

University of North Carolina at Chapel Hill

The University of North Carolina at Chapel Hill Department of
Biostatistics Technical Report Series

Year 2012

Paper 29

Reader Reaction: On Variance Estimation for the Fine-Gray Model

Chenxi Li* Robert J. Gray[†]

Jason P. Fine[‡]

*University of North Carolina at Chapel Hill, chenxil@email.unc.edu

[†]Harvard University, gray@jimmy.harvard.edu

[‡]University of North Carolina at Chapel Hill, jfine@bios.unc.edu

This working paper is hosted by The Berkeley Electronic Press (bepress) and may not be commercially reproduced without the permission of the copyright holder.

<http://biostats.bepress.com/uncbiostat/art29>

Copyright ©2012 by the authors.

Reader Reaction: On Variance Estimation for the Fine-Gray Model

Chenxi Li, Robert J. Gray, and Jason P. Fine

Abstract

Geskus (2011, *Biometrics*, 67, 39-49) studied estimation of the Fine-Gray model for the cumulative incidence function with left truncated right censored competing risks data. The limiting distribution for an estimator based on weighting inversely using weights involving estimates of the joint distribution of the truncation and censoring times was derived via classical martingale theory with variance estimation based on martingale results. In this note, we demonstrate that martingale theory is not applicable and that other theoretical arguments, like those in Fine and Gray (1999), are needed to rigorously establish the asymptotic properties of the estimators and to construct valid variance estimators. For inverse probability of censoring weighted estimators, the common wisdom is that martingale theory fails because of estimation of the censoring distribution in the weights. For the Fine-Gray model, alternative theoretical developments are needed even with a known censoring distribution.

Reader Reaction: On Variance Estimation for the Fine-Gray Model

Chenxi Li^{1,*}, Robert J. Gray^{2,**3,**}, and Jason P. Fine^{1,2***}

¹Department of Biostatistics, University of North Carolina at Chapel Hill

Chapel Hill, NC 27599-7420, USA

²Department of Statistics, University of North Carolina at Chapel Hill

Chapel Hill, NC 27599-3260, USA

³Department of Biostatistics, Harvard University

Boston, MA 02115, USA

**email*: cli@bios.unc.edu

***email*: gray@jimmy.harvard.edu

****email*: jfine@bios.unc.edu

SUMMARY: Geskus (2011, *Biometrics*, **67**, 39-49) studied estimation of the Fine-Gray model for the cumulative incidence function with left truncated right censored competing risks data. The limiting distribution for an estimator based on weighting inversely using weights involving estimates of the joint distribution of the truncation and censoring times was derived via classical martingale theory with variance estimation based on martingale results. In this note, we demonstrate that martingale theory is not applicable and that other theoretical arguments, like those in Fine and Gray (1999), are needed to rigorously establish the asymptotic properties of the estimators and to construct valid variance estimators. For inverse probability of censoring weighted estimators, the common wisdom is that martingale theory fails because of estimation of the censoring distribution in the weights. For the Fine-Gray model, alternative theoretical developments are needed even with a known censoring distribution.

KEY WORDS: Filtration; Intensity; Martingale innovation theorem; Predictability; Subdistribution hazard.



This paper has been submitted for consideration for publication in *Biometrics*

1. Introduction

Geskus (2011) studied estimation of the regression parameters in a semiparametric proportional subdistribution hazards model (Fine and Gray, 1999) with competing risks data subject to left truncation and right censoring. The estimation procedure involves inverse weighting using estimates of the joint distribution of the truncation and censoring times. While the weighting scheme is valid, the martingale arguments used to derive the asymptotic distribution of the regression parameter estimates appears to be invalid with the resulting variance estimates lacking theoretical justification. A subtle error occurs in Geskus (2011) when defining the filtration in equation (20) given which the intensity process of a reweighted counting process of a specific event type is a product of the reweighted at risk process in equation (19) and the subdistribution hazard. Geskus (2011) claimed that the compensated process defined by subtracting the cumulative intensity process from the reweighted counting process is a martingale with respect to this filtration and that standard martingale techniques (Andersen and Gill, 1982) are applicable. The difficulty is that the reweighted at risk process is not predictable with respect to the filtration, invalidating the use of martingale techniques.

In Section 2, we clarify this point by formally establishing the nonpredictability of the reweighted at risk process and further demonstrating that enlarging the filtration to yield predictability of this process modifies the martingale structure such that the intensity is no longer defined in terms of the subdistribution hazard. To demonstrate these points, we begin by considering the case where there is no truncation and the censoring distribution is known in the weights. The common wisdom is that the failure of martingale theory when using inverse probability of censoring weighting arise because of estimation of the censoring distribution in the weights. Our discussion exhibits that even when the censoring distribution is known, there is additional information needed by the weighting scheme for the Fine-Gray model which is nonpredictable. It is straightforward to extend this line of reasoning to the

case where the censoring distribution is unknown and where truncation may be present. In Section 3, for didactic clarity, it is shown that for right censored competing risks data where the censoring distribution is known, standard martingale variance estimators are not valid and alternative theoretical arguments, like those in Fine and Gray (1999), are needed for variance estimation.

2. The Predictability Issue

In the sequel, we use notation as in Geskus (2011). Let T denote the event time, let D denote the event type, where $D \in \{1, \dots, K\}$, with K being the number of event types, and let Z be a $p \times 1$ time-independent covariate vector. In addition, let C be the censoring time and let L be the truncation time. Define $\Delta = I(T \leq C)$, where $I(\cdot)$ is the indicator function, and $X = T \wedge C$. The observed data from a left truncated right censored sample of size N are $\{(L_1, X_1, D_1\Delta_1, \Delta_1, Z_1), \dots, (L_N, X_N, D_N\Delta_N, \Delta_N, Z_N)\}$.

The cumulative incidence for the event of interest, type k say, is described by the distribution of the improper random variable

$$T^* = T \times I(D = k) + \infty \times I(D \neq k),$$

which is $F_k(t) = \Pr(T \leq t, D = k)$. Let $\lambda_k^*(t) = dF_k(t)\{1 - F_k(t)\}^{-1}$ denote the subdistribution hazard function of cause k . A popular approach to examining the association between F_k and Z is the proportional subdistribution hazards model, where conditionally on covariates

$$\lambda_k^*(t|Z) = dF_k(t|Z)\{1 - F_k(t|Z)\}^{-1} = \lambda_{k0}^*(t)\exp(Z\beta_0).$$

This model was originally proposed in Fine and Gray (1999), where estimation of β_0 was proposed using inverse probability of censoring weighting of the score equations from a particular partial likelihood. Geskus (2011) developed a modified inverse weighting scheme for left truncated right censored competing risks data in which estimates of both the censoring

and truncation distributions are employed in the estimated weights. This modified weighting scheme reduces to that in Fine and Gray (1999) in the absence of left truncation.

Fine and Gray (1999) demonstrated that with complete data, standard martingale arguments (Andersen and Gill, 1999) could be applied to an unweighted partial likelihood, with the usual martingale variance estimator being valid. Their derivations utilize a special counting process representation for competing risks data which is appropriate when modelling the subdistribution hazard function. Define $T_i^* = T_i \times I(D_i = k) + \infty \times I(D_i \neq k)$ and let $N_i^*(t) = I(T_i^* \leq t)$ denote whether the event of interest, type k say, has occurred for individual i by time t and $Y_i^*(t) = I(T_i^* \geq t)$ denote whether the individual i is either still at risk at t or has previously experienced a competing event. In the absence of either right censoring or left truncation, Fine and Gray (1999) showed that under the assumed regression model $N_i^*(t)$ has compensator $\Lambda_i^*(t) = \int_0^t Y_i^*(u) \lambda_{k,i}^*(u) du$, with respect to the filtration

$$\mathcal{F}^1(t) = \sigma\{N_i^*(u), Y_i^*(u)Z_i, 0 \leq u \leq t, i = 1, \dots, N\},$$

where $\lambda_{k,i}^*(u) = \lambda_{k0}^*(t) \exp(Z_i \beta_0)$. Hence $M_i^*(t) = N_i^*(t) - \Lambda_i^*(t)$ is a martingale with respect to $\mathcal{F}^1(t)$, enabling the application of martingale theory.

In defining weights for estimation of β_0 in the presence of censoring and truncation, Geskus (2011) assumes that the distribution of C , denoted by G , and the distribution of L , denoted by H , are independent of the covariates in Z . The estimated weights involve nonparametric estimators of G and H . To define these estimators, some notation is needed. Let $t_{(1)} < t_{(2)} < \dots < t_{(d)}$ denote the d ordered distinct observed event times of any event type. Similarly let $c_{(1)} < c_{(2)} < \dots < c_{(m)}$ and $l_{(1)} < l_{(2)} < \dots < l_{(k)}$ be the m ordered distinct observed censoring and k ordered distinct truncation times. Let $d(t)$ and $r(t)$ be the number of observed events of any type and the number of subjects at risk at time t , respectively. Let m_i be the number of censorings at $c_{(i)}$ and let w_i be the number of left truncation times at $l_{(i)}$. Define \hat{G} to be the left truncated right censored product limit estimator of $\bar{G} = 1 - G$,

which is obtained by reversing the role of T and C :

$$\hat{G} = \prod_{j:c(j) \leq t} \left(1 - \frac{m_j}{r(c(j))}\right).$$

Similarly, define \hat{H} to be the reverse time right truncated product limit estimator of H , which is obtained by switching the role of L and X :

$$\hat{H}(t) = \prod_{-t(j) < -t} \left(1 - \frac{w_j}{r(l(j))}\right) = \prod_{t(j) > t} \left(1 - \frac{w_j}{r(l(j))}\right).$$

The reweighted at risk process for subject i defined by Geskus (2011, Equation (19)) is

$$\psi_i(u) = \frac{N}{\hat{N}} \times \frac{\mathbb{I}[\{L_i \leq u \leq X_i\} \cup \{L_i \leq T_i^0 \leq (u \wedge C_i)\}]}{\hat{G}\{(u \wedge T_i^0)-\} \times \hat{H}\{(u \wedge T_i^0)-\}}, \quad \text{where}$$

$$\hat{N} = \sum_{i=1}^n \frac{d(t_{(i)})}{\hat{H}(t_{(i)}-)} + \sum_{j=1}^m \frac{m_j}{\hat{H}(c(j)-)}$$

and $T_i^{(0)} = T_i \times I(D_i \neq k) + \infty \times I(D_i = k)$. This quantity simplifies to that in Fine and Gray (1999) without truncation.

To study the asymptotic properties of the inversely weighted partial likelihood estimators, Geskus (2011) defines ${}_{L_i}N_i^*$ to be the counting process started at L_i , that is, ${}_{L_i}N_i^*(t) = N_i^*(t) - N_i^*(t \wedge L_i)$, and defines ${}_{L_i}N_i^{*c}(t) = \int_0^t \mathbb{I}(C_i \geq u) d\{{}_{L_i}N_i^*(u)\}$. In order to account for the estimated quantities in the reweighted at risk process, Geskus augments the filtration defined by ${}_{L_i}N_i^{*c}(t)$ with quantities which are estimated in $\psi_i(u)$. A martingale structure is derived with respect to the filtration defined by Geskus (2011, Equation (20)), which is

$$\widetilde{\mathcal{F}}^*(t) = \sigma\{{}_{L_i}N_i^{*c}(u), \hat{G}(u), \hat{N}\hat{H}(u), Z_i, u \leq t, i = 1, \dots, N\}.$$

A key step in the proof involves expressing the compensator for ${}_{L_i}N_i^{*c}(t)$ in terms of $\psi_i(u)$ and the subdistribution hazard function, which requires that $\psi_i(u)$ is predictable with respect to $\widetilde{\mathcal{F}}^*(t)$. If predictability is satisfied, then

$$\psi_i(t) {}_{L_i}N_i^*(t) - \int_0^t \psi_i(u) \lambda_{1,i}^*(u) du,$$

has the martingale property. We show below that coupling $\hat{N}\hat{H}(t)$ and \hat{G} with ${}_{L_i}N_i^{*c}(t)$ is insufficient to obtain predictability of $\psi_i(u)$.

To demonstrate lack of predictability, we first consider the simplified setting in which there is no left truncation and \hat{G} is replaced by G in ψ_i . This corresponds to fitting the Fine-Gray model with right censoring when the censoring distribution is known. The theoretical arguments are more transparent under this set-up and are later extended to the practically relevant settings with unknown censoring and truncation distributions. In this scenario, $\hat{N} = N$, $H(u) = \hat{H}(u) = 1$ for all u . Following the reasoning in Geskus (2011), there is no additional information needed for predictability of $\psi_i(u)$ and it is unnecessary to augment $N_i^{*c}(t)$ in

$$\widetilde{\mathcal{F}}^*(t) = \sigma\{N_i^{*c}(u), Z_i, 0 \leq u \leq t, i = 1, \dots, N\},$$

where $N_i^{*c} =_{L_i} N_i^{*c}$ with $L_i = 0, i = 1, \dots, N$. By the definition of ψ_i in Equation (19) of Geskus (2011), one can easily show that

$$\begin{aligned} \psi_i(t) &= \frac{1}{\bar{G}(t-)} \mathbf{I}\{t \leq \min(C_i, T_i^0)\} + \frac{1}{\bar{G}(T_i^0-)} [\mathbf{I}\{T_i^0 < t \leq C_i\} + \mathbf{I}\{T_i^0 \leq C_i < t\}] \\ &\quad + 0 \cdot \mathbf{I}\{C_i < \min(T_i^0, t)\}. \end{aligned} \quad (1)$$

given Z_i and $N_i^{*c}(u) = 0, 0 \leq u < t$. Observe that the right side of (1) is not deterministic given Z_i and $N_i^{*c}(u) = 0, 0 \leq u < t$. This occurs because the history of $N_i^{*c}(u)$ is insufficient to completely determine the indicators involving t in $\psi_i(t)$. As a result,

$$\begin{aligned} &E\{\psi_i(t) | N_i^{*c}(u) = 0, Z_i, 0 \leq u < t\} \\ &= \frac{1}{\bar{G}(t-)} \text{pr}\{t \leq \min(C_i, T_i^0) | N_i^{*c}(u) = 0, Z_i, 0 \leq u < t\} \\ &\quad + E\left[\frac{1}{\bar{G}(T_i^0-)} [\mathbf{I}\{T_i^0 < t \leq C_i\} + \mathbf{I}\{T_i^0 \leq C_i < t\}] \middle| N_i^{*c}(u) = 0, Z_i, 0 \leq u < t\right] \\ &\neq \psi_i(t). \end{aligned}$$

Hence, $\psi_i(t)$ is not predictable with respect to the filtration $\widetilde{\mathcal{F}}^*(t)$.

From (1), one observes that for $\psi_i(t)$ to be predictable the filtration $\widetilde{\mathcal{F}}^*(t)$ needs to be augmented with additional event history information. An augmented filtration containing

only that information needed for predictability is

$$\widetilde{\mathcal{F}}_{aug}^*(t) = \sigma\{N_i^{*c}(u), I\{u \leq \min(C_i, T_i^0)\}, I\{C_i < \min(T_i^0, u)\}, Z_i, 0 \leq u \leq t, i = 1, \dots, N\}.$$

Interestingly, it is straightforward to show that the intensity process of $N_i^{*c}(t)$ with respect to $\widetilde{\mathcal{F}}_{aug}^*(t)$ is $I\{X_i \geq t\}\lambda_{1,i}(t)$. Here, $\lambda_{1,i}(t) = dF_{1,i}(t)\{1 - \sum_m F_{m,i}(t)\}^{-1}$, where $F_{m,i}(t) = F_m(t|Z_i)$, is the cause-specific hazard function of cause 1 for subject i and not the subdistribution hazard $\lambda_{1,i}^*(t)$. One should recognize that it is not possible to derive the limiting distribution of the weighted partial likelihood estimator for regression parameters in the proportional subdistribution hazard model using this cause-specific hazard martingale defined with respect to the augmented filtration.

Now, suppose that the censoring distribution G is unknown, as is usually the case in practice, and one uses \hat{G} in the weight function. The augmented filtration $\widetilde{\mathcal{F}}_{aug}^*(t)$ does not need to be further augmented to obtain the predictability of $\psi_i(t)$. The reason is that $\hat{G}(u)$ is completely determined by the information in $\widetilde{\mathcal{F}}_{aug}^*(t)$, so that adding \hat{G} provides no additional information. Note that augmenting $\widetilde{\mathcal{F}}^*(t)$ with \hat{G} instead of using the filtration $\widetilde{\mathcal{F}}_{aug}^*(t)$ is insufficient to obtain predictability of $\psi_i(u)$ because the extra history in $\widetilde{\mathcal{F}}_{aug}^*(t)$ is not completely determined by \hat{G} . The same reasoning can be used to demonstrate these predictability issues in the presence of both left truncation and right censoring.

Because of the lack of predictability of the reweighted at risk process $\psi_i(u)$, the asymptotic results in Geskus (2011, Theorem 1) do not appear to be correct. In particular, the asymptotic covariance matrix of $\hat{\beta}$ cannot be derived using martingale theory. Alternative arguments are needed, like those in Fine and Gray (1999) for weighted estimating equations with right censored data.

3. Variance Estimation When There is No Left Truncation and the Right Censoring Distribution is Known

Fine and Gray (1999) derived the asymptotic properties of their estimator with estimated G using empirical process techniques, with the corresponding variance not equalling that which would ordinarily be obtained using martingale techniques. It was suggested that the usual martingale theory was not applicable because of the estimation of G , with the implicit implication that martingale theory is applicable with known G . Following the discussion of predictability in Section 2, in this section, we provide explicit details of the method of proof for the case where G is known. This discussion further clarifies the lack of applicability of martingale theory, with either known or unknown G .

The estimating equation for randomly right censored data in Fine and Gray (1999), which is equivalent to that in Geskus (2011) without left truncation, is

$$U(\beta) = \sum_{i=1}^N \int_0^{\infty} \left\{ Z_i - \frac{\sum_j w_j(u) Y_j^*(u) Z_j \exp\{Z_j^T \beta\}}{\sum_j w_j(u) Y_j^*(u) \exp\{Z_j \beta\}} \right\} w_i(u) dN_i^*(u), \quad (2)$$

where $w_i(u) = I(C_i \geq T_i \wedge u) \hat{G}(u-) / \hat{G}((X_i \wedge u)-)$. If G is known, one may replace the estimate of \hat{G} in the estimating equation (2) by its known value, yielding the following estimating equation

$$\tilde{U}(\beta) = \sum_{i=1}^N \int_0^{\infty} \left\{ Z_i - \frac{\sum_j \tilde{w}_j(u) Y_j^*(u) Z_j \exp\{Z_j \beta\}}{\sum_j \tilde{w}_j(u) Y_j^*(u) \exp\{Z_j \beta\}} \right\} \tilde{w}_i(u) dN_i^*(u), \quad (3)$$

where $\tilde{w}_i(u) = I\{C_i \geq T_i \wedge u\} \bar{G}(u-) \{\bar{G}((X_i \wedge u)-)\}^{-1}$. The estimating equation (3) may be centered, as in Fine and Gray (1999), giving

$$\tilde{U}(\beta) = \sum_{i=1}^N \int_0^{\infty} \left\{ Z_i - \frac{\sum_j \tilde{w}_j(u) Y_j^*(u) Z_j \exp\{Z_j \beta\}}{\sum_j \tilde{w}_j(u) Y_j^*(u) \exp\{Z_j \beta\}} \right\} \tilde{w}_i(u) dM_i^*(u, \beta). \quad (4)$$

For censoring-complete data for which C_i is always observed, Fine and Gray (1999) obtained a martingale-type estimating function from the score of the censoring-complete-data partial likelihood,

$$U_{1^*}(\beta) = \sum_{i=1}^N \int_0^{\infty} \left[Z_i - \frac{\sum_j Y_j^{1^*}(u) Z_j \exp(Z_j \beta)}{\sum_j Y_j^{1^*}(u) \exp(Z_j \beta)} \right] dM_i^{1^*}(u, \beta), \quad (5)$$

where $Y_i^{1*}(t) = I(C_i \geq t)Y_i^*(t)$ and

$$M_i^{1*}(t, \beta) = \int_0^t I(C_i \geq u) dN_i^*(u) - \int_0^t Y_i^{1*}(u) \lambda_{10}^*(u) \exp\{Z_j \beta\} du.$$

This covers the complete data situation without censoring, where $C_i = \infty$, all i . Fine and Gray (1999) showed $M_i^{1*}(t, \beta_0)$ is a martingale under the censoring-complete filtration,

$$\mathcal{F}^{1*}(t) = \sigma\{I(C_i \geq u), \int_0^u I(C_i \geq s) dN_i^*(s), Y_i^{1*}(u), Z_i, u \leq t, i = 1, \dots, N\}.$$

Under $\mathcal{F}^{1*}(t)$, $U_{1*}(\beta_0)$ is a sum of martingale integrals with respect to the locally bounded predictable process

$$H_i^*(u) = Z_i - \frac{\sum_j Y_j^{1*}(u) Z_j \exp\{Z_j \beta_0\}}{\sum_j Y_j^{1*}(u) \exp\{Z_j \beta_0\}}.$$

Using these facts, similar arguments to Andersen and Gill (1982) establish the consistency and asymptotic normality of the estimate of β from the estimating equation $U_{1*}(\beta) = 0$.

As argued in Fine and Gray (1999), for censoring-complete data, as $N \rightarrow \infty$, martingale results give that $N^{-1}U_{1*}(\beta)$ converges uniformly in probability to a continuous and deterministic function of β that has a unique 0 at β_0 and is bounded in a neighborhood of β_0 . For consistency with randomly right censored data when there is no left truncation and G is known, it suffices to prove that $N^{-1}\{\tilde{U}(\beta) - U_{1*}(\beta)\}$ goes in probability to 0 uniformly for β in a compact neighborhood of β_0 . Similarly to Section 2, one may demonstrate that $\tilde{w}_i(u)$ is not predictable with respect to the complete data filtration $\mathcal{F}^1(u)$, under which $M_i^*(u, \beta_0)$ satisfies the definition of a martingale, and that augmenting the filtration to obtain predictability is invalid. However, using empirical process techniques like those in Fine and Gray (1999), one can show that

$$N^{-1}\{\tilde{U}(\beta) - U_{1*}(\beta)\} = N^{-1} \sum_{i=1}^N \int_0^\infty \left\{ Z_i(u) - \frac{\mathbf{s}^{(1)}(\beta_0, u)}{\mathbf{s}^{(0)}(\beta_0, u)} \right\} \{\tilde{w}_i(u) - I(C_i \geq u)\} dM_i^*(u, \beta) + o_p(1), \quad (6)$$

where

$$\mathbf{s}^{(p)}(\beta, u) = \lim_{N \rightarrow \infty} N^{-1} \sum_{i=1}^N Y_i^*(u) Z_i(u)^{\otimes p} \exp\{Z_i^T(u) \beta\}, \quad p = 0, 1, 2.$$

Furthermore, the right-hand side of (6) is continuous and dominated by a bounded function, so $N^{-1}\{\tilde{U}(\beta) - U_{1*}(\beta)\}$ vanishes uniformly for β in a compact neighborhood of β_0 (Newey and McFadden, 1986, Lemma 2.4). Accordingly, $\hat{\beta}$ satisfying $\tilde{U}(\hat{\beta}) = 0$, is consistent.

To obtain the asymptotic normality of $\hat{\beta}$, one may mimic Fine and Gray (1999). A first order Taylor expansion of $\tilde{U}(\hat{\beta})$ around β_0 gives that:

$$N^{1/2}(\hat{\beta} - \beta_0) \approx \Omega^{-1}\{N^{-1/2}\tilde{U}(\beta_0)\}, \quad (7)$$

where Ω^{-1} is the limit of the negative of the inverse of the partial derivative matrix of the score function evaluated at β_0 and is equal to the variance of the censoring-complete regression coefficients or, equivalently, the limit of $-N\{d\{U_{1*}(\beta)\}/d\beta\}^{-1}$ at β_0 . It is important to recognize that with censoring-complete data, the implicit weight $I(C_i \geq t)$ has the same conditional expectation as $\tilde{w}_i(t)$ given Z_i . Specifically,

$$\Omega = \int_0^\infty \left\{ \frac{\mathbf{s}^{(2)}(\beta_0, u)}{s^{(0)}(\beta_0, u)} - \bar{\mathbf{z}}(\beta_0, u)^{\otimes 2} \right\} \bar{G}(u-)s^{(0)}(\beta_0, u)\lambda_{10}^*(u)du,$$

where $\bar{\mathbf{z}}(\beta, u) = \mathbf{s}^{(1)}(\beta, u)\{s^{(0)}(\beta, u)\}^{-1}$. However, because of the lack of predictability of \tilde{w}_i which occurs even when G is known, martingale theory cannot be used. Empirical process arguments like those in Fine and Gray (1999) are needed to show that $N^{-1/2}\tilde{U}(\beta_0) = N^{-1/2}\sum_{i=1}^N \eta_i + o_p(1)$, where

$$\eta_i = \int_0^\infty \left\{ Z_i(u) - \frac{\mathbf{s}^{(1)}(\beta_0, u)}{s^{(0)}(\beta_0, u)} \right\} \tilde{w}_i(u)dM_i^*(u, \beta_0). \quad (8)$$

The influence function η_i equals the first term of the influence function in Fine and Gray (1999). By the multivariate central limit theorem, $N^{-1/2}\tilde{U}(\beta_0)$ is asymptotically normal with covariance matrix $\Sigma = E\{\eta_i\eta_i^T\}$. Consequently, $N^{1/2}(\hat{\beta} - \beta_0)$ has a limiting normal distribution with covariance matrix $\Omega^{-1}\Sigma\Omega^{-1}$. In general, because of the lack of a martingale structure, $\Omega \neq \Sigma$ and martingale variance estimators are invalid. The variance may be estimated as discussed in Fine and Gray (1999), with $\hat{\Omega}$ computed empirically with $\beta = \hat{\beta}$ and $\hat{\Sigma}$ calculated ignoring those terms in the influence function which arise because of variability in the estimation of G .

REFERENCES

- Andersen, P. K. and Gill, R. D. (1982). Cox's regression model for counting processes: A large sample study. *The Annals of Statistics* **10**, pp. 1100–1120.
- Fine, J. P. and Gray, R. J. (1999). A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association* **94**, 496–509.
- Geskus, R. B. (2011). Cause-specific cumulative incidence estimation and the fine and gray model under both left truncation and right censoring. *Biometrics* **67**, 39–49.
- Newey, W. K. and McFadden, D. (1986). Large sample estimation and hypothesis testing. In Engle, R. F. and McFadden, D., editors, *Handbook of Econometrics*, volume 4 of *Handbook of Econometrics*, chapter 36, pages 2111–2245. Elsevier.

Received October 2004. Revised February 2005.

Accepted March 2005.

