

# Mark-specific proportional hazards model with multivariate continuous marks and its application to HIV vaccine efficacy trials

YANQING SUN

*Department of Mathematics and Statistics, University of North Carolina at Charlotte, Charlotte, NC  
28223, USA*

MEI LI\*

*School of Public Health, Zhejiang University, Hangzhou, China  
missmeili@hotmail.com*

PETER B. GILBERT

*Department of Biostatistics, University of Washington, and Fred Hutchinson Cancer Research Center,  
Seattle, WA 98109, USA*

## SUMMARY

For time-to-event data with finitely many competing risks, the proportional hazards model has been a popular tool for relating the cause-specific outcomes to covariates (Prentice *and others*, 1978. The analysis of failure time in the presence of competing risks. *Biometrics* 34, **541–554**). Inspired by previous research in HIV vaccine efficacy trials, the cause of failure is replaced by a continuous mark observed only in subjects who fail. This article studies an extension of this approach to allow a multivariate continuum of competing risks, to better account for the fact that the candidate HIV vaccines tested in efficacy trials have contained multiple HIV sequences, with a purpose to elicit multiple types of immune response that recognize and block different types of HIV viruses. We develop inference for the proportional hazards model in which the regression parameters depend parametrically on the marks, to avoid the curse of dimensionality, and the baseline hazard depends nonparametrically on both time and marks. Goodness-of-fit tests are constructed based on generalized weighted martingale residuals. The finite-sample performance of the proposed methods is examined through extensive simulations. The methods are applied to a vaccine efficacy trial to examine whether and how certain antigens represented inside the vaccine are relevant for protection or anti-protection against the exposing HIVs.

*Keywords:* Competing risks; Failure time data; Goodness-of-fit test; HIV vaccine trial; Hypothesis testing; Mark-specific relative risk; Multivariate data; Partial likelihood estimation; Semiparametric model; STEP trial.

\*To whom correspondence should be addressed.

## 1. INTRODUCTION

Around 30 years ago, [Prentice and others \(1978\)](#) developed the mark-specific proportional hazard (PH) model for discrete marks, and since then a great deal of work has been done in discrete-mark failure time analysis, cf., [Kalbfleisch and Prentice \(1980\)](#), [Sun \(2001\)](#) and [Scheike and others \(2008\)](#). Alternatively, [Huang and Louis \(1998\)](#) first considered a continuous mark and developed the nonparametric maximum likelihood estimator of the joint distribution of the failure time and the mark. Motivated by applications in preventative HIV vaccine randomized placebo-controlled efficacy trials, [Gilbert and others \(2004, 2008\)](#) developed procedures for testing dependence of mark-specific hazards rates and relative risks (RRs) on the mark variable. [Sun and others \(2009\)](#) developed estimation and hypothesis testing methods for the mark-specific PH model with a univariate continuous mark. The mark variable of interest is a measure of the protein sequence distance between an HIV sequence sampled from a volunteer infected in the trial, to an HIV sequence represented inside the tested vaccine construct. As such, the scientific question is if and how the vaccine's effect to reduce the hazard of HIV acquisition depends on the sequence distance mark; answers to this question may guide vaccine development, as discussed extensively in the above-cited papers and elsewhere.

However, the previous work did not account for multivariate marks. This is a serious limitation given that all of the candidate HIV vaccines tested in HIV vaccine efficacy trials have contained multiple antigens/immunogens, with rationale to attempt to elicit multiple types of immune response that recognize and block different types of HIV viruses. The greater the number of virus types that can be recognized and killed by vaccine-induced immune responses, the greater the potential overall vaccine efficacy. In the first two efficacy trials, the HIV vaccine construct contained two Envelope (Env) protein antigens, based on two distinct strains of HIV, such that a two-dimensional mark variable is of interest ([Flynn and others, 2005](#); [Pitisuttithum and others, 2006](#)). The vaccine construct evaluated in the third and fourth efficacy trials contained Gag, Pol, and Nef protein antigens, making a three-dimensional mark variable of interest ([Buchbinder and others, 2008](#); [Gray and others, 2009](#)). Lastly, the fifth and most recent efficacy trial tested a vaccine that contained Gag, Pol, and Nef protein antigens, as well as three distinct Env protein antigens, making a six-dimensional mark variable of interest ([Rerks-Ngarm and others, 2009](#)).

The previous work dealt with the multivariate mark issue by collapsing the multiple distances from an infected subject into a univariate mark—the minimum of the distances to each vaccine antigen. This approach is reasonable under the premise that the only thing that matters for protection is the nearness of the exposing HIV to the closest antigen represented inside of the vaccine (e.g. [Gilbert and others, 2008](#)). However, there are many ways in which this assumption may fail. For example, based on host genetics (e.g. HLA type or Fc- $\gamma$ -receptor type), one subgroup may be protected through immune responses that recognize HIV peptides that are similar to HIV peptides represented in antigen 1, whereas another subgroup may be protected through immune responses that are similar to HIV peptides represented in antigen 2; in this case, the vaccine efficacy depends on both individual distances and less so on the minimum distance. For a second example, there are many ways to define a protein sequence distance to be putatively immunologically relevant [several distances were used in the sequence analysis of the [Buchbinder and others \(2008\)](#) efficacy trial reported in [Rolland and others \(2011\)](#)], and if two distances are used such that the first considers many HIV sites irrelevant for protection whereas the second sagely restricts attention to key HIV sites that are contained in epitopes that cause protection, then the first distance could be shorter even though vaccine efficacy depends only on the second. Therefore, a more general approach to assessing and modeling how vaccine efficacy depends on multiple sequence distances is needed, which does not pre-assume a particular way to collapse the multivariate distances into a univariate distance. Outside of the survival analysis field, [Gilbert \(2000\)](#) studied such a general approach with multivariate marks, based on a semi-parametric biased sampling model. However, this method is limited by the fact that the model conditions on infection, so that conditional odds ratios but not prospective RRs of infection can be estimated, and

by the fact that the model treats HIV infection as a binary outcome, not accounting for the time to HIV infection.

Let  $\lambda(t, v|z)$  be the conditional mark-specific hazard function, defined as  $\lambda(t, v|z) = \lim_{h_1, h_2 \rightarrow 0} P\{T \in [t, t + h_1), V \in [v, v + h_2) | T \geq t, Z(t) = z\} / h_1 h_2$ , where  $T$  is the failure time,  $V$  is a continuous mark variable, and  $Z(t)$  is a time-dependent  $p$ -dimensional covariate. [Sun and others \(2009\)](#) studied the mark-specific PH model

$$\lambda(t, v|z) = \lambda_0(t, v) \exp\{\beta^T(v)z(t)\}, \quad (1.1)$$

for evaluating the mark-specific vaccine efficacy. Under this model, the ratio of hazard functions for any two individuals is a function of the mark independent of time. This assumption may not always be met in practice, and can be relaxed through stratification ([Dabrowska, 1997](#)). Here we propose to study the stratified mark-specific PH model with multivariate marks where the baseline hazard function can vary with stratum. Under the stratified mark-specific PH model, the ratio of hazard functions for any two individuals within same stratum is independent of time. However, it may be time dependent for two individuals from different strata. For simplicity, we consider a two-dimensional mark variable  $v = (v_1, v_2)$ . The methods for general multivariate mark follow a similar outline. Note that the mark-specific RR function  $\beta(v) = \beta(v_1, v_2)$  is a  $p$ -dimensional function depending on  $v$ . Estimation of  $\beta(v)$  without any structural assumptions is possible following the procedure for one-dimensional mark variable by [Sun and others \(2009\)](#). However, this would require a very large sample size due to the curse of dimensionality. In this paper, we consider a parametric model for  $\beta(v)$  as the first-order Taylor approximation of  $\beta(v)$  plus an interaction term that allows investigation of the effects of both marks as well as the interactions. We develop an estimation procedure for the stratified mark-specific PH model under this parametric structure for  $\beta(v)$ . We also develop some testing procedures examining how the vaccine efficacy depends on the marks and how relevant the two marks are for protection against HIV infection. The goodness of fit of the proposed model is tested based on the generalized weighted martingale residuals.

The rest of the paper is organized as follows. In Section 2, we develop a partial likelihood procedure for estimating  $\beta(v)$ . The asymptotic properties of the proposed estimator are derived. Likelihood-based statistical tests are proposed to examine a variety of null hypotheses to understand whether and how the vaccine efficacy depends on the marks. In Section 3, we propose a procedure for testing goodness of fit of the multivariate mark-specific proportional hazard model based on the generalized weighted martingale residuals. The proposed methods are applied to a vaccine efficacy trial in Section 4. Extensive simulations are conducted to examine the finite sample performance of the proposed methods in Section 5. Proofs of the main results are placed in the supplementary material available at *Biostatistics* online.

## 2. MARK-SPECIFIC PH MODEL WITH MULTIVARIATE MARKS

### 2.1 Mark-specific PH model

Let  $\lambda_k(t, v|z)$  be the conditional mark-specific hazard function at time  $t$  with mark  $v$  for a subject in stratum  $k$  with covariate  $z$  ([Sun and others, 2009](#)). We assume that the mark  $V$  takes value in the interval  $[0, 1]^2$ , rescaled if necessary. The stratified multivariate mark-specific PH model speculates that

$$\lambda_k(t, v|z) = \lambda_{0k}(t, v) \exp\{\beta^T(v)z(t)\}, \quad \text{for } t \geq 0, 0 \leq v \leq 1, \quad (2.1)$$

for  $k = 1, 2, \dots, K$ , where  $K$  is the number of the strata,  $\lambda_{0k}(t, v)$  is an unspecified baseline mark-specific hazards function for  $k$ th stratum,  $z$  is a  $p$ -dimensional covariate, and  $\beta(v)$  is a  $p$ -dimensional vector of

regression functions. Here we assume that  $\beta(v)$  is given by

$$\beta(v) = \beta_0 + \beta_1 v_1 + \beta_2 v_2 + \beta_{12} v_1 v_2, \quad (2.2)$$

where each of  $\beta_0, \beta_1, \beta_2$ , and  $\beta_{12}$  is a  $p$ -dimensional vector. Thus  $\beta(v)$  is completely specified by the  $4 \times p$  parameters denoted by  $\bar{\beta} = (\beta_0^T, \beta_1^T, \beta_2^T, \beta_{12}^T)^T$ . Let  $z = (z_1, z_2(t))^T$ , where  $z_1$  is the vaccine group indicator, with  $z_1 = 1$  for the vaccine group and  $z_1 = 0$  for the placebo group, while  $z_2(t)$  is other possibly time-dependent covariates. Under the mark-specific PH model (2.1), the mark-specific RR is  $RR(v) = \lambda_k(t, v|z_1 = 1)/\lambda_k(t, v|z_1 = 0) = \exp(b_1(v))$ , where  $b_1(v)$  is the coefficient corresponding  $z_1$ . The mark-specific vaccine efficacy  $VE(v) = 1 - RR(v)$  represents the percentage reduction for the vaccine versus placebo in risk of failure at time  $t$  with mark  $v$  after adjusting for covariates  $z_2$ . It does not depend on time  $t$  under model (2.1).

Let  $T_k$  be the failure time for an individual in the  $k$ th stratum,  $V_k$  be the mark observed at failure and  $Z_k$  be the associated  $p$ -dimensional covariate. Assume that  $(T_k, V_k, Z_k)$  follows model (2.1). Under right censoring, the observed random variables are  $(X_k, \delta_k, \delta_k V_k, Z_k)$ , where  $X_k = \min(T_k, C_k)$ ,  $\delta_k = I(T_k \leq C_k)$  and  $C_k$  is the censoring random variable, which is assumed to be independent of  $(T_k, V_k)$  given  $Z_k$ . The mark variable  $V_k$  is observed if the failure time is uncensored. Let  $(X_{ki}, \delta_{ki}, \delta_{ki} V_{ki}, Z_{ki})$ ,  $i = 1, \dots, n_k$ , be independent identically distributed (iid) replicates of  $(X_k, \delta_k, \delta_k V_k, Z_k)$  from stratum  $k$ , for  $k = 1, 2, \dots, K$ . Let  $\tau$  be the end of follow-up time, where failure times beyond  $\tau$  are considered censored. The estimation of model (2.1) for  $\beta(v)$  under parametric structure (2.2) is given next.

## 2.2 Estimation of the mark-specific PH model

Let  $N_{ki}(t, v) = I(X_{ki} \leq t, V_{ki} \leq v, \delta_{ki} = 1)$  be the marked counting process for  $i$ th individual in stratum  $k$ , where the relation  $V_{ki} \leq v$  indicates that the inequality holds for each component of the multivariate mark. Let  $Y_{ki}(t) = I(X_{ki} \geq t)$  be the at-risk process. Under (2.2),  $\beta(v)$  is completely specified by  $\bar{\beta} = (\beta_0^T, \beta_1^T, \beta_2^T, \beta_{12}^T)^T$ . Let  $\bar{v} = (1, v_1, v_2, v_1 v_2)^T$  and  $\tilde{Z}_{ki}(t, v) = Z_{ki}(t) \otimes \bar{v}$  with  $\otimes$  being the Kronecker product. Similar to Kalbfleisch and Prentice (1980) for the competing risks model with finite number of causes, the log-partial likelihood function for  $\bar{\beta}$  can be expressed as

$$l(\bar{\beta}) = \sum_{k=1}^K \sum_{i=1}^{n_k} \int_0^\tau \int_0^1 \left[ \beta^T(v) Z_{ki}(s) - \log \left\{ \sum_{l=1}^K \sum_{j=1}^{n_k} Y_{lj}(s) \exp(\beta^T(v) Z_{lj}(s)) \right\} \right] N_{ki}(ds, dv). \quad (2.3)$$

Since  $\bar{\beta}^T \tilde{Z}_{ki}(s, v) = (\beta_0^T, \beta_1^T, \beta_2^T, \beta_{12}^T)(Z_{ki}^T(s), Z_{ki}^T(s)v_1, Z_{ki}^T(s)v_2, Z_{ki}^T(s)v_1 v_2)^T = \beta^T(v) Z_{ki}(s)$ , we have

$$l(\bar{\beta}) = \sum_{k=1}^K \sum_{i=1}^{n_k} \int_0^\tau \int_0^1 \left[ \bar{\beta}^T \tilde{Z}_{ki}(s, v) - \log \left\{ \sum_{l=1}^K \sum_{j=1}^{n_k} Y_{lj}(s) \exp(\bar{\beta}^T \tilde{Z}_{lj}(s, v)) \right\} \right] N_{ki}(ds, dv). \quad (2.4)$$

The maximum partial likelihood estimator (MPLE)  $\hat{\bar{\beta}}$  for  $\bar{\beta}$  is obtained by maximizing  $l(\bar{\beta})$  given in (2.4). Taking the derivative of  $l(\bar{\beta})$  with respect to  $\bar{\beta}$ , the score function can be written as

$$U(\bar{\beta}) = \frac{\partial l(\bar{\beta})}{\partial \bar{\beta}} = \sum_{k=1}^K \sum_{i=1}^{n_k} \int_0^\tau \int_0^1 \left[ \tilde{Z}_{ki}(s, v) - \frac{S_k^{(1)}(s, v, \bar{\beta})}{S_k^{(0)}(s, v, \bar{\beta})} \right] N_{ki}(ds, dv), \quad (2.5)$$

where  $S_k^{(j)}(s, v, \bar{\beta}) = n_k^{-1} \sum_{i=1}^{n_k} Y_{ki}(s) \exp\{\bar{\beta}^T \tilde{Z}_{ki}(s, v)\} \tilde{Z}_{ki}(s, v)^{\otimes j}$  for  $j = 0, 1, 2$ . Here for a vector  $z$ ,  $z^{\otimes 0} = 1$ ,  $z^{\otimes 1} = z$ , and  $z^{\otimes 2} = zz^T$ . Let  $Q_k(t, v, \bar{\beta}) = S_k^{(2)}(t, v, \bar{\beta})/S_k^{(0)}(t, v, \bar{\beta}) - (S_k^{(1)}(t, v, \bar{\beta})/$

$S_k^{(0)}(t, v, \bar{\beta})^{\otimes 2}$ . The information matrix is the derivative of  $U(\bar{\beta})$  with respect to  $\bar{\beta}$  given as

$$I(\bar{\beta}) = \sum_{k=1}^K \sum_{i=1}^{n_k} \int_0^\tau \int_0^1 Q_k(s, v, \bar{\beta}) N_{ki}(ds, dv). \quad (2.6)$$

The MPLE  $\hat{\beta}$  for  $\bar{\beta}$  is the solution to  $U(\bar{\beta}) = 0$ . Under condition 1 in Section 2.3, the matrix  $I(\bar{\beta})$  is positive definite with probability 1 as  $n_k \rightarrow \infty$  for  $1 \leq k \leq K$ . Thus  $\hat{\beta}$  exists almost surely and is unique for large sample sizes. The baseline function  $\lambda_{0k}(t, v)$  can be estimated by smoothing the increments of the following estimator of the doubly cumulative baseline function  $\Lambda_{0k}(t, v) = \int_0^t \int_0^v \lambda_{0k}(s, u) ds du$ :

$$\hat{\Lambda}_{0k}(t, v) = \int_0^t \int_0^v \frac{N_k(ds, du)}{n_k S_k^{(0)}(s, u, \hat{\beta})}, \quad (2.7)$$

where  $N_k(t, v) = \sum_{i=1}^{n_k} N_{ki}(t, v)$ . A kernel estimator of  $\lambda_{0k}(t, v)$  is given by  $\hat{\lambda}_{0k}(t, v) = \int_0^\tau \int_0^1 K_{h_1}^{(1)}(t-s) K_{h_2}^{(1)}(v-u) \hat{\Lambda}_{0k}(ds, du)$ , where  $K_{h_1}^{(1)}(x) = K^{(1)}(x/h_1)/h_1$  and  $K_{h_2}^{(2)}(x) = K^{(2)}(x/h_2)/h_2$ ,  $K^{(1)}(\cdot)$  and  $K^{(2)}(\cdot)$  are kernel functions, and  $h_1$  and  $h_2$  are the corresponding bandwidths. The cause-related cumulative incidence function  $P(T_k \leq t, V_k \leq v | Z_k) = \int_0^t \int_0^v S_k(s | Z_k) \exp\{\beta^T(u) Z_k\} \Lambda_{k0}(ds, du)$  can be estimated by  $\int_0^t \int_0^v \hat{S}_k(s | Z_k) \exp\{\hat{\beta}^T(u) Z_k\} \hat{\Lambda}_{k0}(ds, du)$ , where  $\hat{S}_k(s | Z_k) = \exp\{-\int_0^s \int_0^1 \exp\{\hat{\beta}^T(u) Z_k\} \hat{\Lambda}_{k0}(dx, du)\}$  and  $\hat{\beta}^T(u) Z_k = \hat{\beta}^T Z_k \otimes \bar{v}$ .

### 2.3 Asymptotic results

Let  $\bar{\beta}_0$  be the true value of  $\bar{\beta}$  under models (2.1) and (2.2). Let  $s_k^{(j)}(t, v, \bar{\beta}) = E\{S_k^{(j)}(t, v, \bar{\beta})\}$ , for  $j = 0, 1, 2$ , and  $q_k(t, v, \bar{\beta}) = s_k^{(2)}(t, v, \bar{\beta})/s_k^{(0)}(t, v, \bar{\beta}) - (s_k^{(1)}(t, v, \bar{\beta})/s_k^{(0)}(t, v, \bar{\beta}))^{\otimes 2}$ . Let  $n = \sum_{k=1}^K n_k$ . We make use of the following regularity conditions.

*Condition 1.* The covariate process  $Z_k(t)$  is left continuous with bounded variation and satisfies the moment condition  $\sup_{0 \leq t \leq \tau} E\{\|Z_k(t)\|^4 \exp(M\|Z_k(t)\|)\} < \infty$ , where  $\|\cdot\|$  is the Euclidean norm and  $M$  is a positive constant such that  $(\beta_0^T, \beta_1^T, \beta_2^T, \beta_{12}^T) \in (-M, M)^{4p}$  for  $1 \leq k \leq K$ .

*Condition 2.* For  $1 \leq k \leq K$ ,  $\lambda_{0k}(t, v)$  is continuous on  $[0, \tau] \times [0, 1]^2$ ,  $s_k^{(0)}(t, v, \bar{\beta}) > 0$  and each component of  $s_k^{(j)}(t, v, \bar{\beta})$  is continuous on  $[0, \tau] \times [0, 1]^2 \times \mathcal{B}$ , where  $\mathcal{B}$  is a neighborhood of  $\bar{\beta}_0$ .

*Condition 3.* The limit  $n_k/n \rightarrow p_k$  exists as  $n \rightarrow \infty$  for  $0 < p_k < 1$  for  $1 \leq k \leq K$ . The matrix  $\Sigma(\bar{\beta}) = \sum_{k=1}^K p_k \int_0^\tau \int_0^1 q_k(t, v, \bar{\beta}) s_k^{(0)}(t, v, \bar{\beta}) \lambda_{0k}(t, v) dt dv$  is positive definite for  $\bar{\beta} \in \mathcal{B}$ .

The asymptotic results for  $\hat{\beta}$  and  $\hat{\Lambda}_{0k}(\cdot, \cdot)$ ,  $1 \leq k \leq K$ , are presented in the following theorems.

**THEOREM 2.1** Under conditions 1–3,  $\hat{\beta}$  converges in probability to  $\bar{\beta}_0$  as  $n \rightarrow \infty$ .

**THEOREM 2.2** Under conditions 1–3,  $n^{1/2}(\hat{\beta} - \bar{\beta}_0) \xrightarrow{\mathcal{D}} N(0, \Sigma^{-1}(\bar{\beta}_0))$  as  $n \rightarrow \infty$ , where  $\Sigma(\bar{\beta}_0)$  can be consistently estimated by  $n^{-1}I(\hat{\beta})$ .

THEOREM 2.3 Under conditions 1–3, the following decomposition holds uniformly in  $(t, v) \in [0, \tau] \times [0, 1]$  for  $1 \leq k \leq K$  as  $n \rightarrow \infty$ :

$$\begin{aligned} & \sqrt{n}\{\hat{\Lambda}_{0k}(t, v) - \Lambda_{0k}(t, v)\} \\ &= -\sqrt{n}(\hat{\beta} - \bar{\beta}_0) \int_0^t \int_0^v \frac{[s_k^{(1)}(s, u, \bar{\beta}_0)]}{s_k^{(0)}(s, u, \bar{\beta}_0)} \lambda_{0k}(s, u) ds du + \sqrt{n} \int_0^t \int_0^v \frac{M_k(ds, du)}{n_k s_k^{(0)}(s, u, \bar{\beta}_0)} + o_p(1). \end{aligned}$$

$\sqrt{n}(\hat{\beta} - \bar{\beta}_0)$  is asymptotically independent of the processes  $\sqrt{n} \int_0^t \int_0^v \{n_k s_k^{(0)}(s, u, \bar{\beta}_0)\}^{-1} M_k(ds, du)$ ,  $k = 1, \dots, K$ , with the latter being asymptotically independent mean-zero Guassian random fields with variances  $\int_0^t \int_0^v \{p_k s_k^{(0)}(s, u, \bar{\beta}_0)\}^{-1} \lambda_{k0}(s, u) ds du$  and with independent increments.

## 2.4 Hypothesis testing

We propose some statistical tests for evaluating whether and how the vaccine efficacy depends on the marks. The following null hypotheses are examined:  $H_{10} : \beta_1 = \beta_2 = \beta_{12} = 0$ ;  $H_{20} : \beta_{12} = 0$ ;  $H_{30} : \beta_2 = \beta_{12} = 0$  and  $H_{40} : \beta_1 = \beta_{12} = 0$ . The null hypothesis  $H_{10}$  indicates that the RRs do not depend on the marks;  $H_{20}$  implies that the marks  $v_1$  and  $v_2$  do not have interactive effects on RRs;  $H_{30}$  implies that RRs are not affected by  $v_2$ ; while  $H_{40}$  implies that RRs are not affected by  $v_1$ .

Likelihood-based tests such as the likelihood ratio test (LRT), Wald test, and score test are commonly used in the parametric settings. Here we adopt these tests for model (2.1) with  $\beta(v)$  having the parametric structure (2.2). The tests are constructed based on the log-partial likelihood function  $l(\bar{\beta})$  given in (2.4). Let  $\hat{\beta}$  be the MPLE maximizing  $l(\bar{\beta})$ . Denote  $H_0$  as one of the null hypotheses  $H_{10}$ ,  $H_{20}$ , or  $H_{30}$ . Let  $\hat{\beta}_{H_0}$  be the estimator of  $\bar{\beta}$  under  $H_0$ , which is the maximizer of  $l(\bar{\beta})$  under  $H_0$ . For example, for  $H_{10}$ ,  $\hat{\beta}_{H_0}$  is the maximizer of  $l(\bar{\beta})$  under the restriction  $\beta_1 = \beta_2 = \beta_{12} = 0$ . The LRT statistic is  $T_l = 2\{l(\hat{\beta}) - l(\hat{\beta}_{H_0})\}$ . The Wald test statistic is given by  $T_w = (\hat{\beta} - \hat{\beta}_{H_0})^T [I(\hat{\beta})] (\hat{\beta} - \hat{\beta}_{H_0})$ , where the information matrix  $I(\hat{\beta})$  is defined in (2.6). The score test statistic is given by  $T_s = U^T(\hat{\beta}_{H_0}) I(\hat{\beta}_{H_0})^{-1} U(\hat{\beta}_{H_0})$ , where the score function  $U(\hat{\beta}_{H_0})$  and information matrix  $I(\hat{\beta})$  are defined in (2.5) and (2.6), respectively.

Routine analysis following Serfling (1980) shows that under  $H_0$ ,  $T_l$ ,  $T_w$ , and  $T_s$  converge in distribution to a chi-square distribution with degrees of freedom equal to the number of parameters specified under  $H_0$ . The LRT rejects  $H_0$  if  $T_l > \chi_{3p, \alpha}^2$ , the upper  $\alpha$  quantile of the chi-square distribution with  $3p$  degrees of freedom. The corresponding critical values for testing  $H_{20}$ ,  $H_{30}$ , and  $H_{40}$  are  $\chi_{p, \alpha}^2$ ,  $\chi_{2p, \alpha}^2$ , and  $\chi_{2p, \alpha}^2$ , respectively. Similar decision rules hold for the Wald test with test statistic  $T_w$  and the score test with test statistic  $T_s$ . The parameters  $\beta_1$ ,  $\beta_2$ , and  $\beta_{12}$  specified in the null hypotheses are  $p$ -dimensional vectors. The tests can be carried out for the hypotheses that only concern a single component of the coefficients to examine how one particular covariate effect is modified by the marks.

## 3. GOODNESS-OF-FIT TESTS

The estimation and testing procedures developed in Section 2 are developed under model (2.1) with  $\beta(v)$  having the parametric structure (2.2). The validity of these procedures depends on goodness of fit of the multivariate mark-specific proportional hazard model. This section develops some goodness-of-fit tests of model (2.1) under the parametric structure (2.2) for  $\beta(v)$  following Lin and others (1993). Let  $M_{ki}(t, v) = N_{ki}(t, v) - \int_0^t \int_0^v Y_{ki}(s) \exp(\bar{\beta}^T \tilde{Z}_{ki}(s, u)) \lambda_{0k}(s, u) ds du$ . We derive the model checking test

statistics based on the cumulative martingale residuals defined as

$$\hat{M}_{ki}(t, v) = \int_0^t \int_0^v N_{ki}(ds, du) - Y_{ki}(s) \exp(\hat{\beta}^T \tilde{Z}_{ki}(s, u)) \hat{\Lambda}_{0k}(ds, du), \quad (3.1)$$

where  $\hat{\beta}$  is the MPLE given in Section 2, and  $\hat{\Lambda}_{0k}(t, v)$  is defined in (2.7).  $\hat{M}_{ki}(t, v)$  may be interpreted as the difference up to time  $t$  between the observed and the predicted number of events with marks at failures less than  $v$  for the  $i$ th subject in  $k$ th stratum. Thus the martingale residuals are informative about model misspecifications. It is easy to check that  $\sum_{k=1}^K \sum_{i=1}^{n_k} \hat{M}_{ki}(t, v) = 0$ .

For  $1 \leq k \leq K$ , let

$$W_k(t, v, z) = n^{-1/2} \sum_{i=1}^{n_k} g_k(Z_{ki}, z) \hat{M}_{ki}(t, v), \quad (3.2)$$

where  $g_k(Z_{ki}, z)$  is a  $1 \times r$ -vector of known bounded functions of  $Z_{ki}$  and  $z$ . For example, one may take  $g_k(Z_{ki}, z) = f_k(Z_{ki})I(Z_{ki} \leq z)$ , where  $f_k(\cdot)$  is a known function and  $I(Z_{ki} \leq z) = \{I(Z_{1ki} \leq z_1), \dots, I(Z_{pki} \leq z_p)\}$ , in which case  $r = p$ . We construct goodness-of-fit test statistics based on the test process  $W(t, v, z) = (W_1(t, v, z), \dots, W_K(t, v, z))$ . If models (2.1) and (2.2) hold, the process  $W(t, v, z)$  fluctuates randomly about zero. Various test statistics can be constructed by selecting different weight functions  $f_k(\cdot)$  and using different functionals of the process  $W(t, v, z)$ . Here we propose the supremum test statistics to test the overall fit of the model:

$$S = \sup_{1 \leq k \leq K} \sup_{t, v, z} |W_k(t, v, z)|. \quad (3.3)$$

The distribution of  $S$  under the null hypothesis is complicated and intractable. To calculate the critical value of the proposed test statistic, we consider using the Gaussian multiplier method (Lin and others, 1993) that can be applied to approximate the distribution of the process  $W(t, v, z)$ . The key step toward the application of this method is to approximate  $W(t, v, z)$  with the sum of iid processes as shown next.

Let  $S_{kg}^{(0)}(t, v, z, \bar{\beta}) = n_k^{-1} \sum_{i=1}^{n_k} Y_{ki}(t) \exp\{\bar{\beta}^T \tilde{Z}_{ki}(t, v)\} g_k(Z_{ki}, z)$  and  $S_{kg}^{(1)}(t, v, z, \bar{\beta}) = n_k^{-1} \sum_{i=1}^{n_k} Y_{ki}(t) \exp\{\bar{\beta}^T \tilde{Z}_{ki}(t, v)\} \tilde{Z}_{ki}(t, v) \otimes g_k(Z_{ki}, z)$ , where  $A \otimes B$  is the Kronecker product of matrices  $A$  and  $B$ . Let  $s_{kg}^{(0)}(t, v, z, \bar{\beta}) = E\{S_{kg}^{(0)}(t, v, z, \bar{\beta})\}$  and  $s_{kg}^{(1)}(t, v, z, \bar{\beta}) = E\{S_{kg}^{(1)}(t, v, z, \bar{\beta})\}$ . The proof of the following decomposition is given in the supplementary material at *Biostatistics* online.

**THEOREM 3.1** Assume conditions 1–3. For  $1 \leq k \leq K$ , we have

$$\begin{aligned} W_k(t, v, z) &= n^{-1/2} \sum_{l=1}^K \sum_{i=1}^{n_l} \int_0^\tau \int_0^v I(l=k) I(s \leq t) \left[ g_l(Z_{li}, z) - \frac{s_{lg}^{(0)}(s, u, z, \bar{\beta}_0)}{s_l^{(0)}(s, u, \bar{\beta}_0)} \right] M_{li}(ds, du) \\ &\quad + (R_k(t, v, z))^T (\Sigma(\bar{\beta}_0))^{-1} n^{-1/2} \\ &\quad \times \sum_{l=1}^K \sum_{i=1}^{n_l} \left\{ \int_0^\tau \int_0^1 \left[ \tilde{Z}_{li}(s, u) - \frac{s_l^{(1)}(s, u, \bar{\beta}_0)}{s_l^{(0)}(s, u, \bar{\beta}_0)} \right] M_{li}(ds, du) \right\}^T + o_p(1), \end{aligned} \quad (3.4)$$

as  $n \rightarrow \infty$  where

$$R_k(t, v, z) = p_k \int_0^t \int_0^v \left[ \frac{s_k^{(1)}(s, x, \bar{\beta}_0) \otimes s_{kg}^{(0)}(s, x, z, \bar{\beta}_0)}{s_k^{(0)}(s, x, \bar{\beta}_0)} - s_{kg}^{(1)}(s, x, z, \bar{\beta}_0) \right] \lambda_{0k}(s, x) ds dx.$$

Expression (3.4) shows that the process  $W_k(t, v, z)$  is asymptotically equivalent to the sum of iid terms involving the integrations with respect to  $M_{li}(s, u)$ . Donsker's theorem (cf., [van der Vaart, 1998](#)) on the weak convergence of empirical processes can be used to show that the process  $W(t, v, z) = (W_1(t, v, z), \dots, W_K(t, v, z))$  converges weakly to a multi-dimensional Gaussian random field.

Let  $\{\xi_{li}, i = 1, \dots, n_l, l = 1, \dots, K\}$  be iid standard normal random variables. For  $1 \leq k \leq K$ , let

$$\begin{aligned} W_k^*(t, v, z) &= n^{-1/2} \sum_{l=1}^K \sum_{i=1}^{n_l} \int_0^\tau \int_0^v I(l=k)I(s \leq t) \left[ g_l(Z_{li}, z) - \frac{S_{lg}^{(0)}(s, u, z, \hat{\beta})}{S_l^{(0)}(s, u, z, \hat{\beta})} \right] \xi_{li} N_{li}(ds, du) \\ &\quad + n^{-1/2} (\hat{R}_k(t, v, z))^T (n^{-1} I(\hat{\beta}))^{-1} \\ &\quad \times \sum_{l=1}^K \sum_{i=1}^{n_l} \left\{ \int_0^\tau \int_0^1 \left[ \tilde{Z}_{li}(s, u) - \frac{S_l^{(1)}(s, u, \bar{\beta})}{S_l^{(0)}(s, u, \bar{\beta})} \right] \xi_{li} N_{li}(ds, du) \right\}^T, \end{aligned} \quad (3.5)$$

where

$$\hat{R}_k(t, v, z) = (n_k/n) \int_0^t \int_0^v \left[ \frac{S_k^{(1)}(s, x, \hat{\beta}) \otimes S_{kg}^{(0)}(s, x, z, \hat{\beta})}{S_k^{(0)}(s, x, \hat{\beta})} - S_{kg}^{(1)}(s, x, z, \hat{\beta}) \right] d\hat{\Lambda}_{0k}(s, x).$$

Following an application of [Sun and Wu \(2005, Lemma 1\)](#), the distribution of  $W(t, v, z)$  can be approximated by the conditional distribution of  $W^*(t, v, z) = (W_1^*(t, v, z), \dots, W_K^*(t, v, z))$  given the observed data sequence. The continuous mapping theorem entails that the distribution of  $S$  can be approximated by the empirical distribution of  $S^* = \sup_{1 \leq k \leq K} \sup_{t, v, z} |W_k^*(t, v, z)|$  obtained through repeatedly generating independent sets of  $\{\xi_{li}, i = 1, \dots, n_l, l = 1, \dots, K\}$  while holding the observed data fixed. We reject model (2.1) under the parametric structure (2.2) for  $\beta(v)$  at the significance level  $\alpha$  if  $S$  is greater than the critical value  $S_\alpha$ —the upper  $\alpha$  quantile of the copies of  $S^*$  described above.

#### 4. APPLICATION TO A VACCINE EFFICACY TRIAL

We now illustrate our methods with the analysis of data from a vaccine efficacy trial conducted in North and South America ([Buchbinder and others, 2008](#)). The ‘Step’ trial randomized 1836 HIV-negative men to receive either the Merck Adenovirus 5 (Ad5) vaccine (MRKAd5) or placebo. Women were also enrolled, but only one became HIV infected. Of the 1836 men, 87 acquired HIV infection. The sequencing labs attempted to derive single-genome-amplification HIV sequences from all 87 subjects, and were only successful for 65. The 22 men with no sequence data were excluded from the analysis, and therefore for the method to provide consistent estimation the missing completely at random (MCAR) assumption is needed. While MCAR seems reasonable in the application (partial support comes from an analysis of the predictors of missing sequence data, which uncovered no significant predictors), it cannot be assured. [Sun and Gilbert \(2012\)](#) developed an augmented inverse probability-weighted approach to provide consistent estimation and inference in the mark-specific proportional hazards model under a missing at random (MAR) assumption, but this method considered only a univariate mark variable. It would be of interest to extend their method to the multivariate mark setting. Alternatively, it would be of interest to extend the multiple imputation approach of [Lu and Tsiatis \(2001\)](#) from the discrete mark setting to the continuous mark setting. The randomization was stratified by whether a volunteer had pre-existing immunity to the Ad5 vector that was used in the vaccine, defined by an Ad5 neutralization titer greater than 200. The method is implemented accounting for these two strata, where 1058 (778) volunteers had Ad5 below (above) 200. In the data analysis, we focus on men only.

For those with  $Ad5 \leq 200$ , there were 54 total infections: 29 of 522 in the vaccine group (with 7 missing marks) and 25 of 536 in the placebo group (with 8 missing marks). The annual incidences were 5.6% for the vaccine group and 4.7% for the placebo group. For  $Ad5 > 200$ , there were 33 total infections: 24 of 392 in the vaccine group (with 7 missing marks) and 9 of 386 infections in the placebo group (with 0 missing marks). The annual incidences were 6.1% for the vaccine group and 2.3% for the placebo group.

Two factors motivate the need to consider multiple protein sequence distances for the sieve analysis (Gilbert, 2000). First, the MRKAd5 vaccine contained three HIV-1 proteins: Gag, Pol, and Nef. Each of these proteins stimulates immune responses in vaccine recipients that could plausibly make the vaccine effect on infection vary with genetic distance to the protein. As a control, it is also of interest to evaluate genetic distances to all of the HIV-1 vaccine proteins combined not included in the vaccine, Env-Rev-Tat-Vif-Vpr-Vpu, for which the central HXB2 reference strain was used. Because the vaccine should not be able to stimulate immune responses to these proteins, it is expected that the vaccine effect on infection would not vary with this control genetic distance. Secondly, as described in greater detail in the clinical paper (Rolland and others, 2011), two different bioinformatics methods, NetMHC (Buus and others, 2003) and Epipred (Heckerman and others, 2007), were used to predict, for each infected subject, the set of HIV peptides in the reference sequence that could potentially bind to at least one of his/her HLA alleles. These peptides are the ones that could potentially constitute T-cell epitopes and hence contribute to a vaccine effect on infection.

NetMHC predicts quantitative binding of peptides to four-digit HLA alleles, whereas Epipred identifies known and potential HIV-1 CTL epitope motifs using two-digit HLA information. In addition to the known HLA-restricted epitopes previously reported at the Los Alamos National Laboratory HIV database (HIVDB), we accepted all epitope motifs with a posterior probability of  $> 0.8$ . HLA-specific epitopes were predicted in the Gag-Pol-Nef protein sequence contained in the tested vaccine (the ‘Step’ reference sequence) and in all proteins from the HXB2 reference sequence (available at the HIVDB, <http://www.hiv.lanl.gov>). Using Epipred, the first step in computing the distance between a subject’s sequences and the reference sequence is to compute the nonparametric maximum likelihood estimate (NPMLE) of the number of peptides shared between the reference sequence and the subject’s sequences, defined as the sum of estimated epitope-probabilities across all 8-, 9-, 10-, 11-mers in the reference sequence that are exactly matched in all of the subject’s sequences. Then, the distance is the NPMLE of the percent of peptides mismatched in at least one of the subject’s sequences, defined as one minus the ratio of the NPMLE of the number of shared peptides (computed in the first step) and the NPMLE of the number of peptides in the reference sequence. Because the NetMHC software categorizes quantitative binding readouts as nonbinder, weak binder, or strong binder, we defined the distance as the estimated percent of epitopes mismatched in at least one of the subject’s sequences, the latter defined as the number of weak or strong binding 8-, 9-, 10-, or 11-mers in the reference sequence that mismatch the corresponding peptide in at least one of the subject’s sequences.

We consider distances defined using the reference HIV-1 regions Gag, Pol, Nef, Env-Rev-Tat-Vif-Vpr-Vpu, as well as using the two bioinformatics prediction methods. The five selected marks,  $\{v_j, j = 1, \dots, 5\}$ , are listed in Table 1. We analyze the data with the mark-specific proportional hazards model (2.1), where the covariate  $z$  is the indicator for the treatment, with  $z = 1$  for the vaccine group and 0 for placebo group. The  $K = 2$  strata are the  $Ad5 \leq 200$  and  $Ad5 > 200$  groups. We use the standardized marks in the analysis, which are obtained by subtracting the minimum and dividing the range of the measurements for each mark. We perform the backwards model selection procedure by starting the analysis with all five marks in the model (2.1) with  $\beta(v) = \beta_0 + \sum_{j=1}^5 \beta_j v_j$ . The marks with nonsignificant coefficients are eliminated one at a time, each time the mark with the least significant  $p$ -value is eliminated. The backward selection procedure retains the last two marks  $v_1$  and  $v_2$ .

For the remaining two marks  $v_1$  and  $v_2$ , we fit the model (2.1) with  $\beta(v) = \beta_0 + \beta_1 v_1 + \beta_2 v_2 + \beta_{12} v_1 v_2$ . The goodness of fit of model (2.1) with  $\beta(v)$  given by (2.2) is conducted following the procedure given

Table 1. *The specifications of the five selected marks based on the step data*

Mark	Type	Ref <sup>†</sup>	Protein
$v_1$	Epipred	HXB2	Env-Rev-Tat-Vif-Vpr-Vpu
$v_2$	NetMHC	HXB2	Env-Rev-Tat-Vif-Vpr-Vpu
$v_3$	Epipred	Step	Gag
$v_4$	Epipred	Step	Pol
$v_5$	Epipred	Step	Nef

<sup>†</sup>The reference HIV-1 sequence, either a sequence represented in the vaccine tested in the Step trial (Step) or a commonly studied central HIV reference sequence that was not in the vaccine (HXB2).

Table 2. *The  $p$ -values of the LRT, Wald test, and score test for testing  $H_{10}$ ,  $H_{20}$ , and  $H_{30}$  for the bivariate marks  $(v_1, v_2)$  for the step data*

Mark selection	Hypothesis	LRT	Wald	Score
$(v_1, v_2)$	$H_{10}$	0.0357	0.0728	0.0468
	$H_{20}$	0.0455	0.0779	0.0586
	$H_{30}$	0.0386	0.0857	0.0483
	$H_{40}$	0.0152	0.0329	0.0205

in Section 3 with the test statistic  $S = \sup_{1 \leq k \leq K} \sup_{t, v, z} |W_k(t, v, z)|$ . We set  $f_k(z) = 1$  and  $K = 2$ . The  $p$ -value is calculated as the percentage of 500 copies of  $S^* = \sup_{1 \leq k \leq K} \sup_{t, v, z} |W_k^*(t, v, z)|$  greater than  $S$ . The  $p$ -value for the bivariate mark selection is 0.56. These result supports that the model fit is adequate. The estimates  $(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_{12})$  are  $(3.43, -3.40, -10.61, 13.90)$ . The standard errors of  $(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_{12})$  are  $(1.52, 2.96, 3.63, 5.79)$ . Table 2 shows the  $p$ -values of the LRT, Wald test (Wald), and score test (Score) for testing  $H_{10}$ – $H_{40}$  described in Section 2.4. All four hypotheses are rejected with borderline significance level near 0.05, with consistent  $p$ -values across the three tests.

Because the overall infection rate was higher in the vaccine than placebo group (Cox model RR = 1.50, 95% CI 0.95–2.41,  $p = 0.06$ ), we estimate the mark-specific RR as  $RR(v_1, v_2) = \exp(\beta_1(v_1, v_2))$  instead of the mark-specific vaccine efficacy; thus for the step trial the analysis focuses on whether and how the vaccine selectively increased the susceptibility to HIV infection. Based on the  $\beta$  point estimates, the estimated RR,  $\widehat{RR}(v_1, v_2)$  (its standard error), equals 3.27 (3.14), 0.91 (1.63), and 3.08 (11.21) for  $(v_1, v_2) = (0.2, 0.2)$ ,  $(0.5, 0.5)$ , and  $(0.8, 0.8)$ , respectively. Hence the 95% confidence intervals for  $RR(v_1, v_2)$  is  $(-2.88, 9.42)$ ,  $(-2.28, 4.10)$ , and  $(-18.89, 25.05)$  at  $(v_1, v_2) = (0.2, 0.2)$ ,  $(0.5, 0.5)$ , and  $(0.8, 0.8)$ , respectively. The interpretation is that the vaccine had no effect on infection against HIV-1 when both distances are in the range from 0.2 to 0.8.

## 5. SIMULATION STUDY

### 5.1 Numerical performance of estimation and hypothesis testing procedures

In this section, we conduct a simulation study to examine finite sample properties of the proposed estimation and hypothesis testing procedures. The simulations are set up to mimic the step data from the HIV vaccine efficacy trial conducted in North and South America (Buchbinder and others, 2008). We take  $K = 2$  strata corresponding to two levels of neutralization titer Ad5. Two sample sizes  $n_k = 250$  and 400 are considered for each stratum. The baseline hazard functions  $\lambda_{0k}(t, v)$  are set as constants  $\lambda_{0k}$ , with  $\lambda_{01} = 0.4$

Table 3. *Summary statistics and coverage probabilities of the 95% confidence intervals for  $(\beta_0, \beta_1, \beta_2, \beta_{12})$  under models  $M_{10}$ ,  $M_{20}$ , and  $M_{30}$*

Model	$(n_1, n_2)$	Coefficient	Bias	SSE	ESE	CP
$M_{10}$	(250, 250)	$\beta_0$	-0.0185	0.6904	0.6757	0.960
		$\beta_1$	0.0125	1.2292	1.1744	0.959
		$\beta_2$	0.0169	1.2342	1.1817	0.950
		$\beta_{12}$	-0.0226	2.1348	2.0509	0.951
	(400, 400)	$\beta_0$	-0.0235	0.5244	0.5275	0.953
		$\beta_1$	0.0480	0.8983	0.9154	0.958
		$\beta_2$	0.0283	0.9399	0.9145	0.958
		$\beta_{12}$	-0.0619	1.5955	1.5883	0.956
$M_{20}$	(250, 250)	$\beta_0$	-0.0210	0.7824	0.7884	0.964
		$\beta_1$	0.0174	1.2611	1.2736	0.968
		$\beta_2$	0.0100	1.2865	1.2824	0.954
		$\beta_{12}$	0.0005	2.0566	2.0802	0.958
	(400, 400)	$\beta_0$	-0.0340	0.6226	0.6171	0.957
		$\beta_1$	0.0211	0.9909	0.9945	0.962
		$\beta_2$	0.0308	1.0004	1.0020	0.962
		$\beta_{12}$	0.0080	1.6225	1.6216	0.963
$M_{30}$	(250, 250)	$\beta_0$	-0.0428	0.8601	0.8208	0.957
		$\beta_1$	0.0686	1.3330	1.2942	0.957
		$\beta_2$	-0.0021	1.4785	1.4268	0.960
		$\beta_{12}$	-0.0177	2.2907	2.4680	0.961
	(400, 400)	$\beta_0$	-0.0179	0.6560	0.6356	0.946
		$\beta_1$	0.0043	1.0254	1.0026	0.946
		$\beta_2$	-0.0703	1.1276	1.1045	0.950
		$\beta_{12}$	0.0816	1.7516	1.7407	0.954

and  $\lambda_{02} = 0.6$ . The covariate  $Z_k$  is a Bernoulli random variable with  $Z_k = 1$  corresponding to the vaccine group and  $Z_k = 0$  for the placebo group. We take  $P(Z_k = 1) = 0.5$ . The censoring time  $C_k$  follows an exponential distribution, independent of  $(T_k, V_k)$ . We consider the following selections of  $(\beta_0, \beta_1, \beta_2, \beta_{12})$  under (2.1) and (2.2) to examine the proposed estimation and hypothesis testing procedures:

$$\begin{aligned}
 M_{10} &: (-0.08, 0.0, 0.0, 0.0); & M_{11} &: (-1.65, 0.9, 0.8, 0.0); & M_{12} &: (-1.65, 0.9, 0.8, 0.6); \\
 M_{20} &: (-1.65, 0.9, 0.8, 0.0); & M_{21} &: (-3.5, 0.3, 0.1, 5.5); & M_{22} &: (-3.5, 0.3, 0.1, 6.0); \\
 M_{30} &: (-1.65, 1.2, 0.0, 0.0); & M_{31} &: (-1.65, 0.6, 0.8, 1.0); & M_{32} &: (-1.65, 0.4, 0.9, 1.2).
 \end{aligned}$$

Here  $M_{10}$  is a model under the null hypothesis  $H_{10}$  which states that  $\beta(v)$  does not depend on either mark  $v_1$  or  $v_2$ . Under  $M_{11}$ , the mark-specific RR depends on both marks but not the interaction of these marks. Under  $M_{12}$ , the mark-specific RR depends on both marks and their interaction.  $M_{20}$  is a model under the null hypothesis  $H_{20}$  where only  $\beta_{12}$  is zero, and  $M_{30}$  is a model under the null hypothesis  $H_{30}$  where  $\beta(v)$  does not depend on  $v_2$ . The overall RRs of the vaccine versus the placebo in most of the settings are around 50%. The percentage of censoring in all settings is about 70% at the end of follow-up  $\tau = 2$ .

Table 3 lists the bias, the sample standard error of the estimates (SSE), the average of the estimated standard errors (ESE), and the coverage probability (CP) for  $\beta$  for sample size  $n_k = 250$  and 400. Table 3 shows that the biases of the estimators are small, ESE approximates SSE well, and the coverage probabilities are very close to their nominal level of 0.95. Table 4 summarizes the empirical sizes and powers of the LRT, Wald test, and the score test for testing  $H_{10}$ ,  $H_{20}$ , and  $H_{30}$  at the significance level  $\alpha = 0.05$ .

Table 4. The empirical sizes and powers of the LRT, Wald test, and score test for testing the null hypotheses  $H_{10}$ ,  $H_{20}$ , and  $H_{30}$  at the significance level 0.05

Model	$n_1(Vx, Plb)$	$n_2(Vx, Plb)$	LRT	Wald	Score
Testing $H_{10} : \beta_1 = \beta_2 = \beta_{12} = 0$					
$M_{10}$	250 (24,46)	250 (34,60)	6.4	4.8	6.2
	400 (39,73)	400 (54,97)	4.7	3.6	4.4
$M_{11}$	250 (25,46)	250 (35,60)	56.4	53.2	54.9
	400 (41,74)	400 (51,97)	74.1	73.2	73.8
$M_{12}$	250 (30,46)	250 (42,60)	76.4	72.9	75.9
	400 (49,74)	400 (67,97)	91.7	91.2	91.5
Testing $H_{20} : \beta_{12} = 0$					
$M_{20}$	250 (25,46)	250 (36,60)	4.9	4.2	4.6
	400 (41,74)	400 (57,96)	3.9	3.8	3.8
$M_{21}$	250 (23,46)	250 (32,60)	34.0	39.3	38.2
	400 (37,74)	400 (52,97)	51.0	55.2	54.2
$M_{22}$	250 (30,46)	250 (41,60)	40.0	43.3	42.0
	400 (48,74)	400 (67,97)	53.7	57.6	56.8
Testing $H_{30} : \beta_2 = \beta_{12} = 0$					
$M_{30}$	250 (20,46)	250 (29,60)	5.6	4.8	5.5
	400 (33,74)	400 (47,97)	3.9	3.6	3.9
$M_{31}$	250 (29,46)	250 (41,60)	62.8	60.2	61.8
	400 (48,74)	400 (65,96)	83.0	82.1	82.6
$M_{32}$	250 (30,46)	250 (41,60)	74.6	73.3	74.2
	400 (48,74)	400 (66,97)	91.3	90.9	91.2

The average numbers of cases for the vaccine (Vx) and placebo (Plb) groups for are also reported. Each entry of Tables 3 and 4 is calculated based on 1000 simulations. Table 4 shows that the empirical sizes of all these tests are close to 0.05. The LRT has better power than the Wald and score tests. The powers of the tests increase as the simulation model moves from  $M_{11}$  to  $M_{12}$ , representing an increased departure from  $H_{10}$ . Similarly, the powers of these tests for testing  $H_{20}$  increase in the direction from  $M_{21}$  to  $M_{22}$ . The powers for testing  $H_{30}$  also increase in the direction from  $M_{31}$  to  $M_{32}$ , and the powers increase as the sample size increases. The coverage probabilities for  $\beta_0$ ,  $\beta_1$ ,  $\beta_2$ , and  $\beta_{12}$  are also listed to demonstrate that the proposed maximum partial likelihood methods work very well.

### 5.2 Numerical performance of goodness-of-fit procedure

In this section, we conduct a simulation study to check the finite sample performance of the proposed goodness-of-fit testing procedure based on the test statistic  $S = \sup_{1 \leq k \leq K} \sup_{t, v, z} |W_k(t, v, z)|$  defined in (3.3). We set  $f_k(z) = 1$  and  $K = 2$ . The size of the test is examined using the following simple stratified mark-specific proportional hazards model:

$$M_{40} : \lambda_k(t, v|z) = \lambda_{0k} \exp\{(\beta_0 + \beta_1 v_1 + \beta_2 v_2 + \beta_{12} v_1 v_2)z\}, \quad k = 1, 2, \quad (5.1)$$

where  $\lambda_{01} = 0.4$ ,  $\lambda_{02} = 0.6$  and  $(\beta_0, \beta_1, \beta_2, \beta_{12}) = (-1.65, 0.4, 0.9, 1.2)$ . The covariate  $Z_{ki}$  is a Bernoulli random variable with  $P(Z_{ki} = 1) = 0.5$ . The censoring times are generated from an exponential distribution, independent of  $(T_{ki}, V_{ki})$ . The first block in Table 5 shows the empirical sizes of the test for sample

Table 5. *The empirical sizes and powers of goodness-of-fit tests at the significance level 0.05*

Model	$n_1(Vx, Plb)$	$n_2(Vx, Plb)$	Size/power
$M_{40}$	50 (5,9)	50 (8,12)	6.2
	100 (12,18)	100 (16,24)	5.2
	150 (18,27)	150 (25,36)	6.0
$M_{41}$	50 (4,8)	50 (8,13)	49.8
	100 (9,16)	100 (16,27)	79.0
	150 (14,24)	150 (25,40)	87.4
$M_{42}$	50 (5,10)	50 (10,15)	61.0
	100 (11,20)	100 (20,31)	87.0
	150 (17,30)	150 (31,47)	92.4
$M_{43}$	50 (7,12)	50 (11,17)	69.4
	100 (13,24)	100 (23,35)	92.4
	150 (21,36)	150 (36,52)	97.0

sizes of  $n = n_1 + n_2 = 100, 200,$  and  $300$ . The empirical sizes are calculated based on 1000 simulations and 500 Gaussian multiplier samples, and they are very close to the 0.05 nominal level.

To evaluate the power of proposed test, we consider the model

$$\lambda_k(t, v|z) = \lambda_{0k}(t, v) \exp\{-at^b + 0.4v_1 + 0.6v_2\}z, \quad k = 1, 2, \quad (5.2)$$

where  $\lambda_{0k}(t, v) = abkt^{(b-1)} \exp\{at^b - 0.4v_1 - 0.6v_2\}$ , and  $a$  and  $b$  are the parameters whose values are to be selected. Again we take  $Z_{ki}$  as a Bernoulli random variable with  $P(Z_{ki} = 1) = 0.5$ . Model (5.2) is not a mark-specific proportional hazards model since the relative hazard ratio  $\lambda_k(t, v|z = 1)/\lambda_k(t, v|z = 0) = \exp\{at^b - 0.4v_1 - 0.6v_2\}$  changes with time. We consider the alternative models specified by the choices of  $(a, b)$ , defined as  $M_{41} : (a, b) = (0.20, 0.30)$ ;  $M_{42} : (a, b) = (0.25, 0.40)$ ; and  $M_{43} : (a, b) = (0.30, 0.50)$ . As  $a$  and  $b$  increase, the rate of change with respect to  $t$  of the relative hazard ratio increases, which represents an increased departure from the null hypothesis. For each of the models  $M_{41}$ ,  $M_{42}$ , and  $M_{43}$ , random right censoring times are generated from an exponential distribution, independent of  $(T_{ki}, V_{ki})$ . The empirical powers of the test at the significance level 0.05 under the models  $M_{41}$ ,  $M_{42}$ , and  $M_{43}$  at sample sizes  $n = n_1 + n_2 = 100, 200,$  and  $300$  are given in Table 5. Each entry of the table is based on 1000 simulations and 500 Gaussian multiplier samples. The power of the test increases with increasing  $(a, b)$ , and also increases with sample size. The limited simulation study demonstrates the validity of the proposed goodness-of-fit testing procedure. The test provides a valuable tool to check the adequacy of the stratified mark-specific proportional hazards model specified in (2.1) and (2.2).

#### SUPPLEMENTARY MATERIAL

Supplementary material is available at <http://biostatistics.oxfordjournals.org>.

#### ACKNOWLEDGEMENTS

The authors thank the referees for their valuable inputs that have improve the paper. They also thank the HIV Vaccine Trials Network (HVTN) and Merck for providing the data analyzed in this article. The

HVTN is supported through a cooperative agreement with the National Institutes of Health Division of AIDS. *Conflict of Interest*: None declared.

#### FUNDING

This research was partially supported by the National Institutes of Health (grant number R37 AI054165-09). The research of Y.S. was also partially supported by the National Science Foundation (grant number DMS-0905777).

#### REFERENCES

- BUCHBINDER, S. P., MEHROTRA, D. V., DUERR, A., FITZGERALD, D. W., MOGG, R., LI, D., GILBERT, P. B., LAMA, J. R., MARMOR, M., DEL RIO, C. *and others* (2008). Efficacy assessment of a cell-mediated immunity HIV-1 vaccine (the Step Study): a double-blind, randomised, placebo-controlled, test-of-concept trial. *Lancet* **372**, 1881–1893.
- BUUS, S., LAUEMOLLER, S. L., WORNING, P., KESMIR, C., FRIMURER, T., CORBET, S., FOMSGAARD, A., HILDEN, J., HOLM, A., BRUNAK, S. (2003). Sensitive quantitative predictions of peptide-MHC binding by a ‘Query by Committee’ artificial neural network approach. *Tissue Antigens* **62**, 378–384.
- DABROWSKA, D. M. (1997). Smoothed Cox regression. *The Annals of Statistics* **25**, 1510–1540.
- FLYNN, N. M., FORTHAL, D. N., HARRO, C. D., JUDSON, F. N., 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.
- 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., MCKEAGUE, I. W. AND SUN, Y. (2004). Tests for comparing mark-specific hazards and cumulative incidence functions. *Lifetime Data Analysis* **10**, 5–28.
- GILBERT, P. B., MCKEAGUE, I. W. AND SUN, Y. (2008). The two-sample problem for failure rates depending on a continuous mark: an application to vaccine efficacy. *Biostatistics* **9**, 263–276.
- GRAY, G. E., BEKKER, L., CHURCHYARD, G. J., NCHABELENG, M., MLISANA, K., DE BRUYN, G., ROUX, S., MATHEBULA, M., LATKA, M., BENNIE, T. *and others* (2009). Did unblinding affect HIV risk behaviour and risk perception in the HVTN503/Phambili study? *Retrovirology* **6** (Suppl 3), 209.
- HECKERMAN, D., KADIE, C. AND LISTGARTEN, J. (2007). Leveraging information across HLA alleles/supertypes improves epitope prediction. *Journal of Computational Biology* **14**, 736–746.
- HUANG, Y. AND LOUIS, T. A. (1998). Nonparametric estimation of the joint distribution of survival time and mark variables. *Biometrika* **85**, 785–798.
- KALBFLEISCH, J. D. AND PRENTICE, R. L. (1980). *The Statistical Analysis of Failure Time Data*. New York: Wiley.
- 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.
- LU, K. AND TSIATIS, A. A. (2001). Multiple imputation methods for estimating regression coefficients in the competing risks model with missing cause of failure. *Biometrics* **57**, 1191–1197.
- PITISUTTITHUM, P., GILBERT, P., GURWITH, M., HEYWARD, W., MARTIN, M., VAN GRIENSVEN, F., HU, D., TAPPERO, J. W., CHOOPANYA, K. (2006). Randomized, double-blind, placebo-controlled efficacy trial of a bivalent recombinant glycoprotein 120 HIV-1 vaccine among injection drug users in Bangkok, Thailand. *Journal of Infectious Diseases* **194**, 1661–1671.
- PRENTICE, R. L., KALBFLEISCH, J. D., PETERSON, A. V., FLUORNOY, N., FAREWELL, V. T. AND BRESLOW, N. E. (1978). The analysis of failure time in the presence of competing risks. *Biometrics* **34**, 541–554.

- RERKS-NGARM, S., PITISUTTITHUM, P., NITAYAPHAN, S., KAEWKUNGWAL, J., CHIU, J., PARIS, R., PREMSRI, N., NAMWAT, C., DE SOUZA, M., ADAMS, E. *and others* (2009). Vaccination with ALVAC and AIDSVAX to prevent HIV-1 infection in Thailand. *The New England Journal of Medicine* **361**, 2209–2220.
- ROLLAND, M., TOVANABUTRA, S., DE CAMP, A. C., FRAHM, N., GILBERT, P., SANDERS-BUELL, E., HEATH, L., MAGARET, C. A., BOSE, M., BRADFIELD, A. *and others* (2011). Genetic impact of vaccination on breakthrough HIV-1 sequences from the Step trial. *Nature Medicine* **17**, 366–371.
- SCHEIKE, T., ZHANG, M. AND GERDS, T. A. (2008). Predicting cumulative incidence probability by direct binomial regression. *Biometrika* **95**, 205–220.
- SERFLING, R. J. (1980). *Approximation Theorems of Mathematical 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.
- SUN, Y. AND GILBERT, P. B. (2012). Estimation of stratified mark-specific proportional hazards models with missing marks. *Scandinavian Journal of Statistics*, **39**, 34–52.
- SUN, Y., GILBERT, P. B. AND MCKEAGUE, I. W. (2009). Proportional hazards models with continuous marks. *The Annals of Statistics* **37**, 394–426.
- SUN, Y. AND WU, H. (2005). Semiparametric time-varying coefficients regression model for longitudinal data. *Scandinavian Journal of Statistics* **32**, 21–47.
- VAN DER VAART, A. W. (1998). *Asymptotic Statistics*. Cambridge: Cambridge University Press.

[Received December 1, 2011; revised April 6, 2012; accepted for publication May 22, 2012]