The two-sample problem for failure rates depending on a continuous mark: An application to vaccine efficacy

PETER B. GILBERT*

Department of Biostatistics, University of Washington and Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, Seattle, Washington 98109, USA

pgilbert@scharp.org

IAN W. McKEAGUE

Department of Biostatistics, Columbia University, Mailman School of Public Health, 722 West 168th Street, 6th Floor, New York, NY 10032

im2131@columbia.edu

YANQING SUN

Department of Mathematics and Statistics, University of North Carolina at Charlotte, 9201 University City Boulevard, Charlotte, North Carolina 28223, USA

yasun@uncc.edu

SUMMARY

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 on the HIV sequences that infect vaccine efficacy trial participants to 1) test for vaccine efficacy more powerfully than procedures that ignore the sequence data; and 2) evaluate the dependence of vaccine efficacy on the divergence of infecting HIV strains 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 genetic divergence as a continuous mark variable that accompanies each failure (infection) time. Problems 1) and 2) can then be approached by testing whether the ratio of the mark-specific hazard functions for the vaccine and placebo groups is unity or independent of the mark. We develop nonparametric and semiparametric tests for these null hypotheses, and nonparametric techniques for estimating the mark-specific relative risks. The asymptotic properties of the procedures are established. In addition the methods are studied in simulations and are applied to HIV genetic sequence data collected in the first HIV vaccine efficacy trial.

Keywords: Competing risks; Genetic data; Mark variable; Nonparametric statistics;

Proportional hazards; Survival analysis.

*To whom correspondence should be addressed. Phone: 206 667-7299. Fax: 206 667-4812.

1. Introduction

In many studies involving the comparison of survival data from two treatment groups, a mark variable is measured only in failures, and it is of interest to account for this mark in comparing the failure experience. In this article, we develop testing and estimation procedures to assess mark-specific relative risks. This departs from our recent work (Gilbert *et al.*, 2004) 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). The two-sample problem addressed here is scientifically compelling, as elaborated below, whereas the one-sample problem has limited relevance to clinical trials. Furthermore our approach to the two-sample problem entails major differences in hypotheses and inferential procedures compared to those used for the one-sample problem, and here we expand the scope of research to include estimation as well as testing, and semiparametric as well as nonparametric hypothesis testing.

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

From 1998 to 2003 VaxGen Inc. conducted the world's first HIV vaccine efficacy trial (Flynn *et al.*, 2005). HIV uninfected volunteers at high risk for acquiring HIV were randomized to receive the vaccine AIDSVAX ($n_1 = 3,598$) or placebo ($n_2 = 1,805$).

Subjects were monitored for 3 years for the primary study endpoint HIV infection. For each subject who became HIV infected, the envelope glycoprotein (gp120) region of the infecting virus was sequenced. Of the 368 subjects who acquired HIV, the sequence data were collected for 336 subjects (217 of 241 vaccine; 119 of 127 placebo). VaxGen hypothesized that the level of vaccine efficacy would be higher against HIVs with gp120 amino acid sequences that were relatively similar to either of the two HIV strains (named MN and GNE8) that were represented in the vaccine. The distance of each infecting virus to MN and to GNE8 was measured by the percent mismatch in the aligned amino acid sequences (i.e., Hamming distance) for three sets of positions hypothesized to be important for neutralizing HIV (Wyatt *et al.*, 1998): (1) the neutralizing face core of gp120 that was crystalized; (2) the neutralizing face core plus the variable loop V2/V3 regions; and (3) the V3 loop. For each metric and infecting virus, the mark is defined as the minimum of the two distances to the MN and GNE8 reference sequences.

Gilbert *et al.* (1999) and Gilbert (2000) developed a semiparametric biased sampling model as a tool for studying vaccine efficacy as a function of a continuous mark, which parametrically specifies the relationship between vaccine efficacy and the mark, and leaves the distribution of the mark in the infected placebo group unspecified. However, there are no data available for suggesting the correct parametric model, so nonparametric methods are desirable. Furthermore, the earlier work is limited by conditioning on infection, so odds ratios but not relative risks of infection can be estimated, and the model treats HIV infection as a binary outcome, ignoring the time to HIV infection. The methods presented here were developed because they are free from these limitations, as they are nonparametric (though semiparametric procedures are also considered), prospective, and incorporate the failure times.

We introduce tests for the hypothesis that the mark-specific risks in the two groups coincide, and for the hypothesis that the relative mark-specific risk between the 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 joint 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. Since the mark is only observed for failures, it cannot be studied as a covariate in evaluating risk. We assume that each mark variable V_k has known and bounded support; rescaling V_k if necessary, this support is taken to be $[0, 1]$. 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 $[0, \tau]$. These functions are natural extensions of the cause-specific hazard function and cumulative incidence function that have been extensively studied for a discrete mark variable (e.g., Prentice *et al.*, 1978). Similar to the discrete mark case, 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 for comparing cumulative incidence functions among groups, at a specified value of the mark variable.

A standard measure of vaccine efficacy to prevent infection at time t is the relative reduction in hazard due to vaccination: $\text{VE}(t) = 1 - \lambda_1(t)/\lambda_2(t)$, see Halloran *et al.* (1997). It is natural to extend this definition to allow the vaccine efficacy to depend on viral divergence: $\text{VE}(t, v) = 1 - \lambda_1(t, v)/\lambda_2(t, v)$. This parameter has a useful approximate interpretation as the relative multiplicative reduction (vaccine versus placebo) in the probability that one exposure to strain v at time t (by a sexual or needle contact) leads to infection (Halloran *et al.*, 1992). This interpretation follows by randomization, blinding, and the fact that HIV infection is a rare event in HIV vaccine efficacy trials (typically occurring in fewer than 10-15% of subjects), which together imply that the amount of exposure to HIV strains of distance v is approximately equal in the vaccine and placebo groups throughout the follow-up period.

To account for the mark in testing for vaccine efficacy, we develop tests for the null

hypothesis

$$H_0^0\text{: } \lambda_1(t,v) = \lambda_2(t,v) \text{ for } (t,v) \in [0,\tau] \times [0,1]$$

against the following alternative hypotheses:

$$H_1^0\text{: } \lambda_1(t,v) \leq \lambda_2(t,v) \text{ for all } \ (t,v) \in [0,\tau] \times [0,1];$$

$$H_2^0\text{: } \lambda_1(t,v) \neq \lambda_2(t,v) \text{ for some } \ (t,v) \in [0,\tau] \times [0,1]$$

with strict inequality for some $(t,v) \in [0,\tau] \times [0,1]$ in H_1^0. Testing H_0^0 evaluates $\text{VE}(t,v) = 0$ for all t and v, i.e., whether there is any vaccine efficacy against any HIV strain. As we show in simulations, tests of H_0^0 can have much greater power than standard tests of vaccine efficacy that ignore the mark, i.e., that evaluate the null hypothesis $\lambda_1(t) = \lambda_2(t)$ for all $t \in [0,\tau]$. A test ignoring the mark should be done in conjunction with a test of H_0^0, however, to assess the overall clinical/public health benefit of the vaccine. To illustrate the importance of carrying out both tests, if $\text{VE}(t) = 0$ and $\text{VE}(t,v)$ is positive (negative) for $v \leq (>)\ 0.5$, then the vaccine clearly should not be declared effective, yet the analysis accounting for the mark would lead to follow-up studies of the mechanism by which the vaccine impacted mark-specific infection risk.

If H_0^0 is rejected, then it is of interest to assess if vaccine efficacy varies with strain distance. Accordingly, we also develop tests for

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

against the following alternative hypotheses:

$$H_1\text{: } \lambda_1(t,v_1)/\lambda_2(t,v_1) \ \leq \ \lambda_1(t,v_2)/\lambda_2(t,v_2) \text{ for all } v_1 \leq v_2,\ t \in [0,\tau];$$

$$H_2\text{: } \lambda_1(t,v_1)/\lambda_2(t,v_1) \ \neq \ \lambda_1(t,v_2)/\lambda_2(t,v_2) \text{ for some } v_1 \leq v_2,\ t \in [0,\tau]$$

with strict inequality for some t, v_1, v_2 in H_1. Because H_0 and H_1 can be re-expressed as $H_0 : \text{VE}(t,v) = \text{VE}(t)$ for all t, v and $H_1 : \text{VE}(t,v_1) \leq \text{VE}(t,v_2)$ for all $t, v_1 \geq v_2$ (with $<$ for some $v_1 > v_2$), testing H_0 versus H_1 assesses whether vaccine efficacy decreases with HIV sequence divergence. Scientifically this is the most interesting hypothesis to assess.

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. In Section 2 we introduce the proposed procedures for testing H_0^0 and H_0. Large sample results and a simulation technique needed to implement the test procedures are developed in Section 3. In Section 4 we discuss nonparametric estimation of the mark-specific vaccine efficacy. We report the results of a simulation experiment in Section 5, and an application to data from the VaxGen trial is provided in Section 6. Section 7 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 $\Lambda_k(t,v) = \int_0^v \int_0^t \lambda_k(s,u)\,ds\,du$, $k = 1, 2$. 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 (MLE) of $\Lambda_k(t,v)$ is provided by the Nelson–Aalen-type estimator

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

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

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

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

Our tests of H_0^0 are based on comparing $\hat{\Lambda}_1(t,v)$ and $\hat{\Lambda}_2(t,v)$, and of H_0 are based on comparing the nonparametric MLE of $\Lambda_1(t,v) - \Lambda_2(t,v)$ with an estimate under H_0. Since H_0 is equivalent to $\Lambda_1(t,v) = \int_0^t [\lambda_1(s)/\lambda_2(s)]\Lambda_2(ds,v)$ for all t, v, under H_0 we may estimate the difference $\Lambda_1(t,v) - \Lambda_2(t,v)$ 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. Alternatively, under a proportional marginal hazards assumption, $\lambda_1(t)/\lambda_2(t) = \exp(\beta)$, this difference may be estimated by $\int_0^t [\exp(\hat{\beta}) - 1]\hat{\Lambda}_2(ds,v)$, where $\hat{\beta}$ is the maximum partial likelihood estimator of β, which leads to a semiparametric test for H_0. The nonparametric

approach makes minimal assumptions but requires smoothing, whereas the semiparametric approach avoids smoothing and in principle may provide greater power when the proportional hazards assumption holds.

For the nonparametric approach we estimate each hazard function $\lambda_k(t)$ by kernel smoothing:

$$\hat{\lambda}_k(t) = \frac{1}{b_k} \int_0^{\tau+\delta} K\left(\frac{t-s}{b_k}\right) d\hat{\Lambda}_k(s)\,, \tag{2.2}$$

where $\hat{\Lambda}_k(s) = \int_0^t (1/Y_k(s))\, dN_k(s)$ is the 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 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 $\hat{\Lambda}_k(t)$ is smoothed, and converges to zero as $n_k \to \infty$. We choose kernel esimators because they are uniformly consistent under assumptions (see Theorem IV.2.2 in Andersen *et al.*, 1993), a property that is needed for the theoretical justification given later.

2.1 Test Processes and Test Statistics

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

$$L_n^r(t, v) = \sqrt{\frac{n_1 n_2}{n}} \int_a^t H_n(s)\ \left[\hat{\Lambda}_1(ds, v) - \hat{r}(s)\hat{\Lambda}_2(ds, v)\right] \tag{2.3}$$

for $t \ge 0, 0 \le v \le 1$, where $H_n(\cdot)$ is a suitable weight process converging to a non-random function $H(\cdot)$ and $a \ge 0$. The superscript r reflects the choice of process $\hat{r}(s)$ in the test process and indicates whether it is used to test H_0^0 (indicated by r as 1, corresponding to $\hat{r}(s) = 1$), to test H_0 nonparametrically (indicated by r as np; $\hat{r}(s) = \hat{\lambda}_1(s)/\hat{\lambda}_2(s)$) or to test H_0 semiparametrically (indicated by r as sp; $\hat{r}(s) = \exp(\hat{\beta})$). A simple calculation shows that for r as np, $[\cdot]$ in (2.3) compares $\hat{\Lambda}_1(ds, v) - \hat{\Lambda}_2(ds, v)$ to the nonparametric estimate of $\Lambda_1(ds, v) - \Lambda_2(ds, v)$ under H_0 described above. The parallel result holds for r as sp, using the semiparametric estimate of $\Lambda_1(ds, v) - \Lambda_2(ds, v)$ under H_0.

A variety of test statistics can be formulated as functionals of $L_n^r(t, v)$. We develop integration type and supremum type statistics. With $w_V(v)$ a known nonnegative weight function, large values of the following test statistics provide evidence against H_0^0 in the direction of H_0^1 (first two statistics) or H_0^2 (second two statistics):

$$\hat{U}_1^1 \;=\; L_n^1(\tau, 1), \qquad\qquad \hat{U}_2^1 = \int_0^1 w_V(v) L_n^1(\tau, v) dv, \tag{2.4}$$

$$\hat{U}_3^1 = |L_n^1(\tau,1)|, \qquad \hat{U}_4^1 = \int_0^1 w_V(v)(L_n^1(\tau,v))^2 dv. \tag{2.5}$$

For testing H_0, let $y_k(t) = P(X_k \geq t)$, let $\tilde{\tau} = \sup\{t\colon y_1(t) > 0 \text{ and } y_2(t) > 0\}$, 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 $L_n^r(t,v)$ over $a \leq t \leq \tau, 0 \leq v \leq 1$. In practice, however, there would be no harm in taking $a = 0$ in order to use as much of the data as possible (this is done in the simulations and application).

Set $\Delta_n^r(t,v_1,v_2) = L_n^r(t,v_1) + L_n^r(t,v_2) - 2L_n^r(t,(v_1+v_2)/2)$. For r as np or sp, the following test statistics measure departures from H_0 in the direction of H_1 ($\hat{U}_1^r$) or H_2 ($\hat{U}_2^r$):

$$\hat{U}_1^r = \sup_{v_1<v_2} \sup_{0\leq t_1<t_2<\tau} \left[\Delta_n^r(t_2,v_1,v_2) - \Delta_n^r(t_1,v_1,v_2)\right], \tag{2.6}$$

$$\hat{U}_2^r = \sup_{v_1<v_2} \sup_{0\leq t_1<t_2<\tau} \left|\Delta_n^r(t_2,v_1,v_2) - \Delta_n^r(t_1,v_1,v_2)\right|. \tag{2.7}$$

To motivate the test statistics $\hat{U}_1^r$ and $\hat{U}_2^r$, we note from the proof of Theorem 2 in the Appendix that $(n/n_1n_2)^{1/2}[\Delta_n^r(t_2,v_1,v_2) - \Delta_n^r(t_1,v_1,v_2)]$ converges in probability to $\delta(t_1,t_2,v_1,v_2) =$

$$\int_{t_1}^{t_2}\int_{\frac{v_1+v_2}{2}}^{v_2} H(s)(\lambda_1(s,v) - r(s)\lambda_2(s,v))\,dv\,ds - \int_{t_1}^{t_2}\int_{v_1}^{\frac{v_1+v_2}{2}} H(s)(\lambda_1(s,v) - r(s)\lambda_2(s,v))\,dv\,ds,$$

where $r(s) = \lambda_1(s)/\lambda_2(s)$ or $\exp(\beta)$. Under H_0, $\delta(t_1,t_2,v_1,v_2) = 0$ for all $t_1,t_2 \in [0,\tau]$ and $v_1,v_2 \in [0,1]$. Under H_1 and some smoothness conditions, $\delta(t_1,t_2,v_1,v_2) > 0$ for some $t_1 < t_2 \in [0,\tau]$ and $v_1 < v_2 \in [0,1]$. Therefore large values of $\hat{U}_1^r$ ($\hat{U}_2^r$) provide evidence against H_0 in the direction of H_1 (H_2).

In the next section, we provide results that all three processes $L_n^r(t,v)$ (indexed by r) converge weakly to a Gaussian process under the appropriate null hypothesis. We also state results (with proofs given in the Appendix) on the consistency of the proposed tests against their alternatives, and describe a simulation procedure for determining the critical values of the $\hat{U}_j^r$.

3. Large-Sample Results

We present the asymptotic results for the nonparametric tests of H_0. Parallel results for the semiparametric tests of H_0 and the tests of H_0^0 follow by similar but simplified arguments; these results are briefly stated at the end of this section.

We begin by defining notation that is used in the sequel. Let $\gamma_k(t,v) = P(X_k \le t, \delta_k = 1, V_k \le v), k = 1,2$. 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 $S_{T_k}(t) = P(T_k \ge t)$ 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 $D(I)$ the Skorohod space of the functions on a K-dimensional rectangle I that are continuous from above with limits from below (Bickel and Wichura, 1971), and $C(I)$ be the subspace of 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

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

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

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

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 *Assume the following conditions: The weight process $H_n(t)$ is left continuous with total variation over $t \in [0, \tau + \delta]$ bounded in probability, for $\tau + \delta < \tilde{\tau}$, $\delta > 0$. There exists a uniform continuous function $H(t)$ with bounded variation such that* $\sup_{0 \le t \le \tau+\delta} |H_n(t) - H(t)| \xrightarrow{P} 0$. *$\lambda_k(t)$ for $k = 1, 2$ are continuously differentiable and bounded away from zero on $[0, \tau + \delta]$, $\lambda_2(t, v) > 0$ and $\partial^2 \Lambda_2(t, v)/\partial t^2$ is continuous on $[0, \tau + \delta] \times [0, 1]$. $P(C_k \ge \tau + \delta) > 0$. The symmetric kernel function $K(t)$ has bounded support on $[-1, 1]$ and is twice continuously differentiable. $n_k b_k^3 / \log b_k^{-1} \to \infty$ and $n_k b_k^4 \to 0$ for $k = 1, 2$. Then, under H_0*

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

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 $b_k = n_k^{-\nu}$ for $1/4 < \nu < 1/3$.

Let U_j^r be defined the same as $\hat{U}_j^r$ in (2.6)-(2.7), with $L_n^r(t, v)$ replaced with $L^r(t, v)$. By the continuous mapping theorem, $\hat{U}_j^{np} \xrightarrow{\mathcal{D}} U_j^{np}$ under H_0, so $P(\hat{U}_j^{np} > c_{j\alpha}) \to \alpha$, where $c_{j\alpha}$ is the upper α-quantile of U_j^{np}. However, the $c_{j\alpha}$ are unknown and very difficult to estimate due to the complicated nature of the limit process $L^{np}(t, v)$. In the next section we provide a Monte Carlo procedure to obtain each $c_{j\alpha}$.

Theorem 2 establishes that each $\hat{U}_j^{np}$ is consistent against its alternative.

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 $H(t) > 0$ on $[0, \tau] \times [0, 1]$. Then, $P(\hat{U}_1^{np} > c_{1\alpha}) \to 1$ as $n \to \infty$ under H_1, and $P(\hat{U}_2^{np} > c_{2\alpha}) \to 1$ as $n \to \infty$ under H_2.*

Theorems 1 and 2 also hold for L_n^{sp} and $\hat{U}_j^{sp}, j = 1, 2$, under the same conditions except that the conditions on $\lambda_k(t)$ are replaced by the proportional marginal hazards assumption $\lambda_1(t)/\lambda_2(t) = \exp(\beta)$. Theorem 1 holds for L_n^1 under the same conditions minus any assumptions about $\lambda_k(t)$. We note that the tests $\hat{U}_j^1$ are not consistent tests since they are based on $L_n^1(\tau, v)$ — by integrating over $t \in [0, \tau]$, differences between the two mark-specific hazard functions may cancel in a case that the marginal hazards cross.

Consistent supremum versions of these statistics are easily constructed, however. By accumulating the contrast at the end of follow-up τ, the tests based on $\hat{U}_j^1$ presented here may be more powerful than their supremum counterparts, in cases that the marginal hazards do not strongly cross.

3.2 Gaussian Multipliers Simulation Procedure

We now describe a Gaussian multipliers technique for simulating each of the test processes $L_n^{np}(t,v), L_n^{sp}(t,v)$, and $L_n^1(t,v)$ under the null hypothesis, cf. Lin *et al.* (1993). Note that $\gamma_k(ds,v)/y_k(s) = \int_0^v \lambda_k(s,u)\,duds$. By (A.7) in the Appendix and the continuous mapping theorem, we obtain the result that $\int_a^t y_k^{-1}(s)\,\sqrt{n}_k(N_k(ds,v) - Y_k(s)\Lambda_k(ds,v))$

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

Define the process $\tilde{L}^{np}(t,v)$ by replacing $Z_k(t,v)$, $k=1,2$, in $L^{np}(t,v)$ 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 $\tilde{L}^{np}(t,v) \xrightarrow{\mathcal{D}} L^{np}(t,v)$. 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

$$\begin{aligned}
\tilde{L}^{np}(t,v) &= \sqrt{n_2/n}n_1{}^{-1/2}\sum_{i=1}^{n_1} h_{1i}(t,v) - \sqrt{n_1/n}n_2{}^{-1/2}\sum_{i=1}^{n_2} h_{2i}(t,v); \qquad (3.5)\\
h_{1i}(t,v) &= \int_a^t H(s)y_1^{-1}(s)\,(N_{1i}(ds,v) - Y_{1i}(s)\Lambda_1(ds,v))\\
&\quad - \int_a^t H(s)a(s)\Lambda'_{2s}(s,v)y_1^{-1}(s)\,(N_{1i}(ds,1) - Y_{1i}(s)\Lambda_1(ds,1))\\
h_{2i}(t,v) &= \int_a^t H(s)r(s)y_2^{-1}(s)\,(N_{2i}(ds,v) - Y_{2i}(s)\Lambda_2(ds,v))\\
&\quad - \int_a^t H(s)b(s)\Lambda'_{2s}(s,v)y_2^{-1}(s)\,(N_{2i}(ds,1) - Y_{2i}(s)\Lambda_2(ds,1))
\end{aligned}$$

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 $Y_k(s)/n_k$, $a(s)$ 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

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

We show that the conditional weak limit of the process $L_n^{np*}(t,v)$ given the observed data is the same as the weak limit of $L_n^{np}(t,v)$ 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 $L_n^{np*}(t,v)$ converges to the covariance of $L^{np}(t,v)$. It is left to show that the process $L_n^{np*}(t,v)$ is tight (see Appendix).

Theorem 3 *Under the conditions of Theorem 1, conditional on the observed data,* $L_n^{np*}(t,v)\xrightarrow{\mathcal{D}}$
$L^{np}(t,v)$ *in* $D([a,\tau]\times[0,1])$ *under* H_0 *as* $n\to\infty$, *where* $L^{np}(t,v)$ *is given in (3.2).*

Theorem 3 also holds for the semiparametric tests of H_0, using the following modified test processes. By the proof of Proposition 1, under H_0

$$\begin{aligned} L_n^{sp}(t,v) &= \sqrt{\frac{n_2}{n}}\int_0^t H_n(s)\hat{Z}_1(ds,v)-\sqrt{\frac{n_1}{n}}\int_0^t H_n(s)\exp(\hat{\beta})\hat{Z}_2(ds,v) \\ &\quad -\sqrt{\frac{n_1n_2}{n}}\int_0^t H_n(s)[\exp(\hat{\beta})-\exp(\beta)]\Lambda_2(ds,v). \end{aligned} \tag{3.7}$$

Let $U_n(\beta)$ and $J_n(\beta)$ be the score function and information matrix under the proportional marginal hazards model. It is easy to obtain that

$$\begin{aligned} U_n(\beta) &= \sum_{i=1}^{n_1}\int_0^\tau \frac{\sum_{j=1}^{n_2}Y_{2j}(s)}{\sum_{j=1}^{n_1}Y_{1j}(s)\exp(\beta)+\sum_{j=1}^{n_2}Y_{2j}(s)}N_{1i}(ds,1) \\ &\quad -\sum_{i=1}^{n_2}\int_0^\tau \frac{\sum_{j=1}^{n_1}Y_{1j}(s)\exp(\beta)}{\sum_{j=1}^{n_1}Y_{1j}(s)\exp(\beta)+\sum_{j=1}^{n_2}Y_{2j}(s)}N_{2i}(ds,1) \\ &\equiv \sum_{i=1}^{n_1}U_{1i}(\beta)-\sum_{i=1}^{n_2}U_{2i}(\beta); \\ J_n(\beta) &= \sum_{k=1}^{2}\sum_{i=1}^{n_k}\int_0^\tau \frac{\sum_{j=1}^{n_1}Y_{1j}(s)\sum_{j=1}^{n_2}Y_{2j}(s)\exp(\beta)}{(\sum_{j=1}^{n_1}Y_{1j}(s)\exp(\beta)+\sum_{j=1}^{n_2}Y_{2j}(s))^2}N_{ki}(ds,1). \end{aligned}$$

A routine delta method and likelihood analysis yields $n^{1/2}(\exp(\hat{\beta})-\exp(\beta))=\exp(\beta)\times(n^{-1}J_n(\beta))^{-1}n^{-1/2}U_n(\beta)+o_p(1)$. From this result and (3.7), following the arguments of Section , the distribution of $L_n^{sp}(t,v)$ under H_0 can be approximated by the following process $L_n^{sp*}(t,v)$ given the observed data,

$$L_n^{sp*}(t,v) = \sqrt{\frac{n_2}{n}}{n_1}^{-1/2}\sum_{i=1}^{n_1}\Big[\int_0^t H_n(s)(n_1Y_1^{-1}(s))\left(N_{1i}(ds,v)-Y_{1i}(s)\Lambda_1(ds,v)\right)$$

$$-\frac{n_1}{n}\exp(\hat{\beta})(n^{-1}J_n(\hat{\beta}))^{-1}\hat{U}_{1i}(\hat{\beta})\int_0^t H_n(s)\hat{\Lambda}_2(ds,v)\Big]W_{1i}$$
$$-\sqrt{\frac{n_1}{n}}{n_2}^{-1/2}\sum_{i=1}^{n_2}\Big[\int_0^t H_n(s)\exp(\hat{\beta})(n_2Y_2^{-1}(s))\,(N_{2i}(ds,v)-Y_{2i}(s)\Lambda_2(ds,v))$$
$$-\frac{n_2}{n}\exp(\hat{\beta})(n^{-1}J_n(\hat{\beta}))^{-1}\hat{U}_{2i}(\hat{\beta})\int_0^t H_n(s)\hat{\Lambda}_2(ds,v)\Big]W_{2i},$$

where $\hat{U}_{1i}(\hat{\beta})$ and $\hat{U}_{2i}(\hat{\beta})$ are obtained from $U_{1i}(\beta)$ and $U_{2i}(\beta)$, respectively, with β, $\Lambda_1(ds,1)$ and $\Lambda_2(ds,1)$ replaced by $\hat{\beta}$, $\hat{\Lambda}_1(ds,1)$ and $\hat{\Lambda}_2(ds,1)$, respectively.

Similarly, the distribution of $L_n^1(t,v)$ under H_0^0 can be approximated by $L_n^{1*}(t,v)$ given the observed data, where

$$L_n^{1*}(t,v)=\sqrt{\frac{n_1}{n}}{n_2}^{-1/2}\sum_{i=1}^{n_2}\int_0^t H_n(s)(n_2Y_2^{-1}(s))\,(N_{2i}(ds,v)-Y_{2i}(s)\Lambda_2(ds,v))W_{2i}$$
$$-\sqrt{\frac{n_2}{n}}{n_1}^{-1/2}\sum_{i=1}^{n_1}\int_0^t H_n(s)(n_1Y_1^{-1}(s))\,(N_{1i}(ds,v)-Y_{1i}(s)\Lambda_1(ds,v))W_{1i}.$$

3.3 Choice of Weight Process

In exploratory work it can be useful to examine a plot of the test process $L_n^r(t,v)$ 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 $L_n^{r*}(t,v)$. Large values of $|L^1(t,v)|$ for some v and t suggest a departure from H_0^0. Large values of $L_n^{np}(t_1,v)-L_n^{np}(t_2,v)$ for some v and some $t_1<t_2$, as compared with the same contrast in $L_n^{np*}(t,v)$, suggest a departure from H_0 in the direction of H_1. Large absolute differences in $L_n^{np}(t,v)$ across different marks v (as compared with the reference process) would suggest H_2. This graphical procedure is illustrated in Section 6.

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, and we suggest

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

The weight can also be chosen to increase power against specific alternatives (Sun, 2001). Furthermore the test can be made invariant to the order of the two groups by including $\hat{r}(s)^{-1/2}$ in $H_n(s)$.

4. Estimation of Mark-Specific Vaccine Efficacy

Precise estimation of $\mathrm{VE}(t,v)$ introduced in Section 1 requires huge sample sizes, because smoothing is required in both v and t, and generally efficacy trials do not provide sufficient samples (Gilbert *et al.*, 2002). Accordingly, we consider an alternative notion of mark-specific vaccine efficacy defined 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 $\mathrm{VE}^c(t,v)$.

A nonparametric estimator of $\mathrm{VE}^c(t,v)$ is given by $\widehat{\mathrm{VE}}^c(t,v) = 1 - \hat{F}_1(t,v)/\hat{F}_2(t,v)$, where

$$\hat{F}_k(t,v) = \frac{1}{b_k}\int_0^1\int_0^t \frac{\hat{S}_k(s-)}{Y_k(s)} K\left(\frac{v-u}{b_k}\right) N_k(ds,du), \tag{4.1}$$

$\hat{S}_k(t)$ 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 $b_k > 0$ 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 (Prentice *et al.*, 1978).

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 $\log(F_1(t,v)/F_2(t,v))$:

$$1 - \left(1 - \widehat{\mathrm{VE}}^c(t,v)\right) \exp\left(\pm z_{\alpha/2}\sqrt{\frac{\widehat{\mathrm{Var}}\{\hat{F}_1(t,v)\}}{\hat{F}_1(t,v)^2} + \frac{\widehat{\mathrm{Var}}\{\hat{F}_2(t,v)\}}{\hat{F}_2(t,v)^2}}\right); \tag{4.2}$$

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

To estimate $\mathrm{VE}^{dc}(t,v)$, each $P(T_k \le t, V_k \le v)$ is simply estimated by $\int_0^t \left(\hat{S}_k(s-)/Y_k(s)\right) N_k(ds,v)$, the standard estimator for the discrete cumulative incidence function for cause of failure defined by $V \le v$, and its variance is estimated by $\int_0^t \left(\hat{S}_k(s-)/Y_k(s)\right)^2 N_k(ds,v)$.

5. Simulation Experiment

The simulations are based on the features of the VaxGen trial described in the Introduction. We study performance of the test statistics $\hat{U}_j^1, j = 1,2,3,4; \hat{U}_j^{np}$ and $\hat{U}_j^{sp}$, $j = 1,2$; and of $\widehat{\mathrm{VE}}^c(\tau, v)$, with $\tau = 3$ years. For $\mathrm{VE}^c(\tau, v)$ we focus on the end of follow-up $t = \tau$ because it is most important scientifically to understand durability of vaccine efficacy.

The functions $\mathrm{VE}(t, v)$ and $\mathrm{VE}^c(t, v)$ can be complicated functions of v even when T_k and V_k are independent. Although independence may be unlikely in applications, to keep the simulation model simple we focus on this case. Limited simulations under dependence showed comparable performance of the procedures. The cumulative incidence function for group k is taken as $F_k(t, v) = P\{T_k \leq t\} f_{Vk}(v)$, where f_{Vk} is the density of V_k. In the first set of simulations 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. In the second set of simulations we specify non-proportional hazards, to examine the effect of violating the assumption used by the semiparametric tests. In this case V_k and T_2 are distributed the same as above, and T_1 is set as $T_1 = \sqrt{X_1}$, where X_1 is exponential with parameter λ_1, implying $\lambda_1(t) = 2\lambda_1 t$.

We consider two true values of $\mathrm{VE}^c(\tau)$, 0.67 and 0.33. To evaluate the size of the tests of H_0^0 we also consider $\mathrm{VE}^c(\tau) = 0.0$. We select λ_2 so that 50% of placebo recipients are expected to be infected by $\tau = 36$ months. In addition we generate 10% random drop-out before 36 months.

Next, we specify

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

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 increasing levels of dependence between $\lambda_k(t, v)$ and v, with

$E(V_k) = 2/3$ and $4/5$, respectively. The cumulative vaccine efficacy is given by

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

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 H_0. 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), (0.25, 1.0), (0.50, 0.50), or (0.25, 0.25). We also consider a two-sided alternative with $f_{V2}(v)$ a uniform density and $f_{V1}(v) = \frac{16}{3} v I(v < \frac{1}{2}) + (\frac{8}{3} - \frac{8}{3} v) I(v \geq \frac{1}{2})$. Results for this case are given under the heading "2-sided" in Tables 1 and 2.

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 of $\hat{\lambda}_k$ (Andersen *et al.*, 1993, p. 240) separately for $k = 1, 2$, and the method of Gasser and Müller (1979) is used to correct for bias in the tails. An alternative approach would jointly optimize (b_1, b_2) for estimating the hazard ratio. Based on Kelsall and Diggle (1995), 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 some efficacy (i.e., the hazards are unequal). For this reason we optimized b_k for 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_{vk} were chosen to achieve reasonably smooth estimates, which was accomplished with $b_{v1} = b_{v2} = 0.20$. Bias, coverage probability of the 95% confidence intervals (4.2), 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 $n = 100, 200$ or 400 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 Table 1 indicate that the tests of H_0^0 have appropriate sizes and high powers. When $\mathrm{VE}(t,v)$ declines with v, they have greater power than the Cox model Wald test of $\mathrm{VE}(t)=0$. Therefore accounting for the mark variable can substantially improve efficiency. This is especially the case for $\hat{U}_2^1$, although this test has less power than the Cox model test if $\mathrm{VE}(t,v)$ is constant in v (i.e., $\beta_1=\beta_2$). In contrast, the power of $\hat{U}_1^1$ is less sensitive to how strongly $\mathrm{VE}(t,v)$ varies in v. The corresponding 2-sided tests $\hat{U}_3^1$ and $\hat{U}_4^1$ show a similar comparative pattern but with lower power for the one-sided alternatives.

The results in Table 2 show that the tests of H_0 perform well at moderate sample sizes. Somewhat surprisingly, for small/moderate samples the semiparametric tests did not provide greater power than the nonparametric tests in the case that the failure times had proportional hazards. To explain this, note that the nonparametric and semiparametric test processes involve contrasts $\hat{\Lambda}_1(dt,v)-\hat{r}(t)\hat{\Lambda}_2(dt,v)$, with $\hat{r}(t)=\hat{\lambda}_1(t)/\hat{\lambda}_2(t)$ and $\exp(\hat{\beta})$, respectively, and the alternative hypothesis involves changes of $\lambda_1(t,v)/\lambda_2(t,v)$ in v— but *not* in t. Since $\hat{\Lambda}_k(dt,v)$ and $\hat{\lambda}_k(t)$ approximately "track" each other in t, the nonparametric test process can reduce the noise from perturbations of $\hat{\lambda}_1(t)/\hat{\lambda}_2(t)$ in t, whereas the semiparametric test process cannot dampen this noise.

The small simulation study under non-proportional hazards, with H_0 true with $(\beta_1,\beta_2)=(1.0,1.0)$, $(0.5,0.5)$, or $(0.25,0.25)$, demonstrates (as predicted from the theory) that the semiparametric tests are not valid when the marginal proportional hazards condition is not met. The empirical sizes of the tests frequently missed 0.05 by an amount more than 2 or 3 Monte Carlo standard deviations (results not shown).

The results in Table 3 show that for moderate samples, $\widehat{\mathrm{VE}}^c(36,v)$ is unbiased at some parameter configurations and biased at others, and that the bias becomes negligible as the number of infections grows large. 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.

6. Application

We apply the methods to the data from the VaxGen trial described in the Introduction. Figure 1 shows boxplots of the three percent amino acid mismatch distances of the infecting HIV viruses to the nearest virus (MN or GNE8) represented in the tested vaccine. 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 b_k for the estimated hazards of infection $\hat{\lambda}_1(\cdot)$ and $\hat{\lambda}_2(\cdot)$ were $b_1 = 1.83$ months and $b_2 = 2.10$ months. For the neutralizing face core distances, the four tests of $H_0^0 : \mathrm{VE}(t, v) = 0$ gave p-values spanning 0.05 to 0.32 (Figure 1(d)), with $\hat{U}_2^1$ rejecting H_0^0 at level 0.05. Based on this evidence (albeit weak) that $\mathrm{VE}(t, v) \neq 0$, we go on to test $H_0 : \mathrm{VE}(t, v) = \mathrm{VE}(t)$. Neither nonparametric test rejected H_0 (Figure 1(d)). The proportional hazards assumption seemed reasonable based on a goodness-of-fit test ($p = 0.35$), justifying the semiparametric tests of H_0, which gave nonsignificant results (Figure 1(d)). To illustrate the graphical procedure, Figure 2 shows the test process $L_n^{np}(t, v)$ together with 8 randomly selected realization of the null test process $L_n^{np*}(t, v)$, using a unit weight process $H_n(\cdot) = 1$. The maximum absolute deviation of $L_n^{np}(t, v)$ in t is larger than that for all but one of the null test processes. Figure 1 panels (e)-(f) show p-values of the tests for the other two distances, which all exceeded 0.05.

With bandwidths b_{v1} and b_{v2} separately optimized using 2-fold cross-validation, we next estimated $\mathrm{VE}^c(36, v)$ and $\mathrm{VE}^{dc}(36, v)$ (Figure 3). 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 close-

ness 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. The borderline significant result is intriguing, because 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).

7. Concluding Remarks

Nonparametric and semiparametric methods have been developed for testing and estimation of relative risks taking into account a continuous mark variable observed only at uncensored failure times, and for evaluating the relationship between the relative risk and the mark. We showed that if the mark-specific relative risk varies with the mark, then a standard Cox model test of a unit hazard ratio (ignoring the mark) is less powerful (and sometimes much less) than the newly developed nonparametric procedures that test the null $H_0^0 : \lambda_1(t,v)/\lambda_2(t,v) = 1$ of a unit mark-specific hazard ratio. This finding raises the novel idea to consider accounting for the mark variable in secondary hypothesis tests in clinical trials for which there are strong reasons to suspect that the mark-specific relative risk is monotone in the mark. Among the statistics developed for testing H_0^0, $\hat{U}_1^1$ or $\hat{U}_2^1$ are recommended, with the choice between them depending on how strongly the mark-specific relative risk varies with the mark in the alternative hypothesis of interest.

For testing dependency of the mark-specific relative risk on the mark, $H_0 : \lambda_1(t,v)/\lambda_2(t,v) = \lambda_1(t)/\lambda_2(t)$, the simulations suggest that the nonparametric procedures perform better than their semiparametric counterparts that assume proportional marginal hazards. The test based on $\hat{U}_1^{np}$ is recommended.

Although the methods were motivated by a particular scientific problem (the question in HIV vaccine efficacy trials of if and how efficacy of the tested vaccine varies with the genetic distance of the infecting HIV strain), we emphasize that they provide a general solution to the two-sample survival analysis problem with a continuous mark variable, which may have many applications. An appeal of the procedures devel-

oped here 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.

Acknowledgements

The authors gratefully acknowledge David Jobes and VaxGen Inc. for providing the HIV sequence data, and thank Per Andersen for helpful suggestions. This research was partially supported by NIH grant 2 R01 AI54165-04 (Gilbert), NSF grant DMS-0505201 (McKeague), and NSF grant DMS-0304922 and DMS-0604576 (Sun). Part of the research for this paper was done while the second author was visiting the Institute for Mathematical Sciences, National University of Singapore in 2005/2006. The visit was supported by the Institute.

Appendix: Proofs of Theorems

Proofs of the results in Section 3.2 are presented for the nonparametric tests of H_0, involving $L_n^{np}(t,v)$ and $\hat{U}_j^{np}$ with $r(t) = \lambda_1(t)/\lambda_2(t)$ and $\hat{r}(t) = \hat{\lambda}_1(t)/\hat{\lambda}_2(t)$. The proofs are similar and simpler for the other tests and are omitted.

The following results summarized in Lemma 1 and Lemma 2 are to be used in the proof of Proposition 1. Lemma 1 applies to either of the kernel estimators $\hat{\lambda}_k(t)$ defined in (2.2), $k = 1, 2$. For simplicity, we state the lemma for a single estimator of this form, with the subscript k suppressed throughout.

Lemma 1 *Assume that $\lambda(t)$ is continuously differentiable and bounded away from zero on $t \in [0, \tau+\delta]$, $P(C \geq \tau+\delta) > 0$ and that the symmetric kernel function $K(t)$ has support on $[-1,1]$ and is twice continuously differentiable. Let $a > 0$ and $b \to 0$.*
(a) Suppose $nb^3/\log b^{-1} \to \infty$. Then the total variation of $\hat{\lambda}(t)$ is bounded in probability on $[a, \tau]$.
(b) Suppose $nb/\log b^{-1} \to \infty$. Then $\sup_{a \leq t \leq \tau} |\hat{\lambda}(t) - \lambda(t)| \xrightarrow{P} 0$.

Proof of Lemma 1.

Let $\lambda^*(t) = \int_0^{\tau+\delta} b^{-1}K((t-s)/b)\lambda(s)P(Y(s) \neq 0)\,ds$. Then there exists an n_0 such that $\lambda^*(t) = \int_{-1}^{1} K(x)\lambda(t-bx)P(Y(t-bx) \neq 0)\,dx$, for $t \in [a,\tau]$, $n \geq n_0$. Since $Y(\cdot)$ is a non-increasing risk process and $\lambda(t)$ has total variation bounded in probability, it is easy to see that $\lambda^*(t)$ has bounded variation on $[a,\tau]$ uniformly in n. Note that $\lambda^*(t) = E\hat{\lambda}(t)$ and decompose

$$\hat{\lambda}(t) - \lambda(t) = \hat{\lambda}(t) - \lambda^*(t) + \lambda^*(t) - \lambda(t).$$

To prove part (a), it is sufficient to show that $\hat{\lambda}(t) - \lambda^*(t)$ has total variation bounded in probability. By the mean value theorem, it suffices to show that the derivative

$$v_n(t) = \frac{d}{dt}(\hat{\lambda}(t) - \lambda^*(t)) = \int_0^{\tau+\delta} b^{-2}K'((t-s)/b)\,(d\hat{\Lambda}(s) - \lambda(s)P(Y(s) \neq 0)ds)$$

is uniformly bounded in probability over $[a,\tau]$, where $K'(x)$ is the derivative of $K(x)$.

Let $V(t) = \int_{-1}^{1}(K'(s))^2\,ds\lambda(t)/y(t)$, where $y(t) = EY(t)/n$. Let $W_n(t) = (nb/V(t))^{1/2}$ $bv_n(t)$ and $M_n = \sup_{0\leq t\leq\tau}|W_n(t)|$. We also define $r_n = (2\log(\tau_n/b))^{1/2}$ and $d_n = r_n + (\log w^*)/r_n$, where $w^* = (2\pi)^{-1}[\int_{-1}^{1}(K''(x))^2\,dx/\int_{-1}^{1}(K'(x))^2\,dx]^{1/2}$ and τ_n is the highest order statistic in $[0,\tau]$. Thus $\tau_n \xrightarrow{P} \tau$. Following the proofs of Proposition 3.1, Proposition 3.2 and Theorem 4.B of Yandell (1983), it can be shown that for all x,

$$P\{r_n(M_n - d_n) < x\} \to \exp(-2e^{-x}),$$

as $nb \to \infty$. The only thing that we need to note here is that we can treat $K'(t)$ as the weight function w of Yandell (1983). The conditions $K'(t) \geq 0$ and $\int_{-1}^{1} K'(s)\,ds = 1$ are not needed in Yandell's proofs of Proposition 3.1, Proposition 3.2 and Theorem 4.B. Therefore $M_n = O_p(d_n) = O_p((\log b^{-1})^{1/2})$ and $v_n(t) = O_p((\log b^{-1}/nb^3)^{1/2})$ uniformly in $t \in [0,\tau]$. This proves part (a).

Part (b) follows directly from Proposition 3.1, Proposition 3.2 and Theorem 4.B of Yandell (1983). □

Lemma 2 *Let $I = [a,b]\times I_0 \subset R^K$ where I_0 is a $K-1$-dimensional rectangle. Let $D(I)$ be the Skorohod space of the functions on I that are continuous from above with limits*

from below. Assume that the random processes $g_n(\cdot,\cdot)$ *and* $g(\cdot,\cdot)$ *have sample paths in* $D(I)$, $\sup_{(t,x)\in I}|g_n(t,x)-g(t,x)|\xrightarrow{P}0$, *and* $g_n(\cdot,x)$ *and* $g(\cdot,x)$ *have total variations bounded in probability, uniformly in* x. *Assume that the processes* $Q_n(t,x)\xrightarrow{\mathcal{D}}Q(t,x)$ *in* $D(I)$. *Then*

$$\sup_{(t,x)\in I}|\int_a^t Q_n(s,x)\,g_n(ds,x)-\int_a^t Q_n(s,x)\,g(ds,x)|\xrightarrow{P}0 \tag{A.1}$$

$$\sup_{(t,x)\in I}|\int_a^t g_n(s-,x)\,Q_n(ds,x)-\int_a^t g(s-,x)\,Q_n(ds,x)|\xrightarrow{P}0, \tag{A.2}$$

where $g(s-,\cdot)=\lim_{u\to s,u<s}g(u,\cdot)$ *and* $g_n(s-,\cdot)$ *is defined similarly.*

Proof of Lemma 2.

Since $Q_n(t,x)\xrightarrow{\mathcal{D}}Q(t,x)$ in $D(I)$, $Q_n(t,x)$ is tight by Bickel and Wichura (Theorem 2, 1971). It follows that for any $\epsilon>0$, $\eta>0$, there exists a partition $a=t_1<t_1<\ldots<t_{m+1}=b$ and n_0 such that

$$P(\max_{1\le j\le m}\sup_{x\in I_0}\sup_{t\in[t_j,t_{j+1})}|Q_n(t,x)-Q_n(t_j,x)|>\epsilon)<\eta,\quad \text{for } n\ge n_0. \tag{A.3}$$

Also, since $Q_n(t,x)$ converges weakly and $g_n(\cdot,x)$ and $g(\cdot,x)$ have bounded variations in probability uniformly in x, for any $\eta>0$, there exists $B>0$ and $n_1\ge n_0$ such that, for $n\ge n_1$,

$$P(\sup_{(t,x)\in I}|Q_n(t,x)|>B)<\eta \tag{A.4}$$

$$P\Big(\sup_{x\in I_0}\Big[\int_a^b|g_n(ds,x)|+\int_a^b|g(ds,x)|\Big]>B\Big)<\eta. \tag{A.5}$$

For each m, define $Q_n^m(t,x)=Q_n(t_j,x)$ for $(t,x)\in[t_j,t_{j+1})\times I_0$, $j=1,\ldots,m$. Let $C_n=\sup_{(t,x)\in I}|Q_n(t,x)-Q_n^m(t,x)|$. Then $C_n=\max_{1\le j\le m}\sup_{x\in I_0}\sup_{t\in[t_j,t_{j+1})}|Q_n(t,x)-Q_n(t_j,x)|$. Note that

$$\begin{aligned}
&\sup_{(t,x)\in I}\Big|\int_a^t Q_n(s,x)\,g_n(ds,x)-\int_a^t Q_n(s,x)\,g(ds,x)\Big|\\
&\le\sup_{(t,x)\in I}\Big|\int_a^t(Q_n(s,x)-Q_n^m(s,x))\,g_n(ds,x)\Big|+\sup_{(t,x)\in I}\Big|\int_a^t(Q_n(s,x)-Q_n^m(s,x))\,g(ds,x)\Big|\\
&\quad+\sup_{(t,x)\in I}\Big|\int_a^t Q_n^m(s,x)\,(g_n(ds,x)-g(ds,x))\Big|\\
&\le C_n\Big(\sup_{x\in I_0}\int_a^b|g_n(ds,x)|+\sup_{x\in I_0}\int_a^b|g(ds,x)|\Big)+2m\sup_{(t,x)\in I}|Q_n(t,x)|\sup_{(t,x)\in I}|g_n(t,x)-g(t,x)|.
\end{aligned}$$

Thus, for any $\delta > 0$ and $\eta > 0$, choose B as in (A.4) and (A.5). Take ϵ in (A.3) to be $\delta/(2B)$ and n_0 corresponding to the ϵ. We have, for $n \geq n_1$,

$$
\begin{aligned}
&P\Big\{ \sup_{(t,x)\in I} \Big| \int_a^t Q_n(s,x)\, g_n(ds,x) - \int_a^t Q_n(s,x)\, g(ds,x) \Big| > \delta \Big\} \\
&\leq P\Big\{ C_n \Big(\sup_{x\in I_0} \int_a^b |g_n(ds,x)| + \sup_{x\in I_0} \int_a^b |g(ds,x)| \Big) > \delta/2 \Big\} \\
&\quad + P\Big\{ 2m \sup_{(t,x)\in I} |Q_n(t,x)| \sup_{(t,x)\in I} |g_n(t,x) - g(t,x)| > \delta/2 \Big\} \\
&\leq P\Big\{ \sup_{x\in I_0} \int_a^b |g_n(ds,x)| + \sup_{x\in I_0} \int_a^b |g(ds,x)| > B \Big\} + P\{C_n > \delta/(2B)\} \\
&\quad + P\Big\{ \sup_{(t,x)\in I} |Q_n(t,x)| > B\} + P\{ \sup_{(t,x)\in I} |g_n(t,x) - g(t,x)| > \delta/(4mB) \Big\} \\
&\leq 3\eta + P\Big\{ \sup_{(t,x)\in I} |g_n(t,x) - g(t,x)| > \delta/(4mB) \Big\}.
\end{aligned}
$$

Since $\sup_{(t,x)\in I} |g_n(t,x) - g(t,x)| \xrightarrow{P} 0$, we have

$$
\limsup_{n\to\infty} P\Big\{ \sup_{(t,x)\in I} \Big| \int_a^t Q_n(s,x)\, g_n(ds,x) - \int_a^t Q_n(s,x)\, g(ds,x) \Big| > \delta \Big\} \leq 3\eta.
$$

Since η is arbitrary, this proves (A.1). The proof of (A.2) follows by integration by parts.

Proposition 1 *Given the conditions of Theorem 1,*

$$
L_n^{np}(t,v) - \sqrt{\frac{n_1 n_2}{n}} \int_a^t H_n(s)[\Lambda_1(ds,v) - r(s)\Lambda_2(ds,v)] \xrightarrow{\mathcal{D}} L^{np}(t,v) \tag{A.6}
$$

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

Proof of Proposition 1.

Using the central limit theorem for empirical processes (cf. Gilbert *et al.*, 2004, (A.4)),

$$
\sqrt{n}_k(N_k(t,v)/n_k - \gamma_k(t,v), Y_k(t)/n_k - y_k(t)) \xrightarrow{\mathcal{D}} (G_1^{(k)}(t,v), G_2^{(k)}(t)) \tag{A.7}
$$

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

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

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 and Kosorok (2001), we have

$$\hat{Z}_k(t,v) \xrightarrow{\mathcal{D}} Z_k(t,v) \tag{A.8}$$

in $D([0,\tau]\times[0,1])$, where the two processes $Z_1(t,v)$ and $Z_2(t,v)$ are independent.

The test process can be decomposed as follows:

$$\begin{aligned}
&L_n^{np}(t,v) = \sqrt{\frac{n_1n_2}{n}}\int_a^t H_n(s)[\hat{\Lambda}_1(ds,v) - \Lambda_1(ds,v)] \\
&-\sqrt{\frac{n_1n_2}{n}}\int_a^t H_n(s)\hat{r}(s)[\hat{\Lambda}_2(ds,v) - \Lambda_2(ds,v)] + \sqrt{\frac{n_1n_2}{n}}\int_a^t H_n(s)[\Lambda_1(ds,v) - \hat{r}(s)\Lambda_2(ds,v)] \\
&= \sqrt{\frac{n_2}{n}}\int_a^t H_n(s)\hat{Z}_1(ds,v) - \sqrt{\frac{n_1}{n}}\int_a^t H_n(s)\hat{r}(s)\hat{Z}_2(ds,v) \\
&+\sqrt{\frac{n_1n_2}{n}}\int_a^t H_n(s)[r(s)-\hat{r}(s)]\Lambda_2(ds,v) + \sqrt{\frac{n_1n_2}{n}}\int_a^t H_n(s)[\Lambda_1(ds,v) - r(s)\Lambda_2(ds,v)].
\end{aligned}$$

Under H_0, the last term equals zero. By Lemma 1, $\sup_{a\le s\le\tau}|\hat{r}(s)-r(s)|\xrightarrow{P}0$ and $\hat{r}(s)$ has bounded variation in probability on $[a,\tau]$. By Lemma 2, the continuity of $\hat{\lambda}_k(t)$ and the conditions on $H_n(t)$, we have

$$\begin{aligned}
&L_n^{np}(t,v) = \sqrt{\frac{n_2}{n}}\int_a^t H(s)\hat{Z}_1(ds,v) - \sqrt{\frac{n_1}{n}}\int_a^t H(s)r(s)\hat{Z}_2(ds,v) \qquad \text{(A.9)} \\
&+\sqrt{\frac{n_1n_2}{n}}\int_a^t H_n(s)[r(s)-\hat{r}(s)]\Lambda_2(ds,v) + \sqrt{\frac{n_1n_2}{n}}\int_a^t H(s)[\Lambda_1(ds,v) - r(s)\Lambda_2(ds,v)] + o_p(1).
\end{aligned}$$

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 (A.9) equals

$$\sqrt{\frac{n_1n_2}{n}}\int_a^t H_n(s)[-\hat{a}(s)(\hat{\lambda}_1(s)-\lambda_1(s)) + \hat{b}(s)(\hat{\lambda}_2(s)-\lambda_2(s))]\Lambda_2(ds,v). \tag{A.10}$$

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

$$\frac{1}{b_k}\int_0^{\tau+\delta} K\left(\frac{t-s}{b_k}\right)d\Lambda_k(s) = \lambda_k(t) + O(b_k^2),$$

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 $[-1,1]$,

$$\sqrt{\frac{n_1n_2}{n}}\int_a^t H_n(s)\hat{a}(s)(\hat{\lambda}_1(s)-\lambda_1(s))\Lambda_2(ds,v) \tag{A.11}$$

$$= \sqrt{\frac{n_1 n_2}{n}} \int_0^{\tau+\delta} \left[\int_a^t \frac{1}{b_1} K\left(\frac{s-u}{b_1}\right) H_n(s)\hat{a}(s)\Lambda_2(ds,v) \right] d(\hat{\Lambda}_1(u) - \Lambda_1(u)) + O(\sqrt{n}b_1^2)$$
$$= \sqrt{\frac{n_1 n_2}{n}} \int_{a-b_1}^{t-b_1} \left[\int_a^t \frac{1}{b_1} K\left(\frac{s-u}{b_1}\right) H_n(s)\hat{a}(s)\Lambda_2(ds,v) \right] d(\hat{\Lambda}_1(u) - \Lambda_1(u))$$
$$+ \sqrt{\frac{n_1 n_2}{n}} \int_{t-b_1}^{t+b_1} \left[\int_a^t \frac{1}{b_1} K\left(\frac{s-u}{b_1}\right) H_n(s)\hat{a}(s)\Lambda_2(ds,v) \right] d(\hat{\Lambda}_1(u) - \Lambda_1(u)) + O(\sqrt{n}b_1^2).$$

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 $\Lambda'_{2u}(u,v) = \partial\Lambda_2(u,v)/\partial u$. By Lemma 1, $\hat{a}(s)$ has bounded variation in probability over $[a,\tau]$. Since $H_n(s)$ has uniform bounded variation as well, we have that the process $\int_a^t b_1^{-1} K((s-u)/b_1) H_n(s)\hat{a}(s)\Lambda'_{2s}(s,v)\,ds$ is of bounded variation in u in probability uniformly in $v \in [0,1]$ and $t \in [0,\tau]$, and $H(u)a(u)\Lambda'_{2u}(u,v)$ is of bounded variation uniformly in $v \in [0,1]$. Since $\sqrt{n}_1(\hat{\Lambda}_1(u) - \Lambda_1(u))$ converges weakly, it follows from Lemma 2 that (A.11) equals

$$\sqrt{\frac{n_1 n_2}{n}} \int_{a-b_1}^{t-b_1} H(u)a(u)\Lambda'_{2u}(u,v)\, d(\hat{\Lambda}_1(u) - \Lambda_1(u)) + O(\sqrt{n}b_1^2) + O(b_1)$$
$$= \sqrt{\frac{n_2}{n}} \int_a^t H(s)a(s)\Lambda'_{2s}(s,v)\, \hat{Z}_1(ds,1) + O(\sqrt{n}b_1^2) + o_p(1). \tag{A.12}$$

Similarly,

$$\sqrt{\frac{n_1 n_2}{n}} \int_a^t H_n(s)\hat{b}(s)(\hat{\lambda}_2(s) - \lambda_2(s))\Lambda_2(ds,v)$$
$$= \sqrt{\frac{n_1}{n}} \int_a^t H(s)b(s)\Lambda'_{2s}(s,v)\, d\hat{Z}_2(ds,1) + O(\sqrt{n}b_2^2) + o_p(1). \tag{A.13}$$

By (A.9)–(A.13), under $\sqrt{n}b_k^2 \to 0$, as $n \to \infty$ for $k = 1, 2$, we have

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

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

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

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,

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

we have

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

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 $t\in[0,\tau]$ there exists a $v_t\in[0,1]$ such that

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

We choose v_t to be the smallest v satisfying this equality. It follows that v_t is a continuous function of t and

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

Note that the inequality under H_1 is strict for some (t,v), $\lambda_k(t)=\int_0^1\lambda_k(t,v)\,dv$ and the functions $\lambda_1(t,v)$ and $\lambda_2(t,v)$ are continuous. There exists an open neighborhood of t such that $0<v_t<1$. Let $a>0$ and $s_1<s_2$ be such that $v_t-a, v_t+a\in(0,1)$ for $t\in[s_1,s_2]$. Then

$$\int_{v_t}^{v_t+a}H(t)(\lambda_1(t,v)-r(t)\lambda_2(t,v))\,dv-\int_{v_t-a}^{v_t}H(t)(\lambda_1(t,v)-r(t)\lambda_2(t,v))\,dv>0,$$

for $t\in[s_1,s_2]$. Since the integrals above are uniform continuous functions of (t,v_t) and v_t is uniform continuous, there exists a neighborhood $[t_1,t_2]\subset[s_1,s_2]$ and $[v_1,v_2]\subset[0,1]$ such that

$$\int_{t_1}^{t_2}\int_{\frac{v_1+v_2}{2}}^{v_2}H(s)(\lambda_1(s,v)-r(s)\lambda_2(s,v))\,dv\,ds-\int_{t_1}^{t_2}\int_{v_1}^{\frac{v_1+v_2}{2}}H(s)(\lambda_1(s,v)-r(s)\lambda_2(s,v))\,dv\,ds\ge c,$$

where c is a positive constant. Let $\delta_n(t_1,t_2,v_1,v_2)$ be the left side of the above expression with $H(s)$ replaced by $H_n(s)$. Since $H_n(t)\xrightarrow{P}H(t)>0$ uniformly in $t\in[0,\tau]$, we have

$\sqrt{\frac{n_1 n_2}{n}}\delta_n(t_1,t_2,v_1,v_2) \xrightarrow{P} \infty$, as $n \to \infty$. By Proposition 1,

$$
\begin{aligned}
&\Delta_n^r(t_2,v_1,v_2) - \Delta_n^r(t_1,v_1,v_2) - \sqrt{\frac{n_1 n_2}{n}}\delta_n(t_1,t_2,v_1,v_2) \\
&= \left[(L_n^r(t_2,v_2) - L_n^r(t_1,v_2)) - (L_n^r(t_2,\frac{v_1+v_2}{2}) - L_n^r(t_1,\frac{v_1+v_2}{2}))\right] \\
&-\left[(L_n^r(t_2,\frac{v_1+v_2}{2}) - L_n^r(t_1,\frac{v_1+v_2}{2})) - (L_n^r(t_2,v_1) - L_n^r(t_1,v_1))\right] - \sqrt{\frac{n_1 n_2}{n}}\delta_n(t_1,t_2,v_1,v_2) \\
&\xrightarrow{\mathcal{D}} \left[(L^r(t_2,v_2) - L^r(t_1,v_2)) - (L^r(t_2,\frac{v_1+v_2}{2}) - L^r(t_1,\frac{v_1+v_2}{2}))\right] \\
&-\left[(L^r(t_2,\frac{v_1+v_2}{2}) - L^r(t_1,\frac{v_1+v_2}{2})) - (L^r(t_2,v_1) - L^r(t_1,v_1))\right] \qquad \text{(A.14)}
\end{aligned}
$$

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

We note that, under H_2, there exist $[t_1,t_2]$ and $[v_1,v_2]$ such that

$$\left|\int_{t_1}^{t_2}\int_{\frac{v_1+v_2}{2}}^{v_2} H(s)(\lambda_1(s,v) - r(s)\lambda_2(s,v))\,dv\,ds - \int_{t_1}^{t_2}\int_{v_1}^{\frac{v_1+v_2}{2}} H(s)(\lambda_1(s,v) - r(s)\lambda_2(s,v))\,dv\,ds\right| \geq c,$$

where c is a positive constant. Otherwise, $H(s)(\lambda_1(s,v) - r(s)\lambda_2(s,v))$ is a constant function of (s,v), which would be zero since $\lambda_k(t) = \int_0^1 \lambda_k(t,v)\,dv$, $k = 1,2$. Since $H_n(t) \xrightarrow{P} H(t) > 0$ uniformly in $t \in [0,\tau]$, it follows that $\sqrt{\frac{n_1 n_2}{n}}|\delta_n(t_1,t_2,v_1,v_2)| \xrightarrow{P} \infty$, as $n \to \infty$. By (A.14) and Slutsky's Theorem, we have $\hat{U}_2^r \xrightarrow{P} \infty$ as $n \to \infty$. This completes the proof. □

Proof of the tightness for $L_n^{np*}(t,v)$ (remaining piece of Theorem 3 proof).

To show tightness of $L_n^{np*}(t,v)$ 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

$$E\left\{\Delta^2(B)\Big|\{\text{observed data}\}\right\} \le \mu_0(B) + o_p(1) \tag{A.15}$$

$$E\left\{\Delta^2(B)\Delta^2(G)\Big|\{\text{observed data}\}\right\} \le \mu_0(B)\mu_0(G) + o_p(1), \tag{A.16}$$

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 (A.15) and (A.16) 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

$$E\left\{\Delta^2(B)\Big|\{\text{observed data}\}\right\} \;\le\; n_1^{-1}\sum_{i=1}^{n_1}(\hat{h}_{1i}(B))^2 + o_p(1) \tag{A.17}$$

$$E\left\{\Delta^2(B)\Delta^2(G)\Big|\{\text{observed data}\}\right\} \;\le\; 6n_1^{-2}\sum_{i=1}^{n_1}(\hat{h}_{1i}(B))^2\sum_{i=1}^{n_1}(\hat{h}_{1i}(G))^2 + o_p(1). \tag{A.18}$$

Then (A.15) and (A.16) follow from working with each of the four terms of $\hat{h}_{1i}$ in (A.17) and (A.18). 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.

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

FLYNN, N.M., FORTHAL, D.N., HARRO, C.D., MAYER, K.H., PARA, M.F. AND THE rgp120 HIV VACCINE STUDY GROUP (2005). Placebo-controlled trial of a recombinant glycoprotein 120 vaccine to prevent HIV infection. *Journal of Infectious Diseases* **191**, 654–665.

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

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. AND KOSOROK, M. R. (2001). Simultaneous inferences in nonparametric functional estimation. *Technical report*, University of Washington, Seattle, Washington.

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., 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. AND SUN, Y. (2004). Tests for comparing mark-specific 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**, 207–221.

GRAY, R. J. (1988). A class of k-sample tests for comparing the cumulative incidence of a competing risk. *Annals of Statistics* **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., WEI, L. J. AND YING, Z. (1993). Checking the Cox model with cumulative sums of martingale-based residuals. *Biometrika* **80**, 557–572.

McKEAGUE, I. W. AND ZHANG, M. J. (1994). Identification of nonlinear time series from first order cumulative characteristics. *Annals of Statistics* **22**, 495–514.

PRENTICE, R. L., KALBFLEISCH, J. D., PETERSON, A. V., FLOURNEY, N., FAREWELL, V. T. AND BRESLOW, N. E. (1978). The analysis of failure times in the presence of competing risks. *Biometrics* **34**, 541–554.

SCHEIKE, T. H. AND ZHANG, M. J. (1998). Cumulative regression function tests for longitudinal data. *Annals of Statistics* **26**, 1328-1355.

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.

WYATT, R., KWONG, P. D., DESJARDINS, E., SWEET, R. W., ROBINSON, J., HENDRICKSEN, W. A. AND 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. *Annals of Statistics* **11**, 1119–1135.

Table and Figure Legends

Table 1. Empirical Power ($\times$ 100%) for Testing H_1^0 and H_2^0

Table 2. Empirical Power ($\times$ 100%) for Testing H_1 and H_2

Table 3. Bias of $\widehat{\text{VE}}^c(36, v)$ and 95% Coverage Probability of $\text{VE}^c(36, v)$

Figure 1. The top panel shows boxplots of amino acid distances in HIV gp120 between the infecting viruses and the nearest vaccine strain MN or GNE8, for the three studied HIV distances. The bottom panel shows p-values of the studied tests: Cox corresponds to the Wald test in the Cox model; 11, 12, 13, 14 correspond to $\hat{U}_1^1, \hat{U}_2^1, \hat{U}_3^1, \hat{U}_4^1$; n1, n2, correspond to $\hat{U}_1^{np}$, $\hat{U}_2^{np}$; s1, s2, correspond to $\hat{U}_1^{sp}$, $\hat{U}_2^{sp}$.

Figure 2. For the neutralizing face core distances, the top-left panel shows the observed test process $L_n^{np}(t, v)$ and the other panels show 8 randomly selected realizations of the simulated null test process $L_n^{np*}(t, v)$.

Figure 3. The left panels show point and 95% confidence interval estimates of $\text{VE}^c(36, v) = 1 - F_1(36, v)/F_2(36, v)$ versus the HIV gp120 amino acid distance between infecting viruses and the nearest vaccine antigen MN or GNE8, for the three studied HIV distances. The right panels show corresponding point and interval estimates of $\text{VE}^{dc}(36, v) = 1 - P(T_1 \leq 36, V_1 \leq v)/P(T_2 \leq 36, V_2 \leq v)$. The dashed horizontal line is the overall vaccine efficacy estimate $\widehat{\text{VE}}^c(36) = 0.048$.

Table 1. Empirical Power (× 100%) for Testing H_1^0 and H_2^0

			VE(τ) = 0	VE(τ) = 0.33				VE(τ) = 0.67			
			β_1	β_1				β_1			
n_k	*Test*	*Altern.*	1	1	0.5	0.25	*2-sided*	1	0.5	0.25	*2-sided*
100	Cox[1]		5.2	65.1	65.1	65.1	61.8	99.9	99.9	99.9	99.8
(48)[2]	$\hat{U}_1^1$	H_1^0	7.9	68.1	72.3	78.8	58.7	99.8	100	100	96.8
	$\hat{U}_2^1$	H_1^0	7.7	58.5	81.0	97.8	56.5	97.8	100	100	97.7
	$\hat{U}_3^1$	H_2^0	5.9	55.4	60.2	69.7	47.3	98.9	99.5	100	94.8
	$\hat{U}_4^1$	H_2^0	6.7	47.6	71.8	94.8	43.1	96.8	99.3	100	94.6
200	Cox		5.0	90.6	90.6	90.6	100	100	100	100	100
(95)[2]	$\hat{U}_1^1$	H_1^0	5.0	92.7	94.3	97.2	91.5	100	100	100	100
	$\hat{U}_2^1$	H_1^0	5.3	86.0	98.4	100	88.1	100	100	100	100
	$\hat{U}_3^1$	H_2^0	7.0	87.5	90.3	94.7	84.7	100	100	100	100
	$\hat{U}_4^1$	H_2^0	5.3	81.0	95.4	100	79.4	100	100	100	100
400	Cox		5.8	99.7	99.7	99.7	100	100	100	100	100
(190)[2]	$\hat{U}_1^1$	H_1^0	6.6	99.9	99.9	100	99.5	100	100	100	100
	$\hat{U}_2^1$	H_1^0	6.0	99.0	100	100	98.8	100	100	100	100
	$\hat{U}_3^1$	H_2^0	5.3	99.6	99.9	100	99.0	100	100	100	100
	$\hat{U}_4^1$	H_2^0	5.2	97.9	100	100	97.6	100	100	100	100

[1]Test statistic is a Wald Z-statistic based on the standard Cox model that ignores the mark.

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

Table 2. Empirical Power (× 100%) for Testing H_1 and H_2

			VE(τ) = 0.33				VE(τ) = 0.67			
			β_1				β_1			
n_k	*Test*	*Altern.*	1	0.5	0.25	*2-sided*	1	0.5	0.25	*2-sided*
100	$\hat{U}_1^{np}$	H_1	6.4	21.8	59.0	42.7	7.1	17.0	35.2	22.9
(48)[1]	$\hat{U}_2^{np}$	H_2	6.2	15.9	47.7	43.3	6.7	12.2	26.1	20.4
	$\hat{U}_1^{sp}$	H_1	6.2	18.3	52.9	35.8	5.7	12.8	30.2	17.8
	$\hat{U}_2^{sp}$	H_2	4.4	11.1	41.4	38.8	3.5	7.3	18.7	15.3
200	$\hat{U}_1^{np}$	H_1	6.3	32.4	87.0	78.3	6.7	21.0	62.7	48.8
(95)[1]	$\hat{U}_2^{np}$	H_2	6.8	23.0	81.4	80.9	6.5	14.3	54.2	51.4
	$\hat{U}_1^{sp}$	H_1	5.6	29.7	84.8	76.8	5.5	20.0	61.1	46.3
	$\hat{U}_2^{sp}$	H_2	5.4	20.8	79.5	81.4	4.8	13.2	49.6	45.6
400	$\hat{U}_1^{np}$	H_1	5.8	48.2	99.5	98.3	6.2	33.7	93.3	87.4
(190)[1]	$\hat{U}_2^{np}$	H_2	5.2	35.8	98.6	98.7	5.8	25.4	89.2	90.4
	$\hat{U}_1^{sp}$	H_1	5.4	46.7	99.0	98.3	5.5	32.7	92.9	86.1
	$\hat{U}_2^{sp}$	H_2	4.8	35.3	98.5	98.7	5.1	23.8	87.9	89.4

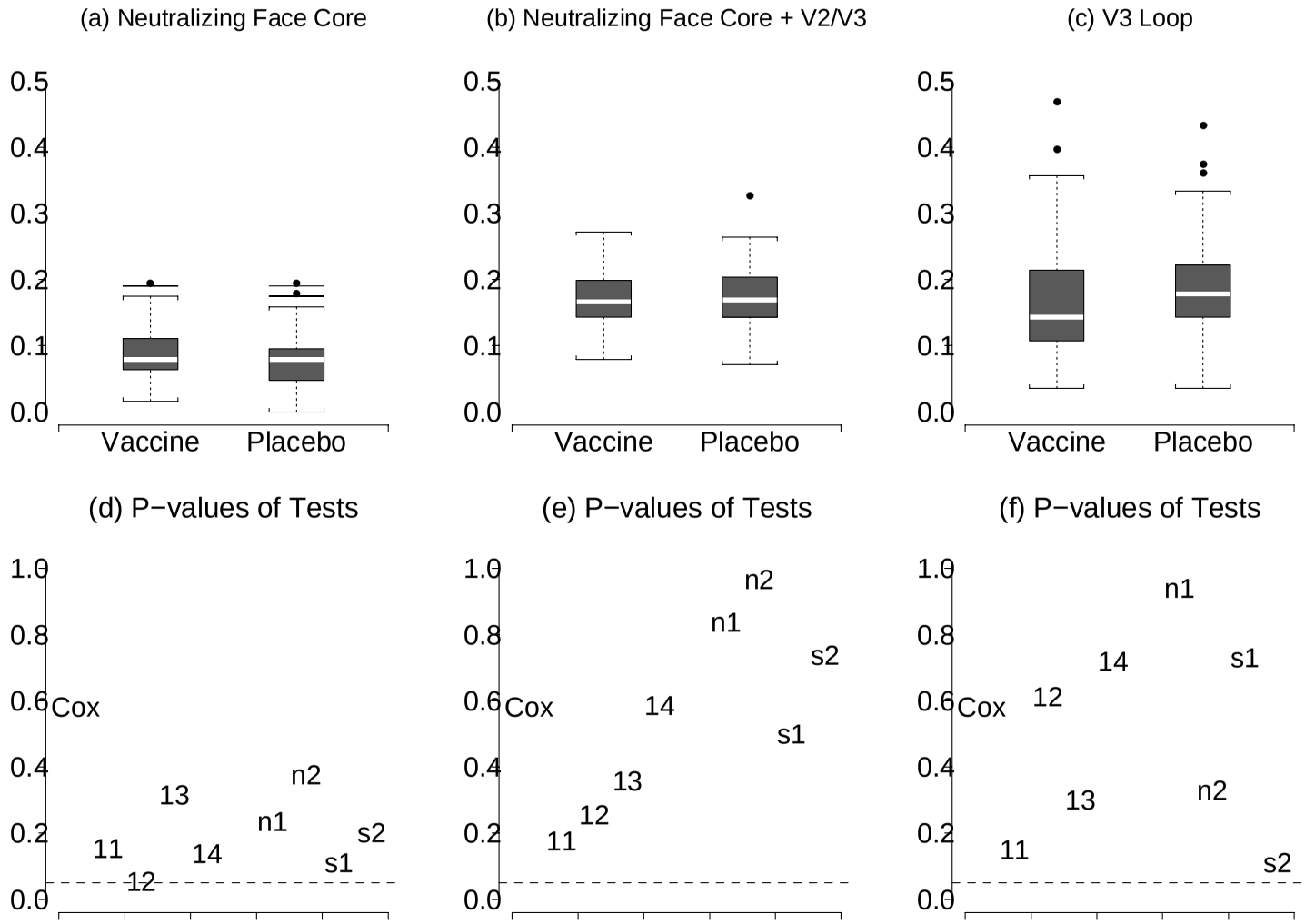
[1]Average number of subjects infected in group 2 (placebo).

Table 3. Bias of $\widehat{\mathrm{VE}}^c(36, v)$ and 95% Coverage Probability of $\mathrm{VE}^c(36, v)$

		$\mathrm{VE}(\tau) = 0.0$	$\mathrm{VE}(\tau) = 0.67$			$\mathrm{VE}(\tau) = 0.33$		
		β_1	β_1			β_1		
n_k	v	1	1	0.5	0.25	1	0.5	0.25
			Average Bias × 100					
100 (48)[1]	0.3	-6.3	−2.3	−6.3	−31.6	−2.5	−5.0	−20.8
	0.5	-5.8	−1.3	−2.6	−13.7	−3.6	−3.6	−9.0
	0.8	-6.3	−3.7	−3.0	−3.6	−5.2	−5.1	−9.6
200 (95)[1]	0.3	-2.8	−0.1	−1.6	−13.0	−0.9	−1.6	−9.0
	0.5	-1.6	−0.0	−0.9	−4.8	−1.0	−2.2	−6.0
	0.8	-3.5	−0.5	−0.6	−1.5	−2.1	−2.7	−5.4
400 (190)[1]	0.3	-1.4	−0.0	−0.4	−3.7	−0.2	−0.1	−3.0
	0.5	-1.1	−0.1	−0.8	−3.6	−0.0	−0.9	−4.6
	0.8	-0.8	−0.3	0.1	−0.9	−0.3	−0.2	−2.4
			Coverage Probability × 100%					
100 (48)[1]	0.3	97.6	97.9	96.0	73.9	97.2	97.3	86.6
	0.5	97.7	98.6	97.5	90.0	97.5	97.9	95.2
	0.8	94.7	96.0	96.2	95.4	94.6	94.9	96.1
200 (95)[1]	0.3	96.7	96.5	96.8	77.1	97.8	97.1	88.0
	0.5	97.2	96.7	97.5	93.8	96.8	97.5	96.5
	0.8	94.9	94.4	95.3	95.8	94.5	95.6	95.9
400 (190)[1]	0.3	96.8	95.4	96.4	87.8	96.8	97.3	92.2
	0.5	96.4	96.3	95.9	93.6	96.5	97.2	96.4
	0.8	96.9	96.0	96.3	96.7	96.2	96.8	96.8

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

Figure 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
(d) P−values of Tests
(e) P−values of Tests
(f) P−values of Tests
1.0
0.8
0.6
0.4
0.2
0.0
Cox
11
12
13
14
n1
n2
s1
s2

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

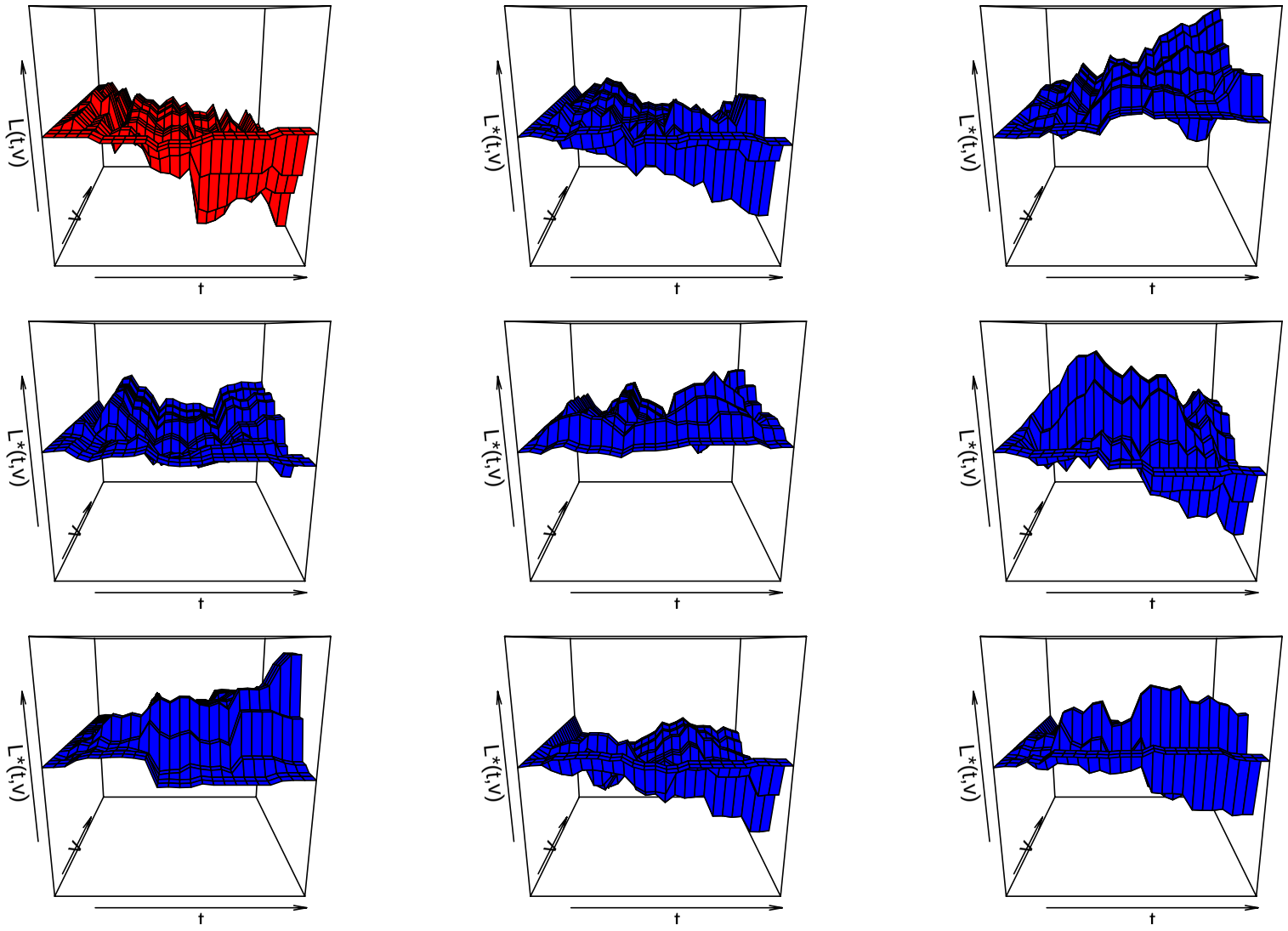
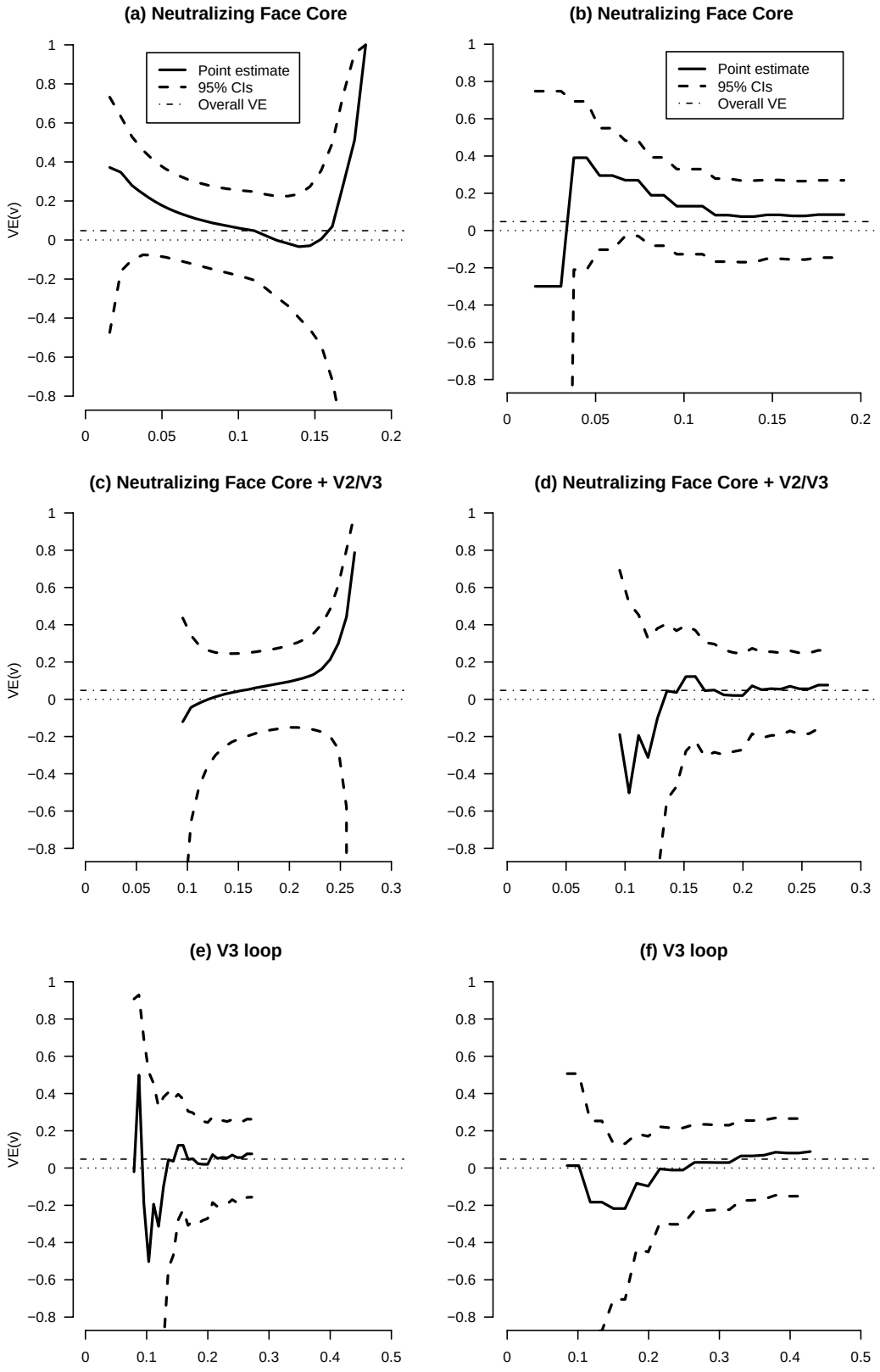


Figure 3


(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
VE(v)