

# Testing mark-specific vaccine efficacy with missing marks

Peter B. GILBERT<sup>1</sup> and Yanqing SUN<sup>2,3</sup>

<sup>1</sup>University of Washington and Fred Hutchinson Cancer Research Center,  
Seattle, WA, USA

<sup>2</sup>University of North Carolina at Charlotte, Charlotte, NC, USA

<sup>3</sup>Corresponding author: Yanqing Sun, e-mail: yasun@uncc.edu

## Abstract

This article develops hypothesis testing procedures for the stratified mark-specific proportional hazards model in the presence of missing marks. The motivating application is placebo-controlled preventive human immunodeficiency virus (HIV) vaccine efficacy trials, which have objective to test if the mark-specific relative hazard rate (vaccine versus placebo) is unity for all mark values, and to test whether it changes with the mark (the mark is the genetic distance of an infecting HIV sequence to an HIV sequence represented inside the tested vaccine). These tests inform on whether the vaccine affects the rate of HIV infection for any HIV genotype and whether the vaccine effect differs by HIV genotype, respectively, and guide vaccine development. One difficulty with these assessments is that the mark may be missing from many HIV infected subjects, predominantly due to rapid evolution of the infecting HIV. The test statistics are constructed based on a two-stage efficient estimator which utilizes auxiliary predictors of the missing marks. The asymptotic properties of the testing procedures are investigated. In addition, their finite-sample performances are investigated in simulations, which verify the double-robustness property under missing at random marks and demonstrate effectiveness of the predictive auxiliaries to recover efficiency. One of the simulations models the recent landmark trial in Thailand, which was the first trial to demonstrate partial efficacy of an HIV vaccine. The new methods are applied to the real data set.

**Keywords:** Auxiliary marks, competing risks failure time, genetic data, HIV vaccine efficacy trial, augmented inverse probability weighted complete-case estimator, mark-specific vaccine efficacy

## 1 Introduction

The primary objective of a preventive HIV vaccine efficacy trial is to assess vaccine efficacy ( $VE$ ) to prevent HIV infection, where typically  $VE$  is defined as one minus the hazard ratio (vaccine/placebo) of HIV infection diagnosis. However, the great genetic variability of HIV poses a central challenge to developing a highly efficacious HIV vaccine. The trial population is exposed to many HIV genotypes but the vaccine only contains a few, which represent particular HIV sequences isolated from infected individuals, and the vaccine is less likely to protect against HIVs with greater genetic distance from the sequences inside the vaccine

(Gilbert, Lele, and Vardi, 1999). Moreover, the trial has objectives to assess whether and how the vaccine impacts the infection rate with any HIV genotype and whether and how the vaccine effect varies by HIV genotype; assessment of this objective as been named, ‘sieve analysis’ (Gilbert, Self, and Ashby, 1998). Gilbert, McKeague and Sun (2008), Sun, Gilbert and McKeague (2009), and Sun and Gilbert (2012) developed sieve analysis methods using the competing risks failure time framework (Prentice et al., 1978), with the competing risk treated not as a discrete categorical variable as usual but rather as a continuous spectrum of genotypes. In particular, they attached a continuous ‘mark’ variable to the infecting genotype that measures the genetic distance of an infecting HIV sequence to a sequence inside the vaccine. The goal of the sieve analysis methods is evaluation of mark-specific vaccine efficacy, defined as one minus the mark-specific hazard ratio (vaccine/placebo) of infection.

The Gilbert, McKeague, and Sun (2008) and Sun, Gilbert, and McKeague (2009) methods assumed no missing mark data in infected subjects, whereas the Sun and Gilbert (2012) paper allowed missing at random (MAR) missing marks. In practice there are missing marks, for example in the Vax004 trial 32 of 368 infected subjects had no HIV sequence data (Gilbert, McKeague, and Sun, 2008), due to drop-out or to inability of the HIV sequencing technology to measure the infecting HIV sequence. Sun and Gilbert (2012) is the only paper on sieve analysis that accommodates missing continuous marks. Sun and Gilbert (2012) restricted attention to estimation methods, and this article is a sequel that develops corresponding inferential/hypothesis testing methods based on the augmented IPW estimator.

## 2 The stratified mark-specific proportional hazards (PH) model

Let  $T$  be the failure time,  $V$  a continuous mark variable with bounded support  $[0, 1]$ , and  $Z(t)$  a possibly time-dependent  $p$ -dimensional covariate. Under the competing risks model, the mark  $V$  is only defined and observable when  $T$  is observed, whereas if  $T$  is right-censored, the mark is undefined and meaningless. Suppose that the conditional mark-specific hazard function at time  $t$  given the covariate history  $Z(s)$ , for  $s \leq t$ , only depends on the current value  $Z(t)$ . Let  $\lambda(t, v|z)$  be the conditional mark-specific hazard function at  $(T, V) = (t, v)$  given  $Z(t) = z$  defined by Sun, Gilbert and McKeague (2009). We consider the stratified mark-specific proportional hazards (PH) model

$$\lambda_k(t, v|z(t)) = \lambda_{0k}(t, v) \exp \{ \beta(v)^T z(t) \}, \quad k = 1, \dots, K, \quad (1)$$

where  $\lambda_k(t, v|z(t))$  is the conditional mark-specific hazard function given covariate  $z(t)$  for an individual in the  $k$ th stratum,  $\lambda_{0k}(\cdot, v)$  is the unspecified baseline hazard function for the  $k$ th stratum,  $\beta(v)$  is the  $p$ -dimensional unknown regression coefficient function of  $v$ , and  $K$  is the number of strata. Model (1) allows different baseline functions for different strata. Sun and Gilbert (2012) developed estimation procedures for model (1) that incorporate auxiliary covariates and/or auxiliary mark variables that inform about the probability  $V$  is observed and about the distribution of  $V$ .

Arranging the first component of  $z(t)$  to be the treatment (vaccine) group indicator and letting  $\beta_1(v)$  be the corresponding regression coefficient, the covariate and stratum adjusted mark-specific vaccine efficacy equals  $VE(v) = 1 - \exp(\beta_1(v))$ . The current article develops parallel hypothesis testing procedures to assess the vaccine efficacy  $VE(v)$  as a function of  $v$ . The two objectives are to assess if the vaccine efficacy ever deviates from

0 (i.e., test  $VE(v) = 0$ ) and to assess if the vaccine efficacy changes with the mark (i.e., test  $VE(v) = VE$ ). Gilbert, McKeague, and Sun (2008) provide additional discussion on the value of testing these hypotheses.

### 3 Testing mark-specific vaccine efficacy

#### 3.1 Missing data assumptions

The right-censored mark-specific failure time is represented by  $(X, \delta, \delta V)$  and  $Z(\cdot)$  is the covariate process, where  $X = \min(T, C)$ ,  $T$  is the failure time of interest,  $C$  is the censoring time,  $\delta$  is the indicator of observed failure and  $V$  the mark variable. Let  $R$  be the indicator of whether all possible data are observed for a subject;  $R = 1$  if either  $\delta = 0$  (right-censored) or if  $\delta = 1$  and  $V$  is observed; and  $R = 0$  otherwise. Auxiliary variables  $A$  may be helpful for predicting missing marks. Since the mark can only be missing for failures, supplemental information is potentially useful only for failures, for predicting missingness and for informing about the distribution of missing marks.

We assume that the censoring time  $C$  is conditionally independent of  $(T, V)$  given  $Z(\cdot)$  for an individual in the  $k$ th stratum. We also assume the mark  $V$  is MAR (Rubin, 1976); that is, given  $\delta = 1$  and  $W = (T, Z(T), A)$  of an individual in the  $k$ th stratum, the probability that the mark  $V$  is missing depends only on the observed  $W$ , not on the value of  $V$ ; this assumption is expressed as

$$r_k(W) \equiv P(R = 1 | \delta = 1, W) = P(R = 1 | V, \delta = 1, W). \tag{2}$$

Let  $\pi_k(Q) = P(R = 1 | Q)$  where  $Q = (\delta, W)$ . Then  $\pi_k(Q) = \delta r_k(W) + (1 - \delta)$ . The MAR assumption (2) also implies that  $V$  is independent of  $R$  given  $Q$ :

$$\rho_k(v, W) \equiv P(V \leq v | \delta = 1, W) = P(V \leq v | R = 1, \delta = 1, W). \tag{3}$$

For an observed value  $w$  of  $W$  of an individual in the  $k$ th stratum, we write  $r_k(w) = P(R = 1 | \delta = 1, W = w)$  and  $\rho_k(v, w) = P(V \leq v | \delta = 1, W = w)$ . The stratum-specific definitions of  $r_k(w)$  and  $\rho_k(v, w)$  leave the options for the models of the probability of complete-case and mark distribution to be different for different strata.

Suppose that  $\tau$  is the end of the follow-up period. Let  $n_k$  be the number of subjects in the  $k$ th stratum; the total sample size is  $n = \sum_{k=1}^K n_k$ . Let  $\{X_{ki}, Z_{ki}(\cdot), \delta_{ki}, R_{ki}, \delta_{ki}V_{ki}, A_{ki}; i = 1, \dots, n_k\}$  be iid replicates of  $\{X, Z(\cdot), \delta, R, \delta V, A\}$  from the  $k$ th stratum. The observed data are denoted by  $\{O_{ki}; i = 1 \dots, n_k, k = 1, \dots, K\}$ , where  $O_{ki} = \{X_{ki}, Z_{ki}(\cdot), R_{ki}, \delta_{ki}V_{ki}, A_{ki}\}$  for  $\delta_{ki} = 1$  and  $O_{ki} = \{X_{ki}, Z_{ki}(\cdot), R_{ki} = 1\}$  for  $\delta_{ki} = 0$ . We assume that  $\{O_{ki}; i = 1 \dots, n_k, k = 1, \dots, K\}$  are independent for all subjects.

#### 3.2 Hypotheses to test

In the context of the vaccine trial application, let  $z(t) = (z_1, z_2^T(t))^T$ , where  $z_1$  is the treatment assignment (1=vaccine; 0=placebo) and  $z_2$  are other related explanatory variables. Let  $\beta(v) = (\beta_1(v), \beta_2^T(v))^T$ , so that  $\beta_1(v)$  is the coefficient for vaccination status and  $\beta_2(v)$  for other covariates. The mark-specific vaccine efficacy  $VE(v)$  can be expressed as  $VE(v) = 1 - \exp(\beta_1(v))$ . Sun and Gilbert (2012) developed procedures for estimating  $\beta(v)$  in model (1) and for constructing pointwise confidence intervals for  $VE(v)$ , for  $0 < v < 1$ .

Here we are interested in testing the following two sets of hypotheses. Let  $[a, b] \subset (0, 1)$ . The first set of hypotheses is

- $H_{10} : \text{VE}(v) = 0$  for  $v \in [a, b]$
- vs.  $H_{1a} : \text{VE}(v) \neq 0$  for some  $v$  (general alternative)
- or  $H_{1m} : \text{VE}(v) \geq 0$  with strict inequality for some  $v$  (monotone alternative).

The second set of hypotheses is

- $H_{20} : \text{VE}(v)$  does not depend on  $v \in [a, b]$
- vs.  $H_{2a} : \text{VE}(v)$  depends on  $v$  (general alternative)
- or  $H_{2m} : \text{VE}(v)$  decreases as  $v$  increases (monotone alternative).

The null hypothesis  $H_{10}$  implies the vaccine affords no protection against any HIV genotype. The alternative  $H_{1m}$  indicates that the vaccine provides protection for at least some of the HIV genotypes, while  $H_{1a}$  states that the vaccine provides protection and/or increased risk for some HIV genotypes. The null hypothesis  $H_{20}$  implies there is no difference in vaccine protection for different HIV genotypes, measured by their distances  $v$  to an HIV sequence represented in the vaccine. The ordered alternative  $H_{2m}$  states that vaccine efficacy decreases with  $v$  and the alternative  $H_{2a}$  indicates that the vaccine efficacy changes with  $v$ . Let  $\beta_1(v)$  be the first component of  $\beta(v)$ . The hypotheses  $H_{10}$  and  $H_{20}$  can be formulated in terms of  $\beta_1(v)$ .

### 3.3 Hypothesis testing procedures

The hypothesis testing procedures concerning the HIV vaccine efficacies are developed based on the augmented inverse probability weighted (AIPW) estimator developed by Sun and Gilbert (2012). Let  $\hat{\beta}^{aug}(v)$  be the AIPW estimator of  $\beta(v)$  for model (1) of Sun and Gilbert (2012). The estimator of the cumulative function  $B(v) = \int_0^v \beta(u) du$  is given by  $\hat{B}^{aug}(v) = \int_0^v \hat{\beta}^{aug}(u) du$ . The covariate-adjusted vaccine efficacy  $\text{VE}(v)$  is defined through the first component of  $\beta(v)$ . Let  $B_1(v)$  the first component of the cumulative coefficient function  $B(v)$ . The hypotheses tests concerning  $\text{VE}(v)$  are constructed based on the first component  $\hat{B}_1^{aug}(v)$  of the AIPW estimator  $\hat{B}^{aug}(v)$ .

Let  $W_B(v) = n^{1/2}\{\hat{B}^{aug}(v) - \hat{B}^{aug}(a)\} - n^{1/2}\{B(v) - B(a)\}$  for  $v \in [a, b]$ . The distribution of  $W_B(v)$ , for  $v \in [a, b]$ , can be approximated by its Gaussian multipliers version  $W_B^*(v)$ ,  $v \in [a, b]$  using the Gaussian multipliers resampling method. Let  $W_{B_1}(v)$  and  $W_{B_1}^*(v)$  be the first component of  $W_B(v)$  and  $W_B^*(v)$ , respectively. With the Gaussian multipliers method, the variance  $\text{Var}\{\hat{B}_1^{aug}(v) - \hat{B}^{aug}(a)\}$  can be consistently estimated by  $\widehat{\text{Var}}\{\hat{B}_1^{aug}(v) - \hat{B}^{aug}(a)\} = n^{-1}\text{Var}^*(W_{B_1}^*(v))$ , where  $\text{Var}^*(W_{B_1}^*(v))$  is the first component on the diagonal of the conditional covariance  $W_B^*(v)$  given the observed data.

#### Testing the null hypothesis $H_{10}$

Consider the test process  $Q^{(1)}(v) = n^{1/2}\{\hat{B}_1^{aug}(v) - \hat{B}_1^{aug}(a)\}$ ,  $v \in [a, b]$  for testing  $H_{10}$ . Then  $Q^{(1)}(v) = W_{B_1}(v) + n^{1/2}\{B_1(v) - B_1(a)\}$ ,  $v \in [a, b]$ . Let  $G(v)$  be the limiting Gaussian process of  $W_{B_1}(v)$ ,  $a \leq v \leq b$ , as  $n \rightarrow \infty$ . Under  $H_{10}$ ,  $B_1(v) - B_1(a) = 0$  for  $v \in [a, b]$ . Hence  $Q^{(1)}(v) \xrightarrow{\mathcal{D}} G(v)$ ,  $v \in [a, b]$ , as  $n \rightarrow \infty$ . Under  $H_{10}$ , the distribution of  $Q^{(1)}(v)$ ,  $v \in [a, b]$ , can be approximated by the conditional distribution of  $W_{B_1}^*(v)$ ,

$v \in [a, b]$ , given the observed data sequence. We propose the following test statistics for testing  $H_{10}$ :  $T_{a1}^{(1)} = \sup_{v \in [a, b]} |Q^{(1)}(v)|$ ,  $T_{a2}^{(1)} = \int_a^b \{Q^{(1)}(v)\}^2 d\text{Var}^* \{W_{B_1}^*(v)\}$ ,  $T_{m1}^{(1)} = \inf_{v \in [a, b]} Q^{(1)}(v)$  and  $T_{m2}^{(1)} = \int_a^b Q^{(1)}(v) d\text{Var}^* \{W_{B_1}^*(v)\}$ .

By the continuous mapping theorem,  $T_{a1}^{(1)} \xrightarrow{\mathcal{D}} \sup_{v \in [a, b]} |G(v)|$ ,  $T_{a2}^{(1)} \xrightarrow{\mathcal{D}} \int_a^b \{G(v)\}^2 d\text{Var}\{G(v)\}$ ,  $T_{m1}^{(1)} \xrightarrow{\mathcal{D}} \inf_{v \in [a, b]} G(v)$ , and  $T_{m2}^{(1)} \xrightarrow{\mathcal{D}} \int_a^b G(v) d\text{Var}\{G(v)\}$  under  $H_{10}$  as  $n \rightarrow \infty$ . The test statistics  $T_{a1}^{(1)}$  and  $T_{a2}^{(1)}$  capture general departures  $H_{1a}$ , while the test statistics  $T_{m1}^{(1)}$  and  $T_{m2}^{(1)}$  are sensitive to the monotone departures  $H_{1m}$ . It is easy to derive that all the test statistics  $T_{a1}^{(1)}$ ,  $T_{a2}^{(1)}$ ,  $T_{m1}^{(1)}$  and  $T_{m2}^{(1)}$  are consistent against their respective alternative hypotheses. The distributions of  $T_{a1}^{(1)}$ ,  $T_{a2}^{(1)}$ ,  $T_{m1}^{(1)}$  and  $T_{m2}^{(1)}$  under  $H_{10}$  can be approximated using the Gaussian multipliers method.

### Testing the null hypothesis $H_{20}$

Let  $Q^{(2)}(v) = (v-a)^{-1}n^{1/2}\{\hat{B}_1^{aug}(v) - \hat{B}_1^{aug}(a)\} - (b-a)^{-1}n^{1/2}\{\hat{B}_1^{aug}(b) - \hat{B}_1^{aug}(a)\}$ . Then for  $a < v \leq b$ ,  $Q^{(2)}(v) = \Gamma(v, W_{B_1}) + n^{1/2}\Gamma(v, B_1)$ , where  $\Gamma(v, F_1) = (v-a)^{-1}\{F_1(v) - F_1(a)\} - (b-a)^{-1}\{F_1(b) - F_1(a)\}$  is a transformation of the function  $F_1(\cdot)$ . We note that  $\Gamma(\cdot, B_1) = 0$  under  $H_{20}$  and  $\Gamma(\cdot, B_1) \neq 0$  under the alternatives. This motivates us to consider  $Q^{(2)}(v)$  as the test process for testing  $H_{20}$  and the following test statistics:  $T_{a1}^{(2)} = \sup_{v \in [a', b]} |Q^{(2)}(v)|$ ,  $T_{a2}^{(2)} = \int_{a'}^b \{Q^{(2)}(v)\}^2 d\text{Var}^* \{W_{B_1}(v)\}$ ,  $T_{m1}^{(2)} = \inf_{v \in [a', b]} Q^{(2)}(v)$  and  $T_{m2}^{(2)} = \int_{a'}^b Q^{(2)}(v) d\text{Var}^* \{W_{B_1}(v)\}$ , where  $a < a' < b$ . We choose  $a' > a$  to avoid zero in the denominator of  $Q^{(2)}(v)$ . In practice, one can choose  $a'$  close to  $a$  to make use of available data and for the tests to be consistent.

Applying the continuous mapping theorem, we have under  $H_{20}$ ,  $T_{a1}^{(2)} \xrightarrow{\mathcal{D}} \sup_{v \in [a', b]} |\Gamma(v, G)|$ ,  $T_{a2}^{(2)} \xrightarrow{\mathcal{D}} \int_{a'}^b \{\Gamma(v, G)\}^2 d\text{Var}\{G(v)\}$ ,  $T_{m1}^{(2)} \xrightarrow{\mathcal{D}} \inf_{v \in [a', b]} \Gamma(v, G)$ , and  $T_{m2}^{(2)} \xrightarrow{\mathcal{D}} \int_{a'}^b \Gamma(v, G) d\text{Var}\{G(v)\}$ , as  $n \rightarrow \infty$ . The tests  $T_{a1}^{(2)}$  and  $T_{a2}^{(2)}$  capture general departures  $H_{2a}$  while the tests  $T_{m1}^{(2)}$  and  $T_{m2}^{(2)}$  are sensitive to the monotone departure  $H_{2m}$ . The test statistics  $T_{m1}^{(2)}$  and  $T_{m2}^{(2)}$  are expected to be negative when  $H_{2m}$  holds. The distributions of  $T_{a1}^{(2)}$ ,  $T_{a2}^{(2)}$ ,  $T_{m1}^{(2)}$  and  $T_{m2}^{(2)}$  under  $H_{20}$  can be approximated using the Gaussian multipliers method.

## 4 Simulation study and data analysis

Our simulation studies show that all of the tests have satisfactory empirical sizes close to the nominal level 0.05. The powers of the tests increase with sample size and they are not overly sensitive to the selected bandwidths. The powers of the tests for testing  $H_{10}$  increase as the model moves in the direction representing the increased departure from the null hypothesis  $H_{10}$ . The powers of the tests for testing  $H_{20}$  increase as the model moves in the direction representing the increased departure from the null hypothesis  $H_{20}$ . Our simulation studies also show that the tests utilizing the auxiliary marks have higher power than those without using the auxiliary marks. The powers are expected to increase with the strength of correlation between the auxiliary marks and the mark of interest. Finally, our simulation studies reflects the double robustness property of the AIPW estimator. The empirical sizes are also close to the nominal level 0.05 when one of  $r_k(w)$  ( the conditional probability of the complete-case indicator  $R_{ki}$ ) and  $g_k(a|t, v, z)$  (the conditional density of  $A_{ki}$  given



$(T_{ki}, Z_{ki}, V_{ki})$ ) is mis-specified. When only  $r_k(w)$  is mis-specified and MAR holds, the empirical powers closely track the corresponding powers under correct model specifications. The empirical powers are lower than those corresponding powers when  $g_k(a|t, v, z)$  is mis-specified or when both  $r_k(w)$  and  $g_k(a|t, v, z)$  are mis-specified. Additional simulations are conducted to gain insight about the power available for the Thai trial.

The method was applied to the Thai trial. In particular, we assessed how the vaccine efficacy against subtype E HIV infection depends on weighted Hamming distance (re-scaled to values between 0 and 1) between the subtype E vaccine-insert sequences and the infecting subtype E HIV. Our method shows a clear significant evidence that the mark-specific vaccine efficacy is greater than 0 for some marks and declines with the mark, in the region  $v \in [0, 0.5]$ . These analyses suggest that the vaccine protected against HIVs closely matched to the vaccine strain HIVs in the monoclonal antibody contact sites, but failed to protect against HIVs with many mismatches in these sites. These results may guide future vaccine research by suggesting to modify future vaccine candidates to include HIV sequences more closely matched to circulating HIVs in the monoclonal antibody contact sites.

### Acknowledgements

The research of Yanqing Sun was partially supported by NSF grants DMS-0905777 and DMS-1208978, NIH NIAID grant 2 R37 AI054165-10. The research of Peter Gilbert was partially supported by NIH NIAID grant 2 R37 AI054165-10.

### References

- 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., McKeague, I. W., 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.
- Gilbert, P. B., Self, S. and Ashby, M. (1998). Statistical methods for assessing differential vaccine protection against human immunodeficiency virus types. *Biometrics* **54**, 799–814.
- 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.
- Robins, J. M., Rotnitzky, A. and Zhao, L. P. (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association* **89**, 846–866.
- 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.