Journal of the American Statistical Association

ISSN: 0162-1459 (Print) 1537-274X (Online) Journal homepage: www.tandfonline.com/journals/uasa20

A Proportional Hazards Model for the Subdistribution of a Competing Risk

Jason P. Fine & Robert J. Gray

To cite this article: Jason P. Fine & Robert J. Gray (1999) A Proportional Hazards Model for the Subdistribution of a Competing Risk, Journal of the American Statistical Association, 94:446, 496-509, DOI: 10.1080/01621459.1999.10474144

To link to this article: https://doi.org/10.1080/01621459.1999.10474144

Published online: 17 Feb 2012.

A Proportional Hazards Model for the Subdistribution of a Competing Risk

Jason P. FINE and Robert J. GRAY

With explanatory covariates, the standard analysis for competing risks data involves modeling the cause-specific hazard functions via a proportional hazards assumption. Unfortunately, the cause-specific hazard function does not have a direct interpretation in terms of survival probabilities for the particular failure type. In recent years many clinicians have begun using the cumulative incidence function, the marginal failure probabilities for a particular cause, which is intuitively appealing and more easily explained to the nonstatistician. The cumulative incidence is especially relevant in cost-effectiveness analyses in which the survival probabilities are needed to determine treatment utility. Previously, authors have considered methods for combining estimates of the cause-specific hazard functions under the proportional hazards formulation. However, these methods do not allow the analyst to directly assess the effect of a covariate on the marginal probability function. In this article we propose a novel semiparametric proportional hazards model for the subdistribution. Using the partial likelihood principle and weighting techniques, we derive estimation and inference procedures for the finite-dimensional regression parameter under a variety of censoring scenarios. We give a uniformly consistent estimator for the predicted cumulative incidence for an individual with certain covariates; confidence intervals and bands can be obtained analytically or with an easy-to-implement simulation technique. To contrast the two approaches, we analyze a dataset from a breast cancer clinical trial under both models.

KEY WORDS: Hazard of subdistribution; Martingale; Partial likelihood; Transformation model.

1. MOTIVATION

Competing-risks data are inherent to medical research in which response to treatment can be classified in terms of failure from disease processes and/or non–disease-related causes. Historically, the cumulative incidence function (also known as the subdistribution, the marginal probability function, the crude incidence, and the absolute cause-specific risk) has provided information that is secondary to that contained in overall survival (Benichou and Gail 1990; Gaynor et al. 1993; Korn and Dorey 1992). In recent years, this quantity has become important in its own right. Quality-of-life studies, which take as input the absolute risks of the different failure types, have become central to cost-effectiveness analyses. Understanding the effects of therapy on different subgroups, intervention can be targeted for those populations most likely to benefit at a reasonable expense and with nominal risk of complications. The cumulative incidence function, which is intuitively appealing and well suited to graphical display, is essential to the decision-making process (Pepe and Mori 1993).

As an example, consider clinical trial E1178 conducted by the Eastern Cooperative Oncology Group. This study compared 2 years of tamoxifen therapy to placebo in elderly ($\geq$ age 65) breast cancer patients with positive axillary nodes. For particular groups of patients, direct information on the probability of different endpoints being observed over time may be more relevant to clinical management of breast cancer than information on cause-specific hazards. That is, the overall probabilities of breast cancer recurrence and the overall probabilities of dying from unrelated causes are directly relevant to different treatment options. In this article we model the cumulative incidence functions for recurrence and nonrecurrence death conditional on treatment, age, tumor size, and number of postive nodes. After adjusting for patient-specific risk factors, tamoxifen appears to decrease the probability of failure from tumor recurrence, but has no effect on failure from unrelated causes. We summarize the absolute risks for several groups of women by predicting their cause-specific failure probabilities under the estimated models for the cumulative incidence functions. In one group, the absolute risk of breast cancer recurrence is high; in another, the risk of recurrence is lower than the risk of dying without recurrence. This information has implications for proper clinical management of the different groups of patients.

With covariates, the standard analysis for competing-risks data involves modeling the cause-specific hazard functions of the different failure types under a proportional hazards assumption (Larson 1984; Prentice et al. 1978). Prediction of the marginal probability function for an individual with certain covariates can be accomplished by combining estimates of the cause-specific hazard functions from the partial likelihood approach (Andersen, Borgan, Gill, and Keiding 1993, pp. 512–515). However, many authors have noted that the effect of a covariate on the cause-specific hazard function of a particular failure type may be very different from the effect of the covariate on the corresponding cumulative incidence function (Gray 1988; Pepe 1991). In the most extreme case, a covariate may have strong influence on the cause-specific hazard function but no effect on the cause-specific subdistribution. Thus, under the cause-specific hazard formulation, testing for covariate effects on the subdistribution is not possible, and issues of model se-

Jason P. Fine is Assistant Professor, Department of Statistics and Department of Biostatistics and Medical Informatics, University of Wisconsin, Madison, WI 53706 (E-mail: *fine@biostat.wisc.edu*). Robert J. Gray is Senior Lecturer, Department of Biostatistics, Harvard University, Boston, MA 02115 (E-mail: *gray@jimmy.harvard.edu*). This research was partially supported by National Institutes of Health grants T32-MH17119 (J. P. Fine) and CA57253 and CA39929 (R. J. Gray). The first author is grateful to his thesis advisor L. J. Wei for encouraging this study. The authors thank the Eastern Cooperative Oncology Group for permission to use the breast cancer data. A documented S-PLUS function can be obtained from the authors.

Journal of the American Statistical Association
June 1999, Vol. 94, No. 446, Theory and Methods

lection and efficient prediction cannot be directly addressed. To this end, a parsimonious semiparametric model for the crude incidence, akin to the Cox model for univariate failure times, would be very useful.

Previous work on the cumulative incidence function has focused on nonparametric tests for group effects when covariates can be discretized into a finite number of configurations (Gray 1988; Pepe 1991). In Section 2 we present our notation and introduce a *proportional hazards model for the subdistribution* that is grounded in the log(−log) transformation model commonly used with univariate survival data (Cox 1972). In Section 3 we show that the partial likelihood approach to estimation and prediction is applicable when censoring is absent or when censoring is present but always observed. In Section 4 we develop a weighted score function using the results from Section 3 that allows estimation of the regression parameters under right censoring. In Section 5 we outline a simulation-based procedure for constructing confidence intervals and bands for the cumulative incidence function for an individual with certain covariates. In Section 6 we carry out two numerical studies to assess the operating characteristics of the weighted score function from Section 4. The first study compares the weighted score function to the partial likelihood approach from Section 3; the second considers the relative efficiency of the weighted to Gray's K-sample statistic in the two-sample situation. In Section 7 we analyze the breast cancer dataset under both the proportional hazards model for the subdistribution and the Cox model for the cause-specific hazard functions. Finally, in Section 8 we discuss relevant issues regarding the procedures presented in this article. We provide details of asymptotic results in the appendixes.

2. THE MODEL

Let T and C be the failure and censoring times, let $\varepsilon \in (1, \ldots, K)$ be cause of failure (for which the K causes are assumed to be observable), and let $\mathbf{Z}$ be a $p \times 1$ bounded time-independent covariate vector. For the usual right-censored data, we observe $X = \min(T, C)$, $\Delta = I(T \leq C)$, and $\mathbf{Z}$, where $I(\cdot)$ is the indicator function. Assume that $\{X_i, \Delta_i, \Delta_i \varepsilon_i, \mathbf{Z}_i\}$ are independent and identically distributed for $i = 1, \ldots, n$. Here our interest is modeling the cumulative incidence function for failure from cause 1 conditional on the covariates, $F_1(t; \mathbf{Z}) = \Pr(T \leq t, \varepsilon = 1 | \mathbf{Z})$.

It seems reasonable to consider the class of semiparametric transformation models that has been rigorously studied with univariate survival data (Cheng, Wei, and Ying, 1995, 1997; Cox 1972; Cuzick 1988; Dabrowska and Doksum 1988; Fine, Ying, and Wei 1998; Murphy, Rossini, and Van der Vaart 1997). Assume that for some known increasing function $g(\cdot)$, $g\{F_1(t; \mathbf{Z})\} = h_0(t) + \mathbf{Z}^T \beta_0$, where $h_0(\cdot)$ is a completely unspecified, invertible, and monotone increasing function and β_0 is a $p \times 1$ parameter vector. For two individuals with covariate vectors $\mathbf{Z}_1$ and $\mathbf{Z}_2$ the conditional cumulative incidence functions satisfy a vertical shift model after transformation. That is, $g\{F_1(t; \mathbf{Z}_1)\} - g\{F_1(t; \mathbf{Z}_2)\} = (\mathbf{Z}_1 - \mathbf{Z}_2)^T \beta_0$ for all t. On the scale of $g(\cdot)$, the regression coefficients are a measure of distance from the baseline marginal probability function, $g^{-1}\{h_0(t)\}$, for which the covariates are identically 0.

A natural first step is to investigate $g = \log\{-\log(1-u)\}$, corresponding to the popular proportional hazards model. To directly estimate this transformation model without simultaneously estimating models for $F_k, k = 2, \ldots, K$, we utilize the hazard of the subdistribution as originally described by Gray (1988), who constructed K-sample tests for differences in the cumulative incidence function based on integrated differences of nonparametric estimates of the within-group subdistribution hazard functions. To formalize the subdistribution hazard, we restate Gray's definition:

$$\begin{aligned}\lambda_1(t; \mathbf{Z}) &= \lim_{\Delta t \to 0} \frac{1}{\Delta t} \Pr\{t \leq T \leq t + \Delta t, \varepsilon = 1 \\ &\qquad | T \geq t \cup (T \leq t \cap \varepsilon \neq 1), \mathbf{Z}\} \\ &= \{dF_1(t; \mathbf{Z})/dt\}/\{1 - F_1(t; \mathbf{Z})\} \\ &= -d \log\{1 - F_1(t; \mathbf{Z})\}/dt.\end{aligned}$$

One can think of λ_1 as the hazard function for the improper random variable $T^* = I(\varepsilon = 1) \times T + \{1 - I(\varepsilon = 1)\} \times \infty$. The implied failure time T^* has distribution function equal to $F_1(t; \mathbf{Z})$, $t < \infty$, and a point mass at $t = \infty$ that is just $\Pr(T^* = \infty | \mathbf{Z}) = \Pr(T < \infty, \varepsilon \neq 1 | \mathbf{Z}) = 1 - F_1(\infty; \mathbf{Z})$.

Clearly, the risk set associated with the hazard λ_1 is unnatural, as in reality those individuals who have already failed from causes other than $\varepsilon = 1$ prior to time t are not "at risk" at t. Note, however, that this construction does have precedent in the cure model in which a subgroup of patients who will never develop disease are included in the risk set for survival calculations. There, failure from other causes (i.e., cure) is unobservable, and estimation of overall survival is tantamount to estimation of the subdistribution for individuals who will eventually experience the event of interest. In the general competing-risks setting, failure from other causes is observable, which leads to difficulty in conceptualizing a risk set in which some individuals are known to have zero probability of $\varepsilon = 1$. Therefore, interpretation of a g-transformation model for the cumulative incidence function is problematic if viewed in terms of the corresponding hazard function.

Nonetheless, under a proportional hazards specification with $\lambda_1(t; \mathbf{Z}) = \lambda_{10}(t) \exp(\mathbf{Z}^T \beta_0)$, where $\lambda_{10}(t)$ is a completely unspecified, nonnegative function in t, the log(−log) transformation model results with $h_0(t) = \log\{\int_0^t \lambda_{10}(s)\, ds\}$. Thus the regression coefficients and baseline hazard from the Cox transformation model for F_1 have a straightforward interpretation that does not depend on the probabilistic structure of the subdistribution hazard. In applications, we anticipate time × covariate interactions. To address this issue, we extend the model to the case of time-varying covariates $\mathbf{Z}(t)$, which are functions of the original, time-independent covariates $\mathbf{Z}$ and t. Let

$$\lambda_1\{t; \mathbf{Z}\} = \lambda_{10}(t) \exp\{\mathbf{Z}^T(t) \beta_0\}, \tag{1}$$

so that $F_1(t;\mathbf{Z}) = 1 - \exp[-\int_0^t \lambda_{10}(s)\exp\{\mathbf{Z}^T(s)\boldsymbol{\beta}_0\}\,ds]$. With the inclusion of interaction terms, the regression coefficients for time-independent covariates now satisfy the transformation model within stratum defined by baseline covariates present in the interactions.

The familiar form of this proportional hazards model is intended to be a convenient empirical representation for the cumulative risk of a competing risk and should be evaluated on the extent to which it permits the analyst to assess the effect of covariates on the cumulative incidence function, via easy-to-implement testing, model selection, and prediction procedures. As we show, many of the techniques applicable with survival data when there is only a single cause of failure will also be useful in modeling the cumulative incidence function in the competing-risks setting.

3. ESTIMATION AND PREDICTION WITH COMPLETE DATA

3.1 Complete Data

In this section we present a modification of the partial likelihood for the subdistribution for use with *complete* data. (By complete data, we mean that T and ε are observed on all individuals.) We show that this procedure yields estimates for the regression parameters that are consistent and asymptotically normal. A version of Breslow's estimator (Breslow 1974) provides a consistent estimate for $\Lambda_{10}(t) = \int_0^t \lambda_{10}(s)\,ds$ that is asymptotically equivalent to a mean 0 Gaussian process. Predicting $F_1(t;\mathbf{Z})$ is straightforward; confidence intervals and bands can be constructed using adaptations of the usual methods.

Without censoring, the partial likelihood approach is applicable to $\lambda_1\{t;\mathbf{Z}\}$. However, the risk sets will not be the same as those for the partial likelihood of the marginal or cause-specific hazards. We define R_i, the risk set at the time of failure for the ith individual, as $\{j\colon (T_j \geq T_i) \cup (T_j \leq T_i \cap \varepsilon_j \neq 1)\}$. An individual who has not failed from the cause of interest by time t is at risk. This includes two distinct groups: those who have not failed from any cause and those who have previously failed from another cause. Although the risk set is unconventional, it leads to a proper partial likelihoood for the improper distribution, $F_1(t;\mathbf{Z})$:

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{n}\left[\frac{\lambda_{10}(T_i)\exp\{\mathbf{Z}_i^T(T_i)\boldsymbol{\beta}\}\Delta T_i}{\sum_{j\in R_i}\lambda_{10}(T_i)\exp\{\mathbf{Z}_j^T(T_i)\boldsymbol{\beta}\}\Delta T_i}\right]^{I(\varepsilon_i=1)}$$

$$= \prod_{i=1}^{n}\left[\frac{\exp\{\mathbf{Z}_i^T(T_i)\boldsymbol{\beta}\}}{\sum_{j\in R_i}\exp\{\mathbf{Z}_j^T(T_i)\boldsymbol{\beta}\}}\right]^{I(\varepsilon_i=1)}.$$

The log partial likelihood is

$$\log\{L(\boldsymbol{\beta})\} = \sum_{i=1}^{n} I(\varepsilon_i = 1) \times \left(\mathbf{Z}_i^T(T_i)\boldsymbol{\beta} - \log\left[\sum_{j\in R_i}\exp\{\mathbf{Z}_j^T(T_i)\boldsymbol{\beta}\}\right]\right).$$

It is clear that the maximizer of $\log\{L(\boldsymbol{\beta})\}$, $\hat{\boldsymbol{\beta}}$, is consistent and asymptotically normal under the standard regularity conditions. To see this, consider the score obtained by differentiating the log partial likelihood with respect to $\boldsymbol{\beta}$:

$$\mathbf{U}_1(\boldsymbol{\beta}) = \sum_{i=1}^{n} I(\varepsilon_i = 1) \times \left[\mathbf{Z}_i(T_i) - \frac{\sum_{j\in R_i}\mathbf{Z}_j(T_i)\exp\{\mathbf{Z}_j^T(T_i)\boldsymbol{\beta}\}}{\sum_{j\in R_i}\exp\{\mathbf{Z}_j^T(T_i)\boldsymbol{\beta}\}}\right].$$

We can reformulate $\mathbf{U}_1(\boldsymbol{\beta})$ in terms of counting processes. Letting $N_i(t) = I(T_i \leq t, \varepsilon_i = 1)$ and $Y_i(t) = 1 - N_i(t-)$,

$$\mathbf{U}_1(\boldsymbol{\beta}) = \sum_{i=1}^{n}\int_0^\infty \left[\mathbf{Z}_i(s) - \frac{\sum_j Y_j(s)\mathbf{Z}_j(s)\exp\{\mathbf{Z}_j^T(s)\boldsymbol{\beta}\}}{\sum_j Y_j(s)\exp\{\mathbf{Z}_j^T(s)\boldsymbol{\beta}\}}\right] \times dN_i(s). \quad (2)$$

With complete data, $N_i(t)$ has compensator $A_i(t,\boldsymbol{\beta}) = \int_0^t Y_i(u)\lambda_{10}(u)\exp\{\mathbf{Z}_i^T(u)\boldsymbol{\beta}\}\,du$ with respect to the complete data filtration, $\mathcal{F}^1(t)$, the increasing sequnce of sigma algebras, $\sigma\{N_i(u), Y_i(u)\mathbf{Z}_i(u), u \leq t, i = 1,\ldots,n\}$. It is easy to see that $M_i^1(t,\boldsymbol{\beta}_0) = N_i(t) - A_i(t,\boldsymbol{\beta}_0)$ satisfies the definition of a martingale under $\mathcal{F}^1(t)$, but will not be a martingale with respect to the natural filtration for the cause-specific hazard function.

The form of the estimating equation for the regression parameters under model (1) for the subdistribution is identical to that for the marginal or cause-specific proportional hazards models, the only difference being the definitions of the risk sets and the martingale processes. Thus $\mathbf{U}_1(\boldsymbol{\beta})$ can be rewritten as a martingale integral when evaluated at $\boldsymbol{\beta}_0$. Under mild regularity conditions, the martingale central limit theorem is applicable and the score converges weakly to a Gaussian process (Rebolledo 1978). The usual results obtain: $n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ is asymptotically normal with limiting covariance matrix, $\boldsymbol{\Omega}^{-1}$, where

$$\boldsymbol{\Omega} = \int_0^\infty \left\{\frac{\mathbf{s}^{(2)}(\boldsymbol{\beta}_0,u)}{s^{(0)}(\boldsymbol{\beta}_0,u)} - \bar{\mathbf{z}}(\boldsymbol{\beta}_0,u)^{\otimes 2}\right\} s^{(0)}(\boldsymbol{\beta}_0,u)\,d\Lambda_{10}(u),$$

$$\bar{\mathbf{z}}(\boldsymbol{\beta},u) = \frac{\mathbf{s}^{(1)}(\boldsymbol{\beta},u)}{s^{(0)}(\boldsymbol{\beta},u)},$$

$$\mathbf{s}^{(p)}(\boldsymbol{\beta},u) = \lim_{n\to\infty} n^{-1}\sum_{i=1}^{n} Y_i(u)\mathbf{Z}_i(u)^{\otimes p}\exp\{\mathbf{Z}_i^T(u)\boldsymbol{\beta}\}, \quad p = 0,1,2, \quad (3)$$

and, for a vector $\mathbf{v}$, $\mathbf{v}^{\otimes 0} = 1$, $\mathbf{v}^{\otimes 1} = \mathbf{v}$, and $\mathbf{v}^{\otimes 2} = \mathbf{v}\mathbf{v}^t$. Inferences about $\boldsymbol{\beta}$ can be made using a normal distribution with covariance matrix $\hat{\boldsymbol{\Omega}}^{-1}$, a consistent estimate for the asymptotic variance, which can be computed from

$$\hat{\boldsymbol{\Omega}} = \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\mathbf{S}_1^{(2)}(\hat{\boldsymbol{\beta}},T_i)}{S_1^{(0)}(\hat{\boldsymbol{\beta}},T_i)} - \bar{\mathbf{Z}}(\hat{\boldsymbol{\beta}},T_i)^{\otimes 2}\right\} I(\varepsilon_i = 1),$$

where

$$\bar{\mathbf{Z}}(\boldsymbol{\beta}, u) = \frac{\mathbf{S}_1^{(1)}(\boldsymbol{\beta}, u)}{S_1^{(0)}(\boldsymbol{\beta}, u)}$$

and

$$\mathbf{S}_1^{(p)}(\boldsymbol{\beta}, u) = \frac{1}{n}\sum_{i=1}^{n} Y_i(u)\mathbf{Z}_i(u)^{\otimes p}\exp\{\mathbf{Z}_i^T(u)\boldsymbol{\beta}\}, \qquad p = 0, 1, 2.$$

A variation on Breslow's estimator is

$$\hat{\Lambda}_{10}(t) = \frac{1}{n}\sum_{i=1}^{n}\int_0^t \frac{1}{S_1^{(0)}(\hat{\boldsymbol{\beta}}, u)}\, dN_i(u).$$

To predict the cumulative incidence at time t for an individual with covariates $\mathbf{Z} = \mathbf{z}_0$, we estimate $\Lambda_1(t; \mathbf{z}_0)$ with

$$\hat{\Lambda}_1(t; \mathbf{z}_0) = \frac{1}{n}\sum_{i=1}^{n}\int_0^t \frac{\exp\{\mathbf{z}_0^T(u)\hat{\boldsymbol{\beta}}\}}{S_1^{(0)}(\hat{\boldsymbol{\beta}}, u)}\, dN_i(u).$$

Analytic and simulation-based confidence intervals and bands for $F_1(t; \mathbf{z}_0)$ can be constructed with existing techniques (Hall and Wellner 1980; Lin, Fleming, and Wei 1994). Following the results of Lin et al. (1994), $n^{1/2}\{\hat{\Lambda}_1(t; \mathbf{z}_0) - \Lambda_1(t; \mathbf{z}_0)\}$ is asymptotically equivalent to

$$n^{-1/2}\sum_{i=1}^{n}\left[\int_0^t \frac{\exp\{\mathbf{z}_0^T(u)\boldsymbol{\beta}_0\}\, dM_i^1(u, \boldsymbol{\beta}_0)}{S_1^{(0)}(\boldsymbol{\beta}_0, u)} + \mathbf{h}^t(t, \mathbf{z}_0)\boldsymbol{\Omega}^{-1} \times \int_0^\infty \{\mathbf{Z}_i(u) - \bar{\mathbf{Z}}(\boldsymbol{\beta}_0, u)\}\, dM_i^1(u, \boldsymbol{\beta}_0)\right],$$

where

$$\mathbf{h}(t, \mathbf{z}_0) = \int_0^t \{\mathbf{z}_0(u) - \bar{\mathbf{z}}(\boldsymbol{\beta}_0, u)\} \times \exp\{\mathbf{z}_0^T(u)\boldsymbol{\beta}_0\}\, d\Lambda_{10}(u). \quad (4)$$

3.2 Censoring Complete Data

In smartly designed clinical trials, censoring results only from administrative loss-to-follow up; that is, patients have not failed by the time the data are analyzed. Under this condition, the potential censoring time is always observed, even on individuals who die prior to the time of analysis. We call these data *censoring complete*. We redefine the risk set for the ith individual to include the censoring information

$$R_i = \{j : (C_j \wedge T_j \geq T_i) \cup (T_j \leq T_i \cap \varepsilon_j \neq 1 \cap C_j \geq T_i)\},$$

where $i \wedge j$ denotes $\min(i, j)$. In our hypothetical cohort, an individual with $\varepsilon \neq 1$ is still "at risk" for failure from the cause of interest until time C, when $T < C$. If (T, ε) and C are conditionally independent given the covariate, then the "crude" subdistribution hazard function with censoring-complete data, $\lambda_{1*}\{t; \mathbf{Z}\}$, is equivalent to the "net" subdistribution hazard function with complete data, $\lambda_1\{t; \mathbf{Z}\}$. This result is similar to that for the "crude" and "net" hazards in the univariate survival time setting. As proof, observe that

$$\begin{aligned}
&\lambda_{1*}(t; \mathbf{Z}) \\
&= \lim_{\Delta t \to 0}\frac{1}{\Delta t}\Pr\{t \leq T \leq t + \Delta t, \varepsilon = 1 | C \wedge T \geq t \cup (T \leq t \cap \varepsilon \neq 1 \cap C \geq t), \mathbf{Z}\} \\
&= \lim_{\Delta t \to 0}\frac{\frac{1}{\Delta t}\{\Pr(t \leq T \leq t + \Delta t, \varepsilon = 1, C \geq t | \mathbf{Z})\}}{\Pr[\{T \geq t \cup (T \leq t \cap \varepsilon \neq 1)\} \cap C \geq t | \mathbf{Z}]} \\
&= \frac{d\{\Pr(T \leq t, \varepsilon = 1 | \mathbf{Z})\}\Pr(C \geq t | \mathbf{Z})}{\Pr\{(T \leq t, \varepsilon \neq 1) \cup T \geq t | \mathbf{Z}\}\Pr(C \geq t | \mathbf{Z})} \\
&= \lambda_1\{t; \mathbf{Z}\}.
\end{aligned}$$

Using the censoring-complete risk sets, the partial likelihood principle can again be applied to the multiplicative model for $\lambda_1\{t; \mathbf{Z}\}$. Differentiating the log partial likelihood with respect to $\boldsymbol{\beta}$ and reexpressing in terms of counting processes, we obtain a martingale-type estimating function,

$$\mathbf{U}_{1*}(\boldsymbol{\beta}) = \sum_{i=1}^{n}\int_0^\infty \left[\mathbf{Z}_i(s) - \frac{\sum_j Y_j^*(s)\mathbf{Z}_j(s)\exp\{\mathbf{Z}_j^T(s)\boldsymbol{\beta}\}}{\sum_j Y_j^*(s)\exp\{\mathbf{Z}_j^T(s)\boldsymbol{\beta}\}}\right] \times dM_i^{1*}(s, \boldsymbol{\beta}), \quad (5)$$

where

$$Y_i^*(t) = I(C_i \geq t)\{1 - N_i(t-)\}$$

and

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

When C_i is always observed, one can check that $M_i^{1*}(t, \boldsymbol{\beta}_0)$ satisfies the definition of a martingale under the censoring-complete filtration, $\mathcal{F}^{1*}(t)$, the increasing sequence of sigma-algebras, $\sigma\{I(C_i \geq u), I(C_i \geq u)N_i(u), Y_i^*(u), Y_i^*(u)\mathbf{Z}_i(u), u \leq t, i = 1, \ldots, n\}$. Under $\mathcal{F}^{1*}(t)$, $\mathbf{U}_{1*}(\boldsymbol{\beta}_0)$ is a sum of martingale integrals with respect to the locally bounded predictable process

$$\mathbf{H}_i^*(s) = \mathbf{Z}_i(s) - \frac{\sum_j Y_j^*(s)\mathbf{Z}_j(s)\exp\{\mathbf{Z}_j^T(s)\boldsymbol{\beta}_0\}}{\sum_j Y_j^*(s)\exp\{\mathbf{Z}_j^T(s)\boldsymbol{\beta}_0\}}.$$

The asymptotic results for censoring-complete data estimation and prediction follow from the complete data derivations, which are inherited from the ordinary Cox model. In the following section we discuss extensions of this methodology with censored data for which either T_i or C_i (not both) can be observed.

4. A WEIGHTED SCORE FUNCTION FOR INCOMPLETE DATA

When the usual right censoring is present, we can adapt inverse probability of censoring weighting (IPCW) techniques (Robins and Rotnitzky 1992) to construct an unbiased estimating function from the score of the complete-data partial likelihood. In the sequel, we assume for simplicity that C is independent of T, ε, and $\mathbf{Z}$ and that the censoring random variables are independently distributed with $\Pr(C \geq t) = G(t)$. However, we note that IPCW can be generalized to allow for dependence between C and $\mathbf{Z}$. For instance, when censoring depends on a subset of discrete covariates, we can nonparametrically estimate the necessary weights separately for each covariate configuration. With continuous covariates, we might assume that a proportional hazards model holds for the conditional distribution of C given $\mathbf{Z}$.

Knowledge of vital status on individual i at time t is denoted by $r_i(t) = I(C_i \geq T_i \wedge t)$. If $r_i(t) = 1$, then $Y_i(t)$ and $N_i(t)$ are computable in terms of the observed data up to time t (i.e., not missing). Censored individuals are observed until time C_i; thereafter, vital status is uncertain. Individuals observed to fail remain in the risk set indefinitely, as long as they have not failed from cause 1. Although $Y_i(t)$ and $N_i(t)$ are not observable when $r_i(t) = 0$, $r_i(t)Y_i(t)$ and $r_i(t)N_i(t)$ are computable for $r_i(t) = 0, 1$. The quantity $r_i(t)/G(X_i \wedge t)$ has expectation 1 conditional on T_i, ε_i, and $\mathbf{Z}_i$. We can multiply this term by any other deterministic function of time, $D(t)$, with the expectation of the product being $D(t)$. At time t, we associate the time-dependent weight $w_i(t) = r_i(t)\hat{G}(t)/\hat{G}(X_i \wedge t)$ with individual i, where $\hat{G}(\cdot)$ is the Kaplan–Meier estimate of the survival function of the censoring random variable calculated using $\{X_i, 1 - \Delta_i, i = 1, \ldots, n\}$. This forms the basis for IPCW, which is applied to (2):

$$\mathbf{U}_2(\boldsymbol{\beta}) = \sum_{i=1}^{n} \int_0^{\infty} \left\{ \mathbf{Z}_i(s) - \frac{\sum_j w_j(s) Y_j(s) \mathbf{Z}_j(s) \exp\{\mathbf{Z}_j^T(s)\boldsymbol{\beta}\}}{\sum_j w_j(s) Y_j(s) \exp\{\mathbf{Z}_j^T(s)\boldsymbol{\beta}\}} \right\} \times w_i(s)\, dN_i(s). \quad (6)$$

If there is only a single cause of failure, then $\mathbf{U}_2(\boldsymbol{\beta})$ reduces to the typical score function for the Cox model. This happens because $\Delta_i = 1$ implies $\varepsilon_i = 1$, so that $w_i(t)Y_i(t) = I(C_i \geq T_i \wedge t, T_i \geq t)\hat{G}(t)\{\hat{G}(X_i \wedge t)\}^{-1} = I(X_i \geq t)\hat{G}(t)\{\hat{G}(X_i \wedge t)\}^{-1} = I(X_i \geq t)$ and $w_i(t)dN_i(t) = I(C_i \geq T_i \wedge t)d\{I(T_i \leq t)\} = d\{I(T_i \leq t, \Delta_i = 1)\}$ are just the risk set indicator and counting process for individual i in the univariate setting.

In Appendix A we show that $\hat{\boldsymbol{\beta}}$, the solution to $\mathbf{U}_2(\boldsymbol{\beta}) = 0$, is consistent for $\boldsymbol{\beta}_0$. Taking a Taylor series of $\mathbf{U}_2(\hat{\boldsymbol{\beta}})$ around $\boldsymbol{\beta}_0$, a first order approximation holds:

$$n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \approx \boldsymbol{\Omega}^{-1}\{n^{-1/2}\mathbf{U}_2(\boldsymbol{\beta}_0)\},$$

where $\boldsymbol{\Omega}^{-1}$ is the limit of the negative of the inverse of the partial derivative matrix of the score function evaluated at $\boldsymbol{\beta}_0$ and is equal to the variance of the censoring-complete regression coefficients or, equivalently, the limit of $n^{-1}d\{\mathbf{U}_{1*}(\boldsymbol{\beta})\}/d\boldsymbol{\beta}$ at $\boldsymbol{\beta}_0$. This follows from the fact that with censoring-complete data, the implicit weight $I(C_i \geq t)$ has the same conditional expectation as $r_i(t)G(t)/G(X_i \wedge t)$. With right-censored data, a consistent estimate for $\boldsymbol{\Omega}$ is

$$\hat{\boldsymbol{\Omega}} = \frac{1}{n}\sum_{i=1}^{n} \left\{ \frac{\hat{\mathbf{S}}_2^{(2)}(\hat{\boldsymbol{\beta}}, X_i)}{\hat{S}_2^{(0)}(\hat{\boldsymbol{\beta}}, X_i)} - \bar{\mathbf{Z}}(\hat{\boldsymbol{\beta}}, X_i)^{\otimes 2} \right\} \Delta_i I(\varepsilon_i = 1),$$

$$\bar{\mathbf{Z}}(\boldsymbol{\beta}, u) = \frac{\hat{\mathbf{S}}_2^{(1)}(\boldsymbol{\beta}, u)}{\hat{S}_2^{(0)}(\boldsymbol{\beta}, u)},$$

and

$$\hat{\mathbf{S}}_2^{(p)}(\boldsymbol{\beta}, u) = \frac{1}{n}\sum_{i=1}^{n} w_i(u) Y_i(u) \mathbf{Z}_i(u)^{\otimes p} \exp\{\mathbf{Z}_i^T(u)\boldsymbol{\beta}\}, \quad p = 0, 1, 2. \quad (7)$$

We also show in Appendix A that $n^{-1/2}\mathbf{U}_2(\boldsymbol{\beta}_0)$ can be expressed as the sum of n independent and identically distributed random variables and has a normal limiting distribution with variance–covariance matrix, $\boldsymbol{\Sigma}$, which can be consistently estimated with the empirical covariance matrix

$$\hat{\boldsymbol{\Sigma}} = n^{-1}\sum_{i=1}^{n} (\hat{\boldsymbol{\eta}}_i + \hat{\boldsymbol{\psi}}_i)^{\otimes 2},$$

where

$$\hat{\boldsymbol{\eta}}_i = \int_0^{\infty} \left\{ \mathbf{Z}_i(u) - \frac{\hat{\mathbf{S}}_2^{(1)}(\hat{\boldsymbol{\beta}}, u)}{\hat{S}_2^{(0)}(\hat{\boldsymbol{\beta}}, u)} \right\} w_i(u)\, d\hat{M}_i^1(u),$$

$$\hat{\boldsymbol{\psi}}_i = \int_0^{\infty} \frac{\hat{\mathbf{q}}(u)}{\hat{\pi}(u)}\, d\hat{M}_i^c(u),$$

$$\hat{\mathbf{q}}(u) = -n^{-1}\sum_{i=1}^{n} \int_0^{\infty} \left\{ \mathbf{Z}_i(s) - \frac{\hat{\mathbf{S}}_2^{(1)}(\hat{\boldsymbol{\beta}}, s)}{\hat{S}_2^{(0)}(\hat{\boldsymbol{\beta}}, s)} \right\} \times w_i(s)\, d\hat{M}_i^1(s) I(s \geq u > X_i),$$

$$\hat{\pi}(u) = n^{-1}\sum_{i=1}^{n} I(X_i \geq u),$$

$$\hat{M}_i^1(t) = I(T_i \leq t, \varepsilon_i = 1) - \int_0^t \{1 - I(T_i < s, \varepsilon_i = 1)\} \times \exp\{\mathbf{Z}_i^T(s)\hat{\boldsymbol{\beta}}\} d\hat{\Lambda}_{10}(s),$$

$$\hat{M}_i^c(t) = I(X_i \leq t, \Delta_i = 0) - \int_0^t I(X_i \geq u)\, d\hat{\Lambda}^c(u),$$

$$\hat{\Lambda}_{10}(t) = \frac{1}{n}\sum_{i=1}^{n} \int_0^t \frac{1}{\hat{S}_2^{(0)}(\hat{\boldsymbol{\beta}}, u)} w_i(u)\, dN_i(u),$$

and

$$\hat{\Lambda}^c(t) = \int_0^t \frac{\sum_i d\{I(X_i \leq u, \Delta_i = 0)\}}{\sum_i I(X_i \geq u)}. \quad (8)$$

The estimator for the cumulative hazard of the censoring time, $\hat{\Lambda}^c(\cdot)$, is the commonly used Nelson–Aalen estimator (Nelson 1972). The baseline hazard for the subdistribution is estimated using a modified version of Breslow's estimator, $\hat{\Lambda}_{10}(\cdot)$, and converges uniformly in probability to the true baseline hazard of the subdistribution on the interval $[0, \tau)$, where τ is chosen such that $\Pr(X \geq \tau) > 0$. In the absence of covariates, it is easy to prove that $1-\exp\{-\hat{\Lambda}_{10}(\cdot)\}$ is asymptotically equivalent to the nonparametric maximum likelihood estimator for $F_1(\cdot)$. In Section 5 we show how to use $\hat{\boldsymbol{\beta}}$ and $\hat{\Lambda}_{10}(\cdot)$ to make inferences about the predicted cumulative incidence function for an individual with arbitrary covariates.

The distribution of $n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ can be approximated by a normal distribution with variance $\hat{\boldsymbol{\Omega}}^{-1}\hat{\boldsymbol{\Sigma}}\hat{\boldsymbol{\Omega}}^{-1}$. Inferences about covariate effects on the cumulative incidence function can be based on this limiting result. Tests for covariate effects may also be based on score-type tests. In the case of complete or censoring-complete data, score and partial likelihood ratio tests that are asymptotically equivalent to the corresponding Wald tests are possible. Note that with complete data, the score test for H_0: $\boldsymbol{\beta} = 0$ in the two-sample problem is identical to Gray's test statistic with weight $K(t) \propto \hat{F}_{10}(t-)\{\hat{F}_{11}(t-)\hat{F}_{12}(t-)\}^{-1}$, where $\hat{F}_{10}(t)$ is the pooled nonparametric estimate of $F_1(t)$ and $\hat{F}_{1i}(t), i = 1, 2$, are nonparametric estimates of $F_1(t)$ in group i.

5. PREDICTING CUMULATIVE INCIDENCE WITH INCOMPLETE DATA

To predict cumulative incidence at time t for an individual with covariates $\mathbf{Z} = \mathbf{z}_0$, we estimate the cumulative subdistribution hazard by $\hat{\Lambda}_1(t; \mathbf{z}_0) = \int_0^t \exp\{\mathbf{z}_0^T(u)\hat{\boldsymbol{\beta}}\}\, d\hat{\Lambda}_{10}(u)$, where $\hat{\boldsymbol{\beta}}$ and $\hat{\Lambda}_{10}(t)$ can be found in Section 4. The predicted cumulative incidence is then $\hat{F}_1(t; \mathbf{z}_0) = 1 - \exp\{-\hat{\Lambda}_1(t; \mathbf{z}_0)\}$. The consistency of $\hat{\boldsymbol{\beta}}$ for $\boldsymbol{\beta}_0$ and uniform convergence in probability of $\hat{\Lambda}_{10}(t)$ to $\Lambda_{10}(t)$ ensure the uniform convergence of $\hat{F}_1(t; \mathbf{z}_0)$ to $F_1(t; \mathbf{z}_0)$ on a suitably chosen interval. In Appendix B we demonstrate that $n^{1/2}\{\hat{\Lambda}_1(t; \mathbf{z}_0) - \Lambda_1(t; \mathbf{z}_0)\}$ converges weakly to a Gaussian process on $[0, \tau)$, where $\Pr(X \geq \tau) > 0$. Combining this result with the functional delta method, we conclude that for a known, monotone, absolutely continuous transformation $m(\cdot)$, $n^{1/2}(m[\hat{F}_1\{t; \mathbf{z}_0\}] - m[F_1\{t; \mathbf{z}_0\}])$ converges weakly to a Gaussian process that has the same limiting distribution as

$$
\begin{aligned}
&J_1\{t; \mathbf{z}_0\} \\
&\quad = \dot{m}[F_1\{t; \mathbf{z}_0\}] \exp[-\Lambda_1\{t; \mathbf{z}_0\}] \\
&\qquad \times \Bigg(n^{-1/2} \sum_{i=1}^{n} \int_0^t \frac{\exp\{\mathbf{z}_0^T(u)\boldsymbol{\beta}_0\}}{s^{(0)}(\boldsymbol{\beta}_0, u)} \frac{r_i(u)G(u)}{G(X_i \wedge u)}\, dM_i^1(u, \boldsymbol{\beta}_0) \\
&\qquad\quad + \mathbf{h}^t\{t, \mathbf{z}_0\}\boldsymbol{\Omega}^{-1} n^{-1/2} \\
&\qquad\quad \times \sum_{i=1}^{n} \Bigg[\int_0^\infty \left\{ \mathbf{Z}_i(u) - \frac{\mathbf{s}^{(1)}(\boldsymbol{\beta}_0, u)}{s^{(0)}(\boldsymbol{\beta}_0, u)} \right\} \\
&\qquad\qquad \times \frac{r_i(u)G(u)}{G(X_i \wedge u)}\, dM_i^1(u, \boldsymbol{\beta}_0) + \int_0^\infty \frac{\mathbf{q}(u)}{\pi(u)}\, dM_i^c(u) \Bigg] \\
&\qquad\quad + n^{-1/2} \sum_{i=1}^{n} \int_0^\infty \frac{v\{u, t, \mathbf{z}_0\}}{\pi(u)}\, dM_i^c(u) \Bigg) + o_p(1),
\end{aligned}
$$

where $\dot{m}(s) = d\{m(s)\}(ds)^{-1}$, $\mathbf{s}^{(i)}, i = 0, 1, 2$, and $\boldsymbol{\Omega}$ are defined in (3), $\mathbf{h}\{t, \mathbf{z}\}$ is given in (4), $\pi(u)$ and $\mathbf{q}(u)$ are defined in (A.3), and

$$
\begin{aligned}
v\{u, t, \mathbf{z}\} = &- \lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} \int_0^t \frac{\exp\{\mathbf{z}^T(s)\boldsymbol{\beta}_0\}}{s^{(0)}(\boldsymbol{\beta}_0, s)} \\
&\times \frac{r_i(s)G(s)}{G(X_i \wedge s)}\, dM_i^1(s, \boldsymbol{\beta}_0) I(s \geq u > X_i). \quad (9)
\end{aligned}
$$

The transformation m is chosen to stabilize the variance and to ensure that pointwise and simultaneous confidence intervals for the probability $F_1\{t; z_0\}$ are bounded between 0 and 1. The limiting process J_1 is quite complicated and, other than pointwise asymptotic normality, its properties are difficult to obtain analytically. To construct confidence intervals and bands for the predicted cumulative incidence function, we approximate the distribution of $J_1\{t; \mathbf{z}_0\}$ by a mean 0 Gaussian process, $\hat{J}_1\{t; \mathbf{z}_0\}$, whose distribution can be generated through simulation. To compute $\hat{J}_1\{t; \mathbf{z}_0\}$, we replace $M_i^1(u, \boldsymbol{\beta}_0)$ and $M_i^c(u)$ by $\hat{M}_i^1(u)\mathbf{A}_i$ and $\hat{M}_i^c(u)\mathbf{A}_i$, where $\{\mathbf{A}_i\}$ is a random sample from the standard normal distribution and $\hat{M}_i^1(u)$ and $\hat{M}_i^c(u)$ can be found in (8). We also replace other theoretical quantities in $J_1\{t; \mathbf{z}_0\}$ with their observed values, yielding

$$
\begin{aligned}
&\hat{J}_1\{t; \mathbf{z}_0\} \\
&\quad = \dot{m}[\hat{F}_1\{t; \mathbf{z}_0\}] \exp[-\hat{\Lambda}_1\{t; \mathbf{z}_0\}] \\
&\qquad \times \Bigg(n^{-1/2} \sum_{i=1}^{n} \int_0^t \frac{\exp\{\mathbf{z}_0^T(u)\hat{\boldsymbol{\beta}}\}}{\hat{S}_2^{(0)}(\hat{\boldsymbol{\beta}}, u)}\, w_i(u)\, d\hat{M}_i^1(u)\mathbf{A}_i \\
&\qquad\quad + \hat{\mathbf{h}}^t\{t, \mathbf{z}_0\}\hat{\boldsymbol{\Omega}}^{-1} n^{-1/2} \\
&\qquad\quad \times \sum_{i=1}^{n} \Bigg[\int_0^\infty \left\{ \mathbf{Z}_i(u) - \frac{\hat{\mathbf{S}}_2^{(1)}(\hat{\boldsymbol{\beta}}, u)}{\hat{S}_2^{(0)}(\hat{\boldsymbol{\beta}}, u)} \right\} \\
&\qquad\qquad \times w_i(u)\, d\hat{M}_i^1(u)\mathbf{A}_i \\
&\qquad\qquad + \int_0^\infty \frac{\hat{\mathbf{q}}(u)}{\hat{\pi}(u)}\, d\hat{M}_i^c(u)\mathbf{A}_i \Bigg] \\
&\qquad\quad + n^{-1/2} \sum_{i=1}^{n} \int_0^\infty \frac{\hat{v}\{u, t, \mathbf{z}_0\}}{\hat{\pi}(u)}\, d\hat{M}_i^c(u)\mathbf{A}_i \Bigg).
\end{aligned}
$$

Note that conditional on the observed data, the only random components in $\hat{J}_1\{t; \mathbf{z}_0\}$ are $\{\mathbf{A}_i, i = 1, \ldots, n\}$. A proof similar to that given by Cheng et al. (1997) or Lin et al. (1994) permits the claim that conditional on the observed data, $\hat{J}_1\{t; \mathbf{z}_0\}$ converges weakly to the unconditional limiting distribution of $J_1\{t; \mathbf{z}_0\}$. Using the consistency of the empirical quantities $\hat{\boldsymbol{\beta}}, \hat{G}, \hat{\mathbf{h}}, \hat{\boldsymbol{\Omega}}, \hat{v}, \hat{\Lambda}_{10}$, and $\hat{\Lambda}_c$ and the

known distribution of $\{\mathbf{A}_i, i = 1, \cdots, n\}$, it is straightforward to show that as $n \to \infty$, the covariance function of $\hat{J}_1$ converges almost surely to that of J_1. For any finite set of time points $(t_1, \cdots, t_m)$, the conditional limiting distribution of $\{\hat{J}_1(t_1; \cdot), \cdots, \hat{J}_1(t_m; \cdot)\}^t$ is the same as the unconditional limiting distribution of $\{J_1(t_1; \cdot), \cdots, J_1(t_m; \cdot)\}^t$. Because the only "random" components in $\hat{J}_1$ are the normal variates, which are time independent, simple moment inequalities can be applied to give the tightness of $\hat{J}_1$ as a process in t.

A procedure for calculating $(1 - 2\alpha_1)$ pointwise confidence intervals and a $(1 - 2\alpha_2)$ confidence band for $F_1\{t; \mathbf{z}_0\}$ is as follows. Generate B samples $\{\mathbf{A}_i, i = 1, \ldots, n\}$ and for each realization, $k = 1, \ldots, B$, compute $\hat{J}_{1k}\{t; \mathbf{z}_0\}$. Next, use the simulated distribution of $\hat{J}_1\{t; \mathbf{z}_0\}$ to compute an estimate of the variance function, $\sigma\{t; \mathbf{z}_0\} = \text{var}\{n^{1/2}(m[\hat{F}_1\{t; \mathbf{z}_0\}] - m[F_1\{t; \mathbf{z}_0\}])\}$:

$$\hat{\sigma}\{t; \mathbf{z}_0\} = \frac{1}{B} \sum_{k=1}^{B} \hat{J}_{1k}^2\{t; \mathbf{z}_0\}.$$

Pointwise confidence intervals for $F_1\{t; \mathbf{z}_0\}$ can be constructed using the large-sample properties of $J_1\{t; \mathbf{z}_0\}$. For example, a $(1 - 2\alpha_1)$ confidence interval for $F_1\{t; \mathbf{z}_0\}$ is

$$m^{-1}(m[\hat{F}_1\{t; \mathbf{z}_0\}] \pm n^{-1/2} \phi_{\alpha_1} \hat{\sigma}^{1/2}\{t; \mathbf{z}_0\}),$$

where ϕ_α is the 100α percentile of the standard normal distribution. For $(1 - 2\alpha_2)$ simultaneous confidence intervals, use the simulated distribution of $\hat{J}_1\{t; \mathbf{z}_0\}$ and the variance estimates, $\hat{\sigma}\{t; \mathbf{z}_0\}$, to find $C\{\alpha_2, \mathbf{z}_0\}$ such that

$$\Pr\left[\sup_{t \in [0,\tau]} \frac{|\hat{J}_1\{t; \mathbf{z}_0\}|}{\hat{\sigma}^{1/2}\{t; \mathbf{z}_0\}} \le C\{\alpha_2, \mathbf{z}_0\}\right] = 1 - 2\alpha_2.$$

A $(1 - 2\alpha_2)$ confidence band for $F_1\{t; \mathbf{z}_0\}$ is

$$m^{-1}(m[\hat{F}_1\{t; \mathbf{z}_0\}] \pm n^{-1/2} C\{\alpha_2, \mathbf{z}_0\} \hat{\sigma}^{1/2}\{t; \mathbf{z}_0\}).$$

6. SIMULATION EXPERIMENTS

In this section we present the results of two numerical investigations. In the first study, fairly extensive simulations were used to compare the estimators from the weighted score function (6) to the censoring-complete estimators, described in Section 3.2, which require knowledge of the potential censoring times for subjects observed to fail with $\varepsilon_i \neq 1$. Next, in a smaller experiment, the new procedures were compared to Gray's K-sample test (1988) in the two-sample setting.

In both sets of simulations, data were generated from the following model. There were two covariates $\mathbf{Z}_i = (Z_{i1}, Z_{i2})$, with new covariate values generated for each sample. The subdistributions for type 1 failures were given by

$$\Pr(T_i \le t, \varepsilon_i = 1 | \mathbf{Z}_i) = 1 - [1 - p\{1 - \exp(-t)\}]^{\exp(Z_{i1}\beta_{11} + Z_{i2}\beta_{12})},$$

which is a unit exponential mixture with mass $1 - p$ at ∞ when $\mathbf{Z}_i = (0, 0)$, and uses the proportional subdistribution hazards model to obtain the subdistribution for nonzero covariate values. The subdistribution for type 2 failures was then obtained by taking $\Pr(\varepsilon_i = 2 | \mathbf{Z}_i) = 1 - \Pr(\varepsilon_i = 1 | \mathbf{Z}_i)$ and using an exponential distribution with rate $\exp(Z_{i1}\beta_{21} + Z_{i2}\beta_{22})$ for $\Pr(T_i \le t | \varepsilon_i = 2, \mathbf{Z}_i)$. Censoring times were generated from the uniform $[a, b]$ distribution.

Both the censoring-complete and the weighted estimating equations are derivatives of well-behaved objective functions. The solutions were calculated by using a modified Newton algorithm to maximize the objective functions. Variance estimators were also calculated for each sample. For the censoring-complete estimator, the variance estimator was the inverse of the matrix of derivatives of the estimating equations; for the weighted estimating equation, the estimator described in Section 4 was used.

We first assumed that the covariates were independent and identically distributed standard normal variates and that the true parameter values were $(p, \beta_{11}, \beta_{12}, \beta_{21}, \beta_{22}) = (.3, .5, .5, -.5, .5)$. We used four different degrees of censoring, with 1,000 samples generated for each case. The parameter values gave 33% type 1 failures and 67% type 2 failures, in the absence of censoring. Table 1 gives for each estimator $\text{E}(\hat{\beta})$, estimated with the average of $\hat{\beta}$ from the 1,000 samples; $\text{var}(\hat{\beta})$, estimated with the empirical variance of $\hat{\beta}$ from the 1,000 samples; and $\text{E}(\hat{\text{var}})$, the average of the 1,000 variance estimators. Standard errors for $\text{E}(\hat{\text{var}})$ were substantially smaller than for $\text{var}(\hat{\beta})$ and are omitted.

Next, we generated the covariates as Bernoulli(.5) variates. The true parameter values were $(p, \beta_{11}, \beta_{12}, \beta_{21}, \beta_{22}) = (.6, 1, -1, 1, 1)$, and the same levels of censoring were

Table 1. Comparison of the Censoring Complete (CC) and Weighted (W) Estimating Equation Estimators

[a, b]	Percent censored	Eq.	E($\hat{\beta}$) β_{11}	E($\hat{\beta}$) β_{12}	Var($\hat{\beta}$) β_{11}	Var($\hat{\beta}$) β_{12}	E($\hat{\text{var}}$) β_{11}	E($\hat{\text{var}}$) β_{12}
[∞, ∞]	0	W	.507(.004)	.510(.004)	.017(.0009)	.017(.0008)	.017	.016
		CC	.507(.004)	.510(.004)	.017(.0009)	.017(.0008)	.017	.017
[1, 2]	25	W	.509(.005)	.507(.005)	.021(.001)	.022(.001)	.021	.021
		CC	.509(.005)	.507(.005)	.021(.001)	.022(.001)	.021	.021
[.5, 1]	46	W	.507(.006)	.508(.005)	.032(.002)	.030(.001)	.029	.029
		CC	.506(.006)	.509(.005)	.032(.002)	.030(.001)	.030	.030
[0, .77]	68	W	.518(.007)	.512(.007)	.055(.003)	.054(.002)	.052	.052
		CC	.518(.007)	.514(.007)	.056(.003)	.055(.002)	.054	.054

NOTE: Normally distributed covariates and $(p, \beta_{11}, \beta_{12}, \beta_{21}, \beta_{22}) = (.3, .5, .5, -.5, .5)$. Sample size is 200. Standard errors are given in parentheses.

Table 2. Comparison of the Censoring Complete (CC) and Weighted (W) Estimating Equation Estimators

[a, b]	Percent censored	Eq.	$E(\hat{\beta})$		$Var(\hat{\beta})$		$E(\hat{var})$	
			β_{11}	β_{12}	β_{11}	β_{12}	β_{11}	β_{12}
$[\infty, \infty]$	0	W	1.010(.006)	−1.007(.006)	.040(.002)	.039(.002)	.037	.038
		CC	1.010(.006)	−1.007(.006)	.040(.002)	.039(.002)	.038	.038
[.5, 1.7]	23	W	1.005(.008)	−1.014(.008)	.056(.002)	.057(.003)	.055	.055
		CC	1.006(.008)	−1.016(.008)	.057(.003)	.058(.003)	.057	.057
[0, 1.1]	47	W	1.024(.010)	−1.021(.010)	.100(.005)	.094(.005)	.091	.090
		CC	1.025(.010)	−1.018(.010)	.102(.005)	.096(.005)	.094	.093
[0, 0.4]	71	W	1.048(.015)	−1.054(.015)	.24(.01)	.23(.01)	.22	.22
		CC	1.049(.015)	−1.053(.015)	.24(.01)	.23(.01)	.22	.22

NOTE: Bernoulli distributed covariates and $(p, \beta_{11}, \beta_{12}, \beta_{21}, \beta_{22}) = (.6, 1, -1, 1, 1)$. Sample size is 200. Standard errors are given in parentheses.

used with 1,000 replicates for each scenario. Under the assumed model, the parameters gave 60% type 1 failures in the absence of censoring. In this case, convergence was not obtained for two samples, which were excluded from the results displayed in Table 2.

In both sets of simulations, the performance of the censoring-complete and the weighted estimating equations are nearly identical. Overall, both estimators exhibit a slight small-sample bias toward values larger in magnitude, especially for heavily censored samples. This is to be expected; technically, the moments are not finite in finite samples, so the results should be interpreted as conditional on the set of samples in which the estimators are finite. The variance estimators perform well in all cases.

We performed additional simulations to compare the two-group score tests using the censoring-complete and the weighted scores to the test proposed by Gray (1988). Data were generated as before, with a single binary covariate z_{i1} and with $p = .5$, $\beta_{21} = 1$, and β_{11} and the censoring as specified in Table 3. The results in Table 3 indicate that the performance of the three tests is almost indistinguishable. Not only was the average performance in agreement, but the correlations of the test statistics were $> .97$ for all cases. There is a close connection between the weighted score $U_2(0)$ and the score used by Gray (1988). With two groups, if $\hat{G}$ using the pooled data is replaced by the Kaplan–Meier estimates computed separately within each group, then, at least with no ties, and if left- and right-continuous estimates are used in the proper places, the two scores become identical.

7. A REAL DATA EXAMPLE

There were 167 eligible patients in the E1178 clinical trial. The number of positive nodes ranged from 1 to 34, with a median of 3; tumor size ranged from .3 cm to 17 cm, with a median of 2.5 cm; and age ranged from 65 to 84, with a median of 71. Of the 82 patients on placebo, 59 had recurrence of breast cancer and 19 died without recurrence; of the 85 on tamoxifen, 42 had recurrence of breast cancer and 23 died without recurrence. Median follow-up on those still disease-free and alive was 13.8 years.

Table 3. Rejection Probabilities of the Censoring Complete (CC) and Weighted (W) Score Tests and of the Test of Gray (1988) (G) for a Single Bernoulli Covariate

[a, b]	Percent censored	β_{11}	Test		
			CC	W	G
[1, 2]	18	0	.044(.005)	.044(.005)	.045(.005)
[0, 1]	56	0	.052(.005)	.048(.005)	.048(.005)
[1, 2]	18	.5	.66(.01)	.66(.01)	.65(.01)
[0, 1]	57	.75	.76(.01)	.75(.01)	.75(.01)

NOTE: Estimated from 2,000 samples of size 200. Standard errors are given in parentheses.

We begin our analysis by considering the effect of treatment while ignoring other covariates. Figure 1 gives nonparametric estimates of the cumulative incidence of breast cancer recurrence as described by Kalbfleisch and Prentice (1980, pp. 169) (see also Aalen 1978). Figure 1 also gives the estimated cumulative incidence curves (described in sec. 5) from the proportional hazards model (1), using a binary covariate Z for treatment. The proportional hazards estimate shows some indication of lack of fit, with the curves separating more gradually than for the nonparametric estimate. Because there is no intrinsic reason to expect proportional subdistribution hazards, a general diagnostic method to examine the proportional hazards assumption is helpful. Following Schoenfeld (1982), we define residuals u_j to be the contribution to $U_2(\beta)$ in (7) at the jth type 1 failure time t_j. For each covariate, plotting u_j versus t_j gives a method to check whether the model fits across time. The residuals should locally have mean 0 across time, and patterns other than a constant local average indicate lack of fit. Figure 2 plots these residuals for the proportional subdistribution hazards model, and indicates substantial lack of fit.

Based on the proportional hazards results, we then generalized the model to $\lambda_1(t; Z) = \lambda_{10}(t)\exp(\beta_0 Z + \beta_1 Zt + \beta_2 Zt^2)$, allowing the hazard ratio to be quadratic in time. The residual plot from this model is also given in Figure 2, and the estimated cumulative incidence is given in Figure 1. Both indicate that the quadratic in time fits the data well. Comparing the two models, the test that the linear and quadratic terms are both 0 has p value .03.

We next added the covariates log number of positive nodes, tumor size, and age at entry to the model. Each was added as a linear proportional hazards term. Residual plots for these variables indicated that the proportional hazards assumption was adequate, and that the quadratic in time continued to fit well for the treatment variable. Adding quadratic terms in the covariates also did not indicate significant departures from linearity in the covariate

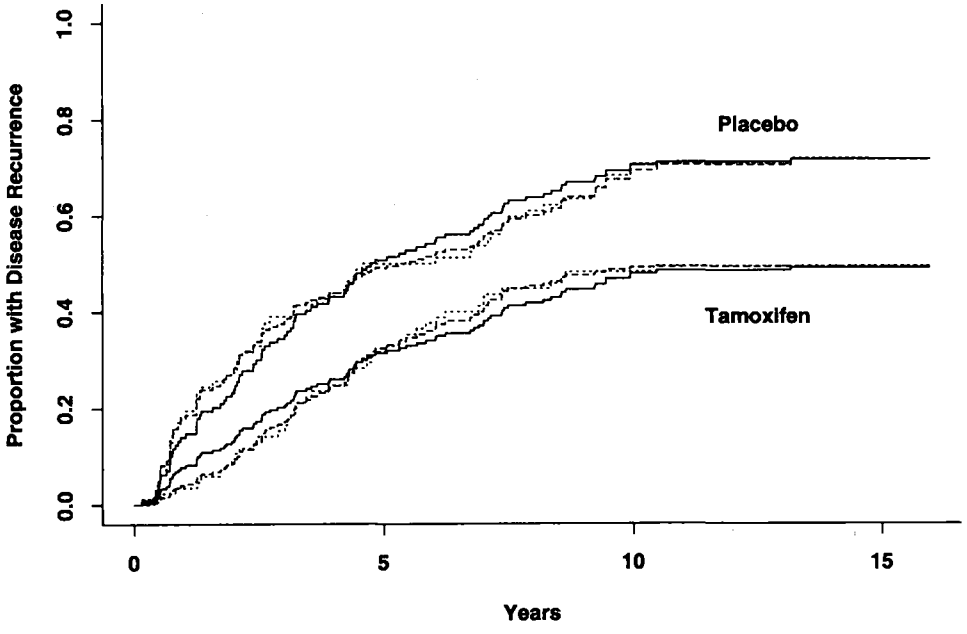

Figure 1. Predicted Cumulative Incidence of Breast Cancer Recurrence. · · ·, Nonparametric; ——, proportional hazards; – – –, quadratic in time.

effects. The estimated coefficients and standard errors from this model are given in Table 4. All variables except age are significant. In a similar analysis of deaths without recurrence, only age was found to be significant, with older patients having higher probabilities of dying without recurrence.

Using the model from Table 4, we can estimate the probability of developing breast cancer recurrence over time for patients with different characteristics. A woman diagnosed at age 80 with 1 node positive and a 2 cm tumor has an estimated probability of .27 of breast cancer recurrence within 10 years if treated with tamoxifen, considerably better than the estimate of .43 for placebo. With the same characteristics, the probability of dying prior to recurrence within 10 years is estimated to be about .44. Thus for these relatively good-risk patients treated with tamoxifen, breast cancer may no longer be the primary health concern. On the other hand, for a woman age 80 at diagnosis with 10 nodes positive and an 8 cm tumor, the probability of developing a breast cancer recurrence within 10 years is estimated to be .69 when treated with tamoxifen, whereas the estimated 10-year probability of death without recurrence is .20, and so for these patients breast cancer recurrence remains a major concern.

Estimates of this type can also be calculated by combining regression analyses of the cause-specific hazards. The results for the cause-specific hazard analysis of breast cancer recurrence are very similar to those for the subdistribution hazards analysis. Schoenfeld residuals from a proportional hazards fit for breast cancer recurrence show a similar pattern of nonproportionality in the treatment effect to that seen in Figure 2. The parameter estimates for the recurrence model with a quadratic in time for the treatment effect are given in Table 4. Except for age, the relative differences are very small, and age is not significant in either analysis. As might be expected, age is the only variable that has a significant effect on the cause-specific hazard for death without recurrence (coefficient .106, standard error .035). Residual plots based on age do not suggest nonproportionality for nonrecurrence death, and additional plots based on the nonsignificant covariates do not indicate nonlinear covariate effects. Because those covariates that were significant for breast cancer recurrence do not have much effect on the cause-specific hazard for death without recurrence, it is not surprising that the cause-specific hazard and the subdistribution analyses for breast cancer recurrence give similar results, as large differences in the competing causes generally are needed to create substantial differences in these two analyses. It is worth noting that although the results of the analysis based on the cause-specific hazard for breast cancer recurrence appear to agree with those for the $\log\{-\log(\ )\}$ transformation model for the cumulative incidence function, it is not possible to formally test for covariate effects on the subdistribution in the cause-specific hazard analysis.

Combining the cause-specific hazard estimates for the model given in Table 4 for breast cancer recurrence and the model for nonrecurrence death including age only, the estimated cumulative 10-year probability of breast cancer recurrence for an 80-year-old woman treated with tamoxifen is .28 with 1 node positive and a 2 cm tumor and .72 with 10 nodes positive and an 8 cm tumor. The estimated 10-year cumulative probabilities of dying without recurrence are .42 and .23 for these same groups. Again, these values are very similar to the estimates obtained from the direct analysis of the subdistribution functions. The main advantage of the subdistribution methodology is that through simple testing and model selection procedures, we can see the direct effect of each covariate on the cumulative incidence curves.

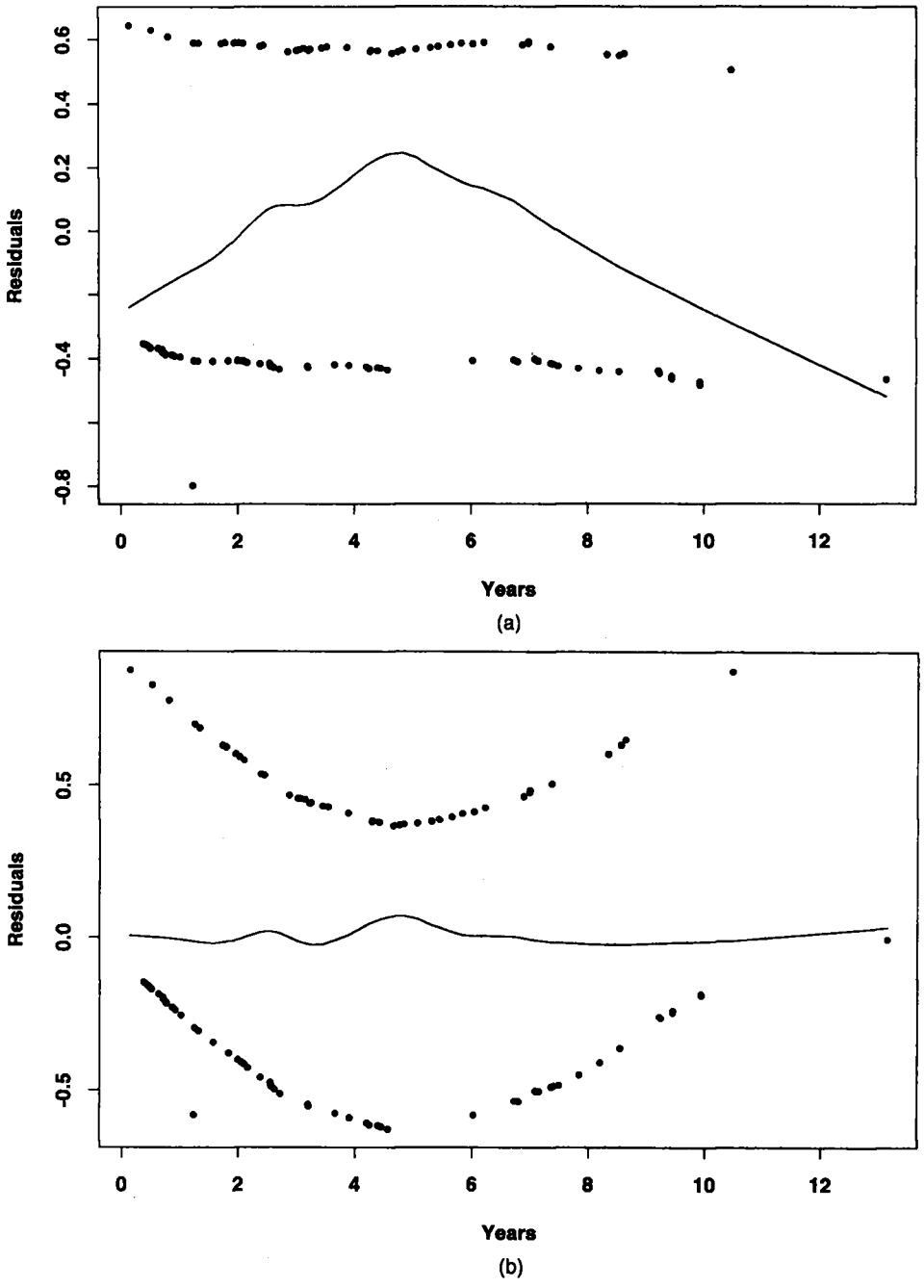

Figure 2. Schoenfeld-Type Residual Plots for Proportional Subdistribution Hazards Model. (a) Treatment effect only; (b) treatment effect with linear and quadratic time interactions. In both plots, the solid line is a locally weighted regression smooth with span = .5.

8. CONCLUDING REMARKS

Methods for estimation and prediction under a new semiparametric log(−log) transformation model for the subdistribution have been presented. The partial likelihood principle and a familiar weighting technique yielded valid procedures under a variety of censoring scenarios, and martingale techniques were used to derive the asymptotic properties of the procedures. Simulation studies indicated that under the assumed model, the weighted estimating equation with right-censored data can be as efficient as the censoring-complete score function based on the partial likelihood principle. In the two-sample setting, a simple modification of the weighted score function yields a test that is identical to Gray's K-sample statistic based on the nonparametric maximum likelihood estimator of the subdistribution hazard. An analysis of a breast cancer dataset illustrated how simple and straightforward the new methodology is relative to the cause-specific proportional hazards formulation with multiple continuous and discrete covariates.

As with the Cox model for univariate right-censored survival data, the complete-data partial likelihood is semiparametric efficient under the filtration $\mathcal{F}^1(t)$; see Appendix C for a sketch of the proof. This efficiency result is premised on a direct modeling framework for F_1 in which no assumptions are made about F_k, $k \neq 1$. Under this condi-

Table 4. Estimated Coefficients and Standard Errors in Models for Breast Cancer Recurrence

	Log(nodes)	Tumor size	Age	Treatment	Treatment × t	Treatment × t^2
			Subdistribution hazards			
Coefficient	.27	.11	−.037	−2.03	.86	−.087
Standard Error	.11	.04	.026	.62	.31	.031
			Cause specific hazards			
Coefficient	.29	.14	−.012	−1.99	.84	−.084
Standard Error	.11	.04	.026	.62	.31	.031

tion, only information from failures for which $\varepsilon = 1$ can be utilized and the risk set at time t fails to distinguish between individuals who are still alive and those who have already failed from another cause. As mentioned previously, an intriguing result is that in the two-sample problem, the score test based on $\mathbf{U}_2(0)$ is nearly identical to Gray's test (1988) based on nonparametric estimates of the subdistribution hazards. Thus it would be interesting to know whether other methods requiring models for $F_k, k \neq 1$ can recover much information.

To gain efficiency, one might consider techniques for simultaneously estimating $F_k, k = 1, \cdots, K$, via the cause-specific hazard functions, $\lambda_k^{\mathrm{CS}}(t; \mathbf{Z}), k = 1 \ldots, K$, which are completely specified by K log(− log) transformation models for the cumulative incidence functions. That is, $\lambda_k^{\mathrm{CS}}(t; \mathbf{Z}) = dF_k(t; \mathbf{Z})\{1 - \sum_{i=1}^K F_i(t; \mathbf{Z})\}^{-1}$, assuming censoring is noninformative. Instead of separately estimating models for each cumulative incidence function, one can jointly estimate the implied models for $\lambda_k^{\mathrm{CS}}(t; \mathbf{Z}), k = 1, \ldots, K$. Because the cause-specific hazards identify the joint distribution of (T, ε) (Prentice et al. 1978), one may maximize a single nonparametric likelihood constructed from $\lambda_k^{\mathrm{CS}}, k = 1, \ldots, K$, over the unknown regression parameters and baseline subdistribution hazard functions in the models for $F_k, k = 1, \ldots, K$. Here, all failures and all censorings contribute directly to the likelihood of the observed data; seemingly, no information is lost. This sort of procedure is likely to be quite complicated, requiring optimization over function spaces, and is less robust to model misspecification than the estimator from $\mathbf{U}_2(\boldsymbol{\beta})$. Note in particular that partial likelihood methods could not generally be used to estimate the regression parameters, unlike in the proportional cause-specific hazards model. At the same time, the likelihood method has the added flexibility of allowing one to specify g-transformation models other than the log(− log) model for the subdistribution. That is, one could assume $g_k\{F_k(t; \mathbf{Z})\} = h_{k0}(t) + \mathbf{Z}^T\boldsymbol{\beta}_k, k = 1, \ldots, K$, where g_k may vary across causes.

One alternative to the proportional hazards model for the subdistribution is the linear regression model first proposed for the survival distribution by Aalen (1980). Recently, well-justified semiparametric methodology for this log-linear transformation model has been developed (Lin and Ying 1994). The weighting techniques described in Section 4 can be easily adapted to the martingale-type estimating equations used for semiparametric analysis of the additive risk model with $g = -\log(1 - u)$ in the univariate setting. This marginal approach, in which $F_k, k \neq 1$, are unspecified, will be communicated in a separate report.

APPENDIX A: CONSISTENCY AND ASYMPTOTIC NORMALITY OF $\hat{\beta}$ WITH RIGHT-CENSORED DATA

Using the consistency results for censoring-complete data, $n^{-1}\mathbf{U}_{1*}(\boldsymbol{\beta})$ converges uniformly in probability to a continuous and deterministic function of $\boldsymbol{\beta}$ that has a unique 0 at $\boldsymbol{\beta}_0 \in \mathcal{R}^p$ and is bounded in a neighborhood of $\boldsymbol{\beta}_0$. Thus, for consistency with right-censored data, it is sufficient to prove that $n^{-1}\{\mathbf{U}_2(\boldsymbol{\beta}) - \mathbf{U}_{1*}(\boldsymbol{\beta})\}$ goes in probability to 0 uniformly for $\boldsymbol{\beta}$ in a compact neighborhood of $\boldsymbol{\beta}_0$. As with the complete and the censoring-complete scores, (6) can be centered:

$$\mathbf{U}_2(\boldsymbol{\beta}) = \sum_{i=1}^n \int_0^\infty \left[\mathbf{Z}_i(u) - \frac{\sum_j w_j(u)Y_j(u)\mathbf{Z}_j(u)\exp\{\mathbf{Z}_j^T(u)\boldsymbol{\beta}\}}{\sum_j w_j(u)Y_j(u)\exp\{\mathbf{Z}_j^T(u)\boldsymbol{\beta}\}}\right] \times w_i(u)\, dM_i^1(u, \boldsymbol{\beta}).$$

Using algebraic manipulations and the uniform convergence of $\hat{G}(\cdot)$ to $G(\cdot)$ (Fleming and Harrington 1991, chap. 6),

$$n^{-1}\mathbf{U}_2(\boldsymbol{\beta}) = n^{-1}\sum_{i=1}^n \int_0^\infty \left\{\mathbf{Z}_i(u) - \frac{\hat{\mathbf{S}}_2^{(1)}(\boldsymbol{\beta}, u)}{\hat{S}_2^{(0)}(\boldsymbol{\beta}, u)}\right\} \times \tilde{w}_i(u)\, dM_i^1(u, \boldsymbol{\beta}) + n^{-1}\sum_{i=1}^n \mathbf{R}_i(\boldsymbol{\beta}) + o_p(1), \quad \text{(A.1)}$$

where

$$\mathbf{R}_i(\boldsymbol{\beta}) = \int_0^\infty \left\{\frac{\hat{G}(u)}{\hat{G}(X_i \wedge u)} - \frac{G(u)}{G(X_i \wedge u)}\right\} \times \left[\mathbf{Z}_i(u) - \frac{\sum_j \tilde{w}_j(u)Y_j(u)\mathbf{Z}_j(u)\exp\{\mathbf{Z}_j^T(u)\boldsymbol{\beta}\}}{\sum_j \tilde{w}_j(u)Y_j(u)\exp\{\mathbf{Z}_j^T(u)\boldsymbol{\beta}\}}\right] \times r_i(u)\, dM_i^1(u, \boldsymbol{\beta}),$$

and $\tilde{w}_j(u) = r_j(u)G(u)\{G(X_i \wedge u)\}^{-1}$. Manipulating the martingale representation of the Kaplan–Meier estimator (Gill 1980), we can express

$$\frac{\hat{G}(t)}{\hat{G}(X_i \wedge t)} - \frac{G(t)}{G(X_i \wedge t)}$$

$$= -\frac{G(t)I(X_i < t)}{G(X_i)} \sum_{j=1}^{n} \int_{X_i}^{t} \frac{1}{\sum_{k=1}^{n} I(X_k \geq u)}$$

$$\times dM_j^c(u) + o_p(1),$$

where $M_i^c(u) = I(X_i \leq u, \Delta_i = 0) - \int_0^u I(X_i \geq t)\, d\Lambda^c(t)$ is the martingale associated with the censoring process and $\Lambda^c(u)$ is the common cumulative hazard of the censoring distribution. The second term in (A.1) is just an average of martingale integrals with respect to the censoring filtration, $\mathcal{F}^c(u) = \{I(X_i \geq t), I(X_i \leq t, \Delta_i = 0), \mathbf{Z}_i(t), t \leq u, \varepsilon_i, i = 1, \ldots, n\}$. Under regularity conditions, $n^{-1}\sum_i \mathbf{R}_i(\beta)$ is dominated by a bounded function and will converge in probability to 0 uniformly for β in a compact neighborhood of β_0 (Newey and McFadden 1994, lem. 2.4).

The compensated counting process, $M_i^1(u, \beta)$, retains the martingale property with respect to the complete-data filtration when evaluated at β_0. However, it is important to recognize that $\hat{\mathbf{S}}_2^{(k)}, k = 0, 1$, and $r_i(u)/G(X_i \wedge u)$ are not adapted with respect to $\mathcal{F}^1(u)$ and the first term in (A.1) is not a martingale integral at β_0. Under regularity conditions, $\mathbf{s}^{(k)}$ can be substituted for $\hat{\mathbf{S}}_2^{(k)}$ in the first term, yielding

$$n^{-1}\mathbf{U}_2(\beta) = n^{-1}\sum_{i=1}^{n} \int_0^{\infty} \left\{ \mathbf{Z}_i(u) - \frac{\mathbf{s}^{(1)}(\beta, u)}{s^{(0)}(\beta, u)} \right\}$$

$$\times \tilde{w}_i(u)\, dM_i^1(u, \beta) + o_p(1), \quad \text{(A.2)}$$

because for $k = 0, 1, 2$,

$$\mathbf{t}^{(k)}(\beta, u) = \lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} \frac{r_i(u)\hat{G}(u)}{\hat{G}(X_i \wedge u)}\, Y_i(u)\mathbf{Z}_i(u)^{\otimes k}$$

$$\times \exp\{\mathbf{Z}_i^T(u)\beta\} = G(u)\mathbf{s}^{(k)}(\beta, u).$$

It is still true that $E\{\tilde{w}_i(u)dM_i^1(u, \beta);\, \mathbf{Z}_i\} = E[E\{r_i(u)G(u)\, dM_i^1(u, \beta)/G(X_i \wedge u);\, \mathbf{Z}_i, T_i, \varepsilon_i\}] = G(u)E\{dM_i^1(u, \beta); \mathbf{Z}_i\}$ for all β and u. In addition, $dM_i^{1*}(u, \beta) = I(C_i \geq u)dM_i^1(u, \beta)$. Using these facts, the asymptotic results for $\mathbf{U}_{1*}(\beta)$, and the boundedness of the covariates,

$$n^{-1}\{\mathbf{U}_2(\beta) - \mathbf{U}_{1*}(\beta)\} = \sum_{i=1}^{n} \int_0^{\infty} \left\{ \mathbf{Z}_i(u) - \frac{\mathbf{s}^{(1)}(\beta, u)}{s^{(0)}(\beta, u)} \right\}$$

$$\times \{\tilde{w}_i(u) - I(C_i \geq u)\}\, dM_i^1(u, \beta) + o_p(1)$$

is continuous and dominated by a bounded function and so will vanish uniformly in probability for β in a compact neighborhood of β_0 (Newey and McFadden 1994, lem. 2.4).

Similar manipulations requiring basic empirical process theory for justification show

$$n^{-1/2}\mathbf{U}_2(\beta_0)$$

$$= n^{-1/2}\sum_{i=1}^{n} \int_0^{\infty} \left\{ \mathbf{Z}_i(u) - \frac{\mathbf{s}^{(1)}(\beta_0, u)}{s^{(0)}(\beta_0, u)} \right\} \tilde{w}_i(u) dM_i^1(u, \beta_0)$$

$$+ n^{-1/2}\sum_{i=1}^{n} \mathbf{R}_i(\beta_0) + o_p(1).$$

Lenglart's inequality (Lenglart 1977) and other approximations can be applied to the second term, giving

$$n^{-1/2}\mathbf{U}_2(\beta_0)$$

$$= n^{-1/2}\sum_{i=1}^{n} \int_0^{\infty} \left\{ \mathbf{Z}_i(u) - \frac{\mathbf{s}^{(1)}(\beta_0, u)}{s^{(0)}(\beta_0, u)} \right\} \tilde{w}_i(u)\, dM_i^1(u, \beta_0)$$

$$+ n^{-1/2}\sum_{i=1}^{n} \int_0^{\infty} \frac{\mathbf{q}(u)}{\pi(u)}\, dM_i^c(u) + o_p(1),$$

where

$$\mathbf{q}(u) = -\lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} \int_0^{\infty} \left\{ \mathbf{Z}_i(s) - \frac{\mathbf{s}^{(1)}(\beta_0, s)}{s^{(0)}(\beta_0, s)} \right\}$$

$$\times \tilde{w}_i(s)\, dM_i^1(s, \beta_0) I(s \geq u > X_i)$$

and

$$\pi(u) = \lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} I(X_i \geq u). \quad \text{(A.3)}$$

In the limit, $n^{-1/2}\mathbf{U}_2(\beta_0) = n^{-1/2}\sum_i(\eta_i + \psi_i) + o_p(1)$ can be viewed as the sum of n independent and identically distributed random variables, where

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

and

$$\psi_i = \int_0^{\infty} \frac{\mathbf{q}(u)}{\pi(u)}\, dM_i^c(u).$$

Utilizing the multivariate central limit theorem, $n^{-1/2}\mathbf{U}_2(\beta_0)$ is asymptotically normal with covariance matrix $\Sigma = E\{(\eta_i + \psi_i)(\eta_i + \psi_i)^t\}$. An empirical covariance estimator, in which all unknown quantities are replaced with their observed counterparts, converges in probability to Σ; see Section 4.

APPENDIX B: WEAK CONVERGENCE OF $J_1\{t; \mathbf{z}_0\}$

In this appendix we show that for an individual with covariates $\mathbf{Z} = \mathbf{z}_0$, $J_1\{t; \mathbf{z}_0\} = n^{1/2}(m[\hat{F}_1\{t; \mathbf{z}_0\}] - m[F_1\{t; \mathbf{z}_0\}])$ converges weakly to a Gaussian process on $[0, \tau)$. First, weak convergence of $P\{t; \mathbf{z}_0\} = n^{1/2}[\hat{\Lambda}_1\{t; \mathbf{z}_0\} - \Lambda_1\{t; \mathbf{z}_0\}]$ is established. Because $P\{t; \mathbf{z}_0\}$ cannot be expressed as a martingale integral, verification of finite-dimensional convergence and tightness will follow from first principles. The functional delta method then provides the desired result.

We start by noting that

$$n^{1/2}[\hat{\Lambda}_1\{t; \mathbf{z}_0\} - \Lambda_1\{t; \mathbf{z}_0\}]$$

$$= n^{-1/2}\sum_{i=1}^{n} \int_0^{t} \frac{\exp\{\mathbf{z}_0^T(u)\hat{\beta}\}}{\hat{S}_2^{(0)}(\hat{\beta}, u)}\, w_i(u)\, dN_i(u)$$

$$- n^{-1/2}\sum_{i=1}^{n} \int_0^{t} \frac{w_i(u)Y_i(u)\exp\{\mathbf{Z}_i^T(u)\beta_0\}}{\hat{S}_2^{(0)}(\beta_0, u)}$$

$$\times \exp\{\mathbf{z}_0^T(u)\beta_0\}\, d\Lambda_{10}(u), \quad \text{(B.1)}$$

where $\hat{\mathbf{S}}_2^{(k)}, k = 0, 1, 2$, are defined in (7). Taking a Taylor series around β_0 and using Slutsky's theorem for stochastic processes, the first term in (B.1), $n^{1/2}\hat{\Lambda}_1\{t; \mathbf{z}_0\}$, is asymptotically equivalent

to

$$n^{-1/2}\sum_{i=1}^{n}\int_0^t \frac{\exp\{\mathbf{z}_0^T(u)\boldsymbol{\beta}_0\}}{\hat{S}_2^{(0)}(\boldsymbol{\beta}_0,u)}\frac{r_i(u)\hat{G}(u)}{\hat{G}(X_i\wedge u)}\,dN_i(u)$$

$$+\,\mathbf{h}(t,\mathbf{z}_0)n^{1/2}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0) \quad \text{(B.2)}$$

where $\mathbf{h}\{t,\mathbf{z}_0\}$ is given in (4). Observe that the second term in (B.1) is just $n^{-1/2}d\Lambda_1\{t;\mathbf{z}_0\}$ integrated with respect to $\hat{S}_2^{(0)}(\boldsymbol{\beta}_0,u)/\hat{S}_2^{(0)}(\boldsymbol{\beta}_0,u)=1$. Combining this expression with the first term in (B.2), we can use Gill's martingale representation of $\hat{G}(\cdot)$, the results for $n^{1/2}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)$ from Section 4, and other asymptotic devices to show that

$$\begin{aligned}P\{t;\mathbf{z}_0\} &= n^{-1/2}\sum_{i=1}^{n}\int_0^t \frac{\exp\{\mathbf{z}_0^T(u)\boldsymbol{\beta}_0\}}{s^{(0)}(\boldsymbol{\beta}_0,u)}\,\tilde{w}_i(u)\,dM_i^1(u,\boldsymbol{\beta}_0)\\ &\quad+\mathbf{h}^t\{t,\mathbf{z}_0\}\boldsymbol{\Omega}^{-1}n^{-1/2}\\ &\quad\times\sum_{i=1}^{n}\left[\int_0^\infty\left\{\mathbf{Z}_i(u)-\frac{\mathbf{s}^{(1)}(\boldsymbol{\beta}_0,u)}{s^{(0)}(\boldsymbol{\beta}_0,u)}\right\}\tilde{w}_i(u)\right.\\ &\quad\left.\times\,dM_i^1(u,\boldsymbol{\beta}_0)+\int_0^\infty\frac{\mathbf{q}(u)}{\pi(u)}\,dM_i^c(u)\right]\\ &\quad+n^{-1/2}\sum_{i=1}^{n}\int_0^\infty\frac{v\{u,t,\mathbf{z}_0\}}{\pi(u)}\,dM_i^c(u)+o_p(1),\end{aligned}$$

where $\mathbf{s}^{(k)}, k=0,1,2$, and $\boldsymbol{\Omega}$ can be found in (3), $v\{u,t,\mathbf{z}_0\}$ is given in (9), and $\pi(u)$ and $\mathbf{q}(u)$ are defined in (A.3). From the preceding formula, it is clear that $P\{t;\mathbf{z}_0\}$ is asymptotically equivalent to the sum of n independent and identically distributed random processes. Finite-dimensional convergence follows from the multivariate central limit theorem. As for tightness, the second and third terms are functions of time only through the nonrandom function $\mathbf{h}\{t,\mathbf{z}_0\}$ and are naturally tight. The fourth term is a martingale integral with respect to $v\{u,t,\mathbf{z}_0\}/\pi(u)$, a deterministic function of t, and is also tight. Tightness of the first term follows from basic properties of empirical processes. Thus $P\{t;\mathbf{z}_0\}$, which is a sum of tight functions, is tight. With an application of the functional delta method, $J_1\{t;\mathbf{z}_0\}=n^{1/2}(m[\hat{F}_1\{t;\mathbf{z}_0\}]-m[F_1\{t;\mathbf{z}_0\}])=\dot{m}[F_1\{t;\mathbf{z}_0\}]\exp[-\Lambda_1\{t;\mathbf{z}_0\}]P\{t;\mathbf{z}_0\}+o_p(1)$ converges weakly to a Gaussian process. The covariance function of this process is very complicated, however. In Section 5 we show how one can easily simulate from a mean 0 Gaussian process, $\hat{J}_1\{t;\mathbf{z}_0\}$, that converges weakly to the same limiting process as $J_1\{t;\mathbf{z}_0\}$.

APPENDIX C: EFFICIENCY CONSIDERATIONS

In this appendix we present a heuristic argument that the complete-data partial likelihood outlined in Section 2 is semiparametric efficient across all parametric submodels based on the information in $\mathcal{F}^1(t)$. Details are provided to clarify the foundations of likelihood-based methodology for the subdistribution. The result follows closely from the proof for the univariate Cox model.

To begin, an extra condition is necessary to account for the dependent nature of the latent failure times. We artificially assume that the subdistributions are orthogonal to one another (condition 1); that is, the parameter spaces on which the subdistributions are defined do not overlap. In theory, the condition is not satisifed as the subdistribution functions are restricted, so that $\sum_k F_k(t;\mathbf{Z})\le 1$ for all t, $\mathbf{Z}$, where $F_k(t;\mathbf{Z})=\Pr(T\le t,\varepsilon=k|\mathbf{Z})$ is the subdistribution for the kth failure type. The cause-specific hazard parameterization, which violates condition 1, enforces this restriction. However, condition 1 is in the spirit of the direct modeling approach taken in this article; only information contained in $\mathcal{F}^1(t)$ is used. With this in mind, the results of the exercise should be interpreted as follows: When estimating the regression parameters in model (1) without making any assumptions about the form of the other subdistribution functions, the complete-data semiparametric efficient score is given in (2).

Without censoring, an individual likelihood contribution is

$$\prod_{k=1}^{K} dF_k\{T;\mathbf{Z}\}^{I(\varepsilon=k)}.$$

Under condition 1, the pertinent contribution to the likelihood for estimation of the subdistribution for cause 1 is

$$\begin{aligned}&dF_1\{T;\mathbf{Z}\}^{I(\varepsilon=1)}\\ &=\left(\lambda_1(T,\boldsymbol{\gamma})\exp\{\mathbf{Z}^T(T)\boldsymbol{\beta}\}\right.\\ &\quad\left.\times\exp\left[-\int_0^T\lambda_1(u,\boldsymbol{\gamma})\exp\{\mathbf{Z}^T(u)\boldsymbol{\beta}\}\,du\right]\right)^{I(\varepsilon=1)}.\end{aligned}$$

Here $\lambda_1(u,\boldsymbol{\gamma}_0)=\lambda_{10}(u)$ at the true value of $\boldsymbol{\gamma}\in\mathcal{R}^s$, an s-dimensional nuisance parameter, $s<\infty$. Therefore, $\lambda_1(u,\boldsymbol{\gamma})$ is a parametric submodel. The semiparametric efficient score can be determined via projection theory, and the derivation is identical to that for univariate survival data. The efficient score is given by

$$\int_0^\infty\left\{\mathbf{Z}(u)-\frac{\mathbf{s}^{(1)}(\boldsymbol{\beta},u)}{s^{(0)}(\boldsymbol{\beta},u)}\right\}dM^1(u,\boldsymbol{\beta}),$$

which is equivalent to the complete-data influence function for the maximum partial likelihood estimator for $\boldsymbol{\beta}$ from model (1).

[Received July 1997. Revised October 1998.]

REFERENCES

Aalen, O. O. (1978), "Nonparametric Estimation of Partial Transition Probabilities in Multiple Decrement Models," *The Annals of Statistics*, 6, 534–545.

——— (1980), *A Model for Nonparametric Regression Analysis of Counting Processes* Lecture Notes in Statistics 2, New York: Springer-Verlag, pp. 1–25.

Andersen, P. K., Borgan, O., Gill, R. D., and Keiding, N. (1993), *Statistical Models Based on Counting Processes*, New York: Springer-Verlag.

Benichou, J., and Gail, M. H. (1992), "Estimates of Absolute Cause-Specific Risk in Cohort Studies," *Biometrics*, 46, 813–826.

Breslow, N. E. (1974), "Covariance Analysis of Censored Survival Data," *Biometrics*, 30, 89–99.

Cheng, S. C., Wei, L. J., and Ying, Z. (1995), "Analysis of Transformation Models with Censored Data," *Biometrika*, 82, 835–846.

——— (1997), "Predicting Survival Probabilities With Semiparametric Transformation Models," *Journal of the American Statistical Association*, 92, 227–235.

Cox, D. R. (1972), "Regression Models and Life Tables" (with discussion), *Journal of the Royal Statistical Society*, Ser. B, 34, 187-220.

Cuzick, J. (1988), "Rank Regression," *The Annals of Statistics*, 16, 1369–1389.

Dabrowska, D. M., and Doksum, K. A. (1988), "Estimation and Testing in a Two-Sample Generalized Odds-Rate Model," *Journal of the American Statistical Association*, 83, 744–749.

Fine, J. P., Ying, Z., and Wei, L. J. (1998), "On the Linear Transformation Model With Censored Data," *Biometrika*, 85, pp. 980–986.

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

Gaynor, J. J., Feuer E. J., Tan, C. C., Wu, D. H., Little, C. R., Strauss, D. J., Clarkson, B. D., and Brennan, M. F. (1993), "On the Use of Cause-Specific Failure and Conditional Failure Probabilities: Examples From

Clinical Oncology Data," *Journal of the American Statistical Association*, 88, 400–409.

Gill, R. (1980), *Censoring and Stochastic Integrals*, Mathematical Centre Tracts 124, Amsterdam: Mathematisch Centrum.

Gray, R. J. (1988), "A Class of *K*-Sample Tests for Comparing the Cumulative Incidence of a Competing Risk," *The Annals of Statistics*, 16, 1141–1154.

Hall, W. J., and Wellner, J. A. (1980), "Confidence Bands for Survival Curves Under the Proportional Hazards Model," *Biometrika*, 80, 557–572.

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

Korn, E. L., and Dorey, F. J. (1992), "Applications of Crude Incidence Curves," *Statistics in Medicine*, 11, 813–829.

Larson, M. G. (1984), "Covariate Analysis of Competing Risks Models with Log-Linear Models," *Biometrics*, 40, 459–469.

Lenglart, E. (1977), "Relation de Domination Entre Deux Processus," *Annales Institut Henri Poincaré*, 13, 171–179.

Lin, D. Y., Fleming, T. R., and Wei, L. J. (1994), "Confidence Bands for Survival Curves Under the Proportional Hazards Model," *Biometrika*, 81, 73-81.

Lin, D. Y., and Ying, Z. (1994), "Semiparametric Analysis of the Additive Risk Model," *Biometrika*, 81, 61-71.

Murphy, S. A., Rossini, A. J., and Van der Vaart, A. W. (1997), "Maximum Likelihood Estimation in the Proportional Odds Model," *Journal of the American Statistical Association*, 92, 968–976.

Nelson, W. B. (1972), "Theory and Application of Hazard Plotting for Censored Failure Data," *Technometrics*, 14, 945–965.

Newey, W., and McFadden, D. (1993). "Estimation in Large Samples," in *Handbook of Econometrics*, Vol. 5, eds. D. McFadden and R. Engler, Amsterdam: North-Holland.

Pepe, M. S. (1991), "Inference for Events with Dependent Risks in Multiple Endpoint Studies," *Journal of the American Statistical Association*, 86, 770–778.

Pepe, M. S., and Mori, M. (1993), "Kaplan-Meier, Marginal or Conditional Probability Curves in Summarizing Competing Risks Failure Time Data?," *Statistics in Medicine*, 12, 737–751.

Prentice, R. L., Kalbfleisch, J. D., Peterson, A. V., Flournoy, N., Farewell, V. T., and Breslow, N. E. (1978), "The Analysis of Failure Times in the Presence of Competing Risks," *Biometrics*, 34, 541–554.

Rebolledo, R. (1978), "Sur les Applications de la Theorie des Martingales a L'etude Statistique D'une Famille de Processus Ponctuels," *Springer Lecture Notes in Mathematics 636*, Berlin: Springer-Verlag.

Robins, J. M., and Rotnitzky, A. (1992), "Recovery of Information and Adjustment for Dependent Censoring Using Surrogate Markers," in *AIDS Epidemiology-Methodological Issues*, eds. N. Jewell, K. Dietz, and V. Farewell, Boston: Birkhauser, pp. 24–33.

Schoenfeld, D. (1982), "Partial Residuals for the Proportional Hazards Regression Model," *Biometrika*, 69, 139–241.