Competing Risks Regression for Stratified Data

Bingqing Zhou[1,*], **Aurelien Latouche**[2,**], **Vanderson Rocha**[3,***], and **Jason Fine**[4,****]

[1]Division of Biostatistics, School of Public Health, Yale University, New Haven, Connecticut 06520, U.S.A.

[2]Université Versailles St-Quentin, EA2506, Versailles, France

[3]Acute Leukemia Working Party and Eurocord, European Blood and Marrow Transplant Group, Hopital Saint Louis, Paris University, 75010 Paris, France

[4]Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599-7420, U.S.A.

Summary

For competing risks data, the Fine–Gray proportional hazards model for subdistribution has gained popularity for its convenience in directly assessing the effect of covariates on the cumulative incidence function. However, in many important applications, proportional hazards may not be satisfied, including multicenter clinical trials, where the baseline subdistribution hazards may not be common due to varying patient populations. In this article, we consider a stratified competing risks regression, to allow the baseline hazard to vary across levels of the stratification covariate. According to the relative size of the number of strata and strata sizes, two stratification regimes are considered. Using partial likelihood and weighting techniques, we obtain consistent estimators of regression parameters. The corresponding asymptotic properties and resulting inferences are provided for the two regimes separately. Data from a breast cancer clinical trial and from a bone marrow transplantation registry illustrate the potential utility of the stratified Fine–Gray model.

Keywords

Clustering; Dependent censoring; Hazard of subdistribution; Inverse weighting; Martingale; Multicenter trials; Partial likelihood

1. Introduction

Competing risks data arise when subjects may fail from several distinct types including disease and nondisease related causes. The observed data consist of a failure time and a failure type. In this setting, the cumulative incidence (also referred to as the subdistribution) may be of prime interest as it quantifies the absolute risks of different failure types. The focus of this article is evaluating covariate effects on the cumulative incidence across levels of a stratifying variable using the popular Fine–Gray regression model.

[*]bingqing.zhou@yale.edu
[**]aurelien.latouche@uvsq.fr
[***]vanderson.rocha@sls.aphp.fr
[****]jfine@bios.unc.edu

To compare the cumulative incidence of a particular type of failure among different groups, Gray developed a class of K-sample tests based on weighted averages of the subdistribution hazards for that particular failure type of interest (Gray, 1988). Pepe (1991) proposed a nonparametric two sample test for group effects. No covariates other than the group factor can be considered in these tests. A semiparametric proportional hazards model was proposed by Fine and Gray (1999) to assess the effect of covariates or prognostic factors on the cumulative incidence function. The Fine–Gray model captures the cause-specific failure probabilities after adjusting for patient-specific risks and provides a way to directly examine the influence of risk factors on the absolute risks. Alternative models and methods of estimation for cumulative incidence regression have been studied in, for example, Klein and Andersen (2005); Scheike and Zhang (2008); and Scheike, Zhang, and Gerds (2008), including nonproportional hazards models and goodness-of-fit methods for assessing the proportional hazards assumption. In this article, we focus on a simple extension of the Fine–Gray model, which is currently the most widely used cumulative incidence regression methodology in practice, adapting techniques that are widely employed with proportional hazards modeling under independent censoring.

In real applications, the proportional subdistribution hazards assumption may not hold for certain covariates. In these scenarios, naive application of the Fine–Gray model, either omitting such covariates or including them assuming they satisfy the proportional hazards assumption, could lead to biased estimation and tests, and potential loss of power. We propose a stratified Fine–Gray model that adjusts for such discrete factors, without estimating their effects on the sub-distribution hazard. The main idea is to allow the baseline hazard function to vary across levels of stratification variables.

In classical survival analysis, stratifying is a common data analytic strategy. For instance, in proportional hazards regression with independent censoring, there are frequently important factors, the different levels of which produce hazard functions that differ markedly from proportionality. Stratification on these factors may yield a simpler and more flexible analysis than modeling interactions parametrically with functions of time defined through time-dependent covariates (Kalbfleisch and Prentice, 2002). Notably, stratification is a standard approach to account for varying patient populations such as in multicenter clinical trials, where baseline hazards vary across centers (Therneau and Grambsch, 2000). Both Splus and SAS statistical software offer the strata option in fitting the Cox (1972) regression model, but these functions may not be suitable for the Fine–Gray model.

We consider two typical data sets that exhibit different stratification regimes where cumulative incidence is a primary endpoint. We say data are regularly stratified if there are a small number of large groups (strata) and that data are highly stratified if the number of groups (strata) is large compared to the strata sizes.

Clinical trial E1178 conducted by the Eastern Cooperative Oncology Group compared 2 years of treatments tamoxifen therapy to placebo in elderly (⩾ age 65) breast cancer patients with positive auxiliary nodes. The cumulative incidence of breast cancer recurrence is of great interest. Nonproportional hazards of the treatment effect has previously been observed using time-dependent covariates (Fine and Gray, 1999), suggesting stratification on treatment when testing effects of other risk factors. Within each treatment arm, there are a large number of patients, as with regularly stratified data.

A bone marrow transplant registry is maintained by the European Blood and Marrow Transplant Group. The primary endpoint is time from graft to first occurrence of acute or chronic graft-versus-host-disease (GvHD) whereas death and relapse (free of GvHD) are the competing events. This is a multicenter design. Katsahian et al. (2006) proposed a frailty

model for the subdistribution hazard to assess the heterogeneity across clusters and to incorporate such an effect when testing other risk factors. A stratified competing risks regression can be applied here to account for correlation within center due to unobserved center effects without estimating those effects. The number of centers is much larger than the number of patients from each center, yielding highly stratified data.

Existing work on the cumulative incidence function for clustered and/or stratified data includes, but is not limited to modified nonparametric Gray type tests (Chen et al., 2008), mixed proportional hazards models (Katsahian et al., 2006), and introducing interactions using time-dependent covariates to the Fine-Gray (1999) model. The modified tests focus on formally testing group effects, which limits the introduction of continuous covariates and the quantification of covariate effects. The mixed proportional hazards model is a frailty regression model tailored to competing risks data. Although the approach seems promising, the statistical properties of the resulting analysis are unclear (Katsahian et al., 2006). Moreover, if the main goal is to investigate covariate effects, introducing a dependence parameter in modeling the joint distribution within each cluster does not seem to have much advantage. Stratified (Gross and Huber, 1987) and marginal proportional hazards approaches (Lee, Wei, and Amato, 1992) have been advocated for such scenarios with independent censoring. Such models avoid introducing time-dependent covariates to capture cluster effects, simplifying the explanation of the noncluster covariate effects.

In Section 2, we introduce the stratified Fine–Gray model. In Section 3, we discuss partial likelihood inferences for the stratified model in the absence of independent censoring, along with inverse weighting estimation equations, which permit additional independent censoring. Inferential issues are presented in Section 4, with results for the two stratification regimes in separate subsections. Prediction of the cumulative incidence is discussed in Section 5. Section 6 discusses some simulation studies, with the analysis of the two motivating data sets following in Section 7. We conclude with a few remarks. Technical details are included in the Web Appendices.

2. Data and Model

Let T, C be the failure and censoring times, $\varepsilon \in \{1, \ldots, m\}$ be the cause of failure, and $\mathbf{Z}$ be a $I \times 1$ vector of covariates. For right censored data, one observes $\{X = T \wedge C, \Delta = I(T \leq C), \Delta \varepsilon, \mathbf{Z}\}$, where $a \wedge b = min(a, b)$. Here, as in Fine and Gray (1999), we permit $\mathbf{Z}$ to include time-varying covariates that are known, deterministic functions of time and time-independent covariates, and hence are fully observed. In the sequel, we suppress the dependence on time, when there is no loss of clarity. We are interested in assessing the effect of covariates on the cumulative incidence function for failure from cause 1, conditional on the covariates: $F_1(t, \mathbf{Z}) \equiv \Pr(T \leq t, \varepsilon = 1 | \mathbf{Z})$. We assume $(T, \varepsilon, \mathbf{Z})$ and C are independent given $\mathbf{Z}$; assumptions for the dependence of C on stratum are described below.

Fine and Gray (1999) proposed a proportional hazards model for the subdistribution hazard $\lambda_1(t; \mathbf{Z}) = dF_1(t, \mathbf{Z})/\{1 - F_1(t, \mathbf{Z})\}$. The model relates the covariates with $\lambda_1(\cdot)$ for independent and identically distributed (i.i.d.) observations $\{X_i, \Delta_i, \Delta_i \varepsilon_i, \mathbf{Z}_i\}_{i=1, \ldots, n}$ by

$\lambda_1(t; Z_i) = \lambda_{10}(t) \exp\left(\beta_0' Z_i\right)$, where $\lambda_{10}(\cdot)$ is an unspecified, nonnegative function denoting the baseline subdistribution hazard when covariate $\mathbf{Z} = 0$; $\boldsymbol{\beta}_0$ is a $I \times 1$ vector of unknown regression parameters.

The goal of our study is to introduce stratification to the Fine–Gray model. With stratification, we observe $\{X_{ki}, \Delta_{ki}, \Delta_{ki} \varepsilon_{ki}, \mathbf{Z}_{ki}\}_{k=1, \ldots, s; i=1, \ldots, n_k}$, where k denotes the stratum, i denotes the subject within strata, and $\sum_{k=1}^{s} n_k = n$. For regularly stratified data, we

assume i.i.d. observations within strata, whereas for highly stratified data, we will assume that information observed within a cluster is i.i.d. across strata; see condition 3 in Web Appendix A. When appropriate, the n observations can be denoted using single index: $\{X_l, \Delta_l, \Delta_l \varepsilon_l, \mathbf{Z}_l, K_l\}_{l=1,\ldots,n}$, where $K = k \in \{1,\ldots,s\}$ gives the stratification information. We define the cumulative incidence function for stratum k as $F_{1k}(t, \mathbf{Z}) \equiv \Pr(T \leq t, \varepsilon = 1 \mid \mathbf{Z}, K = k)$. Correspondingly, the proportional hazards specification for the subdistribution hazard for kth stratum $\lambda_{1k}(t, \mathbf{Z}) = dF_{1k}(t, \mathbf{Z})/\{1 - F_{1k}(t, \mathbf{Z})\}$ is

$$\lambda_{1k}(t; Z_{ki}) = \lambda_{1k0}(t) \exp\left(\beta_0' Z_{ki}\right), \qquad (1)$$

where λ_{1k0} is the baseline subdistribution hazard in stratum $k(= 1,\ldots, s)$ and β_0 is the regression coefficient, which is assumed common to all strata. An important point is that no assumptions are made about the relationships between the baseline hazard functions.

3. Estimation

The estimation procedure is adapted from Fine and Gray (1999). The data are said to be complete when failure time T and failure cause ε are observed for all individuals; the data are censoring complete (CC) when the failure time is right censored but potential censoring time C is always observed. Starting from a modification of the partial likelihood for the subdistribution for complete or CC data, we then extend the estimation equation to classical right censored data.

3.1 Complete and Censoring Complete Data

With complete data, the risk set at time of failure of the ith subject in the kth stratum is $R_{ki} = \{i' : (T_{ki'} \geq T_{ki}) \cup (T_{ki'} \leq T_{ki} \cap \varepsilon_{ki'} \geq 1)\}$. When the data are CC, the associated risk set is modified to $R_{ki} = \{i' : (T_{ki'} \wedge C_{ki'} \geq T_{ki}) \cup (T_{ki'} \leq T_{ki} \cap \varepsilon_{ki'} \geq 1 \cap C_{ki'} \geq T_{ki})\}$. We will detail the derivation of the partial likselihood approach for the CC data, because the *complete data* are special cases of the CC data by letting the censoring times be larger than the maximum failure time.

Let $N_{ki} = I(T_{ki} \leq t, \varepsilon_{ki} = 1)$ and $Y_{ki}(t) = 1 - N_{ki}(t-)$ denote the counting process and risk process for the complete data, respectively. When the data are CC, the risk process is modified to $Y_{ki}^*(t) = I(C_{ki} \geq t) Y_{ki}(t)$. Applying the partial likelihood approach to $\lambda_{1k}(t \mid \mathbf{Z})$, we obtain the corresponding partial likelihood

$$L^*(\beta) = \prod_{k=1}^{s} \prod_{i=1}^{n_k} \left\{ \frac{e^{\beta' Z_{ki}(X_{ki})}}{\sum_{i=1}^{n_k} Y_{ki'}^*(X_{ki}) e^{\beta' Z_{ki'}(X_{ki})}} \right\}^{\Delta_{ki} I(\varepsilon_{ki}=1)} = \prod_{k=1}^{s} \prod_{i=1}^{n_k} \left\{ \frac{e^{\beta' Z_{ki}(X_{ki})}}{n_k S_k^{(0)}(\beta, X_{ki})} \right\}^{\Delta_{ki} I(\varepsilon_{ki}=1)},$$

where

$$S_k^{(p)}(\beta, t) = n_k^{-1} \sum_{i=1}^{n_k} Y_{ki}^*(t) Z_{ki}(t)^{\otimes p} e^{\beta' Z_{ki}(t)}, \ p = 0, 1, 2,$$

with $a^{\otimes 0} = 1$, $a^{\otimes 1} = a$, and $a^{\otimes 2} = aa^T$. Differentiating the log partial likelihood in β yield the following estimating equation:

$$U_1^*(\beta, t) = \sum_{k=1}^{s} \sum_{i=1}^{n_k} \int_0^t \left\{ Z_{ki}(u) - \overline{Z}_k(\beta, u) \right\} I(C_{ki} \geq u) \, dN_{ki}(u) \qquad (2)$$

Where $\overline{Z}_k(\beta, t) = S_k^{(1)}(\beta, t) / S_k^{(0)}(\beta, t)$. The estimator $\hat{\beta}$, which maximizes $L^*(\beta)$ may be obtained as a solution to $U_1^*(\beta, \tau) = 0$.

3.2 Weighted Estimating Equation for Right Censored Data

When classical right censoring is present, we can adapt inverse probability of censoring weighting techniques (Robins and Rotnitzky, 1992) to construct an unbiased estimating function, as proposed by Fine and Gray (1999). Let $G(\cdot)$ be the survival function of the censoring variable with highly stratified data and $G_k(\cdot)$ be the survival function of the censoring variable in the kth stratum with the regular stratification. The implicit weight $I(C_{ki} \geq t)$ with the CC data is replaced by $w_{ki}(t) = I(C_{ki} \geq T_{ki} \wedge t) G_k(t) / G_k(X_{ki} \wedge t)$, where $G_k(t) = \Pr(C_{ki} \geq t)$, $i = 1, \ldots, n_k$, for each k with regularly stratified data, assuming the censoring distribution is stratum dependent. For highly stratified data, the replacement would be $w_{ki}(t) = I(C_{ki} \geq T_{ki} \wedge t) G(t) / G(X_{ki} \wedge t)$, and $G(t) = \Pr(C_{ki} \geq t)$, $k = 1, \ldots, s$; $i = 1, \ldots, n_k$, assuming random vectors $(n_k, C_{k1}, \ldots, C_{kn_k})$ are i.i.d. across k, but allow the dependence in the censoring times in a given stratum. That is, arbitrary dependence among $C_{k1}, \ldots, C_{kn_k}$ would be allowed.

The difference in the assumptions for the two data regimes has to do with the consistency of the potential estimator of the censoring time distribution. For regularly stratified data, the size of each stratum goes to infinity. We can afford to allow for the dependence between censoring time C and strata K. Therefore, $G_k(\cdot)$ is used. If strata sizes are finite, we cannot consistently estimate the censoring distribution in each strata. Information must be pooled across strata via the assumption of a single G, as is employed with highly stratified data.

Because the distribution of the censoring random variable is unknown in either stratification regime, $G_k(\cdot)(G(\cdot))$ needs to be estimated: $\hat{G}_k(\cdot)$, the Kaplan–Meier estimate of the survival function of the censoring random variable in the kth stratum is used with regular stratification and $\hat{G}(\cdot)$, the Kaplan–Meier estimate of the survival function of the censoring random variable, is used with highly stratified data. Only patients from the kth stratum are used in the calculation of $\hat{w}_{ki}(t)$ for regularly stratified data because the stratum sizes go to infinity. In contrast, for highly stratified data, patients from all strata are used in $\hat{G}$.

The estimated weights are $\hat{w}_{ki}(t) = I(C_{ki} \geq T_{ki} \wedge t) \hat{G}_k(t) / \hat{G}_k(X_{ki} \wedge t)$ or $\hat{w}_{ki}(t) = I(C_{ki} \geq T_{ki} \wedge t) \hat{G}(t) / \hat{G}(X_{ki} \wedge t)$ such that $\hat{w}_{ki}(t) \to w_{ki}(t)$ as $n \to \infty$ for both stratification schemes. That is,

$$\widehat{w}_{ki}(t) = \begin{cases} 0 & \text{if right censored} \\ \widehat{G}_k(t|X_{ki}) \text{ or } \widehat{G}(t|X_{ki}) & \text{if failed} \\ 1 & \text{otherwise.} \end{cases}$$

The resulting weighted score function is:

$$U_1(\beta,t) = \sum_{k=1}^{s}\sum_{i=1}^{n_k}\int_0^t \left\{ Z_{ki}(u) - \frac{\sum_{i'=1}^{n_k}\widehat{w}_{ki'}(u)\,Y_{ki'}(u)\,Z_{ki'}(u)\,e^{\beta' Z_{ki'}(u)}}{\sum_{i'=1}^{n_k}\widehat{w}_{ki'}(u)\,Y_{ki'}(u)\,e^{\beta' Z_{ki'}(u)}} \right\}$$

$$\widehat{w}_{ki'}(u)\,dN_{ki}(u) = \sum_{k=1}^{s}\sum_{i=1}^{n_k}\int_0^t \left\{ Z_{ki}(u) - \widehat{Z}_k(\beta,u) \right\}\widehat{w}_{ki}(u)\,dN_{ki}(u), \tag{3}$$

where $\widehat{S}_k^{(p)}(\beta,t) = n_k^{-1}\sum_{i=1}^{n_k}\widehat{w}_{ki}(t)\,Y_{ki}(t)\,Z_{ki}(t)^{\otimes p}e^{\beta' Z_{ki}(t)}$,

$\widehat{Z}_k(\beta,t) = \widehat{S}_k^{(1)}(\beta,t)/\widehat{S}_k^{(0)}(\beta,t)$, $p=0,1,2$. The estimator of $\widehat{\boldsymbol{\beta}}$ is obtained by zeroing the estimating equation (3).

4. Inference

In this section, we consider the asymptotic properties of the estimators for different data regimes. In each case, the consistency of $\widehat{\boldsymbol{\beta}}$ comes first. Then the asymptotic normality of $n^{-\frac{1}{2}}\boldsymbol{U}_1^*(\beta_0,\tau)$ or $n^{-\frac{1}{2}}\boldsymbol{U}_1(\beta_0,\tau)$ is shown. Asymptotic normality of $n^{\frac{1}{2}}\left(\widehat{\beta}-\beta_0\right)$ is obtained by Taylor series approximation and the first two results. Variance estimation follows.

4.1 Regularly Stratified Data

In this scenario, the number of strata s is finite. When $n\to\infty$, $n_k\to\infty$, for each $k=1,\dots,s$.

We start from the CC data. Under the regularity conditions stated in Web Appendix A, we can show that $\widehat{\boldsymbol{\beta}}$ is consistent by adapting the consistency result of $\widehat{\boldsymbol{\beta}}$ in Andersen and Gill (1982). Then by Taylor series expansion and the consistency of $\widehat{\beta}$, $n^{\frac{1}{2}}\left(\widehat{\beta}-\beta_0\right)\approx \boldsymbol{\Omega}_r^{-1}\left\{ n^{-\frac{1}{2}}\boldsymbol{U}_1^*(\beta_0,\tau) \right\}$, where $\boldsymbol{\Omega}_r$ is the limit of the negative of the partial derivative matrix of the score function $\boldsymbol{U}_1^*(\beta,\tau)$ evaluated at $\boldsymbol{\beta}_0$. Clearly, $\boldsymbol{U}_1^*(\beta,t) = \sum_{k=1}^{s}\boldsymbol{U}_{1k}^*(\beta,t)$, where each component $\boldsymbol{U}_{1k}^*(\beta,t) = \sum_{i=1}^{n_k}\int_0^t\left\{ Z_{ki}(u) - \overline{Z}_k(\beta,u) \right\}I(C_{ki}\ge u)\,dN_{ki}(u)$ and by Fine and Gray (1999), as $n_k\to\infty$, $n_k^{-\frac{1}{2}}\boldsymbol{U}_{1k}^*(\beta_0,\tau)\xrightarrow{D}N(0,\boldsymbol{\Omega}_{rk})$ with

$$\boldsymbol{\Omega}_{rk} = \int_0^\tau \left\{ s_k^{(2)}(\beta_0,t)/s_k^{(0)}(\beta_0,t) - e_k(\beta_0,t)^{\otimes 2} \right\}\times s_k^{(0)}(\beta_0,t)\,\lambda_{0k}(t)\,dt,$$

where $s_k^{(p)}(\beta,t) = \lim_{n_k\to\infty}n_k^{-1}\sum_{i=1}^{n_k}Y_{ki}^*(t)\,Z_{ki}(t)^{\otimes p}e^{\beta' Z_{ki}}$, and $e_k(\beta,t) = s_k^{(1)}(\beta,t)/s_k^{(0)}(\beta,t)$, $p=0,1,2$. It follows that $\sum_{k=1}^{s}\pi_k\boldsymbol{\Omega}_{rk}=\boldsymbol{\Omega}_r$, Where $p_k = n_k/n\to\pi_k$. Thus, as $n\to\infty$, $n^{-\frac{1}{2}}\boldsymbol{U}_1^*(\beta_0,\tau) = \sum_{i=1}^{s}p_k^{\frac{1}{2}}n_k^{-\frac{1}{2}}\boldsymbol{U}_{1k}^*(\beta_0,\tau)\xrightarrow{D}N(0,\boldsymbol{\Omega}_r)$. Hence, $n^{-\frac{1}{2}}\left(\widehat{\beta}-\beta_0\right)\xrightarrow{D}N\left(0,\boldsymbol{\Omega}_r^{-1}\right)$.

A consistent estimate of the covariance matrix for the regression coefficients can be obtained by replacing the unknown quantities with their observed values. We have

$$\widehat{\boldsymbol{\Omega}}_r = \frac{1}{n}\sum_{k=1}^{s}\sum_{i=1}^{n_k}\Delta_{ki}I\left(\varepsilon_{ki}=1\right)\times\left[\frac{\boldsymbol{S}_k^{(2)}\left(\widehat{\beta},X_{ki}\right)}{\boldsymbol{S}_k^{(0)}\left(\widehat{\beta},X_{ki}\right)}-\left\{\frac{\boldsymbol{S}_k^{(1)}\left(\widehat{\beta},X_{ki}\right)}{\boldsymbol{S}_k^{(0)}\left(\widehat{\beta},X_{ki}\right)}\right\}^{\otimes 2}\right].$$

When the data are right censored, by applying results of weighted score function for incomplete data (Fine and Gray, 1999) to each component of $\boldsymbol{U}_1(\beta,t):\boldsymbol{U}_{1k}(\beta,t)=\sum_{i=1}^{n_k}\int_0^t\left\{Z_{ki}(u)-\widehat{Z}_k(\beta,u)\right\}\widehat{w}_{ki}(u)\,dN_{ki}(u),$ the asymptotic properties for right censored stratified data may be established. One may show that as $n\to\infty, n^{-1}\left\{\boldsymbol{U}_1(\beta,\tau)-\boldsymbol{U}_1^*(\beta,\tau)\right\}=\sum_{i=1}^{s}p_k n_k^{-1}\left\{\boldsymbol{U}_{1k}(\beta,\tau)-\boldsymbol{U}_{1k}^*(\beta,\tau)\right\}$ goes in probability to 0 uniformly for $\boldsymbol{\beta}$ in a compact neighborhood of $\boldsymbol{\beta}_0$. Therefore, $\widehat{\boldsymbol{\beta}}$, the solution to $\boldsymbol{U}_1(\boldsymbol{\beta},\tau)=0$ is consistent for $\boldsymbol{\beta}_0$. Furthermore, for each $n_k^{-\frac{1}{2}}\boldsymbol{U}_{1k}(\beta_0,\tau)\xrightarrow{D}N\left(0,\sum_{rk}\right)$, where $\Sigma_{rk}=\mathrm{E}\{(\boldsymbol{\eta}_{ki}+\boldsymbol{\psi}_{ki})^{\otimes 2}\}$. The details of $\boldsymbol{\eta}_{ki}$ and $\boldsymbol{\psi}_{ki}$ for each k are omitted here because they are identical to $\boldsymbol{\eta}_i$ and $\boldsymbol{\psi}_i$ in Fine and Gray (1999), with the added subscript k. Because $\boldsymbol{U}_{1k}$ are independent, $n^{-\frac{1}{2}}\boldsymbol{U}_1(\beta_0,\tau)=\sum_{k=1}^{s}p^{\frac{1}{2}}n_k^{-\frac{1}{2}}\boldsymbol{U}_{1k}(\beta_0,\tau)$ is asymptotically normally distributed as $N(0,\Sigma_r)$, where $\sum_r=\sum_{k=1}^{s}\pi_k\sum_{rk}$. Similar to CC data, $n^{\frac{1}{2}}\left(\widehat{\beta}-\beta_0\right)\approx\boldsymbol{\Omega}_r^{-1}\left\{n^{-\frac{1}{2}}\boldsymbol{U}_1(\beta_0,\tau)\right\}$, where $\boldsymbol{\Omega}_r^{-1}$ has the same form as the variance in the CC case. Hence, the distribution of $n^{\frac{1}{2}}\left(\widehat{\beta}-\beta_0\right)$ is asymptotically normal with covariance matrix $\boldsymbol{\Omega}_r^{-1}\sum_r\boldsymbol{\Omega}_r^{-1}$.

To estimate the covariance matrix for the regression coefficients, we need to find a consistent estimator of $\boldsymbol{\Omega}_r$ and Σ_r. The estimator of $\boldsymbol{\Omega}_r$ is $\widehat{\boldsymbol{\Omega}}_r$ with $\widehat{\boldsymbol{S}}_k^{(p)}$ replacing $\boldsymbol{S}_k^{(p)}$ for each k, $p=0,1,2$. Each component of Σ_r, Σ_{rk} can be estimated with the empirical covariance matrix $\widehat{\sum}_{rk}=n_k^{-1}\sum_i\left(\widehat{\eta}_{ki}+\widehat{\psi}_{ki}\right)^{\otimes 2}$. For brevity, we do not show the details of $\widehat{\boldsymbol{\eta}}_{ki}$ and $\widehat{\boldsymbol{\psi}}_{ki}$, which can be obtained by adding subscript k to $\widehat{\boldsymbol{\eta}}_i$ and $\widehat{\boldsymbol{\psi}}_i$ in Fine and Gray (1999).

Therefore, the distribution of $n^{\frac{1}{2}}\left(\widehat{\beta}-\beta_0\right)$ can be approximated by a normal distribution with variance $n\left\{\sum_{k=1}^{S}n_k\widehat{\boldsymbol{\Omega}}_{rk}\right\}^{-1}\left\{\sum_{k=1}^{S}n_k\widehat{\sum}_{rk}\right\}\left\{\sum_{k=1}^{S}n_k\widehat{\boldsymbol{\Omega}}_{rk}\right\}^{-1}$.

4.2 Highly Stratified Data

In this scenario, the strata size n_k is finite, $k=1,\ldots,s$, $s\to\infty$ as $n\to\infty$.

As in the regularly stratified data, we consider the CC situation first. In Web Appendix B, we show $\widehat{\boldsymbol{\beta}}$ is consistent under regularity conditions listed in Web Appendix A. We can further show asymptotic normality of $n^{-\frac{1}{2}}\boldsymbol{U}_1^*(\beta_0,\tau)$ as follows. Because $M_{ki}^*(\beta,t)=\int_0^t I(C_{ki}\geq u)\,dN_{ki}(t)-\int_0^t Y_{ki}^*(u)\lambda_{k0}(u)\,e^{\beta'Z_{ki}(u)}du$ is a martingale for CC data filtration $F^*(t)=\sigma\left\{I(C_{ki}\geq u),I(C_{ki}\geq u)N_{ki}(u),Y_{ki}^*(u),Z_{ki}(u),u\leq t,k=1,\ldots,s;i=1,\ldots,n\right\}$ (Fine and Gray, 1999), we can reexpress $\boldsymbol{U}_1^*(\beta,t)$ as a martingale-type estimation equation:

$$U_1^*(\beta, t) = \sum_{k=1}^{S} \sum_{i=1}^{n_k} \int_0^t \left\{ Z_{ki}(u) - \bar{Z}_k(\beta, u) \right\} dM_{ki}^*(\beta, u).$$

Applying the martingale central limit theorem (Rebolledo, 1980), $n^{-\frac{1}{2}} U_1^*(\beta_0, \cdot)$ converges in distribution to a continuous Gaussian process. At time $t=\tau$, $n^{-\frac{1}{2}} U_1^*(\beta_0, \tau) \xrightarrow{P} N(0, \Omega_h)$, where

$$\Omega_h = \lim_{n \to \infty} n^{-1} \sum_{k=1}^{S} \mathrm{E} \left[\sum_{k=1}^{n_k} \int_0^\tau \left\{ Z_{ki}(\mathbf{u}) - \bar{Z}_k(\beta_0, \mathbf{u}) \right\}^{\otimes 2} \times Y_{ki}^*(\mathbf{u}) \, e^{\beta'_0 Z_{ki}(\mathbf{u})} \lambda_{k0}(\mathbf{u}) \, d\mathbf{u} \right]. \qquad (4)$$

Some algebra proves that equation (4) satisfies

$$\Omega_h = \lim_{n \to \infty} n^{-1} \sum_{k=1}^{S} \mathrm{E} \left[\sum_{i=1}^{n_k} \int_0^\tau \left\{ Z_{ki}(\mathbf{u}) - \bar{Z}_k(\beta_0, \mathbf{u}) \right\}^{\otimes 2} \times Y_{ki}^*(\mathbf{u}) \, e^{\beta'_0 Z_{ki}(\mathbf{u})} \frac{d\bar{N}_k(\mathbf{u})}{S_k^{(0)}(\beta_0, \mathbf{u})} \right].$$

Similar to the regularly stratified data, Ω_h is equal to the limit of the negative of the partial derivative matrix of the score function evaluated at β_0. In addition, $n^{\frac{1}{2}}(\widehat{\beta} - \beta_0) \approx \Omega_h^{-1} \left\{ n^{-\frac{1}{2}} U_1^*(\beta_0, \tau) \right\}$. Thus, $n^{-\frac{1}{2}}(\widehat{\beta} - \beta_0) \xrightarrow{D} N(0, \Omega_h^{-1})$. Note that $\bar{m}\Omega_h$ equals to I defined in condition (4) of Web Appendix A, where $\bar{m}$ is the average strata size as $n \to \infty$.

Replacing the unknown quantities in Ω_h with their observed values gives

$$\widehat{\Omega}_h = \frac{1}{n} \sum_{k=1}^{S} \sum_{i=1}^{n_k} \Delta_{ki} I(\varepsilon_{ki}=1) \times \left[\frac{S_k^{(2)}(\widehat{\beta}, X_{ki})}{S_k^{(0)}(\widehat{\beta}, X_{ki})} - \left\{ \frac{S_k^{(1)}(\widehat{\beta}, X_{ki})}{S_k^{(0)}(\widehat{\beta}, X_{ki})} \right\}^{\otimes 2} \right],$$

an analogous formulation to that for $\widehat{\Omega}_r$.

When the data are right censored, $\widehat{\beta}$ is consistent as shown in Web Appendix C.1. We also establish that $n^{\frac{1}{2}}(\widehat{\beta} - \beta_0) \approx \widetilde{\Omega}_h^{-1} \left\{ n^{-\frac{1}{2}} U_1(\beta_0, \tau) \right\}$, where

$$\widetilde{\Omega}_h = \lim_{n \to \infty} n^{-1} \sum_{k=1}^{S} \mathrm{E} \left[\sum_{i=1}^{n_k} \int_0^\tau \left\{ Z_{ki}(\mathbf{u}) - \tilde{Z}_k(\beta_0, \mathbf{u}) \right\}^{\otimes 2} \times w_{ki}(\mathbf{u}) Y_{ki}(\mathbf{u}) \, e^{\beta'_0 Z_{ki}(\mathbf{u})} \frac{d\bar{N}_k(\mathbf{u})}{\tilde{S}_k^{(0)}(\beta_0, \mathbf{u})} \right].$$

We show in Web Appendix C.2 that $S^{-\frac{1}{2}} U(\beta_0, \tau) = S^{-\frac{1}{2}} \sum_{k=1}^{S} (\eta_k + \psi_k) + o_p(1)$ is asymptotically normal with co-variance matrix $\Sigma_h = \mathrm{E}\{(\eta_k + \psi_k)(\eta_k + \psi_k)^T\}$, where

$$\psi_k = \sum_{i=1}^{n_k} \int_0^\infty \mathrm{q}\,(\mathbf{y})\,/\,\{\overline{\mathbf{m}}\pi\,(\mathbf{y})\}\,\mathbf{dM^c_{ki}}\,(\mathbf{y})\,,$$

$$\eta_k = \sum_{i=1}^{n_k} \int_0^t \left\{Z_{ki}\,(u) - \tilde{Z}_k\,(\beta_0, u)\right\} w_{ki}\,(u)\,dN_{ki}\,(u)\,,$$

$$\pi\,(y) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^n I\,(X_t \geq y)\,,$$

$$\mathrm{q}\,(\mathbf{y}) = \mathrm{E}\left[\sum_{\mathbf{i=1}}^{\mathbf{n_k}} \sum_{\mathbf{j=1}}^{\mathbf{n_k}} \int_\mathbf{0}^\mathbf{t} \mathbf{A_{kj}}\,(\beta_\mathbf{0}, \mathbf{u}) \times \mathbf{I}\left(\mathbf{X_{kj}} < \mathbf{y} \leq \mathbf{u}\right) \mathbf{w_{kj}}\,(\mathbf{u})\,\mathbf{dN_{ki}}\,(\mathbf{u})\right]\,,$$

, and

$$A_{kj}\,(\beta_0, u) = \frac{Y_{kj}\,(u)\,e^{\beta'_0 Z_{kj}(t)}}{n\tilde{S}_k^{(0)}\,(\beta_0, u)} \times \left\{Z_{kj}\,(u) - \tilde{Z}_k\,(\beta_0, u)\right\}\,.$$

Here, $\tilde{\mathbf{Z}}_k(\boldsymbol{\beta}_0, \cdot)$ is $\hat{\mathbf{Z}}_k(\boldsymbol{\beta}_0, \cdot)$ with $\hat{w}_{ki}(\cdot)$ being replaced by $w_{ki}(\cdot)$, and $M_{ki}^c(\cdot)$ is the martingale associated with the censoring process.

It follows that $n^{\frac{1}{2}}\left(\hat{\beta} - \beta_0\right)$ is asymptotically normal with covariance matrix $\tilde{\boldsymbol{\Omega}}_h^{-1}\left(\overline{m}^{-1}\sum_h\right)\tilde{\boldsymbol{\Omega}}_h^{-1}$.

The estimator of $\tilde{\boldsymbol{\Omega}}_h$ is $\hat{\boldsymbol{\Omega}}_h$ with $\widetilde{S}_k^{(p)}$ replacing $S_k^{(p)}$ for all k and p as in the regularly stratified case. The inner matrix in the variance, Σ_h, can be estimated empirically by $1/S\sum_{k=1}^S\left(\widehat{\eta}_k + \widehat{\psi}_k\right)^{\otimes 2}$, where

$$\widehat{\eta}_k = \sum_{i=1}^{n_k} \int_0^\infty \left\{Z_{ki}\,(u) - \widehat{Z}_k\left(\widehat{\beta}, u\right)\right\} \widehat{w}_{ki}\,(u)\,dN_{ki}\left(\widehat{\beta}, u\right)\,,$$

$$\widehat{\psi}_k = \sum_{i=1}^{n_k} \int_0^\infty \widehat{Q}\left(\widehat{\beta}, y\right)\left\{\sum_{l=1}^n I\,(X_l \geq y)\,/\,s\right\}^{-1} d\widehat{M}_{ki}^c\,(y)\,,$$

$$\widehat{M}^c (y) = I (X \le y, \Delta{=}0) - \int_0^y I (X \ge t) \, d\widehat{\wedge}^c (t),$$

$$\widehat{\wedge}^c (t) = \int_0^t \left\{ \sum_{l=1}^n I (X \ge y) \right\}^{-1} \sum_{l=1}^n dI (X \le u, \Delta{=}0),$$

$$\widehat{Q}\left(\widehat{\beta}, y\right) = \mathbf{s}^{-1} \sum_{k=1}^{\mathbf{S}} \sum_{i=1}^{n_k} \sum_{j=1}^{n_k} \int_0^\infty \widehat{A}_{kj}\left(\widehat{\beta}, u\right) I\left(X_{kj}{<}y \le u\right) \times \widehat{w}_{kj}(u)\, \widehat{w}_{ki}(u)\, dN_{ki}(u),$$

and $\hat{A}_{kj}$ is defined analogous to A_{kj} with $\mathbf{Z}^{\tilde{}}$ replaced by $\hat{\mathbf{Z}}$ and S˜ replace by Ŝ.

This variance estimator can be unstable in small sample sizes, owing to variability in $\hat{G}$ in the tails and the small within cluster sample sizes. As an alternative, bootstrap variance estimation is introduced for highly stratified right censored data. Adapting the simple bootstrap sampling for censored data (Efron, 1981), we have the following scheme:

1. Draw a bootstrap sample by independently sampling s times with replacement from the s strata. This corresponds to drawing repeated from the empirical distribution of the strata, which puts equal mass, $1/s$, on each stratum;

2. Let data* represent this artificial data set, calculate $\hat{\boldsymbol{\beta}}^*$;

3. Independently repeat the above steps B times, obtaining B regression coefficient estimates, denoted $\widehat{\beta}_i^*, i{=}1, \ldots, B$;

4. Calculate the sample standard deviation of the $\widehat{\beta}_i^*, i{=}1, \ldots, B$, and use this as an estimate of the standard error of $\hat{\boldsymbol{\beta}}$.

This resampling approach is valid under the assumption that data within cluster are i.i.d. across clusters.

5. Predicting Cumulative Incidence

To predict cumulative incidence at time t for a patient with covariates $\mathbf{Z} = \mathbf{z}$, we need an estimator of the baseline cumulative subdistribution hazard. This estimator can be obtained using a variation of Breslow's (1974) estimator for regularly stratified data, but is infeasible for highly stratified data due to finite strata sizes. In this section, we will briefly discuss the estimators of the cumulative subdistribution hazard for regularly stratified data and the associated cumulative incidence estimators for individuals with certain covariate values.

With CC data, for each stratum k, the Breslow type estimator is

$$\widehat{\wedge}_{k0}^* (t) = \frac{1}{n_k} \sum_{i=1}^{n_k} \int_0^t \frac{1}{S_k^{(0)}\left(\widehat{\beta}, u\right)} \Big| (C_{ki} \ge u)\, dN_{ki}(u), \ k{=}1, \ldots, s,$$

where $\hat{\boldsymbol{\beta}}$ is defined in Section 4. Thus, the cumulative subdistribution hazard for a patient with covariates $\mathbf{Z} = \mathbf{z}$ at time t is estimated by

$$\widehat{\Lambda}_k^* (t, z) = \int_0^t \exp\left\{\widehat{\beta}' (u)\right\} d\widehat{\Lambda}_{ki} (u) = \frac{1}{n_k} \sum_{i=1}^{n_k} \int_0^t \frac{\exp\left\{\widehat{\beta}' z(\mathbf{u})\right\}}{S_k^{(0)} \left(\widehat{\beta}, u\right)} \big| (C_{ki} \geq u) \, dN_{ki} (u) \,, \; k=1, \ldots, s.$$

When the data are right censored, our estimator of the baseline cumulative subdistribution hazard for stratum k is

$$\widehat{\Lambda}_{k0} (t) = \frac{1}{n_k} \sum_{i=1}^{n_k} \int_0^t \frac{\widehat{w}_{ki} (u)}{\widehat{S}_k^{(0)} \left(\widehat{\beta}, u\right)} dN_{ki} (u) \,. \tag{5}$$

This is essentially the Breslow estimator, after incorporating the inverse probability of censoring weights to account for independent right censoring. The corresponding cumulative subdistribution hazard is estimated by

$$\widehat{\Lambda}_k (t, z) = \frac{1}{n_k} \sum_{i=1}^{n_k} \int_0^t \frac{\exp\left\{\widehat{\beta}' z(\mathbf{u})\right\}}{\widehat{S}_k^{(0)} \left(\widehat{\beta}, u\right)} \widehat{w}_{ki} (u) \, dN_{ki} (u) \,, \; k=1, \ldots, s.$$

Now that we have obtained the estimators of the cumulative subdistribution hazards at time t for a patient with covariates $\mathbf{z}$, we can predict the cumulative incidence by $\widehat{F}_{1k} (t, z) = 1 - \exp\left\{-\widehat{\Lambda}_k^* (t, z)\right\}$ for CC data, or $\hat{F}_{1k} (t, \mathbf{z}) = 1 - \exp\{-\hat{\Lambda}_k (t, \mathbf{z})\}$ for right censored data. Confidence intervals and bands for $\widehat{F}_{1k} (t, z) , \widehat{\Lambda}_k^* (t, z$ or $\widehat{\Lambda}_k (t, z)$ can be constructed along the lines of Fine and Gray (1999). They are easily obtained by adding stratum information k to the earlier results. The details are omitted.

As stated previously, we are not able to predict cumulative incidence using highly stratified data. We can only evaluate the effect of risk factors on the cumulative incidence.

6. Simulation Experiments

Numerical investigations were conducted to assess the performance of the proposed weighted estimation approach. We compared the estimators from the stratified weighted score function (3) to the stratified censoring complete (CCS) estimators (described in Section 3.1) and the CC estimators (described in Section 3.2 of Fine and Gray, 1999). The objective is to assess potential biases in ignoring the stratification information when naively using the unstratified analyses versus the stratified analyses, which appropriately account for such stratification.

In all sets of simulations, data were generated repeatedly 1000 times. Within each replicate, we employed the algorithm used by Fine and Gray (1999) to generate the data for each stratum. Two competing risks were considered. The subdistribution for type 1 failure was given by $F_{1k} (t; Z = \Pr\left(T_{ki} \leq t, \varepsilon_{ki}=1 | Z_{ki} = 1 - \left\{1 - p + p e^{-\rho_k t \gamma k}\right\}^{\exp(\beta'_1 Z_{ki})}\right.$, which is a weibull mixture with mass $1 - p$ at ∞ when $\mathbf{Z}_{ki} = 0$, and uses the proportional subdistribution hazards for nonzero covariate values. The subdistribution for type 2 failure was then obtained by taking $\Pr(\varepsilon_{ki} = 2 \mid \mathbf{Z}_{ki}) = 1 - \Pr(\varepsilon_{ki} = 1 \mid \mathbf{Z}_{ki})$ and using an exponential

distribution with rate $\exp\left(\beta'_2 Z_{ki}\right)$ for $\Pr(T_{ki} \quad t| \varepsilon_{ki} = 2, \mathbf{Z}_{ki})$. Censoring times were generated from the uniform [a,b] distribution.

6.1 Simulation of Regularly Stratified Data

We present two sets of simulations to demonstrate the performance of the weighted estimation function when the number of strata is fixed. In both sets of simulations, we considered 3 strata; covariates are i.i.d. given stratum; $(\rho_1, \rho_2, \rho_3) = (1,.3,2)$, $(\gamma_1, \gamma_2, \gamma_3) = (.5,2,1)$; and $p = 0.6$.

First, we generated data by assuming the true parameter values (β_1, β_2) to be $(0, -0.5)$ for normal covariates, which have unit variance and varying means of 0, 1, or 2; and assuming (β_1, β_2) to be $(0, 1)$ for Bernoulli covariates with means of .3, .5, or .7 for the three strata, respectively. Censoring times were generated from a uniform [a,b] distribution with a and b being specified to reach the targeted percentage of censored observations. We used two different percentages of censoring for each. Table 1 gives the empirical sizes of CC, CCS, and stratified weighted (WS) score tests for sample sizes of 100, 250, 500, and 1000 at the nominal level of 0.05. As the sample size increases, the empirical size of the unstratified test (CC) deviates substantially from the nominal level, especially for the normal covariates. The stratified weighted tests and the CCS tests both achieve close to the nominal level.

Next, we changed (β_1, β_2) to $(0.5, -0.5)$ for normal covariates; and $(1,1)$ for Bernoulli covariates, keeping other items unchanged. The sample size was 250 for each replicate. Table 2 gives $E(\hat{\beta}_1)$, estimated with the average of the $\hat{\beta}_1$ from the 1000 replicates; $var(\hat{\beta}_1)$, estimated with the empirical variance of $\hat{\beta}_1$; and $E(\hat{var})$, the average of the variance estimators of $\hat{\beta}_1$. As expected, the performances of the two stratified approaches are better than the unstratified approach, which exhibits substantial bias. Both of the stratified approaches gave very similar results, with small biases and similar variances.

6.2 Simulation for Highly Stratified Data

We assumed that the number of strata was 50; the strata sizes were uniformly distributed from $\{3,4,5\}$; $(\rho_1,\ldots,\rho_k,\ldots,\rho_{50}) = (0.1,\ldots, 0.1k,\ldots, 5)$, $(\gamma_1,\ldots,\gamma_k,\ldots,\gamma_{50}) = (5,\ldots, 5 -0.1k, \ldots, 0.1)$; and $p = 0.6$. The covariates were two dimensional. The first component was independently distributed as a normal or Bernoulli covariate. The mean of the distributions was strata dependent with three levels as in the regularly stratified case. The second component was i.i.d. uniformly distributed independent of strata. We considered the true parameter values to be $(\beta_{11}, \beta_{12}, \beta_{21}, \beta_{22}) = (0, 1, -0.5, .5)$ and $(\beta_{11}, \beta_{12}, \beta_{21}, \beta_{22}) = (1, 1, -0.5, .5)$, respectively.

Table 3 gives the results for β_{11} and β_{12} for the various approaches. The weighted stratified (WS) approach for right censored data and the CCS approach produce very similar and precise estimators of the parameters and their variances, for both β_{11} and β_{12}. For right censored data, both the model based plug-in (WSh) and bootstrap (WSb) variance estimators $(E(\hat{var}))$ are provided for the weighted estimating equation. They have correlations of between 96% and 98% in the four settings. The former leads to slightly lower empirical coverage than the latter, although both methods, as well as CCS inferences, have coverages which are close (within 0.02) to the nominal level. In contrast, the approach without stratification (CC) provides biased results and coverages that are greatly reduced. When $\beta_{11} = 0, \hat{\beta}_{11}$ departs substantially from the truth; when $\beta_{11} = 1, \hat{\beta}_{12}$ and β_{12} are very discrepant.

7. Real Data Examples

7.1 The Eastern Cooperative Oncology Group Study

In this study, there were 167 eligible patients. Of the 82 patients on placebo, 59 had breast cancer recurrence, 19 died without recurrence, and 4 were censored; of the 85 patients on tamoxifen, 42 had breast cancer recurrence, 23 died without recurrence, and 20 were censored. In addition to treatment group to be considered, there were three prognostic factors: number of positive nodes, tumor size, and age at treatment. We were interested in the cumulative incidence function of breast cancer recurrence for the two treatment groups with the competing event being death without recurrence.

The data were previously analyzed by Fine and Gray (1999). The model they suggested for breast cancer recurrence allows the subdistribution hazards ratio of treatment to be quadratic in time due to substantial lack of fit in the proportional subdistribution hazards model. The other covariates were considered in the analysis in the form of linear proportional hazards terms, which did not suggest lack of fit.

Based on the above, we considered model (1) with covariates for ith patient in kth group $\mathbf{Z}_{ki}$ = $(Z_{1ki}, Z_{2ki}, Z_{3ki})$ = $(log(nodes), tumor\ size, age)_{ki}$, where k =1, 2 denotes the treatment group (1 for placebo and 2 for tamoxifen); i = 1,…, 82 for k = 1; i = 1,…, 85 for k = 2. Instead of being a covariate, treatment was a stratification variable. The estimated coefficients and standard errors for this model, as well as those from Fine and Gray's analysis are displayed in Table 4. The coefficient estimates and standard errors of the 3 covariates from the two models agree. The log of the number of nodes and the tumor size are significant for the subdistribution of the breast cancer recurrence, whereas age is insignificant.

Using the estimators described in Section 5, we are able to compare the estimated baseline cumulative subdistribution hazards for the two treatment groups when covariates $\mathbf{Z} = \mathbf{0}$. Figure 1a depicts the estimated baseline cumulative subdistribution hazards of breast cancer recurrence for each group using our stratified competing risk regression model described above. Figure 1b gives the log of the ratio of baseline cumulative subdistribution hazards. The nonconstancy of the curve in Figure 1b is clear evidence of nonproportionality of the subdistribution hazards of treatment. One could argue that interpreting the treatment effect is more intuitive using such plots, as opposed to fitting more complex models with time-dependent covariates to capture the nonproportional treatment effect.

7.2 Acute Myeloid Leukemia Data

The data arise from an ongoing registry by the European Blood and Marrow Transplant Group. The event of interest was the time from graft to the first occurrence of either acute GvHD grade 2 or chronic GvHD. Death and relapse without GvHD are the competing causes of failure. Katsahian et al. (2006) proposed a frailty model for the subdistribution hazard to test the prognostic factors while adjusting for the center effect. A subset of the data was used in their frailty model analyses, with the reference date being 1 January 2002 as well as the following inclusion criteria for patients: (1) received either genoidentical or matched unrelated donor (MUD) stem cell transplant; (2) were more than 16 years old at time of transplant; (3) had acute myeloid leukemia in first complete remission; (4) received a transplant between 1 January 1994 and 31 December 2004; and (5) did not receive reduced intensity regimen nor T cell depleted transplant. Centers with only one patient enrolled were excluded. A total of 1022 patients from 121 clusters was analyzed.

In our analysis, we used the same registry, but with data extracted up to July 2008, while maintaining the inclusion criteria from Katsahian et al. (2006). The median follow-up was

1250 days, comprised of patients still alive without relapse and disease. Thus, we have a total of 2952 patients from 244 centers, with 1385 GvHD events and 629 competing causes of failure observed. The median patients per center was 6. About 1/3 of centers had only two or three patients.

Because the patient populations are remarkably different across centers, which was evident from Katsahian et al. (2006), we consider the highly stratified Fine–Gray model (1), stratifying on center. The four covariates are specified as in Katsahian et al. (2006). Let $\mathbf{Z}_{ki}$ = $(Z_{1ki}, Z_{2ki}, Z_{3ki}, Z_{4ki})$ for the ith subject in the kth center, where $k = 1,\ldots, 244$; $i = 1,\ldots, i_k$; and $i_k = 2,\ldots, 92$. Here, Z_{1ki} = I(female donor to male recipient match), Z_{2ki} = I(source of stem cells is peripheral blood), Z_{3ki} =I(French-American-British [FAB] classification of acute myeloid leukemia is M5, M6, or M7), and Z_{4ki} = I(type of transplant is matched unrelated donor). We first considered univariate models (1), similarly to Katsahian et al. (2006), followed by a multiple covariate analysis. Unstratified Fine–Gray models were also applied.

The coefficient estimates are reported in Table 5. In all cases, sex matching between donor and recipient (female donor to male recipient versus others) is a significant prognostic factor in the subdistribution hazard of GvHD occurrence, whereas the other three factors are not significant. Despite similar overall conclusions from the unstratified and stratified analysis, attention needs to be paid to some of the covariate estimates. For source of stem cells (peripheral blood or bone marrow), the coefficient estimates from the two approaches have opposite sign, suggesting an interaction between stem cell transplants and centers. We note that the univariate model (called simple stratified regression) and the model with multiple covariates (called multiple stratified regression) give very similar results from both an unstratified and stratified approach. This suggests that there is modest confounding of the covariate effects under consideration by heterogeneity across centers.

8. Concluding Remarks

Recently, Ruan and Gray (2008) proposed a generic Kaplan– Meier multiple imputation method that recovers the missing potential censoring information for the analysis of cumulative incidence functions using standard analysis, which can potentially be applied in the setting of stratified analysis. Although the methods performed well empirically, the statistical properties of the approach were not established. Such imputation-based procedures have not been used very often in survival settings, in part because of the ad hoc nature of the resulting inferences and a lack of understanding regarding when such inferences are valid. A goal of this article was to develop rigorous methodology for stratified Fine–Gray models, similar to that for stratified Cox models for independently censored data.

We proposed both bootstrap and plug-in formulas to obtain variance estimators for the highly stratified situation. The closed form variance estimator does not require estimation of the baseline hazard in each stratum, unlike for regularly stratified data and the original Fine and Gray (1999) model. This is accomplished by employing an alternative derivation of the influence function, which is not applicable in the other settings. In the simulation studies, we find that with realistic sample sizes, the bootstrap variance estimator slightly outperforms the closed form estimator. General bootstrap theory should be valid with highly stratified data, assuming the data within cluster (including cluster size) are i.i.d. across strata. A rigorous proof would entail more careful consideration of the regularity conditions.

There are alternative modeling strategies to tackle the same inferential problem. In our applications, the covariate effects are of primary interest, with the correlations within clusters serving as a nuisance. For situations where the correlations are of genuine interest,

the random effects model introduced by Katsahian et al. (2006) might be of greater utility, where the frailty is introduced on the stratum level hazard. Scheike et al. (2010) proposed a closely related frailty model, which implies a semiparametric additive model for the marginal cumulative incidence function. Another possibility is to model the covariate effects unconditionally on stratum, without specifying the correlation structure. The model proposed by Chen et al. (2008) assumes common baseline subdistribution hazards within strata but different subdistribution hazards across covariate groups instead of across strata. The goal is to test the equivalence of such hazards nonparametrically in the two sample set-up, adjusting for correlations within clusters. Such marginal approaches require further development to accommodate general regression models. This work is beyond the scope of the current article but merits further research.

The implementation of the proposed methodology may, in some cases, be carried out using existing software. For complete censored competing risks data, one can perform a stratified Cox regression analysis of a modified data set where individuals observed to fail from causes other than the cause of interest (say 1) are given a censored observation at the time of the observed censoring time (Andersen, Abildstrom, and Rosthøj, 2002). The R function for standard Cox regression, COXPH with the option STRATA, can be applied to this modified data set. For censored competing risks data, we have designed a R function stratified competing risks regression (CRRS) for the weighted estimating equation procedures described in Section 4; see Supplementary Materials.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

Andersen PK, Gill RD. Cox's regression model for counting processed: A large sample study. Annals of Statistics. 1982; 10:1100–1120.

Andersen PK, Abildstrom SZ, Rosthøj S. Competing risks as a multi-state model. Statistical Methods in Medical Research. 2002; 11:203–215. [PubMed: 12040697]

Breslow NE. Covariate analysis of censored survival data. Biometrics. 1974; 30:89–99. [PubMed: 4813387]

Chen BE, Kramer JL, Greene MH, Rosenberg PS. Competing risks analysis of correlated failure time data. Biometrics. 2008; 64:172–179. [PubMed: 17680835]

Cox DR. Regression models and life tables (with discussion). Journal of the Royal Statistical Society. 1972; 34:187–200.Series B

Efron B. Censored data and the bootstrap. Journal of the American Statistical Association. 1981; 76:312–319.

Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. Journal of the American Statistical Association. 1999; 94:496–509.

Gray RJ. A class of K-sample tests for comparing the cumulative incidence of a competing risk. Annals of Statistics. 1988; 16:1141–1154.

Gross ST, Huber G. Matched pair experiments: Cox and maximum likelihood estimation. Scandinavian Journal of Statistics. 1987; 14:27–41.

Kalbfleisch, JD.; Prentice, RL. The Statistical Analysis of Failure Time Data. New York: Wiley; 2002.

Katsahian S, Resche–Rigon M, Chevret S, Porcher R. Analysing multicenter competing risks data with a mixed proportional hazards model for the subdistribution. Statistics in Medicine. 2006; 25:4267–4278. [PubMed: 16960919]

Klein JP, Andersen PK. Regression modeling of competing risks data based on pseudo-values of the cumulative incidence function. Biometrics. 2005; 61:223–229. [PubMed: 15737097]

Lee, EW.; Wei, LJ.; Amato, DA. Cox-type regression analysis for large numbers of small groups of correlated failure time observations. In: Klein, JP.; Goel, PK., editors. Survival Analysis: State of the Art. Dordrecht, The Netherlands: Kluwer; 1992. p. 237-247.

Pepe MS. Inference for events with dependent risks in multiple endpoint studies. Journal of the American Statistical Association. 1991; 86:770–778.

Rebolledo R. Central limit theorems for local martingales. Probability Theory and Related Fields. 1980; 51:269–286.

Robins, JM.; Rotnitzky, A. Recovery of information and adjustment for dependent censoring using surrogate markers. In: Jewell, N.; Dietz, K.; Farewell, V., editors. AIDS Epidemiology—Methodological Issues. Boston: Birkhauser; 1992. p. 297-331.

Ruan PK, Gray RJ. Analyses of cumulative incidence functions via non-parametric multiple imputation. Statistics in Medicine. 2008; 27:5709–5724. [PubMed: 18712779]

Scheike TH, Zhang MJ. Flexible competing risks regression modeling and goodness-of-fit. Lifetime Data Analysis. 2008; 14:464–483. [PubMed: 18752067]

Scheike TH, Zhang MJ, Gerds T. Predicting cumulative incidence probability by direct binomial regression. Biometrika. 2008; 95:205–220.

Scheike TH, Sun Y, Zhang MJ, Jensen TK. A semiparametric random effects model for multivariate competing risks. Biometrika. 2010; 97:133–145.

Therneau, TM.; Grambsch, PM. Modeling Survival Data: Extending The Cox Model. New York: Springer; 2000.

Wei LJ, Lin DY, Weissfeld L. Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. Journal of the American Statistical Association. 1989; 84:1065–1073.
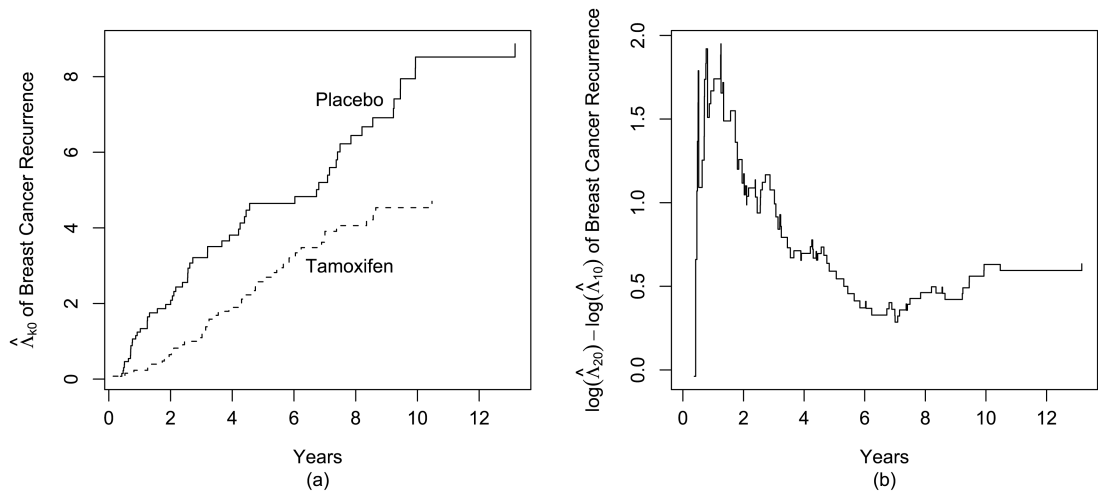
Figure 1.
Estimated baseline cumulative subdistribution hazards for tamoxifen ($k = 1$) and placebo groups ($k = 2$).

Table 1

Comparison of empirical sizes of tests from CC, CCS, weighted stratified (WS) estimating equations (regularly stratified data)

Z	[a,b]	n	s	cen	β_1	CC	CCS	WS
Normal	[∞, ∞]	100	3	0	0	0.075	0.048	0.060
		250	3	0	0	0.081	0.059	0.058
		500	3	0	0	0.123	0.055	0.058
		1000	3	0	0	0.143	0.051	0.055
Normal	[0, 1.7]	100	3	0.55	0	0.089	0.059	0.071
		250	3	0.55	0	0.099	0.054	0.057
		500	3	0.54	0	0.251	0.047	0.045
		1000	3	0.55	0	0.340	0.052	0.052
Bernoulli	[∞, ∞]	100	3	0	0	0.061	0.045	0.052
		250	3	0	0	0.071	0.043	0.045
		500	3	0	0	0.078	0.048	0.050
		1000	3	0	0	0.106	0.045	0.046
Bernoulli	[0.5, 1.5]	100	3	0.36	0	0.056	0.043	0.052
		250	3	0.36	0	0.062	0.056	0.057
		500	3	0.34	0	0.104	0.051	0.051
		1000	3	0.35	0	0.178	0.052	0.050

Table 2

Comparison of parameter estimates from CC, CCS, stratified weighted (WS) estimating equations (regularly stratified data)

Z	[a,b]	cen	Eq.	β_1	$\hat{\beta}_1$	$var(\hat{\beta}_1)$	$E(va\hat{r}_1)$
Normal	[∞, ∞]	0	CC	0.5	0.551	0.005	0.004
			CCS	0.5	0.504	0.007	0.006
			WS	0.5	0.504	0.007	0.006
Normal	[0.5, 1.5]	0.37	CC	0.5	0.533	0.007	0.006
			CCS	0.5	0.498	0.010	0.010
			WS	0.5	0.498	0.010	0.010
Bernoulli	[∞, ∞]	0	CC	1	1.027	0.023	0.024
			CCS	1	1.003	0.025	0.026
			WS	1	1.003	0.025	0.026
Bernoulli	[0, 1.7]	0.44	CC	1	0.942	0.045	0.043
			CCS	1	0.990	0.052	0.050
			WS	1	0.989	0.050	0.048

Table 3

Comparison of parameter estimates from CC, CCS, stratified weighted (WSh for plug-in variance estimator and WSb for bootstrap variance estimator) estimating equations (highly stratified data)

Z_{11}	$[a, b]$	Percentage censored	β_1		E($\hat{\beta}$)[1]		var($\hat{\beta}$)		E($\hat{var}$)		coverage	
					β_{11}	β_{12}	β_{11}	β_{12}	β_{11}	β_{12}	β_{11}	β_{12}
Normal	[0.5, 1.5]	.27	(0,1)	CC	.287	.911	.006	.119	.005	.103	.019	.927
				CCS	.005	.992	.017	.205	.016	.194	.946	.945
				WSh	.008	.990	.016	.198	.015	.184	.941	.942
				WSb	.008	.990	.016	.198	.016	.201	.951	.952
Normal	[0, 1.5]	.28	(1,1)	CC	1.009	.507	.010	.110	.008	.101	.934	.645
				CCS	1.042	1.055	.046	.295	.039	.283	.937	.952
				WSh	1.041	1.050	.045	.286	.037	.267	.931	.948
				WSb	1.041	1.050	.045	.286	.043	.302	.950	.960
Binary	[0, 1.7]	.37	(0,1)	CC	.385	.884	.039	.125	.040	.123	.521	.938
				CCS	−.003	1.048	.101	.275	.093	.259	.943	.941
				WSh	.004	1.039	.095	.261	.088	.247	.939	.937
				WSb	.004	1.039	.095	.261	.099	.280	.953	.953
Binary	[0, 1.2]	.41	(0,1)	CC	1.239	.650	.046	.132	.047	.119	.833	.805
				CCS	1.033	1.038	.135	.314	.120	.288	.944	.949
				WSh	1.037	1.034	.132	.307	.117	.280	.938	.936
				WSb	1.037	1.034	.132	.307	.165	.320	.964	.941

Wsh & WSb differ only in the variance computation, whereas the estimates are the same.

Table 4

Estimation of coefficients and standard errors in models for breast cancer recurrence

	Fine and Gray			Stratified regression		
	$\hat{\beta}$	se($\hat{\beta}$)	p	$\hat{\beta}$	se($\hat{\beta}$)	p
Log(nodes)	.274	.111	.014	.272	.111	.014
Tumorsize	.109	.040	.007	.107	.040	.007
Age	−.037	.028	.180	−.036	.027	.190
Treatment	−2.035	.644	.002	–	–	–
Treatment × *t*	.857	.334	.010	–	–	–
Treatment × *t²*	−.086	.034	.013	–	–	–

Table 5

Estimation of coefficients and standard errors in models for acute GvHD or chronic GvHD occurrence

| | Unstratified | | | | Stratified | | | |
| | Simple | | Multiple | | Simple | | Multiple | |
	β̂	p	β̂	p	β̂	p	β̂	p
Female to male versus others	.363(.060)	<.01	.368(.063)	<.01	.343(.073)	<.01	.372(.075)	<.01
PBSC versus BMT	−.051(.054)	.35	−.060(.057)	.29	.076(.075)	.31	.060(.084)	.48
FAB M5, M6, M7 versus others	.089(.067)	.19	.099(.069)	.15	.109(.063)	.08	.109(.066)	.10
Matched unrelated donor versus genoidentical	.049(.081)	.54	.086(.087)	.33	.119(.116)	.33	.120(.122)	.34

PBSC = peripheral blood stem cells; BMT = bone marrow transplant.