Published in final edited form as:

Lifetime Data Anal. 2018 July ; 24(3): 407–424. doi:10.1007/s10985-017-9400-9.

Group and within-group variable selection for competing risks data

Kwang Woo Ahn,

Division of Biostatistics, Medical College of Wisconsin, 8701 Watertown Plank Road, Milwaukee, WI 53226, U.S.A

Anjishnu Banerjee,

Division of Biostatistics, Medical College of Wisconsin, 8701 Watertown Plank Road, Milwaukee, WI 53226, U.S.A

Natasha Sahr, and

Division of Biostatistics, Medical College of Wisconsin, 8701 Watertown Plank Road, Milwaukee, WI 53226, U.S.A

Soyoung Kim

Division of Biostatistics, Medical College of Wisconsin, 8701 Watertown Plank Road, Milwaukee, WI 53226, U.S.A

Abstract

Variable selection in the presence of grouped variables is troublesome for competing risks data: while some recent methods deal with group selection only, simultaneous selection of both groups and within-group variables remains largely unexplored. In this context, we propose an adaptive group bridge method, enabling simultaneous selection both within and between groups, for competing risks data. The adaptive group bridge is applicable to independent and clustered data. It also allows the number of variables to diverge as the sample size increases. We show that our new method possesses excellent asymptotic properties, including variable selection consistency at group and within-group levels. We also show superior performance in simulated and real data sets over several competing approaches, including group bridge, adaptive group lasso, and AIC/BIC-based methods.

Keywords

Adaptive penalty; clustered data; competing risks data; group bridge

1 Introduction

Competing risks are one of the more common types of time-to-event data studied in clinical research. Classically, competing risks have been defined as the scenario when there is a

Tel.: +1-414-955-7387, Fax: +1-414-955-0176, kwoohahn@mcw.edu.

Supplementary Materials

The proofs of Theorem 2 and Corollary 1 and a list of variables for the bone marrow transplant data are available online.

possibility of more than one type of event occurring at a given time, with occurrence of any one event precluding the occurrence of all others (Fine and Gray, 1999). For example, Shaw et al (2010) studied pediatric patients, who were less than 18 years old, undergoing allogeneic Treplete, myeloablative bone marrow transplantation between 1993 and 2006. They studied outcomes including relapse, treatment-related mortality (TRM), and graft-versus-host disease (GVHD), where relapse and TRM are competing risks to each other and death is a competing risk for GVHD.

The Cox proportional hazards model can be used to model cause-specific hazards (Prentice et al, 1978) for competing risks data. However, the effect of a factor on the cause specific hazard for a particular type of failure can be quite different than its effect on the cumulative incidence of that type of failure (Gray, 1988). The Fine-Gray model proposed by Fine and Gray (1999) overcomes this limitation of the cause-specific hazards model to fit the cumulative incidence function. Zhou et al (2012) recently extended the Fine-Gray model to the marginal Fine-Gray model for clustered competing risks data.

There is rich literature on group variable selection for linear regression and the Cox proportional hazards model. Yuan and Lin (2006) proposed the group lasso for linear regression models, where it selects variables at group level, but not at within-group individual level. Huang et al (2009), Seetharaman (2013), and Huang et al (2014) studied the group bridge for linear regression and the Cox proportional hazards models. Although the models proposed by Huang et al (2009) and Huang et al (2014) appeared to select variables at both group and within-group individual variable levels in their simulation studies, they only studied group variable selection consistency, while ignoring within-group variable selection consistency. An alternative approach by Wu and Wang (2013) considered the Cox regression model with double penalties: lasso and group lasso penalties. However, this approach introduces considerable additional computation burden, due to the need to simultaneously search two optimal tuning parameters. Additionally, asymptotic properties of this model were not studied by the authors.

There are several literature on variable selection for the Fine-Gray model. Kuk and Varadhan (2013) proposed an AIC/BIC/BICc-based variable selection for the Fine-Gray model. This method can select variables at individual level for continuous and binary variables, and at group level for categorical variables, but it does not select at individual level within-group variables. Ha et al (2014) studied a penalized-hierarchical-likelihood-based variable selection method at individual levels, not at group levels, for subdistribution hazard frailty models. Recently, Fu et al (2016a) and Fu et al (2016b) extensively studied lasso, adaptive lasso, SCAD, and MCP for individual variable selection and their group variable selection versions for independent and clustered competing risks data. However, they did not address within-group variable selection. In addition, their oracle properties were limited to a fixed number of variables.

We propose an adaptive group bridge method which alleviates several of the deficiencies in the previously mentioned literature for competing risks data. Our method enables variable selection at both group and within-group variable levels. We also show the consistency of such selection at both levels while allowing the number of variables to diverge as the sample

size increases. Our model is computationally efficient, relying on one tuning parameter. Additionally, our model is flexible enough to be applicable in case of both independent and clustered competing risks data. The adaptive group bridge shows several orders of magnitude of improvement in variable selection accuracy with simulated data when compared to some competing approaches. We demonstrate similar improvement with a real data example in the context of bone marrow transplant patients.

The rest of the paper is organized as follows: In Section 2, we elaborate on the details of our proposed method and study its theoretical properties. Next in Section 3, we implement our method with different scenarios, along with some competing approaches and compare performance. We finally consider our real data example and end with a brief conclusion.

2 Methods

2.1 Notations

We begin with some notations that are used through the rest of the paper. Without loss of generality, we consider two causes of failure $e \in \{1, 2\}$. We assume that there are m clusters and the i th cluster has n_i individuals, where $n = \sum_{i=1}^m n_i$ is the total sample size. We allow the number of variables d_m to increase as m increases. Let T_{ij} , C_{ij} , e_{ij} and $\mathbf{Z}_{ij} = (Z_{ij1}, \dots, Z_{ijd_m})^T$ be the event time, censoring time, cause of failure, and variable vector of individual j in cluster i , respectively, for $i = 1, \dots, m$ and $j = 1, \dots, n_i$. Let $\mathbf{T}_i = \{T_{ij}, j = 1, \dots, n_i\}$, $\mathbf{C}_i = \{C_{ij}, j = 1, \dots, n_i\}$, $\mathbf{e}_i = \{e_{ij}, j = 1, \dots, n_i\}$, and $\mathbf{Z}_i = \{\mathbf{Z}_{ij}, j = 1, \dots, n_i\}$. Suppose that $(\mathbf{T}_i, \mathbf{e}_i, \mathbf{C}_i, \mathbf{Z}_i)$ are independent and identically distributed. We assume that the C_{ij} 's do not depend on the $\mathbf{Z}_{ij}$'s and the (T_{ij}, e_{ij}) 's are independent of the C_{ij} 's given $\mathbf{Z}_{ij}$ for $i = 1, \dots, m$ and $j = 1, \dots, n_i$. This setting allows the event times to be correlated within the same cluster. Similarly, the censoring times may be correlated within the same cluster. The C_{ij} 's are assumed to have a common distribution G and $\hat{G}(t)$ denotes the Kaplan-Meier estimates of $G(t)$. Let $X_{ij} = T_{ij} \wedge C_{ij}$ be the observed time and $\delta_{ij} = \mathbb{I}(T_{ij} < C_{ij})\mathbb{I}(e_{ij} = 1)$, where $a \wedge b = \min(a, b)$. Let $N_{ij}(t) = \mathbb{I}(T_{ij} \leq t, \delta_{ij} = 1)$ and $Y_{ij}(t) = 1 - N_{ij}(t-)$. Let $F_1(t|\mathbf{Z}_{ij})$ be the cumulative incidence of cause 1 at time t given $\mathbf{Z}_{ij}$, where $F_1(t|\mathbf{Z}_{ij}) = P(T_{ij} \leq t, e_{ij} = 1 | \mathbf{Z}_{ij})$. Define the subdistribution hazard of cause 1 given variable $\mathbf{Z}_{ij}$:

$$\lambda_1(t | \mathbf{Z}_{ij}) = - \frac{d[\log \{1 - F_1(t | \mathbf{Z}_{ij})\}]}{dt}.$$

Then, $F_1(t | \mathbf{Z}_{ij}) = 1 - \exp\{-\int_0^t \lambda_1(u | \mathbf{Z}_{ij}) du\}$. Zhou et al (2012) extended Fine and Gray (1999) to the marginal subdistribution hazards model for clustered competing risks data: $\lambda_1(t|\mathbf{Z}_{ij}) = \lambda_{10}(t) \exp(\boldsymbol{\beta}^T \mathbf{Z}_{ij})$, where $\lambda_{10}(t)$ is an unspecified baseline subdistribution hazard function and $\boldsymbol{\beta} = (\beta_1, \dots, \beta_{d_m})^T$ is a parameter vector of interest. Let $\boldsymbol{\beta}_0 = (\beta_{10}, \dots, \beta_{d_{m0}})^T$ be the true parameter vector.

Next, we define some notations on group variables and their memberships. Assume that we have K groups of variables. Let $A_1, \dots, A_K$ be subsets of $\{1, \dots, d_m\}$ representing group memberships of variables, where A_k 's may overlap. Define $\boldsymbol{\beta}_A = (\beta_j, j \in A)^T$ and $\boldsymbol{\beta}_{A0} = (\beta_{j0}, j \in A)^T$ for a set A . To distinguish the individual memberships between non-zero β_{j0} 's and

zero β_{j0} 's, we define B_1 and B_2 such that $\beta_{j0} = 0$ if $j \in B_1$ and $\beta_{j0} = 0$ if $j \in B_2$. To distinguish the group memberships between non-zero β_{A_k0} 's and zero β_{A_k0} 's, without loss of generality we further define E_1 and E_2 such that $E_1 = \cup_{k=1}^{K_1} A_k$ and $E_2 = \cup_{k=K_1+1}^K A_k$, where $\beta_{A_k0} = 0$ for $1 \leq k \leq K_1$ and $\beta_{A_k0} = 0$ for $K_1 + 1 \leq k \leq K$. The L_b norm of $\mathbf{a}$ is denoted by $\|\mathbf{a}\|_b$.

2.2 The group bridge

Following Zhou et al (2012), we assume that there exists a τ such that i) $t \in [0, \tau]$; and ii) $P(T_{ij} > \tau) > \delta > 0$ and $P(C_{ij} > \tau) > \delta_c > 0$ for all i and j . Let $l(\beta) = -\log L(\beta)$, where

$$L(\beta) = \prod_{i=1}^m \prod_{j=1}^{n_i} \left(\frac{\exp\{\beta^T \mathbf{Z}_{ij}(X_{ij})\}}{\sum_{i'=1}^m \sum_{j'=1}^{n_{i'}} Y_{i'j'}(X_{ij}) \hat{w}_{i'j'}(X_{ij}) \exp\{\beta^T \mathbf{Z}_{i'j'}(X_{ij})\}} \right)^{\Delta_{ij}},$$

where

$$\hat{w}_{ij}(t) = \frac{I(C_{ij} \geq T_{ij} \wedge t) \hat{G}(t)}{\hat{G}(X_{ij} \wedge t)}.$$

Let $U_1(\beta) = -l'(\beta)/\beta$. Zhou et al (2012) proposed the inverse weighted estimating equation to make inference on β for the marginal subdistribution hazards model:

$$U_1(\beta) = \sum_{i=1}^m \sum_{j=1}^{n_i} \int_0^\tau \{\mathbf{Z}_{ij}(u) - \mathbf{H}(\beta, u)\} \hat{w}_{ij}(u) dM_{ij}(\beta, u),$$

where $M_{ij}(\beta, t) = \int_0^t dN_{ij}(u) - \int_0^t Y_{ij}(u) \lambda_{10}(u) e^{\beta^T \mathbf{Z}_{ij}(u)} du$. Zhou et al (2012) studied the consistency and the asymptotic normality of the estimator of β when d_m is fixed with respect to m .

For the group bridge method, we propose the following group penalized function to estimate β :

$$\ell_m(\beta) = l(\beta) + \lambda_m \sum_{k=1}^K c_k \left(\sum_{j \in A_k} |\beta_j| \right)^\gamma, \quad (1)$$

where $0 < \gamma < 1$ and $\lambda_m > 0$. In practice, we set $c_k \propto |A_k|^{1-\gamma}$ (Huang et al, 2014), where $|A|$ is the cardinality of A . The group bridge (1) can be used for variable selection at the group

and individual variable levels simultaneously. Following Huang et al (2009) and Huang et al (2014), we can formulate minimizing $\ell_m(\boldsymbol{\beta})$ to minimizing

$$\ell(\boldsymbol{\beta}) + \sum_{k=1}^K \theta_k^{1-1/\gamma} c_k^{1/\gamma} \sum_{j \in A_k} |\beta_j| + \zeta_m \sum_{k=1}^K \theta_k,$$

where $\lambda_m = \zeta_m^{1-\gamma} \gamma^{-\gamma} (1-\gamma)^{\gamma-1}$. As Huang et al (2009) discussed, a weighted L_1 penalty allows a sparse solution in $\boldsymbol{\beta}$. In addition, small θ_k will force $\boldsymbol{\beta}_{A_k} = \mathbf{0}$ so it allows group variable selection.

We have the following theorem for the group bridge estimator:

Theorem 1—*In addition to Conditions (C1)–(C5) in the Appendix, we further assume*

$$\begin{aligned} (C6') \quad & \frac{\lambda_m^2}{m} \sum_{k=1}^K c_k^2 \|\boldsymbol{\beta}_{0A_k}\|_{2\gamma-2}^2 |A_k| \leq d_m M_m, M_m = O_p(1), \text{ where } \min_{1 \leq k \leq K} c_k \rightarrow 1 \\ & \text{and } \lambda_m / \{m^{\gamma/2} \kappa_{\max}(\Omega) d_m^{1-\gamma/2}\} \rightarrow \infty \text{ as } m \rightarrow \infty; \\ (C7') \quad & \lambda_m m^{-1/2} \rightarrow 0, \\ & \lambda_m m^{-1/2} \rightarrow 0, 1/\kappa_{\min}(\Omega) + \kappa_{\max}(\Omega) + \sum_{k=1}^K c_k^2 = O(1), \lambda_m / (m^{\gamma/2} d_m^{1-\gamma/2}) \rightarrow \infty \text{ as } \\ & m \rightarrow \infty \end{aligned}$$

Then, we have

1. Consistency: $\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|_2 = O_p(\sqrt{d_m/m})$;
2. Group variable selection consistency: $P(\hat{\boldsymbol{\beta}}_{E_2} = \mathbf{0}) \rightarrow 1$;
3. Asymptotic distribution: for fixed E_1 ,

$$\sqrt{m}(\hat{\boldsymbol{\beta}}_{E_1} - \boldsymbol{\beta}_{E_1}) \rightarrow_d N(\mathbf{0}, \Omega_{11}^{*-1} \Gamma_{11}^* \Omega_{11}^{*-1}),$$

where Ω_{11}^* and Γ_{11}^* are the leading $|E_1| \times |E_1|$ submatrices of Ω and Γ , respectively.

Conditions (C1)–(C5), (C6'), and (C7') are similar to Conditions (C1)–(C7) of Huang et al (2014). Therefore, the proof of Theorem 1 can be similarly shown as in Theorems 1 and 2 of Huang et al (2014) and thus omitted. Theorem 1 shows the group variable selection consistency and $\sqrt{m/d_m}$ -consistency of the group bridge method.

2.3 The adaptive group bridge

Although the group bridge can be used for bi-level variable selection, it still lacks the within-group variable selection consistency. To overcome this limitation of the group bridge, we propose the following group penalized function to estimate $\boldsymbol{\beta}$:

$$\mathcal{L}_m(\boldsymbol{\beta}) = l(\boldsymbol{\beta}) + \lambda_m \sum_{k=1}^K c_k \left(\sum_{j \in A_k} \frac{|\beta_j|}{|\tilde{\beta}_j|^\nu} \right)^\gamma, \quad (2)$$

where $0 < \gamma < 1$, $\lambda_m > 0$, $\nu > 0$, and $\tilde{\beta}_j$ is a $\sqrt{m/d_m}$ -consistent estimator for β_j . In practice, $\tilde{\boldsymbol{\beta}} = (\tilde{\beta}_1, \dots, \tilde{\beta}_{d_m})^T$ can be obtained from the marginal Fine-Gray model of Zhou et al (2012) or the group bridge, see below for the consistency of the marginal Fine-Gray model. Let $\hat{\boldsymbol{\beta}}$ be the estimator of $\boldsymbol{\beta}$, where $\hat{\boldsymbol{\beta}}$ is obtained by minimizing $\mathcal{L}_m(\boldsymbol{\beta})$. The case that $|A_k| = 1$ for all k can be handled as a special case. We call this group penalty as adaptive group bridge penalty. Because $l(\boldsymbol{\beta})/\|\boldsymbol{\beta}\| = \mathbf{U}_1(\boldsymbol{\beta})$, the proposed method can be considered as the penalized marginal Fine-Gray model with the adaptive group bridge penalty.

Remark 1—The group bridge has $\nu = 0$ in (2). Thus, given group k , the group bridge equally penalizes β_j 's for $j \in A_k$ in the L_1 penalty. On the other hand, motivated by the adaptive lasso (Zou, 2006), the adaptive group bridge imposes different penalties, which are proportional to $1/|\tilde{\beta}_j|^\nu$, on different variables within group for some $\nu > 0$. As m grows, $1/|\tilde{\beta}_j|^\nu$ for $\beta_{j0} = 0$ goes to infinity while $1/|\tilde{\beta}_j|^\nu$ for non-zero β_{j0} converges to $\beta_{j0}^{-\nu}$ that is a constant. Thus, $1/|\tilde{\beta}_j|^\nu$ assigns larger penalties to zero parameters so they are shrunk more towards zero. This added flexibility enables non-zero variables within group to be selected more consistently for an appropriate λ_m . In practice, $\nu = 1$ to 2 is used (Zou, 2006; Wang et al, 2013).

Following Huang et al (2014), we can formulate minimizing $\mathcal{L}_m(\boldsymbol{\beta})$ to minimizing

$$Q_m(\boldsymbol{\beta}, \boldsymbol{\theta}) = l(\boldsymbol{\beta}) + \sum_{k=1}^K \theta_k^{1-1/\gamma} c_k^{1/\gamma} \sum_{j \in A_k} \frac{|\beta_j|}{|\tilde{\beta}_j|^\nu} + \zeta_m \sum_{k=1}^K \theta_k. \quad (3)$$

Then, we have the following lemma:

Lemma 1—Assume that $\lambda_m = \zeta_m^{1-\gamma} \gamma^{-\gamma} (1-\gamma)^{\gamma-1}$. Then, $\hat{\boldsymbol{\beta}}$ minimizes $\mathcal{L}_m(\boldsymbol{\beta})$ if and only if $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}})$ minimizes $Q_m(\boldsymbol{\beta}, \boldsymbol{\theta})$, where $\theta_k > 0$ and $\hat{\theta}_k > 0$ for $k = 1, \dots, K$.

Lemma 1 can be shown similarly to Proposition 1 of Huang et al (2009), so its proof is omitted.

We study the asymptotics of the adaptive group bridge for complete censored data and right-censored data. Censoring in complete censored data is only from administrative loss-to-follow up; that is, patients have not failed by the time the data are analyzed (Fine and Gray, 1999). The asymptotics for complete censored data can be similarly done to right-censored data, so it is omitted.

Denote $\rightarrow_d$ as convergence in distribution. Under Conditions (C1)–(C7) in the Appendix, we have the following theorem:

Theorem 2

1. *Consistency:* $\|\hat{\beta} - \beta_0\|_2 = O_p(\sqrt{d_m/m})$.
2. *Individual variable selection consistency:* $P(\hat{\beta}_{B_2} = 0) \rightarrow 1$.
3. *Asymptotic distribution:* for fixed B_1 , we have

$$\sqrt{m}(\hat{\beta}_{B_1} - \beta_{B_1}) \rightarrow_d N(0, \Omega_{11}^{-1} \Gamma_{11} \Omega_{11}^{-1}),$$

where Ω_{11} and Γ_{11} are the leading $|B_1| \times |B_1|$ submatrices of Ω and Γ , respectively.

The proof of Theorem 2 can be found in online Supplemental Materials. Theorem 2 provides the oracle property of the adaptive group bridge. In particular, it shows that the adaptive group bridge consistently selects not only non-zero group variables, but also non-zero individual variables within group. On the other hand, the group bridge of Huang et al (2014) and the methods of Fu et al (2016a) and Fu et al (2016b) did not have or did not show within-group variable selection consistency. We have the following corollary from Theorem 2:

Corollary 1—Let $\tilde{\beta}$ be the estimator from the marginal Fine-Gray model of Zhou et al (2012). Then, under Conditions (C1)–(C5) in the Appendix, we have $\|\hat{\beta} - \beta_0\|_2 = O_p(\sqrt{d_m/m})$.

The proof of Corollary 1 is provided in online Supplemental Materials.

To obtain $\hat{\beta}$, we minimize $Q_m(\beta, \theta)$ of (3). Define $\nabla \ell(\beta) = \ell(\beta)/\beta$ and $\nabla^2 \ell(\beta) = \nabla^2 \ell(\beta)/(\beta \beta^T)$. Define $\mathbf{X}$ such that $\nabla^2 \ell(\beta) = \mathbf{X}^T \mathbf{X}$, where $\mathbf{X}$ can be obtained by Cholesky decomposition. Let $\mathbf{Y} = (\mathbf{X}^T)^{-1} \{\nabla^2 \ell(\beta) \beta - \nabla \ell(\beta)\}$. We approximate $\ell(\beta)$ by $(\mathbf{Y} - \mathbf{X}\beta)^T (\mathbf{Y} - \mathbf{X}\beta)/2$ and minimize

$$\tilde{Q}_m(\beta, \theta) = \frac{1}{2}(\mathbf{Y} - \mathbf{X}\beta)^T (\mathbf{Y} - \mathbf{X}\beta) + \sum_{k=1}^K \theta_j^{1-1/\gamma} c_j^{1/\gamma} \sum_{j \in A_k} \frac{|\beta_j|}{|\tilde{\beta}_j|^\nu} + \zeta_m \sum_{k=1}^K \theta_k.$$

Then, the optimization algorithm is as follows:

1. Obtain an consistent estimator $\tilde{\beta}$ from the marginal Fine-Gray model or the group bridge.
2. Obtain the initial value $\beta^{(0)}$ from the marginal Fine-Gray model or the group bridge.
3. Compute $\mathbf{Y}$ and $\mathbf{X}$ using $\beta^{(j)}$ at the j th step.

4. Compute

$$\theta_k^{(i)} = c_k \left(\frac{1-\gamma}{\zeta_m^\gamma} \right)^\gamma \left(\sum_{j \in A_k} \frac{|\beta_j^{(i)}|}{|\tilde{\beta}_j|^\nu} \right)^\gamma, k = 1, \dots, K.$$

5. Obtain $\beta^{(i+1)}$ by minimizing $\tilde{Q}_m(\beta, \theta^{(i)})$ with respect to β .

6. Repeat (3)–(5) until $\|\beta^{(i+1)} - \beta^{(i)}\|_1 < 10^{-4}$.

The generalized cross validation (GCV) can be used to choose a tuning parameter ζ_m . See Huang et al (2014) for the detailed formulation of GCV.

To obtain the standard errors of $\hat{\beta}$, we use a quadratic approximation as in Fan and Li (2002) and Huang et al (2014). Let

$$\Xi(\hat{\beta}, \hat{\theta}) = \text{diag} \left\{ \sum_{j \in A_k} \hat{\theta}_k^{1-1/\gamma} c_k^{1/\gamma} \frac{I(\hat{\beta}_j \neq 0)}{|\hat{\beta}_j| |\tilde{\beta}_j|^\nu}, j = 1, \dots, d_m \right\},$$

where

$$\hat{\theta}_k = c_k \left(\frac{1-\gamma}{\zeta_m^\gamma} \right)^\gamma \left(\sum_{j \in A_k} \frac{|\hat{\beta}_j|}{|\tilde{\beta}_j|^\nu} \right)^\gamma, k = 1, \dots, K.$$

Then, the covariance matrix of $\hat{\beta}$ can be estimated by the sandwich formula:

$$\{ \nabla^2 l(\hat{\beta}) + \Xi(\hat{\beta}, \hat{\theta}) \}^{-1} \widehat{\text{Cov}} \{ \nabla l(\hat{\beta}) \} \{ \nabla^2 l(\hat{\beta}) + \Xi(\hat{\beta}, \hat{\theta}) \}^{-1}. \quad (4)$$

3 Simulation

We performed simulation studies under two group variable settings: i) group variables consisting of continuous variables; and ii) group variables consisting of continuous variables and categorical variables. For each setting, we considered both independent and clustered competing risks data. A positive stable frailty was used to generate correlated event times as in Logan et al (2011), where generating independent data was handled as a special case. Censoring times and event times were independently generated. Censoring times were correlated within a cluster. We considered the number of clusters $m = 50, 100$, and 200 with the same size of cluster $n_i = 4$. For each cluster, two independent random effects w and w_c were generated from a positive stable frailty distribution with parameter ψ , where the Laplace transformation of the standard positive stable distribution was $L(s) = \exp(-s^\psi)$ with $0 < \psi \leq 1$. Smaller ψ was used to generate a larger correlation between event times within the cluster. Independent event times were generated with $\psi = 1$. Three ψ values were used:

0.25 and 0.5 for clustered data, and 1 for independent data. These values of ψ are values of Kendall's tau 0.75, 0.5 and 0, respectively, under the survival setting (Logan et al, 2011). Because the simulation results with $\psi = 0.5$ and $\psi = 0.25$ were similar, the results with $\psi = 0.5$ were omitted in the paper. We generated competing risks event times and censoring times for each cluster from

$$\begin{aligned} F_1(t | w) &= 1 - \{1 - p(1 - e^{-t})\}^{\text{wexp}(\beta^T \mathbf{Z})}, \\ F_2(t | w) &= (1 - p)^{\text{wexp}(\beta^T \mathbf{Z})} (1 - e^{-t \exp(\alpha^T \mathbf{Z})}), \\ G(t | w_c) &= \exp(-w_c \lambda_c t). \end{aligned}$$

We selected p and λ_c to generate 40% cause 1 events, 30% cause 2 events, and 30% censoring. Each simulation was performed 500 iterations. We evaluated the mean squared error that was calculated by

$$\text{MSE} = \frac{1}{500} \sum_{i=1}^{500} \|\hat{\beta}^i - \beta_0\|_2^2,$$

where $\hat{\beta}^i$ is the estimator of β_0 at the i th iteration. We used the group bridge and the marginal Fine-Gray model to estimate $\tilde{\beta}$. For the adaptive group bridge, $\nu = 1, 1.5$, and 2 were examined. Because the results with $\nu = 1, 1.5$, and 2 were very similar, we reported the results with $\nu = 1$ only in the paper. The performance of the proposed method was compared to the methods of Fu et al (2016b), the group bridge, and BIC-based method in Sections 3.1 and 3.2. The AIC/BICc-based methods were also conducted. However, their results were worse than the BIC-based method and thus omitted in the paper. The BIC-based variable selection of Kuk and Varadhan (2013) was performed using the R package “crrstep” (Varadhan and Kuk, 2015). For the methods of Fu et al (2016b), $\ell(\beta)$ with the adaptive lasso and the adaptive group lasso penalties was considered using the R package “crrp” (Fu, 2015). In Section 3.3, we examine the performance of the proposed standard error formula (4).

3.1 Simulation 1: group variables consisting of continuous variables

In this section, we performed a simulation study on selection of group variables consisting of continuous variables. For $m = 50$, there were 8 groups each of which contains 3 continuous variables. Given each group, 3 continuous variables were generated from $N(\mu, \Sigma)$, where

$$\mu = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} 1 & 0.5 & 0.5 \\ 0.5 & 1 & 0.5 \\ 0.5 & 0.5 & 1 \end{pmatrix}.$$

Different groups were assumed to be independent. The true β and α were

$$\begin{aligned}
 (\beta_1, \dots, \beta_{12})^T &= (0.8, -0.8, 0.5, 0.7, -0.7, 0, -0.8, 0, 0, 0, 0, 0)^T, \\
 (\alpha_1, \dots, \alpha_{12})^T &= (0, 0, 0, -0.8, 0.7, 0.9, 0.8, -0.8, 0, 0.9, 0, 0)^T, \\
 (\beta_{10}, \dots, \beta_{24})^T &= (0, \dots, 0)^T, \quad (\alpha_{10}, \dots, \alpha_{24})^T = (0, \dots, 0)^T.
 \end{aligned}$$

Thus, Group 1 was non-zero for cause 1, but zero for cause 2. Groups 2 and 3 were non-zero for causes 1 and 2. Group 4 was zero for cause 1, but non-zero for cause 2. To examine Condition (C5) $d_m^A/m \rightarrow 0$, we added one more group with $(\beta_{25}, \beta_{26}, \beta_{27})^T = (0, 0, 0)^T$ and $(\alpha_{25}, \alpha_{26}, \alpha_{27})^T = (0, 0, 0)^T$ for $m = 100$ and two more groups with $(\beta_{25}, \dots, \beta_{30})^T = (0, \dots, 0)^T$ and $(\alpha_{25}, \dots, \alpha_{30})^T = (0, \dots, 0)^T$ for $m = 200$. These allow that d_m grows as m increases. The number of the true non-zero groups and non-zero individual variables of the underlying model were 3 and 6, respectively, for each setting. For the BIC-based method and the adaptive lasso, variables were selected individually. Table 1 summarizes the simulation results. “AGB-GB” and “AGB-FG” indicate the adaptive group bridge with $\tilde{\beta}$ from the group bridge and the marginal Fine-Gray model, respectively. AGLasso, ALasso, and BIC represent the adaptive group lasso, the adaptive lasso, and the BIC-based variable selection method, respectively. “% Corr. Group” and “% Corr. Individual” represent the proportions that the corresponding variable selection method correctly identified the non-zero groups and non-zero individual variables of the underlying model, respectively. Thus, “% Corr. Individual” indicates the proportion that the variable selection method correctly identified the underlying model, that is, all non-zero individual variables and all zero individual variables. “Group Size” and “Model Size” indicate the mean number of groups and individual variables selected by each variable selection method. The numbers in parentheses for “Group Size” and “Model Size” are the number of true non-zero groups and non-zero individual variables. “MSER” is the ratio of the MSE of each variable selection method to the MSE of the oracle estimator, where the oracle estimator was obtained from the marginal Fine-Gray model with the true non-zero β_0 's. Thus, MSER values closer to 1 mean better estimation of β_0 .

Table 1 summarizes the results. The adaptive group bridge methods with $\tilde{\beta}$ from the group bridge and the marginal Fine-Gray model performed very similarly. The adaptive group bridge, the group bridge, and the adaptive group lasso identified true non-zero and zero groups very well and performed better than the adaptive lasso and the BIC-based method in group variable selection. While the group bridge, the adaptive lasso, and the BIC performed poorly in within-group variable selection, the adaptive group bridge identified non-zero and zero individual variables very well in both independent and clustered data. The adaptive group lasso was not able to select non-zero within-group variables as expected. The mean group sizes of the adaptive group bridge, the group bridge, and the adaptive group lasso were very close to 3. On the other hand, the adaptive lasso and the BIC-based method over-selected group variables. As m increased, the mean group sizes and the mean model sizes of the adaptive group bridge became closer to 3 and 6, respectively. The adaptive group bridge also selected a smaller set of individual variables compared to the other methods. The MSERs of the adaptive group bridge and the BIC-based method were similar to each other

and were lower than the other methods. In addition, the MSERs of the adaptive group bridge got smaller as m increased.

3.2 Simulation 2: group variables consisting of continuous variables and categorical variables

In this section, we performed a simulation study on selecting group variables consisting of continuous variables and categorical variables. For $m = 50$, there were 8 groups: 4 groups consisting of continuous variables (Groups 1 to 4) and 4 groups consisting of categorical variables (Groups 5 to 8). Groups 1 and 3 contained 3 continuous variables each and Groups 2 and 4 were comprised of 6 continuous variables each. The pairwise correlation among continuous variables within group was 0.5 as in Section 3.1. There was no correlation between continuous variables if they belonged to different groups. Groups 5 and 6 consisted of 4 categories each (that is, 3 indicator variables each) and Groups 7 and 8 (that is, 6 indicator variables each). The reference group for each categorical variable was set to 0. Thus, there were 36 individual variables in total. The true β and α were

$$\begin{aligned} (\beta_1, \dots, \beta_9)^T &= (0.7, -0.7, 0, 0.8, -0.8, 0, \dots, 0)^T, & (\beta_{10}, \dots, \beta_{18})^T &= (0, \dots, 0)^T, \\ (\beta_{19}, \dots, \beta_{24})^T &= (1, -0.9, 0, 0, 0, 0)^T, & (\beta_{22}, \dots, \beta_{36})^T &= (0, \dots, 0)^T, \\ (\alpha_1, \dots, \alpha_9)^T &= (0, 0, 0 - 0.9, 0.9, 0, 0.8, 0, -0.8)^T, & (\alpha_{10}, \dots, \alpha_{18})^T &= (0, \dots, 0)^T, \\ (\alpha_{19}, \dots, \alpha_{24})^T &= (0, 0, 0, -0.8, 0.9, 0)^T, & (\beta_{22}, \dots, \beta_{36})^T &= (0, \dots, 0)^T. \end{aligned}$$

Thus, Groups 1 and 5 were non-zero for cause 1, but zero for cause 2. Group 2 was non-zero for causes 1 and 2. Groups 2 and 6 were zero for cause 1, but non-zero for cause 2. To examine Condition (C5) $d_m^4/m \rightarrow 0$, we added one more group consisting of continuous variables with $(\beta_{37}, \beta_{38}, \beta_{39})^T = (a_{37}, a_{38}, a_{39})^T = (0, 0, 0)^T$ for $m = 100$. For $m = 200$ we added one more group consisting of continuous variables with $(\beta_{37}, \beta_{38}, \beta_{39})^T = (a_{37}, a_{38}, a_{39})^T = (0, 0, 0)^T$ and one more categorical variable with $(\beta_{40}, \beta_{41}, \beta_{42})^T = (a_{40}, a_{41}, a_{42})^T = (0, 0, 0)^T$ for $m = 200$. Thus, the number of non-zero groups and non-zero individual variables of the underlying model were 3 and 6, respectively, for each setting. For the BIC-based method, continuous variables were selected individually and categorical variables were selected as a group. The adaptive lasso selected variables individually.

Table 2 provides the simulation results. The adaptive group bridge methods with $\tilde{\beta}$ from the marginal Fine-Gray model identified non-zero variables slightly better than that with $\tilde{\beta}$ from the group bridge. The adaptive group bridge, the group bridge, and the adaptive group lasso selected true nonzero groups very well and performed better than the adaptive lasso and the BIC-based methods in group variable selection. The adaptive lasso performed poorly in variable selection at both levels. The adaptive group bridge identified true non-zero and zero individual variables very well in both independent and clustered data. On the other hand, the other methods over-selected individual variables. As m increased, the mean group sizes and the mean model sizes of the adaptive group bridge became closer to 3 and 6, respectively, as in Section 3.1. The MSERs of the adaptive group bridge and the BIC-based methods were

comparable and lower than the other methods. Furthermore, the MSERs of the adaptive group bridge got smaller as m increased.

3.3 Evaluation of the sandwich formula

We examined the accuracy of the proposed sandwich formula (4) based on the setting of Section 3.1 with $\tilde{\beta}$ from the marginal Fine-Gray model. Table 3 shows the summary statistics. “SW” indicates the mean estimated standard deviations from the sandwich formula (4). “SD” represents the empirical standard deviation. Although the proposed sandwich formula appeared to slightly underestimate the true standard deviations of $\hat{\beta}$, it estimated the standard deviations of $\tilde{\beta}$ reasonably well. As m increased, the accuracy of the sandwich formula was improved. We also examined (4) using $\tilde{\beta}$ from the group bridge and obtained very similar results to Table 3. Thus, we did not present the results in the paper.

4 Bone marrow transplant data

We applied the adaptive group bridge to bone marrow transplant data in this section. The data were collected by the Center for International Blood and Marrow Transplant Research (Shaw et al, 2010). It consisted of pediatric patients, which were less than or equal to 18 years old, undergoing allogeneic T-replete, myeloablative bone marrow transplantation between 1993 and 2006. The data contained 1591 patients from 108 transplant centers. The outcome of interest was chronic graft-versus-host disease (CGVHD) and death was a competing risk. The rates of CGVHD, death, and censoring were 21%, 36%, and 43%, respectively. The random effect score test of Commenges and Andersen (1995) was used to test center effects in CGVHD based on the cause-specific hazards model and its censoring times based on the proportional hazards models. Both p-values were less than 0.01.

There were 12 variables: 3 continuous variables and 9 categorical variables. The 3 continuous variables include year of transplant, time from diagnosis to transplant, and patient age. The 9 categorical variables are disease type, karnofsky performance status, total body irradiation, antithymocyte globulin (ATG), sex donor/recipient match, cytomegalovirus donor/recipient match, disease stage, donor type, and GVHD prophylaxis. Each of the continuous variables formed Groups 1 to 3. The 9 categorical variables formed from Group 4 to Group 12. More details on these variables can be found in online Supplementary Materials.

To select significant variables, we used the adaptive group bridge and compared it to the group bridge, the adaptive group lasso, and the BIC-based method. The adaptive lasso was not used because of its poor performance in the simulation studies. The adaptive group bridge with $\tilde{\beta}$ from the group bridge and the marginal Fine-Gray model also selected the same variables. Thus, we report the results for the adaptive group bridge with $\tilde{\beta}$ from the marginal Fine-Gray model, the group bridge, the adaptive group lasso, and the BIC-based method only. Table 4 shows the list of variables selected by each method and their parameter estimates with standard errors. “AGB”, “GB”, “AGLasso” and “BIC” represent the adaptive group bridge with $\tilde{\beta}$ from the marginal Fine-Gray model, the group bridge, the adaptive group lasso, and the BIC-based methods, respectively. “MFG” shows the estimates and their p-values from the marginal Fine-Gray model fitting the non-zero group variables selected by

at least one of the three methods. ‘ P -value’ and ‘Overall p -value’ are p -values from testing whether individual variables and group variables were equal to zero, respectively. ‘Ref’ and ‘Est.’ stand for the reference group and parameter estimates, respectively. The numbers in parentheses are the estimated standard errors of the non-zero estimates. We set the coefficient of the reference group to zero in estimating the parameters of the categorical variables.

First of all, the adaptive group bridge and group bridge selected 5 group variables of which overall p -values were all less than 0.05 in the marginal Fine-Gray model. However, the adaptive group lasso and the BIC-based method did not select sex match, which is why all of its parameter estimates are zeros in Table 4. The estimates of the group bridge for Female/Male in Sex match and 8/8 matched URD in Donor type were non-zero; however, their p -values were greater than 0.05 in the marginal Fine-Gray model. In addition, none of the within-group variables from the adaptive group lasso were zeros as expected. On the other hand, all non-zero coefficients of the adaptive group bridge were significant at the significance level 0.05 in the marginal Fine-Gray model. The variables identified by the adaptive group bridge match existing literature. The ATG use has been effective to prevent CGVHD (Kroger et al, 2016). HLA-identical siblings are known to be superior to the other donors (Shaw et al, 2010). In addition, the combination of CsA and MTX after transplantation is considered as the gold standard for GVHD prophylaxis (Rubio et al, 2015). For the donor/recipient sex match, male recipients with female donors were worse than the other groups, which is consistent to a recent report of Kim et al (2016).

5 Conclusion

We have proposed the adaptive group bridge to select variables for the Fine-Gray model with independent and clustered competing risks data. The proposed method not only consistently selects group variables, but also consistently selects within-group variables. The proposed method performed better than the group bridge, the adaptive lasso, the adaptive group lasso, and AIC/BIC-based methods in the simulation study. In particular, the adaptive group bridge identified true non-zero and zero individual variables better than the group bridge.

The proposed method was limited to when the number of parameters are less than the number of observations. Developing a group variable selection method when $d_m > n$ is an interesting future research problem. A two-step variable selection procedure may be developed for this: once we screen group variables at the first step, we may use the adaptive group bridge to obtain a further parsimonious list of non-zero variables at the second step. Although the adaptive group bridge may allow censoring to depend on some variables, we have not studied its asymptotics. Studying the adaptive group bridge with variable-dependent censoring would be another important problem to explore in the future.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The US National Cancer Institute (U24CA076518) partially supported this work. The authors would like to thank the Associate Editor and two reviewers for their helpful comments that significantly improved the paper.

References

- Cai J, Fan J, Li R, Zhou H. Variable selection for multivariate failure time data. *Biometrika*. 2005; 92:303–316. [PubMed: 19458784]
- Commenges D, Andersen PK. Score test of homogeneity for survival data. *Lifetime Data Analysis*. 1995; 1:145–156. [PubMed: 9385097]
- Fan J, Li R. Variable selection for Cox's proportional hazards model and frailty properties. *Journal of the American Statistical Association*. 2002; 30:74–99.
- 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.
- Fu, Z. *crrp*: Penalized Variable Selection in Competing Risks Regression. 2015. URL <http://CRAN.R-project.org/package=crrp>, r package version 1.0
- Fu Z, Parikh CR, Zhou B. Penalized variable selection in competing risks regression. *Lifetime Data Analysis*. 2016a; doi: 10.1007/s10,985-016-9362-3
- Fu Z, Ma S, Lin H, Parikh CR, Zhou B. Penalized variable selection for multi-center competing risks data. *Statistics in Biosciences*. 2016b; doi: 10.1007/s12,561-016-9181-9
- Gray RJ. A class of K-sample tests for comparing the cumulative incidence of a competing risk. *The Annals of Statistics*. 1988; 16:1141–1154.
- Ha ID, Lee M, Oh S, Jeong JH, Sylvester R, Lee Y. Variable selection in subdistribution hazard frailty models with competing risks data. *Statistics in Medicine*. 2014; 33:4590–604. [PubMed: 25042872]
- Huang J, Ma S, Xie H, Zhang CH. A group bridge approach for variable selection. *Biometrika*. 2009; 96:339–355. [PubMed: 20037673]
- Huang J, Li L, Liu Y, Zhao X. Group selection in the Cox model with a diverging number of covariates. *Statistica Sinica*. 2014; 24:1787–1810.
- Kim HT, Zhang MJ, Woolfrey AE, Martin AS, Chen J, Saber W, Perales MA, Armand P, Eapen M. Donor and recipient sex in allogeneic stem cell transplantation: what really matters. *Haematologica*. 2016; 101:1260–1266. [PubMed: 27354023]
- Kroger N, Solano C, Wolschke C, et al. Antilymphocyte globulin for prevention of chronic graft-versus-host disease. *The New England Journal of Medicine*. 2016; 374:43–53. [PubMed: 26735993]
- Kuk D, Varadhan R. Model selection in competing risks regression. *Statistics in Medicine*. 2013; 32:3077–3088. [PubMed: 23436643]
- Logan B, Zhang MJ, Klein JP. Marginal models for clustered time to event data with competing risks using pseudovalue. *Biometrics*. 2011; 67:1–7. [PubMed: 20377579]
- Prentice RL, Kalbfleisch JD, Peterson AV, Flournoy N, Farewell VT, Breslow NE. The analysis of failure times in the presence of competing risks. *Biometrics*. 1978; 34:541–554. [PubMed: 373811]
- Rubio MT, Labopin M, Blaise D, et al. The impact of graft-versus-host disease prophylaxis in reduced-intensity conditioning allogeneic stem cell transplant in acute myeloid leukemia: a study from the acute leukemia working party of the european group for blood and marrow transplantation. *Haematologica*. 2015; 100:683–689. [PubMed: 25769546]
- Seetharaman, I. Master's thesis. Kansas State University; KS, USA: 2013. Consistent bi-level variable selection via composite group bridge penalized regression.
- Shaw PJ, Kan F, Ahn KW, Spellman SR, Aljurf M, Ayas M, et al. Outcomes of pediatric bone marrow transplantation for leukemia and myelodysplasia using matched sibling, mismatched related, or matched unrelated donors. *Blood*. 2010; 116:4007–4015. [PubMed: 20671124]

- Varadhan, R., Kuk, D. *crrstep*: Stepwise Covariate Selection for the Fine and Gray Competing Risks Regression Model. 2015. URL <http://CRAN.R-project.org/package=crrstep>, r package version 2015-2.1
- Wang HJ, Zhou J, Li Y. Variable selection for censored quantile regression. *Statistica Sinica*. 2013; 23:145–167. [PubMed: 23847407]
- Wu TT, Wang S. Doubly regularized Cox regression for high-dimensional survival data with group structures. *Statistics and Its Interface*. 2013; 6:175–186.
- Yuan M, Lin Y. Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B*. 2006; 68:49–67.
- Zhou B, Fine J, Latouche A, Labopin M. Competing risks regression for clustered data. *Biostatistics*. 2012; 13:371–383. [PubMed: 22045910]
- Zou H. The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*. 2006; 101:1418–1429.

Appendix

To describe the conditions to justify the asymptotics, we define

$$w_{ij}(t) = \frac{I(C_{ij} \geq T_{ij} \wedge t)G(t)}{G(X_{ij} \wedge t)}$$

$$\mathbf{q}(u) = - \lim_{m \rightarrow \infty} \frac{1}{m} \sum_{i=1}^m \sum_{j=1}^{n_i} \int_0^\tau \{Z_{ij}(t) - \mathbf{e}(\beta_0, t)\} w_{ij}(t) I(X_{ij} < u \leq t) dM_{ij}(\beta_0, t),$$

$$\pi(u) = \lim_{m \rightarrow \infty} \frac{1}{m} \sum_{i=1}^m \sum_{j=1}^{n_i} I(X_{ij} \geq u).$$

where $\mathbf{a}^{\otimes 0} = \mathbf{1}$, $\mathbf{a}^{\otimes 1} = \mathbf{a}$, and $\mathbf{a}^{\otimes 2} = \mathbf{a}\mathbf{a}^T$. Let $N_{ij}^c(t) = I(C_{ij} \leq t)$ and $\Lambda^c(t)$ be the cumulative hazard function by treating censored observations as events. Define $M_{ij}^c(t) = N_{ij}^c(t) - \int_0^t I(X_{ij} \geq u) d\Lambda^c(u)$. Consider the following conditions:

- (C1) $\int_0^\tau \lambda_{10}(u) du < \infty$ and $|Z_{ijq}(0)| + \int_0^\tau |dZ_{ijq}(u)| \leq M$ for all i, j, q , where M is a positive constant;
- (C2) There exists a neighborhood $\mathcal{B}$ of β_0 and $\sup_{t \in [0, \tau], \beta \in \mathcal{B}} \|\mathbf{S}^{(r)}(\beta, \theta) - \mathbf{s}^{(r)}(\beta, \theta)\|_2$ converges in probability to 0 for $r = 0, 1, 2$. There also exists a matrix $\Gamma = \Gamma(\beta_0)$ such that $\|m^{-1} \sum_{i=1}^m \text{Var}(\mathbf{D}_i) - \Gamma\| \rightarrow 0$, where

$$\mathbf{D}_i = \sum_{j=1}^{n_i} \left[\int_0^\tau \{Z_{ij}(u) - \mathbf{e}(\beta_0, u)\} w_{ij}(u) dM_{ij}(\beta_0, u) + \int_0^\tau \sum_{i=1}^m \mathbf{q}(u) / \pi(u) dM_{ij}^c(u) \right].$$

In addition, there exist constants C_1 and C_2 such that $0 < C_1 < \kappa_{\min}(\Gamma)$ $\kappa_{\max}(\Gamma) < C_2 < \infty$ for all m , where $\kappa_{\min}(\mathbf{A})$ and $\kappa_{\max}(\mathbf{A})$ are the minimal and maximal eigenvalues of a matrix $\mathbf{A}$, respectively;

- (C3) For $r = 0, 1, 2$, $\mathbf{s}^{(r)}(\beta, t)$'s are continuous in $\beta \in \mathcal{B}$ uniformly in $t \in [0, \tau]$ and are bounded on $\mathcal{B} \times [0, \tau]$. $\mathbf{s}^{(0)}(\beta, t)$ is bounded away from 0 on $\mathcal{B} \times [0, \tau]$. The true

parameter β_{j0} is bounded away from 0 for $j \in B_1$. Let

$\Omega = \int_0^\tau v(\beta_0, u) s^{(0)}(\beta_0, u) \lambda_{10}(u) du$. There exists C_3 and C_4 such that $0 < C_3 < \kappa_{\min}(\Omega) \leq \kappa_{\max}(\Omega) < C_4 < \infty$;

(C4) There exists a constant C_5 such that $\sup_{1 \leq i \leq m} E(D_{ij}^2 D_{ij'}^2) \leq C_5 < \infty$ for all $1 \leq j, j' \leq d_m$. $C_m^* = \max_j \sum_{k=1}^K I(j \in A_k)$ is bounded;

(C5) $d_m^A/m \rightarrow 0$;

(C6) $\sum_{k=1}^{K_1} c_k \left\{ \left(\sum_{j \in A_k \cap B_1} |\beta_{j0}|^{1-\nu} \right)^{\gamma-1} \sum_{j \in A_k \cap B_1} |\beta_{j0}|^\nu \right\} \leq M_m$, where $M_m = O_p(1)$;

(C7) $\lambda_m/\sqrt{m} \rightarrow 0$, $\sqrt{m/d_m} \tilde{\beta}_j = O_p(1)$, and $\min(\lambda_m m^{(\nu-1)/2} d_m^{-(1+\nu)/2}, \lambda_m m^{\gamma(\nu-1)/2} d_m^{-1+\gamma(1-\nu)/2}) \rightarrow \infty$.

Conditions (C1)–(C3) are standard conditions for the marginal subdistribution hazards model (Zhou et al, 2012). Conditions (C1)–(C5) are similar to the conditions of Cai et al (2005) and Huang et al (2014) so that they guarantee local asymptotic quadratic property of $\mathcal{I}(\beta)$ and the existence of local minimizer of $\mathcal{L}_m(\beta)$. Once Conditions (C3) and (C4) are met, Condition (C6) is satisfied if the true non-zero β_{j0} 's are bounded above. Condition (C7) controls λ_m to have the oracle property.

Table 1

Simulation results for group variables consisting of continuous variables.

m	ψ	Method	% Corr. Group	% Corr. Individual	Group Size (3)	Model Size (6)	MSER
50	0.25	AGB-GB	0.75	0.61	3.18	6.33	1.52
		AGB-FG	0.73	0.59	3.23	6.43	1.53
		GB	0.75	0.27	3.13	7.00	1.67
		AGL _{lasso}	0.70	0.00	3.34	10.03	1.83
		AL _{lasso}	0.34	0.26	4.01	7.19	1.86
		BIC	0.69	0.62	3.40	6.51	1.78
100	0.25	AGB-GB	0.86	0.77	3.07	6.15	1.58
		AGB-FG	0.85	0.72	3.09	6.23	1.57
		GB	0.86	0.40	3.04	6.69	2.26
		AGL _{lasso}	0.88	0.00	3.10	9.30	2.57
		AL _{lasso}	0.52	0.42	3.61	6.71	2.51
		BIC	0.68	0.62	3.38	6.52	1.71
200	0.25	AGB-GB	0.90	0.82	3.11	6.21	1.23
		AGB-FG	0.89	0.79	3.12	6.25	1.25
		GB	0.89	0.30	3.11	7.11	1.65
		AGL _{lasso}	0.75	0.00	3.28	9.83	1.71
		AL _{lasso}	0.56	0.52	3.60	6.71	1.70
		BIC	0.78	0.73	3.27	6.37	1.37
	1	AGB-GB	0.96	0.92	3.04	6.08	1.25
		AGB-FG	0.96	0.91	3.04	6.09	1.26
		GB	0.97	0.47	3.03	6.74	2.07
		AGL _{lasso}	0.91	0.00	3.10	9.29	1.93
		AL _{lasso}	0.72	0.70	3.32	6.36	1.93
		BIC	0.77	0.74	3.28	6.35	1.36
200	0.25	AGB-GB	0.97	0.95	3.03	6.05	1.18
		AGB-FG	0.97	0.93	3.03	6.07	1.18
		GB	0.95	0.40	3.05	6.90	1.80
		AGL _{lasso}	0.81	0.00	3.20	9.61	1.60

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

m	ψ	Method	% Corr. Group	% Corr. Individual	Group Size (3)	Model Size (6)	MSER
1		ALasso	0.73	0.70	3.33	6.37	1.56
		BIC	0.78	0.77	3.24	6.29	1.26
		AGB-GB	0.98	0.96	3.02	6.04	1.26
		AGB-FG	0.99	0.95	3.01	6.05	1.26
		GB	0.99	0.45	3.01	6.70	2.48
		AGLasso	0.95	0.00	3.05	9.14	1.90
		ALasso	0.87	0.84	3.14	6.17	1.81
		BIC	0.82	0.78	3.19	6.25	1.30

Table 2

Simulation results for groups with continuous and categorical variables.

m	ψ	Method	% Corr. Group	% Corr. Individual	Group Size (3)	Model Size (6)	MSER
50	0.25	AGB-GB	0.80	0.52	2.88	5.74	2.61
		AGB-FG	0.74	0.48	2.79	5.34	3.16
		GB	0.72	0.28	2.65	5.91	3.37
		AGLasso	0.80	0.00	2.76	11.03	4.78
		AL	0.01	0.00	5.34	10.30	4.55
		BIC	0.66	0.00	3.11	7.06	2.93
100	0.25	AGB-GB	0.53	0.18	3.02	6.00	3.23
		AGB-FG	0.47	0.18	2.75	5.11	3.75
		GB	0.54	0.11	2.72	6.25	3.14
		AGLasso	0.77	0.00	2.94	11.80	3.07
		AL	0.00	0.00	5.84	12.38	5.69
		BIC	0.52	0.00	3.01	6.66	3.39
200	0.25	AGB-GB	0.92	0.69	3.08	6.29	1.78
		AGB-FG	0.94	0.77	3.02	6.02	2.33
		GB	0.98	0.35	3.02	7.03	2.11
		AGLasso	0.96	0.00	3.03	12.11	2.80
		AL	0.00	0.00	5.36	10.42	3.82
		BIC	0.81	0.00	3.19	7.30	1.71
	1	AGB-GB	0.88	0.58	3.12	6.45	2.21
		AGB-FG	0.88	0.67	3.06	6.04	2.97
		GB	0.93	0.28	3.04	7.25	2.57
		AGLasso	0.94	0.00	3.05	12.16	3.18
		AL	0.00	0.00	5.41	10.82	4.75
		BIC	0.79	0.00	3.19	7.30	1.92
200	0.25	AGB-GB	0.99	0.89	3.01	6.12	1.70
		AGB-FG	1.00	0.96	3.00	6.02	1.89
		GB	1.00	0.42	3.00	6.89	2.80
		AGLasso	0.99	0.00	3.01	12.04	2.83

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

m	ψ	Method	% Corr. Group	% Corr. Individual	Group Size (3)	Model Size (6)	MSER
1		AL	0.02	0.02	4.99	8.82	3.62
		BIC	0.87	0.00	3.14	7.19	1.67
		AGB-GB	0.96	0.81	3.04	6.22	1.77
		AGB-FG	0.98	0.91	3.02	6.07	2.09
		GB	0.99	0.35	3.01	7.07	2.52
		AGLasso	0.94	0.00	3.06	12.18	2.63
		AL	0.01	0.00	5.34	9.60	3.81
		BIC	0.88	0.00	3.12	7.19	1.64

Table 3

Simulation results for the accuracy of the sandwich formula (4).

m	ψ	Method	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	$\hat{\beta}_7$
50	0.25	SW	0.179	0.179	0.188	0.165	0.164	0.160
		SD	0.221	0.231	0.253	0.205	0.201	0.205
		SW	0.143	0.143	0.145	0.133	0.133	0.121
100	0.25	SD	0.170	0.182	0.174	0.159	0.156	0.156
		SW	0.127	0.127	0.132	0.116	0.116	0.112
		SD	0.149	0.148	0.149	0.138	0.141	0.130
200	0.25	SW	0.101	0.101	0.103	0.094	0.094	0.085
		SD	0.112	0.118	0.121	0.109	0.115	0.102
		SW	0.089	0.090	0.093	0.083	0.082	0.079
400	0.25	SD	0.093	0.094	0.105	0.088	0.091	0.090
		SW	0.071	0.071	0.073	0.067	0.067	0.060
		SD	0.072	0.078	0.077	0.072	0.075	0.068

SW indicates the mean estimated standard deviations from the sandwich formula (4). SD represents the empirical standard deviation.

Table 4

List of selected variables.

Variables	Description	AGB		GB		AGLasso		BIC		MFG	
		Est.		Est.		Est.		Est.		Est.	
Patient age	Continuous	0.063 (0.011)		0.061 (0.011)		0.048 (0.011)		0.062 (0.012)		0.060 (0.011)	
ATG	No (Ref)	0		0		0		0		0	
	Yes	-0.563 (0.208)		-0.473 (0.203)		-0.294 (0.203)		-0.827 (0.224)		-0.824 (0.231)	
Sex match	Male/Male (Ref)	0		0		0		0		0	
	Female/Male	0.310 (0.118)		0.416 (0.130)		0		0		0.447 (0.175)	
	Male/Female	0		0.227 (0.145)		0		0		0.277 (0.168)	
	Female/Female	-		0		0		0		-0.051 (0.153)	
Donor type	1-antigen mismatched related donors (Ref)	0		0		0		0		0	
	Phenotypic matched related donors	0		0		-0.183 (0.328)		-0.270 (0.294)		-0.324 (0.291)	
	HLA-identical siblings	-0.983 (0.152)		-0.858 (0.146)		-0.723 (0.193)		-1.053 (0.167)		-1.056 (0.158)	
	8/8 matched unrelated donors	0		0.152 (0.185)		0.022 (0.214)		-0.013 (0.190)		0.068 (0.191)	
GVHD prophylaxis	CSA/FK506+/-others (Ref)	0		0		0		0		0	
	MTX+/-CSA/FK506	-0.404 (0.190)		-0.370 (0.189)		-0.280 (0.201)		-0.498 (0.178)		-0.528 (0.173)	

Sex match shows donors' gender and patients' gender. For example, 'Male/Female' indicates that a donor was a male and the corresponding patient was a female. The numbers in parentheses are the estimated standard errors of the non-zero estimates. '-' indicates 'Not Applicable'.