Comparison of Mark-specific Relative Risks with Application to Viral Divergence in Vaccine Efficacy Trials

Peter B. GILBERT, Ian W. MCKEAGUE, and Yanqing SUN

November 12, 2004

Abstract

The efficacy of an HIV vaccine to prevent infection is likely to depend on the genetic variation of the exposing virus. This paper addresses the problem of using data from an HIV vaccine efficacy trial to detect such dependence in terms of the divergence of infecting HIV viruses in trial participants from the HIV strain that is contained in the vaccine. Because hundreds of amino acid sites in each HIV genome are sequenced, it is natural to treat the divergence (defined in terms of Hamming distance say) as a continuous mark variable that accompanies each failure (infection) time. The problem can then be approached by testing whether the ratio of the mark-specific hazard functions for the vaccine and placebo groups is independent of the mark. We develop nonparametric tests for this null hypothesis, using test statistics sensitive to ordered and two-sided alternatives. The test statistics are functionals of a bivariate test process that contrasts Nelson–Aalen-type estimates of cumulative mark-specific hazard functions for the two groups. Asymptotically correct critical values are obtained through a Gaussian multipliers simulation technique. Techniques for estimating mark-specific vaccine efficacy based on the cumulative mark-specific incidence functions are also developed. Numerical studies show good performance of the procedures. The methods are illustrated with application to HIV genetic sequence data collected in the first HIV vaccine efficacy trial.

Some key words: Competing risks; genetic data; nonparametric statistics; survival analysis.

[1]Peter B. Gilbert is Research Associate Professor, Department of Biostatistics, University of Washington and Associate Member, Fred Hutchinson Cancer Research Center, Seattle, WA 98109 (E-mail: pgilbert@scharp.org); Ian W. McKeague is Professor of Biostatistics, Columbia University Mailman School of Public Health, 722 West 168th Street, 6th Floor, New York, NY 10032 (E-mail: im2131@columbia.edu); and Yanqing Sun is Professor of Statistics, University of North Carolina at Charlotte, Charlotte, NC 28223 (E-mail: yasun@uncc.edu).

1 INTRODUCTION

In many longitudinal studies involving the comparison of survival data from two treatment groups, the hazard of an endpoint event is related to a mark variable observed at the endpoint, and it is of interest to determine whether the relative risk between the two groups depends on the mark. In this article, we develop testing and estimation procedures to address this problem. Our approach is based on recent work in which we developed a test for the dependence of a single mark-specific hazard rate on the mark variable (i.e., the "one-sample" problem), see Gilbert, McKeague and Sun (2004).

We are motivated by applications in HIV vaccine efficacy trials. The broad genetic sequence diversity of HIV poses one of the greatest challenges to developing an effective AIDS vaccine (cf., Nabel 2001, Graham 2002). Vaccine efficacy to prevent infection, defined in terms of the hazard ratio between vaccine and placebo recipients, may decrease with the genetic divergence of a challenge HIV from the virus or viruses represented in the vaccine construct (Gilbert, Lele and Vardi, 1999). Detecting such a decrease can help guide the development of new vaccines to provide greater breadth of protection (Gilbert et al., 2001). The relevance of our mark-specific hazard function approach is that the "distance" between a subject's infecting strain and the nearest vaccine strain [defined based on the comparison of the two genetic sequences, as in Gilbert, Lele and Vardi (1999) and Wu, Hsieh and Li (2001)] can be viewed as a mark variable that is only observed in subjects who experience the endpoint event (HIV infection).

VaxGen Inc. conducted the world's first HIV vaccine efficacy trial, in North America and the Netherlands during 1998–2003. At the start of the trial, 5,403 HIV uninfected volunteers at high risk for acquiring HIV were randomized to receive 7 injections of the investigational vaccine AIDSVAX ($n_1 = 3,598$) or of placebo ($n_2 = 1,805$). Subjects were followed for occurrence of the primary study endpoint HIV infection every six months for 3 years. For each subject who became HIV infected during the trial, blood was drawn on the date of infection diagnosis to use for sequencing the envelope glycoprotein (gp120) region of the infecting virus. Of the 368 subjects who acquired HIV, the sequence data were collected for 336 subjects (217 of 241 infected vaccine recipients; 119 of 127 infected placebo recipients). The vaccine contains two genetically engineered HIV gp120 envelope glycoprotein molecules, based on two HIV isolates (named MN and GNE8), and VaxGen hypothesized that the level of vaccine efficacy would be higher against exposing HIVs with gp120 amino acid sequences that were relatively similar to at least one of the HIV strains represented in the vaccine. Three metrics were pre-specified for comparing an infecting virus to the MN and GNE8 strains: the percent mismatch in the aligned amino acid sequences (i.e., Hamming

distance) for three sets of positions. The first set comprises approximately 30 discontinuous amino acids representing the neutralizing face core of gp120 that was crystalized (Wyatt et al., 1998). The second set consists of those positions used for the first distance plus approximately 80 amino acids in the variable loop V2/V3 regions, which are expected to be part of the neutralizing face but could not be crystalized. The third set is the approximately 33 amino acids in the V3 loop, which contains important neutralizing determinants (Wyatt et al., 1998). For each metric and infecting virus, the distance was computed as the minimum of the two distances to the MN and GNE8 sequences.

Gilbert, Lele and Vardi (1999) and Gilbert (2000) developed semiparametric biased sampling models as a tool for studying vaccine efficacy as a function of a continuous mark. However, these methods are limited by the facts that (i) the models condition on infection, so that odds ratios but not relative risks of infection can be estimated; (ii) the relationship between vaccine efficacy and the mark is specified parametrically, with scant data available for suggesting the correct parametric model; and (iii) the models treat HIV infection as a binary outcome, and do not account for the time to HIV infection. The procedures developed here are free from these limitations, as they are prospective, nonparametric, and incorporate the event times.

We introduce nonparametric tests of whether the mark-specific relative risk between the two groups is independent of the mark. The time T_k to endpoint and the mark variable V_k for a representative individual in group k are assumed to be jointly absolutely continuous with density $f_k(t,v)$. We only get to observe $(X_k, \delta_k, \delta_k V_k)$, where $X_k = \min\{T_k, C_k\}$, $\delta_k = I(T_k \leq C_k)$, and C_k is a censoring time assumed to be independent of both T_k and V_k, $k = 1, 2$. When the failure time T_k is observed, $\delta_k = 1$ and the mark V_k is also observed, whereas if T_k is censored, the mark is unknown. We assume that each variable V_k has known and bounded support; rescaling V_k if necessary, this support is taken to be $[0, 1]$. This replicates the one-sample setup of Gilbert, McKeague and Sun (2004). The mark-specific hazard rate in group k is

$$\lambda_k(t,v) = \lim_{h_1,h_2 \to 0} P\{T_k \in [t, t+h_1), V_k \in [v, v+h_2) | T_k \geq t\}/h_1 h_2 \tag{1.1}$$

and the mark-specific cumulative incidence function is

$$F_k(t,v) = \lim_{h_2 \to 0} P\{T_k \leq t, V_k \in [v, v+h_2)\}/h_2, \tag{1.2}$$

$k = 1, 2$, with t ranging over a fixed interval $[0, \tau]$. The functions (1.1) and (1.2) are related by the equation $F_k(t,v) = \int_0^t \lambda_k(s,v) S_k(s)\, ds$, where $S_k(t)$ is the survival function for group k, and are estimable from the observed group k competing risks failure time data. In the case of a discrete mark variable, Gray

(1988) developed a nonparametric test of the null hypothesis of equal cumulative incidence functions across groups, at a specified value of the mark.

The null hypothesis of interest in our case is

$$H_0\text{: } \lambda_1(t,v)/\lambda_2(t,v) \text{ does not depend on } v \text{ for } t \in [0,\tau]$$

which is to be tested against the following alternative hypotheses:

$$\begin{aligned} H_1\text{: } \lambda_1(t,v_1)/\lambda_2(t,v_1) &\leq \lambda_1(t,v_2)/\lambda_2(t,v_2) \text{ for all } v_1 \leq v_2,\ t \in [0,\tau]; \\ H_2\text{: } \lambda_1(t,v_1)/\lambda_2(t,v_1) &\neq \lambda_1(t,v_2)/\lambda_2(t,v_2) \text{ for some } v_1 \neq v_2,\ t \in [0,\tau] \end{aligned}$$

with strict inequality for some $t, v_1 < v_2$ in H_1. To develop suitable test statistics, we will exploit the observation that H_0 holds if and only if the mark-specific relative risk coincides with the ordinary relative risk, i.e., $\lambda_1(t,v)/\lambda_2(t,v) = \lambda_1(t)/\lambda_2(t)$ for all t, v, where $\lambda_k(t) = \int_0^1 f_k(t,v)\,dv/S_k(t) = \int_0^1 \lambda_k(t,v)\,dv$ is the group-k hazard irrespective of the mark.

Testing H_0 versus the monotone alternative H_1 allows us to assess whether the instantaneous relative risk (vaccine/placebo) of HIV infection increases as a function of the divergence v of the exposing virus. A standard measure of vaccine efficacy to prevent infection at time t is the relative reduction in hazard due to vaccination: $\mathrm{VE}(t) = 1 - \lambda_1(t)/\lambda_2(t)$ (Halloran, Struchiner, and Longini, 1997). It is natural to extend this definition to allow the vaccine efficacy to depend on viral divergence: $\mathrm{VE}(t,v) = 1 - \lambda_1(t,v)/\lambda_2(t,v)$. Then, the above hypotheses can be re-expressed as $H_0 : \mathrm{VE}(t,v) = \mathrm{VE}(t)$ for all t, v; $H_1 : \mathrm{VE}(t,v_1) \leq \mathrm{VE}(t,v_2)$ for all $t, v_1 \geq v_2$ (with $<$ for some $v_1 > v_2$); and $H_2 : \mathrm{VE}(t,v_1) \neq \mathrm{VE}(t,v_2)$ for some $t, v_1 \neq v_2$. That is, testing H_0 versus H_1 assesses whether vaccine efficacy decreases with divergence. These tests are biologically meaningful because, under the assumption of an equal distribution of exposure to HIV strains with divergence v for the vaccine and placebo arms at all times up to t (defensible by randomization and double-blinding), $\mathrm{VE}(t,v)$ approximately equals the relative multiplicative reduction in susceptibility to strain v for vaccine versus placebo recipients under a fixed amount of exposure to strain v at time t. To make this approximate interpretation of $\mathrm{VE}(t,v)$ exact requires both the assumption of equal exposure to strain v for the vaccine and placebo arms and that the probability of infection conditional on exposure to strain v is homogeneous among subjects within each study arm (Halloran, Haber and Longini, 1992).

An alternative notion of mark-specific vaccine efficacy is given in terms of cumulative incidences:

$$\mathrm{VE}^c(t,v) = 1 - F_1(t,v)/F_2(t,v),$$

which we call cumulative vaccine efficacy. This represents a time-averaged– rather than instantaneous–measure of vaccine efficacy and is much easier to estimate than $\mathrm{VE}(t, v)$. We also consider the doubly cumulative vaccine efficacy

$$\mathrm{VE}^{dc}(t, v) = 1 - P(T_1 \le t, V_1 \le v)/P(T_2 \le t, V_2 \le v),$$

which can be estimated without smoothing and with greater precision than $\mathrm{VE}^{c}(t, v)$.

In Section 2 we introduce the proposed test procedure and discuss estimation of the cumulative and doubly cumulative vaccine efficacies. Large sample results and a simulation technique needed to implement the test procedure are developed in Section 3. We report the results of a simulation experiment in Section 4, and an application to data from the VaxGen trial is provided in Section 5. Section 6 contains concluding remarks. Proofs of the main results are collected in the Appendix.

2 TEST PROCEDURE

We base our approach on estimates of the doubly cumulative mark-specific hazard functions $\Lambda_k(t, v) = \int_0^v \int_0^t \lambda_k(s, u)\, ds\, du,\ k = 1, 2$. The idea is to compare a nonparametric estimate of $\Lambda_1(t, v) - \Lambda_2(t, v)$ with an estimate under H_0.

Given observation of i.i.d. replicates $(X_{ki}, \delta_{ki}, \delta_{ki}V_{ki}),\ i = 1, \ldots, n_k$, of $(X_k, \delta_k, \delta_k V_k)$, $k = 1, 2$, the nonparametric maximum likelihood estimator of $\Lambda_k(t, v)$ is provided by the Nelson–Aalen-type estimator

$$\hat{\Lambda}_k(t, v) = \int_0^t \frac{N_k(ds, v)}{Y_k(s)},\ t \ge 0,\ v \in [0, 1], \tag{2.1}$$

where $Y_k(t) = \sum_{i=1}^{n_k} I(X_{ki} \ge t)$ is the size of the risk set for group k at time t, and

$$N_k(t, v) = \sum_{i=1}^{n_k} I(X_{ki} \le t, \delta_{ki} = 1, V_{ki} \le v)$$

is the marked counting process with jumps at the uncensored failure times X_{ki} and associated marks V_{ki}, cf. Huang and Louis (1998, (3.2)).

Notice that H_0 holds if and only if $\lambda_1(t, v)/\lambda_2(t, v) = \lambda_1(t)/\lambda_2(t)$ for all t, v, which is equivalent to $\Lambda_1(t, v) = \int_0^t [\lambda_1(s)/\lambda_2(s)]\Lambda_2(ds, v)$ for all t, v. Thus, under H_0 we may estimate the contrast $\Lambda_1(t, v) - \Lambda_2(t, v)$ by $\int_0^t [(\hat{\lambda}_1(s)/\hat{\lambda}_2(s)) - 1]\hat{\Lambda}_2(ds, v)$, where $\hat{\lambda}_k(t)$ is a nonparametric estimator of $\lambda_k(t)$, as discussed below.

We estimate each hazard function $\lambda_k(t)$ by smoothing the increments of the Nelson–Aalen estimator, a technique developed by Rice and Rosenblatt (1976), Yandell (1983), Ramlau-Hansen (1983), and Tanner and Wong (1983). The estimator of $\lambda_k(t)$ is given by

$$\hat{\lambda}_k(t) = \frac{1}{b_k} \int_0^{\tau+\delta} K\left(\frac{t-s}{b_k}\right) d\hat{\Lambda}_k(s)\,,$$

where $\hat{\Lambda}_k(s) = \int_0^t (1/Y_k(s))\, dN_k(s)$ is the ordinary Nelson–Aalen estimator of $\Lambda_k(t) = \int_0^t \lambda_k(s)\, ds$, with $N_k(t) = \sum_{i=1}^{n_k} I(X_{ki} \le t, \delta_{ki} = 1)$. The kernel function K is a bounded symmetric function with support $[-1, 1]$ and integral 1. The bandwidth b_k is a positive parameter that indicates the window $[t - b_k,\ t + b_k]$ over which the Nelson–Aalen estimator is smoothed, and converges to zero as $n_k \to \infty$. We choose kernel esimators because they are nonparametric and they are uniformly consistent under assumptions, a property that is needed for the theoretical justification given later. Specifically, if $[t_1, t_2]$ is an interval satisfying $0 < t_1 < t_2 \le \tau$, λ_k is continuous on $[0, \tau + \delta]$, and

$$\inf_{s\in[0,\tau+\delta]} b_k^2 Y_k(s) \xrightarrow{P} \infty \qquad \text{as} \quad n \to \infty,$$

then $\hat{\lambda}_k$ converges uniformly in probability to λ_k on $[t_1, t_2]$ (see Theorem IV.2.2 in Andersen et al. 1993).

2.1 Test Processes and Test Statistics

Based on the above discussion, we now introduce test processes of the form

$$L_n(t, v) = \sqrt{\frac{n_1 n_2}{n}} \int_a^t H_n(s) \left[\hat{\Lambda}_1(ds, v) - \frac{\hat{\lambda}_1(s)}{\hat{\lambda}_2(s)} \hat{\Lambda}_2(ds, v)\right] \tag{2.2}$$

for $t \ge 0, 0 \le v \le 1$, where $H_n(\cdot)$ is a suitable weight process converging to $H(t)$ and $a \ge 0$. Note that the statistic can be made symmetric by incorporating $\hat{\lambda}_2(\cdot)$ into $H_n(\cdot)$.

Let $y_k(t) = P(X_k \ge t)$ and $\tilde{\tau} = \sup\{t\colon y_1(t) > 0 \ \text{ and } \ y_2(t) > 0\}$ and assume $\tau < \tilde{\tau}$. With kernel smoothing, the bias term of $\hat{\lambda}_k(t)$ is of order $O(b_k^2)$ for the inner points in $[b_k, \tilde{\tau} - b_k]$ and of order $O(b_k)$ for the boundary points in $(0, b_k)$ or $(\tilde{\tau} - b_k, \tilde{\tau})$. To simplify the proofs and the conditions on the rates of convergence concerning b_k, we take $a > 0$ and construct the test statistics from the process $L_n(t, v)$ over $a \le t \le \tau, 0 \le v \le 1$. In practice, however, there would be no harm in taking $a = 0$, or close to zero in order to use as much of the data as possible (this is done in the simulations and application below). The following test statistics are proposed to measure departures from H_0 in the directions of H_1 and H_2:

$$\hat{U}_1 \ = \ \sup_{0\le v\le 1} \sup_{a<t_1<t_2<\tau} (L_n(t_1, v) - L_n(t_2, v)), \tag{2.3}$$

$$\hat{U}_2 \quad = \quad \sup_{0 \le v_1 < v_2 \le 1} \sup_{a < t < \tau} |L_n(t, v_2) - L_n(t, v_1)|. \tag{2.4}$$

Simple calculation shows that $(n/n_1 n_2)^{1/2}(L_n(t_1, v) - L_n(t_2, v))$ converges to

$$\Gamma(v, t_1, t_2) = \int_0^v \int_{t_1}^{t_2} H(s)\lambda_2(s, u)[r(s) - \lambda_1(s, u)/\lambda_2(s, u)]\, ds\, du,$$

where $r(s) = \lambda_1(s)/\lambda_2(s)$. Under H_1 and suitable smoothness conditions (for the details, see the proof of Theorem 2 in the Appendix), it can be shown that $r(s) > \lambda_1(s, u)/\lambda_2(s, u)$ for all s in some interval $[t_1, t_2]$ and all u in some interval $[0, v]$, implying that $\Gamma(v, t_1, t_2) > 0$. However, under H_0, we have $\Gamma(v, t_1, t_2) = 0$ for all $v,\ t_1,\ t_2$. Thus a large value of $\hat{U}_1$ provides evidence against H_0 in the direction of H_1. A similar argument motivates $\hat{U}_2$ as a natural test statistic for detecting the two-sided alternative H_2. Several other functionals of $L_n(t, v)$ were evaluated as test statistics, and $\hat{U}_1$ and $\hat{U}_2$ were found to have the greatest power.

In the next section, we show that $L_n(t, v)$ converges weakly to a Gaussian process under H_0. We also show that the proposed test based on $\hat{U}_j$ is consistent against its alternative, $j = 1, 2$. Since each $\hat{U}_j$ above is a continuous functional of $L_n(t, v)$, its limiting null distribution is the distribution of the corresponding functional of the limiting Gaussian process. These distributions are intractable, however, so we determine the critical values of the $\hat{U}_j$ with a simple simulation procedure, described in Section 3.2

2.2 Estimation of Cumulative Vaccine Efficacy

Confidence intervals for $\mathrm{VE}(t, v)$ are too wide to discern patterns in mark-specific efficacy within efficacy trials, see Gilbert et al. (2002). On the other hand, the cumulative vaccine efficacy $\mathrm{VE}^c(t, v) = 1 - F_1(t, v)/F_2(t, v)$ can be estimated relatively easily because smoothing is not required over t, only over v. A nonparametric estimator is given by $\widehat{\mathrm{VE}}^c(t, v) = 1 - \hat{F}_1(t, v)/\hat{F}_2(t, v)$, where

$$\hat{F}_k(t, v) = \frac{1}{b_{vk}} \int_0^1 \int_0^t \frac{\hat{S}_k(s-)}{Y_k(s)} K\left(\frac{v - u}{b_{vk}}\right) N_k(ds, du), \tag{2.5}$$

$\hat{S}_k(t)$ is the Kaplan–Meier estimate of $S_k(t)$, $K(\cdot)$ is a bounded symmetric kernel function with support $[-1, 1]$ and integral 1, and $b_{vk} > 0$ is a bandwidth. The estimator $\hat{F}_k(t, v)$ is the continuous analog of the estimator that has been used for a discrete mark (Fine and Gray, 1999; McKeague, Gilbert and Kanki, 2001).

If $F_1(t,v) \neq 0$ and $F_2(t,v) \neq 0$, a $100(1-\alpha)\%$ pointwise confidence interval for $\mathrm{VE}^c(t,v)$ can be computed by transforming symmetric confidence limits about $\log(\hat{F}_1(t,v)/\hat{F}_2(t,v))$:

$$1 - \left(1 - \widehat{\mathrm{VE}}^c(t,v)\right) \exp\left(\pm z_{\alpha/2} \sqrt{\frac{\widehat{\mathrm{Var}}\{\hat{F}_1(t,v)\}}{\hat{F}_1(t,v)^2} + \frac{\widehat{\mathrm{Var}}\{\hat{F}_2(t,v)\}}{\hat{F}_2(t,v)^2}}\right), \tag{2.6}$$

where

$$\widehat{\mathrm{Var}}\{\hat{F}_k(t,v)\} = \frac{1}{b_{vk}^2} \int_0^1 \int_0^t \left[\frac{\hat{S}_k(s-)}{Y_k(s)} K\left(\frac{v-u}{b_{vk}}\right)\right]^2 N_k(ds,du).$$

To estimate the doubly cumulative vaccine efficacy $\mathrm{VE}^{dc}(t,v)$, each $P(T_k \leq t, V_k \leq v)$ is simply estimated by $\hat{F}_{k1}(t) = \int_0^t \left(\hat{S}_k(s-)/Y_k(s)\right) N_k(ds,v)$, the estimator for the cumulative incidence function with the discrete cause of failure 1 defined by $V \leq v$, and its variance is estimated by $\int_0^t \left(\hat{S}_k(s-)/Y_k(s)\right)^2 N_k(ds,v)$. Similarly as for $\mathrm{VE}^c(t,v)$, a confidence interval for $\mathrm{VE}^{dc}(t,v)$ can be constructed by transforming symmetric confidence limits about $\log(P(T_1 \leq t, V_1 \leq v)/P(T_2 \leq t, V_2 \leq v))$, where the estimated variance of the log ratio is obtained via the delta method.

3 LARGE-SAMPLE RESULTS

We begin by defining notation that is used in the sequel. Let $\gamma_k(t,v) = P(X_k \leq t, \delta_k = 1, V_k \leq v), k = 1,2$. By the Glivenko–Cantelli Theorem, $N_k(t,v)/n_k$ and $Y_k(t)/n_k$ converge almost surely to $\gamma_k(t,v)$ and $y_k(t)$, uniformly in $(t,v) \in [0,\infty) \times [0,1]$ and $t \in [0,\infty)$, respectively. Note that we may write $\lambda_k(t,v) = f_k(t,v)/S_{T_k}(t)$, where $S_{T_k}(t) = P(T_k \geq t)$ and $f_k(t,v)$ is the joint density of (T_k, V_k) for group k. Also, $\lambda_k(t) = f_{T_k}(t)/S_{T_k}(t)$, where $f_{T_k}(t)$ is the density of T_k for group k. Let $D(I)$ be the set of all uniformly bounded, real-valued functions on a K-dimensional rectangle I, endowed with the uniform metric. Let $C(I)$ be the subspace of uniformly bounded, continuous functions on I.

3.1 Asymptotic Distributions of the Test Statistics

Let $Z_1(t,v)$ and $Z_2(t,v)$ be two independent Gaussian processes defined by

$$Z_k(t,v) = \int_0^t \frac{1}{y_k(s)} G_1^{(k)}(ds,v) - \int_0^t \frac{G_2^{(k)}(s)}{y_k(s)^2} \gamma_k(ds,v), \quad k = 1,2, \tag{3.1}$$

where $G_1^{(k)}(t,v)$ and $G_2^{(k)}(t)$ are continuous mean zero Gaussian processes with covariances

$$\mathrm{Cov}(G_1^{(k)}(s,u), G_1^{(k)}(t,v)) = \gamma_k(s \wedge t, u \wedge v) - \gamma_k(s,u)\gamma_k(t,v),$$

$$\mathrm{Cov}(G_2^{(k)}(s), G_2^{(k)}(t)) = y_k(s \vee t) - y_k(s)y_k(t),$$
$$\mathrm{Cov}(G_1^{(k)}(s,u), G_2^{(k)}(t)) = (\gamma_k(s,u) - \gamma_k(t-,u))I(t \le s) - \gamma_k(s,u)y_k(t).$$

Let $r(t) = \lambda_1(t)/\lambda_2(t)$, $a(t) = 1/\lambda_2(t)$ and $0 < \kappa = \lim_{n\to\infty} n_1/n < 1$. Define

$$\begin{aligned} L(t,v) &= \sqrt{1-\kappa}\left[\int_a^t H(s)Z_1(ds,v) - \int_a^t H(s)a(s)\Lambda'_{2s}(s,v)\, Z_1(ds,1)\right] \\ &\quad -\sqrt{\kappa}\left[\int_a^t H(s)r(s)Z_2(ds,v) - \int_a^t H(s)r(s)a(s)\Lambda'_{2s}(s,v)\, dZ_2(ds,1)\right], \end{aligned} \tag{3.2}$$

where $\Lambda'_{2t}(t,v) = \partial\Lambda_2(t,v)/\partial t$.

Our first result describes the limiting null distribution of the test process and the test statistics.

Theorem 1. *Let the weight process $H_n(t)$ be a continuous functional of the processes $N_k(t,1)$ and $Y_k(t)$, $k = 1, 2$, $t \in [0, \tau+\delta]$, $\tau+\delta < \tilde{\tau}$ for some $\delta > 0$. Assume there exists a uniformly continuous function $H(t)$ such that $\sup_{0\le t\le\tau+\delta} |H_n(t) - H(t)| \xrightarrow{\text{a.s.}} 0$ and both H_n and H have bounded variation independent of n almost surely. Assume $\lambda_k(t)$ is twice continuously differentiable over $[0, \tau+\delta]$, $k = 1, 2$, $\lambda_2(t)$ is bounded away from zero on $[a/2, \tau+\delta]$, $\lambda_2(t,v) > 0$ and $\partial^2\Lambda_2(t,v)/\partial t^2$ is continuous on $[0,\tau+\delta]\times[0,1]$. Also assume the kernel function $K(\cdot)$ has bounded variation. Suppose $nb_k^2 \to \infty$ and $nb_k^6 \to 0$ for $k = 1, 2$. Then, under H_0*

$$L_n(t,v) \xrightarrow{\mathcal{D}} L(t,v) \tag{3.3}$$

in $D([a,\tau]\times[0,1])$ as $n \to \infty$.

The proof of Theorem 1 immediately follows from Proposition 1 given in the Appendix. The conditions on the rates of convergence are satisfied if, for example, $b_k = n_k^{-\alpha}$ for $1/6 < \alpha < 1/2$.

Let U_1 and U_2 be defined the same as $\hat{U}_1$ and $\hat{U}_2$ in (2.3) and (2.4), respectively, with $L_n(\cdot)$ replaced with $L(\cdot)$. By the continuous mapping theorem, $\hat{U}_j \xrightarrow{\mathcal{D}} U_j$ under H_0, so $P(\hat{U}_j > c_{j\alpha}) \to \alpha$, where $c_{j\alpha}$ is the upper α-quantile of U_j. However, the $c_{j\alpha}$ are unknown and very difficult to estimate due to the complicated nature of the limit process $L(t,v)$. In the next section we provide a Monte Carlo procedure to obtain each $c_{j\alpha}$. Before proceeding, we show that the test statistics $\hat{U}_j$ are consistent against their respective alternatives.

Theorem 2. *In addition to the conditions given in Theorem 1, assume that $\lambda_1(t,v)$ and $\lambda_2(t,v)$ are continuous and that $H(t,v) > 0$ on $[0,\tau]\times[0,1]$. Then, $P(\hat{U}_1 > c_{1\alpha}) \to 1$ as $n \to \infty$ under H_1 and $P(\hat{U}_2 > c_{2\alpha}) \to 1$ as $n \to \infty$ under H_2.*

3.2 Gaussian Multipliers Simulation Procedure

We now describe a Gaussian multipliers technique for simulating the test process $L_n(t,v)$ under the null hypothesis, cf. Lin, Wei and Ying (1993) and Lin, Wei and Fleming (1994). By (7.2) in the Appendix and the continuous mapping theorem, we obtain

$$\int_0^t \frac{1}{y_k(s)}\sqrt{n}_k(N_k(ds,v)/n_k - \gamma_k(ds,v)) - \int_0^t \frac{1}{y_k(s)^2}\sqrt{n}_k(Y_k(s)/n_k - y_k(s))\,\gamma_k(ds,v)$$
$$\xrightarrow{\mathcal{D}} Z_k(t,v). \tag{3.4}$$

Define the process $\tilde{L}(t,v)$ by replacing $Z_k(t,v)$, $k=1,2$, in $L(t,v)$ given in (3.2) with the term on the left side of (3.4) and replacing κ with n_1/n. Applying the continuous mapping theorem again, we have $\tilde{L}(t,v)\xrightarrow{\mathcal{D}} L(t,v)$. Let $N_{ki}(t,v) = I(X_{ki}\le t, \delta_{ki}=1, V_{ki}\le v)$ and $Y_{ki}(t) = I(X_{ki}\ge t)$, $k=1,2$. It follows that

$$\tilde{L}(t,v) = \sqrt{n_2/n}{n_1}^{-1/2}\sum_{i=1}^{n_1} h_{1i}(t,v) - \sqrt{n_1/n}{n_2}^{-1/2}\sum_{i=1}^{n_2} h_{2i}(t,v), \tag{3.5}$$

where

$$\begin{aligned} h_{1i}(t,v) &= \int_a^t H(s)y_1^{-1}(s)\,(N_{1i}(ds,v) - \gamma_1(ds,v)) \\ &\quad - \int_a^t H(s)y_1^{-2}(s)(Y_{1i}(s) - y_1(s))\,\gamma_1(ds,v) \\ &\quad - \int_a^t H(s)y_1^{-1}(s)a(s)\Lambda'_{2s}(s,v)\,(N_{1i}(ds,1) - \gamma_1(ds,1)) \\ &\quad + \int_a^t H(s)y_1^{-2}(s)a(s)\Lambda'_{2s}(s,v)(Y_{1i}(s) - y_1(s))\gamma_1(ds,v) \end{aligned}$$

and

$$\begin{aligned} h_{2i}(t,v) &= \int_a^t H(s)y_2^{-1}(s)r(s)\,(N_{2i}(ds,v) - \gamma_2(ds,v)) \\ &\quad - \int_a^t H(s)y_2^{-2}(s)r(s)(Y_{2i}(s) - y_2(s))\,\gamma_2(ds,v) \\ &\quad - \int_a^t H(s)y_2^{-1}(s)b(s)\Lambda'_{2s}(s,v)\,(N_{2i}(ds,1) - \gamma_2(ds,1)) \\ &\quad + \int_a^t H(s)y_2^{-2}(s)b(s)\Lambda'_{2s}(s,v)(Y_{2i}(s) - y_2(s))\gamma_2(ds,v), \end{aligned}$$

with $a(s) = 1/\lambda_2(s)$, $b(s) = \lambda_1(s)/(\lambda_2(s))^2$, and $\Lambda'_{2s}(s,v) = \partial\Lambda_2(s,v)/\partial s$.

Define $\hat{h}_{ki}(t,v)$ by replacing, in $h_{ki}(t,v)$, $H(s)$ with $H_n(s)$, $y_k(s)$ with $Y_k(s)/n_k$, $\gamma_k(s,v)$ with $N_k(s,v)/n_k$, $a(s)$ with $\hat{a}(s)$, and $\Lambda'_{2s}(s,v)$ with a suitable smooth uniformly consistent estimate $\hat{\Lambda}'_{2s}(s,v)$

on $[a, \tau] \times [0, 1]$. Let $W_{ki}, i = 1, \ldots, n_k, k = 1, 2$ be i.i.d. standard normal random variables. Let

$$L_n^*(t, v) = \sqrt{\frac{n_2}{n}} {n_1}^{-1/2} \sum_{i=1}^{n_1} \hat{h}_{1i}(t, v) W_{1i} - \sqrt{\frac{n_1}{n}} {n_2}^{-1/2} \sum_{i=1}^{n_2} \hat{h}_{2i}(t, v) W_{2i}. \tag{3.6}$$

We show that the conditional weak limit of the process $L_n^*(t, v)$ given the observed data is the same as the weak limit of $L_n(t, v)$ under the null hypothesis H_0. Note that the two terms in (3.2) and (3.6) are independent. It is easy to show that for any two points (t, v) and (s, w) in $[a, \tau] \times [0, 1]$,

$$n_k^{-1} \sum_{i=1}^{n_k} \hat{h}_{1i}(t, v) \hat{h}_{1i}(s, w) \xrightarrow{P} E[h_{1i}(t, v) h_{1i}(s, w)],$$

since $\hat{h}_{ki}(t, v) \xrightarrow{P} h_{ki}(t, v)$ as $n \to \infty$. Thus, the conditional covariance of $L_n^*(t, v)$ converges to the covariance of $L(t, v)$. It is left to show that the processes $L_n^*(t, v)$ is tight (see Appendix). Therefore, under H_0 the conditional limit process of $L_n^*(t, v)$ given the observed data sequence equals the limit process $L(t, v)$ in distribution.

Theorem 3. *Under the conditions of Theorem 1, conditional on the observed competing risks data sequence,*

$$L_n^*(t, v) \xrightarrow{\mathcal{D}} L(t, v) \tag{3.7}$$

in $D([a, \tau] \times [0, 1])$ under H_0 as $n \to \infty$, where $L(t, v)$ is given in (3.2).

3.3 Choice of Weight Process and a Graphical Procedure

In exploratory work it can be useful to examine a plot of the test process $L_n(t, v)$ with the weight process chosen to be $H_n(t) = 1$, and compare it with plots of (say) 5–20 realizations of the simulated reference process $L_n^*(t, v)$. Large values of the contrast $L_n(t_1, v) - L_n(t_2, v)$ for some v and some $t_1 < t_2$, as compared with the same contrast in $L_n^*(t, v)$, then suggest a departure from H_0 in the direction of H_1. Large absolute differences in $L_n(t, v)$ across different marks v (as compared with the reference process) would suggest H_2. This graphical procedure is illustrated in Section 5.

The test process is more variable at larger failure times, so it is advisable to choose the weight process to downweight the upper tail of the integral. In addition, it is desirable to have a symmetric test process, so we suggest the following choice of weight process:

$$H_n(s) = \hat{\lambda}_2(s) \sqrt{\frac{Y_1(s)}{n_1} \frac{Y_2(s)}{n_2}}. \tag{3.8}$$

The weight process can also be chosen to increase the power of the tests for some specific alternatives, cf. Sun (2001).

If it is of interest to test the hypothesis that the mark-specific hazard ratio is independent of the mark over a subinterval $[v_1, v_2]$, then the testing procedure can be applied with $H_n(s, v)$ made to depend on v and set equal to zero outside of $[v_1, v_2]$.

4 SIMULATION EXPERIMENT

The simulations are based on the features of the VaxGen trial described in the Introduction, in which vaccine and placebo recipients were monitored for infection during a $\tau = 36$ month period after enrollment. We study performance of the test statistics $\hat{U}_1$ and $\hat{U}_2$, and of the cumulative vaccine efficacy estimator $\widehat{\mathrm{VE}}^c(\tau, v)$. The latter is only considered at the end of follow-up $t = \tau$, because it is most important scientifically to understand durability of vaccine efficacy, and precision is maximized at τ.

To set up the simulation experiment, first consider the case with T_k and V_k independent, $k = 1,\ 2$. The cumulative incidence function for group k is then $F_k(t, v) = P\{T_k \leq t\} f_{Vk}(v)$, where f_{Vk} is the density of V_k. We specify T_1 and T_2 to be exponential with parameters $\theta\lambda_2$ and λ_2, respectively, so that the cumulative vaccine efficacy by time τ irrespective of the mark V is given by $\mathrm{VE}^c(\tau) = 1 - (1 - \exp(-\lambda_2\theta\tau))/(1-\exp(-\lambda_2\tau))$, where λ_2 is the constant infection hazard rate in the placebo group. Here θ is the constant infection hazard ratio between groups 1 and 2, which could itself be used to measure overall vaccine efficacy. We consider two true values of $\mathrm{VE}^c(\tau)$, 0.67 and 0.33, corresponding to a moderately and weakly efficacious vaccine, respectively. In addition, we select λ_2 so that 50% of placebo recipients are expected to be infected by $\tau = 36$ months.

Next, we specify

$$f_{Vk}(v) = \frac{1}{\beta_k \left(1.5^{1/\beta_k} - 0.5^{1/\beta_k}\right)} (v + 0.5)^{(1/\beta_k)-1} \qquad \text{for } \ 0 \leq v \leq 1. \tag{4.1}$$

Here $\beta_k = 1$ corresponds to $\lambda_k(t, v)$ not depending on v, with $E(V_k) = 1/2$, and $\beta_k = 0.5,\ 0.25$ correspond to two different levels of dependence between $\lambda_k(t, v)$ and v, with $E(V_k) =$ 2/3 and 4/5, respectively. The degree of dependence of $\lambda_k(t, v)$ on v increases as β_k decreases, and the cumulative vaccine efficacy is given by

$$\mathrm{VE}^c(\tau, v) = 1 - (1 - \mathrm{VE}^c(\tau)) \frac{\beta_2}{\beta_1} \left[\frac{1.5^{1/\beta_2} - 0.5^{1/\beta_2}}{1.5^{1/\beta_1} - 0.5^{1/\beta_1}}\right] (v + 0.5)^{(1/\beta_1)-(1/\beta_2)};$$

this curve and the curve $\mathrm{VE}(\tau, v)$ are depicted in panels (a) and (b) of Figure 1. Note that $\mathrm{VE}(\tau, v) = \mathrm{VE}(\tau)$ and $\mathrm{VE}^c(\tau, v) = \mathrm{VE}^c(\tau)$ if and only if $\beta_1 = \beta_2$, so that setting $\beta_2/\beta_1 = 1.0$ represents the null hypothesis. Furthermore $\beta_2/\beta_1 > 1$ implies $\mathrm{VE}(\tau, v)$ and $\mathrm{VE}^c(\tau, v)$ decrease with v, so the extent of departure from H_0 increases with β_2/β_1. We set the true (β_1, β_2) to be (1.0,1.0), (0.50,1.0), or (0.25,1.0).

We also consider a two-sided alternative with $f_{V2}(v)$ a uniform density and $f_{V1}(v) = \frac{16}{3} v I(v < \frac{1}{2}) + (\frac{8}{3} - \frac{8}{3} v) I(v \geq \frac{1}{2})$. This alternative specifies $\mathrm{VE}(\tau, v)$ and $\mathrm{VE}^c(\tau, v)$ as step functions ((c) and (d) of Figure 1). Results for the two-sided case are given under the heading "2-sided" in Tables 1 and 2.

Next, we consider a case with T_k and V dependent for both groups. For the monotone alternative H_1, we use $F_k(t, v) = P\{T_k \leq t | V_k = v\} f_{Vk}(v) = (1 - \exp(-\theta^{I(k=1)} \lambda t/(v+1))) f_{Vk}(v)$, with $f_{Vk}(v) = (1/\beta_k) v^{\frac{1}{\beta_k} - 1}$. As in the independent case, $\beta_2/\beta_1 = 1.0$ represents the null hypothesis and $\beta_2/\beta_1 > 1.0$ represents the alternative hypotheses with $\mathrm{VE}(t, v)$ and $\mathrm{VE}^c(t, v)$ decreasing with v ((e) and (f) of Figure 1). The true parameter pairs (β_1, β_2) are the same as in the independent case. For a two-sided alternative, we use $F_k(t, v) = (1 - \exp(-\theta^{I(k=1)} \lambda t/(v+1))) f_{Vk}(v)$, with f_{V1} and f_{V2} as in the independent case (see (g) and (h) of Figure 1). For both the 1-sided and 2-sided dependent cases, we select λ such that conditional on $v = 0.5$, 50% of placebo recipients are expected to fail by 36 months, and θ such that $\mathrm{VE}^c(\tau, v = 0.5) = 0.67$ or 0.33.

The weight process $H_n(\cdot)$ of (3.8) is used for the test statistics. For kernel estimation of $\lambda_k(t), k = 1, 2$, the Epanechnikov kernel $K(x) = 0.75(1 - x^2) I(|x| \leq 1)$ is used. For each simulation iteration the optimal bandwidth b_k is chosen to minimize an asymptotic approximation to the mean integrated squared error (Andersen et al., 1993, p. 240), and the method of Gasser and Müller (1979) is used to correct for bias in the tails. An alternative approach to optimizing the bandwidths separately for each hazard function would jointly optimize the bandwidths for estimating the hazard ratio; this issue was investigated by Kelsall and Diggle (1995). Based on their results, joint optimization does not provide appreciable efficiency gains unless the hazards in the two groups are fairly similar. For the HIV vaccine application, it is most interesting to assess the relationship of vaccine efficacy on viral divergence when there is substantial efficacy (i.e., the hazards are unequal), because (tautologically) some degree of protection is necessary for there to be differential protection. For this reason we optimized the hazard functions separately.

The nominal level of the tests is set at 0.05, and critical values are calculated using 500 replicates of the Gaussian multipliers technique described in Section 3.2. For estimation the mark-bandwidths b_{0k} are set at 0.20. Bias, coverage probability of the 95% confidence intervals (2.6), and variance estimation of

$\widehat{\mathrm{VE}}^c(\tau, v)$ are evaluated at the three mark-values $v = 0.30, 0.50, 0.80$. We choose $n = 100$ or 200 and in addition to the 50% administrative censoring for the failure times at 36 months we use a 10% random censoring rate in each arm. The performance statistics are calculated based on 1000 simulated datasets.

The results in Tables 1 and 2 indicate that the tests perform well at moderate sample sizes, although for $\mathrm{VE}^c(\tau) = 0.67$ the procedures are conservative. For HIV vaccine efficacy trials of realistic size ($\sim$190 infections in the placebo arm), the tests have high power to detect the alternative hypotheses considered.

The results in Table 3 show that the bias in $\widehat{\mathrm{VE}}^c(36, v)$ becomes negligible as the number of infections grows large. For small or moderate samples (45 or 90 infections in the placebo arm), the estimator is approximately unbiased under the null $\beta_1 = 0$, is slightly biased when $\beta_1 = 0.5$, is moderately biased when $\beta_1 = 0.25$ and $v = 0.5$, and is largely biased when $\beta_1 = 0.25$ and $v = 0.3$. The large negative bias occurs because for small v, $\mathrm{VE}^c(36, v)$ is near the upper boundary 1.0. The confidence intervals for $\mathrm{VE}^c(36, v)$ have correct coverage probability in large samples and usually perform well at smaller sample sizes, but have too-small coverage probability for the same values of $\mathrm{VE}^c(36, v)$ at which the estimator is substantially negatively biased. The asymptotic variance estimates of $\widehat{\mathrm{VE}}^c(36, v)$ tracked the Monte Carlo variance estimates fairly closely, verifying acceptable accuracy of the variance estimators (not shown).

The simulation study was programmed in Fortran, with pseudorandom-numbers generated with internal Fortran functions. This program and a data analysis program are available upon request.

5 APPLICATION

We apply the methods to the data from the VaxGen trial described in the Introduction. Figure 2 shows boxplots of the three percent amino acid mismatch distances of the infecting HIV viruses to the nearest gp120 sequence (MN or GNE8) in the tested vaccine. The neutralizing face core distances ranged from 0.032 to 0.22 with medians 0.11 and 0.085 in the vaccine and placebo groups, the neutralizing face core plus V2/V3 distances ranged from 0.071 to 0.32 with medians 0.17 in each group, and the V3 loop distances ranged from 0.036 to 0.46 with medians 0.14 and 0.18 in the vaccine and placebo groups, respectively.

The testing procedures were implemented using the same weight function $H_n(\cdot)$, kernel $K(\cdot)$, and procedures for optimal bandwidth selection and tail correction that were used in the simulation experiment. P-values were approximated using 10,000 simulations. The MISE-optimized bandwidths for the estimated hazards of infection $\hat{\lambda}_1(\cdot)$ and $\hat{\lambda}_2(\cdot)$ were 1.83 months and 2.10 months, respectively.

The tests based on $\hat{U}_1$ and on $\hat{U}_2$ gave nonsignificant results for all three distances ($p > 0.05$). In order of the distances presented in Figure 2, $\hat{U}_1$ equaled 0.408 ($p = 0.058$), 0.287 ($p = 0.39$), and 0.214 ($p = 0.68$), respectively, and $\hat{U}_2$ equaled 0.393 ($p = 0.40$), 0.340 ($p = 0.72$), and 0.432 ($p = 0.34$). To illustrate the graphical procedure, for the neutralizing face core distances Figure 3 shows the test process $L_n(t, v)$ together with 8 randomly selected realization of the null test process $L_n^*(t, v)$, using a unit weight function $H_n(\cdot) = 1$. The maximum absolute deviation of $L_n(t, v)$ in t is larger than that for all but one of the null test processes, which is consistent with the fairly small p-value from$\hat{U}_2$ of 0.058.

With bandwidths $b_{v1} = b_{v2}$ set to be one-quarter of the observed range of V for each HIV metric, we next estimated $\mathrm{VE}^c(36, v)$ and $\mathrm{VE}^{dc}(36, v)$ with 95% pointwise confidence intervals (Figure 4). The overall vaccine efficacy estimate $\widehat{\mathrm{VE}}^c(36) = 0.048$ is included for reference. The $\mathrm{VE}^c(36, v)$ curves are estimated with reasonable precision at mark values v not in the tail regions, and $\mathrm{VE}^{dc}(36, v)$ is estimated with reasonable precision for v not in the left tail, with precision increasing with v. For neutralizing face core distances the estimates of $\mathrm{VE}^c(36, v)$ and $\mathrm{VE}^{dc}(36, v)$ in the regions of precision diminished with viral distance, which suggests that the closeness of match of amino acids in the exposing strain versus vaccine strain in the core amino acids may have impacted the ability of the vaccine to stimulate protective antibodies that neutralized the exposing strain. However, because the confidence intervals include both 0 and $\widehat{\mathrm{VE}}^c(36)$ at all marks v, the evidence for decreasing efficacy with viral distance is not significant. This result is consistent with the outome of the testing procedures.

For the neutralizing face core + V2/V3 distances, the estimated vaccine efficacy curves are horizontal in the region of precision, supporting no differential efficacy. In contrast, for the V3 loop distances vaccine efficacy appears to increase with viral distance (Figure 4(e)). However, the confidence intervals are wide for large values of v, and a result of increasing $\mathrm{VE}^c(36, v)$ with v is opposite to the biologically plausible scenario of decreasing $\mathrm{VE}^c(36, v)$ with v.

In conclusion, the testing and estimation procedures do not support that vaccine efficacy varied significantly with any of the three HIV distances studied. This result is expected from the fact that the overall estimate of vaccine efficacy was near zero. It is intriguing that a trend towards decreasing efficacy with larger distances from the vaccine antigens was found for the neutralizing face core distance, as this distance has the soundest biological rationale; three-dimensional structural analysis has demonstrated that the amino acid positions used for this distance constitute conserved neutralizing antibody epitopes (Wyatt et al., 1998).

6 CONCLUDING REMARKS

The problem addressed here, evaluating the relationship between the relative risk of failure and a continuous mark variable observed only at uncensored failure times, is important and has broad application. For HIV vaccine trials, the methods can be used for confirmatory assessments of specific viral metrics hypothesized to be associated with vaccine efficacy, and for exploratory assessments, in which the tests are carried out for many metrics (e.g., based on different sets of sites in the HIV genome and incoporating different weight functions reflecting the relative immunological significance of different amino acid substitutions) to generate hypotheses about what attributes of HIV divergence are most immunologically relevant. Both the confirmatory and exploratory analyses provide critical input into the process of immunogen design to iteratively improve a candidate vaccine's breadth of protective efficacy. The testing procedures can also be used for power calculations in the design of HIV vaccine trials. The test based on $\hat{U}_1$ is preferred for the monotone alternative H_1 and the test based on $\hat{U}_2$ is preferred for the two-sided alternative H_2.

The situation in which a failure time is measured in two groups and the mark characterizes the causal agent, encountered in HIV vaccine trials, occurs in many other disease applications. For example, in an anti-HIV therapeutic trial, subjects randomized to various treatments are followed until treatment failure, and the genetic sequence or phenotypic susceptibility of the HIV is measured at baseline and at the failure time (Gilbert et al., 2000). For each failed subject, a viral distance is calculated between the two time points; this distance is designed to measure the evolution of the virus towards a drug-resistant form. Evaluating the dependency of the relative risk of failure on this accumulated resistance distance assesses whether the metric is more associated with clinical resistance for one treatment than the other. In other settings it is of interest to compare treatment groups by the relationship between the risk of death and a quality-of-life score or a lifetime medical cost. An appeal of the procedures developed here for addressing such problems is that they are based on a mark-specific version of the widely-applied and well-understood Nelson–Aalen-type nonparametric maximum likelihood estimator, and naturally extend the scope of methods that have been developed for competing risks data with discrete (cause-of-failure) marks.

ACKNOWLEDGMENT

The authors gratefully acknowledge David Jobes and VaxGen Inc. for providing the HIV sequence data. This research was partially supported by NIH grant 1 RO1 AI054165-01 (Gilbert), NSF grant DMS-0204688 (McKeague), and NSF grant DMS-0304922 (Sun).

7 APPENDIX: PROOFS OF THEOREMS

Proposition 1. *Given the conditions expressed in Theorem 1,*

$$L_n(t,v) - \sqrt{\frac{n_1 n_2}{n}} \int_a^t H_n(s)[\Lambda_1(ds,v) - r(s)\Lambda_2(ds,v)] \xrightarrow{\mathcal{D}} L(t,v) \tag{7.1}$$

in $D([a,\tau]\times[0,1])$.

Proof of Proposition 1.

Using the central limit theorem for empirical processes (cf. Gilbert, McKeague and Sun, 2004, (A.4)),

$$\sqrt{n}_k(N_k(t,v)/n_k - \gamma_k(t,v), Y_k(t)/n_k - y_k(t)) \xrightarrow{\mathcal{D}} (G_1^{(k)}(t,v), G_2^{(k)}(t)) \tag{7.2}$$

in $D([0,\tau]\times[0,1])\times D[0,\tau]$, where $G_1^{(k)}(t,v)$ and $G_2^{(k)}(t)$ are continuous mean zero Gaussian processes with covariances

$$\begin{aligned}
&\mathrm{Cov}(G_1^{(k)}(s,u), G_1^{(k)}(t,v)) = \gamma_k(s\wedge t, u\wedge v) - \gamma_k(s,u)\gamma_k(t,v),\\
&\mathrm{Cov}(G_2^{(k)}(s), G_2^{(k)}(t)) = y_k(s\vee t) - y_k(s)y_k(t),\\
&\mathrm{Cov}(G_1^{(k)}(s,u), G_2^{(k)}(t)) = (\gamma_k(s,u) - \gamma_k(t-,u))I(t\le s) - \gamma_k(s,u)y_k(t).
\end{aligned}$$

Let $\hat{Z}_k(t,v) = \sqrt{n_k}(\hat{\Lambda}_k(t,v) - \Lambda_k(t,v))$. By the functional delta method as used in (A.7)–(A.8) of Gilbert et al. (2001), we have

$$\hat{Z}_k(t,v) \xrightarrow{\mathcal{D}} Z_k(t,v) \tag{7.3}$$

in $D([0,\tau]\times[0,1])$, where the two processes $Z_1(t,v)$ and $Z_2(t,v)$ are independent. Applying the almost sure representation theorem (Shorack and Wellner, 1986, p. 47) as in the proof of Proposition 2 of Gilbert, McKeague and Sun (2004), we may treat the weak convergence in (7.3) as almost sure convergence uniformly on $[0,\tau]\times[0,1]$.

Let $r(t) = \lambda_1(t)/\lambda_2(t)$ and $\hat{r}(t) = \hat{\lambda}_1(t)/\hat{\lambda}_2(t)$. The test process can be decomposed as follows:

$$\begin{aligned}
&L_n(t,v) = \sqrt{\frac{n_1 n_2}{n}} \int_a^t H_n(s)[\hat{\Lambda}_1(ds,v) - \Lambda_1(ds,v)]\\
&-\sqrt{\frac{n_1 n_2}{n}} \int_a^t H_n(s)\hat{r}(s)[\hat{\Lambda}_2(ds,v) - \Lambda_2(ds,v)] + \sqrt{\frac{n_1 n_2}{n}} \int_a^t H_n(s)[\Lambda_1(ds,v) - \hat{r}(s)\Lambda_2(ds,v)]\\
&= \sqrt{\frac{n_2}{n}} \int_a^t H_n(s)\hat{Z}_1(ds,v) - \sqrt{\frac{n_1}{n}} \int_a^t H_n(s)\hat{r}(s)\hat{Z}_2(ds,v)\\
&+\sqrt{\frac{n_1 n_2}{n}} \int_a^t H_n(s)[r(s) - \hat{r}(s)]\Lambda_2(ds,v) + \sqrt{\frac{n_1 n_2}{n}} \int_a^t H_n(s)[\Lambda_1(ds,v) - r(s)\Lambda_2(ds,v)].
\end{aligned} \tag{7.4}$$

Under H_0, the last term equals zero. Let $\hat{a}(s) = 1/\hat{\lambda}_2(s)$ and $\hat{b}(s) = \lambda_1(s)/(\lambda_2(s)\hat{\lambda}_2(s))$. Let $a(s) = 1/\lambda_2(s)$ and $b(s) = \lambda_1(s)/(\lambda_2(s))^2$. The third term of (7.4) equals

$$\sqrt{\frac{n_1 n_2}{n}} \int_a^t H_n(s)[-\hat{a}(s)(\hat{\lambda}_1(s) - \lambda_1(s)) + \hat{b}(s)(\hat{\lambda}_2(s) - \lambda_2(s))]\Lambda_2(ds, v). \tag{7.5}$$

Next, the third term in (7.4) can be approximated by the integrations with respect to $\hat{Z}_k(t, 1)$, $k = 1, 2$. Note that

$$\hat{\lambda}_k(t) = \frac{1}{b_k} \int_0^{\tau+\delta} K\left(\frac{t-s}{b_k}\right) d\hat{\Lambda}_k(s)$$

and

$$\frac{1}{b_k} \int_0^{\tau+\delta} K\left(\frac{t-s}{b_k}\right) d\Lambda_k(s) = \lambda_k(t) + \frac{1}{2} b_k^2 \lambda_k''(t) \int_{-1}^1 x^2 K(x)\, dx + O(b_k^3),$$

uniformly in $t \in [a, \tau]$. We have, by changing the order of integration and noting the compact support of the kernel function $K(\cdot)$ on $[-1, 1]$,

$$\begin{aligned}
&\sqrt{\frac{n_1 n_2}{n}} \int_a^t H_n(s)\hat{a}(s)(\hat{\lambda}_1(s) - \lambda_1(s))\Lambda_2(ds, v) \qquad (7.6)\\
&= \sqrt{\frac{n_1 n_2}{n}} \int_0^{\tau+\delta} \left[\int_a^t \frac{1}{b_1} K\left(\frac{s-u}{b_1}\right) H_n(s)\hat{a}(s)\Lambda_2(ds, v)\right] d(\hat{\Lambda}_1(u) - \Lambda_1(u)) + O(\sqrt{n} b_1^3)\\
&= \sqrt{\frac{n_1 n_2}{n}} \int_{a-b_1}^{t-b_1} \left[\int_a^t \frac{1}{b_1} K\left(\frac{s-u}{b_1}\right) H_n(s)\hat{a}(s)\Lambda_2(ds, v)\right] d(\hat{\Lambda}_1(u) - \Lambda_1(u))\\
&+ \sqrt{\frac{n_1 n_2}{n}} \int_{t-b_1}^{t+b_1} \left[\int_a^t \frac{1}{b_1} K\left(\frac{s-u}{b_1}\right) H_n(s)\hat{a}(s)\Lambda_2(ds, v)\right] d(\hat{\Lambda}_1(u) - \Lambda_1(u)) + O(\sqrt{n} b_1^3).
\end{aligned}$$

By the uniform convergence of $H_n(s)$ to $H(s)$ and $\hat{a}(s)$ to $a(s)$, and the uniform continuity of $H(s)$ and $a(s)$, we have

$$\frac{1}{b_1} \int_a^t K\left(\frac{s-u}{b_1}\right) H_n(s)\hat{a}(s)\Lambda_2(ds, v) = H(u)a(u)\Lambda_{2u}'(u, v) + o_p(1),$$

uniformly in $u \in (a - b_1, t + b_1)$, $0 \le t \le \tau$, where $\Lambda_{2u}'(u, v) = \partial\Lambda_2(u, v)/\partial u$. Further, the process $\int_a^t b_1^{-1} K((s-u)/b_1) H_n(s)\hat{a}(s)\Lambda_2(ds, v)$ is of bounded variation in u uniformly in n, $v \in [0, 1]$ and $t \in [0, \tau]$, and $H(u)a(u)\Lambda_{2u}'(u, v)$ is of bounded variation uniformly in $v \in [0, 1]$. It follows from Lemma A.1 of Lin and Ying (2001) that (7.6) equals

$$\begin{aligned}
&\sqrt{\frac{n_1 n_2}{n}} \int_{a-b_1}^{t-b_1} H(u)a(u)\Lambda_{2u}'(u, v)\, d(\hat{\Lambda}_1(u) - \Lambda_1(u)) + O(\sqrt{n} b_1^3) + O(b_1)\\
&= \sqrt{\frac{n_2}{n}} \int_a^t H(s)a(s)\Lambda_{2s}'(s, v)\, \hat{Z}_1(ds, 1) + O(\sqrt{n} b_1^3) + o_p(1). \qquad (7.7)
\end{aligned}$$

Similarly,

$$\sqrt{\frac{n_1 n_2}{n}} \int_a^t H_n(s)\hat{b}(s)(\hat{\lambda}_2(s) - \lambda_2(s))\Lambda_2(ds, v)$$
$$= \sqrt{\frac{n_1}{n}} \int_a^t H(s)b(s)\Lambda'_{2s}(s, v)\, d\hat{Z}_2(ds, 1) + O(\sqrt{n}b_2^3) + o_p(1). \tag{7.8}$$

By (7.4), (7.6), (7.7) and (7.8), under $\sqrt{n}b_k^3 \to 0$, as $n \to \infty$ for $k = 1, 2$, we have

$$\begin{aligned} L_n(t, v) &= \sqrt{\frac{n_2}{n}} \left[\int_a^t H_n(s)\hat{Z}_1(ds, v) - \int_a^t H(s)a(s)\Lambda'_{2s}(s, v)\, \hat{Z}_1(ds, 1) \right] \\ &\quad - \sqrt{\frac{n_1}{n}} \left[\int_a^t H_n(s)\hat{r}(s)\hat{Z}_2(ds, v) - \int_a^t H(s)b(s)\Lambda'_{2s}(s, v)\, d\hat{Z}_2(ds, 1) \right] \\ &\quad + \sqrt{\frac{n_1 n_2}{n}} \int_a^t H_n(s)[\Lambda_1(ds, v) - r(s)\Lambda_2(ds, v)] + o_p(1). \end{aligned}$$

By Lemma 1 in Bilias, Gu and Ying (1997), we have

$$\begin{aligned} L_n(t, v) &= \sqrt{\frac{n_2}{n}} \left[\int_a^t H(s)\hat{Z}_1(ds, v) - \int_a^t H(s)a(s)\Lambda'_{2s}(s, v)\, \hat{Z}_1(ds, 1) \right] \\ &\quad - \sqrt{\frac{n_1}{n}} \left[\int_a^t H(s)r(s)\hat{Z}_2(ds, v) - \int_a^t H(s)b(s)\Lambda'_{2s}(s, v)\, d\hat{Z}_2(ds, 1) \right] \\ &\quad + \sqrt{\frac{n_1 n_2}{n}} \int_a^t H_n(s)[\Lambda_1(ds, v) - r(s)\Lambda_2(ds, v)] + o_p(1). \end{aligned}$$

Note that $b(s) = r(s)a(s)$. It follows by the continuous mapping theorem that

$$L_n(t, v) - \sqrt{\frac{n_1 n_2}{n}} \int_a^t H_n(s)[\Lambda_1(ds, v) - r(s)\Lambda_2(ds, v)] \xrightarrow{\mathcal{D}} L(t, v).$$

in $D([a, \tau] \times [0, 1])$. □

Proof of Theorem 2.

Under H_1, the ratio $\lambda_1(t, v)/\lambda_2(t, v)$ increases with v for all $t \in [0, \tau]$. Since $\lambda_k(t) = \int_0^1 \lambda_k(t, v)\, dv$, $k = 1, 2$, and under H_1,

$$\frac{\lambda_1(t, 0)}{\lambda_2(t, 0)} \leq \frac{\lambda_1(t, v)}{\lambda_2(t, v)} \leq \frac{\lambda_1(t, 1)}{\lambda_2(t, 1)},$$

we have

$$\frac{\lambda_1(t, 0)}{\lambda_2(t, 0)} \leq \frac{\lambda_1(t)}{\lambda_2(t)} \leq \frac{\lambda_1(t, 1)}{\lambda_2(t, 1)}.$$

Under the assumptions of Theorem 2, $\frac{\lambda_1(t,v)}{\lambda_2(t,v)}$ is continuous in $v \in [0, 1]$ for every $t \in [0, \tau]$. By the intermediate-value theorem, for every $\in [0, \tau]$ there exists a $v_t \in [0, 1]$ such that

$$r(t) = \frac{\lambda_1(t)}{\lambda_2(t)} = \frac{\lambda_1(t, v_t)}{\lambda_2(t, v_t)}.$$

Since $\lambda_1(t,v)/\lambda_2(t,v)$ increases with v for all $t \in [0,\tau]$, we have

$$\frac{\lambda_1(t,v)}{\lambda_2(t,v)} \geq r(t) \quad \text{for} \quad v \geq v_t \quad \text{and} \quad \frac{\lambda_1(t,v)}{\lambda_2(t,v)} \leq r(t) \quad \text{for} \quad v \leq v_t.$$

Further, since $\int_0^1 H(t)(\lambda_1(t,v) - r(t)\lambda_2(t,v))\,dv = 0$, we have $\int_0^v H(t)(\lambda_1(t,v) - r(t)\lambda_2(t,v))\,dv \leq 0$ for $(t,v) \in [0,\tau] \times [0,1]$. Note that the inequality in H_1 is strict for some (t,v) and the functions $\lambda_1(t,v)$ and $\lambda_2(t,v)$ are continuous. It follows that under H_1, there exists a neighborhood $[t_1,t_2] \times [v_1,v_2]$ such that

$$\int_0^v H(t)(\lambda_1(t,v) - r(t)\lambda_2(t,v))\,dv \leq c < 0.$$

Since $H_n(t) \xrightarrow{P} H(t) > 0$ uniformly in $t \in [0,\tau]$, we have

$$\sqrt{\frac{n_1 n_2}{n}} \sup_{0\leq v\leq 1} \sup_{a\leq t_1\leq t_2\leq \tau} \left(- \int_{t_1}^{t_2} \int_0^v H_n(s)(\lambda_1(s,v) - r(s)\lambda_2(s,v))\,dv\,ds \right) \xrightarrow{P} \infty,$$

as $n \to \infty$. By Proposition 1,

$$\begin{aligned} &L_n(t_2,v) - L_n(t_1,v) - \sqrt{\frac{n_1 n_2}{n}} \int_{t_1}^{t_2}\int_0^v H_n(s)(\lambda_1(s,v) - r(s)\lambda_2(s,v))\,dv\,ds \\ &\xrightarrow{\mathcal{D}} L(t_2,v) - L(t_1,v). \end{aligned}$$

Applying Slusky's Theorem, we have $\hat{U}_1 \xrightarrow{P} \infty$ as $n \to \infty$.

Now, under H_2, by the continuity of the functions, there exist $t \in [0,\tau]$ and $[v_1,v_2]$, such that

$$\left| \int_0^t \int_{v_1}^{v_2} H(s)(\lambda_1(s,v) - r(s)\lambda_2(s,v))\,dv\,ds \right| \geq c > 0.$$

Since $H_n(t) \xrightarrow{P} H(t) > 0$ uniformly in $t \in [0,\tau]$, we have $\sqrt{\frac{n_1 n_2}{n}} |\int_0^t \int_{v_1}^{v_2} H_n(s)(\lambda_1(s,v) - r(s)\lambda_2(s,v))\,dv\,ds|$ $\xrightarrow{P} \infty$ as $n \to \infty$. By Proposition 1,

$$\begin{aligned} &L_n(t,v_2) - L_n(t,v_1) - \sqrt{\frac{n_1 n_2}{n}} \int_0^t \int_{v_1}^{v_2} H_n(s)(\lambda_1(s,v) - r(s)\lambda_2(s,v))\,dv\,ds \\ &\xrightarrow{\mathcal{D}} L(t,v_2) - L(t,v_1). \end{aligned}$$

By Slutsky's Theorem, $|L_n(t,v_2) - L_n(t,v_1)| \xrightarrow{P} \infty$. Therefore $\hat{U}_2 \xrightarrow{P} \infty$ as $n \to \infty$. This completes the proof. □

Proof of the tightness for $L_n^*(t,v)$.

To show tightness of $L_n^*(t,v)$ given the observed data sequence, it suffices to check a slight extension of the moment conditions of Bickel and Wichura (1971) for stochastic processes on the plane, cf. McKeague and Zhang's (1994, page 506) extension of the moment conditions of Billingsley (1968).

It is sufficient to show that ${n_1}^{-1/2}\sum_{i=1}^{n_1}\hat{h}_{1i}(t,v)W_{1i}$ in (3.6) is tight given the observed data sequence. The tightness of the second term follows similarly. Let $B=[t_1,t_2]\times[v_1,v_2]$ and $G=[s_1,s_2]\times[x_1,x_2]$ be any pair of neighboring blocks in $[0,\tau]\times[0,1]$. Let $\hat{h}_{1i}(B)=\hat{h}_{1i}(t_2,v_2)-\hat{h}_{1i}(t_2,v_1)-\hat{h}_{1i}(t_1,v_2)+\hat{h}_{1i}(t_1,v_1)$ and

$$\Delta(B)=n_1^{-1/2}\sum_{i=1}^{n_1}\hat{h}_{1i}(B)W_{1i}.$$

We show that there exists a finite measure μ_0 on $[0,\tau]\times[0,1]$ such that

$$E\left\{\Delta^2(B)\middle|\{\text{observed data}\}\right\}\le\mu_0(B)+o_p(1) \tag{7.9}$$

$$E\left\{\Delta^2(B)\Delta^2(G)\middle|\{\text{observed data}\}\right\}\le\mu_0(B)\mu_0(G)+o_p(1), \tag{7.10}$$

where the $o_p(1)$ term converges to zero in probability independently of (or uniformly in) B and G. Since a simple linear combination of tight processes is tight, it suffices to check the conditions (7.9) and (7.10) for each of the four terms in $\hat{h}_{1i}$. However, for ease of notation we use $\hat{h}_{1i}$ to represent any one of the four terms.

By the uniform convergence of $H_n(s)$, $Y_k(s)$, $N_k(s,v)/n_k$, $\hat{a}(s)$, and $\hat{\Lambda}'_{2s}(s,v)$ on $[a,\tau]\times[0,1]$, a simple probability argument yields that

$$E\left\{\Delta^2(B)\middle|\{\text{observed data}\}\right\}\le n_1^{-1}\sum_{i=1}^{n_1}(\hat{h}_{1i}(B))^2+o_p(1) \tag{7.11}$$

$$E\left\{\Delta^2(B)\Delta^2(G)\middle|\{\text{observed data}\}\right\}\le 6n_1^{-2}\sum_{i=1}^{n_1}(\hat{h}_{1i}(B))^2\sum_{i=1}^{n_1}(\hat{h}_{1i}(G))^2+o_p(1) \tag{7.12}$$

Then (7.9) and (7.10) follow from working with each of the four terms of $\hat{h}_{1i}$ in (7.11) and (7.12). The details are omitted. □

References

Andersen, P. K., Borgan, O., Gill, R. D. and Keiding, N. (1993). *Statistical Models Based on Counting Processes.* New York: Springer.

Bickel, P. J. and Wichura, M. J. (1971). Convergence criteria for multiparameter stochastic processes and some applications. *Annals of Mathematical Statistics*, 42, 1656–1670.

Bilias, Y., Gu, M. and Ying, Z. (1997). Towards a general asymptotic theory for Cox model with staggered entry. *Annals of Statistics*, 25, 662–682.

Billingsley, P. (1968). *Convergence of Probability Measures*, New York, Wiley.

Fine, J. P. and Gray, R. J. (1999). A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association*, 94, 496–509.

Gasser, T. and Müller, H-G. (1979). Kernel estimation of regression functions. In *Smoothing Techniques for Curve Estimation*, Lecture Notes in Mathematics 757, 23–68. Berlin: Springer-Verlag.

Gilbert, P. B. (2000). Large sample theory of maximum likelihood estimates in semiparametric biased sampling models. *Annals of Statistics*, 28, 151–194.

Gilbert, P. B., Lele, S. and Vardi Y. (1999). Maximum likelihood estimation in semiparametric selection bias models with application to AIDS vaccine trials. *Biometrika*, 86, 27–43.

Gilbert, P. B., Hanna, G., De Gruttola, V., Martinez-Picado, J., Kuritzkes, D., Johnson, V., Richman, D., D'Aquila, R. (2000). Comparative analysis of HIV-1 type 1 genotypic resistance across antiretroviral trial treatment regimens. *AIDS Research and Human Retroviruses*, 16, 1325-1336.

Gilbert, P. B., Self, S. G., Rao, M., Naficy, A. and Clemens, J. D. (2001). Sieve analysis: methods for assessing from vaccine trial data how vaccine efficacy varies with genotypic and phenotypic pathogen variation. *Journal of Clinical Epidemiology*, 54, 68–85.

Gilbert, P. B., Wei, L. J., Kosorok, M. R. and Clemens, J. D. (2002). Simultaneous inference on the contrast of two hazard functions with censored observations. *Biometrics*, 58, 773–780.

Gilbert, P. B., McKeague, I. W., Sun, Y. (2004). Tests for comparing mark-specfic hazards and cumulative incidence functions. *Lifetime Data Analysis*, 10, 5–28.

Graham, B. S. (2002). Clinical trials of HIV vaccines. *Annual Review of Medicine*, 53, 207–21.

Gray, R. J. (1988). A class of k-sample tests for comparing the cumulative incidence of a competing risk, *Ann. Statist.* **16**, 1141–1154.

Halloran, M. E., Haber M. J. and Longini I. M. (1992). Interpretation and estimation of vaccine efficacy under heterogeneity. *American Journal of Epidemiology*, 136, 328-343.

Halloran, M.E., Struchiner, C.J. and Longini, I.M. (1997). Study designs for different efficacy and effectiveness aspects of vaccination. *American Journal of Epidemiology*, 146, 789-803.

Huang, Y. and Louis, T. A. (1998). Nonparametric estimation of the joint distribution of survival time and mark variables. *Biometrika*, 85, 785–798.

Kelsall, J. E. and Diggle, P. J. (1995). Kernel estimation of relative risk. *Bernoulli*, 1, 3–16.

Lin, D. Y. and Ying, Z. (2001). Semiparametric and nonparametric regression analysis of longitudinal data (with discussion). *J. Amer. Statist. Assoc.*, 96, 103–113.

Lin, D. Y., Wei, L. J. and Ying, Z. (1993). Checking the Cox model with cumulative sums of martingale-based residuals. *Biometrika* **80**, 557–572. Lin, D. Y., Fleming, T. R. and Wei, L. J. (1994). Confidence bands for survival curves under the proportional hazards model. *Biometrika* **81**, 73–81.

McKeague, I. W. and Zhang, M. J. (1994). Identification of nonlinear time series from first order cumulative characteristics. *The Annals of Statistics* **22** 495–514

McKeague, I. W. Gilbert, P. B. and Kanki, P. J. (2001). Comparison of competing risks with adjustment for covariate effects. *Biometrics* **57** 818-828.

Nabel, G. J. (2001). Challenges and opportunities for development of an AIDS vaccine. *Nature*, 410, 1002–7.

Ramlau-Hansen, H. (1983). Smoothing counting process intensities by means of kernel functions. *The Annals of Statistics* **11**, 453–466.

Rice, J. and Rosenblatt, M. (1976). Estimation of the log survivor function and hazard function. *Sankhyā, Series A* **38**, 60–78.

Shorack, G. R. and Wellner, J. A. (1986). *Empirical Processes with Applications to Statistics*, New York, Wiley.

Sun, Y. (2001). Generalized nonparametric test procedures for comparing multiple cause-specific hazard rates. *Journal of Nonparametric Statistics*, 13, 171-207.

Tanner, M. A. and Wong, W.-H. (1983). The estimation of the hazard function from randomly censored data by the kernel method. *The Annals of Statistics* **11**, 989–993.

Wu, T.-J., Hsieh, Y.-C. and Li, L.-A. (2001). Statistical measures of DNA sequence dissimilarity under Markov chain models of base composition. *Biometrics*, 57, 441-448.

Wyatt, R., Kwong, P. D., Desjardins, E., Sweet, R. W., Robinson, J., Hendrickson, W. A., Sodroski, J. G. (1998). The antigenic structure of the HIV gp120 envelope glycoprotein. *Nature*, 393, 705-711.

Yandell, B.S. (1983). Nonparametric inference for rates with censored survival data. *The Annals of Statistics* **11**, 1119–1135.

Table 1. Empirical power ($\times$ 100%) for testing H_1 and H_2; hazard and mark independent

		$\mathrm{VE}^c(\tau) = 0.67$					$\mathrm{VE}^c(\tau) = 0.33$				
			β_1					β_1			
n_k	*Test*		1	0.5	0.25	*2-sided*		1	0.5	0.25	*2-sided*
100 (48)[1]	U_1	(16)[2]	3.4	13.0	49.3	4.0	(32)[2]	7.1	29.8	85.5	18.0
	U_2		2.8	5.0	14.9	23.5		6.2	11.0	42.5	61.3
200 (95)[1]	U_1	(31)[2]	2.7	20.2	81.8	5.6	(64)[2]	6.5	46.2	99.0	31.9
	U_2		1.9	5.0	36.7	53.4		5.3	16.8	81.3	91.5
400 (190)[1]	U_1	(62)[2]	2.3	29.1	99.3	20.8	(128)[2]	4.4	65.9	100	68.3
	U_2		1.0	11.0	77.3	89.9		4.0	32.7	99.2	99.7

[1]Average number of subjects infected in group 2 (placebo).

[2]Average number of subjects infected in group 1 (vaccine) under H_0.

Table 2. Empirical power ($\times$ 100%) for testing H_1 and H_2; hazard and mark dependent

		$\mathrm{VE}^c(\tau) = 0.67$					$\mathrm{VE}^c(\tau) = 0.33$				
			β_1					β_1			
n_k	*Test*		1	0.5	0.25	*2-sided*		1	0.5	0.25	*2-sided*
100 (48)[1]	U_1	(16)[2]	3.0	25.6	75.0	4.3	(32)[2]	5.6	71.7	99.2	26.0
	U_2		2.8	8.3	35.8	20.8		6.4	34.0	73.7	66.1
200 (95)[1]	U_1	(31)[2]	1.4	47.4	98.0	8.5	(64)[2]	5.7	95.2	100	49.5
	U_2		1.7	18.0	65.1	46.3		6.5	67.2	98.5	92.9
400 (190)[1]	U_1	(62)[2]	0.6	82.2	100	24.4	(128)[2]	4.3	99.9	100	83.6
	U_2		1.8	47.0	95.7	87.0		5.5	94.2	100	99.9

[1]Average number of subjects infected in group 2 (placebo).

[2]Average number of subjects infected in group 1 (vaccine) under H_0.

Table 3. Bias of $\widehat{\mathrm{VE}}^c(36, v)$ and 95% coverage probability of $\mathrm{VE}^c(36, v)$; hazard and mark independent

		$\mathrm{VE}^c(\tau) = 0.67$			$\mathrm{VE}^c(\tau) = 0.33$		
		β_1			β_1		
n_k	v	1	0.5	0.25	1	0.5	0.25
Average Bias $\times$ 100							
100 (48)[1]	0.3	−2.3	−6.3	−31.6	−2.5	−5.0	−20.8
	0.5	−1.3	−2.6	−13.7	−3.6	−3.6	−9.0
	0.8	−3.7	−3.0	−3.6	−5.2	−5.1	−9.6
200 (95)[1]	0.3	−0.1	−1.6	−13.0	−0.9	−1.6	−9.0
	0.5	−0.0	−0.9	−4.8	−1.0	−2.2	−6.0
	0.8	−0.5	−0.6	−1.5	−2.1	−2.7	−5.4
400 (190)[1]	0.3	−0.0	−0.4	−3.7	−0.2	−0.1	−3.0
	0.5	−0.1	−0.8	−3.6	−0.0	−0.9	−4.6
	0.8	−0.3	0.1	−0.9	−0.3	−0.2	−2.4
Coverage Probability $\times$ 100%							
100 (48)[1]	0.3	97.9	96.0	73.9	97.2	97.3	86.6
	0.5	98.6	97.5	90.0	97.5	97.9	95.2
	0.8	96.0	96.2	95.4	94.6	94.9	96.1
200 (95)[1]	0.3	96.5	96.8	77.1	97.8	97.1	88.0
	0.5	96.7	97.5	93.8	96.8	97.5	96.5
	0.8	94.4	95.3	95.8	94.5	95.6	95.9
400 (190)[1]	0.3	95.4	96.4	87.8	96.8	97.3	92.2
	0.5	96.3	95.9	93.6	96.5	97.2	96.4
	0.8	96.0	96.3	96.7	96.2	96.8	96.8

[1]Average number of subjects infected in group 2 (placebo).

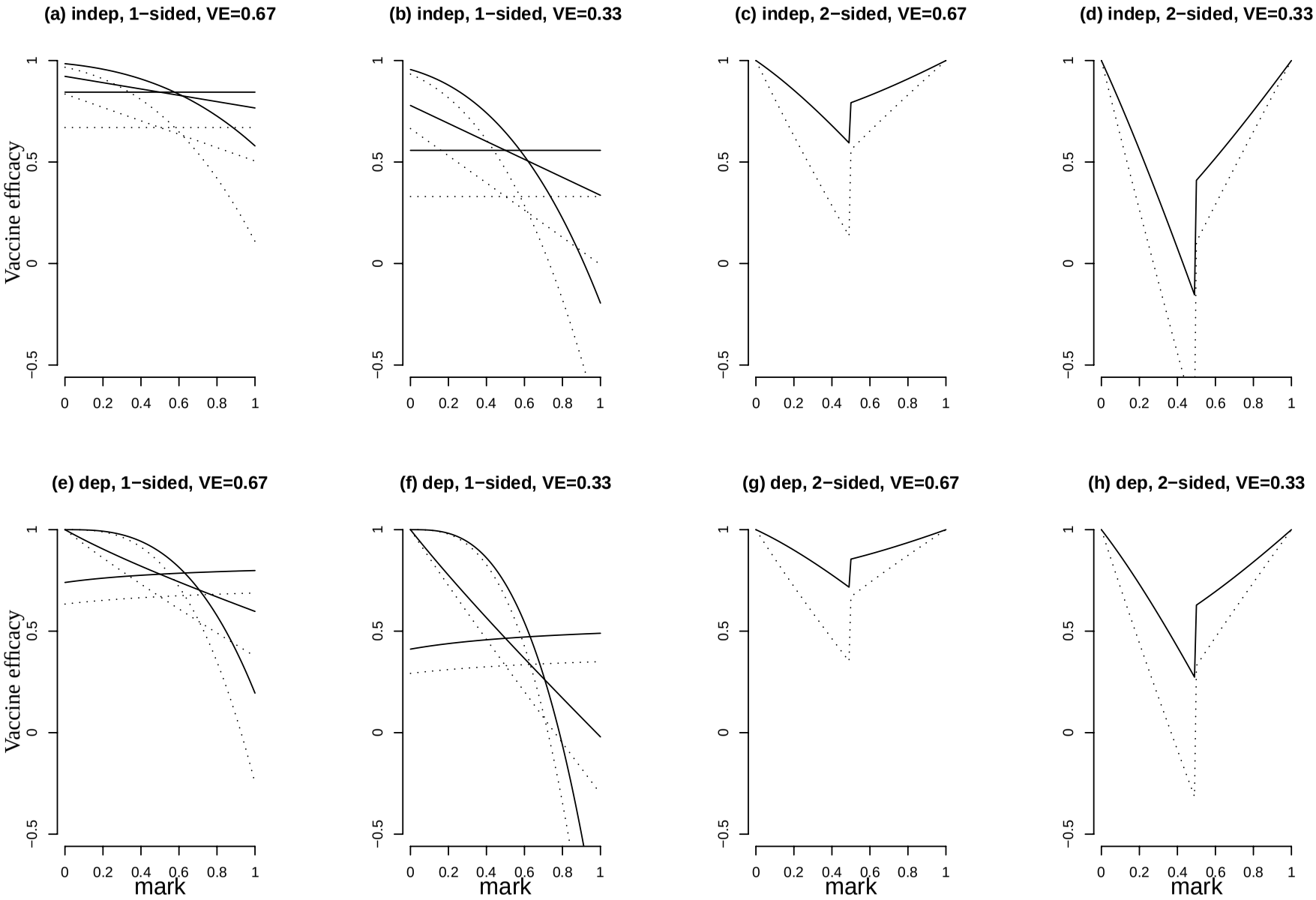
Figure Captions

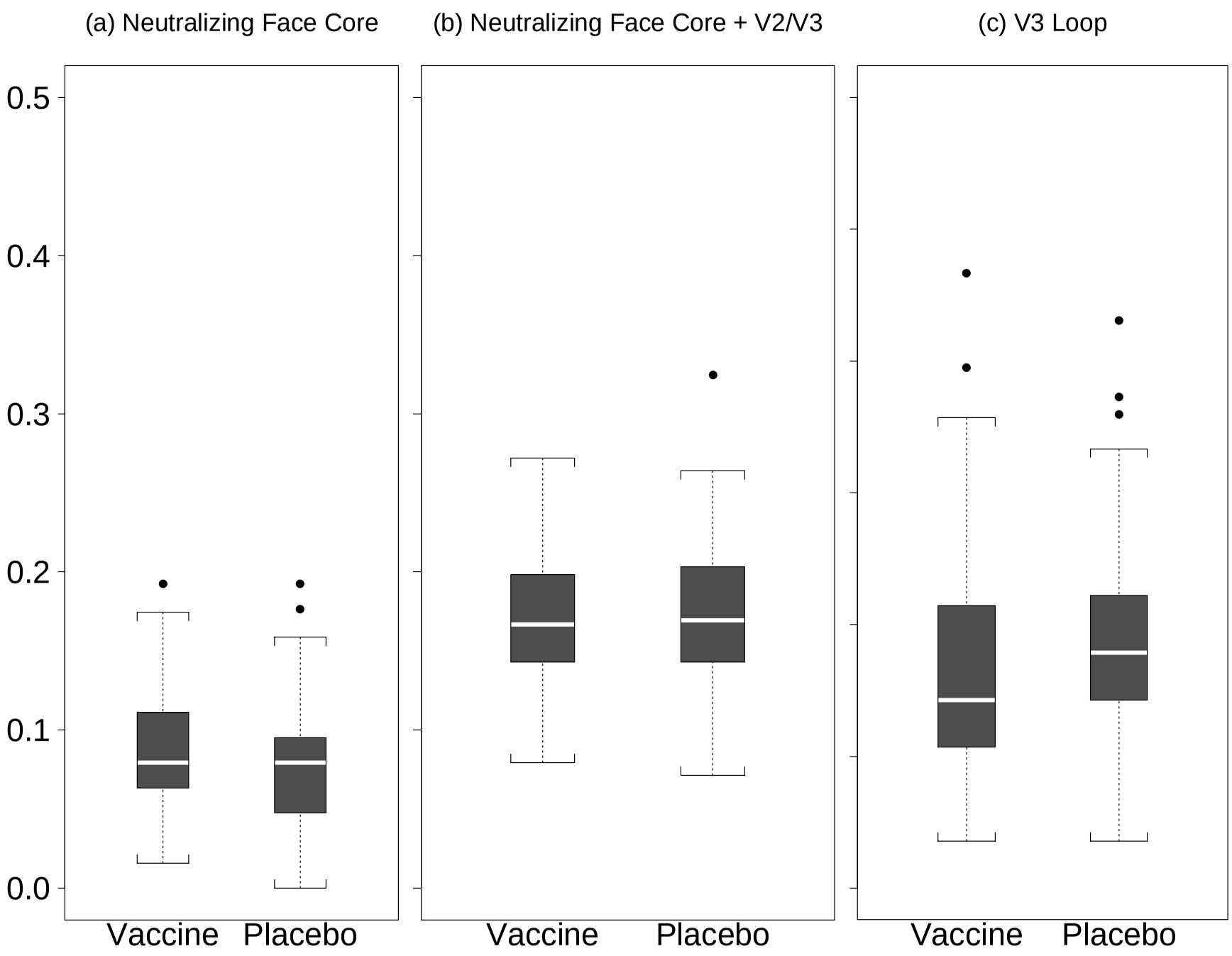
Figure 1. The figure shows the true $\mathrm{VE}(36, v)$ (solid lines) and true $\mathrm{VE}^c(36, v)$ (dashed lines) used in the simulation study for (a) $\mathrm{VE}^c(36) = 0.67$, mark and hazard independent (indep), 1-sided alternative; (b) $\mathrm{VE}^c(36) = 0.33$, indep, 1-sided; (c) $\mathrm{VE}^c(36) = 0.67$, indep, 2-sided; (d) $\mathrm{VE}^c(36) = 0.33$, indep, 2-sided; (e) $\mathrm{VE}^c(36, v = 0.5) = 0.67$, mark and hazard dependent (dep), 1-sided alternative; (f) $\mathrm{VE}^c(36, v = 0.5) = 0.33$, dep, 1-sided; (g) $\mathrm{VE}^c(36, v = 0.5) = 0.67$, dep, 2-sided; (h) $\mathrm{VE}^c(36, v = 0.5) = 0.33$, dep, 2-sided.

Figure 2. For the VaxGen HIV vaccine trial, the figure shows boxplots of amino acid Hamming distances in HIV gp120 between the infecting viruses and the nearest vaccine strain MN or GNE8, for distances computed in (a) the neutralizing face core, (b) the neutralizing face core plus the V2/V3 loops, and (c) the V3 loop.

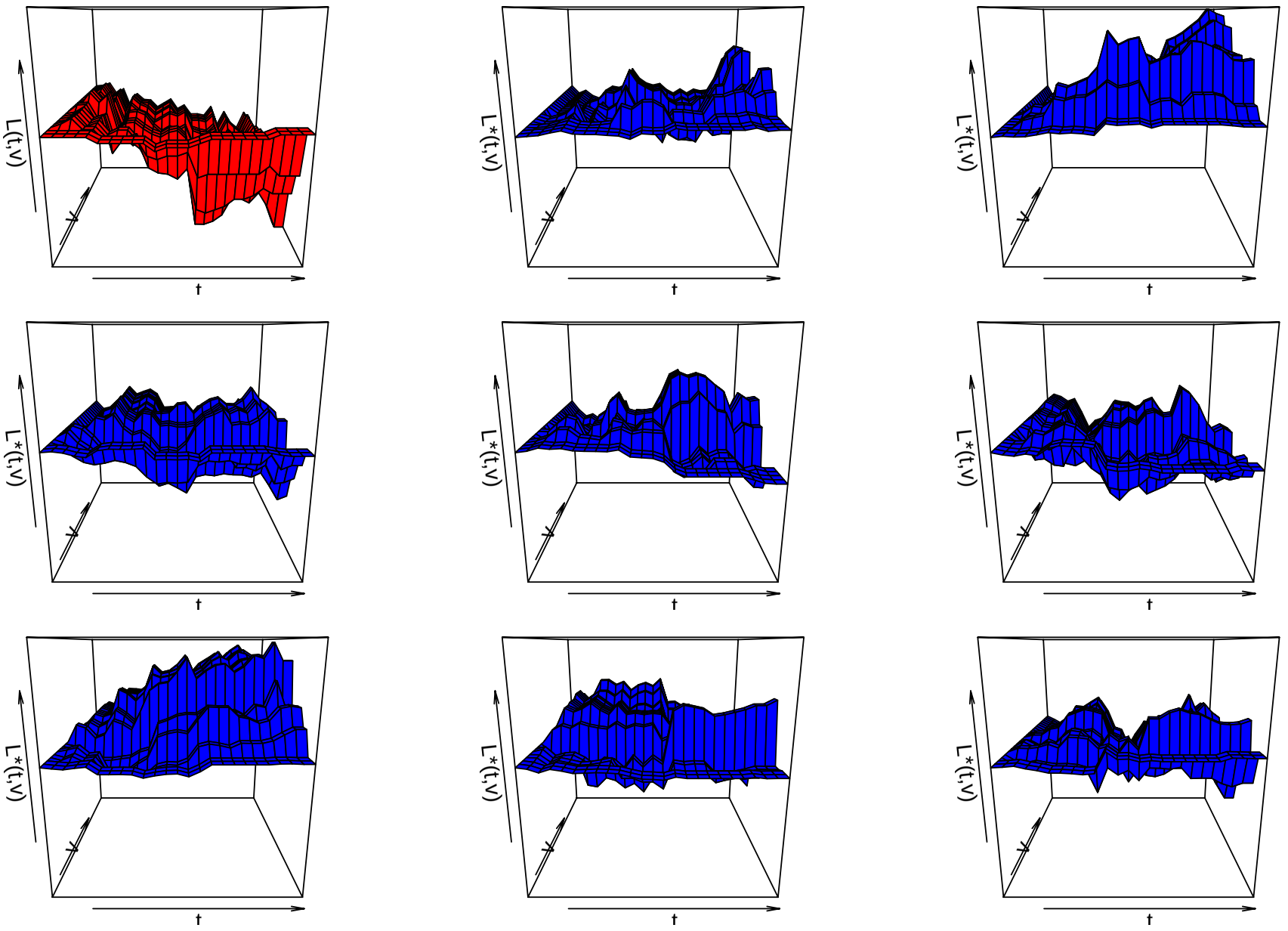
Figure 3. For the VaxGen HIV vaccine trial and neutralizing face core distances, the top-left panel shows the observed test process $L_n(t, v)$ and the other panels show 8 randomly selected realizations of the simulated null test process $L_n^*(t, v)$.

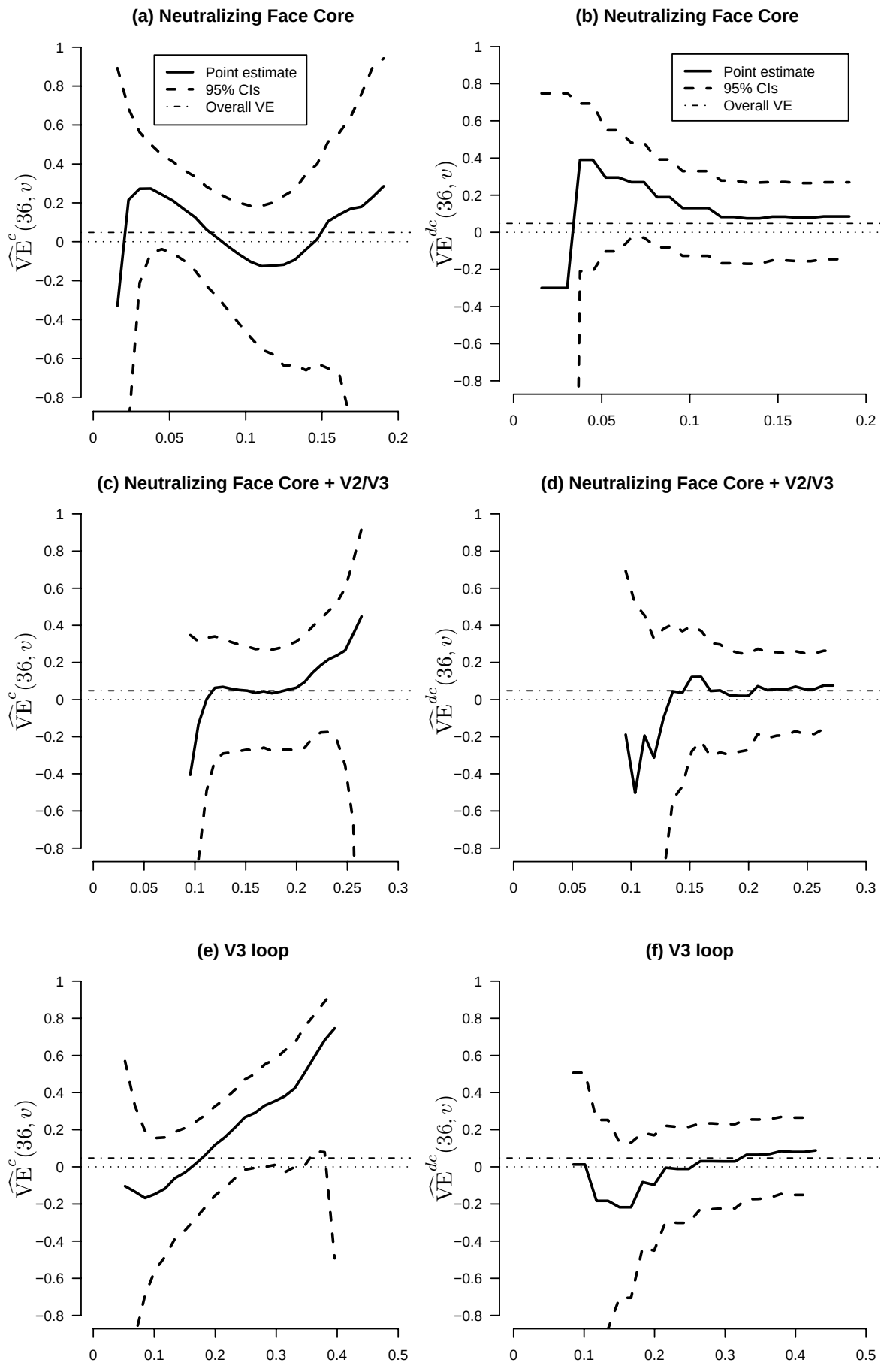
Figure 4. For the VaxGen HIV vaccine trial, the left panels show point and 95% confidence interval estimates of $\mathrm{VE}^c(36, v) = 1 - F_1(36, v)/F_2(36, v)$ versus the HIV gp120 amino acid distance between infecting viruses and the nearest vaccine antigen MN or GNE8, for distances computed in (a) the neutralizing face core, (c) the neutralizing face core plus the V2/V3 loops, and (e) the V3 loop. The right panels show corresponding point and interval estimates of $\mathrm{VE}^{dc}(36, v) = 1 - P(T_1 \leq 36, V_1 \leq v)/P(T_2 \leq 36, V_2 \leq v)$ for these three distances.


(a) indep, 1−sided, VE=0.67
(b) indep, 1−sided, VE=0.33
(c) indep, 2−sided, VE=0.67
(d) indep, 2−sided, VE=0.33
(e) dep, 1−sided, VE=0.67
(f) dep, 1−sided, VE=0.33
(g) dep, 2−sided, VE=0.67
(h) dep, 2−sided, VE=0.33
Vaccine efficacy
mark
−0.5
0
0.5
1
0
0.2
0.4
0.6
0.8
1


(a) Neutralizing Face Core
(b) Neutralizing Face Core + V2/V3
(c) V3 Loop
0.5
0.4
0.3
0.2
0.1
0.0
Vaccine
Placebo
Vaccine
Placebo
Vaccine
Placebo

Test process and 8 simulated test processes for neutralizing face core distance




(a) Neutralizing Face Core
(b) Neutralizing Face Core
(c) Neutralizing Face Core + V2/V3
(d) Neutralizing Face Core + V2/V3
(e) V3 loop
(f) V3 loop
Point estimate
95% CIs
Overall VE
$\widehat{\mathrm{VE}}^{c}(36, v)$
$\widehat{\mathrm{VE}}^{dc}(36, v)$