monotone decreasing level of estimated PE against HIV-1 with 1 mutation and 2 mutations, respectively.

A second type of sieve analysis, phenotypic sieve analysis, compares breakthrough viruses of VRC01 group participants to those of control group participants using an immunological assay. Specifically, data from a “mAb function checkerboard” is filled out, with the rows representing serum samples from a random sample of VRC01 recipients HIV-1 uninfected at the time of sampling, and the columns representing HIV-1 Env proteins created from the breakthrough viruses obtained. The phenotypic sieve analysis compares the sensitivity of the breakthrough viruses to VRC01 recipient sera between the VRC01 and control groups (Montefiori 2014, Gilbert et al. 2010). A second version of the phenotypic sieve analysis compares the sensitivity of the breakthrough viruses to the VRC01 clinical product between the infected VRC01 and infected control groups; this latter analysis is simpler, based on a single column of data instead of a matrix, because it does not use VRC01 recipient sera. For an example of the second type of phenotypic sieve analysis, the TZM-bl neutralization assay could be performed on all control group founder HIV-1s versus VRC01 and on all VRC01 group founder HIV-1s (pooled over the two VRC01 groups and two trials) versus VRC01, and a Wilcoxon rank sum test applied to assess a difference. Reduced TZM-bl neutralization sensitivity of the VRC01 group founder viruses would support that neutralization had a role in protection and that the TZM-bl assay was able to detect this role, and would support using the TZM-bl assay as an endpoint for evaluating future candidate HIV-1 vaccines. The HVTN conducted the first version of neutralization sieve analysis for an HIV-1 vaccine efficacy trial (Gilbert et al. 2010). Figure 10 illustrates a hypothetical genotypic and phenotypic sieve analysis of simulated AMP data sets pooling over the two trials.

Another phenotypic sieve analysis focuses on VRC01 group cases, and measures different *in vitro* VRC01 functions of closest-to-infection participant serum samples against the participant-matched breakthrough HIV-1s. The effector functions that tend to have low

readouts against participant-matched breakthrough viruses compared to a panel of control group breakthrough HIV-1s or a historic panel of HIV-1s are implicated as potential CoPs.

In addition, an important sieve analysis will assess PE against breakthrough HIV-1s that are sensitive to neutralization by VRC01 as measured using the first available post-infection sample (sensitive is defined by positive neutralization based on IC₅₀ or IC₈₀). This PE parameter measuring PE against neutralization sensitive HIV-1s is estimated (with confidence intervals and p-values) in the same way as genotype-specific PE described above. Moreover, this type of sieve analysis is also conducted with sensitive/resistant defined by positive/negative response of other assays besides neutralization (e.g., ADCC, ADCP).

Statistical power for assessing a phenotypic sieve effect—We conduct power calculations based on TZM-bl neutralization IC₅₀ values (in mcg/mL) of VRC07 (a close cousin of VRC01) to 125 viruses representing global HIV-1 diversity (data provided by Mark Louder, Robert Bailer, and John Mascola from the VRC). From the control group we sample 74 HIV-1-infected individuals with IC₅₀ from the VRC07 data set, and from the pooled VRC01 group we sample 61 HIV-1-infected individuals with IC₅₀ values from the VRC07 data set with the sampling biased by IC₅₀-dependent prevention efficacy PE(x), where PE(x) is the probability an HIV-1-infected control infusions recipient would be infected had she or he received VRC01 with IC₅₀ value of $\times$ to the exposing virus. (Seventy and 61 infections are selected such that these calculations consider the same scenario as considered above in power calculations for CoPs.) Figure 11 shows power for comparing the VRC01 IC₅₀ distribution of the pooled VRC01 group HIV-1s versus the control group HIV-1s, based on a two-sample Wilcoxon rank sum test. The effect size indexing a difference between the two groups can be expressed in terms of the PE(x) function (where a horizontal line indicates the null hypothesis of no difference), or equivalently in terms of a difference in the IC₅₀ cumulative distribution function between the two groups. The results show that, pooling over the two trials, there is high power to detect ‘Effect Size 3’ phenotypic sieve effects (second steepest curve in Figure 11). Effect Size 3 is plausibly real, because under the design alternative of 60% PE we expect a transition between full protection against the most sensitive viruses and zero protection against the most resistant viruses. Based on the panel of HIV-1s from the VRC07 data set and IC₅₀ values, Effect Size 3 assumes that VRC01 confers PE = 100% against the 11.6% most neutralization-susceptible HIV-1s and confers PE = 0% against the 2.8% most neutralization-resistant HIV-1s, and confers a monotonically changing gradient of PE from 0% to 100% against the 85.6% of intermediate viruses ranging between the two poles. The wide range of neutralization sensitivities of HIV-1s suggests that this range of prevention efficacies is plausible. In comparison, Effect Size 2 assumes PE = 100% and PE = 0% against <0.1% and <0.1% of the most sensitive and resistant viruses, respectively, whereas Effect Size 4 assumes PE = 100% and PE = 0% against 41.9% and 23.3% of the most sensitive and resistant viruses, respectively.

DISCUSSION

The primary goal of the AMP trials is to learn whether the broadly neutralizing mAb VRC01 prevents HIV-1 acquisition, and to quantify the level of its overall PE, in each of two distinct

study populations— North/South American/Swiss MSM + TG and sub-Saharan African women. To meet this goal, the AMP trials are designed to have high probability of detecting PE at level 60% or greater in each trial. If VRC01 prevents HIV-1 acquisition, then the secondary goal of the AMP trials - to learn about correlates of protection (CoPs) - becomes critically important, as this knowledge is needed to guide next steps of research for developing a broadly neutralizing mAb (or combination of mAbs) into a product, and for defining phase 1 HIV-1 vaccine trial study endpoints and benchmarks for those endpoints that can be used to qualify candidate HIV-1 vaccines for preventive vaccine efficacy trials. In particular, learning about the CoPs entails learning the VRC01 markers that associate with the level of PE, which may define the levels of certain functional assay readouts that are needed to achieve a high level of PE. Learning about the CoPs also entails characterizing the genetic sequences and immunological resistance phenotypes of breakthrough founder HIV-1s, enabling sieve analysis that can determine if the level of VRC01 PE drops off against HIV-1s with certain effector function escape mutations and/or with a relatively large degree of measured neutralization resistance (or other effector function resistance) to VRC01. Moreover, learning about the CoPs entails learning how PE depends jointly on VRC01 markers and on the genetics and immunological resistance phenotypes of breakthrough founder HIV-1s. Identification of markers of PE against specific HIV-1 genotypes or phenotypes could be a particularly powerful tool for generating insights into the mechanistic correlates of protection.

The need to meet both goals shapes the study design that includes three study groups (30 mg/kg, 10 mg/kg, control infusions) in each of two trial cohorts. Because the assessment of CoPs is essentially an interaction/subgroup analysis, it requires a substantially larger sample size to be adequately powered than the assessment of overall PE, such that pooling data over the VRC01 groups and trials/cohorts is needed to power the CoPs analysis. In addition, including two VRC01 dose levels improves power of the CoPs analysis by increasing the amount of potentially protection relevant variability in VRC01 markers over time and among VRC01 recipients. Yet including two VRC01 dose levels does not increase the sample size of the trial compared to a design with a single VRC01 dose group, based on the fact that the primary efficacy analysis compares HIV-1 incidence between the pooled VRC01 groups versus the control group. Moreover, the sample size of each AMP trial is kept to a manageable size in being a phase 2b intermediate-sized efficacy trial instead of a phase 3 trial— as a phase 2b trial it is powered to discriminate 0% from 60% PE, with an average of 2100 participants per trial, whereas in contrast a phase 3 design would be powered to discriminate 30% from 60% or 25% from 50% PE thus requiring more than twice the sample size. While conducting two trials is important for the CoPs analysis, perhaps the most important reason for this design choice is simply to answer whether VRC01 works in each of two distinct at-risk populations of global importance, a question that would be much harder to answer in sequential efficacy trials.

The CoP objective focuses on studying serum markers as CoPs, because serum samples can be readily stored longitudinally in almost all study participants, and hence measured in almost all VRC01 recipients acquiring the HIV-1 infection primary endpoint and in a random sample of HIV-1 uninfected VRC01 recipient controls. Antibody in serum can be directly quantified unlike mucosal secretions collected by either wicks or washes that can be

confounded by difficulty defining the denominator for the measurement. Nevertheless, to understand mechanisms of protection it is important to consider how to study mucosal markers as CoPs. Antibodies could act in the epithelium, lamina propria, submucosa, and/or downstream in draining lymph nodes. Direct assessment of mucosal markers as CoPs in the AMP trials would arguably require taking tissue biopsies from almost all VRC01 group participants (to allow measuring the mucosal markers in almost all individuals who acquire an HIV-1 infection primary endpoint), which the protocol team deemed to be too onerous to ask of AMP trial participants. Therefore, the HVTN is implementing a separate phase 1 trial that administers four 8-weekly infusions of VRC01 at 10 mg/kg or 30 mg/kg to men and women at low risk of HIV-1 infection (N=23 participants per dose level), as well as a single 30 mg/kg VRC01 infusion for N=16 individuals. The men undergo rectal biopsies and the women undergo rectal and cervicovaginal biopsies. Tissue concentrations of VRC01, though at limited time-points, will be assayed in concert with serum concentrations. The identification of strong associations between mucosal and serum VRC01 markers would enable the investigation of mucosal CoPs in the AMP trials by using serum markers and other predictors to impute mucosal markers, and to use statistical techniques that account for the uncertainty in the imputations.

A challenge posed to the trials is the uncertainty in the projected HIV-1 incidence in the two control infusion groups. For the North/South American/Swiss MSM + TG trial the HIV-1 incidence projections are solidly grounded in data and experience except for one important factor—the fraction of study participants who receive PrEP and the level of adherence to PrEP. The trial facilitates free access to Tenofovir-emtricitabine PrEP for all study participants seeking it, and it is difficult to predict the pattern of PrEP use. The trial design allows for up to 50% PYRs during PrEP use at 90% assumed efficacy. If PrEP use is greater, then the trial would be at-risk of being prematurely completed due to an inability to answer the primary efficacy question in a timely manner. Likewise for the sub-Saharan African women trial, there are ample data and experience to project HIV-1 incidence results, except for uncertainty about the changing landscape of HIV-1 prevention packages in sub-Saharan African women, including on the extent of Tenofovir-emtricitabine PrEP use and the use of new interventions that could be discovered to be highly efficacious during the AMP trials. These factors are difficult to predict, and thus the trial design builds in some conservatism in its HIV-1 incidence assumptions.

In conclusion, the AMP trials, which are being conducted by the HIV Vaccine Trials Network and the HIV Prevention Trials Network in close partnership, constitute the first test in humans of the concept that a broadly neutralizing antibody can prevent HIV-1 infection. The AMP trials are designed to answer to this question, and, if the answer is affirmative, to learn about the immunological and virological correlates of protection and hence to guide next steps of HIV-1 vaccine research.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors would like to thank the protocol team of the AMP trials and the many groups and individuals dedicated to planning and conducting these trials, as well as the AMP trial participants. The authors also thank Lindsay Carpp for technical writing.

Funding sources: This work was supported by the National Institute of Allergy and Infectious Disease at the National Institutes of Health [2 R37 AI054165-11 to P.B.G., UM1 AI068635 to P.B.G., UM1 AI068617 to D.J.D., UM1 AI068619 to M.C., UM1 AI068618 to J.M., UM1 AI068614 to L.C.]. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

ABBREVIATIONS

ADCC	antibody-dependent cellular cytotoxicity
ADCP	antibody-dependent cellular phagocytosis
ADCVI	antibody-dependent cell-mediated viral inhibition
AMP	Antibody Mediated Prevention
ARVs	antiretroviral drugs
bnAb	broadly neutralizing antibody
CoP	correlate of protection
DSMB	Data Safety and Monitoring Board
IC50	50% inhibitory concentration
ITT	intention-to-treat
IV	intravenous
mAb	monoclonal antibody
MITT	modified intention-to-treat
MSM	men who have sex with men
NHP	non-human primate
PE	prevention efficacy
PEP	post-exposure prophylaxis
PK	pharmacokinetic
PrEP	pre-exposure prophylaxis
RSV	respiratory syncytial virus
TG	transgender persons
tMLE	targeted minimum loss based estimation

References

- Abdool, Karim Q., Abdool Karim, SS., Frohlich, JA., et al. Effectiveness and safety of tenofovir gel, an antiretroviral microbicide, for the prevention of HIV infection in women. *Science*. 2010; 329(5996):1168–1174. [PubMed: 20643915]
- AIDSinfo. Guidelines for the Use of Antiretroviral Agents in HIV-1-Infected Adults and Adolescents: Limitations to Treatment Safety and Efficacy: Adverse Effects of Antiretroviral Agents. 2014. Last Modified 5/1/2014 Accessed 2/17/2015. <http://aidsinfo.nih.gov/guidelines>
- Asmal M, Sun Y, Lane S, et al. Antibody-dependent cell-mediated viral inhibition emerges after simian immunodeficiency virus SIVmac251 infection of rhesus monkeys coincident with gp140-binding antibodies and is effective against neutralization-resistant viruses. *Journal of Virology*. 2011; 85(11):5465–5475. [PubMed: 21450829]
- Baba TW, Liska V, Hofmann-Lehmann R, et al. Human neutralizing monoclonal antibodies of the IgG1 subtype protect against mucosal simian-human immunodeficiency virus infection. *Nature Medicine*. 2000; 6(2):200–206.
- Baeten, Jared M., Donnell, Deborah, Ndase, Patrick, et al. Antiretroviral Prophylaxis for HIV Prevention in Heterosexual Men and Women. *New England Journal of Medicine*. 2012; 367(5):399–410. [PubMed: 22784037]
- Binley JM, Lybarger EA, Crooks ET, et al. Profiling the specificity of neutralizing antibodies in a large panel of plasmas from patients chronically infected with human immunodeficiency virus type 1 subtypes B and C. *Journal of Virology*. 2008; 82(23):11651–11668. [PubMed: 18815292]
- Burton DR, Stanfield RL, Wilson IA. Antibody vs. HIV in a clash of evolutionary titans. *Proceedings of the National Academy of Sciences*. 2005; 102(42):14943–14948.
- Caskey M, Klein F, Lorenzi JC, et al. Viraemia suppressed in HIV-1-infected humans by broadly neutralizing antibody 3BNC117. *Nature*. 2015; 522(7557):487–91. [PubMed: 25855300]
- Celum C, Wald A, Hughes J, et al. Effect of aciclovir on HIV-1 acquisition in herpes simplex virus 2 seropositive women and men who have sex with men: a randomised, double-blind, placebo-controlled trial. *Lancet*. 2008; 371(9630):2109–2119. [PubMed: 18572080]
- Centers for Disease Control. Pneumocystis pneumonia—Los Angeles. *Morbidity and Mortality Weekly Report*. 1981; 30(21):250–2. [PubMed: 6265753]
- Centers for Disease Control and Prevention. Estimated HIV incidence in the United States, 2007–2010. 2012. Last Modified 12/2012 Accessed 07/08. http://www.cdc.gov/hiv/pdf/statistics_hsr_vol_17_no_4.pdf
- Centers for Disease Control and Prevention. Diagnoses of HIV infection in the United States and dependent areas, 2013. 2015. Last Modified 02/23/2015 Accessed 07/08. http://www.cdc.gov/hiv/library/reports/surveillance/2013/surveillance_Report_vol_25.html
- Cohen MS, Chen YQ, McCauley M, et al. Prevention of HIV-1 infection with early antiretroviral therapy. *New England Journal of Medicine*. 2011; 365(6):493–505. [PubMed: 21767103]
- Cohen, Myron S., Chen, Ying Q., McCauley, Marybeth, et al. Antiretroviral Therapy for the Prevention of HIV-1 Transmission. *New England Journal of Medicine*. 2016; 0(0) null.
- Corti D, Langedijk JP, Hinz A, et al. Analysis of memory B cell responses and isolation of novel monoclonal antibodies with neutralizing breadth from HIV-1-infected individuals. *PLoS One*. 2010; 5(1):e8805. [PubMed: 20098712]
- Cottrell ML, Yang KH, Prince HM, et al. A Translational Pharmacology Approach to Predicting Outcomes of Preexposure Prophylaxis Against HIV in Men and Women Using Tenofovir Disoproxil Fumarate With or Without Emtricitabine. *J Infect Dis*. 2016; 214(1):55–64. [PubMed: 26917574]
- Crooks GE, Hon G, Chandonia JM, et al. WebLogo: a sequence logo generator. *Genome Research*. 2004; 14(6):1188–90. [PubMed: 15173120]
- Edlefsen, PT., Magaret, CA., Poon, A., et al. KeyStone Symposia on Molecular and Cellular Biology HIV Vaccines (X8), Olympic Valley, CA. 2016. Systematic evaluation of HIV infection time and founder inferences using SGA and Illumina sequences.

- Ellenberg, SS., Fleming, TR., DeMets, DL. Data Monitoring Committees in Clinical Trials: A Practical Perspective. Wiley; 2002. Reprint, Not in File
- Emanuel EJ, Miller FG. The ethics of placebo-controlled trials—a middle ground. *New England Journal of Medicine*. 2001; 345(12):915–9. [PubMed: 11565527]
- Emanuel EJ, Wendler D, Grady C. What makes clinical research ethical? *Journal of the American Medical Association*. 2000; 283(20):2701–11. [PubMed: 10819955]
- Falkowska E, Ramos A, Feng Y, et al. PGV04, an HIV-1 gp120 CD4 binding site antibody, is broad and potent in neutralization but does not induce conformational changes characteristic of CD4. *Journal of Virology*. 2012; 86(8):4394–4403. [PubMed: 22345481]
- Fauci AS, Folkers GK, Marston HD. Ending the global HIV/AIDS pandemic: the critical role of an HIV vaccine. *Clinical Infectious Diseases*. 2014; 59(Suppl 2):S80–4. [PubMed: 25151483]
- Fleming TR, Richardson BA. Some design issues in trials of microbicides for the prevention of HIV infection. *Journal of Infectious Diseases*. 2004; 190(4):666–674. [PubMed: 15272392]
- Florese RH, Demberg T, Xiao P, et al. Contribution of nonneutralizing vaccine-elicited antibody activities to improved protective efficacy in rhesus macaques immunized with Tat/Env compared with multigenic vaccines. *Journal of Immunology*. 2009; 182(6):3718–3727.
- Follmann D. Augmented designs to assess immune response in vaccine trials. *Biometrics*. 2006; 62(4): 1161–1169. [PubMed: 17156291]
- Fong Y, Di C, Permar S. Change point testing in logistic regression models with interaction term. *Statistics in Medicine*. 2015; 34(9):1483–94. [PubMed: 25612253]
- Forthal DN, Landucci G, Phan TB, et al. Interactions between natural killer cells and antibody Fc result in enhanced antibody neutralization of human immunodeficiency virus type 1. *Journal of Virology*. 2005; 79(4):2042–2049. [PubMed: 15681406]
- Freidlin B, Korn EL, Gray R. A general inefficacy interim monitoring rule for randomized clinical trials. *Clinical Trials*. 2010; 7(3):197–208. [PubMed: 20423925]
- Gabriel EE, Gilbert PB. Evaluating principal surrogate endpoints with time-to-event data accounting for time-varying treatment efficacy. *Biostatistics*. 2014; 15(2):251–65. [PubMed: 24337534]
- Georgiev IS, Gordon Joyce M, Zhou T, et al. Elicitation of HIV-1-neutralizing antibodies against the CD4-binding site. *Current Opinion in HIV and AIDS*. 2013; 8(5):382–392. [PubMed: 23924998]
- Gilbert PB, Gabriel EE, Miao X, et al. Fold rise in antibody titers by measured by glycoprotein-based enzyme-linked immunosorbent assay is an excellent correlate of protection for a herpes zoster vaccine, demonstrated via the vaccine efficacy curve. *Journal of Infectious Diseases*. 2014; 210(10):1573–81. [PubMed: 24823623]
- Gilbert PB, Janes HE, Huang Y. Power/sample size calculations for assessing correlates of risk in clinical efficacy trials. *Stat Med*. 2016; 35(21):3745–59. [PubMed: 27037797]
- Gilbert P, Wang M, Wrin T, et al. Magnitude and breadth of a nonprotective neutralizing antibody response in an efficacy trial of a candidate HIV-1 gp120 vaccine. *Journal of Infectious Diseases*. 2010; 202(4):595–605. [PubMed: 20608874]
- Gilbert PB. Some design issues in phase 2B vs phase 3 prevention trials for testing efficacy of products or concepts. *Statistics in Medicine*. 2010; 29(10):1061–1071. [PubMed: 20419758]
- Gilbert PB, Grove D, Gabriel E, et al. A Sequential Phase 2b Trial Design for Evaluating Vaccine Efficacy and Immune Correlates for Multiple HIV Vaccine Regimens. *Statistical Communications in Infectious Diseases*. 2011; 3(1)
- Gilbert PB, Hudgens MG. Evaluating candidate principal surrogate endpoints. *Biometrics*. 2008; 64(4):1146–1154. [PubMed: 18363776]
- Graham BS, Ambrosino DM. History of passive antibody administration for prevention and treatment of infectious diseases. *Curr Opin HIV AIDS*. 2015; 10(3):129–34. [PubMed: 25760933]
- Grant RM, Lama JR, Anderson PL, et al. Preexposure chemoprophylaxis for HIV prevention in men who have sex with men. *New England Journal of Medicine*. 2010; 363(27):2587–2599. [PubMed: 21091279]
- Gray ES, Madiga MC, Hermanus T, et al. The neutralization breadth of HIV-1 develops incrementally over four years and is associated with CD4+ T cell decline and high viral load during acute infection. *Journal of Virology*. 2011a; 85(10):4828–4840. [PubMed: 21389135]

- Gray ES, Taylor N, Wycuff D, et al. Antibody specificities associated with neutralization breadth in plasma from human immunodeficiency virus type 1 subtype C-infected blood donors. *Journal of Virology*. 2009; 83(17):8925–8937. [PubMed: 19553335]
- Gray GE, Allen M, Moodie Z, et al. Safety and efficacy of the HVTN 503/Phambili study of a clade-B-based HIV-1 vaccine in South Africa: a double-blind, randomised, placebo-controlled test-of-concept phase 2b study. *Lancet Infectious Diseases*. 2011b; 11(7):507–515. [PubMed: 21570355]
- Gray GE, Moodie Z, Metch B, et al. Recombinant adenovirus type 5 HIV gag/pol/nef vaccine in South Africa: unblinded, long-term follow-up of the phase 2b HVTN 503/Phambili study. *Lancet Infect Dis*. 2014; 14(5):388–396. [PubMed: 24560541]
- Hall HI, Song R, Rhodes P, et al. Estimation of HIV incidence in the United States. *Journal of the American Medical Association*. 2008; 300(5):520–529. [PubMed: 18677024]
- Hammer SM, Sobieszczyk ME, Janes H, et al. Efficacy trial of a DNA/rAd5 HIV-1 preventive vaccine. *New England Journal of Medicine*. 2013; 369(22):2083–2092. [PubMed: 24099601]
- Haynes BF, Gilbert PB, McElrath MJ, et al. Immune-correlates analysis of an HIV-1 vaccine efficacy trial. *New England Journal of Medicine*. 2012; 366(14):1275–1286. [PubMed: 22475592]
- Hessell AJ, Rakasz EG, Poignard P, et al. Broadly neutralizing human anti-HIV antibody 2G12 is effective in protection against mucosal SHIV challenge even at low serum neutralizing titers. *PLoS Pathogens*. 2009; 5(5):e1000433. [PubMed: 19436712]
- Hofmann-Lehmann R, Vlasak J, Rasmussen RA, et al. Postnatal pre- and postexposure passive immunization strategies: protection of neonatal macaques against oral simian-human immunodeficiency virus challenge. *Journal of Medical Primatology*. 2002; 31(3):109–119. [PubMed: 12190851]
- Huang J, Ofek G, Laub L, et al. Broad and potent neutralization of HIV-1 by a gp41-specific human antibody. *Nature*. 2012; 491(7424):406–412. [PubMed: 23151583]
- Huang Y, Gilbert PB, Montefiori DC, et al. Simultaneous Evaluation of the Magnitude and Breadth of a Left and Right Censored Multivariate Response, with Application to HIV Vaccine Development. *Statistics in Biopharmaceutical Research*. 2009; 1(1):81–91. [PubMed: 20072667]
- Huang Y, Gilbert PB, Wolfson J. Design and estimation for evaluating principal surrogate markers in vaccine trials. *Biometrics*. 2013; 69(2):301–309. [PubMed: 23409839]
- Janes H, Gilbert P, Buchbinder S, et al. In pursuit of an HIV vaccine: designing efficacy trials in the context of partially effective nonvaccine prevention modalities. *AIDS Research and Human Retroviruses*. 2013; 29(11):1513–23. [PubMed: 23597282]
- Kalbfleisch, J., Prentice, R. *The Statistical Analysis of Failure Time Data*. Second. Hoboken, New Jersey: John Wiley & Sons, Inc; 2002.
- Ko SY, Pegu A, Rudicell RS, et al. Enhanced neonatal Fc receptor function improves protection against primate SHIV infection. *Nature*. 2014; 514(7524):642–645. [PubMed: 25119033]
- Koblin BA, Mayer KH, Eshleman SH, et al. Correlates of HIV acquisition in a cohort of Black men who have sex with men in the United States: HIV prevention trials network (HPTN) 061. *PLoS One*. 2013; 8(7):e70413. [PubMed: 23922989]
- Kong L, Lee JH, Doores KJ, et al. Supersite of immune vulnerability on the glycosylated face of HIV-1 envelope glycoprotein gp120. *Nature Structural & Molecular Biology*. 2013; 20(7):796–803.
- Kwong PD, Mascola JR. Human antibodies that neutralize HIV-1: identification, structures, and B cell ontogenies. *Immunity*. 2012; 37(3):412–425. [PubMed: 22999947]
- Ledgerwood JE, Coates EE, Yamshchikov G, et al. Safety, pharmacokinetics and neutralization of the broadly neutralizing HIV-1 human monoclonal antibody VRC01 in healthy adults. *Clinical & Experimental Immunology*. 2015; 182(3):289–301. [PubMed: 26332605]
- Li Y, Migueles SA, Welcher B, et al. Broad HIV-1 neutralization mediated by CD4-binding site antibodies. *Nature Medicine*. 2007; 13(9):1032–1034.
- Lynch RM, Boritz E, Coates EE, et al. Virologic effects of broadly neutralizing antibody VRC01 administration during chronic HIV-1 infection. *Science Translational Medicine*. 2015; 7(319):319ra206.
- Marrazzo JM, Ramjee G, Richardson BA, et al. Tenofovir-Based Preexposure Prophylaxis for HIV Infection among African Women. *New England Journal of Medicine*. 2015; 372:509–518. [PubMed: 25651245]

- Mayer K, Seaton K, Huang Y, et al. Clinical safety and pharmacokinetics of IV and SC VRC01, a broadly neutralizing mAb. CROI, Boston. 2016
- McCormack, Sheena, Dunn, David T., Desai, Monica, et al. Pre-exposure prophylaxis to prevent the acquisition of HIV-1 infection (PROUD): effectiveness results from the pilot phase of a pragmatic open-label randomised trial. *The Lancet*. 2016; 387(10013):53–60.
- McLellan JS, Pancera M, Carrico C, et al. Structure of HIV-1 gp120 V1/V2 domain with broadly neutralizing antibody PG9. *Nature*. 2011; 480(7377):336–343. [PubMed: 22113616]
- Meier AS, Gilbert PB. Accuracy and precision of estimating intervention efficacy when the timing of observed events differ by treatment arm. *Contemporary Clinical Trials*. 2005; 26(5):598–610. [PubMed: 16112914]
- Moldt B, Rakasz EG, Schultz N, et al. Highly potent HIV-specific antibody neutralization in vitro translates into effective protection against mucosal SHIV challenge in vivo. *Proceedings of the National Academy of Sciences*. 2012; 109(46):18921–18925.
- Molina, Jean-Michel, Capitant, Catherine, Spire, Bruno, et al. On-Demand Preexposure Prophylaxis in Men at High Risk for HIV-1 Infection. *New England Journal of Medicine*. 2015; 373(23):2237–2246. [PubMed: 26624850]
- Montefiori DC. Importance of neutralization sieve analyses when seeking correlates of HIV-1 vaccine efficacy. *Human Vaccines & Immunotherapeutics*. 2014; 10(8):2507–11. [PubMed: 25424964]
- Moore KL, van der Laan MJ. Application of time-to-event methods in the assessment of safety in clinical trials. U C Berkeley Division of Biostatistics Working Paper Series, Paper 248. 2009
- NIAID. Daily-Use HIV Prevention Approaches Prove Ineffective Among Women in NIH Study. 2013. Last Modified 3/4/2013. <http://www.niaid.nih.gov/news/newsreleases/2013/Pages/VOICEwomen.aspx>
- Nishimura Y, Igarashi T, Haigwood N, et al. Determination of a statistically valid neutralization titer in plasma that confers protection against simian-human immunodeficiency virus challenge following passive transfer of high-titered neutralizing antibodies. *Journal of Virology*. 2002; 76(5):2123–2130. [PubMed: 11836389]
- Null D, Bimle C, Weisman L, et al. Palivizumab, a humanized respiratory syncytial virus monoclonal antibody, reduces hospitalization from respiratory syncytial virus infection in high-risk infants. *Pediatrics*. 1998; 102(3):531–537.
- Padian NS, van der Straten A, Ramjee G, et al. Diaphragm and lubricant gel for prevention of HIV acquisition in southern African women: a randomised controlled trial. *Lancet*. 2007; 370(9583):251–261. [PubMed: 17631387]
- Patterson KB, Prince HA, Kraft E, et al. Penetration of tenofovir and emtricitabine in mucosal tissues: implications for prevention of HIV-1 transmission. *Sci Transl Med*. 2011; 3(112):112re4.
- Pegu A, Yang ZY, Boyington JC, et al. Neutralizing antibodies to HIV-1 envelope protect more effectively in vivo than those to the CD4 receptor. *Science Translational Medicine*. 2014; 6(243):243ra88.
- Plotkin SA, Gilbert PB. Nomenclature for immune correlates of protection after vaccination. *Clinical Infectious Diseases*. 2012; 54(11):1615–1617. [PubMed: 22437237]
- Prejean J, Song R, Hernandez A, et al. Estimated HIV incidence in the United States, 2006–2009. *PLoS One*. 2011; 6(8):e17502. [PubMed: 21826193]
- Qin L, Gilbert PB, Follmann D, et al. Assessing Surrogate Endpoints In Vaccine Trials With Case-Cohort Sampling and the Cox Model. *The Annals of Applied Statistics*. 2008; 2(1):386–407. [PubMed: 19079758]
- Rees H, Delany-Moretlwa S, Lombard C, et al. FACTS 001 Phase III Trial of Pericoital Tenofovir 1% Gel for HIV Prevention in Women. CROI, Seattle, WA. 2015
- Rida W, Fast P, Hoff R, et al. Intermediate-size trials for the evaluation of HIV vaccine candidates: a workshop summary. *Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology*. 1997; 16(3):195–203. [PubMed: 9390572]
- Sather DN, Armann J, Ching LK, et al. Factors associated with the development of cross-reactive neutralizing antibodies during human immunodeficiency virus type 1 infection. *Journal of Virology*. 2009; 83(2):757–769. [PubMed: 18987148]

- Scheid JF, Mouquet H, Ueberheide B, et al. Sequence and structural convergence of broad and potent HIV antibodies that mimic CD4 binding. *Science*. 2011; 333(6049):1633–1637. [PubMed: 21764753]
- Schneider TD, Stephens RM. Sequence logos: a new way to display consensus sequences. *Nucleic Acids Research*. 1990; 18(20):6097–100. [PubMed: 2172928]
- Shibata R, Igarashi T, Haigwood N, et al. Neutralizing antibody directed against the HIV-1 envelope glycoprotein can completely block HIV-1/SIV chimeric virus infections of macaque monkeys. *Nature Medicine*. 1999; 5(2):204–210.
- Shingai M, Donau OK, Plishka RJ, et al. Passive transfer of modest titers of potent and broadly neutralizing anti-HIV monoclonal antibodies block SHIV infection in macaques. *Journal of Experimental Medicine*. 2014; 211(10):2061–2074. [PubMed: 25155019]
- Simek MD, Rida W, Priddy FH, et al. Human immunodeficiency virus type 1 elite neutralizers: individuals with broad and potent neutralizing activity identified by using a high-throughput neutralization assay together with an analytical selection algorithm. *Journal of Virology*. 2009; 83(14):7337–7348. [PubMed: 19439467]
- Stamatatos L, Morris L, Burton DR, et al. Neutralizing antibodies generated during natural HIV-1 infection: good news for an HIV-1 vaccine? *Nature Medicine*. 2009; 15(8):866–870.
- Tudor D, Derrien M, Diomedea L, et al. HIV-1 gp41-specific monoclonal mucosal IgAs derived from highly exposed but IgG-seronegative individuals block HIV-1 epithelial transcytosis and neutralize CD4(+) cell infection: an IgA gene and functional analysis. *Mucosal Immunology*. 2009; 2(5):412–426. [PubMed: 19587640]
- UNAIDS. Report on the Global AIDS Epidemic, 2012 country progress reports. 2013. Accessed 2/20/2013. <http://www.unaids.org/en/dataanalysis/knowyourresponse/countryprogressreports/2012countries/>
- UNAIDS. The GAP Report. 2014. http://www.unaids.org/en/media/unaids/contentassets/documents/unaidspublication/2014/UNAIDS_Gap_report_en.pdf
- Van Damme L, Corneli A, Ahmed K, et al. Preexposure prophylaxis for HIV infection among African women. *New England Journal of Medicine*. 2012; 367(5):411–422. [PubMed: 22784040]
- Van Damme L, Corneli A, Ahmed K, et al. Preexposure Prophylaxis for HIV Infection among African Women. *New England Journal of Medicine*. 2012; 367(5):411–422. [PubMed: 22784040]
- van der Laan, MJ., Rose, S. Targeted Learning: Causal Inference for Observational and Experimental Data. Springer; 2011. Reprint, Not in File
- van der Straten A, Van Damme L, Haberer JE, et al. Unraveling the divergent results of pre-exposure prophylaxis trials for HIV prevention. *AIDS*. 2012; 26(7):F13–F19. [PubMed: 22333749]
- Walker LM, Burton DR. Rational antibody-based HIV-1 vaccine design: current approaches and future directions. *Current Opinion in Immunology*. 2010; 22(3):358–366. [PubMed: 20299194]
- Walker LM, Huber M, Doores KJ, et al. Broad neutralization coverage of HIV by multiple highly potent antibodies. *Nature*. 2011; 477(7365):466–470. [PubMed: 21849977]
- Walker LM, Phogat SK, Chan-Hui PY, et al. Broad and potent neutralizing antibodies from an African donor reveal a new HIV-1 vaccine target. *Science*. 2009; 326(5950):285–289. [PubMed: 19729618]
- Walker LM, Simek MD, Priddy F, et al. A limited number of antibody specificities mediate broad and potent serum neutralization in selected HIV-1 infected individuals. *PLoS Pathogens*. 2010; 6(8):e1001028. [PubMed: 20700449]
- WHO. WHO expands recommendation on oral pre-exposure prophylaxis of HIV infection (PrEP). Policy brief. 2015; 48:2. WHO/HIV/2015.
- WHO in partnership with UNICEF and UNAIDS. Global Report on HIV Treatment 2013: Results, Impact and Opportunities. 2013. Last Modified June 2013. http://apps.who.int/iris/bitstream/10665/85326/1/9789241505734_eng.pdf
- World Medical Association. Declaration of Helsinki: ethical principles for medical research involving human subjects. Note of Clarification on Paragraph 29. 2002 Accessed 07/06/2015.
- Wu X, Yang ZY, Li Y, et al. Rational design of envelope identifies broadly neutralizing human monoclonal antibodies to HIV-1. *Science*. 2010; 329(5993):856–861. [PubMed: 20616233]

Zhou T, Georgiev I, Wu X, et al. Structural basis for broad and potent neutralization of HIV-1 by antibody VRC01. *Science*. 2010; 329(5993):811–7. [PubMed: 20616231]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

	10 mg/kg	30 mg/kg	Overlap
High (>50 mcg/mL)	10% PYRs	50% PYRs	10% PYRs
Medium (10 to 50 mcg/mL)	40% PYRs	40% PYRs	40% PYRs
Low (<10 mcg/mL)	50% PYRs	10% PYRs	10% PYRs
Total overlap = 60% PYRs			

Figure 1.

Projected overlap in serum VRC01 concentrations and person-years at risk (PYRs) for HIV-1 acquisition. 30 mg/kg and 10 mg/kg doses of VRC01 administration are compared. Under perfect infusion adherence, 100% PYRs > 1 mcg/mL for both dose groups, 91% PYRs > 4 mcg/mL for 10 mg/kg and 100% PYRs > 4 mcg/mL for 30 mg/kg doses.

Figure 2.

Predicted VRC01 median serum concentration over time (solid lines) for (A) 10 mg/kg and (B) 30 mg/kg IV infusions of VRC01 at Weeks 0, 8, 16, 24, 32, 40, 48, 56, 64, and 72, with 90% confidence intervals (CIs) (dashed lines) based on a population pharmacokinetic analysis from VRC602 (n=23 VRC01 recipients, using a standard two compartment pharmacokinetic model developed with the program NONMEM). The horizontal lines at 10 and 50 mcg/mL denote the boundaries of the low, medium, and high ranges of the serum concentration marker $S(t)$. PYRs = person-years at risk for HIV-1 acquisition.

Figure 3.

Serum concentration of VRC01 (mcg/mL) measured by ELISA for 24 participants in the HVTN 104 phase 1 trial. (A) 10 mg/kg dose (N=12 participants received the first infusion; N=11 received the second infusion 8 weeks later). (B) 30 mg/kg dose (N=12 received the first infusion; N=9 received the second infusion 8 weeks later). The x-axis labels indicate days post-enrollment of concentration measurements, centered at the expected measurement times.

Figure 4.

Modeled cumulative incidence of HIV-1 infection (VRC01 dose groups pooled vs. control) based on 5000 simulated data sets of the two AMP trials. HIV infection times were simulated as a function of time since prior infusion via a Cox regression model assuming exponential infection times and a 3.0% (5.5%) annual control group incidence in the Americas/Switzerland (Africa) trial. The Cox model coefficients were determined numerically by setting PE = 40% in the 10 mg/Kg dose group and PE = 80% in the 30 mg/Kg dose group, assuming that the instantaneous hazard ratio (VRC01/control) is higher when the current concentration of VRC01 is lower. Participants were assumed to have received all 10 infusions 8-weekly. The expected VRC01 serum concentrations were calculated based on modeling of the observed HVTN 104 data using a two-compartment pharmacokinetics model with an estimated clearance rate of 0.018 Liter/hour, volume of distribution in the central compartment of 5.97 Liter, volume of distribution in the peripheral compartment of 3.07 Liter, and inter-compartment clearance rate of 0.012 Liter/hour. Presented geometric mean concentrations are after 5 infusions of VRC01 in the 10 mg/kg and 30 mg/kg dose groups for a 70 kg participant. AMP, Antibody Mediated Protection; PE, prevention efficacy.

Figure 5.

Total sample size for 90% power as a function of prevention efficacy. The total sample size for 90% power to reject $H_0: PE = 0\%$ in favor of $H_1: PE > 0\%$ in a 1:1:1 allocation design with 80 weeks of follow-up for HIV-1 infection (1-sided $\alpha = 0.025$ test) is shown as a function of prevention efficacy. Upper green line = North/South American/Swiss MSM + TG trial; lower orange line = sub-Saharan African women trial. MSM, men who have sex with men; PE, prevention efficacy; TG, transgender persons.

Figure 6.

HIV-1 incidence in the control infusions group as a function of person-years at-risk under PrEP use. The annual HIV-1 incidence in the North/South American/Swiss MSM + TG control infusions group, assuming 4% annual HIV-1 incidence in the hypothetical scenario of no PrEP use, is shown under different levels of PrEP use and PrEP efficacy. MSM, men who have sex with men; PrEP, pre-exposure prophylaxis; TG, transgender persons.

Figure 7.

Probabilities of stopping early in the North/South American/Swiss MSM + TG trial. The probabilities of stopping early for potential harm, non-efficacy, and high efficacy are shown. The probability of having unconditional power to reject the null hypothesis $H_0: PE = 0\%$ in favor of the alternative hypothesis $H_1: PE > 0\%$ (1-sided $\alpha = 0.025$ test) is also shown. MSM, men who have sex with men; PE, prevention efficacy; TG, transgender persons.

2-phase logistic regression; 2-sided alpha = 0.05; controls:cases = 5:1

Figure 8.

Power curves for anticipated fixed PYRs at risk and PE. Power of a CoP analysis for detecting that PE varies over the 3 VRC01 marker-defined subgroups $S(t)=\text{Low}$, $S(t)=\text{Medium}$, and $S(t)=\text{High}$ at levels PE = 30%, 60%, and 80%, respectively, for a marker with assay noise parameter $\rho = 1$ (perfect marker), 0.9, 0.7, and 0.5 and percentages of person-years at risk in the (Low, Medium, High) marker regions set at (30%, 40%, 30%). The calculations assume 61 VRC01 group HIV-1-infected cases and 305 VRC01 group HIV-1-uninfected controls with data on $S(t)$ and pool over both VRC01 groups and over both trials. The dashed vertical line represents the sample size of the two AMP trials combined. AMP, Antibody Mediated Prevention; CoP, correlate of protection; PE, prevention efficacy; PYRs, person-years at risk.

2-phase logistic regression; 2-sided alpha = 0.05; controls:cases = 5:1

Figure 9.

Power curves for anticipated PYRs at risk and PE with rho fixed at 0.9. Power of a CoP analysis for detecting that PE varies over the 3 VRC01 marker-defined subgroups $S(t)$ = Low, $S(t)$ = Medium, and $S(t)$ = High at levels PE = 30%, 60%, and 80%, respectively, for a marker with assay noise parameter rho = 0.9 and percentages of person-years at risk in the Low and High marker regions varied from 10%, 15%, 20%, 25%, 30%. The calculations assume 61 VRC01 group HIV-1-infected cases and 305 VRC01 group HIV-1-uninfected controls with data on $S(t)$ and pool over both VRC01 groups and over both trials. CoP, correlate of protection; PE, prevention efficacy; PYRs, person-years at risk.

Figure 10.

Hypothetical illustration of genotypic and phenotypic sieve analysis. (A) Distribution of the number of known neutralization escape mutations in breakthrough control and VRC01 group HIV-1 sequences. (B) Percentage of the panel of breakthrough HIV-1s that are neutralized at level 50% (IC_{50}) by the given concentration of VRC01. This illustration pools data across both trials.

Figure 11.

Power calculations of phenotypic sieve analysis. IC₅₀ distributions are compared between VRC01 group breakthrough HIV-1s and control group breakthrough HIV-1s, with difference equivalently expressed in terms of (A) the PE(x) curve or (B) the cumulative IC₅₀ distribution functions. In (B), the black/highest curve represents the distribution for the control group breakthrough HIV-1s. (C) Power of the Wilcoxon test. The scenario when PE₁₀=PE₃₀=60% and pooling over the two VRC01 groups and the two trials is highlighted dark blue (with 74 control group cases). IC₅₀, 50% inhibitory concentration; PE, prevention efficacy.

Table 1

Summary of the statistical approach to sequential monitoring of PE for each AMP trial

Monitoring Type	Hypotheses	Testing Approach	Size of Test	Stopping Boundary	Timing of Interim Analyses
Potential Harm	H0: PE = 0% vs. H1: PE < 0%	Exact 1-sided binomial test of the fraction of infections assigned to receive VRC01	Specified in the monitoring plan	Near-constant 1-sided p-value cut-off controlling the FWER at $\alpha = 0.05$	After every infection from 20 th total until first non-efficacy analysis
Non-Efficacy	H0: PE = 40% vs. H1: PE < 40%	2-sided Wald test and 1:1 corresponding 2-sided 95% CI	1-sided $\alpha = 0.025$	Nominal 2-sided 95% CI below 40% ^a supplemented with conditional power calculations	6-monthly starting based on an event-driven trigger
High Efficacy	H0: PE = 70% vs. H1: PE > 70%	2-sided Wald test and 1:1 corresponding 2-sided 95% CI	1-sided $\alpha = 0.025$	Nominal 2-sided 95% CI above 70%	6-monthly starting at the same time as for non-efficacy

^aBased on the approach of (Freidlin, Korn, and Gray 2010).

Table 2

Cumulative number of HIV-1 infections over 80 weeks of follow-up for the two trials.

Trial	PE10 = PE30	Weeks since first person in	Median number of control infusions group endpoints	Median number of pooled VRC01 group endpoints
Americas/Switzerland	0%/60%	0	0/0	0/0
	0%/60%	52	3/3	7/3
	0%/60%	104	16/16	32/13
	0%/60%	156	25/30	50/24
	0%/60%	208	26/37	55/30
	0%/60%	219	26/37	55/30
Sub-Saharan. Africa	0%/60%	0	0/0	0/0
	0%/60%	52	4/4	8/3
	0%/60%	104	17/17	34/14
	0%/60%	156	26/32	51/26
	0%/60%	208	26/37	54/31
	0%/60%	210	26/37	54/31

Median cumulative endpoints for the two AMP trials under the null hypothesis PE10 = PE30 = 0% and under the design alternative hypothesis PE = PE10 = PE30 = 60%. Medians are computed based on simulating 10,000 efficacy trials under protocol assumptions and study monitoring and potential for early stopping. More control group infections occur for the 60% PE scenario than the 0% PE scenario because fewer trials are completed early.

Table 3

Power comparison of the 10 mg/kg versus 30 mg/kg VRC01 groups.

Trial	PE10	PE30	Power to Select Best VRC01 Dose Group ($\times 100$)^a	Power to Detect PE30 Higher than PE10 ($\times 100$)^b
Sub-Saharan Africa^c	20%	60%	88	57
	30%	60%	86	39
	40%	60%	81	21
	20%	70%	97	83
	30%	70%	97	68
	30%	80%	99.7	91
	40%	85%	99.9	91
	50%	85%	99.7	77
Trials Pooled	20%	60%	99	87
	30%	60%	98	68
	40%	60%	95	40
	20%	70%	>99.9	98
	30%	70%	>99.9	93
	50%	85%	>99.9	93

Power to select the best VRC01 group and power to conclude superiority of the 30 mg/kg dose group, for the Sub-Saharan African women trial and pooled over the two trials are shown.

^aProbability that the VRC01 dose group that is truly better (30 mg/kg group) is correctly selected based on the point estimate for PE30 exceeding that for PE10. Specifically, correct selection entails that the estimated PE30 is greater than the estimated PE10 and PE30 is significantly > 0% (1-sided $\alpha = 0.025$).

^bCumulative hazard-based Wald tests comparing HIV-1 incidence between the two VRC01 groups (1-sided $\alpha = 0.025$). Rejection also requires that PE30 is significantly > 0% (1-sided $\alpha = 0.025$).

^cPower is almost identical for the North/South American/Swiss MSM + TG trial.