Published in final edited form as:

Stat Med. 2018 April 30; 37(9): 1577–1586. doi:10.1002/sim.7619.

Variable selection with group structure in competing risks quantile regression

Kwang Woo Ahn[a] and **Soyoung Kim**[a,*]

[a]Division of Biostatistics, Medical College of Wisconsin, Milwaukee, Wisconsin, U.S.A

Abstract

We study the group bridge and the adaptive group bridge penalties for competing risks quantile regression with group variables. While the group bridge consistently identifies non-zero group variables, the adaptive group bridge consistently selects variables not only at group level, but also at within-group level. We allow the number of covariates to diverge as the sample size increases. The oracle property for both methods is also studied. The performance of the group bridge and the adaptive group bridge is compared in simulation and in a real data analysis. The simulation study shows that the adaptive group bridge selects non-zero within-group variables more consistently than the group bridge. A bone marrow transplant study is provided as an example.

Keywords

Adaptive lasso; Competing risks quantile regression; Group bridge

1. Introduction

Quantile regression provides an alternative method to the Cox proportional hazards model and the accelerated failure time (AFT) model in survival analysis [1]. It is often preferred when the survival distribution is skewed. There is rich literature in survival quantile regression. Peng and Huang [2] proposed a martingale-based estimating equations. Reich and Smith [3] developed a semiparametric Bayesian quantile regression model for censored data. Yin et al. [4] studied a power-transformed quantile regression model for survival data. Yin and Cai [5] proposed quantile regression models for correlated survival data.

Recently quantile regression for competing risks data have had much attention. Peng and Fine [1] proposed a semiparametric model based on the competing risks AFT model. Sun et al. [6] developed a regression model when the failure type is missing in competing risks data. Lee and Fine [7] studied parametric and nonparametric methods to make inference on cumulative incidence quantiles.

In spite of increasing popularity of quantile regression for survival and competing risks data, the current literature on variable selection is somewhat limited. Jiang et al. [8] proposed the

*Correspondence to: skim@mcw.edu.

Supplementary Material

Additional supplementary material may be found in the online version of this article at the publishers web site.

adaptive lasso for a composite quantile regression with randomly censored data. Wang et al. [9] also studied the adaptive lasso for censored quantile regression. They all studied a survival setting, not a competing risks setting. In addition, their proposed methods addressed variable selection at individual level, not at group level. In practice, clinicians often encounter group variables such as categorical variables. For example, Verneris et al. [10] studied the outcomes of the patients having reduced-intensity conditioning allogeneic hematopoietic cell transplantation from 1999 to 2011. They studied competing risks outcomes including relapse and treatment-related mortality (TRM), where relapse and TRM are competing risks to each other. The variables that they considered for analysis consisted of binary and categorical variables.

Several penalties have been proposed to select group variables for linear regression and competing risks settings. Yuan and Lin [11] proposed the group lasso, which selects variables at group level, not at within-group level. Huang et al. [12] developed the group bridge to select both non-zero group and non-zero within-group variables. However, they studied group selection consistency only and did not show within-group variable selection consistency. Zhou and Zhu [13] proposed an adaptive hierarchical lasso having group variable selection consistency and within-group variable selection consistency. Zhao et al. [14] applied the adaptive hierarchical lasso penalty to identify non-zero variables at both levels for quantile linear regression. Fu et al. [15] extensively studied lasso, adaptive lasso, SCAD, and MCP for individual variable selection and their group variable selection versions for the subdistribution hazards model. However, they did not address within-group variable selection. In addition, their oracle property was limited to a fixed number of covariates. Despite extensive work in group variable selection for linear, linear quantile, and subdistribution hazards regression models, there is little literature on group variable selection in competing risks quantile regression. In particular, group and within-group level variable selection techniques remain unexplored in the current literature to the best of the authors' knowledge.

We propose the group bridge and the adaptive group bridge for bi-level variable selection, that is, group and within-group variable selection, under the competing risks quantile regression model of Peng and Fine [1]. While the group bridge consistently identifies non-zero group variables, the adaptive group bridge consistently selects non-zero variables at both group level and within-group level. When there is no group structure for variables, individual variable selection can be handled as a special case of the proposed methods. Based on our knowledge, even individual variable selection has not been studied for the competing risks quantile regression. We study their oracle property while allowing the number of variables to diverge as the sample size increases. We show the adaptive group bridge identifies non-zero within-group variables more consistently than the group bridge in simulation study. In Section 2, we describe the proposed methods and study their theoretical properties. In Section 3, we compare the performance of the adaptive group bridge and the group bridge via simulation study. We illustrate a real data example in Section 4 and have a brief conclusion in Section 5. All the proofs of the theorems and the lemmas in this paper can be found in the online Supplementary Materials.

2. Method

In this section, we propose a penalized competing risks quantile regression model and study its theoretical properties. We begin with some notations. Without loss of generality, we consider two causes of failure $\varepsilon \in \{1,2\}$ with sample size n. We allow the number of covariates d_n to increase as n increases. Let T_i, C_i, ε_i, and $\mathbf{Z}_i = (1, Z_{i1}, \ldots, Z_{id_n})^T$ be the event time, censoring time, cause of failure, and covariate vector of subject i for $i = 1, \ldots, n$. Denote $\boldsymbol{\beta}_0(\tau) = \{\beta_{j,0}(\tau);\ j = 0, \ldots, d_n\}^T$ as the true parameter vector given quantile τ, where $\beta_{0,0}$ is the true intercept coefficient. Let $X_i = T_i \wedge C_i$ be the observed time and $\delta_i = I(T_i \le C_i)I(\varepsilon_i = 1)$, where $a \wedge b = \min(a, b)$. We assume that $(T_i, \varepsilon_i, C_i, \mathbf{Z_i})$ are independent and identically distributed, and the T_i's and C_i's are independent given $\mathbf{Z_i}$ for $i = 1, \ldots, n$. The study period is $[0, L]$. Let $F_1(t|\mathbf{Z_i})$ be the cumulative incidence of cause 1 at time t given $\mathbf{Z}_i$, where $F_1(t|\mathbf{Z_i}) = P(T_i \le t, \varepsilon_i = 1|\mathbf{Z}_i)$. Given covariate $\mathbf{Z}$, we define the τth conditional quantile of $F_1(t|\mathbf{Z})$ as $Q_1(\tau|\mathbf{Z}) = \inf\{t: F_1(t|\mathbf{Z}) \ge \tau\}$. For $\tau\ \varepsilon\ [\tau_L, \tau_U]$ with $0 < \tau_L, \tau_U < 1$, we consider $Q_1(\tau|\mathbf{Z}) = g\{\mathbf{Z}^{\mathrm{T}}\boldsymbol{\beta}(\tau)\}$, where $g(\cdot)$ is a known monotone link function. Let $\|\cdot\|$ be the Euclidean norm and $\mathbf{a}^{\otimes 2} = \mathbf{a}\mathbf{a}^T$ for a vector $\mathbf{a}$.

Let $\tilde{\mathbf{Z}}_i = (Z_{i1}, \ldots, Z_{id_n})^T$. For simplicity, we assume that $\tilde{\mathbf{Z}}_i$'s are fixed over time. Let $N_i^G(t) = I(C_i \le T_i)I(C_i \le t)$ be the counting process for censoring and $Y_i(t) = I(X_i \ge t)$. We use the Cox proportional hazards model to fit censoring time C_i's:

$$\lambda^G(t|\tilde{\mathbf{Z}}_i) = \lambda_0^G(t)e^{\boldsymbol{\alpha}^T\tilde{\mathbf{Z}}_i},$$

Where $\lambda_0^G(t)$ is an arbitrary baseline hazard function for censoring and $\boldsymbol{\alpha}^T$ is the unknown parameter vector. Define

$$\mathbb{S}_G^{(d)}(\boldsymbol{\alpha}, t) = n^{-1}\sum_{i=1}^{n} Y_i(t)\tilde{\mathbf{Z}}_i^{\otimes d} e^{\boldsymbol{\alpha}^T\tilde{\mathbf{Z}}_i},$$

where $d = 0,1$, and 2. The baseline cumulative hazard function for censoring $\Lambda_0^G(t)$ is estimated by the Breslow-type estimator [16]:

$$\hat{\Lambda}_0^G(t;\hat{\boldsymbol{\alpha}}) = \int_0^t \frac{\sum_{i=1}^n dN_i^G(u)}{n\mathbb{S}_G^{(0)}(\hat{\boldsymbol{\alpha}}, u)},$$

where $\hat{\boldsymbol{\alpha}}$ is the estimator of $\boldsymbol{\alpha}$ based on the Cox proportional hazards model. Then, we estimate $G(t|\tilde{\mathbf{Z}}_i)$ as follows:

$$\widehat{G}(t \mid \widetilde{\mathbf{Z}}_i) = \exp\left\{-\int_0^t e^{\widehat{\boldsymbol{\alpha}}^T \widetilde{\mathbf{Z}}_i} d\widehat{\Lambda}_0^G(u : \widehat{\boldsymbol{\alpha}})\right\}.$$

We can obtain the consistency of $\widehat{\boldsymbol{\alpha}}$, $\widehat{\Lambda}_0^G(t:\widehat{\boldsymbol{\alpha}})$, and $\widetilde{G}(t \mid \widetilde{\mathbf{Z}}_i)$ as follows:

Lemma 2.1

Assume Conditions (a)-(e) as in Appendix. Then, we have $\left\|\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}\right\| = O_p\left(\sqrt{d_n/n}\right)$ $\sup_t |\widehat{\Lambda}_0^G(t:\widehat{\boldsymbol{\alpha}}) - \Lambda_0^G(t)| = O_p\left(\sqrt{d_n/n}\right)$, and $\sup_t |\widehat{G}(t|\widetilde{\mathbf{Z}}) - G|(t|\widetilde{\mathbf{Z}})| = O_p\left(\sqrt{d_n/n}\right)$.

When the censoring distribution G does not depend on any covariates, the Kaplan-Meier estimator can be used instead of the Breslow estimator. The proof of Lemma 2.1 can be found in the online Supplemental Materials.

Next, we define some notations on group variables and their memberships. Assume that we have K groups of variables. Let $A_1,\ldots,A_K$ be subsets of $\{1, \ldots, d_n\}$ representing group memberships of variables, where A_k's may overlap. Define $\boldsymbol{\beta}_A(\boldsymbol{\tau}) = \{\beta_j(\boldsymbol{\tau}), j \in A\}^T$ and $\boldsymbol{\beta}_{A,0}(\boldsymbol{\tau}) = \{\beta_{j,0}(\boldsymbol{\tau}); j \in A\}^T$ for a set A. To distinguish the individual memberships between non-zero $\beta_{j,0}(\boldsymbol{\tau})$'s and zero $\beta_{j,0}(\boldsymbol{\tau})$'s, we define B_1 and B_2 such that $\beta_{j,0}(\boldsymbol{\tau}) \neq 0$ if $j \in B_1$ and $\beta_{j,0}(\boldsymbol{\tau}) = 0$ if $j \in B_2$. To distinguish the group memberships between non-zero $\boldsymbol{\beta}_{A_k,0}(\tau)$'s and zero $\boldsymbol{\beta}_{A_k,0}(\tau)$'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 $\boldsymbol{\beta}_{A_k,0}(\tau) \neq 0$ for $1 \le k \le K_1$ and $\boldsymbol{\beta}_{A_k,0}(\tau) = \mathbf{0}$ for $K_1 + 1 \le k \le K$.

To estimate $\boldsymbol{\beta}(\boldsymbol{\tau})$, Peng and Fine [1] considered the estimating equation $\mathbf{S}_n(\mathbf{b}, \boldsymbol{\tau}) = 0$, where

$$\mathbf{S}_n(\mathbf{b}, \tau) = n^{-1/2} \sum_{i=1}^{n} \mathbf{Z}_i \left[\frac{I\left\{X_i \le g(\mathbf{Z}_i^T \mathbf{b})\right\} I(\delta_i = 1)}{\widehat{G}(X_i \mid \widetilde{\mathbf{Z}}_i)} - \tau\right]. \quad (1)$$

To solve $\mathbf{S}_n(\mathbf{b}, \boldsymbol{\tau}) = \mathbf{0}$, Peng and Fine [1] proposed the following L_1-type convex function:

$$U_n(\mathbf{b}, \tau) = \sum_{i=1}^{n} I(\delta_i = 1) \left|\frac{g^{-1}(X_i) - \mathbf{b}^T \mathbf{Z}_i}{\widehat{G}(X_i \mid \widetilde{\mathbf{Z}}_i)}\right| + \left|M - \mathbf{b}^T \sum_{i=1}^{n} \frac{-\mathbf{Z}_i I(\delta_i = 1)}{\widehat{G}(X_i \mid \widetilde{\mathbf{Z}}_i)}\right| + \left|M - \mathbf{b}^T \sum_{i=1}^{n} 2\mathbf{Z}_i \tau\right|,$$

where M is a very large positive number to bound $|\mathbf{b}^T \sum_{i=1}^{n} -\mathbf{Z}_i I(\delta_i = 1)/\widehat{G}(X_i \mid \widetilde{\mathbf{Z}}_i)|$ and $|\mathbf{b}^T \sum_{i=1}^{n} 2\mathbf{Z}_i \tau|$ for all $\mathbf{b}$'s in the parameter space for $\boldsymbol{\beta}_0(\boldsymbol{\tau})$. They studied the consistency and the asymptotic normality of the estimator of $\boldsymbol{\beta}_0(\boldsymbol{\tau})$ obtained by solving $\mathbf{S}_n(\mathbf{b}, \boldsymbol{\tau}) = 0$ when G is non-covariate dependent and d_n is fixed.

To select variables at bi-level, we propose the following penalized function:

$$W_n(\mathbf{b}, \tau) = U_n(\mathbf{b}, \tau) + \lambda_n \sum_{k=1}^{K} c_k \left(\sum_{j \in A_k} \frac{|b_j|}{|\widetilde{\beta}_j(\tau)|^{\nu}} \right)^{\gamma}, \quad (2)$$

where $\widetilde{\beta}_j(\tau)$ is a consistent estimator of $\boldsymbol{\beta}(\tau)$, $\nu \geq 0$, $\lambda_n > 0$, and $0 < \gamma < 1$. Following Huang et al. [12], we set $c_k \propto |A_k|^{1-\gamma}$, where $|A|$ is the cardinality of A. If $\nu = 0$, the penalty term is the group bridge penalty of Huang et al. [12] and Huang et al. [17]. When $\nu > 0$, we call the penalty term as adaptive group bridge penalty. The adaptive group bridge becomes i) individual variable selection when $|A_k| = 1$ for all k; and ii) the adaptive hierarchical lasso penalty of Zhou and Zhu [13] when $\gamma = 1/2$ and $c_k = 1$ for all k.

We can formulate minimizing $W_n(\mathbf{b}, \tau)$ to minimizing

$$\widetilde{W}_n(\mathbf{b}, \boldsymbol{\theta}, \tau) = U_n(\mathbf{b}, \tau) + \sum_{k=1}^{K} \theta_k^{1-1/\gamma} c_k^{1/\gamma} \sum_{j \in A_k} \frac{|b_j|^{\gamma}}{|\widetilde{\beta}_j|^{\nu}} + \zeta_n \sum_{k=1}^{K} \theta_k, \quad (3)$$

where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_K)^T$. By defining

$$\theta_k = c_k \left(\frac{1-\gamma}{\zeta_n \gamma} \right)^{\gamma} \left(\sum_{j \in A_k} \frac{|\beta_j(\tau)|}{|\widetilde{\beta}_j(\tau)|^{\nu}} \right)^{\gamma}, \ k = 1, \ldots, K,$$

we can show the following lemma similarly to Proposition 1 of Huang et al. [12] and thus its proof is omitted:

Lemma 2.2

Assume that $\lambda_n = \zeta_n^{1-\gamma} \gamma^{-\gamma} (1-\gamma)^{\gamma-1}$ for $0 < \gamma < 1$. Then, $\hat{\boldsymbol{\beta}}(\tau)$ minimizes $W_n(\mathbf{b}, \boldsymbol{\tau})$ if and only if $\left\{ \hat{\boldsymbol{\beta}}(\tau), \hat{\boldsymbol{\theta}} \right\}$ minimizes $\widetilde{W}_n(\mathbf{b}, \boldsymbol{\theta}, \tau)$, where $\theta_k > 0$ and $\hat{\theta}_k > 0$ for $k = 1, \ldots, K$.

Define $\widetilde{\mathbf{S}}_n(\mathbf{b}, \tau) = n^{-1/2} \sum_{i=1}^{n} \mathbf{Z}_i [F_1 \left\{ g(\mathbf{Z}_i^T \mathbf{b}) | \mathbf{Z}_i \right\} - \tau]$. Denote $\nabla \widetilde{\mathbf{S}}_n(\mathbf{b}, \tau)$ as the first derivative of $\widetilde{\mathbf{S}}_n(\mathbf{b}, \tau)$ with respect to $\mathbf{b}$. We first study the oracle property of the group bridge estimator given $\boldsymbol{\tau}$. We assume that

- **(C1)** There exists $\omega > 0$ such that $P(C = \omega | \widetilde{\mathbf{Z}}) \geq c > 0$ and $P(C > \omega | \widetilde{\mathbf{Z}}) = 0$ for any $\widetilde{\mathbf{Z}}$.
- **(C2)** Z_{ij} and $\beta_{j,0}(\boldsymbol{\tau})$ are uniformly bounded for $j = 1, \ldots, d_n$.
- **(C3)** $f_1(t|\mathbf{z})$ is bounded above uniformly in t and $\mathbf{z}$, where $f_1(t|\mathbf{z}) = dF_1(t|\mathbf{z})/dt$.

(C4) Define $\mathbf{H}(\mathbf{b}) = E\left\{n^{-1/2}\nabla\widetilde{\mathbf{S}}_n(\mathbf{b},\tau)\right\} = E[\mathbf{Z}^{\otimes 2}f_1\left\{g(\mathbf{Z}^T\mathbf{b})|\mathbf{Z}\right\}_g'(\mathbf{Z}^T\mathbf{b})]$. For some $\rho_0 > 0$, $C_1 > 0$, and $C_2 > 0$, $\inf_{\mathbf{b}\varepsilon\mathscr{B}(\rho_0)}\kappa\left\{\mathbf{H}(\mathbf{b})\right\} \geq C_1$ and $\sup_{\mathbf{b}\varepsilon\mathscr{B}(\rho_0)}\kappa\left\{\mathbf{H}(\mathbf{b})\right\} \leq C_2 < \infty$, where $\mathscr{B}(\rho_0) = \left\{\mathbf{b}\varepsilon\mathbb{R}^{d_n+1} : \left\|\mathbf{b} - \boldsymbol{\beta}_0(\tau)\right\| \leq \rho_0\right\}$ and $\kappa(\mathbf{H})$ is the eigenvalue of a matrix $\mathbf{H}$.

(C5) $\Sigma(\tau) = Var\{\mathbf{S}_n(\mathbf{b},\tau)\}$. There exist $C_3 > 0$ and $C_4 > 0$ such that $\inf_{\boldsymbol{\beta}\varepsilon\mathscr{B}(\rho_0)}\kappa\left[\sum(\tau)\}\right] \geq C_3$ and $\sup_{\boldsymbol{\beta}\varepsilon\mathscr{B}(\rho_0)}\kappa\left[\sum(\tau)\}\right] \leq C_4 < \infty$, where $\rho_0 > 0$.

(C6) There exists a constant $C_5 > 0$ such that $\sup_{\mathbf{b}\varepsilon\mathscr{B}(\rho_0), 0 \leq i \leq d_n} n^{-1}Cov\left\{\nabla\widetilde{S}_{n,ij}(\mathbf{b},\tau), \nabla\widetilde{S}_{n,ij'}(\mathbf{b},\tau)\right\} \leq C_5 < \infty$, for all $0 < j, j' < d_n$, where $\nabla\widetilde{S}_{n,ij}(\mathbf{b},\tau)$ is the (i, j)th entry of $\nabla\widetilde{\mathbf{S}}_n(\mathbf{b},\tau)$.

(C7) $d_n^4/n \to 0$.

(C8) $C_n^* = \max_j \sum_{k=1}^K I(j\varepsilon A_k)$ is bounded and $\lambda_n^2/n\sum_{k=1}^{K_1} c_k^2\left\{\sum_{j\varepsilon A_k}|\beta_{j,0}(\tau)|\right\}^{2\gamma-2}|A_k| \leq d_n M_n$, $M_n = O_p(1)$,where $\lambda_n/\left[n^{\gamma/2}\kappa_{\max}\left\{\sum(\tau)\right\}d_n^{1-\gamma/2}\right] \to \infty$ as $n \longrightarrow \infty$

(C9) $\lambda_n n^{-1/2} \to 0$, $1/\kappa_{\min}\left\{\sum(\tau)\right\} + \kappa_{\max}\left\{\sum(\tau)\right\} + \sum_{k=1}^K c_k^2 = O(1)$, $\lambda_n/(n^{\gamma/2}d_n^{1-\gamma/2}) \to \infty$ as $n \longrightarrow \infty$

(C1)−(C5) are similar to the standard conditions for the competing risks quantile regression of Peng and Fine [1]. Peng and Fine [1] suggested to use a truncated censoring time $C = mim(C, L)$ for ω in (C1) so that (C1) is always satisfied. In practice, ω can be chosen as large as possible so that only small information loss occurs [1]. (C4) − (C6) and (C8) control the behavior of the estimating equation as d_n grows. Similar conditions to (C4) − (C8) were used to allow d_n to diverge as $n \longrightarrow \infty$ in Cai et al. [18], Huang et al. [12], and Huang et al. [17]. (C5), (C6), and (C9) restricts the variability of $Var\{\mathbf{S}_n(\mathbf{b},\tau)\}$ and $Var\left\{\widetilde{\mathbf{S}}_n(\mathbf{b},\tau)\right\}$ as n and d_n increase. (C8) and (C9) control λ_n, the number of variables within group, and the magnitude of the true parameters in non-zero groups, which were used in Huang et al. [17]. The variance matrix $\Sigma(\tau)$ in Condition (C5) can be specified as follows: Define $e_G(\boldsymbol{\alpha}_0, t)$ and $\mathbf{A}(\boldsymbol{\alpha}_0)$ as in Appendix. We further define $\mathbf{h}(t, u, \mathbf{Z}_i) = \exp(\boldsymbol{\alpha}_0^T\mathbf{Z}_i)\int_u^t\left\{\mathbf{Z}_i - e_G(\boldsymbol{\alpha}_0, u)\right\}d\Lambda_0^G(v)$, $M_i^G(t) = N_i^G(t) - \int_0^t Y_i(u)\exp(\boldsymbol{\alpha}_0^T\mathbf{Z}_i)d\Lambda_0^G(u)$ and

$$q(t) = E\left(\frac{1}{n}\sum_{i=1}^n\int_0^L\left[\mathbf{h}^T(t, 0, \mathbf{Z}_i)\mathbf{A}(\boldsymbol{\alpha}_0)^{-1}\left\{\mathbf{Z}_i - e_G(\boldsymbol{\alpha}_0, t)\right\} + \frac{\exp(\boldsymbol{\alpha}_0^T\mathbf{Z}_i)I(u \leq t)}{s_G^{(0)}(\boldsymbol{\alpha}_0, u)}\right]M_i^G(u)\right)$$

$\mathbf{w}_i(\mathbf{b}) = \mathbf{Z}_i I\left\{X_i \le g(\mathbf{Z}_i^T \mathbf{b})\right\} I(\delta_i = 1) q(X_i) / G(X_i | \mathbf{Z}_i)$. Then, $\sum(\tau) = E\left\{\eta_1(\tau)\eta_1(\tau)^T\right\}$, where $\eta_i(\tau) = \mathbf{Z}_i[I\left\{X_i \le g(\mathbf{Z}_i^T \boldsymbol{\beta}_0^T(\tau))\right\} I(\delta_i = 1) / G(X_i | \mathbf{Z}_i) - \tau] + \mathbf{w}_i\left\{\boldsymbol{\beta}_0(\tau)\right\}$. The s includes the detailed derivation of $\eta_i(\boldsymbol{\tau})$ and the asymptotic normality of the estimator obtained by solving $\mathbf{S}_n(\mathbf{b}, \boldsymbol{\tau}) = \mathbf{0}$ for fixed d_n. Denote $\rightarrow_d$ as convergence in distribution.

First of all, the following lemma shows the consistency of the estimator obtained by solving $\mathbf{S}_n(\mathbf{b}, \boldsymbol{\tau}) = \mathbf{0}$ when d_n diverges as $n \longrightarrow \infty$:

Lemma 2.3

Let $\tilde{\boldsymbol{\beta}}(\tau)$ be the estimator obtained by solving $\mathbf{S}_n(\mathbf{b}, \boldsymbol{\tau}) = \mathbf{0}$. Then, under the conditions (C1) − (C7), we have $\left\|\tilde{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}_0(\tau)\right\| = O_p\left(\sqrt{d_n/n}\right)$.

The proof of Lemma 2.3 can be found in the online Supplementary Materials. Peng and Fine [1] studied the consistency of $\tilde{\boldsymbol{\beta}}(\tau)$ for non-covariate dependent censoring with fixed number of covariates. Lemma 2.3 extends their result to covariate-dependent censoring with diverging d_n. Similarly to Huang et al. [17], we have the following theorem for the group bridge estimator given $\boldsymbol{\tau}$:

Theorem 2.4

Assume $\nu = 0$ in (2). Under (C1) − (C9), we have

1. *Consistency:* $\left\|\hat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}_0(\tau)\right\| = O_p\left(\sqrt{d_n/n}\right)$.

2. *Group variable selection consistency:* $P\left\{\hat{\boldsymbol{\beta}}_{E_2}(\tau) = 0\right\} \rightarrow 1$.

3. *Asymptotic distribution: for fixed unknown* $\left\{E_1, \boldsymbol{\beta}_{E_1,0}\right\}$,

$$\sqrt{n}\left\{\hat{\boldsymbol{\beta}}_{E_1}(\tau) - \boldsymbol{\beta}_{E_1,0}(\tau)\right\} \rightarrow_d N[\mathbf{0}, \mathbf{H}_{11}^*\left\{\boldsymbol{\beta}_0(\tau)\right\}^{-1} \sum\nolimits_{11}^*(\tau) \mathbf{H}_{11}^*\left\{\boldsymbol{\beta}_0(\tau)\right\}^{-1}],$$

   where $\mathbf{H}_{11}^*\left\{\boldsymbol{\beta}_0(\tau)\right.$ and $\sum_{11}^*(\tau)$ are the leading $|E_1| \times |E_1|$ submatrices of $\mathbf{H}\{\boldsymbol{\beta}_0(\boldsymbol{\tau})\}$ and $\Sigma(\boldsymbol{\tau})$, respectively.

Using Lemma 2.3, Theorem 2.4 can be shown similarly to the proofs of Theorems 1 and 2 of Huang et al. [17] and thus its proof is omitted. Theorem 2.4 shows the group variable selection consistency $\sqrt{n/d_n}$-consistency of the group bridge estimator.

Although the group bridge can consistently select non-zero group variables, it may not effectively eliminate zero individual variables within non-zero group variables. This may be improved with using $\nu > 0$ in (2), that is, the adaptive group bridge penalty. For the adaptive group bridge, we have the following theorem given $\boldsymbol{\tau}$:

Theorem 2.5

Assume $\nu > 0$ in (2). In addition to (C1) − (C7), we assume

(C8b) For some ν_1 and ν_2 such that $0 < \nu_1 < 1$, $0 < \nu_2$, and $\nu_2/(1-\nu_1) < \nu$, $\min_{j\epsilon B_1} |\beta_{0,j}(\tau)| = O_p\left\{(d_n/n)^{\nu_1/2}\right\}$, $\max_k |A_k \cap B_1| = O\left\{(n/d_n)^{\nu_2/2}\right\}$, and

$$\sum_{k=1}^{K_1} c_k \left\{ \left(\sum_{j\epsilon A_k \cap B_1} |\beta_{j,0}(\tau)|^{1-\nu} \right)^{\gamma-1} \sum_{j\epsilon A_k \cap B_1} \frac{1}{|\beta_{j,0}(\tau)|^{\nu}} \right\} = O_p(\sqrt{d_n}).$$

$(C9b) \lambda_n/\sqrt{n} \to 0$, $\sqrt{n/d_n}\tilde{\beta}_j = O_p(1)$, and $\min\left(\lambda_n n^{(\nu-1)/2} d_n^{-(1+\nu)/2}, \lambda_n n^{\gamma(\nu-1)/2} d_n^{-1+\gamma(1-\nu)/2}\right) \to \infty$.

Then, we have

1. *Consistency:* $\left\|\hat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}_0(\tau)\right\| = O_p(\sqrt{d_n/n})$.

2. *Bi-level variable selection consistency:* $P\left\{\hat{\boldsymbol{\beta}}_{B_2}(\tau) = 0\right\} \to 1$.

3. *Asymptotic distribution:for fixed unknown* $\left\{B_1, \boldsymbol{\beta}_{B_1,0}\right\}$,

$$\sqrt{n}\left\{\hat{\boldsymbol{\beta}}_{B_1}(\tau) - \boldsymbol{\beta}_{B_1,0}(\tau)\right\} \to_d N[\mathbf{0}, \mathbf{H}_{11}\left\{\boldsymbol{\beta}_0(\tau)\right\}^{-1} \Sigma_{11}(\tau) \mathbf{H}_{11}\left\{\boldsymbol{\beta}_0(\tau)\right\}^{-1}],$$

   where $\mathbf{H}_{11}\{\boldsymbol{\beta}_0(\boldsymbol{\tau})\}$ and $\Sigma_{11}(\boldsymbol{\tau})$ are the leading $|B_1| \times |B_1|$ submatrices of $\mathbf{H}\{\boldsymbol{\beta}_0(\boldsymbol{\tau})\}$ and $\Sigma(\boldsymbol{\tau})$, respectively.

The proof of Theorem 2.5 can be found in the online s. (C8b) controls the magnitude of non-zero parameters and the number of non-zero parameters. It requires the smallest magnitude of non-zero parameters does not shrink towards zero too fast. (C9b) controls λ_n and ν as $n \to \infty$ to obtain the oracle property. Theorem 2.5 provides the oracle property of the adaptive group bridge estimator. In particular, it shows that the adaptive group bridge consistently identifies not only non-zero group variables, but also non-zero within-group variables.

To obtain $\hat{\boldsymbol{\beta}}$, we minimize $\tilde{W}_n(\mathbf{b}, \boldsymbol{\theta}, \tau)$ of (3). Then, the optimization algorithm is as follows:

1. Obtain an consistent estimator $\tilde{\boldsymbol{\beta}}(\tau)$ and an initial value $\boldsymbol{\beta}^{(0)}(\boldsymbol{\tau})$ from Peng and Fine [1] or the group bridge.

2. Compute

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

3. Obtain $\boldsymbol{\beta}^{(i+1)}(\boldsymbol{\tau})$ by minimizing $\widetilde{W}_n\left(\mathbf{b}, \boldsymbol{\theta}^{(i)}, \tau\right)$ with respect to **b**.

4. Repeat (2)–(3) until $\|\boldsymbol{\beta}^{(i+1)}(\boldsymbol{\tau}) - \boldsymbol{\beta}^{(i)}(\boldsymbol{\tau})\| < 10^{-4}$.

The minimization in Step 3 can be implemented using R package **quantreg** [19]. To choose a tuning parameter ζ_n in (3), we propose the following BIC-type criterion motivated by Lee et al. [20] and Shows et al. [21]:

$$\frac{2}{n}U_n\left\{\widehat{\beta}(\tau), \tau\right\} + C\log\left(d_n\right)p_n\frac{\log(n)}{2n},$$

where p_n is the number of non-zero estimates given ζ_n and C is some positive number.

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. Censoring times and event times were independently generated. Let $\mathbf{Z} = (1, \widetilde{\mathbf{Z}})^T$, $\boldsymbol{\beta}_0^{-0}(\tau) = \left\{\beta_{1,0}(\tau), \ldots, \beta_{d_n,0}(\tau)\right\}^T$, and $\boldsymbol{\zeta}_0^{-0}(\tau) = \left\{\zeta_{1,0}(\tau), \ldots, \zeta_{d_n,0}(\tau)\right\}^T$. Event times and cause of failure were generated as follows:

$$P(\varepsilon = 1) = p_1,$$

$$P\left\{T \le t | \varepsilon = 1, \widetilde{\mathbf{Z}}\right) = \Phi(\log t - \boldsymbol{\beta}_0^{-0}(\tau)^T\widetilde{\mathbf{Z}}\right\},$$

$$P\left\{T \le t | \varepsilon = 2, \widetilde{\mathbf{Z}}\right) = \Phi(\log t - \boldsymbol{\zeta}_0^{-0}(\tau)^T\widetilde{\mathbf{Z}}\right\},$$

$$\log Q_1(\tau | \mathbf{Z}) = \Phi^{-1}\left(\frac{\tau}{p_1}\right) + \boldsymbol{\beta}_0^{-0}(\tau)^T\widetilde{\mathbf{Z}},$$

$$G(t) = \exp(-\lambda_c\boldsymbol{\alpha}_0^T\widetilde{\mathbf{Z}}t).$$

Thus, $\boldsymbol{\beta}_0(\tau) = \left\{\Phi^{-1}(\tau/p_1), \boldsymbol{\beta}_0^{-0}(\tau)\right\}^T$. We set $\boldsymbol{\beta}_0^{-0}(\tau) = \zeta_0^{-0}(\tau)$. Selecting non-zero $\beta_{j,0}(\boldsymbol{\tau})$ for $j = 1, \ldots, d_n$ was of interest in this simulation study. We selected p_1 and λ_c to generate 40%

cause 1 events, 30% cause 2 events, and 30% censoring. Each simulation was conducted 1000 iterations. The competing risks quantile regression of Peng and Fine [1] and the group bridge were used to estimate $\tilde{\boldsymbol{\beta}}$. The adaptive group bridge with $\nu = 1$ was compared to the group bridge. We evaluated the mean squared error that was calculated by

$$\mathrm{MSE} = \frac{1}{1000}\sum_{i=1}^{1000}\left\|\hat{\boldsymbol{\beta}}^{i,-0}(\tau) - \boldsymbol{\beta}_0^{-0}(\tau)\right\|^2,$$

where $\hat{\boldsymbol{\beta}}^{i,-0}(\tau)$ is the estimator of $\boldsymbol{\beta}_0^{-0}(\tau)$ at the ith iteration given τ. The proposed BIC-type criterion with $C = 1.5$ was used to select the tuning parameter. Two τ values were examined: $\tau = 0.1$ and 0.25. We first considered Setting i) group variables consisting of continuous variables with non-covariate dependent censoring distribution, that is, $\boldsymbol{\alpha}_0 = \mathbf{0}$. We examined $n = 400$, 600, and 800. To generate $\tilde{\mathbf{Z}}$, three correlated continuous variables for each group were generated from $N(\mathbf{0}, \Sigma)$, where

$$\sum = \begin{pmatrix} 1 & 0.5 & 0.5 \\ 0.5 & 1 & 0.5 \\ 0.5 & 0.5 & 1 \end{pmatrix}.$$

Variables were assumed to be independent if they belong to different groups. For $n = 400$, 600, and 800, there were 9, 10, 11 groups, respectively. The true $\boldsymbol{\beta}_0(\tau)$ for $n = 400$ was $\{\beta_{1,0}(\tau), \ldots, \beta_{9,0}(\tau)\}^T = (1, -1, 0, -1, 1, 0, 1, 0, 0)^T$ and $\{\beta_{10,0}(\tau), \ldots, \beta_{27,0}(\tau)\}^T = (0, \ldots, 0)^T$. For n=600, we added $\{\beta_{28,0}(\tau), \beta_{29,0}(\tau), \beta_{30,0}(\tau)\}^T = (0,0,0)^T$. For n=800, we further added $\{\beta_{31,0}(\tau), \beta_{32,0}(\tau), \beta_{33,0}(\tau)\}^T = (0,0,0)^T$. This setting allowed d_n to grow as n increased. The number of non-zero groups and non-zero individual variables of the underlying model were 3 and 5, respectively, for each n.

Table 1 summarizes the simulation results. "AGB-CQ", "AGB-GB", and "GB" indicate the adaptive group bridge with $\tilde{\boldsymbol{\beta}}(\tau)$ from Peng and Fine [1], the adaptive group bridge with $\tilde{\boldsymbol{\beta}}(\tau)$ from the group bridge, and the group bridge, respectively. "% Corr. Group" and "% Corr. Individual" represent the proportions that the corresponding variable selection method correctly identified the non-zero group variables and non-zero individual variables of the underlying model, respectively. "Group Size" and "Model Size" are the mean number of groups and individual variables selected by each variable selection method, respectively. "MSER" is the ratio of the median MSE of each variable selection method to that of the oracle estimator. The adaptive group bridge and the group bridge identified the true non-zero and zero groups very well in group variable selection. The mean group sizes of the adaptive group bridge and the group bridge were very close to 3. However, the group bridge performed poorly in within-group variable selection, that is, individual variable selection. It over-identified individual variables as non-zero variables. On the other hand, the adaptive group bridge correctly identified the true non-zero individual variables well. In addition, as n increased, the mean group sizes and the mean model sizes of the adaptive group bridge became closer to 3 and 5, respectively. The MSERs of the adaptive group bridge with $\tilde{\boldsymbol{\beta}}(\tau)$ from the group bridge was lower than those of the other methods. Furthermore, the MSERs of the adaptive group bridge got smaller as n increased in general. We also conducted a simulation under the same setting except that pairwise correlation between continuous variables was assumed to be 0.2 if they belonged to different groups. We had similar results to Table 1 and thus did not report them.

Next, we performed a simulation study for Setting ii) group variables consisting of continuous variables and categorical variables with non-covariate dependent censoring distribution. We examined 3 sample sizes: $n = 600$, 900, and 1200. For $n = 600$, there were 10 groups: 5 groups consisting of continuous variables (Groups 1 to 5) and 5 groups consisting of categorical variables (Groups 6 to 10). Groups 1 and 2 contained 6 continuous variables each and Groups 3 to 5 were comprised of 3 continuous variables each. The pairwise correlation among continuous variables within group was 0.5. There was no correlation between continuous variables if they belonged to different groups. Groups 6 and 7 consisted of 7 categories each (that is, 6 indicator variables each) and Groups 8 to 10 categories had 4 categories each (that is, 3 indicator variables each). The reference group for each categorical variable was set to 0. Thus, there were 42 variables in total. The true $\boldsymbol{\beta}_0(\tau)$ for $n = 600$ was $\{\beta_{1,0}(\tau), \ldots, \beta_{6,0}(\tau)\}^T = (1, -1, 0, \ldots, 0)^T$, $\{\beta_{7,0}(\tau), \ldots, \beta_{12,0}(\tau)\}^T = (0, \ldots, 0)^T$, $\{\beta_{13,0}(\tau), \beta_{14,0}(\tau), \beta_{15,0}(\tau)\}^T = (1,0,0)^T$, and $\{\beta_{16}(\tau), \ldots, \beta_{21}(\tau)\}^T = (0, \ldots, 0)^T$, $\{\beta_{22,0}(\tau), \ldots, \beta_{27,0}(\tau)\}^T = (1, -1, 0, \ldots, 0)^T$, $\{\beta_{28,0}(\tau), \ldots, \beta_{33,0}(\tau)\}^T = (0, \ldots, 0)^T$, $\{\beta_{34,0}(\tau), \beta_{35,0}(\tau), \beta_{36,0}(\tau)\}^T = (1,0,0)^T$, and $\{\beta_{37}(\tau), \ldots, \beta_{42}(\tau)\}^T = (0, \ldots, 0)^T$. For $n = 900$, we added one more group consisting of 3 continuous variables with pairwise correlation 0.5 and $\{\beta_{43,0}(\tau), \beta_{44,0}(\tau), \beta_{45,0}(\tau)\}^T = (0,0,0)^T$. For $n = 1200$, in addition to $\{\beta_{43,0}(\tau), \beta_{44,0}(\tau), \beta_{45,0}(\tau)\}^T$, we further added a categorical variable having 4 categories, that is, 3 indicator variables: $\{\beta_{46,0}(\tau), \beta_{47,0}(\tau), \beta_{48,0}(\tau)\}^T = (0,0,0)^T$. Thus, the number of non-zero groups and non-zero individual variables of the underlying model were 4 and 6, respectively, for each n.

Table 2 shows the simulation results. The adaptive group bridge identified the true non-zero and zero groups better than the group bridge when $n = 600$ and 900 for $\tau = 0.1$, and $n = 600$ for $\tau = 0.25$. When $n = 1200$, both of the methods selected non-zero group variables very well. The mean group sizes of the adaptive group bridge were very close to 4. The group bridge performed poorly in individual variable selection as in Setting i). On the other hand, the adaptive group bridge correctly identified the true non-zero individual variables proficiently. In addition, as n increased, the mean group sizes and the mean model sizes of the adaptive group bridge became closer to 4 and 6, respectively. The MSERs of the adaptive group bridge with $\widetilde{\boldsymbol{\beta}}(\tau)$ from the group bridge was lower than those of the other methods. In addition, the MSERs of the adaptive group bridge got smaller as n increased in general.

Last, we performed a simulation study for Setting ii) with covariate-dependent censoring distribution. We used the same $\boldsymbol{\beta}_0(\tau)$ as in Setting ii) with non-covariate dependent censoring distribution. The true $\boldsymbol{\alpha}_0$ for $G\left(t|\widetilde{\mathbf{Z}}\right)$ when $n = 600$ was $(\alpha_{1,0}, \ldots, \alpha_{6,0})^T = (1,-1,0, \ldots, 0)^T$, $(\alpha_{7,0}, \ldots, \alpha_{21,0})^T = (0, \ldots, 0)^T$, $(\alpha_{22,0}, \ldots, \alpha_{27,0})^T = (1,-1, 0, \ldots, 0)^T$, and $(\alpha_{28,0}, \ldots, \alpha_{42,0})^T = (0, \ldots, 0)^T$. For $n = 900$ and 1200, we added $(\alpha_{43,0}, \alpha_{44,0}, \alpha_{45,0})^T = (0, 0, 0)^T$ and $(\alpha_{46,0}, \alpha_{47,0}, \alpha_{48,0})^T = (0, 0, 0)^T$, respectively. The Breslow-type estimator was used to estimate $G\left(t|\widetilde{\mathbf{Z}}\right)$. We selected p_1 and λ_c to generate 50% cause 1 events, 20% cause 2 events, and 30%

censoring. Table 3 summarizes the simulation results. In general, the results were similar to Table 2. The adaptive group bridge performed better than the group bridge in terms of individual variable selection and MSER.

4. Bone marrow transplant data example

The adaptive group bridge was applied to a bone marrow transplant data set. Verneris et al. [10] studied the outcomes of the patients having reduced-intensity conditioning allogeneic hematopoietic cell transplantation from 1999 to 2011. We considered 2011 patients with human leukocyte antigen fully-matched unrelated donors. Relapse was the outcome of interest for the analysis. Treatment-related-mortality (TRM) was a competing risk. There were 40.5% of relapse, 26.6% of TRM, and 32.9% of censoring. In addition, 69%, 16%, and 8% of relapse events occurred within 6 months, between 6 and 12 months, and between 12 and 24 months, respectively. Thus, the distribution of relapse events were skewed. The overall relapse rate at 1 year was about 35%. The 13 binary or categorical variables that we considered for variable selection included disease type, recipient age, donor age, donor-recipient sex match, donor-recipient cytomegalovirus (CMV) match, ABO blood type match, donor parity, disease status at transplant, conditioning intensity, total body irradiation, graft type, graft-versus-host disease (GVHD) prophylaxis, and in-vivo T cell depletion. They consisted of 28 indicator variables. The censoring distribution did not depend on any covariates based on the Cox proportional hazards model.

We selected variables for the 0.35th competing risks quantile regression for relapse using the following three selection methods: the group bridge, the adaptive group bridge with $\widetilde{\boldsymbol{\beta}}(\tau)$ from Peng and Fine [1], and the adaptive group bridge with $\widetilde{\boldsymbol{\beta}}(\tau)$ from the group bridge. The reference group was set to zero. Table 4 shows the selected variables and their estimates. The group bridge selected disease status at transplant, CMV match, conditioning intensity, in-vivo T cell depletion, graft type, and GVHD prophylaxis. On the other hand, both of the adaptive group bridge with $\widetilde{\boldsymbol{\beta}}(\tau)$ from Peng and Fine [1] and the adaptive group bridge with $\widetilde{\boldsymbol{\beta}}(\tau)$ from the group bridge selected the same variables: disease status at transplant, CMV match, conditioning intensity, and in-vivo T cell depletion. The adaptive group bridge did not select graft type and GVHD prophylaxis, which is why all of their estimates are zeros in Table 4. The competing risks quantile regression of Peng and Fine [1] was fitted using the variables selected by at least one of the three methods. “CQ” in Table 4 indicates their estimates and p-values from the competing risks quantile regression of Peng and Fine [1]. It suggests that all variables selected by the adaptive group bridge appeared to be significant. However, graft type and GVHD prophylaxis that the group bridge selected appeared not to be significant based on their *p*-values.

5. Conclusion

The group bridge and the adaptive group bridge were proposed to select variables for the competing risks quantile regression. Their oracle property was studied. In particular, the adaptive group bridge not only consistently identifies non-zero group variables, but also consistently selects non-zero within-group variables. We also proposed the BIC-type criterion to choose a tuning parameter. The proposed BIC-type criterion appears to work properly in the simulation study. The adaptive group bridge selected non-zero within-group variables more consistently than the group bridge in the simulation study. A bone marrow transplant example showed the usefulness of the adaptive group bridge.

The proposed method was limited to when $d_n < n$. Developing a group variable selection method when $d_n < n$ would be a crucial research problem. A two-step variable selection procedure may be developed for this: once we screen group variables in the first step, we may use the adaptive group bridge to obtain a further parsimonious list of non-zero variables in the second step. The theoretical justification of the proposed BIC-type criterion needs to be studied in the future.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported in part by Institutional Research Grant #14-247-29 from the American Cancer Society and the MCW Cancer Center, and the US National Cancer Institute (U24CA076518). The authors would like to thank the Associate Editor and two anonymous reviewers for their helpful comments that significantly improved the manuscript.

Appendix

For $G\left(t|\tilde{\mathbf{Z}}\right)$ and the Breslow estimator, we assume as follows:

a. $\int_0^L \lambda_0^G(t)dt < \infty$ and $P\{Y_i(t) = 1\} > 0$ for $t \in [0, L]$, $i = 1, \ldots, n$, and $d_n^4/n \to 0$ as $n \to \infty$.

b. Z_{ij} is bounded almost surely for all i,j and $\boldsymbol{\alpha}^T\tilde{\mathbf{Z}}$ is bounded almost surely for any $\tilde{\mathbf{Z}}$ and $\boldsymbol{\alpha} \in \mathscr{B}$, where $\mathscr{B}$ is a neighborhood $\boldsymbol{\alpha}_0$.

c. For $d = 0, 1, 2$, there exists a neighborhood $\mathscr{B}$ of $\boldsymbol{\alpha}_0$ such that $s_G^{(d)}(\boldsymbol{\alpha}, t)$ are continuous functions and $\sup_{t \in (0, L), \boldsymbol{\alpha} \in \mathscr{B}}\left\|\mathbb{S}_G^{(d)}(\boldsymbol{\alpha}, t) - s_G^{(d)}(\boldsymbol{\alpha}, t)\right\| \to 0$ in probability.

d. The matrix $\mathbf{A}\left(\boldsymbol{\alpha}_0\right) = \int_0^L v_G\left(\boldsymbol{\alpha}_0, t\right)s_G^{(0)}(\boldsymbol{\alpha}_0, t)\lambda_0^G(t)dt$ is positive definite, where $v_G(\boldsymbol{\alpha}, t) = s_G^{(2)}(\boldsymbol{\alpha}, t)/s_G^{(0)}(\boldsymbol{\alpha}, t) - e_G(\boldsymbol{\alpha}, t)^{\otimes 2}$ and $e_G(\boldsymbol{\alpha}, t) = s_G^{(1)}(\boldsymbol{\alpha}, t)/s_G^{(0)}(\boldsymbol{\alpha}, t)$.

e. For all $\boldsymbol{\alpha} \in \mathscr{B}$, $t \in [0, L]$, $\mathbb{S}_G^{(1)}(\boldsymbol{\alpha}, t) = \partial\mathbb{S}_G^{(0)}(\boldsymbol{\alpha}, t)/\partial\boldsymbol{\alpha}$, and $\mathbb{S}_G^{(2)}(\boldsymbol{\alpha}, t) = \partial^2\mathbb{S}_G^{(0)}(\boldsymbol{\alpha}, t)/(\partial\boldsymbol{\alpha}\partial\boldsymbol{\alpha}^T)$, where $\mathbb{S}_G^{(d)}(\boldsymbol{\alpha}, t)$, $d = 0, 1, 2$ are continuous functions of $\boldsymbol{\alpha} \in \mathscr{B}$ uniformly in $t \in [0, L]$ and are bounded on $\mathscr{B} \times [0, L]$, and $s_G^{(0)}$ is bounded away from zero on $\mathscr{B} \times [0, L]$.

References

1. Peng L, Fine JP. Competing risks quantile regression. Journal of the American Statistical Association. 2009; 104:1440–1453.
2. Peng L, Huang Y. Survival analysis with quantile regression models. Journal of the American Statistical Association. 2008; 103:637–649.
3. Reich BJ, Smith LB. Bayesian quantile regression for censored data. Biometrics. 2013; 69:651–660. [PubMed: 23844559]
4. Yin G, Zeng D, Li H. Power–transformed linear quantile regression with censored data. Journal of the American Statistical Association. 2008; 103:1214–1224.
5. Yin G, Cai J. Quantile regression models with multivariate failure time data. Journal of the American Statistical Association. 2005; 61:151–161.
6. Sun Y, Wang HJ, Gilbert PB. Quantile regression for competing risks data with missing cause of failure. Statistica Sinica. 2012; 22:703–728. [PubMed: 23950622]
7. Lee M, Fine J. Inference for cumulative incidence quantiles via parametric and nonparametric approaches. Statistics in Medicine. 2011; 30:3221–3235. [PubMed: 21898525]
8. Jiang R, Qian W, Zhou Z. Variable selection and coefficient estimation via composite quantile regression with randomly censored data. Statistics & Probability Letters. 2012; 82:308–317.
9. Wang HJ, Zhou J, Li Y. Variable selection for censored quantile regression. Statistica Sinica. 2013; 23:145–167. [PubMed: 23847407]
10. Verneris MR, Lee SJ, Ahn KW, Wang HL, Battiwalla M, Inamoto Y, Munker R, Aljurf M, Saber W, Spellman S, et al. HLA-mismatch is associated with worse outcomes after unrelated donor reduced intensity transplantation: An analysis from the CIBMTR. Biology of Blood and Marrow Transplant. 2015; 21:1783–1789.
11. 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.
12. Huang J, Ma S, Xie H, Zhang CH. A group bridge approach for variable selection. Biometrika. 2009; 96:339–355. [PubMed: 20037673]
13. Zhou N, Zhu J. Group variable selection via a hierarchical lasso and its oracle property. Statistics and Its Interface. 2010; 3:557–574.
14. Zhao W, Zhang R, Liu J. Sparse group variable selection based on quantile hierarchical lasso. Journal of Applied Statistics. 2014; 41:1658–1677.
15. Fu Z, Parikh CR, Zhou B. Penalized variable selection in competing risks regression. Lifetime Data Analysis. 2016; 23:353376.
16. Breslow NE. Discussion of the paper by d. r.cox. Journal of the Royal Statistical Society: Series B. 1972; 34:216–217.
17. 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.
18. Cai J, Fan J, Li R, Zhou H. Variable selection for multivariate failure time data. Biometrika. 2005; 92:303–316. [PubMed: 19458784]
19. Koenker, R. quantreg: Quantile Regression. 2016. URL https://CRAN.R-project.org/package=quantreg, r package version 5.26
20. Lee ER, Noh H, Park BU. Model selection via bayesian information criterion for quantile regression models. Journal of the American Statistical Association. 2014; 109:216–229.
21. Shows JH, Lu W, Zhang HH. Sparse estimation and inference for censored median regression. Journal of Statistical Planning and Inference. 2010; 140:1903–1917. [PubMed: 20607110]

Table 1

Simulation results for group variables consisting of continuous variables with $G(X)$.

τ	n	Method	% Corr. Group	% Corr. Individual	Group Size	Model Size	MSER
0.1	400	AGB-CQ	0.995	0.987	3.005	5.013	2.108
		AGB-GB	0.991	0.976	3.001	5.011	1.159
		GB	0.988	0.554	3.012	5.570	2.451
	600	AGB-CQ	0.996	0.996	3.004	5.005	2.281
		AGB-GB	0.996	0.992	2.999	5.000	1.056
		GB	0.994	0.622	3.008	5.472	2.491
	800	AGB-CQ	0.997	0.996	3.003	5.004	1.834
		AGB-GB	0.997	0.995	3.003	5.005	1.024
		GB	0.997	0.627	3.003	5.434	2.315
0.25	400	AGB-CQ	0.917	0.855	3.097	5.185	1.316
		AGB-GB	0.931	0.871	3.076	5.150	1.041
		GB	0.945	0.363	3.062	5.963	1.543
	600	AGB-CQ	0.936	0.884	3.073	5.141	1.293
		AGB-GB	0.948	0.892	3.054	5.115	1.077
		GB	0.960	0.358	3.041	5.977	1.583
	800	AGB-CQ	0.949	0.905	3.059	5.111	1.041
		AGB-GB	0.966	0.929	3.038	5.085	1.006
		GB	0.973	0.404	3.028	5.875	1.463

Table 2

Simulation results for group variables consisting of continuous and categorical variables with $G(X)$.

τ	n	Method	% Corr. Group	% Corr. Individual	Group Size	Model Size	MSER
0.1	600	AGB-CQ	0.776	0.500	3.737	5.354	4.061
		AGB-GB	0.870	0.652	3.861	5.603	1.621
		GB	0.536	0.136	3.348	5.485	8.226
	900	AGB-CQ	0.952	0.809	3.955	5.835	3.216
		AGB-GB	0.989	0.913	3.989	5.934	1.344
		GB	0.830	0.306	3.798	6.386	4.387
	1200	AGB-CQ	0.990	0.910	3.998	5.940	3.283
		AGB-GB	0.999	0.973	4.001	5.989	1.248
		GB	0.955	0.383	3.955	6.704	4.147
0.25	600	AGB-CQ	0.933	0.647	4.015	6.097	1.925
		AGB-GB	0.955	0.768	4.023	6.098	1.421
		GB	0.881	0.187	3.897	7.187	2.660
	900	AGB-CQ	0.974	0.810	4.025	6.163	1.681
		AGB-GB	0.979	0.879	4.019	6.118	1.212
		GB	0.979	0.220	3.997	7.393	2.484
	1200	AGB-CQ	0.965	0.832	4.033	6.167	1.519
		AGB-GB	0.964	0.888	4.034	6.126	1.250
		GB	0.991	0.264	4.005	7.370	2.730

Table 3

Simulation results for group variables consisting of continuous and categorical variables with $G(X \mid \tilde{\mathbf{Z}})$

τ	n	Method	% Corr. Group	% Corr. Individual	Group Size	Model Size	MSER
0.1	600	AGB-CQ	0.769	0.547	3.725	5.393	3.096
		AGB-GB	0.862	0.687	3.850	5.615	1.625
		GB	0.533	0.155	3.372	5.538	7.692
	900	AGB-CQ	0.956	0.830	3.976	5.859	2.504
		AGB-GB	0.981	0.902	3.991	5.915	1.356
		GB	0.796	0.287	3.849	6.411	4.049
	1200	AGB-CQ	0.996	0.912	3.996	5.945	3.250
		AGB-GB	0.998	0.963	4.000	5.982	1.248
		GB	0.949	0.370	3.946	6.669	3.972
0.25	600	AGB-CQ	0.940	0.674	4.007	6.040	1.881
		AGB-GB	0.963	0.803	4.024	6.051	1.346
		GB	0.907	0.210	3.914	7.179	2.376
	900	AGB-CQ	0.966	0.854	4.031	6.099	1.680
		AGB-GB	0.967	0.891	4.031	6.093	1.206
		GB	0.885	0.261	4.088	7.363	2.350
	1200	AGB-CQ	0.980	0.888	4.020	6.099	1.537
		AGB-GB	0.986	0.920	4.014	6.078	1.177
		GB	0.996	0.296	4.000	7.202	2.386

Table 4

Selected variables and estimates. “ref” means the reference group. “CQ” indicates the competing risks quantile regression.

Variable	Subcategory	AGB-CQ	AGB-GB	GB	CQ	
		Est.	Est.	Est.	Est.	*p*-value
Disease status	Early (ref)	0	0	0	0	
	Intermediate	0	0	0	−0.055	0.503
	Advanced	−0.891	−0.891	−0.859	−0.460	< 0.001
CMV match	+/+ (ref)	0	0	0	0	
	+/−	−0.455	−0.445	−0.476	−0.363	0.018
	−/+	0	0	0	−0.034	0.902
	−/−	0	0	0	−0.027	0.774
	Missing	0	0	0	−0.054	0.939
Conditioning intensity	Reduced intensity (ref)	0	0	0	0	
	Nonmyeloablative	−1.056	−1.060	−1.165	−0.536	< 0.001
In-vivo T cell depletion	No (ref)	0	0	0	0	
	Yes	−0.488	−0.488	−0.466	−0.301	0.010
Graft type	Bone marrow (ref)	0	0	0	0	
	Peripheral blood	0	0	0.181	0.130	0.137
GVHD prophylaxis	FK506 ± others (ref)	0	0	0	0	
	Others	0	0	0.175	0.121	0.212