Statistical Communications in Infectious Diseases

Volume 3, *Issue* 1 2011 *Article* 7

Nonparametric and Semiparametric Analysis of Current Status Data Subject to Outcome Misclassification

Victor G. Sal y Rosas, *University of Washington*
James P. Hughes, *University of Washington and Harborview Medical Center*

Recommended Citation:
Sal y Rosas, Victor G. and Hughes, James P. (2011) "Nonparametric and Semiparametric Analysis of Current Status Data Subject to Outcome Misclassification," *Statistical Communications in Infectious Diseases*: Vol. 3: Iss. 1, Article 7.
DOI: 10.2202/1948-4690.1032

©2011 De Gruyter. All rights reserved.

- 10.2202/1948-4690.1032
Downloaded from De Gruyter Online at 09/12/2016 10:40:35PM
via South China University of Technology

Nonparametric and Semiparametric Analysis of Current Status Data Subject to Outcome Misclassification

Victor G. Sal y Rosas and James P. Hughes

Abstract

In this article, we present nonparametric and semiparametric methods to analyze current status data subject to outcome misclassification. Our methods use nonparametric maximum likelihood estimation (NPMLE) to estimate the distribution function of the failure time when sensitivity and specificity are known and may vary among subgroups. A nonparametric test is proposed for the two sample hypothesis testing. In regression analysis, we apply the Cox proportional hazard model and likelihood ratio based confidence intervals for the regression coefficients are proposed. Our methods are motivated and demonstrated by data collected from an infectious disease study in Seattle, WA.

KEYWORDS: current status data, outcome misclassification, sensitivity, specificity

- 10.2202/1948-4690.1032
Downloaded from De Gruyter Online at 09/12/2016 10:40:35PM
via South China University of Technology

1 Introduction

In many epidemiological studies, key objectives are (i) to estimate the distribution function (d.f.) of the time, T, to a particular event of interest, (ii) to test whether the d.f.s. of two groups are equal, and (iii) to measure the association of the failure time with a set of factors via a regression model. However, there are scenarios where T is not observed; instead one only observes whether or not T exceeds a random (or fixed by design) monitoring time Y. Data with this type of structure are known as current status data or type I interval-censored data.

An additional complication arises if the outcome is measured imperfectly. For example, a test for disease may be insensitive and/or nonspecific; self report of weaning may be inaccurate due to social desirability bias; biopsies may miss a tumor; etc. Assuming that the misclassification rates are known, Neuhaus provided an extended study of misclassified binary data in the context of clustered and longitudinal data (Neuhaus, 2002). Balasubramanian and Lagakos (2001, 2003) estimated the risk of vertical transmission of HIV-1 assuming perfect specificity and time-dependent sensitivity. Richardson and Hughes (2000) implemented an EM algorithm to estimate the cumulative probability of a infectious disease infection in a discrete time context with imperfect testing. Meier, Richardson, and Hughes (2003) extended their ideas using a Cox proportional model in discrete time and applied those ideas to study risk factors associated with HIV acquisition among commercial sex workers in Mombasa, Kenya.

Recently, McKeown and Jewell (2010) discussed adjusting for known rates of outcome misclassification with current status data. In particular, they described the NPMLE of F for the one sample problem, under misclassification, and extended their idea to a parametric regression setting. This work extends these ideas in several directions: (a) In the one sample problem, we assume that sensitivity and specificity may vary across group of individuals. For instance, one may want to combine observations that were tested with different laboratory tests, or a proportion of the cohort may be tested with a more accurate test (possibly perfect sensitivity and specificity) and the remaining participants with a less accurate test, (b) for the two sample problem, some key ideas on hypothesis testing are presented, and (c) in the regression context, we develop a semiparametric proportional hazard model for misclassified current status data.

A study conducted in Seattle, WA from 1998 to 2003 motivated our interest in this problem (Golden, Whittington, Handsfield, Hughes, Stamm, Hogben, Clark, Malinski, Helmers, Thomas, and Holmes, 2005). The primary objective of the study was prevention of recurrent gonorrhea or chlamydial infection in patients 3 to 19 weeks after treatment and randomization to standard or expedited partner therapy. Patients in the expedited-treatment group were offered medication to give to their

- 10.2202/1948-4690.1032
Downloaded from De Gruyter Online at 09/12/2016 10:40:35PM
via South China University of Technology

sex partners, or, if the participants preferred, study staff members could contact the partners and provide them with medication without a clinical examination. In this study, participants were observed only once during follow-up and their time of observation varied considerably. The laboratory test used to measure the outcome was 90% sensitive and 100% specific (Carroll, Aldeen, Morrison, Anderson, Lee, and Mottice, 1998).

The outline of this article is as follows. In section 2, we introduce notation, formulate the statistical problem and present inference results for the one sample problem, two sample hypothesis testing and semiparametric regression analysis. In section 3, we present simulation results and in section 4, an example using data from the aforementioned Partners Notification Study (Golden et al., 2005) is described. We conclude with a discussion and future directions of research in section 5. For details of the proofs, for the case when one laboratory test is available, we refer readers to Sal y Rosas and Hughes (2010). For the case of more than one laboratory test, details are available upon request to the authors.

2 Inference

2.1 Data structure

Assume that the failure time T is a random variable on $\mathbb{R}_+$ with d.f. F and Y is a random observation time on $\mathbb{R}_+$ with d.f. G. At Y we observe only an indicator variable $\tilde{\Delta}$ that tells us whether the outcome has occurred ($\tilde{\Delta} = 1$) or not ($\tilde{\Delta} = 0$) according to a laboratory test result. Let k be the number of laboratory tests, and ϕ_j and ψ_j the sensitivity and specificity of the jth test, respectively. Then the available data are $(Y, \tilde{\Delta}, \mathbf{U})$ where $\mathbf{U} = (U_1, U_2, \ldots, U_k)$ is an vector of indicator variables that denotes which laboratory test was used. For example, if $k = 3$ and $\mathbf{U} = (1,0,0)$, then in this case, three tests were available to the researcher and for this specific observation, the first one was used.

Let $Y_{(i)}$ be the ith ordered value of $Y_1, \ldots, Y_n$ and $(\tilde{\Delta}_{(i)}, \mathbf{U}_{(i)})$ are the censoring and test indicator variables associated with $Y_{(i)}$, respectively. There are three main assumptions that will hold throughout the paper: (1) T is independent of Y, (2) ϕ_j and ψ_j are fixed and known with $\phi_j + \psi_j > 1$ for $j = 1, \ldots, k$, and (3) the assignment of a laboratory test is independent of the disease process. The second assumption was also made by Neuhaus (2002), Richardson and Hughes (2000), Meier et al. (2003), among others.

Notice that the joint distribution of $(Y, \mathbf{U})$ does not depend on F, therefore their contribution is ancillary. Thus, inference will be based on the conditional

DOI: 10.2202/1948-4690.1032

- 10.2202/1948-4690.1032
Downloaded from De Gruyter Online at 09/12/2016 10:40:35PM
via South China University of Technology

likelihood given Y and $\mathbf{U}$:

$$l_n(F) = \sum_{i=1}^{n}\sum_{j=1}^{k} U_{(ij)} \left\{ \tilde{\Delta}_{(i)} \log\left[1-\psi_j+(\psi_j+\phi_j-1)F(Y_{(i)})\right] + (1-\tilde{\Delta}_{(i)}) \log\left[1-\psi_j+(\psi_j+\phi_j-1)F(Y_{(i)})\right] \right\} \quad (1)$$

We wish to maximize (1) over the space $\mathscr{F}$ defined as the space of right continuous increasing step functions, bounded by $[0,1]$, with jumps at $Y_{(1)},\ldots,Y_{(n)}$. Groeneboom and Wellner (1992) studied the special case when $\phi_j = \psi_j = 1$ for all j, and McKeown and Jewell (2010) discussed the case when $k = 1$. We will denote the NPMLE of F as $\hat{F}_n$, and the naive estimator assuming no misclassification as $\tilde{F}_n$.

Notice that if $\tilde{\Delta}_{(1)} = 0$ then the value of the NPMLE $\hat{F}_n$ at $Y_{(1)}$ can be set equal to zero without imposing additional constraints on the maximization problem. A similar argument can be made if $\tilde{\Delta}_{(n)} = 1$ but in this case $\hat{F}_n(Y_{(n)})$ will be equal to one. Thus, without loss of generality, we assume for the rest of the paper that $\tilde{\Delta}_{(1)} = 1$ and $\tilde{\Delta}_{(n)} = 0$.

2.2 One sample estimation

When only one laboratory test is used ($k = 1$), the NPMLE of F has an explicit and simple formula and it is given by the following proposition.

Proposition 1. (McKeown and Jewell, 2010) *Assume that $\phi_1 = \phi$ and $\psi_1 = \psi$ are known, then the NPMLE of F at $Y_{(i)}$ is*

$$\hat{F}_n(Y_{(i)}) = \frac{\left\{\left[\tilde{F}_n(Y_{(i)}) \vee (1-\psi)\right] \wedge \phi\right\} + \psi - 1}{\phi + \psi - 1} \quad (2)$$

where $a \vee b = \max(a,b)$, $a \wedge b = \min(a,b)$, and $\tilde{F}_n$ is the "naive estimator"

$$\tilde{F}_n(Y_{(m)}) = \max_{i \le m} \min_{k \ge m} \frac{\sum_{j=i}^{k} \tilde{\Delta}_{(j)}}{k-i+1}\ ,\ m \in \{1,\ldots,n\} \quad (3)$$

In the more general case, when sensitivity and/or specificity vary across two or more group of individuals, it is not possible to express $\hat{F}_n$ explicitly as in (2). However, one can still characterize the NPMLE by using the monotonicity of F and noting that, for a given sample size n, the log likelihood function is concave with respect to $\mathbf{x} = (F(Y_{(1)}),\ldots,F(Y_{(n)}))$.

Proposition 2. *Assume that (ϕ_j, ψ_j) are known, and $\phi_j + \psi_j > 1$ for $j = 1,\ldots,k$, then a point $\hat{\mathbf{x}} = (\hat{F}_n(Y_{(1)}),\ldots,\hat{F}_n(Y_{(n)}))$ is the NPMLE over the set $\{\mathbf{x} =$*

- 10.2202/1948-4690.1032
Downloaded from De Gruyter Online at 09/12/2016 10:40:35PM
via South China University of Technology

$(x_1,\ldots,x_n) \in (0,1)^n \ : x_1 \leq \cdots \leq x_n\}$ *if and only if* $\hat{\mathbf{x}}$ *is the left derivative of the convex minorant of the cumulative sum diagram of* $P_0 = (0,0)$ *and* $P_l = (G_l(\hat{\mathbf{x}}), V_l(\hat{\mathbf{x}}))$ *for* $i = 1,\ldots,n$ *where*

$$G_l(\mathbf{x}) = \sum_{i=1}^{l}\sum_{m=1}^{k} U_{(im)} \left[\frac{\tilde{\Delta}_{(i)}(\phi_m+\psi_m-1)^2}{[1-\psi_m+(\phi_m+\psi_m-1)x_i]^2} + \frac{(1-\tilde{\Delta}_{(i)})(\phi_m+\psi_m-1)^2}{[\psi_m-(\phi_m+\psi_m-1)x_i]^2} \right].$$

and

$$\begin{aligned} V_l(\mathbf{x}) \quad = \quad & \sum_{i=1}^{l} x_i \sum_{m=1}^{k} U_{(im)} \left[\frac{\tilde{\Delta}_{(i)}(\phi_m+\psi_m-1)^2}{[1-\psi_{(m)}+(\phi_m+\psi_m-1)x_i]^2} \right] + \\ & \sum_{i=1}^{l} x_i \sum_{m=1}^{k} U_{(im)} \left[\frac{(1-\tilde{\Delta}_{(i)})(\phi_m+\psi_m-1)^2}{[\psi_{(m)}-(\phi_m+\psi_m-1)x_i]^2} \right] + \\ & \sum_{i=1}^{j}\sum_{m=1}^{k} U_{(im)} \left[\frac{\tilde{\Delta}_{(i)}(\phi_m+\psi_m-1)}{1-\psi_{(m)}+(\phi_m+\psi_m-1)x_i} - \frac{(1-\tilde{\Delta}_{(i)})(\phi_m+\psi_m-1)}{\psi_{(m)}-(\phi_m+\psi_m-1)x_i} \right]. \end{aligned}$$

Therefore to compute $\hat{\mathbf{x}}_n$, one can choose an initial value $\mathbf{x}^{(0)}$, and compute $\mathbf{x}^{(1)}$ as the left derivative of $P_0 = (0,0)$ and $P_l = (G_l(\mathbf{x}^{(0)}), V_l(\mathbf{x}^{(0)}))$. One repeats this procedure until a convergence criteria is achieved.

This type of argument was first introduced by Groeneboom and Wellner (1992) to estimate the d.f. of the time to infection in the case of type II interval censored data (with this type of data, we only know that the failure time has occurred either within some random time interval, or before the left endpoint of the time interval, or after the right end point of the time interval) and they named it the Iterative Convex Minorant (ICM) algorithm. Since then, it has been used for several statistical problems (e.g. Banerjee (2007), Huang (1996), Pan (1999), among others). Jongbloed (1998) provided a detailed study of the algorithm and added the necessary conditions that assure global convergence and call this version the Modified ICM (MICM) algorithm.

The following proposition states the limit distribution of the likelihood ratio test statistic under the null hypothesis $H_0 : F(t_0) = \tau_0 \in (0,1)$, and will provide a tool to compute pointwise confidence intervals for $\hat{F}_n$.

Proposition 3. *Consider the hypothesis testing problem* $H_0 : F(t_0) = \tau_0$ *where* $\tau_0 \in (0,1)$ *and* $t_0 \in (0,\infty)$*, and suppose that* F *and* G *are continuously differentiable in a neighborhood of* t_0 *with* $f(t_0) > 0$ *and* $g(t_0) > 0$*. If* H_0 *holds then*

$$2\log\lambda_n(\tau_0) = 2\log\left[\frac{L_n(\hat{F}_n)}{L_n(\hat{F}_n^0)}\right] = 2[l_n(\hat{F}_n) - l_n(\hat{F}_n^0)] \rightarrow_d \mathscr{D} \tag{4}$$

where λ_n *is the likelihood ratio test statistic,* $\hat{F}_n^0$ *is the NPMLE under* H_0*, and* $\mathscr{D}$ *is a random variable that does not depend on any parameter of the problem .*

DOI: 10.2202/1948-4690.1032

- 10.2202/1948-4690.1032
Downloaded from De Gruyter Online at 09/12/2016 10:40:35PM
via South China University of Technology

As stated in Proposition 3, $\mathscr{D}$ is a random variable that does not depend on F, G, ϕ, ψ or t_0 and a tabulation of the quantiles of this random variable is presented in Banerjee and Wellner (2001). Thus, the set of all τ such that $2\log\lambda_n(\tau)$ is less than d_α, where d_α is the $100(1-\alpha)$th percentile of $\mathscr{D}$, will give us an approximately $100(1-\alpha)\%$ confidence interval for $\hat{F}_n(t_0)$. We will denote such interval as $C_{n,\alpha} = \{\tau \in (0,1) : 2\log\lambda_n(\tau) < d_\alpha\}$.

Remark 1. Our model is part of the family of monotone response models. The idea is that the joint density of $(\Delta, \mathbf{U})$, conditional on the observation time Y, is monotone with respect to F. As a consequence, Proposition 3 follows as a corollary of Theorem 2.2 in Banerjee (2007). The main difference between the cases $k = 1$ and $k > 1$ are that the first one has an explicit formula for $\hat{F}_n$ and the second one requires an iterative algorithm (MICM) to compute $\hat{F}_n$.

Finally, the following algorithm describes how to compute the NPMLE of F under H_0.

Algorithm 1. Same sensitivity and specificity for all observations ($k = 1$).

(a) Find m such that $Y_{(m)} \le t_0 \le Y_{(m+1)}$

(b) For $\left\{Y_{(1)}, \ldots, Y_{(m)}\right\}$, compute the left derivative of the cumulative sum diagram formed by $P_0 = (0,0)$ and $\left\{P_i = \left(i, \sum_{j=1}^{i} \tilde{\Delta}_{(j)}\right)\right\}_{i=1}^{m}$ using (3), and denote this by $\eta = (\eta_1, \ldots, \eta_m)$. Then

$$\hat{F}_n^0(Y_{(i)}) = \left[\frac{\eta_i + \psi - 1}{\phi + \psi - 1} \vee 0\right] \wedge \tau_0$$

for $i = 1, \ldots, m$.

(c) For $\left\{Y_{(m+1)}, \ldots, Y_{(n)}\right\}$, compute the left derivative of the cumulative sum diagram formed by $P_0 = (0,0)$ and $\left\{P_i = \left(i, \sum_{j=1}^{i} \tilde{\Delta}_{(m+j)}\right)\right\}_{i=1}^{n-m}$ using (3), and denote this by $\xi = (\xi_{m+1}, \ldots, \xi_n)$. Then

$$\hat{F}_n^0(Y_{(i)}) = \left[\frac{\xi_i + \psi - 1}{\phi + \psi - 1} \vee \tau_0\right] \wedge 1$$

for $i = m+1, \ldots, n$.

Algorithm 2. More than one test - varying sensitivity and specificity ($k > 1$).

(a) Find m such that $Y_{(m)} \le t_0 \le Y_{(m+1)}$.

(b) For $\left\{Y_{(1)}, \ldots, Y_{(m)}\right\}$, compute $\eta = (\eta_1, \ldots, \eta_m)$ that maximizes the log likelihood of only the first m observations (instead of all n) using the MICM algorithm. Then $\hat{F}_n^0(Y_{(i)}) = \eta_i \wedge \tau_0$ for $i = 1, \ldots, m$.

- 10.2202/1948-4690.1032
Downloaded from De Gruyter Online at 09/12/2016 10:40:35PM
via South China University of Technology

(c) For $\{Y_{(m+1)},\ldots,Y_{(n)}\}$, compute $\xi = (\xi_{m+1},\ldots,\xi_n)$ that maximizes the log likelihood of only the last $n-m$ observations using the MICM algorithm. Then $\hat{F}_n^0(Y_{(i)}) = \xi_i \vee \tau_0$ for $i = m+1,\ldots,n$.

Remark 2. Our approach to compute pointwise confidence intervals is computationally faster than the "*m*" out of "*n*" bootstrap idea proposed by McKeown and Jewell (2010) because it avoids the problem of choosing the appropriate value of "*m*". In addition, this approach works when sensitivity and specificity varies by subgroups.

2.3 Two-Sample Test

Consider a binary variable Z that denotes whether the person is in the "intervention" group ($Z=1$) or the "control" group ($Z=0$), and where the probability of being in the intervention group is denoted by ρ. Let F_0 and F_1 denote the d.f.s of the intervention and control groups respectively, and assume the observations times for both groups follow a d.f. G. In this section we address the problem of testing for a difference between the d.f. of the two samples, i.e. $H_0 : F_0 = F_1$ where F_0 and F_1 are defined on $[0,M)$ for $M > 0$.

To address this problem, we propose the following test statistic

$$\hat{U}_n = \sqrt{\frac{n_1 n_0}{n}} \int_0^M [\hat{F}_{n_1}(t) - \hat{F}_{n_0}(t)] d\hat{G}_n \tag{5}$$

where $\hat{G}_n$ is the empirical d.f. of the observation times of the combined sample ($n = n_1 + n_0$), n_0 and n_1 are the sample size of the control and intervention group, and $\hat{F}_{n1}, \hat{F}_{n0}$ are the NPMLE of F_1 and F_0 respectively. $\hat{F}_{n1}$ and $\hat{F}_{n0}$ are computed using (2) or the MICM algorithm when $k=1$ or $k>1$, respectively. Let r_i be the number of observations where the ith test was used and let $p_i = r_i/n$, then the limit distribution of the proposed test statistic, under the null hypothesis, is presented below.

Proposition 4. *Assume that H_0 holds, the conditions of Proposition 3 in Sal y Rosas and Hughes (2010) hold, and $n_1/n \to \rho \in (0,1)$ as $n \to \infty$ then*

$$\hat{U}_n \to_d N\left(0, I_k^{-1}(F^{H_0}, G)\right) \tag{6}$$

where F^{H_0} is the common d.f. under H_0, and

$$I_k^{-1}(F^{H_0}, G) = \int_0^M \frac{[1-\bar{\psi}+(\bar{\psi}+\bar{\phi}-1)F^{H_0}(y)][\bar{\psi}-(\bar{\psi}+\bar{\phi}-1)F^{H_0}(y)]}{(\bar{\psi}+\bar{\phi}-1)^2} dG$$

where $\bar{\phi} = \sum_{j=1}^{k} p_j \phi_j$ and $\bar{\psi} = \sum_{j=1}^{k} p_j \psi_j$.

DOI: 10.2202/1948-4690.1032

- 10.2202/1948-4690.1032
Downloaded from De Gruyter Online at 09/12/2016 10:40:35PM
via South China University of Technology

Remark 3. When $k = 1$ and the null hypothesis holds, the constraints on the naive estimator will become irrelevant as the sample size n increases (i.e. $P(\tilde{F}_n^{H_0}(t) < 1-\psi)$ and $P(\tilde{F}_n^{H_0}(t) > \phi)$ tend to zero as $n \to \infty$ if t is an interior point of the domain of F^{H_0}). Therefore, as n increases

$$\hat{F}_n^{H_0} \approx \frac{\tilde{F}_n^{H_0} + \psi - 1}{\phi + \psi - 1}$$

and

$$\frac{\tilde{U}_n}{\sqrt{\tilde{I}_1^{-1}}} \approx \frac{(\phi+\psi-1)\hat{U}_n}{(\phi+\psi-1)\sqrt{\hat{I}_1^{-1}}} = \frac{\hat{U}_n}{\sqrt{\hat{I}_1^{-1}}} \tag{7}$$

where $\tilde{U}_n$ and $\tilde{I}_1^{-1}$ are naive estimators that assume no misclassification. Thus, when ϕ and ψ are constant across all individuals, misclassification can be ignored for testing $H_0 : F_0 = F_1$ (although the estimates $\tilde{F}_{n_0}$ and $\tilde{F}_{n_1}$ are, of course, biased). Because this is an asymptotic result, we explore the behavior of both test statistics for finite sample sizes in section 3.

2.4 Semiparametric regression by the Cox proportional hazards model

The proportional hazard model is

$$\Lambda(T|\mathbf{Z}) = \Lambda_0(T)e^{(\mathbf{Z}'\theta)} \tag{8}$$

where Λ is the cumulative hazard, and the covariate vector $\mathbf{Z} = (Z_1, \ldots, Z_r)$ is assumed to act additively on $\log(\Lambda(T|\mathbf{Z}))$, Λ_0 is the baseline cumulative hazard, and θ=$(\theta_1, \ldots, \theta_r)$ is the vector of log hazard ratios linking $\mathbf{Z}$. Since $F = 1 - e^{-\Lambda}$, we may combine (1) and (8), and the observed log likelihood function, for an i.i.d sample of observations $(Y_1, \tilde{\Delta}_1, \mathbf{U}_1, \mathbf{Z}_1), \ldots, (Y_n, \tilde{\Delta}_n, \mathbf{U}_n, \mathbf{Z}_n)$, is (up to a constant)

$$\begin{aligned} l_n(\theta,\Lambda) \;=\; & \sum_{i=1}^{n}\sum_{m=1}^{k} U_{im}\left\{\tilde{\Delta}_i \log\left[\phi_m - (\phi_m+\psi_m-1)e^{-\Lambda(Y_i)e^{\mathbf{Z}_i'\theta}}\right] + \right. \\ & \left. + (1-\tilde{\Delta}_i)\log\left[1-\phi_m+(\phi_m+\psi_m-1)e^{-\Lambda(Y_i)e^{\mathbf{Z}_i'\theta}}\right]\right\} \end{aligned}$$

where $\theta \in \Theta \subset R^r$ and $\Lambda \in \mathscr{G}$ where $\mathscr{G}$ is the set of nonnegative right-continuous increasing step functions (but bounded over the support of the observation time) with jump points at $Y_{(1)}, \ldots, Y_{(n)}$. Let $\hat{\theta}_n, \hat{\Lambda}_n$ denote the maximum likelihood estimates of θ and Λ.

- 10.2202/1948-4690.1032
Downloaded from De Gruyter Online at 09/12/2016 10:40:35PM
via South China University of Technology

This model is part of a bigger family of semiparametric binary regression models under shape constraints studied by Banerjee, D., and Mishra (2009). Applying those results, the regression coefficient has $\sqrt{n}$ rate of convergence and the likelihood ratio statistic may be used to compute confidence intervals for the regression coefficients.

Proposition 5. *Suppose that conditions (A.1)-(A.6) in Banerjee et al. (2009) hold and* $\phi_m + \psi_m > 1$ *for* $m = 1, \ldots, k$. *Then* $\hat{\theta}_n$ *has an asymptotic normal distribution*

$$\sqrt{n}(\hat{\theta}_n - \theta) \rightarrow_d N(0, I_0^{-1}) \tag{9}$$

Moreover, under $H_0 : \theta = \theta_0 \in R^r$

$$2\log(\lambda_n) \;=\; 2\log\left[\frac{L_n(\hat{\theta}_n, \hat{\Lambda}_n)}{L_n(\theta_0, \hat{\Lambda}_n^{\theta_0})}\right] = 2\left[l_n(\hat{\theta}_n, \hat{\Lambda}_n)) - l_n(\theta_0, \hat{\Lambda}_n^{\theta_0})\right] \rightarrow_d \chi_r^2 \tag{10}$$

where λ_n *is the likelihood ratio statistic and* $\hat{\Lambda}_n^{\theta_0}$ *is the NPMLE of* Λ *under* H_0.

An explicit computation of I_0^{-1} when only one test is available ($k = 1$) can be found in Sal y Rosas and Hughes (2010). The proof for the general case follows a similar argument and it is available upon request. In principle, one could use (9) to compute confidence intervals for the regression coefficients; however, that would involve estimation of several additional nuisance parameters. Instead, we compute likelihood ratio based confidence intervals for the regression coefficient by computing the rejection region of the associated hypothesis testing problem.

Remark 4. We do not specify the limit distribution of the estimated cumulative hazard function ($\hat{\Lambda}_n$). In most of the literature on semiparametric models with order restrictions on the nuisance parameter, the likelihood function is concave with respect to the nuisance parameter. In those scenarios, finding the asymptotic behavior of $\hat{\Lambda}_n$ is possible using techniques from isotonic regression (Robertson, Wright, and Dykstra, 1998) and convex optimization (Rockafellar, 1970). When current status data is subject to outcome misclassification, that concavity property does not always hold and depends on the value of the sensitivity (but not specificity). Therefore, the asymptotic behavior of the cumulative hazard function remains to be found. However, the main objective of this paper is to obtain an accurate estimation of the regression coefficient, adjusted for the baseline hazard. Therefore, we postpone this problem for future research.

We now propose an algorithm to compute $(\hat{\theta}_n, \hat{\Lambda}_n)$. To avoid excessive notation we will consider the case of a single covariate ($r = 1$).

Algorithm 3. Estimation of the Regression Parameters

If the true disease status Δ is observed, then one can estimate θ using the profile approach proposed by Huang (1996) or the joint maximization idea proposed

DOI: 10.2202/1948-4690.1032

- 10.2202/1948-4690.1032
Downloaded from De Gruyter Online at 09/12/2016 10:40:35PM
via South China University of Technology

by Pan (1999). If, instead, one observes $\tilde{\Delta}$ then we propose estimating θ using an EM algorithm (Dempster, Laird, and Rubin, 1977) as described below.

1. **Initial value**. Let $\theta_n^{(0)}$ and $\hat{\Lambda}_n^{(0)}$ be the initial estimates of θ and Λ_0, respectively. For example, one can set $\hat{\Lambda}_n^{(0)}(Y_i) = -\log(1-i/(n+1))$, for $i = 1,\ldots,n$, and then estimate $\theta_n^{(0)}$ by maximizing the complete data log likelihood $l_n^C(\cdot,\hat{\Lambda}_n^{(0)})$ with respect to the regression coefficients, where

$$l_n^C(\theta,\Lambda) = \sum_{i=1}^{n}\left\{\Delta_i \log\left[1-e^{-\Lambda(Y_i)e^{Z_i\theta}}\right] - (1-\Delta_i)\Lambda(Y_i)e^{Z_i\theta}\right\} \qquad (11)$$

   Let $(\hat{\theta}_n^{(k)},\hat{\Lambda}_n^{(k)})$ be the current estimates of (θ,Λ_0) at iteration k. Then at iteration $k+1$ one implement two steps:
2. **Expectation step**. Let $W = (\tilde{\Delta},Y,Z,\mathbf{U} = (0,\ldots,1,\ldots,0))$ where $\mathbf{U}$ is 1 at the mth position. Then

$$E_{(k)}[\Delta \mid W] = \begin{cases} \dfrac{P^{(k)}[\Delta=1 \mid Y,Z]\phi_m}{P^{(k)}[\Delta=1 \mid Y,Z]\phi_m + P^{(k)}[\Delta=0 \mid Y,Z](1-\psi_m)}\ , \ \tilde{\Delta}=1 \\ \dfrac{P^{(k)}[\Delta=1 \mid Y,Z](1-\phi_m)}{P^{(k)}[\Delta=1 \mid Y,Z](1-\phi_m) + P^{(k)}[\Delta=0 \mid Y,Z)]\psi_m}\ , \ \tilde{\Delta}=0 \end{cases} \qquad (12)$$

   where

$$P^{(k)}[\Delta=1 \mid Y,Z] = 1-\exp\left[-\hat{\Lambda}_n^{(k)}(Y)\exp(Z\hat{\theta}_n^{(k)})\right]$$

3. **Maximization step**. Update the parameters according to

$$\begin{aligned}(\hat{\theta}_n^{(k+1)},\hat{\Lambda}_n^{(k+1)}) &= \arg\max_{\theta\in\Theta\ ,\Lambda\in\mathscr{G}} Q\left[(\theta,\Lambda) \mid \hat{\theta}_n^{(k)},\hat{\Lambda}_n^{(k)}\right] \\ &= \arg\max_{\theta\in\Theta\ ,\Lambda\in\mathscr{G}} l_n^C(E_{(k)}[\Delta_i \mid \tilde{\Delta}_i,Y_i,Z_i],Y_i,Z_i)\end{aligned}$$

   For the maximization step we use the profile idea described by Huang (1996).
4. **Stopping criteria**. We alternate between the expectation and maximization steps until the following stopping criteria holds

$$\left|\frac{l_n(\hat{\theta}_n^{(k+1)},\hat{\Lambda}_n^{(k+1)}) - l_n(\hat{\theta}_n^{(k)},\hat{\Lambda}_n^{(k)})}{l_n(\hat{\theta}_n^{(k)},\hat{\Lambda}_n^{(k)})}\right| \leq \varepsilon$$

   where $\varepsilon > 0$ is the tolerance level set in advance.

- 10.2202/1948-4690.1032
Downloaded from De Gruyter Online at 09/12/2016 10:40:35PM
via South China University of Technology

3 Simulations

We conducted simulation studies to (1) assess the bias and misinterpretation of inference results when one ignores outcome misclassification and (2) assess the behavior of the proposed estimators for small sample sizes. These two objectives will be studied for different outcome prevalences, levels of misclassification and observation time distributions.

3.1 Simulations for the one sample problem

For the observation times, we consider continuous uniform and exponential distribution functions. For the distribution of failure times, we use a standard exponential distribution. We consider sample sizes $n = 500$ and $n = 1000$ with $R = 1000$ simulations per scenario. We denote by p the expected proportion of left censoring observations and adjust the distribution of the observation times to achieve a fixed value of p ($p = P(T \leq Y)$). At $t_0 = G^{-1}(0.5)$ (median observation time), we compute the asymptotic percent bias (b), defined as

$$b = 100 \times \frac{1}{R} \sum_{r=1}^{R} \left[\frac{\hat{F}_n^{(r)}(t_0) - F_0(t_0)}{F_0(t_0)} \right]$$

and the nominal coverage of the 95% likelihood ratio-base confidence interval (γ). Table 1 provides percent bias and coverage of selected estimators when the expected number of failures is 10%. As expected, the naive estimator is biased. In particular, for low prevalence diseases (small number of true failures observed) the bias is the greatest when the specificity is low and bias is less affected by low sensitivity. Bias of the adjusted estimator is small and decreases as n increases while the unadjusted estimator remains biased regardless of n. The coverage of the proposed confidence interval is very good and the average length of the confidence interval is shorter, as expected, when a portion of the sample is tested with a perfect laboratory test. Overall, the adjusted estimators have little bias and good coverage.

DOI: 10.2202/1948-4690.1032

- 10.2202/1948-4690.1032
Downloaded from De Gruyter Online at 09/12/2016 10:40:35PM
via South China University of Technology

Table 1: Percent bias and coverage to estimate $F(t_0)$, where $t_0 = G^{-1}(0.5)$ is the median of the observation times distribution. Sample sizes of 500 and 1000 observations, with 10% expected failures, were considered. Results are based on 1000 simulations

	$(n = 500)$						$(n = 1000)$					
(ϕ, ψ)	b^1	b^2	b^3	γ^2	γ^3	$E^{3,2}$	b^1	b^2	b^3	γ^2	γ^3	$E^{3,2}$
A. Uniform observation times												
$(1, 0.8)$	175.6	-1.2	-2.1	0.944	0.948	93.0	177.2	0.8	0.1	0.947	0.945	92.2
$(1, 0.7)$	264.3	-0.6	-5.4	0.944	0.935	90.0	263.6	-1.8	-3.7	0.950	0.944	88.5
$(0.8, 1)$	-24.5	-5.7	-5.4	0.946	0.951	99.0	-23.5	-4.3	-4.2	0.950	0.942	98.8
$(0.7, 1)$	-34.9	-7.0	-6.9	0.947	0.940	99.1	-33.1	-4.4	-4.2	0.949	0.945	98.8
$(0.8, 0.9)$	67.4	-1.3	-3.6	0.946	0.947	92.9	66.7	-2.3	-3.3	0.963	0.941	92.7
$(0.7, 0.8)$	56.4	-3.1	-4.1	0.941	0.945	91.4	56.6	-2.8	-4.2	0.948	0.941	90.1
B. Exponential observation times												
$(1, 0.8)$	252.8	4.4	-1.3	0.952	0.942	92.1	252.0	2.7	-0.5	0.954	0.947	91.2
$(1, 0.7)$	381.8	11.4	1.5	0.948	0.943	89.2	378.4	5.4	-2.4	0.949	0.940	86.9
$(0.8, 1)$	-24.4	-5.5	-5.4	0.942	0.944	99.0	-22.6	-3.3	-3.3	0.948	0.952	98.7
$(0.7, 1)$	-34.1	-5.8	-5.6	0.939	0.935	99.0	-32.3	-3.3	-3.4	0.950	0.952	98.7
$(0.8, 0.9)$	104.7	0.1	-2.6	0.944	0.933	93.2	104.7	-0.2	-2.4	0.948	0.949	91.8
$(0.7, 0.8)$	95.9	2.3	-3.7	0.944	0.942	90.4	94.7	-0.2	-2.6	0.946	0.938	89.8

n = sample size, ϕ = sensitivity, and ψ = specificity
b: Percent bias, γ: Observed coverage for a nominal 95% CI
$E^{3,2}$= Average c.i length [3] / Average c.i length [2]
[1] One lab test; ϕ, ψ assumed to be 1 (i.e. naive estimator)
[2] One lab test; ϕ, ψ assumed to be true values
[3] Two lab tests; first used on 90% has misclassification rates ϕ, ψ (assumed known); the second is a prefect test

Downloaded from De Gruyter Online at 09/12/2016 10:40:35PM
via South China University of Technology
- 10.2202/1948-4690.1032

3.2 Simulations for two sample hypothesis testing

We assume that the failure time distribution in the control group is exponential with hazard rate equal to one. The observation times follow a continuous uniform distribution for both groups, and the expected proportion of observed failures for the control is 13%. We compute the observed proportion of rejections under the null and proportional hazard alternative hypothesis for situations of nondifferential and differential misclassification. Nondifferential (differential) misclassification means that the misclassification rates do not depend (do depend) on the group assignment.

Our simulations, presented in table 2, suggest that for nondifferential misclassification, the adjusted test (controlling for misclassification) is somewhat more conservative than the unadjusted test (ignoring the misclassification). However, as predicted by the asymptotic theory, as the sample sizes increases, this difference diminishes. Still, the unadjusted test behaves better in most of the studied scenarios when misclassification is not differential. The adjusted test is most conservative for low levels of specificity.

This behavior can be understood by studying the $k = 1$ case. For a disease with low prevalence, $\phi = 1$ and $\psi < 1$, (2) reduces to $\hat{F} = [\tilde{F} \wedge (1-\psi) + \psi - 1]\psi^{-1}$ and that implies that $\hat{F}_n = 0$ whenever $\tilde{F}_n < 1 - \psi$ which will be true in many cases for a disease with low prevalence. That means that the adjusted variance estimator $\hat{I}_n^{-1}$ (11) relies less on the data and more on the assumed values of ϕ and ψ in comparison with the unadjusted estimator that only depends on the data. On the other hand, when $\phi < 1$ and $\psi = 1$, (2) reduces to $\hat{F} = \tilde{F}\phi^{-1}$ and $I^{-1} = \phi^{-2}\tilde{I}^{-1}$ and as a result the test statistic will not be affected by the sensitivity of the laboratory test.

Under differential misclassification, the adjusted test statistic preserves the correct type I error rate under the null hypothesis, in contrast to the naive test statistic that does not preserve the type I error rate at all. We observe that the estimated power of the adjusted test statistic is affected the most for low specificity levels and in particular for low sample sizes. However, as the sample size increases, the power of the proposed test increases regardless of the misclassification levels (see table 3).

As a consequence, when misclassification is not differential and only one laboratory test is used, we recommend ignoring misclassification and computing a test statistic and p-value based on the unadjusted data. However, to estimate the d.f. for each group, the adjusted estimators $\hat{F}_{n_1}$ and $\hat{F}_{n_0}$ should be used. If more than one laboratory test are used, then appropriated adjustment is needed and the use of the adjusted test statistics is the appropriated option. If misclassification is differential, we recommend using the adjusted test statistic always.

DOI: 10.2202/1948-4690.1032

- 10.2202/1948-4690.1032
Downloaded from De Gruyter Online at 09/12/2016 10:40:35PM
via South China University of Technology

Table 2: Non differential misclassification: Size and power of the proposed test with samples sizes of 500, 1000, and 1500 observations, 13% of expected failures in the control group, observations are randomized 1:1 to each group, and 1000 repetitions were implemented in each scenario. Naive estimator assumes a perfect test ($(\phi,\psi)=(1,1)$) and the adjusted estimator assumes the correct levels of sensitivity and specificity.

(ϕ,ψ)	$n=500$		$n=1000$		$n=1500$	
	Naive	Adjusted	Naive	Adjusted	Naive	Adjusted
A. One laboratory test test						
A.1. Under H_0						
$(1,0.9)$	0.052	0.040	0.056	0.048	0.055	0.043
$(0.9,1)$	0.058	0.057	0.053	0.052	0.054	0.054
$(0.9,0.9)$	0.043	0.032	0.049	0.039	0.045	0.043
$(0.8,0.9)$	0.061	0.045	0.049	0.039	0.052	0.042
A.2. Under $H_a: HR=0.5$						
$(1,0.9)$	0.224	0.197	0.373	0.356	0.482	0.472
$(0.9,1)$	0.351	0.352	0.546	0.546	0.750	0.750
$(0.9,0.9)$	0.181	0.141	0.323	0.304	0.419	0.411
$(0.8,0.9)$	0.143	0.120	0.261	0.242	0.318	0.302
B. Two laboratory test (50% of the observations with each test)						
B.1. Under H_0						
$(1,0.9)$, $(1,1)$	0.056	0.057	0.065	0.067	0.047	0.047
$(0.9,1)$, $(1,1)$	0.059	0.057	0.048	0.048	0.050	0.050
$(0.9,0.9)$, $(0.9,1)$	0.061	0.052	0.056	0.042	0.062	0.047
$(0.8,0.9)$, $(0.9,0.9)$	0.051	0.044	0.051	0.046	0.055	0.046
B.2. Under $H_a: HR=0.5$						
$(1,0.9)$, $(1,1)$	0.944	0.942	0.999	0.999	1.000	1.000
$(0.9,1)$, $(1,1)$	0.930	0.929	0.999	0.998	1.000	1.000
$(0.9,0.9)$, $(0.9,1)$	0.764	0.780	0.95	0.968	0.995	0.995
$(0.8,0.9)$, $(0.9,0.9)$	0.538	0.498	0.825	0.818	0.932	0.930

- 10.2202/1948-4690.1032
Downloaded from De Gruyter Online at 09/12/2016 10:40:35PM
via South China University of Technology

Table 3: Differential misclassification (Different test for each group): Size and power of the proposed test with samples sizes of 500, 1000, and 1500 observations, 13% of expected failures in the control group, observations are randomized 1:1 to each group, and 1000 repetitions were implemented in each scenario. Naive estimator assumes a perfect test ($(\phi,\psi)=(1,1)$) and the adjusted estimator assumes the correct levels of sensitivity and specificity.

(ϕ,ψ)		$n=500$		$n=1000$		$n=1500$	
		Naive	Adjusted	Naive	Adjusted	Naive	Adjusted
A. Under H_0							
Control	Intervention						
(0.9,0.95)	(0.7,0.95)	0.125	0.048	0.215	0.048	0.287	0.046
(0.95,0.9)	(0.95,0.8)	0.869	0.057	0.993	0.053	1.000	0.053
(0.9,0.8)	(0.9,0.8)	0.517	0.042	0.807	0.050	0.910	0.069
B. Under $H_a: HR=0.5$							
Control	Intervention						
((0.9,0.95)	(0.7,0.95)	0.838	0.592	0.982	0.892	1.000	0.981
(0.95,0.9)	(0.95,0.8)	0.383	0.286	0.651	0.625	0.814	0.804
(0.9,0.8)	(0.9,0.8)	0.102	0.241	0.131	0.582	0.190	0.749

3.3 Simulations for regression models

For our regression simulations, we assume that the baseline hazard is a standard Weibull distribution ($\Lambda_0(t)=(t/\alpha)^k$ where α and k are the rate and scale parameters, respectively). We consider one or two binary covariates, each with probability of success equal to 0.5 and fix the sample size at 500 observations. We generate the observation times from a uniform distribution such that the expected number of left censored observations in the baseline group (for more than one covariate this group is defined by assigning all covariates equal to zero) was 10%. The number of replications is 2000 in all scenarios. We consider the model $\Lambda(t)=(t/\alpha)^k\exp(\theta_1 Z_1+\theta_2 Z_2)$, where Z_1 and Z_2 are independent $Bernoulli(0.5)$ random variables. Finally, when applying the EM algorithm, we did not implement the full maximization step and instead we only require that the M-step estimate increases the likelihood.

Table 4 shows that ignoring outcome misclassification induces attenuation of the regression coefficients. Moreover, the higher the misclassification the stronger

DOI: 10.2202/1948-4690.1032

- 10.2202/1948-4690.1032
Downloaded from De Gruyter Online at 09/12/2016 10:40:35PM
via South China University of Technology

this attenuation is. For low prevalence, we observe that the regression coefficients of the naive estimator are more affected by low levels of specificity than sensitivity.

In our simulations, we observed that between 0.5% and 1% of the generated datasets did not converge. The convergence criteria had two components: (a) the stopping criteria of the EM algorithm (step 4 of **Algorithm 3**), and (b) the Hessian matrix at the EM solution being negative definite). Unexpectedly, among the simulations that converge to a solution, the NPMLE of the regression coefficients (adjusting for misclassification) is slightly biased upwards (based on the mean of the NPMLE across simulations), but the median is closer to the true value. Pan (1999) mentioned that for Cox regression in interval censored data (without misclassification) the ICM was slightly biased upwards. Similarly, Shiboski (1998) described results where the regression coefficients are biased for generalized additive models for current status data. This mean bias appears to arise largely from the result of a small number of simulations where the Hessian at the EM solution is nearly singular.

Table 4: Mean estimates of the Cox regression coefficients and size of the likelihood ratio test, for the regression model $\Lambda(t) = (t/\alpha)^k \exp(\theta_1 Z_1 + \theta_2 Z_2)$, where $(\theta_1, \theta_2) = (\log(0.5), \log(1.5)) = (-0.695, 0.405)$, α and k are the scale and shape parameters of the standard Weibull distribution, Z_1 and Z_2 are independent bernoulli random variables with probability of success 0.5, samples of 500 observations, 10% expected failures for the baseline group ($Z_1 = Z_2 = 0$), and 2000 simulations.

		Naive		Adjusted		
(ϕ, ψ)	(α, k)	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_1^+$	$\hat{\theta}_2^+$	η
(0.90,0.95)	(1,0.5)	-0.437	0.260	-0.766 [-0.719]	0.437 [0.414]	0.060
(0.80,0.95)		-0.413	0.245	-0.784 [-0.712]	0.448 [0.424]	0.060
(0.90,0.90)		-0.304	0.187	-0.788 [-0.711]	0.480 [0.411]	0.053
(0.90,0.95)	(1,1)	-0.442	0.267	-0.776 [-0.729]	0.448 [0.430]	0.075
(0.80,0.95)		-0.412	0.249	-0.771 [-0.717]	0.443 [0.416]	0.058
(0.90,0.90)		-0.307	0.189	-0.784 [-0.719]	0.467 [0.426]	0.055
(0.90,0.95)	(1,1.5)	-0.446	0.264	-0.779 [-0.736]	0.439 [0.414]	0.063
(0.80,0.95)		-0.419	0.249	-0.779 [-0.727]	0.444 [0.402]	0.061
(0.90,0.90)		-0.310	0.189	-0.799 [-0.729]	0.460 [0.426]	0.062

$^+$: Mean [Median]

η: Estimated size of the likelihood ratio test for testing $H_0 : \theta_1 = \log(0.5)$

- 10.2202/1948-4690.1032
Downloaded from De Gruyter Online at 09/12/2016 10:40:35PM
via South China University of Technology

4 Application

The Partner Notification Study was conducted in King County Seattle, WA from September 1998 to March 2003 and enrolled heterosexual men and women who received a diagnosis of gonorrhea or genital chlamydia (Golden et al., 2005). Researchers contacted clinicians who diagnosed and treated the infections to seek permission to contact their participants. To minimize the likelihood of reinfection before randomization, patients who could not be contacted within 14 days after treatment were not eligible for the study. Each participant was randomized to expedited partner treatment (intervention) or standard partner referral (control). The primary outcome was persistent or recurrent gonorrhea and/or chlamydial infection (we consider a composite outcome only) in the original participant at 90 days after enrollment although actual follow up times varied considerably (19 to 116 days) due to difficulty contacting participants and scheduling follow-up visits.

Table 5: Descriptive statistics: Partners Notification study

N(%)	Control N=933 (50.0)	Intervention N=931 (50.0)	Total N=1864 (100.0)
Female	731 (78.3)	736 (78.5)	1467 (78.7)
Age (years) [a]	21 [19-25]	22 [19-26]	21 [19-25]
Initial diagnoses			
Gonorrhea	133 (14.3)	132 (14.2)	265 (14.2)
Genital chlamydia	752 (80.6)	752 (80.8)	1504 (80.7)
Both	48 (5.1)	47 (5.0)	95 (5.1)
Events	122 (13.1)	92 (9.9)	214 (11.5)
Observation time (days) [a]	87 [77-103]	87 [76-104]	87 [77-103]
Regression Analysis			
Factors	Univariate Analysis HR [95 %CI], P-value		Multivariate Analysis HR (P-value)
Intervention	0.744 [0.566,0.970], 0.031		0.534
Gender	1.283 [0.950,1.914], 0.121		1.126
Gender × Intervention	-		1.479 (0.302)

[a] Median [interquartile range], and HR= Hazard ratio

- 10.2202/1948-4690.1032
Downloaded from De Gruyter Online at 09/12/2016 10:40:35PM
via South China University of Technology

Of the 1864 participants, 931 were randomized to the intervention and 933 to the control group. A high proportion of participants were women (80%), the median age was 21 years and a large number of participants who were treated for chlamydia were enrolled in the study (see table 5); however, all these characteristics were balanced in each arm. The sensitivity and specificity of the tests used to diagnose either gonorrhea or chlamydia were approximately 0.9 and 1.0, respectively (Carroll et al., 1998). In order to avoid missing an infection that could have happened between enrollment and the observation time, participants were asked whether they repeated their treatment using medication intended for a partner; only one person acknowledge doing so. The observation times was similar in both groups with median 87 days (IQR: 77-103) for the control group and 87 days (IQR: 76-104) for the intervention group.

We compute the NPMLE separately for the control and intervention group. This is presented in Figure 1. The application of the proposed two sample hypothesis test gave a p-value of 0.024. The result is similar, as we showed in our theoretical results in the previous section, to the naive estimator that ignores outcome misclassification (p-value = 0.024). Therefore, one can conclude that there is evidence that the mean difference between the survival functions the intervention and control groups is not zero.

In an adjusted univariate analysis, participants in the intervention group were 26% less likely to experience reinfection than participants in the control group (HR=0.744 [95%CI: 0.566-0.970], p-value = 0.031). This result is similar to the unadjusted analysis (HR=0.743 95%CI: 0.571-0.970], p-value = 0.031) and confirms that for diseases with low prevalence, low sensitivity does not affect the estimates as much as low specificity.

In an adjusted multivariate analysis, including gender and the interaction of gender and intervention, there was no evidence that the effect of the intervention was different for men and women (p-value of the iteraction = 0.302) and a similar result was obtained in the unadjusted analysis (p-value of the interaction = 0.300). In conclusion, participants in the intervention group have significant lower risk of recurrence of gonorrhea and/or chlamydial infection. Similar results were obtained when ignoring misclassification as low sensitivity has limited effect when analyzing low prevalent diseases.

- 10.2202/1948-4690.1032
Downloaded from De Gruyter Online at 09/12/2016 10:40:35PM
via South China University of Technology

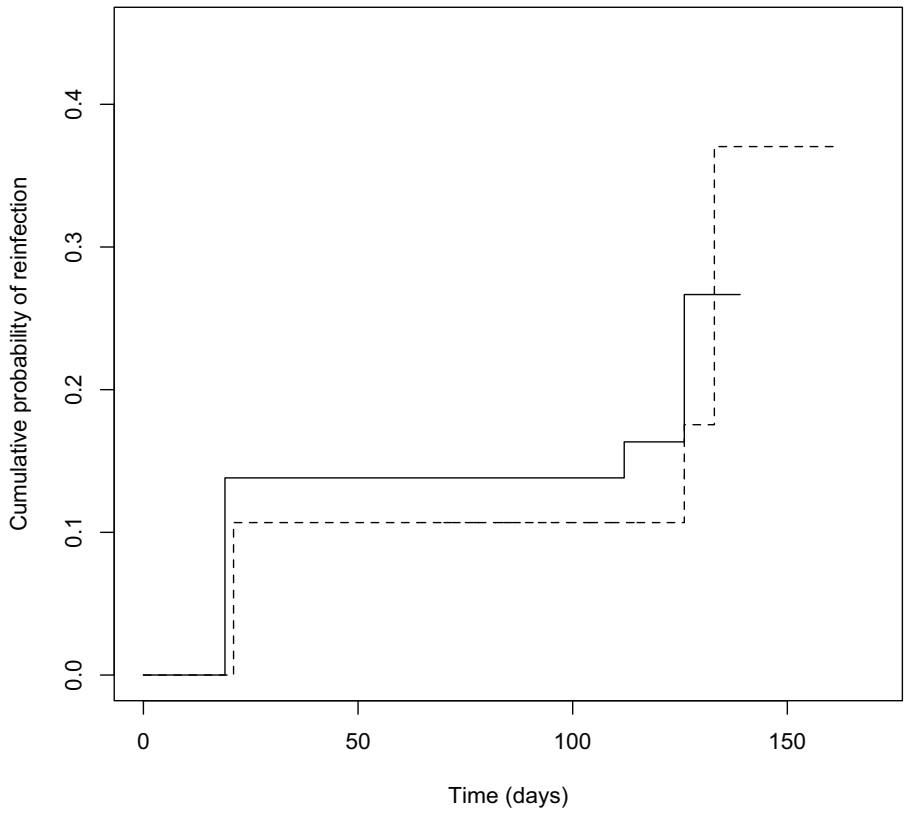

Figure 1: Estimated cumulative probability of reinfection for participants in the placebo (solid line) and intervention (dashed line) arm.

5 Discussion and Future Research

Most laboratory tests that are used to diagnose diseases have sensitivity and specificity less than 1.0 (e.g culture for gonorrhea, sputum-smear test for TB). On average, if the observed disease prevalence during the study follow-up is low (high) and specificity (sensitivity) is less than 1.0 then the use of the standard methodology that does not account for outcome misclassification results in overestimation (underestimation) of the cumulative probability of failure.

We develop methodology for one sample estimation, two sample hypothesis testing and semiparametric regression that account for outcome misclassification in current status data by extending existing models that assume no misclassification, and allowing the misclassification rates to vary by groups. For the two sample

DOI: 10.2202/1948-4690.1032

- 10.2202/1948-4690.1032
Downloaded from De Gruyter Online at 09/12/2016 10:40:35PM
via South China University of Technology

problem, we consider a test statistic based on the difference of survival means, extending the idea described by Andersen and Ronn (1995). This test is not the only one that researchers had studied for current status data, some other examples are log rank type tests (Sun, 1996, Sun and Kalbfleish, 1993, 1996, Sun, 1999), and likelihood ratio or score test based on specific types of alternative hypotheses (Kulikov, 2002).

For the regression problem, we choose the Cox proportional hazard model because of its broadly applicability. However, the same ideas can be studied in other regression models such the proportional odds (Rossini and Tsiatis, 1996), acelerated failure time (Tian and Cai, 2006), linear (Shen, 2000), additive hazard (Lin, Oakes, and Ying, 1998), and generalized additive hazard (Shiboski, 1998). In some situations (e.g predicting survival), one would like to make inference on the cumulative hazard function when considering a regression model. Finding the correct limit distribution of the cumulative hazard function under misclassification remains an open problem.

There is an important limitation in the analysis of our application example. A number of participants reported symptoms at their follow up visit (n=408, 22.3%) but this percentage did not differ meaningfully between the groups (21.1 % in the placebo group vs. 23.5% in the treatment group). This feature of the data may suggest a violation of the assumption of independence of the failure and observation times. However, we also analyzed the data after deleting the symptomatic cases and the results were qualitatively similar (data not shown). Estimation and inference in this setting (adding information on symptoms) is under investigation.

In the hypothesis testing problem, we only consider situations when one can assume that the observation times are not different between groups. It is reasonable to assume that with some work our methods can also be extended to that situation where the observation times differ between groups. For example, for the case of nonmisclassification, Sun (1999) proposed modelling the observation times using a proportional hazard model.

Similar to previous authors, the methodology presented in this paper assumes perfect knowledge of the values of sensitivity and specificity but allows them to differ by groups. We conjecture that allowing sensitivity and specificity to vary at the individual level will require additional assumptions on the limit distribution of the misclassification rates and is a future area of research. Another important potential extension is to incorporate uncertainty or vague prior information about ϕ and ψ, and to estimate ϕ, and ψ when gold standard data, or multiple test results in each sample, are available.

- 10.2202/1948-4690.1032
Downloaded from De Gruyter Online at 09/12/2016 10:40:35PM
via South China University of Technology

References

Andersen, P. K. and B. B. Ronn (1995): "A nonparametric test for comparing two samples where all observations are either left- or right-censored," *Biometrics*, 51, 323–329.

Balasubramanian, R. and S. W. Lagakos (2003): "Estimation of a failure time distribution based on imperfect diagnostic test," *Biometrika*, 90, 171–182.

Balasubramanian, S. and S. W. Lagakos (2001): "Estimation of the timing of perinatal transmission of HIV," *Biometrics*, 57, 1048–1058.

Banerjee, M. (2007): "Likelihood based inference for monotone response models," *The Annals of Statistics*, 35, 931–956.

Banerjee, M., M. D., and S. Mishra (2009): "Semiparametric binary regression models under shape constraints with application to Indian schooling data," *Journal of Econometrics*, 249, 101–117.

Banerjee, M. and J. Wellner (2001): "Likelihood ratio tests for monotone functions," *The Annals of Statistics*, 29, 1699–1731.

Carroll, K. C., W. E. Aldeen, M. Morrison, R. Anderson, D. Lee, and S. Mottice (1998): "Evaluation of the Abbott LCx ligase chain reaction assay for detection of Chlamydia trachomatis and Neisseria gonorrhoeae in urine and genital swab specimens from a sexually transmitted disease clinic population," *Journal of Clinical Microbiology*, 36.

Dempster, A., N. Laird, and D. Rubin (1977): "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society. Series B(Methodological)*, 39, 1–38.

Golden, M., W. Whittington, H. Handsfield, J. Hughes, W. Stamm, M. Hogben, A. Clark, C. Malinski, J. Helmers, K. Thomas, and K. Holmes (2005): "Effect of expedited treatment of sex partners on recurrent or persistent gonorrhea or chlamydial infection," *New England Journal of Medicine*, 352, 676–685.

Groeneboom, P. and J. Wellner (1992): *Information Bounds and Nonparametric Maximum Likelihood Estimation*, Birkhauser Verlag.

Huang, J. (1996): "Efficient estimation for the proportional hazard model with interval censoring," *The Annals of Statistics*, 24, 540 – 568.

Jongbloed, G. (1998): "The iterative convex minorant algorithm for nonparametric estimation," *Journal of Computational and Graphical Statistics*, 7, 310 – 321.

Kulikov, V. N. (2002): *Direct and Indirect Use of Maximum Likelihood*, Ph.D. thesis, Delft University.

Lin, D., D. Oakes, and Z. Ying (1998): "Additive hazard regression with current status data," *Biometrika*, 85, 289 – 298.

McKeown, K. and N. P. Jewell (2010): "Misclassification of current status data," *Lifetime Data Analysis*, 16, 215–230.

DOI: 10.2202/1948-4690.1032

- 10.2202/1948-4690.1032
Downloaded from De Gruyter Online at 09/12/2016 10:40:35PM
via South China University of Technology

Meier, A., B. Richardson, and J. P. Hughes (2003): "Discrete proportional hazard models for mismeasured outcomes," *Biometrics*, 59, 947 – 954.

Neuhaus, J. M. (2002): "Analysis of clustered and longitudinal binary data subject to response misclassification," *Biometrics*, 58, 675 – 683.

Pan, W. (1999): "Extending the iterative convex minorant algorithm to the Cox model for interval-censored Data," *Journal of Computational and Graphical Statistics*, 8, 109–120.

Richardson, B. A. and J. P. Hughes (2000): "Product limit estimation for infections diseases data when the diagnostic test for the outcome is measured with uncertainty," *Biostatistics*, 1, 341–354.

Robertson, T., F. T. Wright, and R. Dykstra (1998): *Order restricted statistical inference*, Chichester ; New York : Wiley.

Rockafellar, R. T. (1970): *Convex Analysis*, Princeton Univ. Press.

Rossini, A. J. and A. A. Tsiatis (1996): "A semiparametric proportional odds regression model for the analysis of current status data," *Journal of the American Statistical Association*, 91, 713 – 721.

Sal y Rosas, V. and J. Hughes (2010): "Nonparametric and semiparametric analysis of current status data subject to outcome misclassification," Working Paper 364, University of Washington.

Shen, X. (2000): "Linear regression with current status data," *Journal of the American Statistical Association*, 95, 842 – 852.

Shiboski, S. (1998): "Generalized additive models for current status data," *Lifetime Data Analysis*, 4, 29–50.

Sun, J. (1996): "A nonparametric test for interval-censored failure time data with application to AIDS studies," *Statistics in Medicine*, 15, 1387 – 1395.

Sun, J. (1999): "A nonparametric test for current status data with unequal censoring," *Journal of the Royal Statistical Society: Series B*, 61, 243 – 250.

Sun, J. and J. D. Kalbfleish (1993): "The analysis of current status data on point process," *Journal of the American Statistical Association*, 88, 1449 – 1454.

Sun, J. and J. D. Kalbfleish (1996): "Nonparametric test of tumor prevalence data," *Biometrics*, 52, 726 – 731.

Tian, L. and T. Cai (2006): "On the accelerated failure time model for current status and interval censored data," *Biometrika*, 93, 329 – 342.

- 10.2202/1948-4690.1032
Downloaded from De Gruyter Online at 09/12/2016 10:40:35PM
via South China University of Technology