Biometrics 59, 947–954
December 2003

Discrete Proportional Hazards Models for Mismeasured Outcomes

Amalia S. Meier,[1,*] Barbra A. Richardson,[2] and James P. Hughes[2]

[1]Program in Infectious Diseases, Fred Hutchinson Cancer Research Center, Seattle, Washington, U.S.A.
[2]Department of Biostatistics, University of Washington, Box 357232, Seattle, Washington, U.S.A.
[*]*email:* ameier@u.washington.edu

Summary. Outcome mismeasurement can lead to biased estimation in several contexts. Magder and Hughes (1997, *American Journal of Epidemiology* **146,** 195–203) showed that failure to adjust for imperfect outcome measures in logistic regression analysis can conservatively bias estimation of covariate effects, even when the mismeasurement rate is the same across levels of the covariate. Other authors have addressed the need to account for mismeasurement in survival analysis in selected cases (Snapinn, 1998, *Biometrics* **54,** 209–218; Gelfand and Wang, 2000, *Statistics in Medicine* **19,** 1865–1879; Balasubramanian and Lagakos, 2001, *Biometrics* **57,** 1048–1058, 2003, *Biometrika* **90,** 171–182). We provide a general, more widely applicable, adjusted proportional hazards (APH) method for estimation of cumulative survival and hazard ratios in discrete time when the outcome is measured with error. We show that mismeasured failure status in a standard proportional hazards (PH) model can conservatively bias estimation of hazard ratios and that inference, in most practical situations, is more severely affected by poor specificity than by poor sensitivity. However, in simulations over a wide range of conditions, the APH method with correctly specified mismeasurement rates performs very well.

Key words: Measurement error; Proportional hazards; Sensitivity; Specificity; Survival.

1. Introduction

1.1 *Motivation*

When binary outcomes are measured with error, the subject may be misclassified. For example, in survival analysis, if the diagnostic tool used to measure failure is not perfectly sensitive and specific, it may not be indicative of the subject's true event status. As we will show, failure to account for this mismeasurement may result in incorrect inferences.

Many types of outcomes have inherent mismeasurement which under ideal conditions can be ignored. More substantial mismeasurement in these contexts may result, due to the following reasons. Diagnostic tools that are highly sensitive and specific may be too expensive to use routinely or may require any of the following for optimal performance: special handling of specimens (refrigeration, immediate processing), technologically advanced instruments, highly trained operators, or stringent laboratory conditions. These conditions may be difficult to achieve in some settings. In other cases, imperfect screening tests are used when the gold standard diagnostic tools are unnecessarily invasive, requiring biopsy or surgery (e.g., PSA screening for prostate cancer), or may only be performed post mortem (e.g., Alzheimer's disease). Still other tests have known rates of failure, but remain standards of care, as no less risky alternatives have been developed (for example, triple serum screening for Down's syndrome, followed by amniocentesis). Methods that can provide accurate inferences when the outcome is measured with error are necessary.

1.2 *Measurement Error in Survival Outcomes*

Mismeasured binary outcomes in survival analysis have been studied in a few contexts. Snapinn (1998) introduced the use of "auxiliary variables" (covariates that help to define the relationship between the true and observed outcomes) in Cox regression analysis with mismeasured outcomes. In this approach, prior distributions for the auxiliary variables conditioned on the true outcomes are required. Snapinn (1998) applied this method to mismeasurement resulting from adjudication by an endpoint committee (for example, assessment of myocardial infarction).

Gelfand and Wang (2000) proposed a method of quantifying the cumulative risk of false positive outcomes when subjects are repeatedly screened for an event. The goal in that analysis was not estimation of the risk of an *event*, but estimation of the false positive rate. Accurate estimation of the cumulative risk of false positive results aided in assessing the appropriate frequency of screening or target populations for screening.

Balasubramanian and Lagakos (2001) developed a procedure for estimating the risk of perinatal transmission of HIV-1 during the late stages of pregnancy, in the context of changing test sensitivity. All testing occurred after the period of exposure had ended (i.e., after birth). The sensitivity of the polymerase chain reaction (PCR) test for HIV-1 was assumed to vary with time since infection. Their method also allowed estimation of risk factors for transmission. In a more recent article (Balasubramanian and Lagakos, 2003.) the authors

developed hazard estimation procedures for a context in which periods of exposure vary. The example used was mother-to-child transmission of HIV during pregnancy or via breast milk, the latter being an exposure that changes over time. Inclusion of covariates was not addressed, except in the discussion section as a possible extension.

Richardson and Hughes (2000) derived an estimation maximization (EM) algorithm for the product limit estimate of the survivor function with no covariates when the binary outcome measure was subject to error. Their methods were designed for discrete-time contexts in which all subjects were tested at predetermined time points, and included confirmatory testing for positive screening test outcomes. Subjects were assumed to enroll prior to failure and to be followed until the first positive screening test outcome.

We extend the discrete proportional hazards model of Kalbfleish and Prentice (1980) to incorporate mismeasured outcomes. Our context is similar to that of Richardson and Hughes (2000). However, in addition to estimating cumulative survival, we estimate covariate effects as well. Conditional on knowing the outcome measure's sensitivity and specificity, the resulting estimates have minimal bias. In addition, missed visits are allowed. Unlike Snapinn's method, auxiliary variables are not required. Further, our method applies to a more common setting than that of Balasubramanian and Lagakos: here, the periods of risk and of testing overlap. All subjects are assumed to enter the study prior to failure.

In Section 2, the notation for all models are given. In Section 3, both the standard proportional hazards (PH) and our adjusted proportional hazards (APH) models are introduced. Simulation results in Section 4 demonstrate that the APH method substantially reduces bias due to mismeasured outcomes. Via simulation, we also show the consequences of inaccurately specifying the sensitivity or specificity of the outcome measure. In Section 5, we consider different assumptions regarding mismeasurement rates. In Section 6, we present an application to data from an observational cohort study in Kenya, and in Section 7 we state conclusions and cite likely areas for future work.

2. Notation

The following notation applies to all models. As outcomes are measured with error, we refer separately to the true and observed event statuses and event times. For $i = 1, \ldots, n$, let $t_i(t_i^{\mathrm{o}})$ be the true (observed) time the ith subject had the event and $d_i(d_i^{\mathrm{o}})$ be the indicator of true (observed) event status (1 = failure, 0 = censoring). The vector of covariates for subject i is denoted by $\boldsymbol{X}_i$ (length p). Let θ and ϕ denote, respectively, the sensitivity and specificity of the outcome measure.

Let the vector $\boldsymbol{I}_i = (I_{i1}, \ldots, I_{iT})$ indicate missed visits (1 = missed, 0 = not missed) for subject i at time points $1, \ldots, T$, where T is the last time of observation. In this notation, censored visits are considered "missed." Missed visits are summarized using the vector $\mathbf{C}_i = (C_{i1}, \ldots, C_{it_i^{\mathrm{o}}})$, where $C_{ij} = \sum_{k=1}^{j-1} I_{ik}$, for $(j = 2, \ldots, t_i^{\mathrm{o}})$ and $C_{i1} = 0$; that is, C_{ij} is the number of visits missed by subject i prior to time point j. In general, the observed data for the ith subject is referred to as $\mathbf{Y}_i^{\mathrm{o}} = \{\boldsymbol{X}_i, t_i^{\mathrm{o}}, d_i^{\mathrm{o}}, \boldsymbol{C}_i\}$. Let M_{ij} be the (unobserved) indicator that the result for subject i at visit j was misclassified (0 = not misclassified; 1 = misclassified).

3. Models

3.1 *Standard Discrete Proportional Hazards*

The distribution function for the discrete proportional hazards model with true event status (d_i) and true failure or censoring times (t_i) (Kalbfleisch and Prentice, 1980) is:

$$f(t_i, d_i; \boldsymbol{X}_i, \boldsymbol{\beta}, \boldsymbol{\lambda}_0) = \prod_{j=1}^{t_i-1} \left\{ (1-\lambda_{0j})^{e^{(\boldsymbol{X}_i'\boldsymbol{\beta})}} \right\} \times \left\{ 1-(1-\lambda_{0t_i})^{e^{(\boldsymbol{X}_i'\boldsymbol{\beta})}} \right\}^{d_i} \times \left\{ (1-\lambda_{0t_i})^{e^{(\boldsymbol{X}_i'\boldsymbol{\beta})}} \right\}^{(1-d_i)}. \quad (1)$$

In equation (1), the baseline hazard is $\boldsymbol{\lambda}_0 = (\lambda_{01}, \lambda_{02}, \ldots, \lambda_{0T})$. The hazard for the ith subject with covariate $\boldsymbol{X}_i$ at the jth time point is: $\lambda(j \mid \boldsymbol{X}_i) = 1-(1-\lambda_{0j})^{e^{(\boldsymbol{X}_i'\boldsymbol{\beta})}}$. Our objective is to estimate $\boldsymbol{\Omega} = \binom{\boldsymbol{\lambda}_0}{\boldsymbol{\beta}}$. When uncertain outcomes are present (e.g., due to use of an imperfect diagnostic test), direct maximization of the likelihood based on (1) is not possible, since the true values of d_i and t_i are not known.

3.2 *The Adjusted Proportional Hazards Model*

3.2.1 *Context* In the discrete time setting, subjects are tested at predetermined time points until the time of the first observed failure. Subjects with a positive failure indicator are excluded from further observation, even though the observed failure may be false.

3.2.2 *Assumptions* The usual assumptions of the proportional hazards model hold. In particular, we assume that, conditional on the covariates $\boldsymbol{X}_i$, the censoring and missing data mechanisms are independent of, and uninformative about, the failure mechanism. We assume the misclassification indicators (the M_{ij}) are mutually independent. In addition, we assume, initially, that the probability of misclassification for subjects who have failed (1-sensitivity) does not depend on time since failure, and that the probability of misclassification for subjects who have not failed (1-specificity) does not depend on time until failure or censoring. That is,

$$p\big(M_{ij} = 1 \mid t_i^{\mathrm{o}} \geq j, t_i \leq j, d_i = 1\big) \text{ is constant over } t_i,$$

and

$$p(M_{ij} = 1 \mid t_i^{\mathrm{o}} \geq j \quad \text{and} \quad (t_i > j,\, d_i = 1 \text{ or } t_i \geq j, d_i = 0)) \text{ is constant over } t_i.$$

Later, we will allow sensitivity (θ) and specificity (ϕ) to depend on true failure time in a specific way.

Only external (fixed or time-dependent) covariates are considered (Kalbfleisch and Prentice, 1980), meaning that $\boldsymbol{X}_i$ may change over time, so long as its value is not generated by the individual under study. Should $\boldsymbol{X}_i$ be internal, then $\boldsymbol{X}_i$ would contain the value of t_i^{o}. As knowledge of t_i^{o} affects the probability that a subject has failed at previous time points, and the density (1) does not make use of t_i^{o}, the expression (1) would not be correct. For simplicity we use the notation $\boldsymbol{X}_i$, rather than $\boldsymbol{X}_{ij}$ $(j = 1, \ldots, T)$.

3.2.3 *Derivation of the marginal density* To avoid placing boundaries on the estimates for the hazard vector $\boldsymbol{\lambda}_0$, we reparameterize the baseline hazard. Let $\gamma_{0j} = \log(\frac{\lambda_{0j}}{1-\lambda_{0j}})$, $j = 1, \ldots, T$.

$$f(t_i, d_i; \boldsymbol{X}_i, \boldsymbol{\beta}, \boldsymbol{\gamma}_0) = \prod_{j=1}^{t_i-1} \left\{ (1+e^{\gamma_{0j}})^{-e^{(\boldsymbol{X}_i'\boldsymbol{\beta})}} \right\} \times \left\{ 1-(1+e^{\gamma_{0t_i}})^{-e^{(\boldsymbol{X}_i'\boldsymbol{\beta})}} \right\}^{d_i} \times \left\{ (1+e^{\gamma_{0t_i}})^{-e^{(\boldsymbol{X}_i'\boldsymbol{\beta})}} \right\}^{(1-d_i)}. \quad (2)$$

To compute the joint density $f(t_i,\ d_i,\ t_i^{\mathrm{o}},\ d_i^{\mathrm{o}}) = f(t_i,\ d_i) \times (t_i^{\mathrm{o}},\ d_i^{\mathrm{o}} \mid t_i,\ d_i)$, we first derive $f(t_i^{\mathrm{o}},\ d_i^{\mathrm{o}} \mid t_i,\ d_i)$. We begin with an example. Imagine subject i missed one visit prior to failing at time point t_i, and missed a second visit that fell after t_i, but prior to testing positive for the first time at t_i^{o}. We assume follow-up ends when a failure is observed. The events would be ordered as follows:

$$\underbrace{1, 2, \ldots, \overbrace{t_i-4}^{\text{missed}}, \ldots, t_i-1}_{t_i-2 \text{ true negatives}}, \underbrace{\overbrace{t_i}^{\text{fail}}, \ldots, \overbrace{t_i+3}^{\text{missed}}, \ldots, t_i^{\mathrm{o}}-1}_{t_i^{\mathrm{o}}-t_i-1 \text{ false negatives}} \overbrace{t_i^{\mathrm{o}}}^{\text{observed failure}}.$$

The indicator for missed visits includes $I_{i(t_i-4)} = I_{i(t_i+3)} = 1$, so $C_{it_i} = \sum_{j=1}^{t_i-1} I_{ij} = 1$ and $C_{it_i^{\mathrm{o}}} = 2$. The probability of these outcomes given failure at t_i is $\phi^{t_i-2}(1-\theta)^{t_i^{\mathrm{o}}-t_i-1}\theta$. In general,

$$f\left(t_i^{\mathrm{o}}, d_i^{\mathrm{o}} \,|\, t_i = t_i^{\mathrm{o}}, d_i = 0, \theta, \phi\right) = \phi^{t_i^{\mathrm{o}} - C_{it_i^{\mathrm{o}}} - 1} \phi^{1-d_i^{\mathrm{o}}} (1-\phi)^{d_i^{\mathrm{o}}} = \Gamma_i,$$

and

$$f\left(t_i^{\mathrm{o}}, d_i^{\mathrm{o}} \,|\, t_i \le t_i^{\mathrm{o}}, d_i = 1, \theta, \phi\right) = \phi^{t_i - C_{it_i} - 1}(1-\theta)^{t_i^{\mathrm{o}} - t_i - C_{it_i^{\mathrm{o}}} + C_{it_i}} \times (1-\theta)^{1-d_i^{\mathrm{o}}} \theta^{d_i^{\mathrm{o}}} = \Delta_{it_i}. \quad (3)$$

Note that for both Γ_i and Δ_{it_i}, the contribution of the last test at t_i^{o} depends on whether or not the subject was censored ($d_i^{\mathrm{o}} = 0$) or observed to fail ($d_i^{\mathrm{o}} = 1$).

The marginal density of the observed data, $f(t_i^{\mathrm{o}},\ d_i^{\mathrm{o}})$, is computed by summing $f(t_i,\ d_i,\ t_i^{\mathrm{o}},\ d_i^{\mathrm{o}})$ over the possible values of t_i and d_i. Note that we assume follow-up ends when an event is observed, so that (true) events occurring after t_i^{o} are censored at t_i^{o}. Thus, we may have $t_i = t_i^{\mathrm{o}}$, $d_i = 0$, or $t_i \le t_i^{\mathrm{o}}$, $d_i = 1$, but not $t_i > t_i^{\mathrm{o}}$. Note that when $\theta = \phi = 1$, the density (4) reduces to a form equivalent to expression (1). However, in the general case, the marginal density is:

$$\begin{aligned} &f\left(t_i^{\mathrm{o}}, d_i^{\mathrm{o}}; \boldsymbol{X}_i, \boldsymbol{\beta}, \boldsymbol{\gamma}_0, \theta, \phi\right) \\ &\quad = f\left(t_i = t_i^{\mathrm{o}}, d_i = 0, t_i^{\mathrm{o}}, d_i^{\mathrm{o}}; \boldsymbol{X}_i, \boldsymbol{\beta}, \boldsymbol{\gamma}_0, \theta, \phi\right) \\ &\qquad + \sum_{k=1}^{t_i^{\mathrm{o}}} f\left(t_i = k, d_i = 1, t_i^{\mathrm{o}}, d_i^{\mathrm{o}}; \boldsymbol{X}_i, \boldsymbol{\beta}, \boldsymbol{\gamma}_0, \theta, \phi\right) \\ &\quad = \prod_{j=1}^{t_i^{\mathrm{o}}} (1+e^{\gamma_{0j}})^{-e^{(\boldsymbol{X}_i'\boldsymbol{\beta})}} \Gamma_i + \sum_{k=1}^{t_i^{\mathrm{o}}} \left[\prod_{j=1}^{k-1} \left\{ (1+e^{\gamma_{0j}})^{-e^{(\boldsymbol{X}_i'\boldsymbol{\beta})}} \right\} \times \left\{ 1-(1+e^{\gamma_{0k}})^{-e^{(\boldsymbol{X}_i'\boldsymbol{\beta})}} \right\} \Delta_{ik} \right]. \end{aligned} \quad (4)$$

3.2.4 *Inference* Parameter estimates are obtained by numerically maximizing the likelihood based on (4) over $\boldsymbol{\Omega}^* = \begin{pmatrix} \boldsymbol{\gamma}_0 \\ \boldsymbol{\beta} \end{pmatrix}$. Estimates are transformed from $\boldsymbol{\Omega}^*$ to $\boldsymbol{\Omega} = \begin{pmatrix} \boldsymbol{\gamma}_0 \\ \boldsymbol{\beta} \end{pmatrix}$. The variance of the maximum likelihood estimator (MLE) is approximated by $I(\hat{\boldsymbol{\Omega}})^{-1}$, where $I(\hat{\boldsymbol{\Omega}})$ is the observed information matrix (see Appendix).

In discrete time PH methods, including APH, the expression $e^{\boldsymbol{\beta}}$ is not equal to the hazard ratio (HR). The HR varies over time with $\boldsymbol{\lambda}_0 : \mathrm{HR}_j = 1-(1-\lambda_{0j})^{e^{\boldsymbol{\beta}}}/\lambda_{0j}$. However, by first-order Taylor series expansion, $\mathrm{HR}_j \approx e^{\boldsymbol{\beta}}$. This approximation is appropriate for small $\boldsymbol{\lambda}_0$, as in time-to-event studies, and hence we base inference on $e^{\boldsymbol{\beta}}$.

4. Model Evaluation

4.1 *Comparing APH to PH*

To assess the relative performance of the APH model compared to the discrete PH model, sets of simulations were performed with varying values of $\boldsymbol{\lambda}_0$, β, θ, and ϕ. Data sets were created with a single binary covariate ($X = 0$ or 1) and with 5 timepoints of observation. Approximately 5% of subjects were censored uniformly between timepoints 1 and 4. Other subjects were simulated to be tested at all 5 visits. Representative results are presented here. For these simulations, $\beta = 1.3$. Sensitivity (θ) ranged from 0.4 to 1 and specificity (ϕ) from 0.9 to 1. When $\theta \in \{0.8,\ 1\}$, the baseline hazard $\boldsymbol{\lambda}_0$ was 0.05 at all time points and 800 observations per simulation were used. When $\theta \in \{0.4,\ 0.6\}$, $\boldsymbol{\lambda}_0 = 0.1$ and 1200 observations per simulation were used. Larger numbers of events were needed at lower sensitivity to have sufficient numbers of observed failures in the simulated data. 1000 repetitions were performed for each set of conditions. The PH estimates were computed by numerically maximizing the likelihood based on (4) at $\theta = \phi = 1$, while the APH estimates used the true values of θ and ϕ.

Table 1 shows bias for $\hat{\beta}$ under both models. Let $S = 1000$ be the number of simulations performed and k the simulation index. Bias was computed as $\overline{\beta} - \beta$, where $\overline{\beta} = (1/S)\sum_{k=1}^{S} \hat{\beta}_k$, and percent bias as $(\overline{\beta} - \beta)/\beta * 100$. Standard error was computed as $\hat{\sigma}(\beta) = ((1/S)\sum_{k=1}^{S}[(\hat{\beta}_k - \overline{\beta})^2])^{(1/2)}$, and mean square error (MSE) using $(1/S)\sum_{k=1}^{S}[(\hat{\beta}_k - \beta)^2]$. Estimation under the PH model resulted in substantial bias, especially at low specificities. The APH method performed well relative to PH over all sets of conditions, but demonstrated some bias with increasing uncertainty. While the APH method is somewhat inefficient relative to PH, total MSE is reduced when using APH, and good coverage for β is maintained. Similar results (not shown) were found in estimating cumulative survival $F_0(T) = \prod_{j=1}^{T}(1 - \lambda_{0j})$. A profile analysis of the likelihood near the maximum likelihood estimators shows that confidence intervals for $\boldsymbol{\lambda}_0$ computed by transforming endpoints of $\mathrm{CI}(\boldsymbol{\gamma}_0)$ differ only slightly from those obtained using $\sigma_{\boldsymbol{\lambda}_0}$.

Table 1
Performance in estimating β (for a single, binary covariate) under the PH and APH models when data are simulated under varying mismeasurement rates, and sensitivity and specificity are correctly specified ($\beta = 1.3$)

		PH/APH			
Sensitivity	Specificity	% Bias	Std. Err.	$\sqrt{\text{MSE}}$	% Coverage
1.0	.98	−17.6%/ .4%	.12/.16	.26/.16	47%/93%
1.0	.95	−34.6%/ .8%	.10/.19	.46/.19	1%/94%
1.0	.90	−51.4%/ 1.6%	.09/.24	.67/.24	0%/96%
.8	1.00	−.1%/ .3%	.13/.13	.13/.13	94%/94%
.8	.98	−19.0%/ .5%	.12/.16	.27/.16	41%/94%
.8	.95	−36.7%/ .8%	.10/.19	.49/.19	1%/96%
.8	.90	−54.1%/ 1.3%	.09/.25	.71/.26	0%/96%
.6	1.00	−4.1%/−.1%	.09/.09	.10/.09	88%/96%
.6	.98	−15.0%/ .6%	.08/.10	.21/.10	32%/95%
.6	.95	−28.3%/ .7%	.08/.11	.38/.12	0%/95%
.6	.90	−44.1%/ 1.2%	.07/.15	.58/.15	0%/94%
.4	1.00	−10.2%/−.1%	.09/.11	.16/.10	67%/94%
.4	.98	−23.3%/ .6%	.09/.12	.32/.12	7%/95%
.4	.95	−38.7%/ .8%	.08/.15	.51/.15	0%/96%
.4	.90	56.1%/ 1.8%	.07/.20	.73/.20	0%/95%

The Monte Carlo precision from 1000 simulations shows APH coverage of $\hat{\beta}$ and $F_0(T)$ (results not shown) is not significantly different from 95%, except when specificity is 90%. APH bias was significantly different from zero under more than half of the conditions, though it is much lower than that found when applying PH. In other simulations containing fewer observations (not included in Table 1), small bias in $\hat{\beta}$ was found to be associated with having no observed events at some time points.

A second group of 1000 simulations was performed under these same conditions with $\beta = 0$, to evaluate type I error. Results are summarized over all sets of conditions. In estimating β, both methods had small average bias (APH, −0.0007; PH, −0.0003), and an appropriate type I error (APH, 4.7%; PH, 5.1%). $\text{MSE}^{(1/2)}$ was higher for APH (0.19) than for PH (0.11), though both were small. However, in estimating cumulative survival $F_0(T)$, APH performed better on average in terms of percent bias (APH, 0.04%; PH, −13.4%), type I error (APH, 5.1%; PH, 77.4%) and $\text{MSE}^{(1/2)}$ (APH, 0.03; PH, 0.13).

4.2 *Misspecification of Sensitivity or Specificity*

As shown in 4.1, the APH model performed very well when θ and ϕ were known. When θ and ϕ were assumed to be 1 (PH model), substantial bias and loss of coverage was introduced. To assess robustness of the APH to milder misspecification of θ or ϕ, we performed simulations. One-thousand sets of 800 observations were created with a specificity of 90% and sensitivity of 100%. A binary indicator served as the only covariate, with $\beta = 1.3$. The baseline hazard was $\lambda_{0j} = 0.1$, $j = 1, 2, \ldots, 5$. Each data set was evaluated using the APH method at several values of specificity while correctly assuming perfect sensitivity.

The dashed line of Figure 1 shows substantial bias was induced, even at specificity values that were quite close to the true value. Overstating the specificity caused conservative bias in $\hat{\beta}$, while understating ϕ led to an inflated $\hat{\beta}$. Estimates of baseline cumulative survival ($F_0(t)$) were also biased (data not shown).

Next, data were simulated as described above with 60% sensitivity and 100% specificity. In Figure 1 (solid line), we plotted percent bias in $\hat{\beta}$ over a range of sensitivity values. Misspecifying sensitivity did not cause as much bias as misspecifying specificity, in either $\hat{\beta}$ or $\hat{F}_0(t)$ for the parameter values we used. This is due to the fact that poor sensitivity affects fewer subjects when failure is a rare event ($\lambda_{0j} = .1$).

5. Model Adaptations

Outcome mismeasurement rates (θ and ϕ) may not be constant across time and subject. In the following, we present models in which mismeasurement depends on observed factors. Potential applications are also suggested.

5.1 *Sensitivity and Specificity Depend on Observed Covariates*

Sensitivity and/or specificity may depend on measured covariates. For example, diagnostic tests used at different sites may vary in accuracy. Let $\boldsymbol{Z}_i$ be the vector of measures on which test accuracy depends. Define $g_1(\boldsymbol{Z}_i) = \theta_{\mathbf{Z}_i}$ and $g_2(\boldsymbol{Z}_i) = \phi_{\mathbf{Z}_i}$. When $\boldsymbol{Z}_i$ $(i = 1, \ldots, n)$, $g_1(\cdot)$ and $g_2(\cdot)$ are known, we replace θ and ϕ in (3) by $\theta_{\mathbf{Z}_i}$ and $\phi_{\mathbf{Z}_i}$. The only portion of the estimation procedure that changes is computation of Γ_i and Δ_{ik}. Note that $\boldsymbol{Z}_i$ is used primarily to determine $\theta_{\mathbf{Z}_i}$ and $\phi_{\mathbf{Z}_i}$ and may not affect risk of failure. However, if any subset of $\boldsymbol{Z}_i$ additionally influences the failure mechanism, it may be included in $\boldsymbol{X}_i$.

5.2 *Sensitivity and Specificity Depend on Time*

Sensitivity or specificity may vary over time. For example, in developing countries, a batch of test kits may decrease in sensitivity due to the unavailability of refrigerated storage. In a long-term study, better tests (having less mismeasurement) may be developed and implemented after subject enrollment has begun. We allow θ and ϕ to be time-dependent by substituting θ_j and $\phi_j (j = 1, \ldots, T)$ for θ and ϕ in the

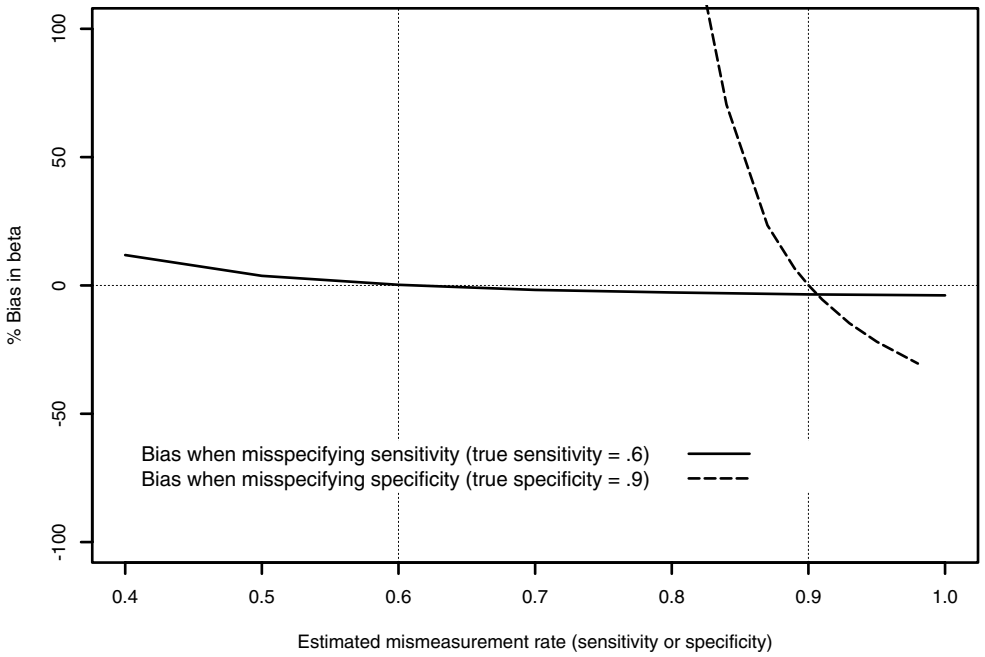

Figure 1. Percent bias in beta using the APH method under two conditions: 1) when simulated sensitivity (θ) is 0.6, but $\hat{\theta}$ varies, and 2) when simulated specificity (ϕ) is 0.9 but $\hat{\phi}$ varies.

derivation of equation (3). The conditional probabilities of observed events are:

$$\Gamma_i = \left\{ \prod_{j=1}^{t_i^o-1} \phi_j^{(I_{ij}=0)} \right\} \phi_{t_i^o}^{(1-d_i^o)} \left(1-\phi_{t_i^o}\right)^{d_i^o},$$

$$\Delta_{it_i} = \left\{ \prod_{j=1}^{t_i-1} \phi_j^{(I_{ij}=0)} \right\} \left\{ \prod_{j=t_i}^{t_i^o-1} (1-\theta_j)^{(I_{ij}=0)} \right\} \left(1-\theta_{t_i^o}\right)^{(1-d_i^o)} \theta_{t_i^o}^{d_i^o}. \tag{5}$$

If test mismeasurement rates depend on both time and other covariates, it is straightforward to build a hybrid from 5.1 and 5.2. One can simply replace θ_j and ϕ_j in (5) with $\theta_{\mathbf{z}_{ij}}$ and $\phi_{\mathbf{z}_{ij}}$.

5.3 *Sensitivity and Specificity Depend on Time since True Failure*

Screening tests for which the accuracy changes with time since true failure may be used. For example, in using ELISA to diagnose HIV infection, the probability of a false negative outcome may decrease with time since infection (increase in sensitivity), as antibodies will increase in number.

For each subject, let θ_k refer to the sensitivity of the screening test performed k timepoints after a subject fails. While θ_k applies to different visit numbers for each subject, we assume its value does not depend on subject. Computation of Γ_i proceeds as in (5), however:

$$\Delta_{it_i} = \left\{ \prod_{j=1}^{t_i-1} \phi_j^{(I_{ij}=0)} \right\} \left\{ \prod_{j=t_i}^{t_i^o-1} \left(1-\theta_{j-t_i+1}\right)^{(I_{ij}=0)} \right\} \times \left(1-\theta_{(t_i^o-t_i+1)}\right)^{(1-d_i^o)} \theta_{(t_i^o-t_i+1)}^{d_i^o}. \tag{6}$$

6. Example

In an observational, prospective, cohort study of commercial sex workers in Mombasa Kenya, subjects were tested regularly for HIV-1 and other sexually transmitted infections (STIs). Of 953 subjects enrolled, 783 returned at least once, resulting in a total of 880 person-years of follow-up. Of interest were the relationships between demographic factors, hormonal contraceptive use, incident STIs, and HIV-1 seroconversion. Complete results, study procedures and participant characteristics are described in detail elsewhere (Martin et al., 1998).

As part of the study protocol, endocervical swab specimens were taken monthly, and were cultured to test for *Neisseria gonorrhoeae*. While specificity of culture is very high, the sensitivity of testing for *Neisseria gonorrhoeae* by culture relative to a gold standard of ligase chain reaction (LCR) has been estimated to be between 50% (Buimer et al., 1996) and 58.3% (Carroll et al., 1998). We applied the APH method to these data setting the sensitivity and specificity, respectively, at 55% and 100%.

We compared the PH and APH methods in estimating risk factors for (first case of) gonorrhea. The cumulative proportion infected was higher among subjects with 10 or fewer years of education. The hazard ratio due to <=10 years of education was lower using the standard PH method ($\hat{\beta} = 0.59$; HR $\approx$ 1.8; p = 0.064) than with the APH ($\hat{\beta} = 0.77$; HR $\approx$ 2.2; p = 0.005). Failure to account for a 45% error rate in test accuracy resulted in an approximate 16% decrease in the magnitude of the hazard ratio and marginal statistical significance. Additional factors that might confound the education/gonorrhea relationship (i.e., number of partners) could be added to the model as well. Cumulative infection rates for the PH method were substantially lower than those of the APH method during the first few months, as poor test sensitivity delayed observation of positive outcomes (Figure 2).

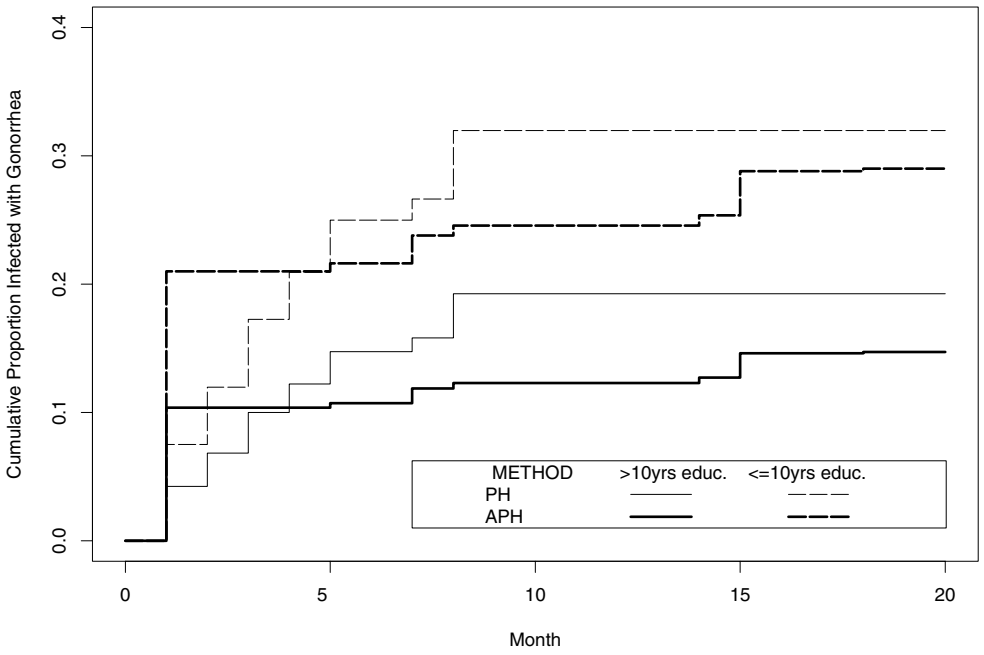

Figure 2. Estimated survival curves (showing cumulative proportion infected with gonorrhea) by education level, for PH and APH methods.

7. Conclusions

When false test results occur, cumulative survival and hazard ratio estimates are biased. In all cases simulated, the bias is toward the null hypothesis. False positive tests induce greater bias in estimation of $F_0(t)$ and $\boldsymbol{\beta}$ than false negative results when the failure rate is low. Conversely, if the failure rate is high, we expect that false negative tests would cause more bias in estimation. Since survival methods often assess rare events, this implies that misspecification of specificity will usually lead to greater bias. For example, in simulation studies, it was shown that the PH method, which assumes perfect sensitivity (θ) and specificity (ϕ), provided only 47% coverage of β under a relatively low false positive rate ($\theta = 1$ and $\phi = .98$). However, with accurate estimates of the test's sensitivity and specificity, and providing that a reasonable amount of data is available, the APH method can estimate both cumulative survival and hazard ratios accurately.

The performance of the APH method is reduced when the test's characteristics are incorrectly specified. Thus, it is important to have good estimates of sensitivity and specificity from external sources. (In general, sensitivity and specificity are not identifiable from the data described.) If these are not available, then use of a validation subsample, in which a portion of the subjects would receive both a screening and a gold standard test at each visit, would allow estimation of sensitivity and specificity for each model described here. Other useful enhancements to the APH method might include techniques appropriate for continuous time with periodic testing (leading to interval censoring) and/or the incorporation of confirmatory testing (i.e., confirming positive tests with a second, independent test).

In research contexts in which good estimates of sensitivity and specificity of time-to-event outcome measures are available, implementation of the APH method can improve accuracy of estimation and inference. In contexts in which false outcome measures are suspected, but the values of sensitivity and specificity are unknown, a sensitivity analysis can be performed using a range of estimated θ and ϕ.

Acknowledgements

This work was funded initially by the University of Washington STD/AIDS Predoctoral and Postdoctoral Training Program (NIAID AI0714P) and later by the Fred Hutchinson Cancer Research Center. Most of the work was done while the first author was pursuing a Ph.D. at the University of Washington. We would like to thank Margaret Pepe and Thomas Lumley for their valuable insights and critiques during the development of these methods. We also thank the two referees whose thoughtful evaluation of this article improved its clarity.

Résumé

La mesure erronée du critère de jugement peut conduire à des estimations biaisées dans plusieurs contextes. Magder et Hughes (1997) ont montré que l'absence de prise en compte de l'erreur sur le critère de jugement dans une analyse logistique peut biaiser de façon conservatrice l'estimation des effets des covariables, même lorsque le taux d'erreur est le même pour chaque valeur de la covariable. D'autres auteurs se sont intéressés à la nécessité de prendre en compte l'erreur dans le recueil du critère de jugement dans les études de survie dans des cas particuliers (Snapinn, 1998; Gelfand et Wang, 2000; Balasubramanian et Lagakos, 2001). Nous proposons une méthode ajustée à risques proportionnels, générale, plus largement utilisable pour l'estimation de la survie cumulée et des risques relatifs dans un contexte de temps discret quand le critère de jugement est recueilli avec erreur. Nous montrons que le fait de recueillir l'événement avec erreur dans un modèle à risques proportionnels standard peut biaiser les estimations

des risques relatifs de façon conservatrice et que l'inférence, dans la plupart des situations, est plus sévèrement influencée par une mauvaise spécificité que par une mauvaise sensibilité. Cependant, dans des simulations portant sur des situations très variées, la méthode proposée avec des taux d'erreurs bien spécifiés a de très bonnes performances.

References

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

Balasubramanian, R. and Lagakos, S. W. (2003). Estimation of a failure time distribution based on imperfect diagnostic tests. *Biometrika* **90,** 71–182.

Buimer, M., van Doornum, G. J., Ching, S., Peerbooms, P. G., Plier, P. K., Ram, D., and Lee, H. H. (1996). Detection of *Chlamydia trachomatis* and *Neisseria gonorrhoeae* by ligase chain reaction–based assays with clinical specimens from various sites: Implications for diagnostic testing and screening. *Journal of Clinical Microbiology* **34,** 2395–2400.

Carroll, K. C., Aldeen, W. E., Morrison, M., Anderson, R., Lee, D., and Mottice, S. (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,** 1630–1633.

Cox, D. R. (1972). Regression models and life-tables. *JRSS, Series B* **34,** 187–220.

Fleming, T. R. and Harrington, D. P. (1991). *Counting Processes and Survival Analysis*. New York: Wiley.

Fuller, W. A. (1987). *Measurement Error Models*. New York: Wiley.

Gelfand, A. E. and Wang, F. (2000). Modelling the cumulative risk for a false-positive under repeated screening events. *Statistics in Medicine* **19,** 1865–1879.

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

Little, R. J. A. and Rubin, D. B. (1987). *Statistical Analysis with Missing Data*. New York: Wiley.

Magder, L. and Hughes, J. P. (1997). Logistic regression when the outcome is measured with uncertainty. *American Journal of Epidemiology* **146,** 195–203.

Martin, H. L., Jr., Nyange, P. M., Richardson, B. A., Lavreys, L., Mandaliya, K., Jackson, D. J., Ndiriya-Achola, J. O., and Kreiss, J. (1998). Hormonal contraception, sexually transmitted diseases, and risk of heterosexual transmissions of human immunodeficiency virus type 1. *Journal of Infectious Diseases* **178,** 1053–1059.

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

Snapinn, S. M. (1998). Survival analysis with uncertain endpoints. *Biometrics* **54,** 209–218.

Received June 2002. *Revised February* 2003.
Accepted February 2003.

Appendix

Information Matrix, APH Method

As we have parameterized the likelihood in terms of $\boldsymbol{\gamma}_0$, not $\boldsymbol{\lambda}_0$, we begin by taking derivatives with respect to each element of $\boldsymbol{\Omega}^* = \begin{pmatrix}\boldsymbol{\gamma}_0\\ \boldsymbol{\beta}\end{pmatrix}$. This simplifies programming, as some expressions are similar to those used in the maximization step.

$$\begin{aligned}
&\frac{\partial L_i\left(\boldsymbol{\Omega}^*; \mathbf{Y}_i^{\mathrm{o}}\right)}{\partial \beta_l} \\
&= \prod_{k=1}^{t_i^{\mathrm{o}}} (1+e^{\gamma_{0k}})^{-e^{(\boldsymbol{X}_i'\boldsymbol{\beta})}} x_{il} e^{(\boldsymbol{X}_i'\boldsymbol{\beta})} \sum_{k=1}^{t_i^{\mathrm{o}}} \log(1+e^{\gamma_{0k}}) \Gamma_i \\
&\quad + \sum_{j=1}^{t_i^{\mathrm{o}}} \left[\Delta_{ij} \prod_{k=1}^{j-1} (1+e^{\gamma_{0k}})^{-e^{(\boldsymbol{X}_i'\boldsymbol{\beta})}} x_{il} e^{(\boldsymbol{X}_i'\boldsymbol{\beta})} \right. \\
&\qquad \times \left\{ (1+e^{\gamma_{0j}})^{-e^{(\boldsymbol{X}_i'\boldsymbol{\beta})}} \sum_{k=1}^{j} \log(1+e^{\gamma_{0k}}) \right. \\
&\qquad \left. \left. - \sum_{k=1}^{j-1} \log(1+e^{\gamma_{0k}}) \right\} \right].
\end{aligned}$$

$$\begin{aligned}
&\frac{\partial L_i\left(\boldsymbol{\Omega}^*; \mathbf{Y}_i^{\mathrm{o}}\right)}{\partial \gamma_{0l}} \\
&= \mathrm{I}\left(l \le t_i^{\mathrm{o}}\right) \left(-e^{(\boldsymbol{X}_i'\boldsymbol{\beta})} \frac{e^{\gamma_{0l}}}{(1+e^{\gamma_{0l}})} \right) \\
&\quad \left[\prod_{k=1}^{t_i^{\mathrm{o}}} (1+e^{\gamma_{0k}})^{-e^{(\boldsymbol{X}_i'\boldsymbol{\beta})}} \Gamma_i - \prod_{k=1}^{l} (1+e^{\gamma_{0k}})^{-e^{(\boldsymbol{X}_i'\boldsymbol{\beta})}} \Delta_{il} \right. \\
&\qquad + \sum_{j=l+1}^{t_i^{\mathrm{o}}} \left[\Delta_{ij} \left\{ 1 - (1+e^{\gamma_{0j}})^{-e^{(\boldsymbol{X}_i'\boldsymbol{\beta})}} \right\} \right. \\
&\qquad \left. \left. \times \prod_{k=1}^{j-1} (1+e^{\gamma_{0k}})^{-e^{(\boldsymbol{X}_i'\boldsymbol{\beta})}} \right] \right].
\end{aligned}$$

Repeat to get 2nd derivatives, using $M = \max(l,\ m)$ and where x_{il} (x_{im}) is the observed value of $\boldsymbol{X}$ for the ith subject at the lth (mth) timepoint.

$$\begin{aligned}
&\frac{\partial^2 L_i\left(\boldsymbol{\Omega}^*; \mathbf{Y}_i^{\mathrm{o}}\right)}{\partial \gamma_{0l} \partial \gamma_{0m}} \\
&= \mathrm{I}(M \le t_i^{\mathrm{o}}) \left[\frac{e^{2(\boldsymbol{X}_i'\boldsymbol{\beta})} e^{\gamma_{0l}} e^{\gamma_{0m}} - \mathrm{I}(l=m) e^{(\boldsymbol{X}_i'\boldsymbol{\beta})} e^{\gamma_{0l}}}{(1+e^{\gamma_{0l}})(1+e^{\gamma_{0m}})} \right] \\
&\quad \left[\prod_{k=1}^{t_i^{\mathrm{o}}} (1+e^{\gamma_{0k}})^{-e^{(\boldsymbol{X}_i'\boldsymbol{\beta})}} \Gamma_i - \prod_{k=1}^{M} (1+e^{\gamma_{0k}})^{-e^{(\boldsymbol{X}_i'\boldsymbol{\beta})}} \Delta_{iM} \right. \\
&\qquad + \sum_{j=M+1}^{t_i^{\mathrm{o}}} \left[\Delta_{ij} \left\{ 1 - (1+e^{\gamma_{0j}})^{-e^{(\boldsymbol{X}_i'\boldsymbol{\beta})}} \right\} \right. \\
&\qquad \left. \left. \times \prod_{k=1}^{j-1} (1+e^{\gamma_{0k}})^{-e^{(\boldsymbol{X}_i'\boldsymbol{\beta})}} \right] \right].
\end{aligned}$$

$$\frac{\partial^2 L_i\left(\mathbf{\Omega}^*; \mathbf{Y}_i^{\mathrm{o}}\right)}{\partial\beta_l \partial\beta_m}$$

$$= x_{il}x_{im}\left[\Gamma_i \prod_{k=1}^{t_i^{\mathrm{o}}}(1+e^{\gamma_{0k}})^{-e^{(\boldsymbol{X}'_i\boldsymbol{\beta})}} \log\prod_{k=1}^{t_i^{\mathrm{o}}}(1+e^{\gamma_{0k}})^{-e^{(\boldsymbol{X}'_i\boldsymbol{\beta})}} \times\left\{1+\log\prod_{k=1}^{t_i^{\mathrm{o}}}(1+e^{\gamma_{0k}})^{-e^{(\boldsymbol{X}'_i\boldsymbol{\beta})}}\right\}\right.$$

$$+\sum_{j=1}^{t_i^{\mathrm{o}}}\left[\Delta_{ij}\prod_{k=1}^{j-1}(1+e^{\gamma_{0k}})^{-e^{(\boldsymbol{X}'_i\boldsymbol{\beta})}} \times\left[-(1+e^{\gamma_{0j}})^{-e^{(\boldsymbol{X}'_i\boldsymbol{\beta})}}\log\prod_{k=1}^{j}(1+e^{\gamma_{0k}})^{-e^{(\boldsymbol{X}'_i\boldsymbol{\beta})}} \times\left\{1+\log\prod_{k=1}^{j}(1+e^{\gamma_{0k}})^{-e^{(\boldsymbol{X}'_i\boldsymbol{\beta})}}\right\}\right.\right.$$

$$\left.\left.\left.+\log\prod_{k=1}^{j-1}(1+e^{\gamma_{0k}})^{-e^{(\boldsymbol{X}'_i\boldsymbol{\beta})}} \times\left\{1+\log\prod_{k=1}^{j-1}(1+e^{\gamma_{0k}})^{-e^{(\boldsymbol{X}'_i\boldsymbol{\beta})}}\right\}\right]\right]\right].$$

$$\frac{\partial^2 L_i\left(\mathbf{\Omega}^*; \mathbf{Y}_i^{\mathrm{o}}\right)}{\partial\gamma_{0l} \partial\beta_m}$$

$$= -x_{im}\mathrm{I}\left(l \le t_i^{\mathrm{o}}\right)\frac{e^{(\boldsymbol{X}_i'\boldsymbol{\beta})}e^{\gamma_{0l}}}{(1+e^{\gamma_{0l}})}$$

$$\times\left[\prod_{k=1}^{t_i^{\mathrm{o}}}(1+e^{\gamma_{0k}})^{-e^{(\boldsymbol{X}'_i\boldsymbol{\beta})}}\Gamma_i\left\{1+\log\prod_{k=1}^{t_i^{\mathrm{o}}}(1+e^{\gamma_{0k}})^{-e^{(\boldsymbol{X}'_i\boldsymbol{\beta})}}\right\}\right.$$

$$\left.-\Delta_{il}\prod_{k=1}^{l}(1+e^{\gamma_{0k}})^{-e^{(\boldsymbol{X}'_i\boldsymbol{\beta})}}\left\{1+\log\prod_{k=1}^{l}(1+e^{\gamma_{0k}})^{-e^{(\boldsymbol{X}'_i\boldsymbol{\beta})}}\right\}\right.$$

$$+\sum_{j=l+1}^{t_i^{\mathrm{o}}}\left[\Delta_{ij}(1+e^{\gamma_{0k}})^{-e^{(\boldsymbol{X}'_i\boldsymbol{\beta})}} \times\left[1-(1+e^{\gamma_{0J}})^{-e^{(\boldsymbol{X}'_i\boldsymbol{\beta})}}\left\{1+\log\prod_{k=1}^{j}(1+e^{\gamma_{0k}})^{-e^{(\boldsymbol{X}'_i\boldsymbol{\beta})}}\right\}\right.\right.$$

$$\left.\left.\left.+\log\prod_{k=1}^{j-1}(1+e^{\gamma_{0k}})^{-e^{(\boldsymbol{X}'_i\boldsymbol{\beta})}}\right]\right]\right].$$

Let Ω_i^* and Ω_m^* be arbitrary elements of the vector $\mathbf{\Omega}^* = \begin{pmatrix}\boldsymbol{\gamma}_0\\ \boldsymbol{\beta}\end{pmatrix}$. Let $\mathcal{L}(\mathbf{\Omega}^*; \mathbf{Y}^{\mathrm{o}}) = \sum_{i=1}^n f(t_i^{\mathrm{o}}, d_i^{\mathrm{o}}; \boldsymbol{\beta}, \boldsymbol{\gamma}_0, \theta, \phi)$, from equation (4). We compute the observed information matrix $-\partial^2\mathcal{L}(\mathbf{\Omega}^*; \mathbf{Y}^{\mathrm{o}})/\partial\Omega_l^*\partial\Omega_m^* = -\sum_{i=1}^n\left[\partial^2\log(L_i)(\mathbf{\Omega}^*; \mathbf{Y}^{\mathrm{o}})/\partial\Omega_l^*\partial\Omega_m^*\right]$ by applying the 1st and 2nd derivatives of $L_i(\mathbf{\Omega}^*; \mathbf{Y}_i^{\mathrm{o}})$ with respect to each element of $\mathbf{\Omega}^*$.

$$\frac{\partial^2\mathcal{L}(\mathbf{\Omega}^*; \mathbf{Y}^{\mathrm{o}})}{\partial\Omega_l^*\partial\Omega_m^*} = \sum_{i=1}^n\left[\frac{1}{L_i\left(\mathbf{\Omega}^*; \mathbf{Y}_i^{\mathrm{o}}\right)}\frac{\partial^2 L_i\left(\mathbf{\Omega}^*; \mathbf{Y}_i^{\mathrm{o}}\right)}{\partial\Omega_l^*\partial\Omega_m^*} - \frac{1}{L_i\left(\mathbf{\Omega}^*; \mathbf{Y}_i^{\mathrm{o}}\right)^2}\frac{\partial L_i\left(\mathbf{\Omega}^*; \mathbf{Y}_i^{\mathrm{o}}\right)}{\partial\Omega_l^*}\frac{\partial L_i\left(\mathbf{\Omega}^*; \mathbf{Y}_i^{\mathrm{o}}\right)}{\partial\Omega_m^*}\right].$$

The chain rule is then applied to compute the observed information matrix of $\mathbf{\Omega}$. To obtain the information with respect to the baseline hazard $\boldsymbol{\lambda}_0$, we apply:

$$\frac{\partial^2\mathcal{L}(\mathbf{\Omega}; \mathbf{Y}^{\mathrm{o}})}{\partial\lambda_{0l}\partial\lambda_{0m}} = \frac{\partial^2\mathcal{L}(\mathbf{\Omega}^*; \mathbf{Y}^{\mathrm{o}})}{\partial\gamma_{0l}\partial\gamma_{0m}}\left[\frac{1}{\lambda_{0l}(1-\lambda_{0l})}\right]\left[\frac{1}{\lambda_{0m}(1-\lambda_{0m})}\right] + I(l=m)\left[\frac{\partial\mathcal{L}(\mathbf{\Omega}^*; \mathbf{Y}^{\mathrm{o}})}{\partial\lambda_{0l}}\right]\left(\frac{2\lambda_{0l}-1}{\lambda_{0l}^2(1-\lambda_{0l})^2}\right),$$

and

$$\frac{\partial^2\mathcal{L}(\mathbf{\Omega}; \mathbf{Y}^{\mathrm{o}})}{\partial\lambda_{0l}\partial\beta_m} = \frac{\partial^2\mathcal{L}(\mathbf{\Omega}^*; \mathbf{Y}^{\mathrm{o}})}{\partial\gamma_{0l}\partial\beta_m}\left[\frac{1}{\lambda_{0l}(1-\lambda_{0l})}\right].$$

Finally, we compute the asymptotic covariance matrix of $\mathbf{\Omega}$ using $\left(-\left[\partial^2\mathcal{L}(\mathbf{\Omega}; \mathbf{Y}^{\mathrm{o}})/\partial\Omega_l\partial\Omega_m\right]\right)^{-1}$ evaluated at $\hat{\mathbf{\Omega}}$.