

# Hypotheses Tests of Strain-Specific Vaccine Efficacy Adjusted for Covariate Effects

Seunggeun Hyun and Yanqing Sun  
Department of Mathematics and Statistics  
University of North Carolina at Charlotte  
9201 University City Boulevard  
Charlotte, NC 28223

## Abstract

In the evaluation of efficacy of a vaccine to protect against disease caused by finitely many diverse infectious pathogens, it is often important to assess if vaccine protection depends on variations of the exposing pathogen. This problem can be formulated under a competing risks model where the endpoint event is the infection and the cause of failure is the infecting strain type determined after the infection is diagnosed. The strain-specific vaccine efficacy is defined as one minus the cause-specific hazard ratio (vaccine/placebo). This paper develops inferences for testing if the vaccine affords protection against various strains and if and how the strain-specific vaccine efficacy depends on the type of exposing strain, adjusting for covariate effects. The Cox proportional hazards model is used to relate the cause-specific outcomes to explanatory variables. The finite sample properties of proposed tests are studied through extensive simulations and are shown to have good performances. The tests developed are applied to the data collected from an oral cholera vaccine trial.

*Key words:* Competing risks model; Cause-specific hazard function; Cox proportional hazards model.

# 1 Introduction

In many medical studies, the evaluation of effectiveness of one treatment often involves comparison of survival data of two treatment groups. This is often done through the ratio of two hazard functions of an endpoint event. In a vaccine efficacy trial, the primary objective is to assess vaccine efficacy (VE) to prevent infection where typically VE is defined as one minus the hazard ratio (vaccine/placebo) of infection. However, it may be quite difficult to achieve an efficacious vaccine due to presence of diverse pathogens. An individual may be exposed to many types of circulating infecting strains and the occurrence of disease is only attributed to one of the exposing strains. The vaccine protection is likely to vary for different exposing strains. When the disease is diagnosed, the type of infecting strains can be determined. This situation can be formulated under a competing risks model where the endpoint event is the infection and the cause of failure is the strain type determined after the infection is diagnosed. The strain-specific vaccine efficacy is defined as one minus the cause-specific hazard ratio (vaccine/placebo). The assessment of vaccine prevention of the infection can be formulated through testing if the vaccine affords protection against various strains and if and how the strain-specific vaccine efficacy depends on the type of exposing strain. Gilbert, McKeague and Sun (2006) has developed some nonparametric inference procedures for testing the mark-specific vaccine efficacy where the mark variable is a continuous variable for strain type without adjusting for covariate variables. The contribution of this paper is to assess the vaccine efficacy adjusted for covariate effects when there are finite many causes of failure (strain types).

Let  $[0, \tau]$  be the follow-up period of a vaccine trial. We use the standard formulation of the competing risks model which assumes the existence of a latent failure time  $T_j$  corresponding to each failure type (or each type of infecting strain in vaccine trial)  $j, j = 1, \dots, k$ . The observed time of failure is given by  $X = \min_{1 \leq j \leq k} T_j$  and the cause of failure  $\delta \in \{1, \dots, k\}$  is observed along with a  $p$ -vector  $\mathbf{Z}$  representing the covariate information. When the failure time  $X$  is observed, the cause  $j$  is also

observed, whereas if  $X$  is censored, the cause is unknown. A number of authors have noticed that the joint and the marginal probability distributions of the latent failure times are generally not identifiable on the basis of competing risks data unless they are independent (Tsiatis 1975). Instead, statistical interest focuses on the conditional cause-specific hazard rate or cumulative incidence function. The conditional cause-specific hazard rate of cause  $j$  is defined by

$$\lambda_j(t|\mathbf{z}) = \lim_{\Delta t \downarrow 0} \frac{1}{\Delta t} P(t \leq X < t + \Delta t, \delta = j | X \geq t, \mathbf{Z} = \mathbf{z}).$$

The function  $\lambda_j(t|\mathbf{z})$  gives the instantaneous failure rate from cause  $j$  at time  $t$  in the presence of the other failure types. The conditional cumulative incidence function is given by

$$F_j(t|\mathbf{z}) = P(X \leq t, \delta = j | \mathbf{Z} = \mathbf{z}).$$

Evidently,  $F_j(t|\mathbf{z})$  does not have a survival interpretation, but has crude interpretation as the expected proportion of persons with covariate  $\mathbf{z}$  who acquire disease by strain  $j$  by time  $t$  in the presence of all circulating strains. Both of them are estimable from the competing risks data (Prentice et al. 1978, Kalbfleisch and Prentice 1980). The problem of testing the equality of cause-specific hazard functions has been developed by Prentice et al. 1978, Aly, Kochar and McKeague 1994, Lunn and McNeil 1995, Sun and Tiwari 1995, Lam 1998, Hu and Tsai 1999, Luo and Turnbull 1999, and Sun 2001 among others. Alternative problem based on the cumulative incidence functions include those of Gray 1988, Benichou and Gail 1990, Fine and Gray 1999, and McKeague, Gilbert and Kanki 2001 among others.

We consider the covariate  $\mathbf{z} = (z_1, \mathbf{z}_2)^T$ , where  $z_1$  is the treatment group indicator and  $\mathbf{z}_2$  the other covariates vector. The strain-specific vaccine efficacy to reduce susceptibility to strain  $j$  is defined as

$$VE_j(t|\mathbf{z}_2) = 1 - \frac{\lambda_j(t|1, \mathbf{z}_2)}{\lambda_j(t|0, \mathbf{z}_2)}.$$

Gilbert (2000) discussed the assumptions needed for the strain-specific vaccine efficacy to have a meaningful biological interpretation. Under the assumption of an equal

distribution, conditional on the covariate  $\mathbf{z}_2$ , of exposure to each strain  $j \in \{1, \dots, k\}$  during the follow-up period  $[0, \tau]$  for vaccine and placebo recipients (justified by randomization and blinding),  $VE_j(t|\mathbf{z}_2)$  approximately equals the relative multiplicative reduction in susceptibility to strain  $j$  for vaccine versus placebo recipients conditional on  $\mathbf{z}_2$  under a fixed amount of exposure to strain  $j$  at time  $t$ .

Among the cause-specific regression models, the Cox's proportional hazards model (Cox 1972) has been widely used. The proportional hazards model specify the strain-specific hazard function  $\lambda_j(t|\mathbf{z})$  given a particular value  $\mathbf{z} = (z_1, \mathbf{z}_2)^T$  by

$$\lambda_j(t|\mathbf{z}) = \lambda_{0j}(t) \exp(\boldsymbol{\beta}_j^T \mathbf{z}) = \lambda_{0j}(t) \exp(\alpha_j z_1 + \boldsymbol{\gamma}_j^T \mathbf{z}_2), \quad (1.1)$$

where  $\lambda_{0j}(\cdot)$  is an unspecified baseline hazard function,  $\boldsymbol{\beta}_j = (\alpha_j, \boldsymbol{\gamma}_j)$  is a  $p$ -vector of regression parameters for the  $j$ th strain. In the special case of a single group variable, ( $z = 1$  indicates the treatment group,  $z = 0$  the placebo group), model (1.1) reduces to

$$\lambda_j(t|z = 0) = \lambda_{0j}(t) \quad \text{and} \quad \lambda_j(t|z = 1) = \lambda_{0j}(t) \exp(\alpha_j).$$

Thus in the vaccine trial  $\alpha_j$  is the log-relative risk among vaccinees and non-vaccinees of disease by strain  $j$  causing clinically significant infection.

In this paper we develop tests for whether the strain-specific vaccine efficacy depends on strains, that is, of the null hypothesis (A)  $H_{A0} : VE_1 = VE_2 = \dots = VE_k$ . The following alternative hypotheses of particular interest will be considered:

$$H_{A1} : VE_1 \geq \dots \geq VE_j \geq \dots \geq VE_k \text{ with strict inequality for some } 1 \leq j \leq k,$$

$$H_{A2} : VE_i \neq VE_j \text{ for at least one pair of } i \text{ and } j, \quad 1 \leq i < j \leq k.$$

$H_{A0}$  implies no differences in vaccine impact to a strain  $j$ . The ordered alternative  $H_{A1}$  says that vaccine efficacy decreases with strains. Under the Cox's proportional model (1.1) the hypotheses (A) can be rewritten as

$$H_{A0} : \alpha_1 = \alpha_2 = \dots = \alpha_k$$

against the following alternative hypotheses

$$H_{A1} : \alpha_1 \leq \cdots \leq \alpha_j \leq \cdots \leq \alpha_k \text{ with strict inequality for some } 1 \leq j \leq k,$$

$$H_{A2} : \alpha_i \neq \alpha_j \text{ for at least one pair of } i \text{ and } j, \quad 1 \leq i < j \leq k.$$

We develop two test statistics for detecting departures from  $H_{A0}$  in the direction of  $H_{A1}$  and  $H_{A2}$ . We also consider tests of the null hypothesis (B) that the vaccine affords no protection against various strains  $H_{B0} : VE_j = 0$  for all  $j = 1, \dots, k$ , that is,

$$H_{B0} : \alpha_j = 0, \text{ for all } j = 1, \dots, k$$

versus one of the following alternative hypotheses

$$H_{B1} : \alpha_j \leq 0 \text{ with strict inequality for some } j,$$

$$H_{B2} : \alpha_j \neq 0 \text{ for some } j.$$

Thus,  $H_{B0}$  implies that the vaccine affords no protection against any strain at all. The alternative  $H_{B1}$  indicates that the vaccine provides protection for at least some of the infecting strain, while  $H_{B2}$  simply says that the vaccine affects the risk of infection.

Gilbert (2000) considered testing differential vaccine efficacy for two strains 1 and 2 using a data duplication method (Lunn and McNeil, 1995) where the logarithm of the ratio  $(1 - VE_2)/(1 - VE_1)$  is reparameterized as regression coefficient in the proportional hazard model for the argumented data. However, it is not appealing to extend this method to test for multiple strain-specific vaccine efficacy since the estimators of these ratios under reparameterization are not asymptotically independent and the covariance structure is not clear. In this paper, we propose some simple test procedures to test for various hypotheses concerning the vaccine efficacy. The tests compare strain variations in vaccine protection and assess if the vaccine provides any protection.

In Section 2 we develop test statistics for detecting departures from  $H_{A0}$  and  $H_{B0}$ . Section 3 describes the results of simulations assessing the level and power of the proposed tests. The tests are applied to the oral cholera vaccine trial in Section 4.

## 2 Test Procedure

### 2.1 Preliminaries

Consider the competing risks model with  $k$  causes of failure. In the absence of censoring, we observe  $(X, \delta, \mathbf{Z})$ , where  $X = \min(T_1, \dots, T_k)$ . The cause of failure  $\delta = j$  when  $X = T_j$ . For simplicity, we restrict attention to time-independent covariates. In the presence of right censoring, we observe  $(\tilde{X}, \tilde{\delta}, \mathbf{Z})$ , where  $\tilde{X} = \min(X, C)$ ,  $\tilde{\delta} = \delta I(X \leq C)$ ,  $C$  is the censoring time, and  $I(\cdot)$  is the indicator function. The value  $\tilde{\delta} = 0$  indicates that the failure time is censored. In this case, neither failure time nor the cause of failure is observed. The right censored competing risks data consist of  $n$  independent replicates of  $(\tilde{X}_i, \tilde{\delta}_i, \mathbf{Z}_i)$  of  $(\tilde{X}, \tilde{\delta}, \mathbf{Z})$ . It is assumed that  $C$  is conditionally independent of  $T_1, \dots, T_k$  given  $\mathbf{Z}$ . The latent failure times  $T_j$  do not have to be independent, but assume that  $P(T_i = T_l) = 0$  for  $i \neq l$  (i.e., assuming distinct failure types).

The conditional cause-specific hazard rates  $\lambda_j(t|\mathbf{z})$  given a particular value  $\mathbf{z} = (z_1, \mathbf{z}_2)^T$  are modelled by separate Cox proportional hazards model

$$\lambda_j(t|\mathbf{z}) = \lambda_{0j}(t) \exp(\boldsymbol{\beta}_j^T \mathbf{z}) = \lambda_{0j}(t) \exp(\alpha_j z_1 + \boldsymbol{\gamma}_j^T \mathbf{z}_2).$$

Then the partial likelihood score function for  $\boldsymbol{\beta}_j$  is

$$U_j(\boldsymbol{\beta}) = \sum_{i=1}^n \Delta_{ji} (\mathbf{Z}_i - \bar{\mathbf{Z}}(\boldsymbol{\beta}, \tilde{X}_i)),$$

where  $\Delta_{ji} = I(\tilde{\delta}_i = j)$  and

$$\bar{\mathbf{Z}}(\boldsymbol{\beta}, t) = \frac{\sum_{i=1}^n Y_i(t) \exp(\boldsymbol{\beta}^T \mathbf{Z}_i) \mathbf{Z}_i}{\sum_{i=1}^n Y_i(t) \exp(\boldsymbol{\beta}^T \mathbf{Z}_i)}, \quad Y_i(t) = I(\tilde{X}_i \geq t).$$

The maximum partial likelihood estimator  $\hat{\boldsymbol{\beta}}_j$  can be obtained from the estimating equation  $U_j(\boldsymbol{\beta}) = \mathbf{0}$ , treating disease by all other strains than  $j$  as censoring. Under some mild regularity conditions (Andersen et al. 1993) the random vector  $\mathcal{I}_j^{1/2}(\hat{\boldsymbol{\beta}}_j)(\hat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_j)$  is asymptotically zero-mean normal with an identity covariance

matrix, where  $\mathcal{I}_j(\boldsymbol{\beta})$  is minus the derivative matrix of  $U_j(\boldsymbol{\beta})$ . Furthermore the  $\widehat{\boldsymbol{\beta}}_j$ ,  $j = 1, 2, \dots, k$ , are asymptotically independent (McKeague et al. 2001).

## 2.2 Test statistics

Our interest is based on the first component of parameter  $\boldsymbol{\beta}_j$ . Let  $\widehat{\alpha}_j$  be the first component of  $\widehat{\boldsymbol{\beta}}_j$ . Then  $\widehat{\alpha}_j$  is asymptotically normal with mean  $\alpha_j$  and variance can be estimated by  $\widehat{\sigma}_j^2$ , the (1,1)th element of  $\mathcal{I}_j^{-1}(\widehat{\boldsymbol{\beta}})$ . The following test statistic  $T_1$  is suggested for detecting the monotone departure  $H_{A1}$  from  $H_{A0}$ :

$$T_1 = \frac{\sum_{j=2}^k w_j (\widehat{\alpha}_j - \widehat{\alpha}_{j-1})}{\sqrt{\sum_{j=1}^k (w_j - w_{j+1})^2 \widehat{\sigma}_j^2}}, \quad (2.1)$$

where  $w_j$  are weights with  $w_1 = w_{k+1} = 0$ . The test statistic  $T_1$  can be rewritten as

$$T_1 = \frac{\sum_{j=1}^k (w_j - w_{j+1}) \widehat{\alpha}_j}{\sqrt{\sum_{j=1}^k (w_j - w_{j+1})^2 \widehat{\sigma}_j^2}}.$$

Since the  $\widehat{\alpha}_j$  are asymptotically independent, it is seen that under  $H_{A0}$  the test statistic  $T_1$  is asymptotically normal with mean zero and variance of one. The  $H_{A0}$  is rejected in favor of  $H_{A1}$  at significance level  $\alpha$  if  $T_1 > z_\alpha$ , where  $z_\alpha$  is  $(1 - \alpha)$ th quantile of the standard normal distribution.

To detect general alternative  $H_{A2}$  from  $H_{A0}$ , we suggest a test statistic  $T_2$

$$T_2 = \begin{pmatrix} w_2(\widehat{\alpha}_2 - \widehat{\alpha}_1) \\ \vdots \\ w_k(\widehat{\alpha}_k - \widehat{\alpha}_{k-1}) \end{pmatrix}^T \boldsymbol{\Sigma}^{-1} \begin{pmatrix} w_2(\widehat{\alpha}_2 - \widehat{\alpha}_1) \\ \vdots \\ w_k(\widehat{\alpha}_k - \widehat{\alpha}_{k-1}) \end{pmatrix}, \quad (2.2)$$



respectively. Here  $U_{12\alpha} = -\sqrt{k}z_\alpha$  and  $U_{11\alpha}$  can be easily obtained through a simple simulation. For example, for  $\alpha = 0.05$   $U_{11\alpha} = -1.955$  when  $k = 2$ , and  $U_{11\alpha} = -2.121$  when  $k = 3$ . We also suggest the following test statistics for detecting departures from  $H_{B0}$  in the direction of  $H_{B2}$ :

$$\begin{aligned} U_{21} &= \sup_{1 \leq j \leq k} \left( \frac{\hat{\alpha}_j}{\hat{\sigma}_j} \right)^2, \\ U_{22} &= \sum_{j=1}^k \left( \frac{\hat{\alpha}_j}{\hat{\sigma}_j} \right)^2. \end{aligned} \tag{2.4}$$

Similarly, the critical values of the test  $U_{21}$  can be easily obtained through simulations and  $U_{22}$  has the asymptotic chi-square distribution with  $k$  degrees of freedom under  $H_{B0}$ .

### 3 Simulation study

In this section we present a simulation study to examine the performance of the proposed test procedures under practical sample sizes. The powers of the tests depend on the specifications of alternative hypotheses. The empirical sizes and powers of the proposed tests are calculated under various models through simulations. We consider a  $p = 2$  dimensional covariate  $\mathbf{Z} = (Z_1, Z_2)$ , where  $Z_1$  is a Bernoulli random variable with probability of success 0.5 and  $Z_2$  is uniformly distributed on  $[0, 1]$ . The latent failure times are taken to be conditionally independent with conditional cause-specific hazard functions given by Cox models of the form

$$\lambda_j(t|\mathbf{z}) = t^{\theta_j} \exp(\alpha_j z_1 + \gamma_j z_2), \quad j = 1, 2, \dots, k, \tag{3.1}$$

where  $\theta_j$ ,  $\alpha_j$  and  $\gamma_j$  are the parameters to be specified. We consider the number of causes  $k$  to be 2 and 3, respectively. The simple random right censoring  $C$  is taken to be exponentially distributed, with parameter adjusted so approximately 25% of the observations are censored. For comparison, we include results for the uncensored case as well. The size and power of the tests at the nominal 0.05 level are estimated from 1000 independent samples. For simplicity we set the weights to be one.

In Table 1–4 we report the results of the simulations for the tests of hypotheses (A). Table 1 and Table 2 present the observed levels and powers of the tests  $T_1$  and  $T_2$  when  $k = 2$  under various choices of  $(\theta_1, \theta_2)$ ,  $(\alpha_1, \alpha_2)$ ,  $(\gamma_1, \gamma_2)$  and different sample sizes. We take  $\theta_1 = 0.0, 0.5$ ,  $\theta_2 = 0.0, 0.5$ ,  $\alpha_1 = -1.0, -1.5, -2.0$ ,  $\alpha_2 = -1.0$ , and the sample size  $n = 100, 200, 300$ , and  $500$ . The results for  $(\gamma_1, \gamma_2) = (1.0, 1.0)$  are collected in Table 1 and the results for  $(\gamma_1, \gamma_2) = (1.0, 1.5)$  in Table 2. Tables show that the observed levels of our tests are close to their nominal 5% level at each sample size for both uncensored and censored data, regardless of choice of  $(\theta_1, \theta_2)$ . The powers of all tests considered increase as the extent of departure from the null hypothesis  $H_{A0}$  increases, i.e. as  $\alpha_1$  changes from  $-1, -1.5$  to  $-2$ . Increasing sample size strengthens the power to detect departures from the  $H_{A0}$ . Compared with the power for  $T_2$ , it is not surprising that the test  $T_1$  shows more power in detecting departure in the direction of  $H_{A1}$ . The performances of the proposed tests when  $k = 3$  are also studied. Table 3 and Table 4 are the observed levels and powers of the tests when  $k = 3$  for various choices of  $(\theta_1, \theta_2, \theta_3)$ ,  $(\alpha_1, \alpha_2, \alpha_3)$ , and  $(\gamma_1, \gamma_2, \gamma_3)$ . The powers in Table 3 are evaluated under the monotone alternative  $H_{A1}$ , while the powers in Table 4 are evaluated under general alternative  $H_{A2}$ . The simulation results in Table 3 lead us to the same conclusions as those for the two causes. The observed levels of tests are all close to 5% nominal level. The test  $T_1$  shows more power in detecting monotone departures  $H_{A1}$  compared with that for  $T_2$ . However, in Table 4 the test  $T_2$  shows more power in detecting non-monotone departures.

We consider the tests of hypotheses (B) in Table 5–7. Table 5 gives the observed levels and powers of the tests  $U_{11}$ ,  $U_{12}$ ,  $U_{21}$ , and  $U_{22}$  with various choices of  $(\theta_1, \theta_2)$ ,  $(\alpha_1, \alpha_2)$  for the sample size  $n = 100, 200, 300, 500$  and for  $(\gamma_1, \gamma_2) = (1.0, 1.0)$ . They performed at near 5% nominal level for both uncensored and censored data, regardless of choice of  $(\theta_1, \theta_2)$ . Compared with the powers for  $U_{21}$  and  $U_{22}$ , the tests  $U_{11}$  and  $U_{12}$  show more powers in detecting departure in the direction of  $H_{B1}$ . Note that  $U_{11}$  shows more power in the direction of  $H_{B1}$  than  $U_{12}$  does, and  $U_{21}$  shows more power

in detecting departure in the direction of  $H_{B2}$  than  $U_{22}$  does. The simulations with  $(\gamma_1, \gamma_2) = (1.0, 1.5)$  are repeated and Table 6 gives the same conclusions as those in Table 5. The performance of the proposed tests  $U_{11}$ ,  $U_{12}$ ,  $U_{21}$ , and  $U_{22}$  with three causes of failures is also summarized in Table 7. The simulation results in Table 7 lead us to similar conclusions as those in the Table 5 and 6.

## 4 Application

In this section, we apply our test procedures to a real data set from an oral cholera vaccine trial to assess the efficacy of two oral cholera vaccines named B subunit killed whole cell (BS-WC) and killed whole-cell-only (WC). The randomized double-blind field trial was conducted in Bangladesh among children aged 2–15 years and adult women aged over 15 years (Clemens et al. 1990, Van Loon et al. 1993). Two cholera biotypes, classical and El Tor, and two cholera serotypes, Inaba and Ogawa, circulated during the trial. The case failure time was the first diarrheal episode in which *Vibrio cholerae* 01 isolated. We analyzed the data collected during the first three years of follow-up. In the study population, children and woman received three doses of the BS-WC vaccine, the WC vaccine, or a placebo. The placebo consisted of heat-inactivated *Escherichia coli* K12 strain. Of the 62,178 persons who were in the study, 20,648 (33.21%) received the WC vaccine, 20,695 (33.28%) persons received the BS-WC vaccine and the rest received the placebo. There were 14,448 (23.24%) boys and 47,730 (76.76%) girls or women. Of the 62,178 study population 523 (0.84%) persons experienced the case episodes; 285 of which excreted classical vibrios and 238 El Tor vibrios by biotypes, and 77 of 523 had Inaba and 446 with Ogawa vibrios by serotypes. Given that the vaccines contain more classical than El Tor antigens, vaccine protection might be expected to differ by biotype. Using the methods developed in this paper, we test, adjusting for gender, if and how vaccine protection varies with biotype, with serotype, and with biotype/serotype jointly.

We let  $z_1$  be the treatment group indicator (each of the two vaccines, BS-WC and WC, is compared to the placebo group) and  $z_2$  the gender. First, we tested for whether the vaccine protection differed according to the biotypes: classical cholera (strain 1) and El Tor cholera (strain 2). Both WC and BS-WC vaccines showed that vaccine protection differed by biotype. The test statistics for hypotheses (A) were  $T_1 = 2.0973$  ( $p$ -value = 0.0080),  $T_2 = 4.3987$  ( $p$ -value = 0.0280) for WC vaccine, and  $T_1 = 1.7453$  ( $p$ -value = 0.0310),  $T_2 = 3.0462$  ( $p$ -value = 0.0790) for BS-WC vaccine. They are all significant at 5% significance level, with an exception of  $T_2$  for BS-WC vaccine which is significant at 10% level of significance. This reveals that both vaccines prevent clinically significant infection better from classical cholera than from El Tor cholera. On the other hand, no differential serotype protection from Inaba and Ogawa cholera was observed for both vaccines (Inaba as strain 1, Ogawa as strain 2). For WC vaccine the test statistics were  $T_1 = -1.0985$  ( $p$ -value = 0.8680) and  $T_2 = 1.2067$  ( $p$ -value = 0.2800). For BS-WC vaccine  $T_1 = 0.4542$  ( $p$ -value = 0.3080) and  $T_2 = 0.2063$  ( $p$ -value = 0.6630).

We also performed how vaccine protection varies with biotype/serotype jointly (classical, Ogawa cholera as strain 1, classical, Inaba cholera and El Tor, Ogawa cholera as strain 2, El Tor, Inaba cholera as strain 3). For WC vaccine the test statistics were  $T_1 = 1.8929$  ( $p$ -value = 0.0200) and  $T_2 = 5.5535$  ( $p$ -value = 0.0640). For BS-WC vaccine  $T_1 = 1.6391$  ( $p$ -value = 0.0400),  $T_2 = 2.7458$  ( $p$ -value = 0.2550). Both vaccines conferred a significant protection difference trend against biotype/serotype jointly, the vaccine protection is better against classical, Ogawa cholera (strain 1), less for classical, Inaba cholera or El Tor, Ogawa cholera (strain 2) and worst for El Tor, Inaba cholera (strain 3). Moreover, all of the test statistics  $U_{11}$ ,  $U_{12}$ ,  $U_{21}$ , and  $U_{22}$  showed highly significant  $p$ -values ( $< 0.00001$ ). This illustrates the significance of vaccine efficacy against biotype/serotype jointly. Our results support that over the entire 3 years each vaccine conferred protection against cholera. Moreover, vaccine efficacy differed notably according to the biotype of the infecting *Vibrio cholerae* 01.

## Acknowledgements

The authors thank Peter Gilbert for helpful discussions and John Clements for providing the cholera vaccine trial data. This research was partially supported by NSF grants DMS-0304922 and DMS-0604576.

## References

- Aly, E. A. A., Kochar, S.C. and McKeague, I. W. (1994). Some tests for comparing cumulative incidence functions and cause-specific hazard rates. *J. Amer. Statist. Assoc.*, **89**, 994–999.
- Andersen, P. K. and Gill, R. D. (1982). Cox’s regression model for counting processes: a large sample study. *Ann. Statist.*, **10**, 1100–1120.
- Andersen, P. K., Borgan, Ø., Gill, R. D., and Keiding, N. (1993). *Statistical models based on counting processes*. Springer-verlag, New York.
- Benichou, J. and Gail, M. H. (1990). Estimates of absolute cause-specific risk in cohort studies. *Biometrics*, **46**, 813–826.
- Cox, D. R. (1972). Regression models and life tables (with discussion). *J. Roy. Statist. Soc. B.*, **34**, 187–220.
- Clemens JD, Sack DA, Harris JR, Van Loon F, Chakraborty J, Ahmed F, Rao MR, Khan MR, Yunus Md, Huda N, Stanton BF, Kay BA, Walter S, Eeckels R, Svennerholm A-M, and Holmgren J (1990). Field trial of oral cholera vaccines in Bangladesh: results from three-year follow-up. *Lancet*, **335**, 270–273.
- Fine, J. P. and Gray, R. J. (1999). A proportional hazards model for the subdistribution of a competing risk. *J. Amer. Statist. Assoc.*, **94**, 496–509.
- Gilbert, B. P. (2000). Comparison of competing risks failure time methods and time-

independent methods for assessing strain variations in vaccine protection. *Biometrics*, **19**, 3065–3086.

Gilbert, P. B., McKeague, I. W. and Sun, Y. (2006). The Two-sample Problem for Failure Rates Depending on a Continuous Mark: An Application to Vaccine Efficacy. Submitted manuscript.

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

Hu, X. S. and Tsai, W. Y. (1999). Linear rank tests for competing risks model. *Statist. Sinica*, **9**, 971–983.

Kalbfleisch, J. D. and Prentice, R. L. (1980). *The Statistical Analysis of Failure Time Data*. Wiley, New York.

Lam, K. F. (1998). A class of tests for the equality of  $k$  cause-specific hazard rates in a competing risks model. *Biometrics*, **85**, 179–188.

Lunn, M. and McNeil, D. (1995). Applying Cox regression to competing risks. *Biometrics*, **51**, 524–532.

Luo, X. and Turnbull, B. (1999). Comparing two treatments with multiple competing risks endpoints. *Statist. Sinica*, **9**, 985–997.

McKeague, I. W., Gilbert, P. B., and Kanki, P. (2001). Omnibus Tests for comparison of competing risks with adjustment for covariate effects. *Biometrics* **57**, 818–828.

Moeschberger, M. L. and David, H. A. (1971). Life tests under competing causes of failure and the theory of competing risks. *Biometrics* **27**, 909–933.

Prentice, R. L., Kalbfleisch, J. D., Perterson, A. V., Flournoy, N., Farewell, V. T. and Breslow, N. E. (1978). The analysis of failure times in the presence of competing risks. *Biometrics* **34**, 541–554.

- Sun, Y. (2001). Generalized nonparametric test procedures for comparing multiple cause-specific hazard rates. *J. of Nonparametric Statistics* **13**, 171–207.
- Sun, Y. and Tiwari, R. C. (1995). Comparing cause-specific hazard rates of a competing risks model with censored data. *IMS Lecture Notes-Monograph Series*, Vol. 27, 255–270. *Analysis of Censored data*, edited by H. L. Koul and J. V. Deshpandé
- Tsiatis, A. (1975). A nonidentifiability aspect of the problem of competing risks. *Proc. Natl. Acad. Sci.* **72**, 20–22.
- Van Loon F, Clemens JD, Chakraborty J, Rao MR, Kay BA, Sack DA, Yunus Md, Ali Md, Svennerholm A-M, and Holmgren J (1993). Field trial of inactivated oral cholera vaccines in Bangladesh: results from 5 years of follow up. *Vaccine* **14**, 162–166.
- Yip, P. and Lam, K. F. (1992). A class of nonparametric tests for the equality of failure rates in a competing risks model, *Comm. Statist. A-Theory Methods* **21**, 2541–2556.

Table 1: Observed levels and powers of tests for  $H_{A0}$  with  $(\gamma_1, \gamma_2) = (1.0, 1.0)$

$(\theta_1, \theta_2)$	$\alpha_1$	$\alpha_2$	$n = 100$		$n = 200$		$n = 300$		$n = 500$	
			$T_1$	$T_2$	$T_1$	$T_2$	$T_1$	$T_2$	$T_1$	$T_2$
(0,0)			(uncensored)							
	-1.0	-1.0	.058	.050	.044	.036	.055	.047	.044	.057
	-1.5	-1.0	.283	.128	.455	.231	.584	.349	.772	.575
	-2.0	-1.0	.638	.392	.895	.727	.973	.906	1.00	.993
			(20%–30%)							
	-1.0	-1.0	.056	.060	.046	.048	.054	.058	.037	.053
	-1.5	-1.0	.241	.113	.395	.178	.522	.289	.690	.468
	-2.0	-1.0	.560	.289	.819	.608	.933	.802	.995	.967
(0,5)			(uncensored)							
	-1.0	-1.0	.053	.054	.056	.054	.059	.048	.067	.042
	-1.5	-1.0	.286	.131	.449	.223	.583	.359	.759	.543
	-2.0	-1.0	.630	.415	.892	.719	.974	.899	.998	.996
			(20%–30%)							
	-1.0	-1.0	.058	.047	.055	.054	.055	.047	.062	.053
	-1.5	-1.0	.248	.128	.376	.185	.509	.298	.667	.465
	-2.0	-1.0	.553	.307	.810	.600	.921	.797	.992	.971
(5,5)			(uncensored)							
	-1.0	-1.0	.058	.050	.045	.036	.055	.047	.044	.056
	-1.5	-1.0	.283	.128	.455	.231	.584	.349	.773	.574
	-2.0	-1.0	.638	.392	.895	.727	.973	.906	1.00	.993
			(20%–30%)							
	-1.0	-1.0	.056	.050	.047	.043	.054	.042	.040	.053
	-1.5	-1.0	.237	.107	.394	.184	.522	.274	.686	.475
	-2.0	-1.0	.550	.281	.810	.605	.931	.798	.994	.956

Table 2: Observed levels and powers of tests for  $H_{A0}$  with  $(\gamma_1, \gamma_2) = (1.0, 1.5)$

$(\theta_1, \theta_2)$	$\alpha_1$	$\alpha_2$	$n = 100$		$n = 200$		$n = 300$		$n = 500$	
			$T_1$	$T_2$	$T_1$	$T_2$	$T_1$	$T_2$	$T_1$	$T_2$
(.0,.0)			(uncensored)							
	-1.0	-1.0	.062	.043	.060	.043	.053	.058	.046	.045
	-1.5	-1.0	.280	.128	.464	.223	.580	.348	.765	.561
	-2.0	-1.0	.612	.376	.880	.702	.964	.878	.997	.988
			(20%-30%)							
	-1.0	-1.0	.061	.065	.058	.058	.063	.041	.043	.053
	-1.5	-1.0	.239	.112	.388	.186	.503	.287	.681	.449
	-2.0	-1.0	.530	.274	.803	.583	.930	.783	.991	.957
(.0,.5)			(uncensored)							
	-1.0	-1.0	.050	.054	.050	.051	.056	.053	.053	.048
	-1.5	-1.0	.274	.141	.428	.231	.578	.342	.763	.546
	-2.0	-1.0	.637	.380	.885	.719	.976	.904	.998	.995
			(20%-30%)							
	-1.0	-1.0	.052	.056	.052	.049	.056	.045	.046	.039
	-1.5	-1.0	.229	.122	.369	.185	.499	.286	.669	.449
	-2.0	-1.0	.537	.295	.819	.614	.933	.802	.995	.971
(.5,.5)			(uncensored)							
	-1.0	-1.0	.062	.043	.060	.044	.053	.058	.046	.045
	-1.5	-1.0	.280	.128	.463	.223	.580	.348	.765	.561
	-2.0	-1.0	.612	.376	.880	.703	.964	.878	.997	.988
			(20%-30%)							
	-1.0	-1.0	.063	.066	.062	.055	.053	.053	.045	.051
	-1.5	-1.0	.235	.123	.375	.177	.503	.277	.669	.452
	-2.0	-1.0	.512	.260	.795	.580	.926	.772	.990	.949

Table 3: Observed levels and powers of tests for  $H_{A0}$  with  $(\gamma_1, \gamma_2, \gamma_3) = (1.0, 1.0, 1.0)$ , where  $\alpha_1, \alpha_2$ , and  $\alpha_3$  are monotone increasing.

$(\theta_1, \theta_2, \theta_3)$	$\alpha_1$	$\alpha_2$	$\alpha_3$	$n = 100$		$n = 200$		$n = 300$		$n = 500$	
				$T_1$	$T_2$	$T_1$	$T_2$	$T_1$	$T_2$	$T_1$	$T_2$
(uncensored)											
(.0,.0,.0)	-1.0	-1.0	-1.0	.055	.049	.057	.047	.053	.042	.046	.046
	-1.6	-1.3	-1.0	.260	.074	.418	.151	.571	.238	.749	.428
	-2.2	-1.6	-1.0	.595	.226	.867	.562	.969	.784	.998	.963
(20%-30%)											
	-1.0	-1.0	-1.0	.059	.040	.053	.045	.055	.053	.045	.046
	-1.6	-1.3	-1.0	.223	.070	.381	.122	.486	.185	.669	.324
	-2.2	-1.6	-1.0	.487	.157	.772	.433	.912	.662	.988	.919
(uncensored)											
(.0,.5,1.0)	-1.0	-1.0	-1.0	.049	.047	.057	.041	.051	.046	.057	.045
	-1.6	-1.3	-1.0	.241	.069	.407	.148	.527	.240	.712	.408
	-2.2	-1.6	-1.0	.580	.225	.842	.533	.947	.775	.994	.956
(20%-30%)											
	-1.0	-1.0	-1.0	.044	.033	.054	.033	.057	.030	.059	.054
	-1.6	-1.3	-1.0	.212	.070	.339	.110	.425	.185	.632	.306
	-2.2	-1.6	-1.0	.484	.160	.745	.411	.888	.628	.974	.871
(uncensored)											
(.5,.5,.5)	-1.0	-1.0	-1.0	.055	.049	.057	.047	.053	.042	.046	.046
	-1.6	-1.3	-1.0	.260	.074	.419	.151	.571	.238	.749	.429
	-2.2	-1.6	-1.0	.595	.226	.867	.562	.969	.784	.998	.963
(20%-30%)											
	-1.0	-1.0	-1.0	.056	.037	.057	.041	.057	.041	.047	.037
	-1.6	-1.3	-1.0	.223	.075	.374	.123	.482	.190	.663	.316
	-2.2	-1.6	-1.0	.476	.151	.770	.429	.914	.648	.986	.903

Table 4: Observed powers of tests for  $H_{A0}$  with  $(\gamma_1, \gamma_2, \gamma_3) = (1.0, 1.0, 1.0)$ , where  $\alpha_1$ ,  $\alpha_2$ , and  $\alpha_3$  are not monotone.

$(\theta_1, \theta_2, \theta_3)$	$\alpha_1$	$\alpha_2$	$\alpha_3$	$n = 100$		$n = 200$		$n = 300$		$n = 500$		
				$T_1$	$T_2$	$T_1$	$T_2$	$T_1$	$T_2$	$T_1$	$T_2$	
(.0,.0,.0)	(uncensored)											
	-1.3	-1.6	-1.0	.133	.074	.180	.142	.252	.249	.348	.410	
	-1.6	-2.2	-1.0	.262	.223	.433	.567	.573	.764	.759	.961	
	(20%-30%)											
	-1.3	-1.6	-1.0	.130	.081	.177	.119	.212	.190	.291	.301	
	-1.6	-2.2	-1.0	.228	.153	.397	.437	.494	.632	.679	.909	
	(.0,.5,1.0)	(uncensored)										
		-1.3	-1.6	-1.0	.125	.060	.180	.106	.224	.153	.319	.269
-1.6		-2.2	-1.0	.256	.147	.417	.398	.535	.606	.733	.872	
(20%-30%)												
-1.3		-1.6	-1.0	.104	.051	.157	.076	.187	.123	.257	.195	
-1.6		-2.2	-1.0	.217	.085	.349	.269	.430	.451	.628	.732	
(.5,.5,.5)		(uncensored)										
		-1.3	-1.6	-1.0	.133	.074	.180	.142	.252	.249	.348	.410
	-1.6	-2.2	-1.0	.262	.223	.434	.567	.573	.764	.759	.961	
	(20%-30%)											
	-1.3	-1.6	-1.0	.118	.075	.172	.114	.212	.190	.285	.312	
	-1.6	-2.2	-1.0	.227	.151	.385	.428	.488	.618	.681	.891	

Table 5: Observed levels and powers of tests for  $H_{B0}$  with  $(\gamma_1, \gamma_2) = (1.0, 1.0)$

$(\theta_1, \theta_2)$	$\alpha_1$	$\alpha_2$	$n = 100$				$n = 200$				$n = 300$				$n = 500$			
			$U_{11}$	$U_{12}$	$U_{21}$	$U_{22}$	$U_{11}$	$U_{12}$	$U_{21}$	$U_{22}$	$U_{11}$	$U_{12}$	$U_{21}$	$U_{22}$	$U_{11}$	$U_{12}$	$U_{21}$	$U_{22}$
(.0,.0)			(uncensored)															
	0.0	0.0	.042	.050	.052	.055	.046	.056	.061	.061	.046	.043	.047	.048	.038	.038	.063	.061
	-0.5	0.0	.398	.302	.195	.202	.649	.479	.426	.416	.814	.636	.628	.603	.958	.826	.875	.868
	-1.0	0.0	.843	.641	.663	.644	.990	.908	.970	.964	1.00	.978	.999	.997	1.00	1.00	1.00	1.00
			(20%-30%)															
	0.0	0.0	.049	.053	.056	.059	.045	.054	.047	.043	.047	.047	.052	.049	.045	.052	.047	.054
	-0.5	0.0	.311	.257	.123	.141	.519	.401	.308	.288	.696	.518	.469	.472	.887	.725	.741	.733
	-1.0	0.0	.720	.521	.492	.478	.962	.803	.873	.845	.996	.920	.979	.973	1.00	.995	.998	.998
(.0,.5)			(uncensored)															
	0.0	0.0	.050	.056	.044	.049	.061	.052	.045	.048	.042	.040	.054	.058	.044	.044	.047	.049
	-0.5	0.0	.500	.363	.271	.260	.780	.586	.569	.547	.918	.744	.773	.772	.989	.917	.958	.948
	-1.0	0.0	.925	.762	.815	.801	.997	.971	.990	.984	1.00	.997	1.00	1.00	1.00	1.00	1.00	1.00
			(20%-30%)															
	0.0	0.0	.046	.054	.045	.040	.050	.042	.048	.047	.048	.051	.055	.052	.035	.040	.038	.039
	-0.5	0.0	.395	.297	.198	.206	.647	.505	.440	.429	.834	.637	.643	.638	.967	.836	.877	.869
	-1.0	0.0	.855	.655	.690	.666	.989	.923	.966	.963	1.00	.979	.998	.997	1.00	1.00	1.00	1.00
(.5,.5)			(uncensored)															
	0.0	0.0	.042	.050	.053	.054	.046	.056	.060	.061	.046	.045	.047	.048	.038	.038	.064	.060
	-0.5	0.0	.398	.302	.195	.202	.649	.478	.426	.416	.815	.637	.628	.604	.958	.826	.875	.867
	-1.0	0.0	.843	.641	.663	.643	.990	.908	.970	.964	1.00	.978	.999	.997	1.00	1.00	1.00	1.00
			(20%-30%)															
	0.0	0.0	.048	.055	.052	.057	.053	.050	.046	.047	.056	.060	.057	.054	.048	.044	.041	.047
	-0.5	0.0	.309	.259	.125	.133	.523	.391	.308	.302	.693	.509	.481	.474	.885	.716	.758	.749
	-1.0	0.0	.733	.522	.493	.482	.966	.809	.874	.866	.998	.926	.977	.969	1.00	.997	.999	.999

Table 6: Observed levels and powers of tests for  $H_{B0}$  with  $(\gamma_1, \gamma_2) = (1.0, 1.5)$

$(\theta_1, \theta_2)$	$\alpha_1$	$\alpha_2$	$n = 100$				$n = 200$				$n = 300$				$n = 500$			
			$U_{11}$	$U_{12}$	$U_{21}$	$U_{22}$												
(.0,.0)	(uncensored)																	
	0.0	0.0	.048	.052	.059	.056	.050	.054	.052	.045	.047	.046	.049	.054	.046	.037	.052	.050
	-0.5	0.0	.339	.262	.162	.179	.593	.434	.371	.349	.762	.579	.568	.540	.930	.779	.820	.800
	-1.0	0.0	.780	.590	.574	.561	.982	.853	.929	.908	.999	.954	.992	.987	1.00	.999	1.00	.999
	(20%-30%)																	
	0.0	0.0	.048	.049	.053	.059	.049	.058	.045	.050	.060	.053	.053	.050	.045	.051	.043	.041
-0.5	0.0	.262	.227	.113	.128	.480	.361	.267	.271	.632	.476	.416	.407	.841	.665	.666	.641	
-1.0	0.0	.650	.453	.400	.387	.921	.741	.797	.774	.984	.879	.947	.931	.999	.983	.997	.996	
(.0,.5)	(uncensored)																	
	0.0	0.0	.038	.054	.056	.058	.057	.049	.039	.044	.040	.044	.053	.055	.042	.037	.047	.050
	-0.5	0.0	.457	.339	.232	.241	.740	.555	.515	.499	.890	.701	.740	.718	.982	.891	.934	.920
	-1.0	0.0	.895	.721	.770	.758	.996	.954	.983	.980	1.00	.994	1.00	1.00	1.00	1.00	1.00	1.00
	(20%-30%)																	
	0.0	0.0	.042	.058	.054	.053	.056	.052	.050	.053	.048	.050	.045	.045	.033	.049	.049	.050
-0.5	0.0	.360	.286	.182	.188	.596	.470	.389	.385	.801	.594	.596	.592	.953	.805	.859	.836	
-1.0	0.0	.817	.625	.636	.622	.982	.891	.953	.944	1.00	.971	.994	.991	1.00	.999	1.00	1.00	
(.5,.5)	(uncensored)																	
	0.0	0.0	.048	.052	.058	.055	.050	.054	.053	.046	.047	.047	.049	.054	.046	.037	.052	.049
	-0.5	0.0	.339	.262	.162	.179	.592	.434	.371	.348	.762	.579	.568	.540	.930	.779	.820	.800
	-1.0	0.0	.781	.589	.574	.561	.982	.853	.929	.908	.999	.954	.992	.987	1.00	.999	1.00	.999
	(20%-30%)																	
	0.0	0.0	.042	.057	.043	.049	.047	.053	.046	.052	.057	.052	.048	.053	.042	.041	.048	.048
-0.5	0.0	.265	.222	.114	.116	.468	.357	.257	.260	.632	.476	.416	.415	.834	.660	.663	.648	
-1.0	0.0	.641	.458	.401	.386	.923	.745	.800	.779	.987	.880	.942	.935	1.00	.987	.997	.996	

Table 7: Observed levels and powers of tests for  $H_{B0}$  with  $(\gamma_1, \gamma_2, \gamma_3) = (1.0, 1.0, 1.0)$

$(\theta_1, \theta_2, \theta_3)$	$n = 100$				$n = 200$				$n = 300$				$n = 500$							
	$\alpha_1$	$\alpha_2$	$\alpha_3$	$U_{11}$	$U_{12}$	$U_{21}$	$U_{22}$	$U_{11}$	$U_{12}$	$U_{21}$	$U_{22}$	$U_{11}$	$U_{12}$	$U_{21}$	$U_{22}$	$U_{11}$	$U_{12}$	$U_{21}$	$U_{22}$	
$(.0, .0, .0)$	(uncensored)																			
	0.0	0.0	0.0	.038	.051	.040	.040	.048	.047	.042	.045	.032	.035	.056	.048	.047	.033	.062	.066	
	-0.5	0.0	0.0	.223	.169	.094	.081	.387	.251	.222	.203	.564	.326	.353	.325	.788	.514	.616	.565	
	-1.0	-0.5	0.0	.709	.712	.420	.474	.942	.920	.841	.877	.992	.981	.966	.977	1.00	1.00	.999	1.00	
	(20%-30%)																			
	0.0	0.0	0.0	.042	.042	.048	.048	.045	.043	.043	.045	.034	.037	.058	.049	.034	.045	.052	.056	
	-0.5	0.0	0.0	.160	.129	.065	.065	.296	.203	.143	.133	.439	.281	.260	.233	.671	.395	.455	.427	
	-1.0	-0.5	0.0	.506	.561	.241	.264	.877	.833	.647	.711	.966	.937	.885	.916	.999	.995	.989	.993	
	$(.0, .5, 1.0)$	(uncensored)																		
		0.0	0.0	0.0	.046	.049	.054	.049	.047	.048	.037	.043	.041	.041	.039	.036	.036	.045	.041	.043
		-0.5	0.0	0.0	.378	.256	.193	.167	.644	.420	.449	.411	.827	.531	.659	.618	.966	.707	.901	.878
		-1.0	-0.5	0.0	.897	.842	.729	.749	.998	.977	.986	.982	1.00	.999	.999	1.00	1.00	1.00	1.00	1.00
(20%-30%)																				
0.0		0.0	0.0	.044	.046	.040	.041	.039	.043	.040	.044	.043	.053	.033	.037	.042	.043	.044	.048	
-0.5		0.0	0.0	.286	.201	.157	.148	.556	.343	.339	.310	.721	.462	.515	.497	.922	.635	.806	.783	
-1.0		-0.5	0.0	.810	.730	.576	.559	.981	.936	.937	.943	1.00	.989	.992	.992	1.00	1.00	1.00	1.00	
$(.5, .5, .5)$		(uncensored)																		
		0.0	0.0	0.0	.038	.051	.041	.040	.048	.047	.042	.045	.032	.035	.055	.047	.047	.033	.062	.065
		-0.5	0.0	0.0	.223	.169	.094	.081	.387	.252	.222	.203	.564	.326	.353	.324	.788	.514	.619	.565
		-1.0	-0.5	0.0	.709	.712	.420	.475	.942	.920	.842	.877	.992	.981	.967	.976	1.00	1.00	.999	1.00
	(20%-30%)																			
	0.0	0.0	0.0	.049	.043	.045	.045	.048	.039	.055	.051	.034	.034	.048	.039	.030	.044	.047	.052	
	-0.5	0.0	0.0	.172	.125	.076	.078	.298	.207	.151	.146	.447	.285	.248	.242	.685	.405	.460	.426	
	-1.0	-0.5	0.0	.514	.561	.247	.283	.871	.836	.658	.720	.970	.944	.886	.919	1.00	.997	.990	.995	