HHS Public Access

Author manuscript

Lifetime Data Anal. Author manuscript; available in PMC 2016 April 01.

Published in final edited form as:

Lifetime Data Anal. 2015 April ; 21(2): 197–217. doi:10.1007/s10985-014-9313-9.

Checking Fine and Gray Subdistribution Hazards Model with Cumulative Sums of Residuals

Jianing Li,
Division of Biostatistics, Medical College of Wisconsin, U.S.A

Thomas H. Scheike, and
Department of Biostatistics, University of Copenhagen, Danmark

Mei-Jie Zhang
Division of Biostatistics, Medical College of Wisconsin, U.S.A

Jianing Li: jili@mcw.edu; Thomas H. Scheike: ts@biostat.ku.dk; Mei-Jie Zhang: meijie@mcw.edu

Abstract

Recently, Fine and Gray (1999) proposed a semi-parametric proportional regression model for the subdistribution hazard function which has been used extensively for analyzing competing risks data. However, failure of model adequacy could lead to severe bias in parameter estimation, and only a limited contribution has been made to check the model assumptions. In this paper, we present a class of analytical methods and graphical approaches for checking the assumptions of Fine and Gray's model. The proposed goodness-of-fit test procedures are based on the cumulative sums of residuals, which validate the model in three aspects: (1) proportionality of hazard ratio, (2) the linear functional form and (3) the link function. For each assumption testing, we provide a *p*-values and a visualized plot against the null hypothesis using a simulation-based approach. We also consider an omnibus test for overall evaluation against any model misspecification. The proposed tests perform well in simulation studies and are illustrated with two real data examples.

Keywords

Competing risk; goodness-of-fit; proportional subdistribution hazard; cumulative residual; link function; omnibus test

1 Introduction

Competing risks data arise in medical studies where patients may experience failure from multiple causes, and failure from one cause precludes observation of failure from the other causes. Fine and Gray (1999) proposed a semi-parametric proportional subdistribution hazards model (FG model). It can be applied to evaluate the effects of covariates on the cumulative incidence function directly. The FG model has been using extensively in cancer studies, epidemiological studies, and many other biomedical studies (Scrucca et al, 2007; Wolbers et al, 2009; Kim, 2007; Lau et al, 2009). Similarly to the Cox regression model for the regular time to event data (Lin et al, 1993), the FG model may also fail in three ways: (1) the proportional hazards assumption does not hold, i.e. the subdistribution hazard ratio is time-variant; (2) the functional forms of individual covariates in the exponent of the model

are misspecified; (3) the link function, i.e. the exponential form for the subdistribution hazard ratio is not appropriate. It is well known that violation of any one of these three model assumptions, required for the Cox proportional hazards model, may lead to severely biased effect for the statistical inference (Lin et al, 1993).

Although it is important to validate the FG model, only limited contributions have been made to check the model assumptions. For checking the proportionality assumption, Zhou et al (2013) proposed a simple goodness-of-fit test by adding a time dependent covariate, and showed through a simulation study that this simple test performed well. However, the power of this simple test may depend on the specific pattern of the covariate effect over time, under the alternative models. Scheike and Zhang (2008) considered a goodness-of-fit test by estimating the regression effect nonparametrically over time based on their direct binomial modeling approach. In general, such nonparametric goodness-of-fit tests may have lower power compared to the specific alternative model based test. Perme and Andersen (2008) proposed using a pseudo-value approach to test the proportionality assumption and a linear functional form of individual covariates for right censored competing risks data. Pseudo-observations are often calculated for a pre-fixed set of time-points. Some computation issues may occur when one needs to calculate the pseudo-observations for a large set of time points. Pseudo-value based goodness-of-fit tests may be difficult to employ when the censoring distribution depends on some continuous covariates.

Martingale residual based goodness-of-fit test has been well developed for checking the Cox model (Andersen et al, 1993). Alternatively, Lin et al (1993) proposed a goodness-of-fit test based on cumulative sums of martingale residuals to check the proportionality assumption, functional forms of covariates effect, and the link function in the Cox model for right censored survival data. In this paper, we extend their approach (Lin et al, 1993) to the right censored competing risks data, using the cumulative sums of residuals to check the FG model. For checking the proportionality assumption, we propose using the score process, which is a special case of the cumulative sums of residuals on time. Compared to the approach of adding a time-dependent variable (Zhou et al, 2013), our proposed test has a better performance when the trend of the covariate effect is not modeled correctly over time. We also propose using the cumulative sums of residuals on covariates values to check the linearity and the link function assumption of the FG model. All tests proposed in this study are based on a simulation based approach for the null limiting distribution of the test process. Besides the numerical results, we provide each test with a diagnostic plot for the visual checking. An omnibus test accumulating the residuals on time and covariate value simultaneously, is considered to evaluate the validation of FG model in an overall view.

In section 2, we briefly introduce the background of FG model and some techniques which will be used for the model checking. In section 3, we examine the performance of the proposed method through simulation studies. In the simulation studies, we test against each of the assumptions with some commonly seen alternatives. We show the application of the proposed approach to two real datasets in section 4 and conclude with discussion in section 5.

2 Model checking techniques

Let T and C be the underlying time to event and censoring, and $X = T \wedge C$. $\Delta = \mathbb{1}\{T \leq C\}$ is the non-censoring indicator, and 3 represents cause of failure. We usually observe i.i.d. (X_i, Δ_i, Δ_i ε_i), where $\Delta_i\varepsilon_i$ contains the information of both censoring and cause of failure. For the competing risk data, one quantity of interest is the cumulative incidence function, the probability of failure due to one specific cause (assuming to be cause 1),

$$F_1(t)=P(T_i \leq t, \varepsilon_i=1)=\int_0^t S(u^-)d\Lambda_1(u),$$

where $S(t)$ is the overall survival function and $\Lambda_1(t)$ is the cumulative cause-specific hazard for cause 1. Fine and Gray (1999) proposed a semi-parametric proportional regression model for the subdistribution hazard function

$$\lambda_1^*(t|\boldsymbol{Z})=\lambda_{10}^*(t)\exp\{\boldsymbol{\beta}^\top \boldsymbol{Z}\}, \quad (1)$$

where $\lambda_{10}^*(t)$ is the baseline subdistribution hazard, $\boldsymbol{\beta}$ are the covariates effect for $\boldsymbol{Z}$. Let $N_i^1(1)=\mathbb{1}\{T_i \leq t, \varepsilon_i=1\}$ be the cause specific counting process, $Y_i^1(t)=1-N_i^1(t^{-1})$ is the risk set associated with the subdistribution hazard. When the right censoring is present, $N_i^1(t)$ and $Y_i^1(t)$ will not be fully observed. Recently, Fine and Gray (1999) proposed using the inverse probability of censoring weighting technique (IPCW) with a time-dependent weight function $w_i(t, G_c) = r_i(t)G_c(t)/G_c(T_i \wedge t)$, where $r_i(t) = \mathbb{1}\{C_i \leq T_i \wedge t\}$ and $G_c(t)$ is the survival function for the censoring distribution. It is easy to see both $w_i(t, \hat{G}_c)dN_i^1(t)$ and $w_i(t, \hat{G}_c)Y_i^1(t)$ are computable for all time t. IPCW can be generalized to allow the dependency between C and $\boldsymbol{Z}$ as suggested in Fine and Gray (1999). In this study, we estimate $G_c(t)$ nonparametrically using the Kaplan-Meier estimator, denoted as $\hat{G}_c(t)$, or $\hat{G}_c$ for simplicity. The score equation and information matrix can be obtained by taking the first and negative second derivative of the partial likelihood:

$$\boldsymbol{U}(\boldsymbol{\beta})=\sum_{i=1}^n \int_0^\infty \left\{\boldsymbol{Z}_i-\frac{\boldsymbol{S}_1(\boldsymbol{\beta},t)}{S_0(\boldsymbol{\beta},t)}\right\} w_i(t,\hat{G}_c)dN_i^1(t),$$
$$\boldsymbol{I}(\boldsymbol{\beta})=\sum_{i=1}^n \int_0^\infty \left[\frac{\boldsymbol{S}_2(\boldsymbol{\beta},t)}{S_0(\boldsymbol{\beta},t)}-\left\{\frac{\boldsymbol{S}_1(\boldsymbol{\beta},t)}{S_0(\boldsymbol{\beta},t)}\right\}^{\otimes 2}\right] w_i(t,\hat{G}_c)dN_i^1(t),$$

where $\otimes$ denotes the Kronecker product and

$$\boldsymbol{S}_k(\boldsymbol{\beta},t)=\sum_{i=1}^n w_i(u,\hat{G}_c)Y_i(u)\boldsymbol{Z}_i^{\otimes k}\exp\{\boldsymbol{\beta}^\top \boldsymbol{Z}_i\}, \text{ for } k=0,1,2.$$

The $\boldsymbol{\beta}$ is estimated by solving the score equation $\boldsymbol{U}(\hat{\boldsymbol{\beta}}) = 0$, and the cumulative baseline subdistribution function, $\Lambda_{10}^*(t)=\int_0^t \lambda_{10}^*(u)du$ is estimated by

$$\hat{\Lambda}_{10}^{*}(t)=\sum_{i=1}^{n}\int_0^t \frac{w_i(u,\hat{G}_c)dN_i^1(u)}{S_0(\hat{\boldsymbol{\beta}},u)}.$$

Under the null hypothesis of FG model, the differences between the weighted counting processes and their respective intensity functions are

$$w_i(u,\hat{G}_c)dM_i^1(t)=w_i(u,\hat{G}_c)dN_i^1(t)-w_i(u,\hat{G}_c)Y_i^1(t)\exp(\boldsymbol{\beta}_0^{\top}\boldsymbol{Z}_i)d\Lambda_{10}^{*}(t). \quad (2)$$

Most importantly, $w_i(u,\hat{G}_c)dM_i^1(t)$ is a zero-mean Gaussian process. The residuals are defined as

$$w_i(t,\hat{G}_c)d\hat{M}_i^1(t)=w_i(t,\hat{G}_c)dN_i^1(t)-w_i(t,\hat{G}_c)Y_i^1(t)\exp(\hat{\boldsymbol{\beta}}^{\top}\boldsymbol{Z}_i)d\hat{\Lambda}_{10}^{*}(t). \quad (3)$$

2.1 Checking the proportional hazards assumption

We use the observed score process $\boldsymbol{U}(\boldsymbol{\beta},\hat{t})$ for checking the proportional hazards assumption. It can be shown that the score process equals to a special form of the cumulative sums of residuals over time:

$$\boldsymbol{U}(\hat{\boldsymbol{\beta}},t)=\sum_{i=1}^{n}\int_0^t \boldsymbol{Z}_i w_i(u,\hat{G}_c)d\hat{M}_i^1(u),$$

Under the null hypothesis that FG model is valid, by involving the equation (2) and (3) we obtain

$$U(\hat{\boldsymbol{\beta}},t)=\sum_{i=1}^{n}\int_0^t \boldsymbol{Z}_i w_i(u,\hat{G}_c)\left\{dN_i^1(u)-Y_i^1(u)\exp(\hat{\boldsymbol{\beta}}^{\top}\boldsymbol{Z}_i)d\hat{\Lambda}_{10}^{*}(u)\right\}$$
$$=\sum_{i=1}^{n}\int_0^t \boldsymbol{Z}_i w_i(u,\hat{G}_c)dM_i^1(u)-\sum_{i=1}^{n}\int_0^t \boldsymbol{Z}_i w_i(u,\hat{G}_c)Y_i^1(u)\left\{\exp(\hat{\boldsymbol{\beta}}^{\top}\boldsymbol{Z}_i)d\hat{\Lambda}_{10}^{*}(u)-\exp(\boldsymbol{\beta}_0^{\top}\boldsymbol{Z}_i)d\hat{\Lambda}_{10}^{*}(u)\right\}$$

By the Taylor expansions of $\boldsymbol{U}(\boldsymbol{\beta},\hat{t})$ at $\boldsymbol{\beta}_0$ and some arguments, the observed score process is shown to be asymptotically equivalent to the summation of *n* independent and identically distributed random processes,

$$\boldsymbol{U}(\hat{\boldsymbol{\beta}},t)=\sum_{i}^{n}\boldsymbol{W}_i^{(p)}(\boldsymbol{\beta}_0,t)+o_p(1),$$

where the explicit form of $\boldsymbol{W}_i^{(p)}(\boldsymbol{\beta},t)$ and a detailed derivation of the limiting distribution under the FG model can be found in the Appendix. The limiting distributional of $w_i(u,\hat{G}_c)M_i^1(u)$ is asymptotically equivalent to the limiting distribution of

$w_i(u, \hat{G}_c)N_i^1(u)V_i$, where $w_i(u, \hat{G}_c)N_i^1(u)$ is the observed counting process and V_i, $i = 1, \ldots, n$, are i.i.d. standard normal variables (Lin et al, 1993). Therefore, $\boldsymbol{W}_i^{(p)}$ is replaced by the product of $\hat{\boldsymbol{W}}_i^{(p)}$ and a normal variable V_i, where $\hat{\boldsymbol{W}}_i^{(p)}$ is a plug-in estimator for $\boldsymbol{W}_i^{(p)}(\boldsymbol{\beta}_0, t)$.

To check the proportionality assumption for the jth covariate, we consider the score process

$$B_j^{(p)}(t) = \{I_{jj}^{-1}(\hat{\boldsymbol{\beta}})\}^{1/2} U_j(\hat{\boldsymbol{\beta}}, t), \text{ for } j = 1, \ldots, m,$$

where U_j and I_{jj}^{-1} are the jth element and jth diagonal element of $\boldsymbol{U}$ and $\boldsymbol{I}^{-1}$, respectively, and m is the total number of covariates. Under the FG model, $B_j^{(p)}(t)$ is, asymptotically, a zero-mean Gaussian process and has the same limiting distributions as

$$\{I_{jj}^{-1}(\hat{\boldsymbol{\beta}})\}^{1/2} \sum_i^n \hat{W}_{ij}^{(p)}(\hat{\boldsymbol{\beta}}, t) V_i,$$

and $\hat{W}_{ij}^{(p)}$ is the jth element of $\hat{\boldsymbol{W}}_i^{(p)} = (\hat{W}_{i1}^{(p)}, \cdots, \hat{W}_{im}^{(p)})$. Resampling techniques can be used for the supremum test over a fixed time interval. Let

$$b_j^{(k)} = \sup_t \left| \{I_{jj}^{-1}(\hat{\boldsymbol{\beta}})\}^{1/2} \sum_i^n \hat{W}_{ij}^{(p)}(\hat{\boldsymbol{\beta}}, t) V_i^{(k)} \right|, \text{ for } k = 1, \ldots, K,$$

where K is the total number of realizations. The p-value of proposed test can be approximated by

$$\frac{1}{K} \sum_{k=1}^{K} \mathbb{1} \left\{ \sup_t \left| B_j^{(p)}(t) \right| \geq b_j^{(k)} \right\}.$$

We can also plot the observed score process $B_j^{(p)}(t)$ versus the follow-up time. Under the null hypothesis, $B_j^{(p)}(t)$ asymptotically equals to a zero-mean Gaussian process, which fluctuates randomly around zero. To access how unusual the observed process is, we may plot it along with a few simulated limiting processes under FG model. If the proportionality assumption is invalid, the observed process will be isolated above or below those simulated processes for some time periods.

To evaluate the overall proportionality assumption, we consider

$$\sup_t \sum_{j=1}^{m} \left| \left\{ I_{jj}^{-1}(\hat{\boldsymbol{\beta}}) \right\}^{1/2} U_j(\hat{\boldsymbol{\beta}}, t) \right|.$$

The overall plot and the *p*-value of the supremum test can be generated exactly the same as we check the proportionality for a single covariate.

2.2 Checking the functional form and the link function

To check the functional form of *j*th covariate, we consider

$$B_j^{(f)}(z_j) = \sum_{i=1}^{n} \int_0^{\infty} \mathbb{1}\{Z_{ij} \le z_j\} w_i(u, \hat{G}_c) d\hat{M}_i^1(u), \text{ for } j = 1, \ldots, m,$$

and to check the link function, we consider

$$B^{(l)}(x) = \sum_{i=1}^{n} \int_0^{\infty} \mathbb{1}\left\{ \hat{\boldsymbol{\beta}}^{\top} \boldsymbol{Z}_i \le x \right\} w_i(u, \hat{G}_c) d\hat{M}_i^1(u).$$

Both processes are cumulative sums of residuals with respect to the covariate values. For the linear functional form, we focus on the change of each single covariate value. For the link function, we catch the value change of all covariates simultaneously by ($\hat{\boldsymbol{\beta}}^{\top}\boldsymbol{Z}$), which reflects on the adequacy of link function. Similarly as the process proposed for checking the proportionality of subdistribution hazards, we can show that, under the FG model, these two processes asymptotically equals to zero-mean Gaussian processes. The limiting processes can be simulated as well (see Appendix for details). Resampling techniques can be used to estimate the *p*-values the supremum tests. Plotting the observed process and a few simulated limiting processes under the FG model versus the covariate values can be used for visually checking the linearity of functional form and the link function.

2.3 An omnibus test for each covariate

Summing up the residuals, with respect to *t* and ***z*** simultaneously, would offer a global checking of the model. Consider

$$\boldsymbol{B}^{(o)}(t, \boldsymbol{z}) = \sum_{i=1}^{n} \int_0^t \mathbb{1}\{\boldsymbol{Z}_i \le \boldsymbol{z}\} w_i(u, \hat{G}_c) d\hat{M}_i^1(u),$$

Since $\boldsymbol{B}^{(o)}(t, \boldsymbol{z})$ is, asymptotically, a zero-mean Gaussian process under the null hypothesis of the FG model (see Appendix for details), we can construct a lack of fit test based on resampling techniques as well. Plotting the observed process over *t* and ***z*** simultaneously would give an omnibus view on model adequacy. But since it is a high-dimensional plot, some technical issues need to be considered, which are not pursued in this study.

3 Simulation study

We evaluated the performance of the proposed approaches by simulation studies. The simulation studies were designed in two aspects: first, we tested the proportional hazards assumption, which used the cumulative residuals with respect to time; second, we tested the linear functional form and the link function, which used the cumulative residual with respect to covariate values (note that only continuous covariate will be considered here). All simulation studies were designed for 15% and 30% censoring, with total sample sizes of 50, 100 and 300, respectively. All censoring time were generated independently from a uniform random variable, $U(0, \tau]$, where τ was used to adjust the censoring rate. For each setting, we replicated 5000 repeated samples for the type I error rate and 2000 samples for the power of the proposed tests. Significance level a was set at 0.05 throughout the simulation study unless otherwise specified.

3.1 proportional hazards assumption

Suppose we have two groups in a univariate model with the only covariate *Z*. Half of the subjects belong to group 1 ($Z = 1$) and the other half belong to group 2($Z = 0$). The type I error rate was evaluated under the null hypothesis with the data generated from the FG model. Respectively, the cumulative incidence functions for cause 1 and cause 2 are:

$$F_1(t|Z)=1-\{1-p_1(1-e^{-t})\}^{\exp(\beta Z)},$$
$$F_2(t|Z)=(1-p_1)^{\exp(\beta Z)}\left[1-e^{\{-t\exp(\beta Z)\}}\right]$$

where $p_1 = F_1(\infty|Z = 0)$ is the cumulative incidence probability of cause 1 for $Z = 0$ as t goes to infinity, and it was set to be 0.66, and β was set to be 0.2 with $F_1(\infty|Z = 1) = 0.73$. We first determined the cause of event by generating a uniform variable $u_i \sim U[0, 1]$. If $u_i \leq F_1(\infty|z_i)$, then the cause of failure was set $\varepsilon_i = 1$, and the failure time was generated from $F_1(t|z_i)/F_1(\infty|z_i)$ using inverse distribution method; otherwise, the failure time was generated from $F_2(t|z_i)/F_2(\infty|z_i)$. The Table 1 shows the type I error under the null hypothesis. The type I error rates were consistently close to the nominal level when the sample sizes were getting large.

Under the alternative hypothesis, the proportional hazards assumption does not hold. We consider following alternative subdistribution hazards for the cause 1,

$$\lambda_1^*(t|Z)=\lambda_{10}^*(t)\exp\{\beta(t)Z\}, \quad (4)$$

where $\beta(t)$ counts the time-varying treatment effect. To study the power of our proposed test, we generated data from two types of $\beta(t)$. We assume the treatment effect has a linear function of time *t*, i.e. $\beta(t) = \beta + \theta t$. The baseline subdistribution hazard function in model (4) was set to

$$\lambda_{10}^*(t)=\gamma\exp(\rho t).$$

A uniform variable u was generated from $U(0, 1)$ to determine the cause of event, and we fixed the probability of cause 1 at 0.66, for simplicity. If $u \leq 0.66$, the failure time was generated from $F_1(t|Z)$; otherwise, the failure time was generated from $F_2(t|Z)$. The cumulative incidence function of cause 1 equals to

$$F_1(t|Z)=1-\exp\{-\Lambda_1^*(t|Z)\}=1-\exp\left[-\int_0^t \gamma \exp\{(\rho+\theta Z)s+\beta Z\}\,ds\right].$$

The failure time of cause 2 was generated from an exponential distribution with rate $\exp(\alpha Z)$, given $F_2(\infty|Z) = 1 - F_1(\infty|Z)$. Here we set $\beta = 0$, $\theta = -8$, $\rho = -5$, $\alpha = 2$ and $\alpha = -0.5$. In parallel to the proposed approach, we also fitted the same data using Fine and Gray's model with $\beta(t) = t$, $\log(t)$, and t^2 (Zhou et al, 2013). Table 2 shows that when the treatment has a linear time-varying effect, fitting time-varying Fine and Gray's model with linear effect gave the highest power. The proposed test and the other two time-varying effect models performed slightly worse than the liner effect model, but all had very reasonable power. From the results, we can see that if the form of treatment effect is correctly specified, it would have the largest power. But when the treatment effect is misspecified, our test would give a comparable and even better power.

In practice, the treatment effect may have a more complicated relationship with time. For example, the treatment has two different effects before and after time t_0, where t_0 is unknown. To evaluate the performance of proposed test under this situation, the data were generated from model (4) with treatment effect $\beta(t) = \beta_1 \mathbb{1}(t \leq t_0) + \beta_2 \mathbb{1}(t > t_0)$. For simplicity, we assume $\lambda_{10}^*(t)=1$, therefore

$$F_1(t|Z)=1-\exp\left[-\int_0^t \exp\{Z\beta_1 \mathbb{1}(u \leq t_0)+Z\beta_2 \mathbb{1}(u>t_0)\}du\right].$$

The treatment effects, β_1 and β_2 were set as 1 and 0.2, respectively, and t_0 was set as 0.5. The results in Table 3 show the power of our proposed test and the tests based on fitting Fine and Gray's model with three different time-varying effects. In this case, our proposed test produced better power than any of the tests based on a pre-specified alternative form of the covariate effect.

3.2 Linear functional form and link function

We use the cumulative sums of residuals over covariate values for testing both the linear functional form and the link function. The former sums up the values for each individual covariate, while the latter sums up the combination of all covariates. To study the type I error rate, the data were generated from model (1) with a single covariate Z, where $Z = \{0, \cdots, 9\}$, with equal proportions. Here we set $\beta = 0.2$. Table 4 shows that the type I error rates were consistently close to the nominal level.

The linear functional form fails when the covariates do not linearly correspond to the log of the subdistribution hazard ratio. To test the validation of functional form, the failure time of cause 1 was generated from

$$\lambda_1^*(t|Z)=\lambda_{10}^*(t)\exp\{\beta_1 Z+\beta_2 Z^2\}, \quad (5)$$

with $\beta_1 = 0.8$ and $\beta_2 = -0.1$. Failure times of cause 2 were generated from the exponential distribution with the rate of exp(−0.5). The rate of cause 1 was pre-set at 0.6. We fitted the data using FG model (1) with the covariate *Z*. Table 4 shows that our proposed test for the validation of the linear functional form has sufficient power and that the power increases when the sample size increases.

The misspecification of the link function means the exponential form in model (1) is inappropriate. For example, if the data is from an additive model, the test should give strong evidence of model misspecification. We generated the failure time of cause 1 from

$$\lambda_1^*(t|Z)=\alpha(t)+\beta Z. \quad (6)$$

For simplicity, we assume $\alpha(t) = 0.03$, which is a constant, and $Z = \{0, \cdots, 9\}$ with equal proportions. We set $\beta = 0.5$. The failure times of cause 2 were generated from an exponential distribution with the rate of exp(−0.5). The rate of cause 1 was pre-set at 0.6. We fitted the data to the FG model (1) with the covariate *Z*. It is clear that this model is misspecified. Table 4 shows the power of our proposed test for validation of the link function of the FG model. We can see the test is very reliable when sample size is getting large.

4 Real data example

We now apply the proposed testing approaches to two real data examples. The *p*-values of the supremum tests throughout this section are based on 1000 replications. In each graphical display, the observed process is indicated by a solid curve and the first 10 simulated processes are plotted in dash curves. Comparing to the simulated curves, we can distinguish the observed curve if the model assumption is not appropriate.

4.1 BMT data

We considered data from the Center for International Blood and Marrow Transplant Research (CIBMTR), on acute myeloid leukemia (AML) patients, who are older than 50 years, and had alternative donor hematopoietic transplants in first complete remission (Weisdorf et al, 2013). The CIBMTR is comprised of clinical and basic scientists who share data on their blood and bone marrow transplant patients, with the CIBMTR Data Collection Center located at the Medical College of Wisconsin. The CIBMTR has a repository of information regarding the results of transplants at more than 450 transplant centers worldwide. For the illustrative purpose, the data set used for our example consists of patients who received a 8/8 HLA-matched unrelated donor transplant (URD: N=440) or an umbilical cord blood transplant (UCB: N=204). The two competing events are leukemia relapse and treatment-related mortality (TRM), which is defined as death without relapse. The CIBMTR

study (Weisdorf et al, 2013) identified that donor type had a time-varying effect on the TRM. The variables considered in this example are donor type (GP (main interest of the study): 1 for 8/8 URD versus 0 for UCB) and patient age (AGE: median=59 (range: 50 – 76) years old). We fit this BMT data to a Fine and Gray's model for TRM with two covariates: GP and AGE. The parameter estimates are $\hat{\beta}(\text{GP}) = -0.287$ ($P = 0.054$) and $\hat{\beta}(\text{AGE}) = 0.012$ ($P = 0.346$).

First, we checked the proportionality of two covariates GP and AGE. The *p*-values of the supremum test are < 0.001 and 0.118, respectively, which indicate that the proportional hazards assumption of AGE is reasonable, but the proportional hazards assumption of GP is strongly violated. Figure 1 displays the score process for the covariates of GP and AGE, respectively. For GP, the observed process severely departs from the simulated process in the early time period since transplant, which indicates that the donor type had a strong time-varying effect on TRM. For the early time departure of the covariate effect (Figure 1), we suggest adding a time-varying covariate $\gamma \times \log(t)$. Applying the *crr*() function in R-*cmprsk* package suggested by Zhou et al (2013), we obtained $\hat{\gamma}(\text{GP}) = 0.565$ ($P < 0.001$). The observed score process of AGE sits well within the simulated score process which indicates that the proportional hazards assumption of AGE does not violate. Also, we noticed that the conclusion from the p-value and graphical approach are consistent. When the proportionality assumption of the FG model is violated, more flexible alternative regression models for the competing risks data need to be considered, which allow some covariates to have time-varying effect (see Scheike et al (2008); Scheike and Zhang (2008)).

We also checked the adequacy of functional form and link function. Both linear functional form and link function are satisfied. We do not recommend testing the functional form for the donor type, since it only has 2 levels and discussing the linearity is not very meaningful. Table 5 gives a summary of the *p*-values for checking each assumption based on the supremum test.

4.2 PBC data

Next, we looked at an example violates the linear functional form. Mayo Clinic developed a database for patients with primary biliary cirrhosis (PBC) of the liver conducted between 1974 and 1984. The data and the description of the PBC dataset were given in Appendix of Fleming and Harrington (1991). Of the 418 patients, 161 died, which was the main interest of the study. 25 patients had a liver transplantation, and 232 patients were censored at the end of the study (without a liver transplantation). This data were used by Dickson et al (1989) building a Cox model with five covariates, log(bilirubin), log(protime), log(albumin), age and odema. Treating a liver transplant and all others as censored observations. Lin et al (1993) demonstrated that the use of the untransformed bilirubin was inappropriate because the fitted model overestimated the hazard at low bilirubin values and suggested a log transformation. Since patients could die after the liver transplantation, and the liver transplant may be correlated with the death rate, we treat the liver transplantation as the competing event for illustrative purpose. In this section, we reanalyze the data in a competing risks setting by using Fine and Gray's subdistribution hazards model, and checking the assumptions of the FG model.

We first consider fitting the Fine and Gray's model (1) with covariates bilirubin, log(protime), log(albumin), age and odema. The parameter estimates are 0.044, 0.966, 0.109, −2.558 and 3.507 respectively, and all are significant, with $P < 0.05$. Figure 2 shows the plot of the cumulative residuals against bilirubin. The fitted model extremely overestimates the subdistribution hazard for the low bilirubin values with $P < 0.001$. The pattern of the plot suggests a log transformation is needed. Figure 3 shows the plot of the cumulative residuals against log(bilirubin), which is a much better functional form ($P = 0.114$). Additional analysis indicates that the functional forms for the remaining covariates and the link function are appropriate (see Table 6).

Our proposed test shows that the proportional hazards assumption was not satisfied for oedema and log(protime) with $P = 0.019$ and 0.007, respectively (see Table 6). The plots of their score processes against time confirm our test conclusion as well (plots are omitted here). The *p*-value of overall proportional test is 0.006, which means the non-proportionality should be corrected. Again, we can consider fitting this PBC data with some more flexible alternative models, which allow some covariates to have time-varying effects (see Scheike et al (2008); Scheike and Zhang (2008)).

Lin et al (1993) checked the Cox model with cumulative sums of martingale-based residuals for the PBC data treating 25 liver transplantations as the censored subjects, and derived similar conclusions since it only counts 13% (25/186) of total events. This shows our proposed goodness-of-fit tests can be used for the model validation for the Fine and Gray's proportional subdistribution hazards model.

5 Conclusion and discussion

In this paper, we presented a dynamic model diagnostics approach for checking the Fine and Gray's model based on cumulative sums of residuals. Specifically, we checked the proportionality of the hazards function, the linearity of the functional form, and the link function. We proposed different test statistics for each specific testing purpose, and derived the asymptotic properties under the null hypothesis that the Fine and Gray's model is valid. The three different test statistics were written into a general form, and more details of this were provided in the Appendix. For each proposed test, we also presented a graphical method for visually checking the model assumptions, which also provided some useful information on how the model assumption was violated.

When any one of the Fine and Gray model assumptions is violated, one needs to consider an alternative regression model. Only a limited goodness-of-fit test has been considered for these flexible models. Scheike et al (2008) and Scheike and Zhang (2008) proposed a class of flexible regression models which allow some covariates to have time-varying effects, and have some alternative link functions. In such situations an important issue would be to subsequently check the adequacy of the extended model. Actually some of the extended models could be also validated using the cumulative sums of residuals. For example, when proportional hazards assumption is invalid, we may consider the time-varying treatment effect as $\beta(t) = \beta + \theta t$. In this case, the cumulative sums of residuals would involve the time-

varying term and the limiting distribution under the null hypothesis needs to be derived. This could be an extension of our current work and further studies are needed.

One issue for using the resampling techniques is the extensive computational time. We use 1000 replicates in most of our studies, but some studies need more replicates to obtain reliable p-values for the supremum-type test (Lin et al, 1993). We implemented the most computational intensive calculation in C++ for its speed and wrapped it by R via the interface between R and C++ for its flexibility. The code implementing the analysis is available upon request, and a user-friendly R package is under development for future release.

References

Andersen, PK.; Borgan, Ø.; Gill, RD.; Keiding, N. Statistical Models Based on Counting Processes. Springer-Verlag; New York: 1993.

Dickson E, Grambsch P, Fleming T, Fisher LD, Langworthy A. Prognosis in primary biliary cirrhosis: model for decision making. Hepatology. 1989; 10:1–7. [PubMed: 2737595]

Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. J Amer Statist Assoc. 1999; 94:496–509.

Fleming, TR.; Harrington, DP. Counting Processes and Survival Analysis. Wiley; New York: 1991.

Gray RJ. A class of *k*-sample tests for comparing the cumulative incidence of a competing risk. Ann Statist. 1988; 16:1141–1154.

Kim H. Cumulative incidence in competing risks data and competing risks regression analysis. Clinical Cancer Research. 2007; 13:559–565. [PubMed: 17255278]

Lau B, Cole S, Gange S. Competing risk regression models for epidemiologic data. American Journal of Epidemiology. 2009; 170:244–256. [PubMed: 19494242]

Lin DY, Wei LJ, Ying Z. Checking the Cox model with cumulative sums of martingale-based residuals. Biometrika. 1993; 80:557–572.

Perme M, Andersen P. Checking hazard regression models using pseudo-observations. Statistics in Medicine. 2008; 27(25):5309–5328. [PubMed: 18712781]

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

Scheike TH, Zhang MJ, Gerds T. Predicting cumulative incidence probability by direct binomial regression. Biometrika. 2008; 95:205–20. pMC Journal-In Process.

Scrucca L, Santucci A, Aversa F. Competing risk analysis using R: an easy guide for clinicians. Bone Marrow Transplant. 2007; 40(4):381–387. [PubMed: 17563735]

Weisdorf D, Eapen M, Ruggeri A, Zhang M, Zhong X, Brunstein C, Ustun C, Rocha V, Gluckman E. Alternative donor hematopoietic transplantation for patients older than 50 years with aml in first complete remission: unrelated donor and umbilical cord blood transplantation outcomes. Blood. 2013; 122:302. [PubMed: 23543458]

Wolbers M, Koller M, Witteman J, Steyerberg E. Prognostic models with competing risks: methods and application to coronary risk prediction. Epidemiology. 2009; 20(4):555–561. [PubMed: 19367167]

Zhou B, Fine J, Laird G. Goodness-of-fit test for proportional subdistribution hazards model. Statistics in Medicine. 201310.:1002/sim.5815

Appendix

Consider the following partial sums of residuals

$$B(t,x)=\sum_{i=1}^{n}\int_0^t f(x,\boldsymbol{Z}_i,\boldsymbol{v},p)w_i(u,\hat{G}_c)d\hat{M}_i^1(u),$$

where $f(x, \boldsymbol{Z}_i, \boldsymbol{v}, p) = (\boldsymbol{v}^\top \boldsymbol{Z}_i)^p \mathbb{1}(\boldsymbol{v}^\top \boldsymbol{Z}_i \le x)$. Here $\boldsymbol{v}$ is a vector with same dimension of covariates $\boldsymbol{Z}$, and $p = 0$ or 1. Under the null hypothesis that the FG model is valid,

$$\begin{aligned} B(t,x)&=\sum_{i=1}^{n}\int_0^t f(x,\boldsymbol{Z}_i,\boldsymbol{v},p)w_i(u,\hat{G}_c)\left\{dN_i^1(u)-Y_i^1(u)\exp(\hat{\boldsymbol{\beta}}^\top \boldsymbol{Z}_i)d\hat{\Lambda}_{10}^*(u)\right\}\\ &=\sum_{i=1}^{n}\int_0^t f(x,\boldsymbol{Z}_i,\boldsymbol{v},p)w_i(u,\hat{G}_c)dM_i^1(u)\\ -\sum_{i=1}^{n}\int_0^t f(x,\boldsymbol{Z}_i,\boldsymbol{v},p)w_i(u,\hat{G}_c)Y_i^1(u)&\left\{\exp(\hat{\boldsymbol{\beta}}^\top \boldsymbol{Z}_i)d\hat{\Lambda}_{10}^*(u)-\exp(\boldsymbol{\beta}_0^\top \boldsymbol{Z}_i)d\Lambda_{10}^*(u)\right\}\\ &=\sum_{i=1}^{n}\int_0^t f(x,\boldsymbol{Z}_i,\boldsymbol{v},p)w_i(u,\hat{G}_c)dM_i^1(u) \end{aligned} \tag{7}$$

$$-\sum_{i=1}^{n}\int_0^t f(x,\boldsymbol{Z}_i,\boldsymbol{v},p)w_i(u,\hat{G}_c)Y_i^1(u)\exp(\boldsymbol{\beta}_0^\top \boldsymbol{Z}_i)\boldsymbol{Z}_i^\top d\hat{\Lambda}_{10}^*(u)\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\right) \tag{8}$$

$$-\sum_{i=1}^{n}\int_0^t f(x,\boldsymbol{Z}_i,\boldsymbol{v},p)w_i(u,\hat{G}_c)Y_i^1(u)\exp(\boldsymbol{\beta}_0^\top \boldsymbol{Z}_i)\left\{d\hat{\Lambda}_{10}^*(u)-d\Lambda_{10}^*(u)\right\}+o_p(1). \tag{9}$$

Taking the Taylor expansion of $\boldsymbol{U}(\boldsymbol{\beta}, \hat{t})$ at $\boldsymbol{\beta}_0$, we can obtain

$$n^{1/2}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)=\boldsymbol{\Omega}^{-1}\{n^{-1/2}\boldsymbol{U}(\boldsymbol{\beta}_0)\}+o_p(1),$$

where $\boldsymbol{\Omega} = \lim_{n\to\infty} \boldsymbol{I}(\boldsymbol{\beta}_0)/n$ and asymptotically $\boldsymbol{U}(\boldsymbol{\beta}_0)$ can be expressed as the sum of n independent and identically distributed random variables, i.e. $n^{-1/2}\boldsymbol{U}(\boldsymbol{\beta}_0)=n^{-1/2}\sum_{i=1}^{n}(\boldsymbol{\eta_i}+\boldsymbol{\psi_i})+o_p(1)$ (for explicit expressions of $\boldsymbol{\eta}_i$ and $\boldsymbol{\psi}_i$ see Fine and Gray (1999)). Since $\boldsymbol{\eta}_i$ contributes the majority of the variability, so we call $\boldsymbol{\eta}_i$ major term and $\boldsymbol{\psi}_i$ minor term. Both $\boldsymbol{\eta}_i$ and $\boldsymbol{\psi}_i$ are zero-mean Gaussian processes. Therefore

$$n^{1/2}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)=\boldsymbol{\Omega}^{-1}\left\{n^{-1/2}\sum_{i=1}^{n}(\boldsymbol{\eta_i}+\boldsymbol{\psi_i})\right\}+o_p(1). \tag{10}$$

Recall that

$$\hat{\Lambda}_{10}^*(t)=\sum_{i=1}^{n}\int_0^t \frac{w_i(u,\hat{G}_c)dN_i^1(u)}{S_0(\hat{\boldsymbol{\beta}},u)},$$

therefore

$$n^{1/2}\left\{\hat{\Lambda}^*_{10}(t)-\Lambda^*_{10}(t)\right\}=n^{1/2}\sum_{i=1}^{n}\int_0^t\left\{\frac{1}{S_0(\hat{\boldsymbol{\beta}},u)}-\frac{1}{S_0(\boldsymbol{\beta}_0,u)}\right\}w_i(u,\hat{G}_c)dN_i^1(u) \quad (11)$$

$$+n^{1/2}\left\{\sum_{i=1}^{n}\int_0^t\frac{1}{S_0(\boldsymbol{\beta}_0,u)}w_i(u,\hat{G}_c)dN_i^1(u)-\Lambda^*_{10}(t)\right\}. \quad (12)$$

Further more, for (11), taking the Taylor expansion of $1/S_0(\hat{\boldsymbol{\beta}},u)$ at $\boldsymbol{\beta}_0$, we can obtain

$$(11)=-\sum_{i=1}^{n}\int_0^t\frac{\boldsymbol{S}_1^{\top}(\boldsymbol{\beta}_0,u)}{\{S_0(\boldsymbol{\beta}_0,u)\}^2}w_i(u,\hat{G}_c)dN_i^1(u)n^{1/2}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)+o_p(1). \quad (13)$$

For (12),

$$\begin{aligned}(12)&=n^{1/2}\sum_{i=1}^{n}\int_0^t\frac{w_i(u,\hat{G}_c)dM_i^1(u)}{S_0(\boldsymbol{\beta}_0,u)}\\&=n^{1/2}\sum_{i=1}^{n}\int_0^t\frac{w_i(u,G_c)dM_i^1(u)}{S_0(\boldsymbol{\beta}_0,u)}+n^{1/2}\sum_{i=1}^{n}\int_0^t\frac{\{w_i(u,\hat{G}_c)-w_i(u,G_c)\}dM_i^1(u)}{S_0(\boldsymbol{\beta}_0,u)}\\&=n^{1/2}\sum_{i=1}^{n}\int_0^t\frac{w_i(u,G_c)dM_i^1(u)}{S_0(\boldsymbol{\beta}_0,u)}-n^{1/2}\sum_{i=1}^{n}\int_0^{\infty}\frac{Q(u,t)}{\sum_{l=1}^{n}(X_i\geq u)}dM_i^c(u)+o_p(1),\end{aligned}$$

where

$$\begin{aligned}Q(u,t)&=\sum_{j=1}^{n}\int_0^t\frac{(X_j\leq u\leq s)}{S_0(\boldsymbol{\beta}_0,s)}w_i(s,\hat{G}_c)dM_i^1(s),\\dM_i^c(u)&=\mathbb{1}(X_i\leq t,\Delta_i=0)-\int_0^t\mathbb{1}(X_i\geq u)d\hat{\Lambda}^c(u),\\\hat{\Lambda}^c(u)&=\int_0^t\frac{d\sum_{i=1}^{n}(X_i\leq t,\Delta_i=0)}{\sum_{i=1}^{n}(X_i\geq u)}.\end{aligned}$$

Plug (11) and (12) into (9) we have

$$(9)=-\sum_{i=1}^{n}\int_0^t f(x,\boldsymbol{Z}_i,\boldsymbol{v},p)\frac{w_i(u,\hat{G}_c)Y_i^1(u)\exp(\boldsymbol{\beta}_0^{\top}\boldsymbol{Z}_i)}{S_0(\boldsymbol{\beta}_0,u)}\sum_{l=1}^{n}w_l(u,\hat{G}_c)dM_l^1(u) \quad (14)$$

$$+\sum_{i=1}^{n}\int_0^t f(x,\boldsymbol{Z}_i,\boldsymbol{v},p)w_i(u,\hat{G}_c)Y_i^1(u)\exp(\boldsymbol{\beta}_0^{\top}\boldsymbol{Z}_i)\frac{\boldsymbol{S}_1^{\top}(\boldsymbol{\beta}_0,u)}{S_0(\boldsymbol{\beta}_0,u)}d\Lambda^*_{10}(u)\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\right) \quad (15)$$

$$-\sum_{i=1}^{n}\int_0^{\infty}f(x,\boldsymbol{Z}_i,\boldsymbol{v},p)w_i(u,\hat{G}_c)Y_i^1(u)\exp(\boldsymbol{\beta}_0^{\top}\boldsymbol{Z}_i)\frac{\sum_{j=1}^{n}Q(u,t)dM_j^c(u)}{\sum_{l=1}^{n}\mathbb{1}(X_i\geq u)}+o_p(1). \quad (16)$$

Under the asymptotic regularity conditions:

$$\lim_{n\to\infty} n^{-1}\boldsymbol{S}_k(\boldsymbol{\beta},t) \xrightarrow{p} \boldsymbol{s}_k(\boldsymbol{\beta},t),\ \lim_{n\to\infty} Q(u,t) \xrightarrow{P} q(u,t),\ \lim_{n\to\infty} n^{-1}\sum_{l=1}^{n}\mathbb{1}(X_i \geq u) \xrightarrow{p} y(t)$$

Exchange summation on (14) and combine with (7). When $n \xrightarrow{p} \infty$,

$$(7)+(14)=n^{-1}\sum_{i=1}^{n}\int_0^t \{f(x,\boldsymbol{Z}_i,\boldsymbol{v},p)-g(\boldsymbol{\beta}_0,u,x)\}\, w_i(u,G_c)dM_i^1(u)+o_p(1),$$

where

$$g(\boldsymbol{\beta}_0,u,x)=\frac{1}{s_0(\boldsymbol{\beta}_0,u)}\sum_{l=1}^{n} f(x,\boldsymbol{Z}_l,\boldsymbol{v},p)w_l(u,G_c)Y_l^1(u)\exp(\boldsymbol{\beta}_0^\top \boldsymbol{Z}_l).$$

Combine (8) and (15). When $n \xrightarrow{p} \infty$,

$$\begin{aligned}(8)+(15)&=-\sum_{i=1}^{n}\int_0^t f(x,\boldsymbol{Z}_i,\boldsymbol{v},p)w_i(u,G_c)Y_i^1(u)\exp(\boldsymbol{\beta}_0^\top \boldsymbol{Z}_i)\left\{\boldsymbol{Z}_i-\tfrac{s_1(\boldsymbol{\beta}_0,u)}{s_0(\boldsymbol{\beta}_0,u)}\right\}^\top d\Lambda_{10}^*(u)\ \left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\right)+o_p(1)\\ &=n^{-1}\sum_{i=1}^{n}\left\{\boldsymbol{C}^\top(\boldsymbol{\beta}_0,t,x)\right\}\ \boldsymbol{\Omega}^{-1}(\boldsymbol{\eta_i}+\boldsymbol{\psi_i})+o_p(1),\end{aligned}$$

where

$$\boldsymbol{C}(\boldsymbol{\beta}_0,t,x)=-\sum_{i=1}^{n}\int_0^t f(x,\boldsymbol{Z}_i,\boldsymbol{v},p)w_i(u,G_c)Y_i^1(u)\exp(\boldsymbol{\beta}_0^\top \boldsymbol{Z}_i)\ \left\{\boldsymbol{Z}_i-\frac{s_1(\boldsymbol{\beta}_0,u)}{s_0(\boldsymbol{\beta}_0,u)}\right\} d\Lambda_{10}^*(u).$$

Exchange summation on (16). When $n \xrightarrow{p} \infty$,

$$(16)=n^{-1}\sum_{i=1}^{n}\int_0^\infty \frac{q(u,t)}{y(u)}v(\boldsymbol{\beta}_0,u,x)dM_i^c(u)+o_p(1)=n^{-1}\sum_{i=1}^{n}q_i^{(f)}(t,x)+o_p(1)$$

where

$$v(\boldsymbol{\beta}_0,u,x)=\sum_{j=1}^{n} f(x,\boldsymbol{Z}_j,\boldsymbol{v},p)w_j(u,\hat{G}_c)Y_j^1(u)\exp(\boldsymbol{\beta}_0^\top \boldsymbol{Z}_j).$$

So

$$B(t,x)=n^{-1}\sum_{i=1}^{n}\{W_i(t,x)\}+o_p(1),$$

where

$$W_i(t,x)=\int_0^t\{f(x,\boldsymbol{Z}_i,\boldsymbol{v},p)-g(\boldsymbol{\beta}_0,u,x)\}\,w_i(u,G_c)dM_i^1(u)+\boldsymbol{C}^\top(\boldsymbol{\beta}_0,t,x)\boldsymbol{\Omega}^{-1}(\boldsymbol{\eta_i}+\boldsymbol{\psi_i})-q_i^{(f)}(t,x), \quad (17)$$

which can be consistently estimated by the plug-in estimators.

Test proportional subdistribution hazards assumption:

To check the proportional subdistribution hazards assumption, we consider the score process $U_j(\hat{\boldsymbol{\beta}},\hat{t})$ for each covariate, which can be written as

$$U_j(\hat{\boldsymbol{\beta}},t)=\sum_{i=1}^{n}\int_0^t Z_{ij}w_i(u,\hat{G}_c)d\hat{M}_i^1(u), j=1,\cdots,m.$$

It is a special case of the general form $\boldsymbol{B}(t,x)$ with $x=\infty$, $p=1$ and $\boldsymbol{v}$ has 1 in jth element and 0 elsewhere. Under the null hypothesis,

$$U_j(\hat{\boldsymbol{\beta}},t)=n^{-1}\sum_{i=1}^{n}\left[\int_0^t\left\{Z_{ij}-\frac{s_{1j}(\boldsymbol{\beta}_0,u)}{s_0(\boldsymbol{\beta}_0,u)}\right\}w_i(u,G_c)dM_i^1(u)+\boldsymbol{C}_j^\top(t)\boldsymbol{\Omega}^{-1}(\boldsymbol{\eta}_i+\boldsymbol{\psi}_i)-q_i^{(f)}(t,x)+o_p(1),\right.$$

where $s_{1j}(\boldsymbol{\beta}_0,u)$ is the jth element of $\boldsymbol{s}_1(\boldsymbol{\beta},u)$ and

$$\boldsymbol{C}_j(t)=-\sum_{i=1}^{n}\int_0^t Z_{ij}w_i(u,G_c)Y_i^1(u)\exp(\boldsymbol{\beta}_0^\top\boldsymbol{Z}_i)\left\{\boldsymbol{Z}_i-\frac{\boldsymbol{s}_1(\boldsymbol{\beta}_0,u)}{s_0(\boldsymbol{\beta}_0,u)}\right\}d\Lambda_{10}^*(u).$$

In practice, we standardized the process by multiplying $I_{jj}^{-1}(\hat{\boldsymbol{\beta}})$, and denoted $B_j^{(p)}(t)=I_{jj}^{-1}(\hat{\boldsymbol{\beta}})U_j(\hat{\boldsymbol{\beta}},t)$ in the text.

Test linear functional form

To test the linear functional form for the jth covariate, we consider

$$B_j^{(f)}(x)=\sum_{i=1}^{n}\int_0^\infty\mathbb{1}\{Z_{ij}\le x\}w_i(u,\hat{G}_c)d\hat{M}_i^1(u), j=1,\cdots,m$$

which is a special case of the general form $\boldsymbol{B}(t, x)$ when $t = \infty$, $p = 0$, and $\boldsymbol{v}$ has 1 in jth element and 0 elsewhere. Under the null hypothesis,

$$B_j^{(f)}(x)=n^{-1}\sum_{i=1}^{n}[\int_0^{\infty}\{\mathbb{1}(Z_{ij}\le x)-g_j(u,x)\}w_i(u,G_c)dM_i(u)+\boldsymbol{C}_j^{\top}(x)\boldsymbol{\Omega}^{-1}(\boldsymbol{\eta}_i+\boldsymbol{\psi}_i)-q_i^{(f)}(t,x)+o_p(1),$$

where

$$g_j(u,x)=\frac{1}{s_0(\boldsymbol{\beta}_0,u)}\sum_{l=1}^{n}\mathbb{1}(Z_{lj}\le x)w_l(u,G_c)Y_l^1(u)\exp(\boldsymbol{\beta}_0^{\top}\boldsymbol{Z}_l),$$
$$\boldsymbol{C}_j(x)=-\sum_{l=1}^{n}\int_0^{\infty}\mathbb{1}(Z_{lj}\le x)w_l(u,G_c)Y_l^1(u)\exp(\boldsymbol{\beta}_0^{\top}\boldsymbol{Z}_l)\left\{\boldsymbol{Z}_l-\frac{\boldsymbol{s}_1(\boldsymbol{\beta}_0,u)}{s_0(\boldsymbol{\beta}_0,u)}\right\}d\Lambda_{10}^{*}(u).$$

Test link function

To test the link function, we consider

$$B^{(l)}(x)=\sum_{i=1}^{n}\int_0^{\infty}\mathbb{1}\left\{\hat{\boldsymbol{\beta}}^{\top}\boldsymbol{Z}_i\le x\right\}w_i(u,\hat{G}_c)d\hat{M}_i^1(u).$$

In this case, $t = \infty$, $p = 0$ and $\boldsymbol{v} = \hat{\boldsymbol{\beta}}$. Under the null hypothesis, the test statistic has exactly the same format as the test statistic in the linear functional form, except the indictor function is replaced by $\mathbb{1}\{\hat{\boldsymbol{\beta}}^{\top}\boldsymbol{Z}_i \le x\}$. Note, if there is only one covariate, checking the function is equivalent to checking the functional form of the covariate.

Omnibus test

Here we consider

$$\boldsymbol{B}^{(o)}(t,\boldsymbol{z})=\sum_{i=1}^{n}\int_0^{t}\mathbb{1}\{\boldsymbol{Z}_i\le\boldsymbol{z}\}w_i(u,\hat{G}_c)d\hat{M}_i^1(u).$$

For the jth covariate is of interest for testing, the jth element in $\boldsymbol{B}^{(o)}(t, x)$ is the statistic we look for,

$$B_j^{(o)}(t,x)=\sum_{i=1}^{n}\int_0^{t}\mathbb{1}\{Z_{ij}\le x\}w_i(u,\hat{G}_c)d\hat{M}_i^1(u),$$

which is a special case of the general form $\boldsymbol{B}(t, x)$ when $p = 0$ and and $\boldsymbol{v}$ has 1 in jth element and 0 elsewhere. The omnibus test also can be viewed as recording each linear function test through the time span. Under the null hypothesis,

$$B_j^{(o)}(t,x)=n^{-1}\sum_{i=1}^{n}\Big[\int_0^t\{\mathbb{1}\,(Z_{ij}\le x)-g_j(u,x)\}w_i(u,G_c)dM_i(u)+\sum_{i=1}^{n}\boldsymbol{C}_j^{\top}(t,x)\boldsymbol{\Omega}^{-1}(\boldsymbol{\eta}_i+\boldsymbol{\psi}_i)-q_i^{(f)}(t,x)+o_p(1),$$

where

$$g_j(u,x)=\frac{1}{s_0(\boldsymbol{\beta}_0,u)}\sum_{l=1}^{n}\mathbb{1}\,(Z_{lj}\le x)\,w_l(u,G_c)Y_l^1(u)\exp(\boldsymbol{\beta}_0^{\top}\boldsymbol{Z}_l),$$

$$\boldsymbol{C}_j(t,x)=-\sum_{l=1}^{n}\int_0^t\mathbb{1}\,(Z_{lj}\le x)\,w_l(u,G_c)Y_l^1(u)\exp(\boldsymbol{\beta}_0^{\top}\boldsymbol{Z})\,\Big\{\boldsymbol{Z}_l-\frac{\boldsymbol{s}_1(\boldsymbol{\beta}_0,u)}{s_0(\boldsymbol{\beta}_0,u)}\Big\}\,d\Lambda_{10}^*(u).$$

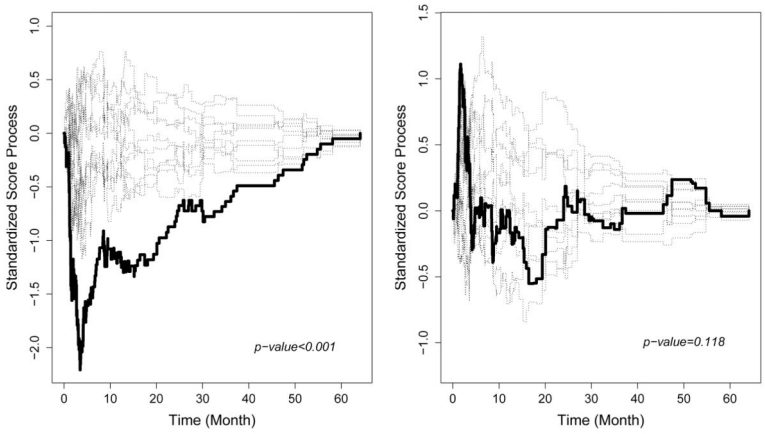

Fig. 1.
Plot of Standardized Score Process for GP and AGE in the BMT Data

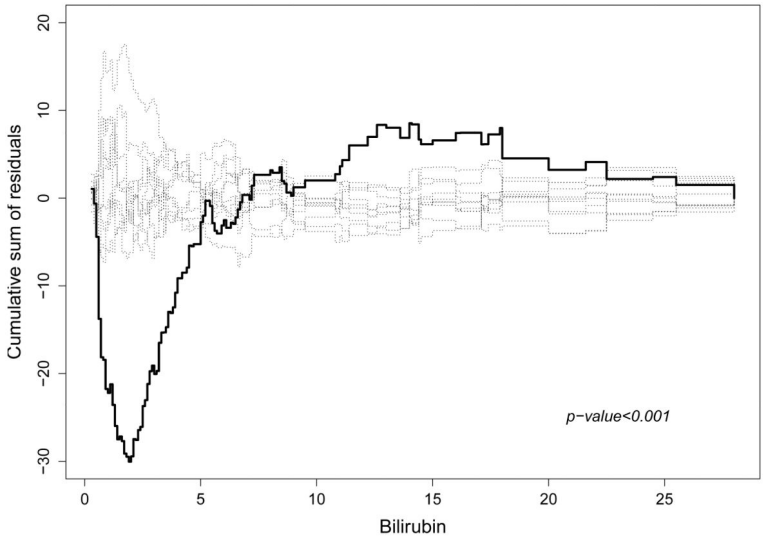

Fig. 2.
Plot of Cumulative Residuals vs bilirubin in the PBC Data

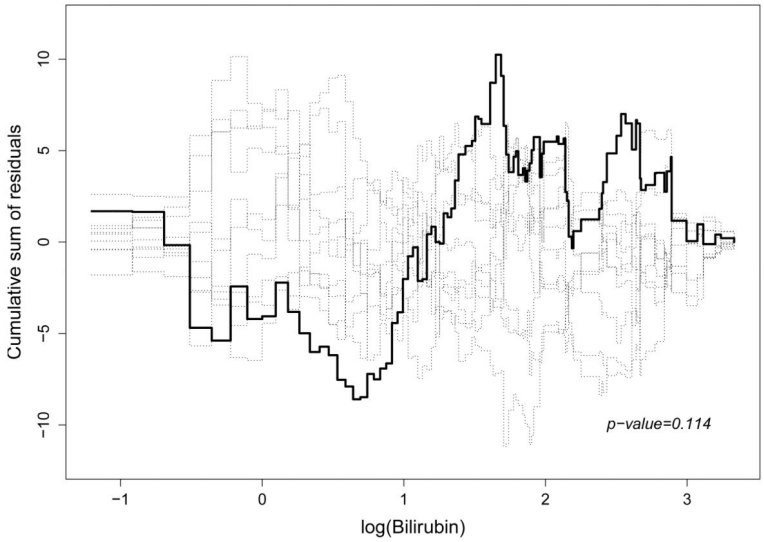

Fig. 3.
Plot of Cumulative Residuals vs log(bilirubin) in the PBC Data

Table 1

Type I Error for Testing the Proportional Hazards Assumption

n	*c*(%)	Type I
50	15	0.0624
50	30	0.0642
100	15	0.0606
100	30	0.0564
300	15	0.0536
300	30	0.0536

Table 2

Estimated Power for Testing Proportional Hazards Assumption

n	c(%)	Proposed	t	t^2	log(t)
50	15	0.3260	0.4305	0.2800	0.2700
50	30	0.2100	0.2815	0.2215	0.1795
100	15	0.5560	0.7840	0.5490	0.4495
100	30	0.3920	0.5870	0.4405	0.3085
300	15	0.9590	0.9985	0.9295	0.8910
300	30	0.8420	0.9780	0.8775	0.7365

Table 3

Estimated Power for Testing Proportional Hazards Assumption

n	c(%)	Proposed	t	t^2	log(t)
50	15	0.3510	0.2450	0.1450	0.2565
50	30	0.2045	0.1680	0.1385	0.1785
100	15	0.5770	0.4350	0.2275	0.4050
100	30	0.3760	0.3215	0.2050	0.2560
300	15	0.9715	0.9125	0.4930	0.8775
300	30	0.8025	0.7620	0.3835	0.5905

Table 4

Type I Error and Estimated Power for Testing the Functional Form and the Link Function

n	*c*(%)	Type I	Test Function Form	Test Link Function
50	15	0.0568	0.1575	0.1325
50	30	0.0556	0.2055	0.1350
100	15	0.0514	0.2505	0.2240
100	30	0.0566	0.3945	0.2170
300	15	0.0478	0.6055	0.7210
300	30	0.0506	0.8645	0.6755

Table 5

p -values of Checking Fine and Gray's Model in BMT Data

	GP	AGE	overall
Proportionality	< 0.001	0.118	0.001
Functional form	–	0.896	-
Link function	-	-	0.954

Table 6

p -values of Checking the Fine and Gray's Model in the PBC Data

	age	oedema	log(bili)	log(albumin)	log(protime)	overall
prop	0.469	0.019	0.071	0.119	0.007	0.006
func	0.180	0.349	0.114	0.515	0.239	-
link	-	-	-	-	-	0.290