Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Contemporary Clinical Trials 27 (2006) 147 – 160
Contemporary Clinical Trials
www.elsevier.com/locate/conclintrial

Statistical evaluation of HIV vaccines in early clinical trials

Zoe Moodie [a,*], A.J. Rossini [a], Michael G. Hudgens [b], Peter B. Gilbert [a], Steven G. Self [a], Nina D. Russell [c]

[a] *Statistical Center for HIV/AIDS Research and Prevention, Fred Hutchinson Cancer Reasearch Center, 1100 Fairview Ave N, LE-400, PO Box 19024, Seattle, WA 98109-1024, USA*
[b] *Department of Biostatistics, University of North Carolina at Chapel Hill, USA*
[c] *Clinical Research Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA*

Received 30 September 2004; accepted 22 November 2005

Abstract

The HIV pandemic is a pressing threat to global public health; HIV vaccine development is critical. Clinical evaluation of HIV vaccine candidates differs from the standard therapeutics trial framework primarily due to the fact that healthy individuals are studied. We present an early stage evaluation program developed for the HIV Vaccine Trials Network (HVTN) motivated by characteristics unique to the vaccine setting. The program consists of 3 prototypical stages (Phase I, Ib, II) that provide a unified yet flexible approach to the safety and immunogenicity evaluation of diverse vaccine regimens. The goal of these early trials is to narrow the number of candidate vaccines to the most promising candidates worthy of further study in efficacy trials.
© 2005 Elsevier Inc. All rights reserved.

Keywords: Phase I/II clinical trials; HIV vaccine; Trial design; Power

1. Introduction

The HIV pandemic represents a monumental challenge to global public health with almost 40 million people living with HIV and over 95% of the 5 million annual incident infections occurring in low or middle income countries where access to treatment is often extremely limited or non-existent [1]. As a result, an estimated 3.1 million deaths were attributed to AIDS in 2004 [1]. Improved access to treatment and prevention is critical, as is the need for an affordable preventive HIV vaccine with proven safety and efficacy to prevent acquisition of HIV or to delay progression and reduce infectiousness of the virus after HIV infection [2,3]. Recent HIV vaccine research initiatives have led to an increase in the number of HIV vaccine products entering clinical trials testing. There are now at least 13 different products at various stages of clinical trials [4] due in part to product development from pharmaceutical companies (e.g. Merck, Chiron) and organizations such as the Vaccine Research Center at the National Institute of Allergies and Infectious Disease (NIAID), National Institutes of Health (NIH) and the International AIDS Vaccine Initiative (IAVI).

Vaccine trials differ from the therapeutic trials and call for a different approach to design and analysis. Vaccine trial participants are drawn from a population of healthy volunteers; recruitment, retention, and follow-up can be more

* Corresponding author. Tel.: +1 206 667 7077; fax: +1 206 667 6888.
E-mail address: zoe@scharp.org (Z. Moodie).

1551-7144/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.cct.2005.11.008

challenging than in therapeutic treatments trials, where volunteers may be more motivated to participate. Another deterrent to participation is that vaccine trial volunteers may subsequently test positive for HIV due to receiving vaccine in the trial. Out of concern that the vaccine will instill a false sense of protection in participants, risk reduction counselling and blinding via placebo arms play essential roles in the trial.

The scientific questions addressed in HIV vaccine trials are complex. HIV vaccine researchers face eminently larger challenges than those of many other vaccine researchers in that genetic mutation and recombination lead to the continual evolution of the virus. At least 10 genetic subtypes of HIV-1 exist within a single group and viruses within subtypes are highly diverse; whether this genetic diversity translates into multiple immunotypes is unknown. Contrast this with the polio virus, which has only three different serotypes. The significance of strain variation within individuals and among populations must be estimated both to develop vaccines and to design trials, particularly later stage trials. Ideally, a preventive vaccine will generate immune responses that protect against all genetic variants of HIV to which an individual might be exposed. Thus research is being done to identify immunogenic conserved regions of HIV genes that are common to all or most subtypes. If these do not exist, a vaccine may need to contain multiple proteins or peptides (i.e., sequences of two or more amino acids) from different HIV isolates. Unlike certain viruses with successful vaccines (e.g., polio, rabies, yellow fever), HIV can exist in the host as free virus as well as within infected cells [5]. This represents one of the most difficult challenges; HIV targets the immune system itself, incorporating its genetic material into that of the host cell such that upon reproduction each new cell also contains HIV genes [6]. Thus the virus can hide its genetic material for a lengthy period of time. In addition, HIV reservoirs in other cells harbor intact viruses that can remain undetected by the immune system [7].

Given these challenges, a fully effective HIV vaccine may not be found. There is an urgent drive to evaluate and optimize many different vaccine regimens in Phase I and II trials to move the most promising forward to efficacy evaluation. As the mode of action of a successful HIV vaccine is still uncertain, early trials must assess multiple types of immunologic endpoints. HIV vaccine trials require standardized, focused, yet flexible methods to adapt to new scientific findings, such as novel technologies for measurement of immune response endpoints or insights on viral diversity. Evaluation of several vaccines in parallel can hasten progress.

The HIV Vaccine Trials Network (HVTN) is an NIH-funded, international clinical trials cooperative group that provides a common evaluation platform for HIV vaccine candidates developed in diverse settings ranging from academic laboratories to large pharmaceutical companies. The statistical group within the HVTN is responsible for the design and analysis of all HVTN trials and hence has defined a common evaluation platform to streamline trial design and facilitate analysis and cross-protocol summaries. In this paper, we describe prototypic designs for early clinical trials testing of HIV vaccines. These designs have become the standard for the HVTN's clinical trials' evaluation program. We begin by briefly highlighting different types of HIV vaccines that might be evaluated in early trials followed by a description of the clinical trials' endpoints typically used in the evaluation of HIV vaccines. Next, we outline the objectives, proposed designs and statistical considerations of Phase I and II HIV vaccine clinical trials and conclude with a discussion of our approach.

2. Vaccine candidates

A multitude of HIV vaccine candidates – varying by type of product, subtype of inserts, and mode of delivery–are now entering or about to enter clinical trials. Recombinant subunit protein vaccines were the first class of HIV immunogen to be tested in humans. Other approaches developed subsequently include viral and bacterial vector-based vaccines, DNA vaccines, peptide-based vaccines, pseudovirions and replicons, and combinations of these. Additionally, both traditional and novel adjuvants, including cytokine-based approaches, are more recent critical variables in multiple trials.

Given the difficulty in obtaining strong, durable immune responses with single modality vaccines, combination approaches to HIV immunization have attracted increasing attention. Immune responses to DNA alone have been relatively weak in humans but prime-boost approaches including adjuvants and recombinant viral vectors have resulted in appreciable HIV-specific $CD8^+$ T cell responses [8]. Combination vaccine regimens generally consist of two components, each seeking to elicit a different type of immune response: cellular or humoral or, less commonly, mucosal. The Food and Drug Administration (FDA) generally requires the establishment of safety of each vaccine component before allowing testing of both components in combination [9].

HIV vaccine candidates within a given class often differ by the subtype and number of their HIV gene inserts. Some manufacturers have focused on genes of different subtypes (e.g., Merck's clade B Gag adenovirus vaccine [8]) while others have included genes of different subtypes (e.g., VRC's multiclade Env/clade B Gag–Pol adenovirus vaccine [10]). Comparative data are not yet available to establish whether an approach which targets a single subtype is more or less immunogenic than a broader approach involving antigens from multiple subtypes.

Vaccination schedules, routes of administration, and modes of delivery are additional variables undergoing evaluation in clinical trials. The number and timing of injections varies depending on the vaccine mechanism of action and its anticipated immunogenicity. The most common route of administration is traditional intramuscular delivery, however intradermal and subcutaneous routes are sometimes considered. Additionally, intramuscular injection by needle is currently being compared to a needle-free intramuscular injection system for DNA-based vaccines (due to evidence with malaria vaccines of improved transgene-specific T cell responses with this approach [11–14]).

With such diversity of vaccine candidates, both across and within classes, a standardized approach is clearly needed to assess and compare the merits of each. Certain HIV vaccine categories, such as DNA vaccines and viral vector vaccines, are represented by a large number of candidates that are similar in their basic structure and mechanism of action. They are being developed and tested concurrently and standardized assessments are required to assess differences across multiple candidates within a particular class, as well as to make comparisons across classes. The next section describes the endpoints used by the HVTN to evaluate the performance of vaccine regimens in early stage clinical trials.

3. Endpoints in Phase I/II trials

Endpoints must be clearly defined and similar across trials to facilitate cross-trial data summary. In Phase I and II trials of HIV vaccine candidates, the primary and secondary objectives focus on standardized safety, social impact, and immunogenicity endpoints. See Ref. [15] for a discussion of endpoints used in Phase III vaccine trials.

3.1. Safety endpoints

Safety data guide decisions about a trial (for example, enrollment in a trial might be halted if safety concerns with the product are observed) and provide information for future trials. The use of standardized safety endpoints and a common reporting format aid in aggregating and comparing information on products across trials.

Safety endpoints fall into two main categories: reactogenicity events and adverse experiences. Reactogenicity events are defined prior to the start of the trial as events that are assumed, due to their nature and timing, to be directly related to vaccine or control administration. They are recorded by clinical staff through questionnaires immediately following vaccine administration until three to seven days afterwards. Reactogenicity events are classified as local events (pain, tenderness, erythema, induration at the injection site) or systemic events (malaise, myalgia, headache, fever, nausea, vomiting, arthralgia). Each sign or symptom is graded on a four-point scale of mild, moderate, severe, potentially life-threatening, based on the NIAID Severity Grades' definition.

Adverse experiences consist of other unfavorable or unintended changes in body structure, body function, or lab result temporally associated with the study vaccine. They may or may not be related to vaccination, and hence are assigned a value for relationship to study product (not, probably not, possibly, probably, or definitely related) in addition to a severity grade. Severity grades are based on a five-point scale (mild, moderate, severe, serious or life-threatening, or death) as defined by the NIAID Severity Grades. Adverse experiences include illnesses, accidents, injuries, or abnormal laboratory values, e.g., blood and urine measurements of various chemical and hematological parameters monitored at regular intervals throughout the trial. Laboratory values falling outside the range of normal are listed as adverse experiences if not associated with a clinical diagnosis. HVTN clinical staff code adverse events into MedDRA lower level terms, which are then linked to MedDRA body system and preferred terms for reporting. A comprehensive summary of safety events in NIAID-sponsored Phase I and II HIV-1 vaccine trials can be found in Ref. [16].

3.2. Social impact endpoints

Social impact endpoints are collected to highlight any social difficulties associated with participation in an HIV vaccine trial, which may carry a stigma. These endpoints include problems with family, friends, travel, work, school,

insurance, housing, or the military. Power of the trial is not typically driven by these endpoints, but they are an important part of trial monitoring.

3.3. Immunological endpoints

Two main objectives are to establish biologic activity (through the use of homologous reagents to the vaccine) and to assess plausible efficacy (through the use of reagents "homologous" to the viral population targeted for protection). Current immunological endpoints in HIV vaccine trials are classified by the type of adaptive immune responses invoked (humoral or cellular) and are measured by various immunologic assays at a select few time points. For the primary analysis, a single key time point (e.g., 2 weeks post-full vaccination) is typically identified to capture peak immunogenicity as well as allow expeditious reporting of results once the key endpoints on all participants have been analyzed. Durability of response may be assessed in a secondary analysis by considering immune response at a later time point or by combining immune responses over several time points [17]. Binary responses (positive/negative) are generally used since the continuous responses are often very low and for many assays have not yet been validated; the lab validation process is simpler for binary outcomes. As vaccine candidates begin to show greater immunogenicity, continuous responses will draw more interest.

3.3.1. Humoral immune responses

Humoral immunity endpoints rely on assays that measure binding and neutralizing antibodies, including primary isolate neutralization. Binding antibodies are capable of attachment to the HIV virus but may or may not hinder it. HIV-specific neutralizing antibodies, however, can prevent infection of host cells by binding to regions of the HIV enveloped required for viral attachment and entry [18]. The antibody assays currently in use in HIV vaccine trials include an ELISA assay to detect binding antibodies and a vital dye neutralization antibody assay [19–21]. For each ELISA assay, duplicate experimental (antigen-containing) and negative control (non-antigen containing) wells are used. An optical density (OD) reading for each well is produced when colored solution absorbs transmitted light proportional to the amount of antibody that has bound. A mean OD is calculated for each pair of replicates and an empirical criterion is used to assess positivity. For example, a sample is declared positive if the difference in mean OD of the experimental and negative control wells exceeds 0.2. For the vital dye neutralization antibody assay the readout is a neutralization titer, defined as the reciprocal of the serum dilution required to produce a 50% reduction in HIV production. Positivity is based on a threshold (e.g., titers greater than 10 are positive). For either assay, whether the positivity criterion indicates a level of humoral immune activity that is clinically relevant to protection is unknown and will remain unknown until an efficacious vaccine is developed.

If the continous OD and titer responses are analyzed, the right-skewness of the data and lower limits of detection must be taken into account. For log-transformed data and left-censored data, standard inferential methods can be applied [22,23].

3.3.2. Cellular immune responses

Cellular immunity is studied through assays that measure effector T lymphocyte responses. Enzyme-linked immunospot (ELISpot) and intra-cellular cytokine staining (ICS) assays are the two most commonly used assays in current HVTN trials.

The ELISpot assay provides a count of cytokine-secreting T cell after stimulation with HIV-specific antigens. The assay is performed on 96-well plates coated with anti-cytokine antibodies processed such that, theoretically, each cytokine-producing cell can be identified by a spot within a well. Raw ELISpot data for statistical analysis consist of the number of spot-forming cells (SFCs) for each experimental and control well; typically three experimental replicates and six control replicates. Experimental wells of interest contain cells stimulated with peptide pools of a given HIV-specific antigen. Each gene-specific response is then categorized as positive or negative based on whether or not the number of SFCs in the experimental wells is significantly greater than the number in the control wells. Various approaches to defining a positivity criterion have been taken [24–31]. Currently, the HVTN uses a permutation test to assess positivity of each antigen-specific response [32]. The test statistic is the difference in means of the experimental and control well responses. A permutation algorithm for step-down max T adjusted p-values [33] based on Algorithms 2.8 and 4.1 in Westfall and Young [34] is used to adjust for the multiple peptide pools.

The ICS assay allows a precise characterization of cells by phenotypic lymphocyte markers. The assay specifies the type of responses ($CD4^+$ or $CD8^+$) more readily than the ELISpot assay and can provide information on the functionality of each cell. ICS is performed on 96-well plates, with a light intensity reading on a scale of 1 to 1024 given to each labelled phenotypic marker for each cell within a well. The output data for statistical analysis consist of a proportion of $CD4^+$ or $CD8^+$ T cells that produce cytokines for each experimental and control well and the total number of $CD4^+$ or $CD8^+$ cells per well. Each experimental well is run only once for each peptide pool; the control wells are currently run in duplicate. Similar to ELISpot assay, each peptide pool-specific response is categorized as positive or negative. As this assay is relatively new, work is ongoing to refine the statistical positivity criterion.

3.4. Accounting for HIV variation in comparing vaccine candidates

To establish biologic activity, standard practice in Phase I and II trials has been to evaluate the immunogenicity of the tested HIV vaccine primarily by its induction of immune responses to the HIV virus represented in the vaccine. For humoral responses this has meant studying neutralization of the HIV isolate in the vaccine and for cell-mediated responses it has meant studying responses to overlapping HIV peptides with sequences that are contained in the vaccine strain sequence. Focusing on such "homologous" immune responses does not permit head-to-head comparisons of the immunogenicity of vaccine candidates, and because a successful vaccine must protect against a diverse array of circulating HIV viruses, it is important to use a set standardized HIV reagents that represent the diversity of HIVs in contemporary populations. Accordingly, the HVTN is currently developing a standardized set of HIV isolates for studying neutralization responses and a standardized set of peptides for studying cellular-mediated responses. We briefly summarize the strategy for constructing these HIV panels.

For neutralization responses, the current strategy is to compare candidates by three tiers of HIV targets, ranging from easiest to hardest to neutralize: 1) well-studied laboratory HIV strains that are known to be easy to neutralize; 2) a panel of 20–40 HIV isolates within the same subtype of the HIV isolate in the vaccine; 3) a panel of 20–40 HIV isolates of a different subtype than that in the vaccine. For panels 2) and 3) the isolates are selected to represent the spectrum of neutralization-sensitivities of HIVs in current populations, for example identified through random sampling of incident HIV infections. To compare candidates by cellular immunogenicity, bioinformatic methods are being used to optimally select a set of HIV peptides (e.g., of length 15 amino acids) that "cover" the largest number of HIV peptides in circulating HIV populations that may contain CTL epitopes [35]. The HVTN is pursuing a variety of approaches, the leading one is described here. A simple approach considers 150 database sequences from each HIV-1 subtype category A, B, C, and non-ABC; the last category includes circulating recombinant forms and unique circulating forms. A peptide set for evaluating immunogenicity can be constructed as the minimum set of peptides of length 10 that cover all "vaccine important peptides (VIPs)" contained in the 600 sequences, where a VIP is defined as a peptide present in at least 15% of the sequences within a subtype category. The rationale is to make the peptide panel sensitive to detecting T cell responses to any peptides that circulate at reasonably high frequency and therefore are of public health importance. A challenge with this approach is obtaining sequences recently sampled so that the VIPs are well-represented in contemporary circulating HIVs.

4. Statistical considerations in clinical evaluation

The goal of the HVTN is to foster cooperation of researchers worldwide and to help streamline the scientific process of HIV vaccine development towards the development of a safe, globally effective preventive HIV vaccine. Phase I and II trials test a range of vaccine candidates with the view of narrowing the number to the best few to move forward to Phase IIb and Phase III trials. The latter are lengthy, costly undertakings and yet, are the only approach for determining efficacy of candidate HIV vaccines.

Clinical evaluation of products in a healthy population requires a framework that ensures the safety of participants to the best extent possible. The HIV vaccine evaluation program begins with a Phase I trial as the first examination of safety in humans as well as to establish some indication of immunogenicity of the vaccine candidate. Once safety and immunogenicity have been demonstrated, safety is further characterized in a larger trial and the vaccine regimen is optimized with respect to immunogenicity endpoints. The safety and immunogencity of the final regimen are then characterized and possibly compared with other competing regimens with sufficient precision to provide evidence for or against initiation of an efficacy trial.

Several elements are common to each stage of trial. Randomization is done in blocks to ensure balance across arms of the trial. For multi-region trials (e.g., US and Southern Africa), separate blocks are used for each region. The interim analyses of each stage of trial are also similar. These consist of unblinded safety data reports prepared at regular intervals for an independent safety monitoring committee. The committee meets every four months to review the safety data from all ongoing HVTN trials. Interim analyses of immunogenicity data from a subset of participants are rarely done hence alpha-spending procedures [36–38] are not currently used.

4.1. Phase I: Initial safety and immunogencity evaluation

Phase I HIV vaccine trials differ from classical Phase I trials that estimate dose–response and maximum tolerated dose. With HIV vaccines, there is a narrow range of possible doses; many vaccine candidates are limited by production or manufacturing considerations rather than by a maximum tolerated dose. In the HVTN evaluation framework, a Phase I trial typically considers three to four different doses of vaccine in a sequential dose-escalation trial with the goal of reaching the maximum dose. If the pre-specified safety stopping rules are met for a dose group, the trial may be halted and the problematic doses excluded from further testing. The sample size is limited to the minimum number needed to quickly assess preliminary safety of the candidate vaccines.

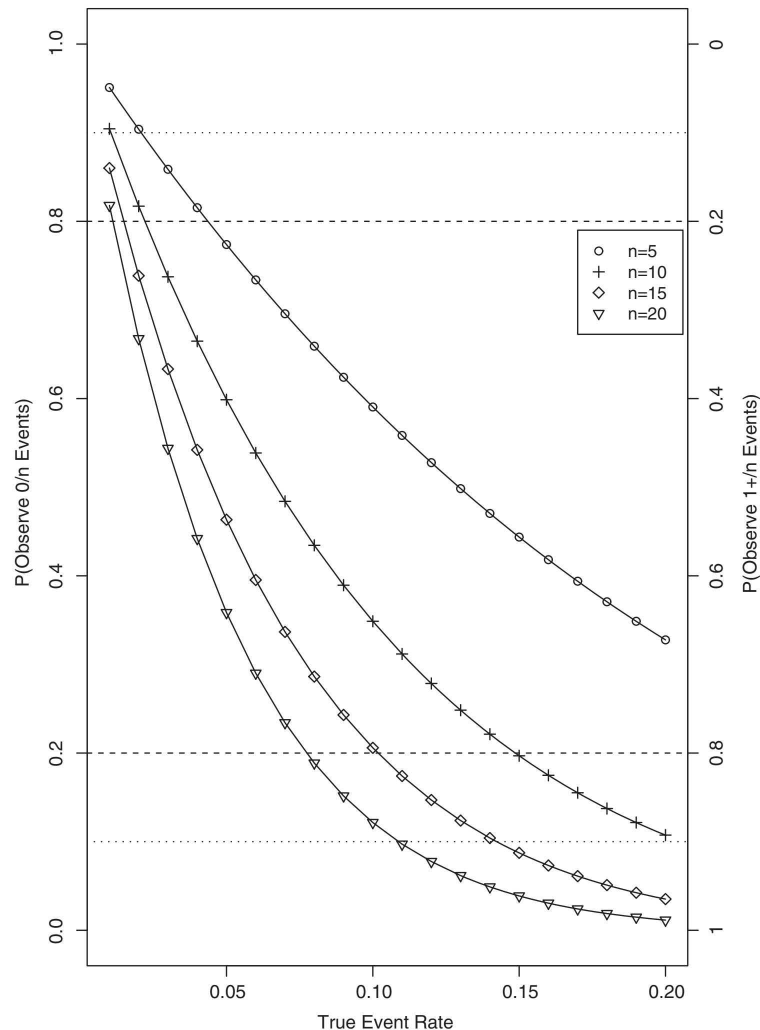

Fig. 1. Probablity of observing $0/n$ or $1+/n$ events for various true event rates.

The standard HVTN Phase I trial design enrolls 10 vaccinees and 2 placebos per group. For example, the HVTN 040 trial used this design to evaluate an alphavirus HIV clade C Gag vaccine in the US and South Africa. Placebos are used primarily for blinding in Phase I trials and are insufficient to adjust for false positive rates, so Phase I trials report crude safety and immunogenicity rates only. The design is powered to observe safety and toxicity outcomes if they are likely (prevalence $\geq 20\%$), with only a low probability of observing rare safety events (prevalence $\leq 3.5\%$). As a result, the trial design permits the reliable detection only of relatively common serious adverse experiences to assure sufficient safety to move the candidate to the next level of evaluation. Reliable detection of a true rate of serious adverse experiences (SAEs) of 20% or more is deemed sufficient. Close safety monitoring in later trials will gather further safety data to make more precise statements about the safety profile. Prophylactic vaccines require an extremely high level of safety given their administration to a healthy population [39–41].

Fig. 1 displays examples of the probability of observing zero events (left margin) or more than one event (right margin) given different prevalence rates. Note that 5 vaccinees are too few to reliably detect even common events. For example, if the prevalence of a vaccine-related adverse event is 20%, the probability of observing at least one such event in an arm with 5 vaccinees is less than 0.70. If $n=0$, the probability is approximately 0.90. There is less gain in moving from 10 to 15 vaccinees per arm, especially considering the cost of increasing the sample size of each arm by 50%.

Table 1 illustrates the precision with which safety and immunogenicity event rates can be estimated for various sample sizes. For example, if we observe a 0 out of 10 immune response rate in a vaccine arm, the 95% confidence interval for the true response rate includes 30%. While this does not provide precise information in the context of safety, an upper bound of approximately 30% for the true immunogenicity response rate may be sufficiently low to exclude a vaccine from further testing at least in its current formulation.

The power calculations for safety described thus far refer to vaccine-related adverse events. In practice, numerous safety events are measured (see Section 3.1). If safety endpoints between groups are compared, one might consider a multiplicity adjustment [34,42–47]. Mehrotra and Heyse proposed a two-step false discovery rate approach to balance potential over-adjustment and no adjustment, applying the methodology to adverse event data from three vaccine trials [48]. However, the HVTN prefers a conservative approach that detects all safety concerns at the risk of an overinflated false positive rate [49,9]. Good judgment is then exercised in interpreting the unadjusted p-values to avoid discarding a safe vaccine.

4.1.1. Initial evaluation of combination vaccines

Combination vaccines refer to vaccination regimens composed of two or more immunogens delivered sequentially or simultaneously. The evaluation of combination HIV vaccine candidates will become increasingly important as none of the currently available single agents generate a full complement of immune responses thought to be important for protection. Two aspects of early clinical evaluation of combination vaccines merit attention here. First is the design of dose-escalation trials. The current standard for dose-escalation of combination vaccine regimens is to first complete

Table 1
Exact two-sided 95% confidence intervals for the response probability given 0, 1, or 2 events are observed for a sample size n

n		Exact two-sided 95% CI
	0 events	
5		[0.00, 0.52]
10		[0.00, 0.31]
15		[0.00, 0.22]
20		[0.00, 0.17]
	1 event	
5		[0.01, 0.72]
10		[0.00, 0.45]
15		[0.00, 0.32]
20		[0.00, 0.25]
	2 events	
5		[0.05, 0.85]
10		[0.03, 0.56]
15		[0.02, 0.40]
20		[0.01, 0.32]

dose-escalation of each component immunogen separately and then proceed with dose-escalation of the combination. For regimens that are delivered by multiple administrations over a six-month period (as is common with the current HIV vaccine candidates) this dose-escalation design is quite time consuming and can require well over a year to complete. A slightly less conservative approach to this design would be to initiate each step in the dose-escalation of the combination immediately after the safety of the single agents is evaluated at the specific doses to be delivered in the combination. Although the dose-escalation of the combination would not be informed by data from the entire dose-escalation of each of the component immunogens, it would be informed by safety of each component at the particular doses being considered in the combination.

A second issue in evaluating combination vaccine regimens involves optimization and characterization of immunogenicity. A natural tendency for clinical researchers is to consider factorial designs for the evaluation of immunogenicity of combination regimens. Such factorial designs target the marginal effects of component immunogens averaged over all other component immunogens with which it could be delivered. The utility of factorial designs typically hinges on an assumption of minimal interaction effects. However, most combination regimens are motivated precisely because of the potential for interaction between the component immunogens and, as such, the marginal effects of the components are rarely the true effects of interest. In addition, detailed discussions with vaccinologists often reveal that only a subset of all possible combinations are of true interest for evaluation. For these reasons, we consider the immunogenicity of combination regimens to be best evaluated in multiarm designs that include only those combinations of particular interest and that are sized to evaluate all pairwise comparisons of regimens of interest.

4.2. Phase Ib: Regimen optimization/selection

Once preliminary safety and some indication of immunogenicity of a candidate HIV vaccine regimen at the desired dose and/or number of injections have been demonstrated in a Phase I trial, the regimen advances to further evaluation of immunogenicity and additional safety testing where factors that enhance immunogenicity are considered. When many vaccine candidates are tested in a single trial and there are clear criteria for moving one or more forward to the next level of evaluation, selection designs are attractive options. Under these circumstances, selection designs are more efficient at screening multiple regimens compared to traditional superiority or non-inferiority trials although the distinctions between these should be made clear. A traditional design is powered to compare pairs of vaccine arms and hence requires a large total sample size to detect even moderate sized differences when there are many vaccines. The selection design requires a smaller sample size since the trial is powered to rank vaccine arms and then select the top ranked vaccine. The HVTN has made a programmatic decision to defer non-inferiority questions to later stages of trial. A selection design can be used appropriately to select among various doses of a vaccine product or the timing of various schedules of the same length of duration. Statistical methods for selection designs are described in detail in Bechhofer et al. [50]. Liu et al. [51] proposed the use of a selection design for a cancer clinical trial of multiple new regimens.

For Phase Ib trials with binary immunogenicity outcomes as primary endpoints, the HVTN often considers a selection design that employs the single-stage procedure described in Chapter 7 of Bechhofer et al. [50]. The procedure, proposed by Sobel and Huyett [52], ranks the response rates of k treatment groups and selects the treatment with the largest observed response rate regardless of the magnitude of the difference between it and the other groups, randomizing in the case of ties. Sample size calculations for the procedure are driven by δ, the smallest difference of interest between the best and next best treatment, and $P(\mathrm{CS})$, the minimum acceptable probability of correct selection. Simulations can be used to calculate the sample size for a given δ and $P(\mathrm{CS})$ to ensure that if the true response rate of the best vaccine exceeds all others by at least δ, the probability of selecting that vaccine is sufficiently high — $P(\mathrm{CS})$ is typically set at 0.90. After ranking the outcomes and selecting the treatment with the largest response, one can assert with confidence of at least $P(\mathrm{CS})$ that the selection is correct whenever the true difference in probabilities is at least as large as δ.

The HVTN often employs a selection design with three to five groups, each containing 30 vaccinees and 6 controls. For example, a selection design will be used in HVTN 055 to select the combination regimen with the highest cumulative HIV-specific T cell response as measured by the ELISpot assay. Table 2 lists the range of response rates, samples sizes (n), and number of treatment groups for which the probability of correct selection is at least 0.90. For trials with at most three treatment groups, $n=30$ per group ensures the probability of correct selection of the vaccine

Table 2
Range of sample sizes, response rates, smallest differences (δ), and number of regimens (k) for which the probability of correct selection of regimen with highest response rate = 90%

n	Highest rate	2nd highest rate	δ	Power (two-sample test)
$k=2$				
20	0.70	0.50	0.20	0.32
30	0.66	0.50	0.16	0.24
40	0.64	0.50	0.14	0.26
50	0.63	0.50	0.13	0.27
$k=3$				
20	0.74	0.50	0.24	0.41
30	0.70	0.50	0.20	0.35
40	0.67	0.50	0.17	0.35
50	0.65	0.50	0.15	0.34
$k=4$				
20	0.76	0.50	0.26	0.47
30	0.72	0.50	0.22	0.41
40	0.69	0.50	0.19	0.42
50	0.67	0.50	0.17	0.42
$k=5$				
20	0.78	0.50	0.28	0.52
30	0.73	0.50	0.23	0.44
40	0.70	0.50	0.20	0.46
50	0.68	0.50	0.18	0.46

For $k>2$, all groups except the highest group had response rates equal to the second highest rate shown in column 2. Contrast with the power to compare the highest and second highest response rates with a standard two-sample test.

that induces the highest response rate is greater than or equal to 0.90 when the difference between the highest and second highest response rate is at least 0.20. Trials with more than three treatment groups should consider larger sized groups so that δ can be set at a value of 0.20 or less. This would call for sample sizes of 40 per group when there are four or five groups under consideration. Choice of δ is a programmatic decision driven by the desire to quickly select a vaccine candidate based on response rate to move forward to characterization and further evaluation in a Phase II trial. Table 2 also illustrates the sample size economy of a selection design over a standard superiority trial design. The last column of the table lists the power of a two-sample t-test to detect various differences in response rates across sample sizes. Power is unacceptably low for all of the scenarios listed in the table; for example, there is only 24% power to detect a difference of 0.16 in the response rates of two vaccine groups with 30 participants in each.

Research is ongoing to refine the basic selection design to the HIV vaccine trials' setting. For example, two new components can be added to the selection rule: (i) an absolute minimum response probability threshold that a vaccine must exceed and (ii) a rule for advancing all candidates that are indistinguishable from the best. A minimum threshold can be incorporated into the selection rule by requiring that the selected candidate(s) have a two-sided $(1-\alpha_1)$ percent confidence interval about the response probability p that exceeds a fixed probability p_{low}. The error rate α_1 and threshold p_{low} should be chosen such that all vaccines that would be predicted to meet the efficacy trial qualifying objective (i.e., $p>0.30$ with $\geq 95\%$ confidence) are advanced. In addition, a rule can be added that advances all candidates for which the $(1-\alpha_2)$ percent confidence interval about $|p-p_{\text{best}}|$ includes 0. The error rate α_2 is based on simulations such that vaccines with p within a tolerance distance Δ_1 of p_{best} (e.g., $\Delta_1=0.1$) are advanced with high probability and vaccines with p further than some clinically relevant distance Δ_2 from p_{best} (e.g., $\Delta_2=0.2$) are not advanced with high probability. For $\alpha_1=0.1$, $p_{\text{low}}=0.2$, and $\Delta_1=0.1$, Table 3 shows incorrect and correct selection probabilities for a three arm Phase Ib trial for different values of (α_2, Δ_2). Research is also being directed towards the use of continuous and multivariate endpoints in selection designs.

Other variations on the selection design might be considered to increase efficiency, if information is available on the magnitude of expected response rates of the best and the next best treatments. An alternative design that places bounds on the theoretical response rates can be used to decrease the required sample size for a given level of power [50]. To date, however, there is insufficient immunogenicity data to make use of this design. Adaptive sequential sampling designs [53] are currently impractical due to the complexity of implementing multisite HIV vaccine trials; they are also hindered by the length of enrollment, the time from enrollment to full vaccination, and the time needed to measure and report key

Table 3
Evaluation of selection designs with $n=30$ vaccinees per arm based on three operating characteristics IS_1, IS_2 and CS

True response rates			$\Delta_2=0.1$						$\Delta_2=0.2$					
			$\alpha_2=0.1$			$\alpha_2=0.3$			$\alpha_2=0.1$			$\alpha_2=0.3$		
p_1	p_2	p_1	IS_1	IS_2	CS	IS_1	IS_2	CS	IS_1	IS_2	CS	IS_1	IS_2	CS
0.2	0.2	0.5	0.22	0.22	0.68	0.04	0.04	0.86	0.22	0.22	0.69	0.04	0.04	0.86
0.2	0.2	0.8	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00
0.2	0.5	0.5	0.08	0.08	0.91	0.01	0.01	0.98	0.08	0.08	0.91	0.01	0.01	0.98
0.2	0.8	0.8	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	1.00	0.00	1.00
0.1	0.1	0.1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.2	0.2	0.2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.5	0.5	0.5	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00
0.8	0.8	0.8	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00

IS_1 is the estimated probability that the trial will incorrectly select an arm with true response rate $p<p_{low}\leq 0.2$; IS_2 is the estimated probability that the trial will incorrectly select an arm with true response rate $p<p_{best}-\Delta_1$ where $\Delta_1=0.1$; and CS is the estimated probability that the trial will correctly advance all arms with true response rate p within Δ_2 of p_{best} and $p>p_{low}$.

immunogenicity endpoints to the statistical center. Primary endpoints for standard immunogenicity outcomes are usually measured two to four weeks after full vaccination and may not be available for months, depending on the length of the vaccination schedule under study and the time required to batch test the samples. Therefore adaptive designs based on immunologic endpoints will be challenging in our setting, where rapid evaluation is highly valued.

In addition to selecting a vaccine regimen, Phase Ib trials also seek to estimate rates of safety and immunogenicity endpoints. If no vaccine-related SAEs are observed in a group of 30 vaccinees, the upper bound on the true rate of SAEs of an exact two-sided 95% confidence interval is 0.12; for $n=50$, the upper bound is 0.07. These are deemed sufficient to expand evaluation to several hundred individuals in a Phase II trial. To illustrate the precision with which Phase Ib trials can estimate immune response rates, Table 4 lists exact two-sided 95% confidence intervals for various sample sizes and observed response rates. Note, for $n=30$ the width of a 95% confidence interval for various observed response rates is between 0.38 and 0.25. The confidence interval is still wide (0.32 to 0.21) if the sample size is increased by 10. Hence if the trial size is limited to vaccine groups of $n\leq 50$ and the placebo group is too small to accurately estimate the rate of false positives, the primary objectives of Phase Ib trial should not include precise estimation of response rates. A placebo group may even be omitted if the primary objective is selection based on crude response rates.

The above discussion of selection designs and sample size calculations refers to decision-making on the basis of a single immune response. In practice, numerous immunological endpoints are of interest with little or no information about which of these, if any, is most predictive of protection from HIV infection or progression to disease. However, investigators may be capable of establishing a hierarchy of key endpoints prior to trial initiation; this is crucial to avoid post hoc data-driven decisions. Furthermore, once quantitative assay outcomes are validated for use as primary endpoints, power calculations may result in different sample size estimates.

4.3. Phase II: Comparison and characterization

The main objective of a Phase II trial of HIV vaccine candidates is to characterize the immune profile of a candidate vaccine to move forward to efficacy testing based on the endpoints described in Section 3.3. A Phase II trial differs

Table 4
Exact two-sided 95% confidence intervals for the response rate given various observed immune response rates and sample sizes

n	Observed immune response rate				
	0.5	0.6	0.7	0.8	0.9
20	[0.27, 0.73]	[0.36, 0.81]	[0.46, 0.88]	[0.56, 0.94]	[0.68, 0.99]
30	[0.31, 0.69]	[0.41, 0.77]	[0.51, 0.85]	[0.61, 0.92]	[0.73, 0.98]
40	[0.34, 0.66]	[0.43, 0.75]	[0.53, 0.83]	[0.64, 0.91]	[0.76, 0.97]
50	[0.36, 0.64]	[0.45, 0.74]	[0.55, 0.82]	[0.66, 0.90]	[0.78, 0.97]

from a Phase I in that it is done in a high risk population likely to be used in a Phase III trial. This is not necessarily true of Phase I trials, which might be restricted to low-risk volunteers. An efficacy trial requires substantial clinical evidence that the vaccine is safe and has shown sufficient immunogenicity to warrant a full-scale efficacy evaluation. Regulatory bodies such as the FDA often require that vaccines tested in efficacy trials must first have been administered to a large number of participants in Phase I and II trials to demonstrate a tolerable safety profile. Hence a Phase II trial of a vaccine candidate that has not been extensively studied in Phase I trials may be larger than what is described here. The sample size of each group may be inflated mainly to increase the total number of participants who will have received the product prior to an efficacy trial with the added benefit of an increase in the precision of the overall safety and immunogencity characterization.

The standard HVTN Phase II design for characterizing the safety and immunogenicity of each vaccine candidate consists of two to four groups, each of which has a vaccine arm of 90 to 150 participants. The total size of the placebo group might be set equal to the size of a single vaccine arm, to allow sufficient precision to adjust the immunogenicity response rates to account for false positive responses and to collect background rates of safety events in populations of interest for Phase III trials. In contrast, the much smaller placebo arms in Phase I and IIb trials serve primarily to blind the trial. A key component of Phase II trials is the demonstration of a stringent safety profile. The larger sample size allows a tighter bound on the rate of vaccine-related SAEs; e.g., the upper bound of an exact two-sided 95% confidence interval is 0.036 for $n=100$. If the earlier phase trial data are pooled with the Phase II safety data, the upper bound is 0.026 for $n=140$ and 0.018 for $n=200$.

Within the HVTN, characterization of immunogenicity is particularly important in Phase II trials, since some minimum level of immunogenicity of the candidate regimen is likely to be a criterion for beginning an efficacy trial. HVTN 203, the HVTN's first Phase II trial, specified that the lower bound of the confidence interval for the net point prevalence cellular response rate must exceed 0.20 or the net cumulative cellular response rate must exceed 0.30 at the key six-month time point. The primary immunogenicity endpoints were cellular responses to Env, Gag, Pol, or Nef as measured by the chromium release and ELISpot assays. The trial's ability to achieve its immunogenicity milestone depended on the false positive rate of the assays, estimated from previous AIDS Vaccine Evaluation Group studies to be 0.02 for point prevalence and 0.06 for cumulative prevalence. Based on these estimates, the 203 trial had over 90% power to detect the point prevalence outcome and approximately 80% power to detect the cumulative prevalence outcome for crude response rates of 0.38 (point prevalence) and 0.54 (cumulative prevalence). Because the immunogenicity milestone was not met, a proposed efficacy trial was called off.

When the false positive rate of an assay is high, detecting a given response rate becomes more challenging (see Table 5). For example, in a trial with $n=90$ per group, there is less than 40% power to detect a net rate of 0.30 in a vaccine group with a true immune response rate of 0.50 and a false positive rate of 0.10. With such a high false positive rate, the true response rate would have to be closer to 0.60 for $n=90$ to allow over 90% power to detect a net response rate of 0.30. These calculations highlight the importance of accurate estimation of the false positive rate of assays used to measure primary immunogenicity endpoints prior to the design of a Phase II trial when immune characterization is a primary objective. The Phase II trial can then in turn provide data on the false positive rate of endpoint assays on blood samples collected from international sites, if this has not yet already been well estimated. Such information will be helpful in planning later stage trials that seek to address immune correlates of protection.

The discussion of immunogenicity endpoints up until this point has focused primarily on response rates. However when the assay being used reliably outputs a quantitative value, the magnitude of response as well as the breadth of response are arguably of greater interest for the comparison of potential vaccine candidates. The measurement of magnitude and breadth cannot be separated; therefore one approach to joint evaluation is to measure magnitude of response to each HIV target in a set (e.g., neutralization titers are measured to each of $N=30$–40 HIV isolates). For a

Table 5
Power to detect a 0.30 net immune response rate for different sample sizes, true response rates, and false positive rates in Phase II trials

True rate	0.40	0.50	0.50	0.50	0.60	0.60
False positive rate	0.01	0.01	0.05	0.10	0.05	0.10
n:						
90	0.38	0.94	0.72	0.39	0.99	0.91
100	0.46	0.97	0.79	0.40	0.99	0.93
120	0.52	0.99	0.86	0.47	0.99	0.97

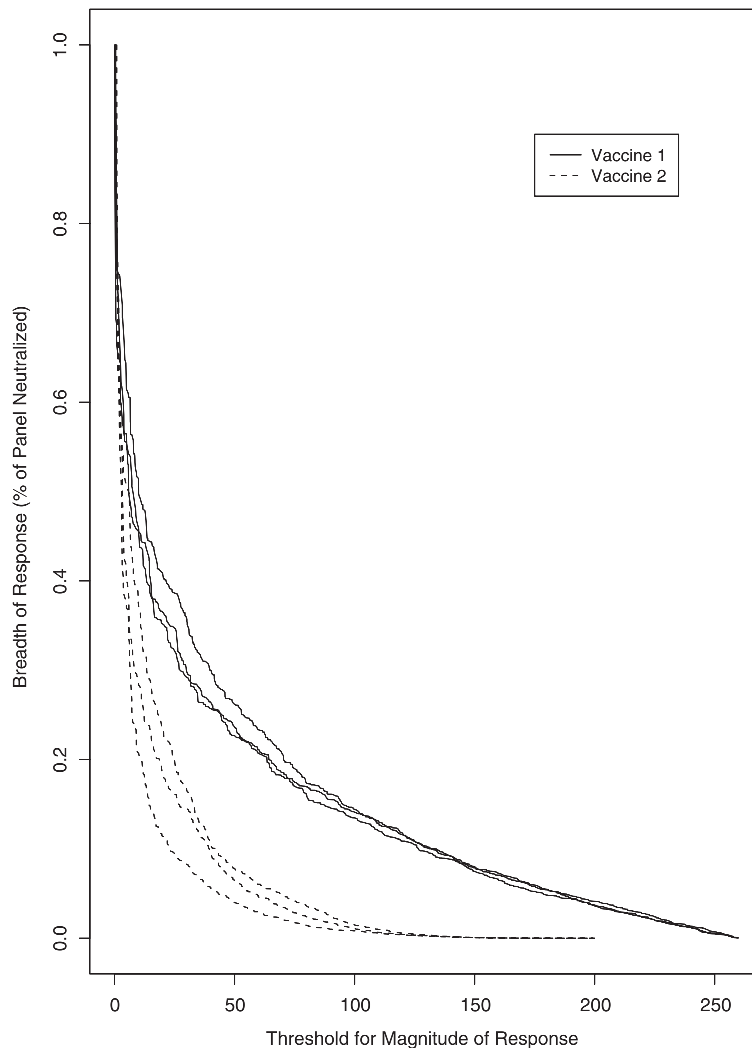

Fig. 2. Magnitude–breadth curves for 3 vaccinees from group 1 and 3 vaccinees from group 2.

given magnitude threshold, k, one can determine the percentage of the N isolates with neutralization titer $>k$. Magnitude and breadth can then be summarized for each participant with a curve (see Fig. 2 for an example). Differences between vaccine groups can be tested by two-sample statistics that contrast summary measures of the curves; for example, a Wilcoxon rank sum test could be used to compare the distributions of the areas under the magnitude–breadth curves.

5. Conclusions

The program outlined in this paper describes an evaluation platform focused on rapidly obtaining results critical to the evaluation of an HIV vaccine regimen for potential success in an efficacy trial. This determination can be made quickly, providing answers to a limited set of key questions using a standardized approach. Flexibility is not lost, as information necessary for further optimization of future vaccine regimens is collected along the way.

The decision to proceed from the first Phase I trial of a given vaccine product to a Phase Ib trial is primarily based on safety endpoints, although lack of any observed immunogenicity may discourage investigators and manufacturers from further study of the vaccine construct. The primary objective of the Phase Ib trial is to narrow the number, while optimizing the regimen, of vaccine products moving into Phase II trials. Further administrative efficiency can be achieved by combining Phase I and Phase Ib trials into a single trial with two distinct parts: dose-escalation and

Table 6
Summary of key features of the typical design for each trial phase

Phase	Primary endpoints	Sample size: V/P	95% CI UB on vaccine safety when 0 SAEs are observed
I	Safety	10/2	31%
Ib	Safety and immunogenicity	30/6	12%
II	Safety and immunogenicity	100/50	3.6%

optimization. Establishment of safety in the first part of the trial would lead directly into the second part without the additional operational burden of implementing a separate protocol. HVTN 065 uses this type of approach to evaluate a DNA vaccine prime with a recombinant modified Vaccinia Ankara boost. Ideally, the subsequent Phase II trial will further narrow the vaccine candidates to a single vaccine regimen that will be moved to testing in a Phase III trial, although this may require a large Phase II trial. Key features of the typical HVTN designs for each trial phase are summarized in Table 6.

The HVTN program is highly standardized for efficiency and facilities cross-trial comparison of vaccine candidates. It is tuned to emphasize rapid evaluation of safety and to optimize a regimen for further evaluation once it meets certain safety and immunogenicity criteria. The issues described in this paper pertain mainly to HIV vaccines but may also apply to other vaccines under development, especially in settings where no efficacious vaccine exists and there are multiple vaccine candidates to evaluate.

Acknowledgments

This research was supported by National Institutes of Health grants 5-U01 AI-46703 (HVTN SDMC). The authors thank Ann Duerr, Barb Metch, Erik Schwab, and Emily Hemminger for helpful comments.

References

[1] UNAIDS/WHO. AIDS Epidemic Update. www.unaids.org. December 2004.
[2] Fauci AS. The AIDS epidemic—considerations for the 21st century. N Engl J Med 1999;314:1046–50.
[3] Varmus H, Nathanson N. Science and the control of AIDS. Science 1998;10:1815.
[4] Desrosiers RC. Prospects for an AIDS vaccine. Nat Med 2001;10:221–3.
[5] Savla U. Reservoir: a dirty word in HIV. J Clin Invest 2004;113:146.
[6] NIAID. How HIV causes AIDS. www.niaid.nih.gov/factsheets/howhiv.htm. November 2004.
[7] HIV InSite Knowledge Base Chapter. Immunopathogenesis of HIV infection. http://hivinsite.ucsf.edu/InSite?page=kb-02-01-04#S4X. October 2004.
[8] Emini E. A potential HIV-1 vaccine using a replication-defective adenoviral vaccine vector. 9th conference on retroviruses and opportunistic infections, Seattle, Washington, February 24–28; 2002.
[9] Horne AD, Clifford J, Goldenthal KL, Kleppinger C. Efficacy and safety of preventive vaccines: FDA evaluation. Proceedings of the annual meeting of the American Statistical Association, Atlanta, Aug. 5–9; 2001.
[10] Letvin NL, Huang Y, Chakrabarti BK, et al. Heterologous envelope immunogens contribute to AIDS vaccine protection in rhesus monkeys. J Virol 2004;78:7490–7.
[11] Wang R, Epstein J, Baraceros FM, et al. Induction of $CD4^+$ T cell dependent $CD8^+$ type 1 responses in humans by a malaria DNA vaccine. Proc Natl Acad Sci U S A 2001;98:10817–22.
[12] Rogers WO, Baird JK, Kumar A, et al. Multistage multiantigen heterologous prime boost vaccine for Plasmodium knowlesi malaria partial protection in rhesus macques. Infect Immun 2001;69:5565–72.
[13] Epstein JE, Gorak EJ, Charoenvit Y, et al. Safety, tolerability, and lack of antibody responses after administration of a pfCSP DNA malaria vaccine via needle or needle-free jet injection, and comparison of intramuscular/intradermal routes. Hum Gene Ther 2002;13:1551–60.
[14] Rogers WO, Weiss WR, Kumar A, et al. Protection of rhesus macaques against lethal Plasmodium knowlesi malaria by a heterologous DNA priming and poxvirus boosting immunization regimen. Infect Immun 2002;70:4329–35.
[15] Hudgens MG, Gilbert PB, Self SG. Endpoints in vaccine trials. Stat Methods Med Res 2004;13:89–114.
[16] Gilbert PB, Chiu YL, Allen M, et al. Long-term safety analysis of preventive HIV-1 vaccines evaluated in AIDS vaccine evaluation group NIAID-sponsored Phase I and II clinical trials. Vaccine 2003;20:2933–47.
[17] Hudgens MG. Estimating cumulative probabilities from incomplete longitudinal binary responses with application to HIV vaccine trials. Stat Med 2003;22:463–79.
[18] HIV InSite Knowledge Base Chapter. Assays to detect host immune response to HIV. http://hivinsite.ucsf.edu/InSite?page=kb=authorsdoc=kb-authorsdoc=kb-02-02-04. July 2003.
[19] Montefiori DC, Graham BS, Kliks S, Wright PF. Serum antibodies to HIV-1 in recombinant vaccinia virus recipients boosted with purified recombinant gp160. NIAID AIDS Vaccine Clinical Trials Network. J Clin Immunol 1992;12:429–39.

[20] Montefiori DC, Robinson Jr WE, Schuffman SS, Mitchell WM. Evaluation of antiviral drugs and neutralizing antibodies to human immunodeficiency virus by a rapid and sensitive microtiter infection assay. J Clin Microbiol 1998;26:2331–5.
[21] Pilgrim AK, Pantaleo G, Cohen OJ, et al. Neutralizing antibody responses to human immunodeficiency virus type I in primary infection and long-term-nonprogressive infection. J Infect Dis 1997;176:929–32.
[22] Klein JP, Moeschberger ML. Survival analysis: techniques for censored and truncated data. Springer-Verlag Inc; 1997.
[23] Hosmer DW, Lemeshow S. Applied survival analysis: regression modeling of time to even data. John Wiley and Sons; 1999.
[24] Larsson M, Jin X, Ramratnam B, et al. A recombinant vaccinia virus based ELIPSOT assay detects high frequencies of Pol-specific CD8 T cells in HIV-1-positive individuals. AIDS 1999;13:767–77.
[25] Subklewe M, Chahroudi A, Bickham K, et al. Presentation of Epstein–Barr virus latency antigens to $CD8^+$, interferon-γ-secreting, T lymphocytes. Eur J Immunol 1999;29:3995–4001.
[26] Alter G, Merchant A, Tsoukas CM, et al. Human immunodeficiency virus (HIV) — specific effector CD8 T cell activity in patients with primary HIV infection. J Infect Dis 2002;185:755–65.
[27] Lalvani A, Brookes R, Hambleton S, Britton WJ, Hill AV, McMichael AJ. Rapid effector function in $CD8^+$ memory T cells. J Exp Med 1997;186:859–65.
[28] Kaul R, Plummer FA, Kimani J, et al. HIV-1-specific mucosal $CD8^+$ lymphocyte responses in the cervix of HIV-1-resistant prostitutes in Nairobi. J Immunol 2000;164:1602–11.
[29] Larsson M, Messner D, Somerson S, et al. Requirement of mature dendritic cells for efficient activation of Influenza A-specific Memory $CD8^+$ T cells. J Immunol 2000;165:1182–90.
[30] Loing E, Andrieu M, Thiam K, et al. Extension f HLA-A *0201-restricted minimal epitope by *N*-palmitoyl-lysine increases the life span of functional presentation to cytotoxic T cells. J Immunol 2000;164:900–7.
[31] Dhodapkar MV, Young JW, Chapman PB, et al. Paucity of functional T-cell memory to melanoma antigens in healthy donors and melanoma patients. Clin Cancer Res 2000;6:4831–8.
[32] Moodie Z, Huang Y, Gu L, Self S, Hural J. Positivity criteria for ELISpot assay data in HIV-1 vaccine trials. AIDS Vaccine 2005, Montreal, vol. 28. Abstract; 2005.
[33] Dudoit S, Shaffer JP, Boldrick JC. Multiple hypothesis testing in microarray experiments. Stat Sci 2003;18:71–103.
[34] Westfall PH, Young SS. Resampling-based multiple testing: examples and methods for *p*-value adjustment. New York: Wiley; 1993.
[35] Li F, Malhotra U, Gilbert PB, Hawkins NR, Duerr AC, McElrath JM, Corey L, Self SG. Peptide selection for human immunodeficiency virus type 1 ctl-based vaccine evaluation (submitted for publication).
[36] O'Brien PC, Fleming TR. A multiple testing procedure for clinical trials. Biometics 1979;35:549–56.
[37] Pocock SJ. Interim analyses for randomized clinical trials: the group sequential approach. Biometrics 1982;38:153–62.
[38] Fleming TR, Harrington DP, O'Brien PC. Designs for group sequential tests. Control Clin Trials 1984;5:348–61.
[39] Ellenberg SS, Braun MM. Monitoring the safety of vaccines: assessing the risks. Drug Saf 2002;25:145–52.
[40] Ellenberg SS. Safety considerations for new development. Pharmacoepidemiol Drug Saf 2002;10:411–5.
[41] Ellenberg SS, Chen RT. The complicated task of monitoring vaccine safety. Public Health Rep 1997;112:10–20.
[42] Benjamini Y, Hochberg Y. Controlling the false discovery rate; a practical and powerful approach to multiple testing. J R Stat Soc Ser B Methodol 1995;57:289–300.
[43] Hochberg Y, Westfall PH. On some multiplicity problems and multiple comparison procedures in biostatistics. Handbook of Statistics, vol. 18; 2000. p. 75–113.
[44] Storey J. A direct approach to false discovery rates. J R Stat B 2002;64:479–98.
[45] Dudoit S, van der Laan MJ, Pollard KS. Multiple testing: Part I. Single-step procedures for control of general type I error rates. Stat Appl Genet Mol Biol 2004;3.
[46] van der Laan MJ, Dudoit S, Pollard KS. Multiple testing: Part II. Step-down procedures for control of the family-wise error rate. Stat Appl Genet Mol Biol 2004;3.
[47] Romano J, Wolf M. Exact and approximate stepdown methods for multiple hypothesis testing. J Am Stat Assoc 2005;100:94–108.
[48] Mehrotra DV, Heyse JF. Use of the false discovery rate for evaluating clinical safety data. Stat Methods Med Res 2004;13:227–38.
[49] Friedman LM, Furberg C, DeMets DL. Fundamentals of clinical trials. New York: Springer-Verlag; 1998.
[50] Bechhofer RE, Santner TJ, Goldsman DM. Design and analysis of experiments for statistical selection, screening, and multiple comparisons. John Wiley and Sons; 1995.
[51] Liu PY, Dahlberg S, Crowley J. Selection designs for pilot studies based on survival. Biometrics 1993;49:391–8.
[52] Sobel M, Huyett MJ. Selecting the best one of several binomial populations. Bell Syst Tech J 1957;36:537–76.
[53] O'Quigley J, Pepe M, Fisher L. Continual reassessment method: a practical design of Phase I clinical trials in cancer. Biometrics 1990;46:33–48.