Surrogates of Protection in Repeated Low-Dose Challenge Experiments

Dustin M. Long[a], Michael G. Hudgens[b]*, Chih-Da Wu[b]

A critical step toward developing a successful vaccine to control the human immunodeficiency virus (HIV) pandemic entails evaluation of vaccine candidates in non-human primates (NHPs). Historically, these studies have usually entailed challenges (i.e., exposures) with very high doses of a simian version of HIV, resulting in infection of all NHPs in the experiment after a single challenge. More recently, researchers have begun to conduct repeated low-dose challenge (RLC) studies in NHPs that are believed to more closely mimic typical exposure in natural human transmission settings. One objective of RLC studies is to assess whether measured immune responses to vaccination can serve as surrogate endpoints for the primary endpoint of interest, namely infection. In this paper, different designs of RLC studies for assessing a binary surrogate of protection are considered. Copyright © 2014 John Wiley & Sons, Ltd.

Keywords: Causal Inference, HIV, Immunogenicity, Infectious Diseases, Transmission Probability, Vaccine

1. Introduction

Experiments in non-human primates (NHPs) play an essential role in the development, screening, and evaluation of preventive human immunodeficiency virus (HIV) vaccines [1, 2]. Historically, these studies have usually entailed challenges (exposures) with very high doses of a hybrid simian HIV, resulting in infection of all NHPs in the experiment. The risk of simian HIV transmission per exposure in such studies is much higher than in human-to-human HIV transmission, so that the failure of a vaccine candidate to protect from infection in this setting is not necessarily indicative of a vaccine without utility in humans [3]. Thus researchers have recently begun to conduct repeated low-dose challenge (RLC) studies in NHPs (e.g., [4–8]) that may more closely mimic typical exposure in natural human transmission settings. It has been shown that these experiments can be adequately powered to detect vaccine efficacy against infection with a feasible numbers of NHPs [9–11]. Because investigators control exposure, RLC experiments can assess vaccine effects of interest that may be difficult to observe in typical human efficacy trials. For example, infection status can be determined after each exposure via viral detection assays, facilitating estimation of the per-exposure or "per-contact" effect of

vaccination. Likewise, challenge studies allow precise characterization of immunological and virological parameters very early after infection, which are practically impossible to evaluate in humans (except possibly in neonates).

In addition to assessing vaccine effects on infection and post-infection endpoints, evaluation of immunological surrogates of protection (SoPs) are an important objective of NHP challenge experiments, providing points of reference to judge new vaccine candidates and for retrospective analyses of candidates evaluated previously in NHP experiments [9]. In general, a SoP is defined to be an immunological variable S (e.g., a T-cell or antibody measurement) such that a vaccine effect on S is predictive of a vaccine effect on the risk of infection or disease [12]. The utility of such a SoP includes guiding vaccine development, providing guidance for regulatory and immunization policy decisions, and bridging efficacy of a vaccine observed in a trial to a new setting. For RLC studies, knowledge of an immunological surrogate can inform comparisons of vaccine candidates in NHPs and support predictions of vaccine efficacy in humans.

Despite the importance of finding immune SoPs, methods for their quantitative assessment are quite limited. Moreover, there exists considerable confusion about what constitutes an immune correlate or surrogate of protection and how it should be appropriately evaluated. Recently Qin *et al.* [13] and Gilbert *et al.* [14] proposed a hierarchical three-tier framework for evaluating immune correlates: correlate of risk, specific SoP, and general SoP. A correlate of risk is an immunological measurement that correlates with the risk of a clinical endpoint (such as infection or disease) in a defined population. A specific SoP is a correlate of risk that is predictive of vaccine efficacy in a particular setting. A general SoP is a specific SoP that is also predictive of vaccine efficacy across different settings (e.g., across populations or across vaccine formulations). Meta-analysis of multiple vaccine studies is required for evaluating a general SoP whereas one study may be sufficient for evaluating a specific SoP. In this paper attention is restricted to the evaluation of a specific SoP (hereafter simply referred to as a 'SoP') from a single RLC study of a candidate HIV vaccine.

Traditionally, identification of potential SoPs has relied on solely assessing whether a immune response was a correlate of risk, i.e., associated with risk of infection or disease. For example, in the first phase III trial of an HIV vaccine, a significant negative association was found between risk of HIV infection and antibody response to the vaccine [15]. Unfortunately, this association-based analysis provides no information to distinguish between two possibilities: (i) a greater vaccine effect on the immune response predicted a greater vaccine effect on infection risk, or (ii) the immune response simply marked an innate ability to escape infection but did not predict vaccine efficacy. In other words, it was difficult based on the analysis results to conclude whether antibody response to the vaccine was a SoP or just a correlate of risk. A similar example is given by Ellenberger *et al.* [6], who found an association between vaccine induced ELISpot Gag responses and risk of simian HIV infection in a RLC study of a candidate HIV vaccine.

Recently, novel experimental designs and corresponding statistical methodology have been proposed for evaluating potential SoPs in the context of human efficacy trials [12, 16]. The central premise behind these designs is to attempt to infer the immune response control NHPs would have had if (counter to fact) they had been vaccinated. The first design, referred to as the baseline immunogenicity predictor (BIP) design, entails measuring a baseline covariate(s) W that is correlated with the immune response that NHPs would have to the HIV vaccine being evaluated. For example, W might be an immune response to a non-HIV vaccine. The missing HIV vaccine immune response for NHPs in the control arm can then be predicted or inferred (perhaps implicitly) from their W and a prediction model based on observed data from the vaccine group. In turn, the association between the vaccine effect on the immune endpoint and the vaccine effect on the infection endpoint can be assessed.

The second study design proposed for evaluation of SoPs in human efficacy trials is the close-out placebo vaccination design, where placebo recipients who are uninfected at the end of the trial are administered the HIV vaccine and their immune response is measured. In the RLC study setting, this design may not be feasible since most, if not all, control NHPs are often infected after repeated challenges, e.g., see [6, 7]. Therefore, in this paper we propose a modification of close-out placebo vaccination wherein after each challenge uninfected control NHPs are randomly assigned to receive vaccine according to a pre-specified crossover probability. This will be referred to as the crossover vaccination (CrV) design.

Simulation studies of large (e.g., Phase III) randomized human efficacy trials have demonstrated that the additional information provided by employing a BIP or close-out placebo vaccination design can enable assessing the extent to which a correlate of risk is a SoP [12, 16]. Similarly, [10] demonstrated the BIP design also holds promise in identifying SoPs in the RLC study setting; however, [10] did not consider a close-out placebo or CrV design. The feasibility of the BIP design relies on the existence of a baseline covariate W that is correlated with the HIV vaccine immune response. Whether such a covariate will be available may not be evident a priori, especially when evaluating new vaccine candidates. The CrV design, on the other hand, does not require a BIP. In order to further compare these two designs, in this paper methods are proposed for assessing potential binary immunological SoPs using either the BIP or CrV design. The operating characteristics of the different designs are then evaluated by simulating RLC studies.

The outline of the rest of this paper is as follows. In Section 2 the methods employed for analysis of RLC studies to determine SoPs using the BIP and CrV design are presented. The finite sample performance of the BIP and CrV designs under various scenarios are assessed in a simulation study in Section 3. The different approaches are applied to a recent RLC study in Section 4. A brief discussion is given in Section 5.

2. Methods

2.1. Notation, Assumptions, and Estimands

Following [10], consider the usual two-arm RLC study design that entails the following three sequential steps: (i) uninfected NHPs are randomized to receive vaccine or control, (ii) immune responses are measured, and (iii) low-dose simian HIV challenges are repeatedly administered until infection or some maximum number of challenges is reached. Let Z_i be the treatment assignment NHP i was randomized to, where $Z_i = 0$ for control and $Z_i = 1$ for vaccine. Using the Neyman potential outcome model, let $S_i(z)$ denote the potential HIV-specific immune response when assigned treatment z for $z = 0, 1$, where $S_i(z) = 0$ indicates the absence of the immune response and $S_i(z) = 1$ indicates the presence of the immune response. Note the immune response is being measured in step (ii), among uninfected NHPs that have yet to challenged with simian HIV. Therefore we assume $S_i(0) = 0$ for all i because vaccine antigens (absent in the control) are necessary to induce an HIV-specific immune response in an uninfected, unexposed NHP. Let $T_i(z)$ be the potential survival time under assignment $z = 0, 1$. Assume the survival time is measured by the number of exposures (i.e., challenges) until infection with simian HIV; thus $T_i(z)$ is a positive integer.

The models and likelihoods developed below rely on several key assumptions. First, assume no interference between NHPs and no variations in the type of vaccine such that the stable unit treatment value assumption holds [17]. The lack of interference between NHPs should hold in this setting since investigators can prevent transmission of simian HIV between NHPs. Second, assume ignorable treatment assignment, i.e., $Z_i \perp (S_i(0), S_i(1), T_i(0), T_i(1))$, which is insured by the use of randomization in assigning NHPs to receive vaccine or control. Third, assume the per-challenge probability of infection is independent of the prior number of challenges.

Let $p(Z_i, S_i(1))$ denote the probability of infection from a single exposure (i.e., the transmission probability) under assignment to treatment Z_i for NHPs with immune response $S_i(1)$ when vaccinated, i.e., $p(Z_i, S_i(1)) = \Pr[T_i(Z_i) = 1 | S_i(1)]$. Following [10, 16], consider the following probit model of the transmission probability

$$p(Z_i, S_i(1); \boldsymbol{\beta}) = \Phi\{\beta_1 + \beta_2 Z_i + \beta_3 S_i(1) + \beta_4 Z S_i(1)\}, \tag{1}$$

where Φ is the standard normal CDF and $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3, \beta_4)$. For NHPs for which the vaccine has no effect on the immune response, i.e., $S_i(1) = 0$, the value of β_1 determines the per-challenge risk of infection when not vaccinated and the parameter β_2 describes the vaccine effect on the per-challenge risk of infection. The parameter β_3 corresponds to differences in the per-challenge infection risk when not vaccinated between NHPs with an immune response to vaccine

(i.e., $S_i(1) = 1$) and NHPs without an immune response to vaccine (i.e., $S_i(1) = 0$). A key parameter of model (1) is β_4 where $\beta_4 \neq 0$ indicates that the vaccine effect on the per-challenge probability of infection differs between NHPs with $S_i(1) = 1$ and NHPs with $S_i(1) = 0$. The parameter β_4 is particularly useful in describing the surrogate value of $S_i(1)$ as discussed below.

Consider the principal stratification [18] of NHPs according to the potential immune response if vaccinated $S_i(1)$. Define the average causal effect of the vaccine on survival (i.e., infection) within principal stratum $S_i(1) = s$ to be $h(E\{T_i(0)|S_i(1) = s\}, E\{T_i(1)|S_i(1) = s\})$ for $s = 0, 1$, where h is some contrast function such that $h(x, y) = 0$ iff $x = y$, and $E\{T_i(z)|S_i(1) = s\}$ is the expected time to infection under treatment z in principal stratum $S_i(1) = s$. Assuming the transmission probability is independent of the previous number of challenges, $T_i(z)|S_i(1) = s$ is a geometric random variable with mean $E\{T_i(z)|S_i(1) = s\} = 1/p(z, s; \boldsymbol{\beta})$. Following [10], choosing $h_1(x, y) = \Phi^{-1}(1/x) - \Phi^{-1}(1/y)$ yields a simple expression for the principal stratum causal effect, namely $-\beta_2 - \beta_4 s$. Alternatively, for contrast function $h_2(x, y) = 1/x - 1/y$ the principal stratum causal effect equals the difference in risk of infection per challenge, which under model (1) is given by

$$p(0, s; \boldsymbol{\beta}) - p(1, s; \boldsymbol{\beta}) = \Phi\{\beta_1 + \beta_3 s\} - \Phi\{\beta_1 + \beta_2 + (\beta_3 + \beta_4)s\}.$$

Another possible contrast function $h_3(x, y) = x - y$ yields the difference in expected survival times within principal strata $S_i(1) = s$, which under model (1) equals

$$E\{T_i(0)|S_i(1) = s\} - E\{T_i(1)|S_i(1) = s\} = \Phi\{\beta_1 + \beta_3 s\}^{-1} - \Phi\{\beta_1 + \beta_2 + (\beta_3 + \beta_4)s\}^{-1}.$$

The proportion associative effect (PAE) statistic has been suggested [10, 12] as a summary measure of the surrogate value of S_i. PAE describes the proportion of the total effect of treatment (e.g., vaccine) represented by the expected associative effect (EAE) compared to the expected dissociative effect (EDE). Specifically, if EAE is the vaccine effect on infection within the stratum where $S_i(1) = 1$ and EDE is the vaccine effect on infection in the stratum where $S_i(1) = 0$, then PAE $= |\text{EAE}|/(|\text{EAE}| + |\text{EDE}|)$. For h_1 defined above EAE $= -\beta_2 - \beta_4$ and EDE $= -\beta_2$, implying

$$\text{PAE} = |\beta_2 + \beta_4|/\{|\beta_2 + \beta_4| + |\beta_2|\}, \tag{2}$$

where in general we adopt the convention $|0|/(|0| + |0|) = 0.5$, which for (2) implies PAE $= 0.5$ if $\beta_2 = \beta_4 = 0$. For contrast function h_2 the expression for PAE has a slightly more complicated form

$$\text{PAE} = \frac{|\Phi\{\beta_1 + \beta_3\} - \Phi\{\beta_1 + \beta_2 + \beta_3 + \beta_4\}|}{|\Phi\{\beta_1 + \beta_3\} - \Phi\{\beta_1 + \beta_2 + \beta_3 + \beta_4)\}| + |\Phi\{\beta_1\} - \Phi\{\beta_1 + \beta_2)\}|} \tag{3}$$

The PAE for h_3 can be expressed similarly. Note the dose in RLC is typically selected by investigators such that $p(Z_i, S_i(1)) \in [0.1, 0.5]$, suggesting for $z, s = 0, 1$ that $(\beta_1 + \beta_2 z + \beta_3 s + \beta_4 zs) \in [\Phi^{-1}(0.1), 0]$. Over this domain $\Phi(\cdot)$ is approximately linear (by Taylor series expansion $\Phi(x) \approx 1/2 + x/\sqrt{2\pi}$), suggesting that (2) and (3) will tend to be similar.

Regardless of the choice of contrast function, PAE $= 0.5$ if and only if EAE and EDE are equal in absolute value, i.e., the magnitude of the vaccine effect on time until infection T is the same for NHPs who would have an immune response to the vaccine compared to those who would not have an immune response. In this case, S has no surrogate value. On the other hand, PAE $= 1$ if and only if EDE $= 0$, indicating there is a vaccine effect on infection if and only if there is a vaccine effect on S; in this case S would be a perfect surrogate for T. Similarly, PAE $= 0$ if and only if EAE $= 0$, indicating there is a vaccine effect on infection if and only if there is no vaccine effect on S; in this case S would be a perfect surrogate for T as well, although in most contexts this scenario might be considered unrealistic. Biomarkers with some surrogate value will have PAE $\neq 0.5$, with the surrogate value increasing as PAE nears 0 or 1.

For simplicity in the sequel we will focus on (2); similar although slightly more complicated considerations apply if a contrast function other than h_1 is chosen by the analyst. Note for (2) that $\beta_4 = 0$ if and only if PAE = 0.5. Therefore we will focus on testing the null $H_0 : \beta_4 = 0$ versus the alternative $H_0 : \beta_4 \neq 0$, or equivalently $H_0 :$ PAE = 0.5 versus $H_A :$ PAE $\neq 0.5$. The likelihood ratio test can be employed to test H_0 versus H_A based on the likelihoods described below.

2.2. Standard RLC Design

The likelihood for the standard RLC design can be constructed as follows. For NHP i, let $T_i = \min\{T_i(0)(1 - Z_i) + T_i(1)Z_i, c_{max}\} = \min\{T_i(Z_i), c_{max}\}$ denote the observed number of exposures during the experiment where c_{max} denotes the right censoring time, i.e., the maximum allowable number of exposures. In general c_{max} may differ from NHP to NHP, but here for simplicity it is assumed c_{max} is the same for all NHPs. Assume c_{max} is a known constant and that $c_{max} > 1$, i.e., each NHP is challenged more than once provided infection does not occur after the first challenge. Let δ_i equal 1 (0) if NHP i is infected (uninfected) by the end of the study, i.e., $\delta_i = I(T_i(Z_i) \leq c_{max})$ where $I(\cdot)$ is the usual indicator function. Let $S_i = S_i(0)(1 - Z_i) + S_i(1)Z_i$ denote the observed immune response. Suppose n iid copies of $(Z_i, T_i, \delta_i, S_i)$ are observed. Letting G denote the marginal distribution function of $S_i(1)$, the likelihood is proportional to

$$L^{rlc}(\beta, G) = \prod_{i=1}^{n} \{\varphi(1, S_i, T_i, \delta_i; \boldsymbol{\beta})dG(S_i)\}^{Z_i} \left\{ \int \varphi(0, s, T_i, \delta_i; \boldsymbol{\beta})dG(s) \right\}^{1-Z_i}, \qquad (4)$$

where $\varphi(Z, S, T, \delta; \boldsymbol{\beta}) = \{1 - p(Z, S; \boldsymbol{\beta})\}^{T-\delta}p(Z, S; \boldsymbol{\beta})^{\delta}$. Because S is assumed binary in this paper, the integral in (4) could be written instead as a sum over $s = 0, 1$; a similar comment applies to the integrals given below as well. The more general form as in (4) is applicable for discrete or continuous S.

Note, because (1) is saturated, the likelihood (4) could equivalently be parameterized directly in terms of the geometric success probability parameters $p(z, s)$ instead of $\boldsymbol{\beta}$. For the vaccine arm, it is straightforward to show that the maximum likelihood estimate of $p(1, s)$ is $\hat{p}(1, s) = \sum_{i:Z_i=1, S_i=s} \delta_i / \sum_{i:Z_i=1, S_i=s} T_i$, i.e., the total number of infections divided by the total number of challenges in vaccinated NHPs with immune response s. Thus maximizing the likelihood is more straightforward using this parameterization, although the maximum likelihood estimator of $p(0, s)$ does not appear to have a closed form. The $\boldsymbol{\beta}$ parameterization has the advantage of summarizing the surrogate value of $S_i(1)$ with a single parameter, namely β_4.

In the standard RLC study design, $S_i(1)$ is only observed when $Z_i = 1$. Thus, intuitively, one might not expect $\boldsymbol{\beta}$ to be fully identifiable without additional information. To the contrary, under the assumptions stated above, $\boldsymbol{\beta}$ is in fact identifiable (provided $\Pr[S_i(1) = 0] \neq \Pr[S_i(1) = 1]$) provided $c_{max} > 2$. A proof is given in the Appendix. We note the proof does not require that (1) be a probit model; any invertible function from the real numbers $\mathbb{R}$ to $[0, 1]$ could be used in place of Φ and identifiability would still hold. The proof in the Appendix also shows that $\boldsymbol{\beta}$ is not identifiable for a single dose challenge design (or for a RLC with at most two challenges $c_{max} = 2$). Identifiability of $\boldsymbol{\beta}$ for the RLC design in turn implies identifiability of PAE. This has two important implications. First, the surprising result that $\boldsymbol{\beta}$ and PAE are identifiable explains why in the simulation study presented below that the standard RLC study design performs relatively well compared to the BIP and CrV designs. Second, that $\boldsymbol{\beta}$ and PAE are identifiable suggests that the surrogate value of a binary S may be assessed using standard RLC study designs, i.e., that the BIP and CrV designs are not necessary (although they may improve power to detect surrogate value).

Intuition regarding why $\boldsymbol{\beta}$ is identifiable for the RLC design but not the single dose challenge design can be gained by considering two coins, one red and one blue. Suppose the probability of landing heads upon a single toss is $p(0, 0; \boldsymbol{\beta}) = \Phi(\beta_1)$ for the red coin and $p(0, 1; \boldsymbol{\beta}) = \Phi(\beta_1 + \beta_3)$ for the blue coin. Consider the following simple experiment. One of the coins is selected randomly, with 2/3 probability of selecting the red coin (and thus 1/3 probability of selecting the blue coin). The selected coin is repeatedly tossed until the coin lands on heads or 10 tosses occur, which ever happens first. The number of tosses is recorded; the color of the selected coin is not recorded. The process of randomly selecting one of

the coins and repeatedly tossing until heads or 10 tosses occur is then repeated a large number of times; each time only the number of tosses is recorded. Now consider the scenario where the two coins have different probabilities of landing on heads; moreover, suppose the red coin is very likely to land on heads ($p(0, 0; \boldsymbol{\beta}) \approx 1$) and the blue coin is very unlikely to land on heads ($p(0, 1; \boldsymbol{\beta}) \approx 0$). Then the empirical distribution of the number of tosses until heads or 10 tosses will be bimodal, with 2/3 of the mass near 1 (corresponding to the red coin with $p(0, 0; \boldsymbol{\beta}) \approx 1$) and the other 1/3 of the mass near 10 (corresponding to the blue coin with $p(0, 1; \boldsymbol{\beta}) \approx 0$). On the other hand, consider the scenario where the two coins have the same probability of landing on heads and $p(0, 0; \boldsymbol{\beta}) = p(0, 1; \boldsymbol{\beta}) \approx 1/2$. In that case, the empirical distribution of the number of tosses will have a unimodal distribution with most of the mass near 1. Thus despite the recorded data not including information about the color of the coin being flipped during each iteration of the experiment, it is possible to distinguish scenarios where the two probabilities are different versus where they are the same. The two coins are analogous to the two strata of unvaccinated NHPs defined by $S_i(1) = 0$ and $S_i(1) = 1$; although we cannot determine strata membership on the basis of the observed data, the RLC experiment affords data that do provide information about the strata specific probabilities of infection per challenge. By the same token, if we imagine repeating the coin experiment described above, but are now only permitted to flip each selected coin a single time, clearly there is no hope of assessing from the recorded data alone whether or not the two coins have different probabilities of landing on heads. Likewise, in a single dose challenge study the data provide no information about the relative probability infection per challenge for the two strata of NHPs.

2.3. Baseline Immunogenicity Predictor Design

Let W_i denote a baseline covariate that might be correlated with $S_i(1)$ and now suppose n iid copies of $(Z_i, T_i, \delta_i, S_i, W_i)$ are observed. Letting G_w denote the conditional distribution of $S_i(1)$ given W_i, the likelihood is proportional to

$$L^{bip}(\beta, G_w) = \prod_{i=1}^{n} \{\varphi(1, S_i, T_i, \delta_i; \boldsymbol{\beta})dG_w(S_i|W_i)\}^{Z_i} \left\{\int \varphi(0, s, T_i, \delta_i; \boldsymbol{\beta})dG_w(s|W_i)\right\}^{1-Z_i}. \tag{5}$$

Note if $S_i(1)$ is assumed independent of W_i, then $G_w = G$ and (4) and (5) are equal. Also note the likelihood (5) relies on the assumption that the per-challenge risk of infection is independent of W_i conditional on $S_i(1)$ and Z_i. BIP models that relax this assumption are considered in [16], [12], and [10].

2.4. Crossover Vaccination Design

Next consider the likelihood for the CrV design. Recall in this design that after each challenge uninfected control NHPs are randomly assigned to receive vaccine according to a pre-specified crossover probability. For NHPs initially assigned to the control arm that cross-over, immune response to the vaccine $S_i(1)$ is measured after vaccination but prior to subsequent challenges. Denote the cross-over probability by c_p. Let T_i^0 denote the observed number of challenges when NHP i was unvaccinated and δ_i^0 be an indicator which equals one if challenge T_i^0 resulted in infection. Define T_i^1 and δ_i^1 similarly for when NHP i was vaccinated. For NHPs randomly assigned to receive vaccine at the start of the experiment (i.e., $Z_i = 1$), let $T_i^0 = \delta_i^0 = 0$. Note if a NHP is initially assigned vaccine ($Z_i = 1$) and becomes infected ($\delta_i^1 = 1$), then $T_i^1 = T_i(1)$. For NHPs randomly assigned to control at the start of the experiment (i.e., $Z_i = 0$) who become infected prior to crossover or reach the end of the experiment prior to crossover, let $T_i^1 = \delta_i^1 = 0$. Note if a NHP is initially assigned control ($Z_i = 0$) and becomes infected prior to crossover, then $T_i^0 = T_i(0)$. Let $\tilde{Z}_i$ indicate vaccination at some point during the study, i.e., $\tilde{Z}_i = Z_i + (1 - Z_i)I[T_i^1 > 0]$. Suppose n iid copies of $(\tilde{Z}_i, T_i^0, \delta_i^0, T_i^1, \delta_i^1, S_i(1)I[\tilde{Z}_i = 1])$ are observed. Note it is being assumed the following two immune responses are equal: (i) the immune response observed when an unvaccinated NHP that has escaped infection from one or more challenges crosses over and is vaccinated, and (ii) the immune response that NHP would have had if, counter to fact, they had been vaccinated at baseline. The likelihood for the CrV design is

proportional to

$$L^{crv}(\beta, G) = \prod_{i=1}^{n} \left\{ (\varphi(0, S_i(1), T_i^0, \delta_i^0; \boldsymbol{\beta}) \varphi(1, S_i(1), T_i^1, \delta_i^1; \boldsymbol{\beta}) dG(S_i(1)) \right\}^{\tilde{Z}_i} \left\{ \int \varphi(0, s, T_i^0, \delta_i^0; \boldsymbol{\beta}) dG(s) \right\}^{1-\tilde{Z}_i} \quad (6)$$

Note if the cross-over probability is set to zero (i.e., $c_p = 0$), then $Z_i = \tilde{Z}_i$ and L^{crv} simplifies to the likelihood for a standard RLC design L^{rlc} given in (4). If the CrV and BIP designs are combined such that W_i is also observed for $i = 1, \ldots, n$, then the likelihood is the same as L^{crv} except that $dG(\cdot)$ is replaced by $dG_w(\cdot|W_i)$ in (6).

3. Simulation study

3.1. Standard RLC Design

A simulation study was conducted to assess whether standardly designed RLC studies of typical moderate sample size can provide adequate power to detect immune responses with high surrogate value. Data were generated assuming $n/2$ NHPs were randomized to each arm (vaccine or control), a maximum number of exposures per NHP of $c_{max} = 10$, and model (1). Simulations were conducted under different scenarios reflecting the degree of association between a binary BIP W_i and immune response $S_i(1)$. In particular, for each NHP, W_i was randomly generated from a Bernoulli distribution with mean 0.5. The immune response under vaccine $S_i(1)$ was then generated under three different scenarios. Under the first scenario, $S_i(1)$ was generated from a Bernoulli distribution with mean 0.3 and independent of W_i such that $\mathrm{OR} = \mathrm{odds}(S_i(1) = 1|W_i = 1)/\mathrm{odds}(S_i(1) = 1|W_i = 0) = 1$. Under the second scenario, $S_i(1)$ was generated from a Bernoulli with mean $0.2 + 0.2W_i$ such that $\mathrm{OR} = 2.67$. Under the third scenario, $S_i(1)$ was generated from a Bernoulli with mean $0.1 + 0.4W_i$ such that $\mathrm{OR} = 9$. Note in all three scenarios the marginal distribution of $S_i(1)$ was Bernoulli with mean 0.3. Values of $\boldsymbol{\beta}$ were selected such that the probability of infection per challenge for controls was 0.5 regardless of $S_i(1)$, i.e., $\beta_1 = \beta_3 = 0$. To reflect immune responses with varying surrogate values, simulated data sets were generated under three scenarios corresponding to $\mathrm{PAE} \in \{0.5, 0.75, 1\}$ for contrast function h_1. Specifically, for the first scenario $\boldsymbol{\beta} = (0, -1, 0, 0)$ such that $\mathrm{PAE} = 0.5$; for the second scenario $\boldsymbol{\beta} = (0, -1, 0, -2)$ such that $\mathrm{PAE} = 0.75$; and for the third scenario $\boldsymbol{\beta} = (0, 0, 0, -3)$ such that $\mathrm{PAE} = 1$. RLC studies were simulated with total sample sizes $n = 30, 40, 50$. For each combination of OR, PAE, and n, 5000 simulated data sets were generated. For each data set the likelihood ratio test was used to assess $H_0 : \beta_4 = 0$ versus $H_A : \beta_4 \neq 0$ at the $\alpha = 0.05$ significance level.

Simulation results are summarized in Tables $1 - 2$. Empirical type I error and power for a standard RLC design (i.e., without or ignoring W_i) are given in Table 1. For $S_i(1)$ with no surrogate value, i.e., $\beta_4 = 0$ and $\mathrm{PAE} = 0.50$, the type I error of the likelihood ratio test did not exceed the nominal level $\alpha = 0.05$ (results were similar under the null for all simulations discussed below). For $S_i(1)$ with moderate surrogate value, i.e., $\beta_4 = -2$ and $\mathrm{PAE} = 0.75$, there was low power (17–49%), even if the total sample size was $n = 50$. For $S_i(1)$ a perfect surrogate, i.e., $\beta_4 = -3$ and $\mathrm{PAE} = 1$, there was decent power (71–82%) provided the total sample size was at least $n = 40$.

Empirical power for a standard RLC design with a (supposed) BIP is given in Table 2. For W_i with a strong association with $S_i(1)$, i.e., $\mathrm{OR} = 9$, the presence of the BIP led to an increase in power, although the increase was often modest. For example, for $n = 50$ and $\mathrm{PAE} = 0.75$ power was 0.49 without the BIP compared to 0.54 with the BIP. For W_i with a moderate association with $S_i(1)$, i.e., $\mathrm{OR} = 2.67$, the presence of a BIP generally did not increase power and in fact, perhaps surprisingly, sometimes led to a decrease in power. For example, for $n = 40$ and $\mathrm{PAE} = 1$ power was 0.71 without a BIP compared to 0.67 with the BIP. Presumably this decrease in power can be attributed to the BIP likelihood (5) requiring estimation of the two conditional distributions G_w for $W_i = 0, 1$ whereas the standard RLC likelihood (4) requires only estimating the marginal distribution G. The cost of estimating the additional parameters associated with G_w is readily apparent by comparing the power when $\mathrm{PAE} = 1$ in Table 1 with Table 2 when $\mathrm{OR} = 1$; in this case W_i provides

no information about $S_i(1)$ and therefore it is better to ignore the putative BIP.

3.2. CrV Design

Simulations studies were also conducted to evaluate the utility of the CrV design as a means to potentially increase the power of RLC studies to detect immune responses with high surrogate value. Different values of the crossover probability c_p were investigated to ascertain the effect of c_p on the power to detect surrogate value. Results for $c_p \in \{0.1, 0.25, 0.5\}$ are given in Table 3. In comparison to Table 1, CrV resulted in an increase in power for $c_p = 0.1$. For example, for $n = 30$ and PAE = 0.75, power was 0.17 for the standard RLC design compared to 0.32 for the CrV design with $c_p = 0.1$, i.e., the CrV design had almost twice the power. Results for $c_p = 0.25$ and $c_p = 0.5$ were mixed, with generally an increase in power for PAE = 0.75 and a decrease in power for PAE = 1 relative to the standard RLC design.

Additional simulation studies were undertaken to examine usefulness of the CrV design in the presence of a BIP W_i having strong association with $S_i(1)$, i.e., OR = 9. These results are given in Table 4. Comparison with Table 2 suggests that CrV is of modest utility in the presence of a strong BIP. For example, for $n = 30$ and PAE = 1, the empirical power was 0.58 without CrV compared to 0.65 for CrV with $c_p = 0.1$. Similarly, comparing Table 3 and 4 indicates a BIP having strong association with $S_i(1)$ does not meaningfully increase power in the CrV design. For instance, for $n = 30$, PAE = 1, and $c_p = 0.1$, the empirical power was 0.61 in the absence of a BIP compared to 0.65 when OR = 9.

The results in Tables 1, 3, and 4 suggest power for the CrV design is a non-monotonic function of c_p. For the simulations presented here $c_p = 0.1$ yielded the greatest power. However, for other parameter configurations a different cross-over probability may be optimal. Therefore it is important that investigators planning a CrV RLC study conduct additional simulation studies for parameter values tailored to the particular setting where the RLC study will take place in order to determine the optimal value of c_p.

One potential drawback of the CrV design is that there are fewer challenges of unvaccinated NHPs, potentially diminishing the power to detect a vaccine effect on the primary endpoint, namely infection. Therefore simulation studies were also conducted to assess how c_p affects the power to detect a vaccine effect on the risk of infection. Let VE $= 1 - p(1)/p(0)$ where $p(z)$ denotes the (marginal) risk of infection per challenge if a NHP is assigned treatment $z = 0, 1$. In words, VE is the relative reduction in the per challenge risk of infection due to vaccination. RLC studies were simulated under different scenarios to assess the power of the likelihood ratio test to reject the null hypothesis of no vaccine effect on the risk of infection, i.e., H_0 : VE = 0. Simulations were performed for VE $\in \{0, 0.25, 0.50, 0.75\}$ and $c_p \in \{0, 0.1, 0.25, 0.5\}$. Results are shown in Table 5. For each scenario the average number of NHPs that crossed-over was in good agreement with the expected number of cross-overs, which in general equals $(n/2) \sum_{j=1}^{c_{max}-1} \{1 - p(0)\}^j (1 - c_p)^{j-1} c_p$. The empirical power of the likelihood ratio test presented in Table 5 demonstrates that the CrV design may result in modest diminution of power to detect a vaccine effect on infection relative to the standard RLC study design. For example, for VE = 0.50 the empirical power was 0.86 when $c_p = 0$ compared to 0.83 when $c_p = 0.25$. These results suggest the CrV design could be implemented without adversely affecting power for the primary endpoint (infection) while potentially increasing power to detect a SoP.

3.3. Additional Simulations

In the simulations described above $\beta_1 = \beta_3 = 0$ such that the probability of infection per challenge for unvaccinated NHPs was 0.5 regardless of $S_i(1)$, i.e., $p(0, 1; \boldsymbol{\beta}) = p(0, 0; \boldsymbol{\beta}) = 0.5$. Additional simulation studies were conducted to assess the power of RLC studies to detect an immune response with surrogate value when the per challenge infection probability differs between unvaccinated NHPs with $S_i(1) = 1$ versus $S_i(1) = 0$, i.e., when $\beta_3 \neq 0$. In particular, the simulation study described above for a standard RLC design without a BIP (the results of which are given in Table 1) was repeated with three different combinations of values for β_1 and β_3 such that $p(0, 1; \boldsymbol{\beta}) < p(0, 0; \boldsymbol{\beta})$ but the marginal probability of infection per challenge when unvaccinated remained 0.5 as in Table 1. For simulations where PAE = 0.5 (i.e., $\beta_4 = 0$), this correspond to the scenario where the immune response $S_i(1)$ is a correlate of infection risk but not predictive of the vaccine effect on

risk of infection. The results for these additional simulations are given in Table 6 and are comparable to those in Table 1, with only a slight reduction in power as the difference in $p(0, 1; \boldsymbol{\beta})$ and $p(0, 0; \boldsymbol{\beta})$ increases.

4. Application

Ellenberger et al. [6] reported results from a RLC experiment with 16 vaccinated NHPs and 14 control NHPs. All animals were challenged weekly with simian HIV and evidence of infection was assessed after each challenge. Previous analyses of these data indicate the vaccine caused a significant reduction in the per challenge probability of infection [6, 10]. Analysis of various immune response measurements from vaccinated NHPs suggested possible correlation between the number of challenges until infection and immune responses to Gag peptide pools as measured by the ELISpot assay (see Figure 3 from [6]). For the analysis here, we let $S_i = 1$ if the total ELISpot response to Gag was at least 1000 per 10^6 PBMC and $S_i = 0$ otherwise. The observed data $(Z_i, T_i, \delta_i, S_i)$ are presented in Table 7.

Fitting model (1) to these data based on (4), the likelihood ratio test of $H_0 : \beta_4 = 0$ yields a two-sided p-value $p = 0.43$, providing no indication that S_i has surrogate value. Suppose, however, contrary to fact, that a BIP W_i had been collected for all NHPs that correlated well (but not perfectly) with S_i in the vaccine arm of the study; such a W_i is given in Table 7. In this case fitting model (1) based on (5), the likelihood ratio test of $H_0 : \beta_4 = 0$ yields $p = 0.04$ and the maximum likelihood estimate of PAE is $\widehat{\mathrm{PAE}} = 0.87$, indicating that S_i has significant surrogate value. While finding a BIP W_i that correlates well with $S_i(1)$ may not always be feasible in practice, this example illustrates the potential utility of a BIP in discerning the surrogate value of $S_i(1)$. Similarly, suppose contrary to fact that the CrV design had been employed in this experiment such that four of the NHPs initially assigned control were randomly crossed-over after escaping infection from at least one challenge prior to vaccination. One possible observed data set from such an experiment is given in Table 8. For these data, fitting model (1) based on (6), the likelihood ratio test of $H_0 : \beta_4 = 0$ yields $p = 0.04$ and $\widehat{\mathrm{PAE}} = 0.87$, indicating that S_i has significant surrogate value. This illustrates the potential utility of the CrV design in the absence of a BIP.

It is important to keep in mind that while in this example the presence of a BIP or employing the CrV design has a substantial effect on the resulting inference drawn about the surrogate value of $S_i(1)$, for other data sets this may not be the case. Moreover, the simulation results from the previous section indicate gains in power due to a BIP or the CrV design tend not to be substantial.

5. Discussion

A challenge in determining the surrogate value of an immune response in a standard RLC study is that $S_i(1)$ is missing for unvaccinated NHPs. Following previous work on human vaccine trials (e.g., [12–14, 16]), in this paper we considered evaluating surrogate value in a standard RLC study design with and without a BIP. We also considered the CrV design, an adaptation of the close-out placebo vaccination design [16], that reveals $S_i(1)$ for some NHPs that initially were unvaccinated. There are four main findings from this paper. First, perhaps surprisingly, it is possible to detect an immune response with surrogate value in a standard RLC study without a BIP; this is not true of a single challenge study design. Second, the presence of a BIP does not appreciably increase power, even if the correlation with the immune response is high. Third, CrV can lead to an increase in power, although this gain is sometimes modest and dependent on the cross-over probability selected. Fourth, regardless of whether a BIP is present or CrV is employed, in order to have adequate power to detect a surrogate of protection (i) the vaccine effect on the immune response must be strongly associated with the vaccine effect on risk of infection (PAE ≈ 1) and (ii) the sample size must be relatively large (e.g., $n = 50$).

There are several areas of possible future research related to evaluating immunological surrogates of protection in RLC studies. First, development of exact-type methods for this setting might be of utility given the relatively small sample sizes typical of RLC studies. While the maximum likelihood approach used in this paper performed well in simulation studies (e.g., the nominal type I error was preserved), it might be preferred to use methods that do not rely on large sample approximations. Indeed the likelihood ratio test performed conservatively in some of the simulation studies presented (e.g., in Tables 1 and 2 the empirical type I error was less than the nominal level), suggesting more powerful tests might be developed. Some possibilities would be to consider higher-order approximations, such as a Bartlett correction [19], to the likelihood ratio test or resampling techniques like the bootstrap as in [10]. Alternatively, a Bayesian approach (e.g., [20, 21]) might be considered if one was willing to entertain the attendant assumptions. A second area of possible future research would be to examine continuous or multivariate S. In this paper we consider only a binary S, which simplifies matters (such as proving identifiability or lack thereof) and is also of practical relevance as binary immune responses (e.g., seroconversion) are often used operationally in vaccine studies as surrogates of protection. If the immune response outcome of interest is continuous (or ordinal), the methods developed in this paper readily apply by choosing different cut-points or positivity thresholds to create a binary S.

The methods in this paper could also be extended to relax the assumption that the transmission probability is independent of the number of previous challenges. One approach in this direction would be to develop a beta-geometric model; Hudgens and Gilbert [10] applied such a model to the Ellenberger et al. RLC experiment and found the time homogeneous transmission probability model fit the data better. Similarly, within the context of the CrV design, methods could be developed to relax the time constancy assumption that (i) the immune response measured after escaping infection from one or more challenges and then being vaccinated is the same as (ii) the immune response that would have been measured if vaccinated prior to any challenges. This assumption would be violated if being challenged but not infected causes the immune response to vaccine to be altered. A lack of association between the observed number of challenges when not vaccinated T_i^0 and observed immune response when vaccinated $S_i(1)$ would be supportive of this assumption. However, presence of such an association does not necessarily indicate the time constancy assumption is invalid because NHPs with missing $S_i(1)$ in the CrV design are not necessarily a random subset of NHPs in the experiment. In the context of human efficacy trials, Follmann [16] suggests some additional approaches to assessing this time constancy assumption which rely on invoking other additional assumptions. Until methods which formally relax this assumption are developed within the context of RLC studies, analysts should rely on subject-matter experts as to the feasibility of this assumption. If the time constancy assumption of the CrV design is not considered feasible, then inferences drawn using the methods described in this paper may be erroneous and should therefore be interpreted with caution.

Finally we note that as high dimensional immune response data becomes more readily available on each NHP, a natural next area to explore would be the application of supervised learning methods for this "large p, small n" setting in identifying immune SoPs.

Appendix

Identifiability of $\boldsymbol{\beta}$ in standard RLC study deisgn

Here we show $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3, \beta_4)$ is identifiable in the standard RLC study design (i.e., without CrV or BIP) for $S_i(1)$ binary. Because treatment assignment is random and $S_i(1)$ is observed in all NHPs randomized to $Z_i = 1$, G is identifiable and can be regarded as fixed and known. Assume $\beta_1, \ldots, \beta_4$ are all finite, so that there is a positive probability of observing (or not) an infection under either treatment assignment. Also assume G is such that $0 < \Pr[S_i(1) = s] < 1$ for $s = 0, 1$. Let $\kappa = \Pr[S_i(1) = 1]$.

Under these assumptions, it is straightforward to show that $\beta_1 + \beta_2$ and $\beta_3 + \beta_4$ are identifiable. To see this, suppose there are two parameterizations $\beta = (\beta_1, \beta_2, \beta_3, \beta_4)$ and $\tilde{\beta} = (\tilde{\beta}_1, \tilde{\beta}_2, \tilde{\beta}_3, \tilde{\beta}_4)$ such that

$$\Pr[T_i = t, \delta_i = d, S_i = s | Z_i = z; \beta] = \Pr[T_i = t, \delta_i = d, S_i = s | Z_i = z; \tilde{\beta}] \tag{7}$$

for all possibly observed combinations of t, d, s, z. Then for $t = 1$, $d = 1$, $s = 0$, $z = 1$, (7) implies $\Phi\{\beta_1 + \beta_2\} = \Phi\{\tilde{\beta}_1 + \tilde{\beta}_2\}$. Because Φ is invertible, this implies $\beta_1 + \beta_2 = \tilde{\beta}_1 + \tilde{\beta}_2$. Similarly it can be shown $\beta_3 + \beta_4 = \tilde{\beta}_3 + \tilde{\beta}_4$. That is, $\beta_1 + \beta_2$ and $\beta_3 + \beta_4$ are identifiable. The more interesting issue germane to assessing the surrogate value of S is whether β_2 and β_4 individually are identifiable from the observable data from a RLC study. To prove this, it will be shown below that β_1 and β_3 are identifiable provided $\Pr[S_i(1) = 1] \neq 1/2$ and $c_{max} > 2$.

As a brief aside, first consider a single dose challenge study (i.e., $c_{max} = 1$). Then for $t = 1$, $d = 1$, $z = 0$, (7) implies

$$\int \Phi(\beta_1 + \beta_3 s) dG(s) = \int \Phi(\tilde{\beta}_1 + \tilde{\beta}_3 s) dG(s) \tag{8}$$

The only other possible pattern of observed data under $Z_i = 0$ would be $t = 1, d = 0, z = 0$, in which case (7) implies

$$\int \{1 - \Phi(\beta_1 + \beta_3 s)\} dG(s) = \int \{1 - \Phi(\tilde{\beta}_1 + \tilde{\beta}_3 s)\} dG(s) \tag{9}$$

Suppose without loss of generality that $\tilde{\beta}_3 \neq 0$, and let $\beta_3 = 0$ and $\beta_1 = \Phi^{-1}\{\int \Phi(\tilde{\beta}_1 + \tilde{\beta}_3 s) dG(s)\}$. Then (8) and (9) hold, yet $\beta_3 \neq \tilde{\beta}_3$. That is, β_1 and β_3 are not identifiable in a single dose challenge study.

One can also show β_1 and β_3 are not identifiable for a RLC study where $c_{max} = 2$. To prove this, let $c_1 = (1 - \kappa)\Phi(\beta_1) + \kappa\Phi(\beta_1 + \beta_3)$ and $c_2 = (1 - \kappa)\Phi(\beta_1)^2 + \kappa\Phi(\beta_1 + \beta_3)^2$. Note $\Pr[T_i = 1, \delta_i = 1, S_i = 0 | Z_i = 0] = c_1$, $\Pr[T_i = 2, \delta_i = 1, S_i = 0 | Z_i = 0] = c_1 - c_2$, and $\Pr[T_i = 2, \delta_i = 0, S_i = 0 | Z_i = 0] = 1 - 2c_1 + c_2$. It is straight-forward to show that $\Phi(\beta_1)$ equals either $c_1 + \{\kappa(c_2 - c_1^2)/(1 - \kappa)\}^{1/2}$ or $c_1 - \{\kappa(c_2 - c_1^2)/(1 - \kappa)\}^{1/2}$. Without loss of generality, assume $\Phi(\beta_1)$ equals the former and let $\tilde{\beta}_1$ be such that $\Phi(\tilde{\beta}_1)$ equals the latter. Let $\tilde{\beta}_3$ be such that $(1 - \kappa)\Phi(\tilde{\beta}_1) + \kappa\Phi(\tilde{\beta}_1 + \tilde{\beta}_3) = c_1$. Then one can show $(1 - \kappa)\Phi(\tilde{\beta}_1)^2 + \kappa\Phi(\tilde{\beta}_1 + \tilde{\beta}_3)^2 = c_2$, implying (7) holds. Yet $c_2 > c_1^2$, implying $\beta_1 \neq \tilde{\beta}_1$.

Now consider an RLC studies where $c_{max} > 2$. Below it is shown that if $\kappa \neq 1/2$, then β_1 and β_3 are identifiable. On the other hand, if $\kappa = 1/2$, then β_1 and β_3 are not identifiable. To begin, note for $z = 0$ that (7) can be equivalently expressed as

$$(1 - \kappa)A^{t-d}(1 - A)^d + \kappa B^{t-d}(1 - B)^d = (1 - \kappa)C^{t-d}(1 - C)^d + \kappa D^{t-d}(1 - D)^d \tag{10}$$

for all t, d where $A = 1 - \Phi(\beta_1)$, $B = 1 - \Phi(\beta_1 + \beta_3)$, $C = 1 - \Phi(\tilde{\beta}_1)$, $D = 1 - \Phi(\tilde{\beta}_1 + \tilde{\beta}_3)$. If $A = C$, then (10) implies $B = D$ and identifiability is proved. So assume by way of contradiction $A \neq C$, and thus $D \neq B$. For $d = 0$ and $t = 1, 2, 3$, (10) implies

$$(1 - \kappa)(A - C) = \kappa(D - B) \tag{11}$$
$$(1 - \kappa)(A^2 - C^2) = \kappa(D^2 - B^2) \tag{12}$$
$$(1 - \kappa)(A^3 - C^3) = \kappa(D^3 - B^3) \tag{13}$$

Dividing (12) and (13) by (11) yields

$$A + C = D + B \tag{14}$$
$$A^2 + AC + C^2 = D^2 + DB + B^2 \tag{15}$$

Squaring (14) and subtracting the result by (15) yields

$$AC = BD \tag{16}$$

Squaring (14) and then subtracting the result by four times (16) on both sides gives

$$(A - C)^2 = (D - B)^2 \tag{17}$$

Assume without loss of generality that $A > C$. Then (11) implies $D > B$ which by (17) in turn implies

$$A - C = D - B. \tag{18}$$

If $\kappa \neq 1/2$, then (11) and (18) cannot simultaneously be true if $A \neq C$ and $B \neq D$, and therefore we have a contradiction. Thus if $\kappa \neq 1/2$, then β_1 and β_3 are identifiable, and likewise β_2 and β_4.

Finally, suppose $\kappa = 1/2$. Note equation (7) for $z = 0$ is equivalent to

$$A^{t-d}(1 - A)^d + B^{t-d}(1 - B)^d = C^{t-d}(1 - C)^d + D^{t-d}(1 - D)^d \tag{19}$$

for all t, d. Suppose without loss of generality that $\tilde{\beta}_3 \neq 0$ and let $\beta_1 = \tilde{\beta}_1 + \tilde{\beta}_3$ and $\beta_3 = -\tilde{\beta}_3$ such that $A = D$ and $B = C$. Then (19) holds for all t, d even though $\beta_1 \neq \tilde{\beta}_1$ and $\beta_3 \neq \tilde{\beta}_3$.

Acknowledgement

References

1. Morgan C, Marthas M, Miller C, Duerr A, Cheng-Mayer C, Desrosiers R, Flores J, Haigwood N, Hu S, Johnson R, *et al.*. The use of nonhuman primate models in HIV vaccine development. *PLoS Medicine* 2008; **5**(8): e173.

2. Fauci A, Johnston M, Dieffenbach C, Burton D, Hammer S, Hoxie J, Martin M, Overbaugh J, Watkins D, Mahmoud A, Greene W. HIV vaccine research: the way forward. *Science* 2008; **321**:530–532.

3. Feinberg MB, Moore JP. AIDS vaccine models: challenging challenge viruses. *Nature Medicine* 2002; **8**:207–210.

4. Roederer M, Keele BF, Schmidt SD, Mason RD, Welles HC, Fischer W, Labranche C, *et al.*. Immunological and virological mechanisms of vaccine-mediated protection against SIV and HIV. *Nature* 2013; **505**:502–508.

5. García-Lerma J, Otten R, Qari S, Jackson E, Cong M, Masciotra S, Luo W, Kim C, Adams D, Monsour M, *et al.*. Prevention of rectal SHIV transmission in macaques by daily or intermittent prophylaxis with emtricitabine and tenofovir. *PLoS Med.* 2008; **5**:e28.

6. Ellenberger D, Otten RA, Li B, Rodriguez V, Sariol CA, Martinez M, Monsour M, Wyatt L, *et al.*. HIV-1 DNA/MVA vaccination reduces the per exposure probability of infection during repeated mucosal SHIV challenges. *Virology* 2006; **352**:216–225.

7. Subbarao S, Otten R, Ramos A, Jackson E, Monsour M, Bashirian S, Kim C, Johnson J, Soriano V, *et al.*. Chemoprophylaxis with tenofovir disoproxil fumarate provided partial protection against simian human immunodeficiency virus infection in macaques given multiple virus challenges. *Journal of Infectious Diseases* 2006; **194**:904–11.

8. Otten R, Adams D, Kim C, Jackson E, Pullium J, Lee K, Grohskopf L, Monsour M, Butera S, Folks T. Multiple vaginal exposures to low doses of R5 simian-human immunodeficiency virus: strategy to study HIV preclinical interventions in nonhuman primates. *Journal of Infectious Diseases* 2005; **191**:164–173.

9. Regoes RR, Longini IM, Feinberg MB, Staprans SI. Preclinical assessment of HIV vaccines and microbicides by repeated low-dose virus challenges. *PLoS Medicine* 2005; **2(8)**:e249.

10. Hudgens MG, Gilbert PB. Assessing vaccine effects in repeated low-dose challenge experiments. *Biometrics* 2009; **65**:1223–1232.

11. Hudgens M, Gilbert P, Mascola J, Wu C, Barouch D, Self S. Power to detect HIV vaccine effects in repeated low-dose challenge experiments. *Journal of Infectious Diseases* 2009; **200**(4):609–613.

12. Gilbert PB, Hudgens MG. Evaluating candidate principal surrogate endpoints. *Biometrics* 2008; **64**:1146–1154.

13. Qin L, Gilbert P, Corey L, McElrath M, Self S. A framework for assessing immunological correlates of protection in vaccine trials. *Journal of Infectious Diseases* 2007; **196**:1304–1312.

14. Gilbert PB, Qin L, Self SG. Evaluating a surrogate endpoint at three levels, with application to vaccine development. *Statistics in Medicine* 2008; **27**:4758–4778.

15. Gilbert PB, Peterson ML, Follmann D, Hudgens MG, Francis DP, Gurwith M, Heyward WL, Jobes DV, Popovic V, Self SG, *et al.*. Correlation between immunologic responses to a recombinant glycoprotein 120 vaccine and incidence of HIV-1 infection in a phase 3 HIV-1 preventive vaccine trial. *Journal of Infectious Diseases* 2005; **191**:666–77.

16. Follmann D. Augmented designs to assess immune response in vaccine trials. *Biometrics* 2006; **62**:1161–1169.

17. Rubin DB. Discussion of "Randomization analysis of experimental data in the Fisher randomization test" by D. Basu. *Journal of the American Statistical Association* 1980; **75**:591–593.

18. Frangakis CE, Rubin DB. Principal stratification in causal inference. *Biometrics* 2002; **58**:21–29.

19. Cordeiro GM, Cribari-Neto, F. *An Introduction to Bartlett Correction and Bias Reduction*. Springer, 2014.

20. Zigler CM, Belin TR. A Bayesian approach to improved estimation of causal effect predictiveness for a principal surrogate endpoint. *Biometrics* 2012; **68**(3):922–932.

21. Li Y, Taylor JM, Elliott MR. A Bayesian approach to surrogacy assessment using principal stratification in clinical trials. *Biometrics*, 2010; **66**(2): 523–531.

Table 1. Empirical type I error and power for standard RLC study design based on simulation study described in Section 3. Each table entry is based on 5000 simulated RLC trials with n NHPs total, of which $n/2$ were randomized to each study arm. Infection outcomes were simulated according to model (1) and PAE is given by (2). Table entries equal the empirical type I error or power for the likelihood ratio test of $H_0 : \beta_4 = 0$ versus $H_A : \beta_4 \neq 0$ at significance level $\alpha = 0.05$.

β_4	PAE	$n = 30$	$n = 40$	$n = 50$
0	0.50	0.01	0.01	0.01
-2	0.75	0.17	0.32	0.49
-3	1.00	0.53	0.71	0.82

Table 2. Empirical type I error and power of likelihood ratio test of $H_0 : \beta_4 = 0$ for standard RLC study design with a binary BIP where $\text{OR} = \text{odds}(S_i(1) = 1|W_i = 1)/\text{odds}(S_i(1) = 1|W_i = 0)$.

			$n = 30$			$n = 40$			$n = 50$	
β_4	PAE	OR=1	OR=2.67	OR=9	OR=1	OR=2.67	OR=9	OR=1	OR=2.67	OR=9
0	0.50	0.01	0.02	0.02	0.01	0.01	0.02	0.01	0.02	0.02
-2	0.75	0.18	0.20	0.29	0.30	0.33	0.43	0.42	0.45	0.54
-3	1.00	0.50	0.53	0.58	0.65	0.67	0.73	0.76	0.77	0.82

Table 3. Empirical type I error and power of likelihood ratio test of $H_0 : \beta_4 = 0$ for the CrV study design without a BIP. After each exposure, uninfected NHPs randomized to the control arm randomly crossed-over to the vaccine arm with probability c_p.

			$n = 30$			$n = 40$			$n = 50$	
β_4	PAE	$c_p = .1$	$c_p = .25$	$c_p = .5$	$c_p = .1$	$c_p = .25$	$c_p = .5$	$c_p = .1$	$c_p = .25$	$c_p = .5$
0	0.50	0.03	0.04	0.06	0.03	0.05	0.06	0.03	0.05	0.06
-2	0.75	0.32	0.36	0.28	0.50	0.49	0.38	0.61	0.61	0.45
-3	1.00	0.61	0.56	0.37	0.79	0.70	0.47	0.89	0.80	0.56

Table 4. Empirical type I error and power of likelihood ratio test of $H_0 : \beta_4 = 0$ for the CrV study design in the presence of a binary BIP with cross-over probability c_p and $\text{OR} = \text{odds}(S_i(1) = 1|W_i = 1)/\text{odds}(S_i(1) = 1|W_i = 0) = 9$.

			$n = 30$			$n = 40$			$n = 50$	
β_4	PAE	$c_p = .1$	$c_p = .25$	$c_p = .5$	$c_p = .1$	$c_p = .25$	$c_p = .5$	$c_p = .1$	$c_p = .25$	$c_p = .5$
0	0.50	0.04	0.04	0.06	0.04	0.05	0.05	0.04	0.04	0.05
-2	0.75	0.36	0.35	0.30	0.50	0.48	0.38	0.62	0.57	0.46
-3	1.00	0.65	0.57	0.41	0.79	0.70	0.51	0.86	0.78	0.59

Table 5. Empirical type I error and power to reject $H_0 : \text{VE} = 0$ where $\text{VE} = 1 - p(1)/p(0)$ is the relative reduction in the per-challenge risk of infection due to vaccination. Simulation results shown are based on $p(0) = 0.5$ and $n = 50$ NHPs total, with 25 randomized to vaccine initially. The average number of control NHPs that crossed-over are given in parentheses. The parenthetical numbers in the column headers denote the (theoretical) expected number of NHPs to crossover.

VE	Crossover probability c_p							
	0	$(0^\dagger)$	0.1	(2.3)	0.25	(5.0)	0.5	(8.3)
0	0.05	$(0^\ddagger)$	0.05	(2.3)	0.05	(5.0)	0.05	(8.4)
0.25	0.26	(0)	0.28	(2.3)	0.25	(4.9)	0.22	(8.3)
0.50	0.86	(0)	0.85	(2.2)	0.83	(5.0)	0.78	(8.3)
0.75	1.00	(0)	1.00	(2.3)	1.00	(5.0)	1.00	(8.3)

† Expected crossovers

‡ Average simulated number of NHPs that crossed-over

Table 6. Empirical type I error and power of likelihood ratio test of $H_0 : \beta_4 = 0$ for standard RLC study design as in Table 1 for different values of the unvaccinated per-infection transmission probabilities $p(0, 1; \boldsymbol{\beta})$ and $p(0, 0; \boldsymbol{\beta})$.

β_4	PAE	$p(0,1;\boldsymbol{\beta}) = 0.45$ $p(0,0;\boldsymbol{\beta}) = 0.52$			$p(0,1;\boldsymbol{\beta}) = 0.40$ $p(0,0;\boldsymbol{\beta}) = 0.54$			$p(0,1;\boldsymbol{\beta}) = 0.35$ $p(0,0;\boldsymbol{\beta}) = 0.56$		
		$n = 30$	$n = 40$	$n = 50$	$n = 30$	$n = 40$	$n = 50$	$n = 30$	$n = 40$	$n = 50$
0	0.50	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.02	0.02
-2	0.75	0.20	0.37	0.50	0.22	0.39	0.52	0.24	0.40	0.52
-3	1.00	0.54	0.71	0.83	0.52	0.70	0.79	0.49	0.67	0.76

Table 7. Data based on Ellenberger et al. [6] RLC experiment, where Z_i indicates vaccination (1) or not (0), T_i denotes the observed number of challenges, δ_i indicates infection (1) or not (0) by the last challenge, S_i denotes immune response to the vaccine, and W_i is a baseline immunogenicity predictor (BIP).

Control				Vaccine				
Z_i	T_i	δ_i	W_i	Z_i	T_i	δ_i	S_i	W_i
0	1	1	1	1	2	1	1	1
0	1	1	0	1	2	1	0	0
0	2	1	0	1	2	1	1	0
0	2	1	1	1	5	1	0	0
0	2	1	0	1	5	1	0	0
0	2	1	0	1	5	1	1	1
0	2	1	0	1	5	1	0	0
0	2	1	1	1	6	1	0	0
0	3	1	0	1	6	1	0	0
0	4	1	1	1	10	1	0	0
0	8	1	0	1	18	1	0	0
0	10	1	0	1	18	0	0	0
0	11	1	0	1	18	0	1	1
0	14	0	0	1	19	1	0	1
				1	19	0	1	0
				1	26	0	1	1

Table 8. Data based on Ellenberger et al. [6] RLC experiment with hypothetical CrV, where Z_i denotes initial randomization to vaccine (1) or not (0), $\tilde{Z}_i$ indicates ever vaccinated (1) or not (0), T_i^0 denotes number of challenges while not vaccinated, δ_i^0 denotes infected (1) or not (0) after T_i^0 challenges, T_i^1 denotes number of challenges while vaccinated, δ_i^1 denotes infected (1) or not (0) after challenges while vaccinated, and $S_i(1)$ indicates potential immune response (1) or not (0) to vaccine. Missing values are denoted by *.

	Control ($Z_i = 0$)						Vaccine ($Z_i = 1$)				
$\tilde{Z}_i$	T_i^0	δ_i^0	T_i^1	δ_i^1	$S_i(1)$	$\tilde{Z}_i$	T_i^0	δ_i^0	T_i^1	δ_i^1	$S_i(1)$
0	1	1	0	0	*	1	0	0	2	1	1
0	1	1	0	0	*	1	0	0	2	1	0
0	2	1	0	0	*	1	0	0	2	1	1
0	2	1	0	0	*	1	0	0	5	1	0
0	2	1	0	0	*	1	0	0	5	1	0
0	2	1	0	0	*	1	0	0	5	1	1
0	2	1	0	0	*	1	0	0	5	1	0
0	2	1	0	0	*	1	0	0	6	1	0
0	3	1	0	0	*	1	0	0	6	1	0
0	4	1	0	0	*	1	0	0	10	1	0
1	6	0	2	1	0	1	0	0	18	1	0
1	8	0	2	1	0	1	0	0	18	0	0
1	5	0	15	0	0	1	0	0	18	0	1
1	2	0	18	0	1	1	0	0	19	1	0
						1	0	0	19	0	1
						1	0	0	26	0	1