Lifetime Data Analysis, 11, 29–40, 2005

Comparison Between Two Partial Likelihood Approaches for the Competing Risks Model with Missing Cause of Failure

KAIFENG LU kaifeng_lu@merck.com
Merck Research Laboratories, Merck & Co., Inc., Rahway, NJ 07065, USA

ANASTASIOS A. TSIATIS tsiatis@stat.ncsu.edu
Department of Statistics, North Carolina State University, Raleigh, NC 27695, USA

Received March 25, 2002; Revised February 2, 2004; Accepted February 24, 2004

Abstract. In many clinical studies where time to failure is of primary interest, patients may fail or die from one of many causes where failure time can be right censored. In some circumstances, it might also be the case that patients are known to die but the cause of death information is not available for some patients. Under the assumption that cause of death is missing at random, we compare the Goetghebeur and Ryan (1995, *Biometrika*, 82, 821–833) partial likelihood approach with the Dewanji (1992, *Biometrika*, 79, 855–857)partial likelihood approach. We show that the estimator for the regression coefficients based on the Dewanji partial likelihood is not only consistent and asymptotically normal, but also semiparametric efficient. While the Goetghebeur and Ryan estimator is more robust than the Dewanji partial likelihood estimator against misspecification of proportional baseline hazards, the Dewanji partial likelihood estimator allows the probability of missing cause of failure to depend on covariate information without the need to model the missingness mechanism. Tests for proportional baseline hazards are also suggested and a robust variance estimator is derived.

Keywords: cause-specific hazard, martingale, missing at random, semiparametric efficiency

1. Introduction

In a typical survival data analysis, a group of individuals are observed from some entry time until the occurrence of some particular event such as death. Often the observation of time to occurrence of the event is right-censored for some individuals as a result of staggered entry, finite study duration, withdrawal from the study, or loss to follow-up. Sometimes, the event can be classified into one of several categories, typically causes of death or other failures. For example, in a clinical trial that compares different therapies for breast cancer, interest may focus on death from breast cancer even though patients may die from other causes. In such cases, the theory of competing risks can be applied to assess the effects of prognostic factors on the cause-specific hazard of interest, for example, perform a standard proportional hazards analysis treating failure types which are not of interest as censored observations (Prentice and Kalbfleisch, 1978; Cox and Oakes, 1984). In some circumstances, patients are known to die but the cause of death information is not available for some individuals, for example, whether death is attributable to the cause of interest or other causes may require documentation with information that is not collected or lost or cause may be difficult for investigators to determine for some patients (Anderson et al., 1996). In such cases, excluding the missing observations from the analysis or treating them as censored may yield biased estimates and erroneous inferences. Under the assumption that the probability of missing cause of death may depend on time but not on covariates and that the baseline cause-specific hazards are proportional, Goetghebeur and Ryan (1995) proposed an approach that utilizes two types of partial likelihood, L and L^*. They showed that the resulting estimator is robust against misspecification of the proportional baseline hazards assumption and retains high efficiency with respect to the estimator based on the Dewanji partial likelihood L^* only. However, when the missingness probability depends also on covariates, the Goetghebeur and Ryan method needs to model the missingness mechanism explicitly (e.g., assuming a logistic model) and estimate the associated parameters along with other model parameters. Therefore, their estimator might be biased under the misspecification of the missingness mechanism.

The partial likelihood L^* has been studied by many investigators (cf., Holt, 1978; Kalbfleisch and Prentice, 1980; Dewanji, 1992). We show that the estimator based on L^* is not only consistent and asymptotically normal but also semiparametric efficient and does not require modeling of the missingness mechanism even when missingness depends on covariate information.

Notation and assumptions similar to those used in Goetghebeur and Ryan (1995) but extended to cover more general cases are described in Section 2. In Section 3, we show that the estimator based on the Dewanji partial likelihood is semiparametric efficient. In Section 4, we demonstrate that the Goetghebeur and Ryan estimator is more robust than the Dewanji partial likelihood estimator against misspecification of proportional baseline hazards, and suggest approaches to model the relationship between two baseline hazards. In Section 5, robust variance estimator is derived for the two partial likelihood estimators under misspecification of the relationship between two baseline hazards. In Section 6, we show that the Dewanji partial likelihood estimator is robust against misspecification of the missingness mechanism provided that cause of failure is missing at random. The article is concluded by a brief discussion.

2. Notation and Assumptions

In this article, we consider a sample of n independent individuals, each of whom can fail or die from one of two possible causes which we refer to as cause two and cause one, respectively, or can be subject to a noninformative censoring mechanism. Typically, the complete data for individual i can be summarized as (T_i, Δ_i, A_i), where T_i is the time to failure or censoring; Δ_i is the failure-censoring indicator taking values 2, 1 and 0 as the ith individual died from cause two, died from cause one, or was censored, respectively. The prognostic covariates A_i are related to the cause-specific hazards and might be time dependent. Furthermore, let $\lambda_\delta(t|a), \delta = 2, 1, 0$ be the cause-specific hazards for failures from cause two, failures from cause one, or censored observations, respectively.

Suppose that the two cause-specific hazards for failures can be modeled by

$$\lambda_\delta(t|a) = \lambda(t) r_\delta(\theta, t, a), \quad \delta = 1, 2, \tag{1}$$

where θ is the unknown vector of regression coefficients. We assume that all information about time common to the two cause-specific hazards has been incorporated into $\lambda(t)$ and without loss of generality, we let $\lambda(t)$ be the unspecified baseline hazard for failures from cause two. No assumptions are made on the cause-specific hazard of censoring or the marginal distribution of covariates. When the competing risks are independent, the cause-specific hazards will be the same as the net-specific hazards, which is implicitly assumed in the sequel.

To make connections with the notation used in Goetghebeur and Ryan (1995), we identify

$$\lambda_2(t|Z, X) = \lambda(t)e^{\phi Z}, \qquad \lambda_1(t|Z, X) = \lambda(t)e^{\xi+\rho X}.$$

Therefore, $A = (Z, X)$, $\theta = (\phi, \rho, \xi)$, $r_2(\theta, t, a) = e^{\phi z}$, $r_1(\theta, t, a) = e^{\xi+\rho x}$.

We could have formulated the model with separate regression parameter vectors for the two failure causes as done in Goetghebeur and Ryan (1995). There may be circumstances, however, where some parameters are common to the two failure causes. For instance, we might be interested in looking at whether the effects of covariates are the same or different across failure types (e.g., the example of time in power for leaders of countries, Chapter 6, Allison, 1995). Therefore, by formulating the model with one parameter vector which contains all different parameters (Andersen et al., 1997, p. 478), we can test the hypothesis of equal covariate effects across different failure types, say, using a likelihood ratio test.

Note that we allow the link functions $\{r_\delta(\cdot), \delta = 2, 1\}$ to depend on both time and covariates through a finite set of parameters. This is a generalization of proportional baseline hazards. For example, if $\lambda_2(t|z, x) = \lambda(t)e^{\phi z}$, $\lambda_1(t|z, x) = \lambda(t)e^{\xi+\gamma t+\rho x}$, then the baseline hazard functions for the two failure types are $\lambda_2(t) = \lambda(t)$, $\lambda_1(t) = \lambda(t)e^{\xi+\gamma t}$, respectively. Therefore, the ratio of the two baseline hazards is equal to $\lambda_1(t)/\lambda_2(t) = e^{\xi+\gamma t}$. If $\gamma \neq 0$, then the ratio is loglinear over time. Another useful example is given by the piecewise constant baseline hazards ratio.

In some circumstances, cause of failure might be missing for some individuals, in which case, we use R_i as the complete-case indicator for individual i, taking values one or zero as cause of failure is known or missing. We assume that cause of failure is missing at random in the sense of Rubin (1976), so that the probability of having a complete case does not depend on the actual cause of failure that might be missing, i.e.

$$P(R_i = 1|T_i, \Delta_i, A_i, \Delta_i > 0) = \pi(T_i, A_i), \tag{2}$$

where π is an unknown function of time and covariates only, taking values in the unit interval. Note that the Goetghebeur and Ryan method is based on the more restrictive missing-at-random assumption which allows the function π to depend on time but not on covariates. If the function π depends on covariates, their method requires joint modeling of missingness mechanism and competing risks. With missing cause of failure, the observed data for the ith individual can be summarized as $\{R_i, T_i, I(\Delta_i = 0), R_i I(\Delta_i = 1), R_i I(\Delta_i = 2), A_i\}$.

For an uncensored individual, one of the following three types of events can occur at the time of failure, i.e., failure from cause one, failure from cause two, or failure with an unknown cause. Let $\{N_{i1}(t), N_{i2}(t), N_{iu}(t)\}$ denote the counting processes of failures, then by (1) and (2), the corresponding intensity processes are given by

$$\lambda^*_{i1}(t, A_i) = Y_i(t)\pi(t, A_i)r_1(\theta_0, t, A_i)\lambda(t),$$
$$\lambda^*_{i2}(t, A_i) = Y_i(t)\pi(t, A_i)r_2(\theta_0, t, A_i)\lambda(t),$$
$$\lambda^*_{iu}(t, A_i) = Y_i(t)\pi^{\mathrm{c}}(t, A_i)r_{\cdot}(\theta_0, t, A_i)\lambda(t),$$

respectively, where $Y_i(t) = I(T_i \geq t)$ is the at-risk indicator,

$$\pi^{\mathrm{c}}(t, A_i) = 1 - \pi(t, A_i),$$

$$r_{\cdot}(\theta, t, A_i) = r_1(\theta, t, A_i) + r_2(\theta, t, A_i),$$

and θ_0 denotes the true value of θ. Let $N_{i\cdot} = N_{i1} + N_{i2} + N_{iu}$ denote the counting process of overall failure, then its intensity process is given by

$$\lambda^*_{i\cdot}(t, A_i) = \lambda^*_{i1}(t, A_i) + \lambda^*_{i2}(t, A_i) + \lambda^*_{iu}(t, A_i) = Y_i(t)r_{\cdot}(\theta_0, t, A_i)\lambda(t).$$

This will be used to motivate the estimation of cumulative baseline hazard for cause two, $\Lambda(t) = \int_0^t \lambda(s)ds$.

Suppose that the missingness mechanism is modeled by $\pi(t, A_i) = \pi(\psi, t, A_i)$, then applying the arguments of Cox (1975), the partial likelihood based on the conditional probabilities of a specific event given that one event of that type occurs from the risk set at that time is given by

$$L = \prod_{t\geq 0}\prod_{i=1}^{n}\left[\prod_{\delta=1}^{2}\left\{\frac{\pi(\psi, t, A_i)r_\delta(\theta, t, A_i)}{\sum_{j=1}^{n} Y_j(t)\pi(\psi, t, A_j)r_\delta(\theta, t, A_j)}\right\}^{dN_{i\delta}(t)}\right]$$
$$\times\left\{\frac{\pi^{\mathrm{c}}(\psi, t, A_i)r_{\cdot}(\theta, t, A_i)}{\sum_{j=1}^{n} Y_j(t)\pi^{\mathrm{c}}(\psi, t, A_j)r_{\cdot}(\theta, t, A_j)}\right\}^{dN_{iu}(t)},$$

where $\prod_{t\geq 0}$ denotes product-integration (Gill and Johansen, 1990). The more informative partial likelihood based on the conditional probabilities of an event of specified type, given that one event occurs, but without conditioning on the type of event, is given by

$$L^* = \prod_{t\geq 0}\prod_{i=1}^{n}\left[\prod_{\delta=1}^{2}\left\{\frac{\pi(\psi, t, A_i)r_\delta(\theta, t, A_i)}{\sum_{j=1}^{n} Y_j(t)r_{\cdot}(\theta, t, A_j)}\right\}^{dN_{i\delta}(t)}\right]\left\{\frac{\pi^{\mathrm{c}}(\psi, t, A_i)r_{\cdot}(\theta, t, A_i)}{\sum_{j=1}^{n} Y_j(t)r_{\cdot}(\theta, t, A_j)}\right\}^{dN_{iu}(t)}.$$

Note that the second partial likelihood can be factorized into two parts, one related to the missingness mechanism, and the other corresponding to the competing risks model, i.e., $L^*(\psi, \theta) = L^*(\psi)L^*(\theta)$, where

$$L^*(\psi) = \prod_{t \geq 0} \prod_{i=1}^{n} \pi(\psi, t, A_i)^{\{dN_{i2}(t)+dN_{i1}(t)\}} \pi^c(\psi, t, A_i)^{dN_{iu}(t)},$$

and

$$L^*(\theta) = \prod_{t \geq 0} \prod_{i=1}^{n} \frac{\left\{\prod_{\delta=1}^{2} r_\delta(\theta, t, A_i)^{dN_{i\delta}(t)}\right\} r.(\theta, t, A_i)^{dN_{iu}(t)}}{\left\{\sum_{j=1}^{n} Y_j(t) r.(\theta, t, A_j)\right\}^{dN_{i\cdot}(t)}}.$$

Therefore, if the parameters modeling the missingness mechanism, ψ, and those for the competing risks model, θ, are separate, one can estimate θ based on $L^*(\theta)$ only.

On the contrary, the function $\pi(\cdot)$ can not be factorized out of the partial likelihood L unless it does not depend on covariates A_i, in which case, $L(\psi, \theta)$ reduces to $L(\theta)$, where

$$L(\theta) = \prod_{t \geq 0} \prod_{i=1}^{n} \left[\prod_{\delta=1}^{2} \left\{ \frac{r_\delta(\theta, t, A_i)}{\sum_{j=1}^{n} Y_j(t) r_\delta(\theta, t, A_j)} \right\}^{dN_{i\delta}(t)} \right] \left\{ \frac{r.(\theta, t, A_i)}{\sum_{j=1}^{n} Y_j(t) r.(\theta, t, A_j)} \right\}^{dN_{iu}(t)}.$$

3. Semiparametric Efficiency

Let $\{r'_\delta(\theta, t, a), r''_\delta(\theta, t, a)\}$ denote the first two partial derivatives of $r_\delta(\theta, t, a)$ with respect to θ, and with similar notation defined for the partial derivatives of $r.(\theta, t, a)$, we write

$$\boldsymbol{m}(\theta, t) = \frac{\sum_{j=1}^{n} Y_j(t) r'.(\theta, t, A_j)}{\sum_{j=1}^{n} Y_j(t) r.(\theta, t, A_j)},$$

then the score vector corresponding to the partial likelihood $L^*(\theta)$ is

$$\boldsymbol{U}(\theta) = \sum_{i=1}^{n} \left[\sum_{\delta=1}^{2} \int \frac{r'_\delta(\theta, t, A_i)}{r_\delta(\theta, t, A_i)} dN_{i\delta}(t) + \int \frac{r'.(\theta, t, A_i)}{r.(\theta, t, A_i)} dN_{iu}(t) - \int \boldsymbol{m}(\theta, t) dN_{i\cdot}(t) \right].$$

Note that $\boldsymbol{U}(\theta_0)$ is the realization of a martingale process at $t = \infty$, and by martingale theory (Fleming and Harrington, 1991), under certain regularity conditions, the resulting estimator of θ from solving the estimating equations, $\boldsymbol{U}(\theta) = 0$, denoted by $\hat{\theta}_n$, is consistent and asymptotically normal (c.f., Andersen and Gill, 1982). It is also straightforward to show that, when evaluated at the truth, the observed information matrix, $\mathcal{I}(\theta) = -\partial \boldsymbol{U}(\theta)/\partial\theta$, has the same expectation as the predictable covariation process of the score vector, i.e., $E\{\mathcal{I}(\theta_0)\} = E\{\langle \boldsymbol{U}(\theta_0), \boldsymbol{U}(\theta_0)\rangle\}$. Consequently, the variance of $\hat{\theta}_n$ can be estimated by $\mathcal{I}^{-1}(\hat{\theta}_n)$.

Using semiparametric theory (e.g., Newey, 1990; Bickel et al., 1993; Robins et al., 1994), we can further show that the influence function of the Dewanji partial likelihood estimator, $\hat{\theta}_n$, is the most efficient influence function among all regular and asymptotically linear (RAL) estimators for θ, so that $\hat{\theta}_n$ is semiparametric efficient within the class of RAL estimators. The proof is outlined in the Appendix A.

4. Proportional Baseline Hazards

Goetghebeur and Ryan (1995) shows that there exists situations where the score test based on their estimating equations is valid while the score test derived from the efficient partial likelihood is biased. Specifically, they considered the score test for $\phi = 0$ in the following model

$$\lambda_2(t|z) = \lambda(t)e^{\phi z}, \qquad \lambda_1(t|z) = \lambda(t)e^{\xi(t)},$$

and score test statistics were constructed under the misspecified model of proportional baseline hazards

$$\lambda_2(t|z) = \lambda(t)e^{\phi z}, \quad \lambda_1(t|z) = \lambda(t)e^{\xi^*}.$$

Assume that cause of failure is missing at random and the missingness probability depends only on t, then as shown in Appendix 1 of their paper, under the null hypothesis, the Goetghebeur and Ryan score statistic has expectation zero, i.e., $E(\partial\ell/\partial\phi) = 0$, while the score statistic derived from the Dewanji partial likelihood has non-zero expectation, i.e., $E(\partial\ell^*/\partial\phi) \neq 0$, where ℓ and ℓ^* are the log partial likelihoods corresponding to $L(\theta)$ and $L^*(\theta)$, respectively. In fact, the compensator of the Dewanji score statistic divided by sample size n converges in probability to

$$D = \int \left(\frac{e^{\xi^*} - e^{\xi(t)}}{1 + e^{\xi^*}} \right) \mu_Z(t) P(T \geq t) \pi(t) \lambda(t) dt,$$

where e^{ξ^*} is the probability limit of $e^{\tilde{\xi}_n^*} = (\sum_i \int dN_{i1})/(\sum_i \int dN_{i2})$, or

$$e^{\xi^*} = \frac{\{\int e^{\xi(t)} P(T \geq t)\pi(t)\lambda(t)dt\}}{\{\int P(T \geq t)\pi(t)\lambda(t)dt\}.}$$

Similarly, $\mu_Z(t)$ is the probability limit of $m_Z(t) = \sum_i Z_i Y_i(t) / \sum_i Y_i(t)$, or

$$\mu_Z(t) = E\{ZP(T \geq t|Z)\}/P(T \geq t).$$

Note that if the censoring distribution does not depend on Z, then by the assumption of noninformative censoring and independent competing risks, under the null hypothesis, $P(T \geq t|Z) = P(T_2 \geq t|Z)P(T_1 \geq t|Z)P(C \geq t|Z) = P(T_2 \geq t)P(T_1 \geq t)$ $P(C \geq t) = P(T \geq t)$ does not depend on Z, hence $\mu_Z(t) = E(Z)$ does not depend on t, so that $D = 0$, which implies that the efficient score statistic is asymptotically unbiased. A small scale simulation confirmed these results.

It is interesting to note that although the Goetghebeur and Ryan score test for $\phi = 0$ is unbiased in this case, if the true value ϕ_0 of ϕ deviates from zero, then the Goetghebeur and Ryan score statistic for $\phi = \phi_0 (\neq 0)$ and hence the Goetghebeur and Ryan estimator, $\tilde{\phi}_n$, will be biased. To see this, let $\theta = (\phi, \xi)$, and assume $\tilde{\theta}_n = (\tilde{\phi}_n, \tilde{\xi}_n)$

solve the Goetghebeur and Ryan estimating equation, $T(\theta) = (\partial\ell/\partial\phi, \partial\ell^*/\partial\xi) = 0$, then, under suitable regularity conditions, $\tilde{\theta}_n \xrightarrow{p} \tilde{\theta} = (\tilde{\phi}, \tilde{\xi})$, where $\tilde{\theta}$ satisfies $\lim_{n\to\infty} n^{-1}T(\theta) = 0$.

Consider the special case where the covariate $Z \sim Bernoulli(0.5)$. Given Z, time to failure from cause two, T_2, and time to failure from cause one, T_1, are conditionally independent, where T_2 follows an exponential distribution with hazard $\lambda_2(t|z) = \lambda e^{\phi z}$, and T_1 follows a Gompertz distribution with hazard $\lambda_1(t|z) = \lambda e^{\xi+\gamma t}$. For simplicity, assume there is no censoring. Let $T = \min(T_2, T_1)$ be the observed failure time, and $\Delta = 2$ or 1 indicate the cause of failure. In addition, assume cause of failure is missing completely at random, i.e., $P(R = 1|T, \Delta, Z) \equiv \pi$. For $\lambda = 1$, $\phi = 0.8$, $\xi = -1$, $\gamma = 1$, and $\pi = 0.5$, the estimating equations yield $\tilde{\phi} = 0.8051$, indicating bias of the Goetghebeur and Ryan estimator.

Similar arguments can be applied to the Dewanji partial likelihood approach. For the special case described above, the Dewanji partial likelihood estimator of ϕ converges in probability to $\phi^* = 0.8482$, indicating a larger bias compared to the Goetghebeur and Ryan estimator.

The bias of the Dewanji partial likelihood estimator results from exploitation of the information that type one failures provide about ϕ through the assumption relating the two baseline cause-specific hazards, which is the very part of the model being misspecified. The better we model the relationship between two baseline hazards, the more precise the Dewanji partial likelihood estimator will get. The multiple imputation method proposed by Lu and Tsiatis (2001) provides a simple way to estimate the two baseline hazards, where we first treat cause two as the cause of interest to get an estimate for $\lambda_2(t) = \lambda(t)$, then treat cause one as the cause of interest to get an estimate for $\lambda_1(t)$. Examination of the plot of $\hat{\lambda}_1(t)$ versus $\hat{\lambda}_2(t)$ may aid us in choosing a plausible parametric model for the ratio of two baseline hazards, for example, log-polynomial or piecewise constant. The score statistic similar to that in Grambsch and Therneau (1994) can then be used to test $\gamma = 0$ in the parametric model, $\xi(t) = \xi + \gamma g(t)$, for some given function of time $g(\cdot)$.

5. Robust Variance Estimator

When the model (1) is incorrect, one can apply the techniques used in the proofs of Lemma 3.1 and Theorem 4.2 of Anderson and Gill (1982) to show that, under sufficient regularity conditions, the Dewanji partial likelihood estimator $\hat{\theta}_n$ converges in probability to a vector of constants θ^*, where θ^* is the unique solution to the system of equations

$$\sum_{\delta=1}^{2} \int E\left[\left\{\frac{r'_\delta(\theta, t, A)}{r_\delta(\theta, t, A)} - \mu(\theta, t)\right\} Y(t)\pi(t, A)\lambda_\delta(t|A)\right] dt$$
$$+ \int E\left[\left\{\frac{r'_\cdot(\theta, t, A)}{r_\cdot(\theta, t, A)} - \mu(\theta, t)\right\} Y(t)\pi^c(t, A)\lambda_\cdot(t|A)\right] dt = 0,$$

where

$$\mu(\theta, t) = \frac{E\{Y(t)r'_{\cdot}(\theta, t, A)\}}{E\{Y(t)r_{\cdot}(\theta, t, A)\}},$$

and $\{\lambda_\delta(t|A), \delta = 1, 2\}$ are the true cause-specific hazards, $\lambda_{\cdot}(t|A) = \lambda_1(t|A) + \lambda_2(t|A)$. A specific example is given in Section 4.

Now, let

$$\hat{\Lambda}(\theta, t) = \int_0^t \frac{\sum dN_{j\cdot}(s)}{\sum Y_j(s) r_{\cdot}(\theta, s, A_j)}$$

be the Breslow estimate of the cumulative baseline hazard for failures from cause two (c.f., Section 4, Goetghebeur and Ryan, 1995), and define

$$W_i(\theta) = \sum_{\delta=1}^{2} \int \frac{r'_\delta(\theta, t, A_i)}{r_\delta(\theta, t, A_i)} dN_{i\delta}(t) + \int \frac{r'_{\cdot}(\theta, t, A_i)}{r_{\cdot}(\theta, t, A_i)} dN_{iu}(t) - \int \boldsymbol{m}(\theta, t) dN_{i\cdot}(t)$$
$$- \int \{r'_{\cdot}(\theta, t, A_i) - \boldsymbol{m}(\theta, t) r_{\cdot}(\theta, t, A_i)\} Y_i(t) d\hat{\Lambda}(\theta, t),$$

then closely following Lin and Wei (1989), one can show that the random vector $n^{1/2}(\hat{\theta}_n - \theta^*)$ is asymptotically normal with mean 0 and with a covariance matrix that can be consistently estimated by $\hat{V}(\hat{\theta}_n) = \hat{A}^{-1}(\hat{\theta}_n)\hat{B}(\hat{\theta}_n)\hat{A}^{-1}(\hat{\theta}_n)$, where $\hat{A}(\theta) = n^{-1}\mathcal{I}(\theta)$, $\hat{B}(\theta) = n^{-1}\sum W_i(\theta)^{\otimes 2}$.

For the Goetghebeur and Ryan estimator $\tilde{\theta}_n = (\tilde{\phi}_n, \tilde{\rho}_n, \tilde{\xi}_n)$, similar arguments can be used to show that, under suitable regularity conditions, $\tilde{\theta}_n$ converges in probability to a vector of constants $\tilde{\theta} = (\tilde{\phi}, \tilde{\rho}, \tilde{\xi})$, where $\tilde{\theta}$ is the unique solution to a system of estimating equations, with a specific example provided by Section 4. The robust variance estimator for $n^{1/2}(\tilde{\theta}_n - \tilde{\theta})$ is given by $\hat{\Sigma}(\tilde{\theta}_n) = \hat{\Gamma}^{-1}(\tilde{\theta}_n)\hat{\Delta}(\tilde{\theta}_n)\hat{\Gamma}^{-1}(\tilde{\theta}_n)$, where $\hat{\Gamma}(\theta) = n^{-1}\boldsymbol{T}'(\theta)$, $\hat{\Delta}(\theta) = n^{-1}\sum K_i(\theta)^{\otimes 2}$, and $K_i(\theta)$ is a three-component column vector with components corresponding to ϕ and ρ similar to those given by Equation (6) of Goetghebeur and Ryan (1995) but with the counting processes subtracted by their estimated compensators based on the partial likelihood L, and the component corresponding to ξ is the same as that of $W_i(\theta)$. Note we have assumed that the probability of missing cause of failure does not depend on covariate information, otherwise, the system of estimating equations, $T(\theta) = 0$, are biased as shown in the next section, in which case, we need to model the missingness mechanism in addition to the competing risks model.

6. Missingness Mechanism

It is apparent from the factorization of $L^*(\theta, \psi)$ in Section 2 that the estimating equations based on the efficient partial likelihood $L^*(\theta)$ yield consistent parameter estimator without the need to model the missingness mechanism provided that cause of failure is missing at random. In contrast, when the missingness depends on covariates, the partial likelihood $L(\theta, \psi)$ does not permit a separation of parameters of interest for the competing risks model, θ, from the nuisance parameters for the missingness mechanism, ψ, and the estimating equations based on the partial likelihood $L(\theta)$ alone are biased. For example, consider the model for the two cause-specific hazards $\lambda_2(t|z) = \lambda(t)e^{\phi z}$, $\lambda_1(t|z) = \lambda(t)e^{\xi}$, and the model for the mechanism of missing cause of failure, $P(R = 1|T = t, Z = z, \Delta = \delta, \Delta > 0) = \pi(t, z)$, and suppose that the missingness model was misspecified as $\pi(t, z) = \pi(t)$ so that the estimating equations, $T(\theta) = (\partial\ell/\partial\phi, \partial\ell^*/\partial\xi) = 0$, were used. Under suitable regularity conditions, the estimator $\tilde{\theta}_n = (\tilde{\phi}_n, \tilde{\xi}_n)$ converges in probability to $\tilde{\theta} = (\tilde{\phi}, \tilde{\xi})$, where $\tilde{\theta}$ satisfies $\lim_{n\to\infty} n^{-1}T(\theta) = 0$.

For example, consider the case where $Z \sim Bernoulli(0.5)$. Given $Z = z$, T_2 and T_1 are conditionally independent with $\lambda_2(t|z) = \lambda e^{\phi z}$, $\lambda_1(t|z) = \lambda e^{\xi}$, where $\lambda = 1$, $\phi = 0.8$, and $\xi = 1$. Assume that there is no censoring. Let $T = \min(T_2, T_1)$ be the observed failure time, and $\Delta = 2$ or 1 indicate the cause of failure. In addition, assume that cause of failure is missing at random with $\pi(t, z) = \pi(z)$, where $\pi(1) = 0.8$, $\pi(0) = 0.5$. Then $\tilde{\phi} = 0.6876$, so that $bias(\tilde{\phi}_n) \to -0.1124$. A small scale simulation confirmed this theoretical result.

7. Discussion

Since the work of Goetghebeur and Ryan (1995), there have been several papers concerning the problem of hypothesis testing or parameter estimation when some failure types are missing. For example, Flehinger et al. (1998) developed a 2-stage approach for the situation in which systems are subject to independent competing risks and the hazards of various risks are proportional to each other. Their approach is applicable when there is a second stage of definitive diagnosis for a small sample of missing causes of failure. Recently, Lu and Tsiatis (2001) and Tsiatis et al. (2002) addressed the problem using multiple imputation method, where they postulate a parametric model for the probability of a failure from cause of interest given a failure occurred and estimate the parameters from complete cases. Theoretically, the performance of the multiple imputation estimator depends on the validity of the parametric model, although their simulation results suggest the approach is quite robust against model misspecification. For situations where one observes a set of possible failure types containing the true type if a failure type is not observed, Dewanji and Sengupta (2003) considered the EM algorithm and proposed a Nelson–Aalen type estimator when certain information on the conditional probability of the true type given the set of possible failure types is available.

Goetghebeur and Ryan (1995) partial likelihood approach has an appealing feature that individuals with known failure types make the same contributions as they would to a standard proportional hazards analysis, while contributions from individuals with unknown failure types are weighted according to the probability that they failed from the cause of interest. In comparison with the Dewanji partial likelihood approach, their approach is quite robust against misspecification of proportional baseline hazards and reduces to the usual estimator when there is no missing causes of failure. However, their approach is not as efficient as the Dewanji partial likelihood approach and may become complicated and subject to bias when probability of missing cause of failure depends on covariate information. It would be interesting to develop doubly robust estimators which remain valid when either the parametric model for the missingness mechanism or the parametric model relating the two competing causes of failure are correctly specified.

Acknowledgments

This work was partially supported by grant CA-51962 from the National Cancer Institute. The authors would like to thank referees for their constructive comments.

Appendix A: Proof of Semiparametric Efficiency of the Dewanji Partial Likelihood Estimator

The model is characterized by the $q \times 1$ parameter of interest θ and the infinite dimensional nuisance parameters $\{\lambda(t), \lambda_0(t|a), p_A(a), \pi(t, a)\}$, where $p_A(a) = P(A = a)$. Similar to Newey (1990), Bickel et al. (1993), and Robins et al. (1994), we consider the Hilbert space $\mathcal{H}$ of all q-dimensional mean zero and square-integrable measurable functions of the observed data for a typical subject $\{R, T, I(\Delta = 0), RI(\Delta = 1), RI(\Delta = 2), A\}$. The nuisance tangent space Λ is the linear subspace of $\mathcal{H}$ spanned by the scores for the nuisance parameters of all parametric submodels and their mean-square closure. It follows from the semiparametric theory that the solution to the estimating equation based on the efficient score is most efficient among all semiparametric estimators, where the efficient score is defined as the residual of the score vector for θ after being projected onto the nuisance tangent space, i.e., $S_{\text{eff}} = S_\theta - \Pi(S_\theta|\Lambda)$. To establish the semiparametric efficiency, we only need to identify the score vector S_θ, the nuisance tangent space Λ, carry out the projection, and verify the asymptotic equivalency of the estimating equation based on the efficient score and the estimating equation used to obtain the efficient partial likelihood estimator.

It is straightforward to show that the log likelihood for a single observation is given by

$$\begin{aligned}\ell(\theta) =& I(R = 1, \Delta > 0) \log \pi(T, A) + I(R = 0) \log\{1 - \pi(T, A)\} \\ &+ \log p_A(A) + I(\Delta = 0) \log \lambda_0(T|A) - \Lambda_0(T|A) + I(\Delta > 0) \log \lambda(T) \\ &+ \sum_{\delta=1}^{2} I(R = 1, \Delta = \delta) \log r_\delta(\theta, T, A) \\ &+ I(R = 0) \log r_{\cdot}(\theta, T, A) - \int \lambda(t) r_{\cdot}(\theta, t, A) Y(t) dt,\end{aligned}$$

where $\{\Lambda_\delta(t|a), \delta = 2, 1, 0\}$ are the cumulative cause-specific hazards.

Since the nuisance parameters are functionally independent and separate from each other in the log likelihood, the nuisance tangent space can be written as a direct sum of four orthogonal spaces,

$$\Lambda = \Lambda_{1s} \oplus \Lambda_{2s} \oplus \Lambda_{3s} \oplus \Lambda_{4s},$$

where Λ_{1s} is associated with $\lambda(t)$, Λ_{2s} associated with $\lambda_0(t|a)$, Λ_{3s} associated with $p_A(a)$, and Λ_{4s} associated with $\pi(t, a)$.

Let $dM_\delta(t) = dN_\delta(t) - \lambda^*_\delta(t|A)dt$ denote the martingale increments for the corresponding counting processes, then standard techniques of semiparametric theory can be used to show that a typical element of Λ_{1s} is given by

$$\int a(t)dM_{\cdot}(t),$$

where $M_{\cdot} = M_1 + M_2 + M_u$, and $a(\cdot)$ is some arbitrary $q \times 1$ function of t.

To simplify notation, write $r_\delta(t, a) = r_\delta(\theta_0, t, a)$, then the score vector for θ evaluated at the truth is given by

$$S_\theta = \sum_{\delta=1}^{2} \int \frac{r'_\delta(t, A)}{r_\delta(t, A)} dM_\delta(t) + \int \frac{r'_{\cdot}(t, A)}{r_{\cdot}(t, A)} dM_u(t).$$

Note that this is orthogonal to $\Lambda_{2s}, \Lambda_{3s}$ and Λ_{4s}. Therefore, by the projection theorem, the efficient score, derived as the residual after projecting S_θ onto Λ, or in this case, Λ_{1s}, is given by

$$S_{eff} = \sum_{\delta=1}^{2} \int \frac{r'_\delta(t, A)}{r_\delta(t, A)} dM_\delta(t) + \int \frac{r'_{\cdot}(t, A)}{r_{\cdot}(t, A)} dM_u(t) - \int a^*(t)dM_{\cdot}(t),$$

where

$$a^*(t) = \frac{E\{Y(t)r'_{\cdot}(t, A)\}}{E\{Y(t)r_{\cdot}(t, A)\}}.$$

The corresponding estimating equation is asymptotically equivalent to $\boldsymbol{U}(\theta) = 0$, so that the Dewanji partial likelihood estimator is semiparametric efficient.

References

P. D. Allison, *Survival Analysis Using the SAS System:* A Practical Guide, The SAS Institute: Cary, 1995.

J. Andersen, E. Goetghebeur, and L. Ryan, "Missing cause of death information in the analysis of survival data," *Statistics in Medicine* vol. 15 pp. 1191–2201, 1996.

P. K. Andersen, O. Borgan, R. D. Gill, and N. Keiding, *Statistical Models Based on Counting Processes*, Springer-Verlag: New York, 1997.

P. K. Andersen and R. D. Gill, "Cox's regression model for counting processes: a large sample study," *Annals of Statistics* vol. 10 pp. 1100–1120, 1982.

P. J. Bickel, C. A. Klaasen, Y. Ritov, and J. A. Wellner, *Efficient and Adaptive Estimation for Semiparametric Models*, John Hopkins University Press: Baltimore, 1993.

D. R. Cox, "Partial likelihood," *Biometrika* vol. 62 pp. 269–276, 1975.

D. R. Cox and D. Oakes, *Analysis of Survival Data*, Chapman & Hall: New York, 1984.

A. Dewanji, "A note on a test for competing risks with missing failure type," *Biometrika* vol. 79 pp. 855–857, 1992.

A. Dewanji and D. Sengupta, "Estimation of competing risks with general missing pattern in failure types," *Biometrics* vol. 59 pp. 1063–1070, 2003.

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

B. J. Flehinger, B. Reiser, and E. Yashchin, "Survival with competing risks and masked causes of failure," *Biometrika* vol. 85 pp. 151–164, 1998.

R. D. Gill and S. Johansen, "A survey of product-integration with a view towards application in survival analysis," *Annals of Statistics* vol. 18 pp. 1501–1555, 1990.

E. Goetghebeur and L. Ryan, "Analysis of competing risks survival data when some failure types are missing," *Biometrika* vol. 82 pp. 821–833, 1995.

P. M. Grambsch and T. M. Therneau, "Proportional hazards tests and diagnostics based on weighted residuals," *Biometrika* vol. 81 pp. 515–526, 1994.

J. D. Holt, "Competing risk analyses with special reference to matched pair experiments," *Biometrika* vol. 65 pp. 159–165, 1978.

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

D. Y. Lin and L. J. Wei, "The robust inference for the Cox proportional hazards model," *Journal of the American Statistical Association* vol. 84 pp. 1074–1078, 1989.

K. Lu and A. A. Tsiatis, "Multiple imputation methods for estimating regression coefficients in the competing risks model with missing cause of failure," *Biometrics* vol. 57 pp. 1191–1197, 2001.

W. K. Newey, "Semiparametric efficiency bounds," *Journal of Applied Econometrics* vol. 5 pp. 99–135, 1990.

R. L. Prentice and J. D. Kalbfleisch, "The analysis of failure times in the presence of competing risks," *Biometrics* vol. 34 pp. 541–554, 1978.

J. M. Robins, A. Rotnitzky, and L. P. Zhao, "Estimation of regression coefficients when some regressors are not always observed," *Journal of the American Statistical Association* vol. 89 pp. 846–866, 1994.

D. R. Rubin, "Inference and missing data," *Biometrika* vol. 63 pp. 581–592, 1976.

A. A. Tsiatis, M. Davidian, and B. McNeney, "Multiple imputation methods for testing different treatment differences in survival distributions with missing cause of failure," *Biometrika* vol. 89 pp. 238–244, 2002.