Received: 9 April 2020 | Revised: 27 August 2021 | Accepted: 28 August 2021

DOI: 10.1002/sim.9197

RESEARCH ARTICLE

Statistics in Medicine WILEY

Penalized variable selection for cause-specific hazard frailty models with clustered competing-risks data

Trias W. Rakhmawati[1] | **Il Do Ha**[2] | **Hangbin Lee**[1] | **Youngjo Lee**[1]

[1]Department of Statistics, Seoul National University, Seoul, South Korea

[2]Department of Statistics, Pukyong National University, Busan, South Korea

Correspondence
Il Do Ha, Department of Statistics, Pukyong National University, Busan 48513, South Korea.
Email: idha1353@pknu.ac.kr

Competing risks data usually arise when an occurrence of an event precludes other types of events from being observed. Such data are often encountered in a clustered clinical study such as a multi-center clinical trial. For the clustered competing-risks data which are correlated within a cluster, competing-risks models allowing for frailty terms have been recently studied. To the best of our knowledge, however, there is no literature on variable selection methods for cause-specific hazard frailty models. In this article, we propose a variable selection procedure for fixed effects in cause-specific competing risks frailty models using a penalized h-likelihood (HL). Here, we study three penalty functions, LASSO, SCAD, and HL. Simulation studies demonstrate that the proposed procedure using the HL penalty works well, providing a higher probability of choosing the true model than LASSO and SCAD methods without losing prediction accuracy. The proposed method is illustrated by using two kinds of clustered competing-risks cancer data sets.

KEYWORDS

competing risks, frailty models, h-likelihood, penalized likelihood, variable selection

1 | INTRODUCTION

Competing risks data are encountered in various research areas including biomedical research, engineering, and econometrics. Two broad classes of competing-risks regression models for analyzing such data are the cause-specific hazard model[1] and the subdistribution hazard model.[2] The former is to model the cause-specific hazard of each event type separately, while the latter is to model the hazard function of subdistribution (subhazard) for a particular event of interest. Recently, the two types of models have been extended to clustered competing risks data, which are correlated within a cluster, via frailty.[3-8]

Recently, variable selection methods using a penalized likelihood allowing for various penalty functions have been widely developed in linear models, generalized linear models, and Cox's proportional hazards (PH) models. The main advantage of this method is to select important covariates and to estimate the regression coefficients, simultaneously; that is, it deletes insignificant variables by estimating their coefficients as zero.[9] In this article, we propose a hierarchical likelihood (h-likelihood[10]) approach for variable selection of fixed effects in cause-specific competing risk frailty models. Unlike the classical likelihood for fixed parameters only, the h-likelihood is constructed for both fixed parameters and unobserved frailties at the same time. The h-likelihood avoids the integration for random effects itself,[11] whereas the marginal likelihood approach often involves intractable integrations when eliminating the frailties. Ha et al[12] have developed a variable selection procedure for fixed effects in standard frailty models using a penalized h-likelihood with a penalty function (eg, LASSO).[13] The penalized h-likelihood method has been extended to the variable selection in subhazard

competing-risks frailty models.[14] Recently, Fu et al[15] and Ahn et al[16] have proposed penalized variable-selection methods under subhazard competing-risks models without frailty term. Furthermore, Hou et al[17] developed a penalized high-dimensional variable selection framework under both cause-specific hazard and subhazard competing-risks models without frailty term. However, their variable selection methods have been mainly developed assuming independence between subjects. To the best of our knowledge, there is no literature on variable selection methods for the cause-specific hazard frailty model. In this article, we further extend the penalized procedures to variable selection in cause-specific hazard frailty models, leading to a fast and efficient procedure. We study three penalty functions, LASSO,[13] SCAD,[9] and HL.[18]

In Section 2, we review the cause-specific competing-risks hazard frailty models, together with the corresponding h-likelihood. In Section 3, we propose a variable selection procedure for the cause-specific hazard frailty models using a penalized h-likelihood. In Section 4, simulation studies are presented to evaluate the performance of the proposed method. In Section 5, the proposed method is illustrated with two kinds of clustered competing-risks cancer data sets. We conclude with a discussion in Section 6.

2 | CAUSE-SPECIFIC HAZARD FRAILTY MODELS AND H-LIKELIHOOD

2.1 | The model

Suppose that there are $i = 1, \dots, q$ clusters (centers) where each center has $j = 1, \dots, n_i$ observations, so that the total sample size is $n = \sum_{i=1}^{q} n_i$. Here q is the number of clusters and n_i is cluster size. For a subject j in cluster i, let T_{ijk} be time to event from cause k ($k = 1, 2, \dots, K$) and let $\epsilon_{ij} \in \{1, 2, \dots, K\}$ be the corresponding cause of event. Let $T_{ij} = \min(T_{ij1}, T_{ij2}, \dots, T_{ijK})$ be time to the first event and let C_{ij} denote the independent censoring time. An unobserved frailty variable for cluster i is denoted by U_i . Suppose that given $U_i = u_i$, C_{ij} is conditionally independent and non-informative of (T_{ij}, ϵ_{ij}) for $j = 1, \dots, n_i$. For simplicity, we consider two types of event ($k = 1, 2$); let T_{ij1} be an event (Type 1) time of interest and T_{ij2} be a competing event (Type 2) time. These results can be easily generalized to K even types. We denote U_{ik} ($k = 1, 2$) be the frailty for Type k event in cluster i.

The cause-specific hazard function conditional on the frailty $U_{ik} = u_{ik}$ for the jth observation in the cluster i which failed from cause k ($k = 1, 2$) is described as[7]

$$\lambda_{ijk}(t|u_{ik}) = \lambda_{0k}(t)\exp(x_{ij}^T\beta_k)u_{ik}, \tag{1}$$

where $\lambda_{0k}(t)$ is unspecified baseline hazard function for event type k and $\beta_k = (\beta_{k1}, \dots, \beta_{kp})^T$ is a $p \times 1$ vector of regression parameters for event type k, and $x_{ij} = (x_{ij1}, \dots, x_{ijp})^T$ is a $p \times 1$ vector of fixed covariates corresponding to β_k. If there is only one event type $K = 1$, then the cause-specific frailty model (1) simply reduces to the standard univariate frailty model.[19] Here, u_{i1} and u_{i2} in the same cluster i can be correlated; a shared bivariate frailty is often considered

$$u_{i1} = u_i \text{ and } u_{i2} = u_i^{\gamma},$$

where γ is a real-valued association parameter that describes the dependency between Type 1 and 2 events. For the distribution of U_i, log-normal distribution[20] or gamma distribution[7,21] has been usually assumed; the corresponding distributions are, respectively, $V_i = \log U_i \sim N(0, \alpha)$ and U_i follows a gamma distribution with mean 1 and variance α.

Let $v_i = \log u_i$. Then the model (1) is expressed as a joint model with a common log-frailty v_i as follows. The event times (T_{ij1}) from Type 1 event follow a cause-specific PH model

$$\lambda_{ij1}(t|v_i) = \lambda_{01}(t)\exp(x_{ij}^T\beta_1 + v_i), \tag{2}$$

and event times (T_{ij2}) from Type 2 event follow a similar model

$$\lambda_{ij2}(t|v_i) = \lambda_{02}(t)\exp(x_{ij}^T\beta_2 + \gamma v_i). \tag{3}$$

If $\gamma > 0$ a cluster with higher frailty will experience an earlier competing event, whereas if $\gamma < 0$ the competing event will be more likely delayed for a cluster with higher frailty.

2.2 | H-likelihood construction

The observed event time and the event indicator are defined as $Y_{ij} = \min(T_{ij}, C_{ij})$ and $\delta_{ijk} = I(Y_{ij} = T_{ijk})$, respectively. Here, $T_{ij} = \min(T_{ij1}, T_{ij2})$ is time to the first event, $\delta_{ijk} = 1$ if Type k event occurs first (ie, $Y_{ij} = T_{ijk}$) and 0 otherwise, and $I(\cdot)$ is the indicator function. Note that δ_{ijk} is often referred to a cause-specific event indicator and that it can also be expressed as

$$\delta_{ijk} = I(T_{ij} \leq C_{ij})I(\epsilon_{ij} = k).$$

Now, we construct the h-likelihood for the cause-specific models with shared log-normal and gamma frailty structures having (2) and (3). Following Lee and Nelder[10] and Ha et al,[19] the h-likelihood for the joint competing-risks model above is defined by

$$h = h(\beta, \gamma, \lambda_0, v, \alpha) = \sum_{ijk} \ell_{1ijk}(\beta_k, \lambda_{0k}, \gamma; y_{ij}, \delta_{ijk}|v_i) + \sum_i \ell_{2i}(\alpha; v_i), \tag{4}$$

where

$$\ell_{1ijk}(\beta_k, \lambda_{0k}, \gamma; y_{ij}, \delta_{ijk}|v_i) = \delta_{ijk}(\log \lambda_{0k}(y_{ij}) + \eta_{ijk}) - \Lambda_{0k}(y_{ij}) \exp(\eta_{ijk})$$

is the conditional log-likelihood of (Y_{ij}, δ_{ij}) given $V_i = v_i$ and

$$\ell_{2i}(\alpha; v_i) = \log\{g(v_i; \alpha)\}$$

is the log-likelihood of V_i with a density function $g(v_i; \alpha)$. For example, for the log-normal frailty $\ell_{2i} = (-1/2)\log(2\pi\alpha) - v_i^2/(2\alpha)$ and for the gamma frailty $\ell_{2i} = (v_i - u_i)\alpha^{-1} + c(\alpha)$, where $c(\alpha) = -\log\Gamma(1/\alpha) - \alpha^{-1}\log\alpha$. Here, $\beta = (\beta_1^T, \beta_2^T)^T$ is a $2p \times 1$ vector of all the regression coefficients for the two event types, $\lambda_0 = (\lambda_{01}(\cdot), \lambda_{02}(\cdot))^T$ is a collection of k $(k = 1, 2)$ baseline hazards, $v = (v_1, \ldots, v_q)^T$, and $\Lambda_{0k}(\cdot)$ is the baseline cumulative hazard function for cause k. Note here that the linear predictors η_{ijk} $(k = 1, 2)$ are formed:

$$\eta_{ijk} = x_{ij}^T \beta_k + v_{ik},$$

where $v_{i1} = v_i$ and $v_{i2} = \gamma v_i$.

Notice that the functional form of baseline hazard function $\lambda_{0k}(t)$ in the h-likelihood (4) is unknown. Following Breslow[22] and Ha et al,[7,12] we consider the cumulative baseline hazard function for event type k as a step function with jumps at the observed event times,

$$\Lambda_{0k}(t) = \sum_{r: y_{(kr)} \leq t} \lambda_{0kr}, \tag{5}$$

where $y_{(k1)} < y_{(k2)} < \cdots < y_{(kD_k)}$ denote the D_k ordered distinct event times of Type k among all of the y_{ij}'s, and $\lambda_{0k} = \lambda_{0k}(y_{(kr)})$. By substituting Λ_{0k} in (5) into (4), the h-likelihood (4) can be rewritten as

$$h = \sum_{k=1}^{2} \left[\sum_{r=1}^{D_k} d_{(kr)} \log \lambda_{0kr} + \sum_{ij} \delta_{ijk}\eta_{ijk} - \lambda_{0kr} \left\{ \sum_{ij \in R_{(kr)}} \exp(\eta_{ijk}) \right\} \right] + \sum_{i=1}^{q} \ell_{2i}(\alpha; v_i), \tag{6}$$

where $d_{(kr)}$ is the number of events of Type k that occur at time $y_{(kr)}$ and $R_{(kr)} = \{ij : y_{ij} \geq y_{(kr)}\}$ is the risk set at time $y_{(kr)}$.

However, the h-likelihood (6) has high-dimensional nuisance parameter λ_{0kr} which increases with the number of events. Following Ha et al (2001), by fixing $(\beta_k, v_i, \alpha, \gamma)$, the nonparametric maximum h-likelihood (MHL) estimator of λ_{0kr}, obtained from $\partial h/\partial\lambda_{0kr} = 0$, is given by

$$\hat{\lambda}_{0kr} = \frac{d_{(kr)}}{\sum_{ij \in R_{(kr)}} \exp(\eta_{ijk})},$$

leading that $\hat{\Lambda}_{0k}(t) = \sum_{r:\ y_{(kr)} \le t} \hat{\lambda}_{0kr}$ $(k = 1, 2)$ is an extension of the Breslow estimator of the baseline cumulative hazard function. Replacing λ_{0kr} in (6) with $\hat{\lambda}_{0kr}$ gives a partial h-likelihood (PHL; denoted by h_p) with λ_{0k} eliminated:

$$h_p = h_p(\beta, \gamma, v, \alpha) = \sum_{ijk} \delta_{ijk}\eta_{ijk} - \sum_{kr} d_{(kr)} \log\left\{\sum_{ij \in R_{(kr)}} \exp(\eta_{ijk})\right\} + \sum_{i=1}^{q} \ell_{2i}(\alpha; v_i). \tag{7}$$

Accordingly, hereafter the h-likelihood inference is based on the PHL h_p.[7]

3 | VARIABLE SELECTION PROCEDURE

For the variable selection of β in the competing-risks models (1), along the lines of Ha et al,[12,14] we propose a penalized h-likelihood, h_v, using h_p and a penalty; it is defined by

$$h_v(\beta, v, \theta) = h_p - n\sum_{j=1}^{p^*} J_\xi(|\beta_j|), \tag{8}$$

where $p^* = Kp$ with $K = 2$, and $J_\xi(|\cdot|)$ is a penalty function that controls model complexity using the tuning parameter ξ. Note that we do not impose any penalty on the dispersion parameters $\theta = (\gamma, \alpha)^T$. Typically, the larger value of ξ tends to choose the simpler model, whereas the smaller value of ξ inclines to more complex model. To get the optimal value of ξ, we use a BIC-type criterion[12] as will be shown in Section 3.2; the generalized cross-validation cannot select the tuning parameter satisfactorily, with a non-ignorable overfitting effect in the resulting model.[23,24]

Various penalty functions have been used in the literature on variable selection in statistical models including the Cox PH model. Here, we consider the following three penalty functions (LASSO, SCAD, and HL), but our results can be applied to other penalty functions.

(i) LASSO:[13]

$$J_\xi(|\beta|) = \xi|\beta|.$$

(ii) SCAD:[9]

$$J'_\xi(|\beta|) = \xi I(|\beta| \le \xi) + \frac{(a\xi - |\beta|)_+}{a-1} I(|\beta| > \xi),$$

where $a = 3.7$ and x_+ denotes the positive part of x.

(iii) HL:[18]

$$J_\xi(|\beta|) \equiv J_{(a,w)}(|\beta|) = \log\Gamma(1/w) + \frac{\log w}{w} + \frac{\beta^2}{2au(|\beta|)} + \frac{(w-2)\log u(|\beta|)}{2w} + \frac{u(|\beta|)}{w},$$

where $u(|\beta|) = [\{8w\beta^2/a + (2-w)^2\}^{1/2} + 2 - w]/4$.

A good penalty function should produce estimates that satisfy the oracle properties (ie, unbiasedness, sparsity, and continuity).[9] LASSO is the most common penalty, but it does not satisfy the oracle properties. Moreover, the LASSO has been criticized on the grounds that it typically ends up selecting a model with too many variables to prevent over-shrinkage of the regression coefficients.[25] However, SCAD and HL satisfy the oracle properties and these can perform well as the oracle procedure in terms of selecting the correct subset models and estimating the true nonzero coefficients at the same time.[9,18] The HL function changes its shape according to the value of w; it becomes a ridge penalty when $w = 0$ and becomes a LASSO penalty when $w = 2$. When $w > 2$, it becomes an unbounded form at the origin.[18] In this article, we use $w = 30$ in the HL penalty as suggested by Lee and Oh.[18]

SCAD provides ordinary least square (OLS) estimates, whereas LASSO and HL give shrinkage estimates which have better prediction error. However, LASSO estimate is often over-shrunken. SCAD and HL give oracle estimates which have better variable selection. By controlling sparsity and shrinkage simultaneously, the HL has a higher probability of choosing the true model than the LASSO and SCAD methods without losing the prediction accuracy.[11,26] In change point problems, the HL method provides a consistent estimation of the number of change points and their locations and sizes, whereas conventional methods such as LASSO and SCAD may not satisfy such property.[27]

3.1 | Estimation procedure for variable selection

For estimating the fixed and random effects (β, v), we use penalized maximum h-likelihood estimators (PMHLEs) which maximize h_v in (8). Given dispersion parameters $\theta = (\gamma, \alpha)^T$, the PMHL estimating equations for $\beta = (\beta_1^T, \beta_2^T)^T$ and v are as follows:

$$\frac{\partial h_v}{\partial \beta} = \frac{\partial h_p}{\partial \beta} - n\sum_{j=1}^{p^*} [J_\xi(|\beta_j|)]' = 0 \tag{9}$$

and

$$\frac{\partial h_v}{\partial v} = \frac{\partial h_p}{\partial v} = 0. \tag{10}$$

Note that (9) is an adjusted estimating equation induced by adding the penalty term, whereas (10) is the same as the standard estimating equation without penalty. Here, the kth $(k = 1, 2)$ component of $\partial h_p/\partial \beta$ and $\partial h_p/\partial v$ are, respectively, given by

$$\frac{\partial h_p}{\partial \beta_k} = X^T(\delta_k - \hat{\mu}_k)$$

and

$$\frac{\partial h_p}{\partial v} = Z^T(\delta_1 - \hat{\mu}_1) + \gamma Z^T(\delta_2 - \hat{\mu}_2) + \frac{\partial \ell_2}{\partial v},$$

where $\partial \ell_2/\partial v = -\alpha^{-1}v$ for the log-normal frailty and $\partial \ell_2/\partial v = \alpha^{-1} - \alpha^{-1}\exp(v)$ for the gamma frailty model, X and Z are $n \times p$ and $n \times q$ model matrices for β and v whose ijth row vectors are x_{ij}^T and z_{ij}^T, respectively, $z_{ij} = (z_{ij1}, z_{ij2}, \ldots, z_{ijq})^T$ is a $q \times 1$ cluster indicator vector such that $z_{ijm} = 1$ if $i = m$ and 0 otherwise, and δ_k is an $n \times 1$ Type k event indicator vector with ijth element δ_{ijk}, Here $\hat{\mu}_k = \hat{\Lambda}_{0k}\exp(\eta_k)$ with $\hat{\Lambda}_{0k}(t) = \sum_{r:\ y_{(kr)} \le t} \hat{\lambda}_{0kr}$ $(k = 1, 2)$, and $\eta_1 = X\beta_1 + Zv$ and $\eta_2 = X\beta_2 + \gamma Zv$.

However, solving directly the estimating equations of β in (9) is difficult because the penalty function $J_\xi(\cdot)$ becomes non-differentiable at the origin and they do not have continuous second-order derivatives. Thus, we use a local quadratic approximation (LQA) for such penalty functions.[9] Given an initial value $\beta^{(0)}$ that is close to the true value of β, the penalty function $J_\xi(\cdot)$ can be locally approximated by a quadratic function as

$$[J_\xi(|\beta_j|)]' = J'_\xi(|\beta_j|)\mathrm{sgn}(\beta_j) \approx \{J'_\xi(|\beta_j^{(0)}|)/|\beta_j^{(0)}|\}\beta_j \quad \text{for } \beta_j \approx \beta_j^{(0)}. \tag{11}$$

Now, to obtain the penalized estimating equations of (β, v) based on the LQA, let us define $\mathbf{X}$, $\mathbf{Z}$, and $\mathbf{W}^*$ as the partition matrices such that

$$\mathbf{X} = \begin{pmatrix} X & \mathbf{0} \\ \mathbf{0} & X \end{pmatrix}, \quad \mathbf{Z} = \begin{pmatrix} Z \\ \gamma Z \end{pmatrix}, \quad \text{and} \quad \mathbf{W}^* = \begin{pmatrix} W_1^* & \mathbf{0} \\ \mathbf{0} & W_2^* \end{pmatrix},$$

where $W_k^* = -\partial^2 h_p/\partial \eta_k \partial \eta_k^T$ $(k = 1, 2)$.[12] Then, following Ha et al,[12,14] we can show that the PMHLEs for (β, v) are obtained from the iterative least squares (ILS) score equations:

$$\begin{pmatrix} \mathbf{X}^T\mathbf{W}^*\mathbf{X} + n\Sigma_\xi & \mathbf{X}^T\mathbf{W}^*\mathbf{Z} \\ \mathbf{Z}^T\mathbf{W}^*\mathbf{X} & \mathbf{Z}^T\mathbf{W}^*\mathbf{Z} + Q \end{pmatrix} \begin{pmatrix} \hat{\beta} \\ \hat{v} \end{pmatrix} = \begin{pmatrix} \mathbf{X}^T\mathbf{w}^* \\ \mathbf{Z}^T\mathbf{w}^* \end{pmatrix}, \tag{12}$$

where $\Sigma_\xi = \text{diag}\{J'_\xi(|\beta_j|)/|\beta_j|\}, \mathbf{w}^* = (w_1^{*T}, w_2^{*T})^T$ with $w_k^* = W_k^*\eta_k + (\delta_k - \mu_k)$, and $Q = \text{diag}(-\partial^2\ell_2/\partial v^2)$ is a $q \times q$ diagonal matrix; $Q = \alpha^{-1}I_q$ with $q \times q$ identity matrix I_q for the log-normal frailty and $Q = \text{diag}(\alpha^{-1}u_i)$ for the gamma frailty.

The score equations (12) are the extensions of the existing estimation procedures. For example, under no penalty (ie, $\Sigma_\xi = 0$) they become the score equations of Christian et al[5] for the standard cause-specific frailty models. For variable selection under the cause-specific hazard model without frailty term v, they reduce to

$$(\mathbf{X}^T\mathbf{W}^*\mathbf{X} + n\Sigma_\xi)\hat{\beta} = \mathbf{X}^T\mathbf{w}^*, \tag{13}$$

implying that the new equations(12) give a special case of the penalized equation (13) for the cause-specific hazard model without frailty term; the resulting estimates $\hat{\beta}$ are the same as those by Hou et al[17] under the LASSO and adoptive LASSO penalties.

The dispersion parameters $\theta = (\gamma, \alpha)^T$ are estimated by maximizing the partial restricted h-likelihood $p_{\beta,v}(h_p)$[14] based on h_p because θ does not depend on the penalty term in h_v of (8); it is given by a function of θ

$$p_{\beta,v}(h_p) = \left[h_p - \frac{1}{2}\log(\det(H_p/2\pi))\right]\Big|_{(\beta,v)=(\hat{\beta}(\theta),\hat{v}(\theta))}, \tag{14}$$

where $H_p = -\partial^2 h_p/\partial(\beta, v)^2 = H_v - \text{blockdiag}(n\Sigma_\xi, 0)$ and $H_v = -\partial^2 h_v/\partial(\beta, v)^2$ is given by the square matrix of left-hand corner of (12). Using the properties of determinations, Equation (14) can be expressed as

$$p_{\beta,v}(h_p) = \hat{h}_p - \frac{1}{2}\log\{\det(\hat{H}_p)\} + \frac{(2p+q)}{2}\log(2\pi), \tag{15}$$

where $\hat{h}_p = h_p(\hat{\beta}(\theta), \hat{v}(\theta))$ is a profile h-likelihood of θ and $\hat{H}_p = H_p(\hat{\beta}(\theta), \hat{v}(\theta))$. Then the estimates of θ are obtained by solving the score equation

$$\partial p_{\beta,v}(h_p)/\partial\theta = 0. \tag{16}$$

Note that this procedure works well for the log-normal frailty, but not for non-lognormal frailty such as the gamma frailty.[7,11] Thus, we use the second-order approximation under gamma frailty,[7,11] given by

$$s_{\beta,v}(h_p) = p_{\beta,v}(h_p) - \{F(h)/24\}, \tag{17}$$

where the detailed form of $F(h)$ is given in Appendix A of the Supplementary Materials.

Accordingly, we see that the proposed procedure is easily implemented via a slight modification to the existing HL procedures for standard frailty models.[12]

3.2 | Standard error and selection of tuning parameter

Following Ha et al,[12,14] an approximated standard error (SE) of $\hat{\beta}$ is obtained from a sandwich formula based on h_p:

$$\text{cov}(\hat{\beta}) = (H_{\beta\beta} + n\Sigma_\xi)^{-1}H_{\beta\beta}(H_{\beta\beta} + n\Sigma_\xi)^{-1}, \tag{18}$$

where $H_{\beta\beta} = \{(X^TW^*X) - (X^TW^*Z)(Z^TW^*Z + Q)^{-1}(Z^TW^*X)\}|_{v=\hat{v}}$. For the choice of tuning parameter ξ, we use a BIC-type criterion based on the penalized HL as in Ha et al,[12] defined by

$$\text{BIC}(\xi) = -2p_v(h_p) + e(\xi)\log(n), \tag{19}$$

where $p_v(h_p)$ is the first-order Laplace approximation to the marginal partial likelihood[12] $m_p(\beta, \theta) = \log\{\int \exp(h_p)dv\}$, and $e(\xi) = \text{tr}[\{H_{\beta\beta} + n\Sigma_\xi\}^{-1}H_{\beta\beta}]$ is the effective number of parameters.[10,11]

Our variable-selection algorithm consists of two loops as follows.

1. Inner loop: We maximize h_v for $\tau = (\beta^T, v^T)^T$ (ie, we solve the ILS equations in (12) for (β, v)) and $p_\tau(h_p)$ in (15) for θ under log-normal frailty, respectively. Here, we maximize $s_\tau(h_p)$ in (17) for θ under gamma frailty.
2. Outer loop: We select ξ that minimizes BIC(ξ) in (19).

Note that $\hat{\xi} = \arg\min_\xi \text{BIC}(\xi)$ is calculated using a simple grid search method. A good initial value is essential to obtain a proper estimate $\hat{\beta}$. Here, we use the solutions under no-penalty (ie, $\Sigma_\gamma = 0$)) for the initial value of LASSO penalty and use the LASSO solution as the initial value for the SCAD and HL penalties.[12,14] After convergence, we compute the estimates of the SEs for $\hat{\beta}$ using (18).

4 | SIMULATION STUDY

We conduct numerical studies, based upon 100 replications of simulated data as in Fan and Li[28] to evaluate the performance of the proposed procedures for the cause-specific frailty models. Here, for the model fitting we first use the shared log-normal frailty. We compare performances of the three variable-selection methods with LASSO, SCAD, and HL penalties.

Following the simulation schemes of Fan and Li,[28] data for the cause-specific hazard frailty model with two types are generated using a technique similar to Beyersmann et al[29] and Christian et al[5] Here, the conditional cause-specific hazard rates for each event type are,

$$\lambda_{ij1}(t|x_{ij}, v_i) = \lambda_{01}(t)\exp(x_{ij}^T\beta_1 + v_i),$$
$$\lambda_{ij2}(t|x_{ij}, v_i) = \lambda_{02}(t)\exp((x_{ij}^T\beta_2 + \gamma v_i).$$

Included in the model above were eight covariates $x_{ij} = (x_{ij1}, \ldots, x_{ij8})^T$ and a shared Normal random effect v_i with mean 0 and variance $\alpha = 0.5$. Here, the eight covariates x_{ij} were generated with an AR(1) structure with a correlation coefficient $\rho = 0.5$. The true regression coefficients for the Type 1 events were set to $\beta_1 = (0.8, 0, 0, 1, 0, 0, 0.6, 0)^T$ and $\beta_2 = -\beta_1$ for the Type 2 events. The association parameter was set to $\gamma = 1$ or -1 as in Huang and Wolfe.[20] Here, we consider two cases for the baseline hazard functions $\lambda_{01}(t)$ and $\lambda_{02}(t)$ based on exponential and Weibull distributions as follows:

$$\text{Case 1}: \ \lambda_{01}(t) = 0.5 \text{ and } \lambda_{02}(t) = 2,$$
$$\text{Case 2}: \ \lambda_{01}(t) = 0.5 \text{ and } \lambda_{02}(t) = 4t.$$

Censoring times are generated from a Uniform(0, c) distribution where the value of c was empirically selected to achieve the approximate right censoring rate 20%. With 20% censoring, the proportions of Type 1 and Type 2 events are about 30% and 50%, respectively. Samples sizes of $n = 250$ and $n = 500$, where $(q, n_i) = (50, 5)$ and $(100, 5)$, are considered. We also added a sample size with $n = 400$, with a larger cluster size having $(q, n_i) = (20, 20)$.

Following Fan and Li[28] and Ha et al[12] the model error (ME) for the cause-specific hazard frailty model is defined by

$$\text{ME}(\hat{\beta}_k) = E\{\exp(-x^T\hat{\beta}_k) - \exp(-x^T\beta_k)\}^2$$

for Type $k = 1, 2$. For the criteria for variable selection, we report the average number of zero coefficients, the probability of choosing the true model (PT), and ME. Let MRME stand for the median of ratios of ME of a selected model to that of the standard estimate under the full model. For model fitting and computation SAS/IML was used. The simulation results under Case 1 are summarized in Table 1. Here, the column labeled "C" (5 is the best) indicates the average number of regression coefficients in each type, of the five true zeros, correctly found to zero, and "IC" (0 is the best) indicates the average number of the three true nonzeros incorrectly found to zero.

From Table 1 one can notice that the SCAD and HL overall perform quite well and they both outperform the LASSO in terms of "C," "PT," and MRME. Both the SCAD and HL methods can be further improved with an increase of the size of q or n_i. In particular, the HL overall outperforms the SCAD in terms of "C," but it consistently outperforms the SCAD in

TABLE 1 Case 1: Simulation results using 100 replications under the cause-specific hazard log-normal frailty model (Type 1 = 30%, Type 2 = 50%, and Censoring = 20%)

		Type 1				Type 2			
(q, n_i)	Method	C	IC	PT	MRME	C	IC	PT	MRME
$\gamma = 1$									
(50, 5)	LASSO	3.31	0	0.14	1.261	3.87	0	0.31	2.154
	SCAD	4.63	0.01	0.70	0.510	4.54	0	0.60	0.654
	HL	4.65	0.03	0.69	0.442	4.75	0	0.76	0.480
(100, 5)	LASSO	3.40	0	0.12	2.449	3.70	0	0.22	2.482
	SCAD	4.80	0	0.82	0.719	4.66	0	0.69	0.610
	HL	4.84	0	0.86	0.282	4.80	0	0.82	0.553
(20, 20)	LASSO	3.36	0	0.14	2.028	3.73	0	0.26	2.238
	SCAD	4.76	0	0.78	0.644	4.63	0	0.70	0.708
	HL	4.75	0	0.77	0.535	4.77	0	0.77	0.508
$\gamma = -1$									
(50, 5)	LASSO	3.20	0	0.11	0.304	3.78	0	0.25	2.045
	SCAD	4.71	0.01	0.75	0.573	4.52	0	0.65	0.547
	HL	4.71	0.01	0.74	0.249	4.76	0	0.80	0.593
(100, 5)	LASSO	3.41	0	0.17	1.776	3.69	0	0.23	2.212
	SCAD	4.78	0	0.79	0.536	4.61	0	0.65	0.617
	HL	4.79	0	0.81	0.313	4.80	0	0.81	0.519
(20, 20)	LASSO	3.56	0	0.17	1.818	3.57	0	0.20	2.489
	SCAD	4.68	0	0.72	0.730	4.49	0	0.60	0.660
	HL	4.71	0	0.74	0.366	4.67	0	0.71	0.649

Note: q, no. of clusters; n_i, cluster size; HL, h-likelihood penalty function; C, average number of coefficients, of the five true zeros, correctly set to zero; IC, average number of the three true nonzero incorrectly set to zero; PT, probability of choosing the true model; MRME, median of relative model errors.

terms of MRME, providing a higher probability (PT) of choosing the true model than LASSO and SCAD methods without losing prediction accuracy. In addition, from 100 replications of simulated data, we also computed the mean of nonzero coefficients of $\hat{\beta}$, their standard deviation (SD), and standard error (SE) which is obtained from the sandwich formula (18). Note here that the SE is the average of 100 estimated standard errors for $\hat{\beta}$ and that the SD is the estimates of the true $\{\text{var}(\hat{\beta})\}^{1/2}$. Though not reported here, we have observed that the bias of the SCAD estimates is the smallest compared to the LASSO and HL, but the HL estimates are improved with q or n_i, and that the SEs in the SCAD and HL substantially improve in that a discrepancy between SE and SD decreases when q or n_i increases; these results confirm the simulation results of Ha et al[12,14]

The simulation results under Case 2 are given in Table 2. We find that the trends of results from Table 2 are similar to those evident in Table 1 under Case 1, and that the HL method still works well as compared to the LASSO and SCAD methods.

Next, with the gamma frailty u_i with mean 1 and variance $\alpha = 0.5$, we have also conducted the same simulation above under Case 2. The results are given in Table S1 of the Supplementary Materials; we again find that their trends are similar to those evident in Tables 1 and 2 under the log-normal frailty. For the frailty models, several authors have shown that misspecifying the random-effect (frailty) distribution such as log-normal or gamma frailty has little effect on the fixed effect estimates (ie, estimated regression parameters), not the frailty variance estimates.[30-33] Moreover, the results from Tables 1 and 2 confirm the simulation results of Park et al[34] that the penalized variable selection procedure of fixed effects in the accelerated failure time models with random effects is not sensitive to the choice of a particular random effect distribution.

TABLE 2 Case 2: Simulation results using 100 replications under the cause-specific hazard log-normal frailty model (Type 1 = 30%, Type 2 = 50%, and Censoring = 20%)

		Type 1				Type 2			
(q, n_i)	Method	C	IC	PT	MRME	C	IC	PT	MRME
$\gamma = 1$									
(50, 5)	LASSO	3.27	0	0.07	1.376	4.00	0	0.31	0.638
	SCAD	4.74	0	0.75	0.685	4.65	0	0.69	0.573
	HL	4.78	0	0.80	0.543	4.78	0	0.80	0.375
(100, 5)	LASSO	3.39	0	0.15	1.052	3.86	0	0.27	1.026
	SCAD	4.73	0	0.75	0.621	4.62	0	0.71	0.690
	HL	4.82	0	0.84	0.476	4.85	0	0.86	0.598
(20, 20)	LASSO	3.11	0	0.09	1.256	3.82	0	0.27	1.119
	SCAD	4.73	0	0.75	0.913	4.60	0	0.68	0.779
	HL	4.69	0	0.76	0.684	4.70	0	0.72	0.528
$\gamma = -1$									
(50, 5)	LASSO	3.23	0	0.12	1.568	3.89	0	0.33	1.581
	SCAD	4.66	0	0.70	0.724	4.53	0	0.63	0.512
	HL	4.81	0.01	0.82	0.359	4.73	0	0.77	0.468
(100, 5)	LASSO	3.43	0	0.15	1.738	3.82	0	0.24	1.466
	SCAD	4.76	0	0.79	0.537	4.50	0	0.62	0.638
	HL	4.84	0	0.84	0.635	4.74	0	0.78	0.569
(20, 20)	LASSO	3.48	0	0.20	1.912	3.98	0	0.30	1.369
	SCAD	4.79	0	0.80	0.833	4.65	0	0.70	0.623
	HL	4.87	0	0.87	0.613	4.74	0	0.76	0.395

Note: q, no. of clusters; n_i, cluster size; HL, h-likelihood penalty function; C, average number of coefficients, of the five true zeros, correctly set to zero; IC, average number of the three true nonzero incorrectly set to zero; PT, probability of choosing the true model; MRME, median of relative model errors.

5 | ILLUSTRATION

For an illustration of the proposed variable selection procedure, we consider two kinds of competing-risks clinical example, that is, two event types and three event types.

5.1 | Bladder cancer data: Two types of events

We consider a multicenter clinical dataset from a bladder cancer trial conducted by the European Organization for Research and Treatment of Cancer (EORTC).[35] We use the subset of the bladder cancer data as considered in Section 1.2.4 of Ha et al[7] The dataset consists of 396 bladder cancer patients from 21 centers, where the numbers of patients per center varied from 3 to 78, with mean 18.9 and median 14. Here we consider two competing endpoints, time to the first bladder recurrence (an event of interest; Type 1 event) and time to the death prior to the recurrence (competing event; Type 2 event). Of 396 patients, 200 patients (50.51%) had recurrence of bladder cancer and 81 patients (20.45%) died prior to the recurrence. One hundred and fifteen patients (29.04%) who were still alive without recurrence were censored at the date of the last available follow-up.

We consider the following 12 categorical covariates (x) of interest:

- Chemotherapy as the main covariate (CHEMO; no = 0, yes = 1),
- Age (0 if Age ≤65 years, 1 if Age >65 years),

- Sex (male = 0, female = 1),
- Prior recurrent rate (PRIORREC; primary, ≤1/yr, >1/yr);
  PRIORREC1 = I(PRIORREC ≤1/yr), PRIORREC2 = I(PRIORREC >1/yr)
- Number of tumors (NOTUM; single, 2-7 tumors, ≥8 tumors);
  NOTUM1 = I(NOTUM = 2-7 tumors), NOTUM2 = I(NOTUM ≥8 tumors);
- Tumor size (TUM3CM; 0 if Tumor size <3 cm, 1 if Tumor size ≥3 cm),
- T category (TLOCC; Ta = 0, T1 = 1),
- Carcinoma in situ (CIS; no = 0, yes = 1),
- G grade (GLOCAL; G1, G2, G3);
  GLOCAL1 = I(GLOCAL = G2), GLOCAL2 = I(GLOCAL = G3).

We fitted the cause-specific hazard models (2) and (3) with a shared log-normal frailty structure, as in Section 4, using the penalized h-likelihood procedure presented in Section 3. The estimated coefficients and their standard errors for Type 1 (ie, bladder cancer recurrence) and Type 2 (ie, death prior to the recurrence) are summarized in Table 3, respectively. First, the estimation results under no penalty are as follows. The estimated association parameter $\hat{\gamma} = 0.406$ shows a positive correlation between the risks of these two events. Meaning that the increment on the risk of bladder cancer recurrence (Type 1 event) increases the risk of dying (Type 2 event). The estimated variance of the frailty is $\hat{\alpha} = 0.101$. For Type 1 event, the following t-values (=Estimate/SE), from six covariates ($x_1, x_5, x_6, x_7, x_{11}, x_{12}$) are significant. In particular, the main covariate, CHEMO (x_1), is the most significant. However for Type 2 event, only three covariates (x_2, x_6, x_7) are significant, but the main covariate, that is, CHEMO (x_1), is not significant. In particular, we find that the CHEMO significantly reduces the risk of recurrence, but there is not enough evidence to conclude that CHEMO increases the risk of death.

Next, we are interested in the important variable selection among 12 covariates under the cause-specific frailty model by using LASSO, SCAD, and HL methods. The selected values of the tuning parameters ξ based on BIC in (19) are 0.013, 0.083, and 0.011 for the LASSO, SCAD, and HL, respectively. The estimates of dispersion parameters $\theta = (\gamma, \alpha)$ for LASSO, SCAD, and HL are similar, with $(1.001, 0.061)$, $(0.785, 0.093)$, and $(0.863, 0.074)$, respectively. The estimated coefficients and their standard errors for Type 1 (ie, bladder cancer recurrence) under the three penalties are given in Table 3. The main covariate, CHEMO (x_1), is very significant in all the three variable selection methods (ie, LASSO, SCAD, and HL). The LASSO method chooses 9 covariates ($x_1, x_2, x_5, x_6, x_7, x_8, x_9, x_{11}, x_{12}$) among 12 covariates, while SCAD and HL choose the same 6 covariates (x_1, x_5, x_6, x_7, x_{11}, x_{12}). Notice here that LASSO chooses the three covariates (x_2, x_8 and x_9) which are not significant under no-penalty. For Type 2, the estimated coefficients and their standard errors are also presented in Table 3. Here, LASSO selects two covariates (x_2 and x_{11}), where x_{11} is not significant under no-penalty. Meanwhile, SCAD and HL choose only one covariate (x_2).

We observe that the nonzero estimates by the SCAD are overall similar to the corresponding estimates without penalty ($\xi = 0$). Note here that SCAD provides OLS estimates and that LASSO and HL give shrinkage estimates for nonzero regression coefficients; LASSO shrinks more than HL does. As expected by Ha et al[12,14] the LASSO selects more covariates as compared to the SCAD and HL. A possible reason may be that the LASSO selects unimportant variables much more than the SCAD and HL methods. These findings indicate that the LASSO may not properly identify important variables in the cause-specific hazard frailty model (1), as evident in the lower "C" value of LASSO in Tables 1 and 2. Accordingly, we see that SCAD and HL methods provide a parsimonious model compared to no-penalty and LASSO, indicating that the former can give better inferences including interpretation and prediction than the latter.[11]

In addition, we fitted the cause-specific hazard models with the gamma frailty. We have found that the estimation results of gamma model (in Table S2 of the Supplementary Materials) are similar to those in Table 3. Below we present a selection procedure between log-normal and gamma frailty models, using the conditional Akaike information criterion (cAIC),[7,36,37] given by

$$\text{cAIC} = -2\ell_p + 2\text{df}_c. \tag{20}$$

Here

$$\ell_p = \sum_{ijk} \delta_{ijk}\eta_{ijk} - \sum_{kr} d_{(kr)} \log\left\{ \sum_{ij \in R_{(kr)}} \exp(\eta_{ijk}) \right\}$$

TABLE 3 Estimated regression coefficients (standard errors) and estimated dispersion parameters in the cause-specific hazard log-normal frailty model for two types of events in the multi-center bladder cancer data

Event	Variable	No-penalty	LASSO	SCAD	HL
Type 1	x_1: CHEMO	−0.876 (0.187)	−0.598 (0.142)	−0.870 (0.182)	−0.696 (0.158)
	x_2: Age	−0.266 (0.147)	−0.131 (0.079)	0 (0)	0 (0)
	x_3: Sex	0.000 (0.210)	0 (0)	0 (0)	0 (0)
	x_4: PRIORREC1	0.289 (0.252)	0 (0	0 (0)	0 (0)
	x_5: PRIORREC2	0.534 (0.200)	0.336 (0.119)	0.426 (0.178)	0.337 (0.127)
	x_6: NOTUM1	0.688 (0.167)	0.455 (0.118)	0.671 (0.163)	0.514 (0.131)
	x_7: NOTUM2	1.217 (0.283)	0.693 (0.171)	1.213 (0.269)	0.863 (0.209)
	x_8: TUM3CM	0.152 (0.176)	0.002 (0.002)	0 (0)	0 (0)
	x_9: TLOCC	0.225 (0.173)	0.183 (0.091)	0 (0)	0 (0)
	x_{10}: CIS	0.246 (0.279)	0 (0)	0 (0)	0 (0)
	x_{11}: GLOCAL1	0.522 (0.166)	0.269 (0.103)	0.540 (0.159)	0.356 (0.119)
	x_{12}: GLOCAL2	0.801 (0.274)	0.276 (0.111)	0.914 (0.249)	0.552 (0.178)
Type 2	x_1: CHEMO	0.381 (0.391)	0 (0)	0 (0)	0 (0)
	x_2: Age	0.854 (0.285)	0.357 (0.131)	0.692 (0.263)	0.444 (0.164)
	x_3: Sex	−0.509 (0.356)	0 (0)	0 (0)	0 (0)
	x_4: PRIORREC1	0.153 (0.400)	0 (0)	0 (0)	0 (0)
	x_5: PRIORREC2	0.506 (0.323)	0 (0)	0 (0)	0 (0)
	x_6: NOTUM1	−0.523 (0.259)	0 (0)	0 (0)	0 (0)
	x_7: NOTUM2	−1.396 (0.552)	0 (0)	0 (0)	0 (0)
	x_8: TUM3CM	−0.139 (0.272)	0 (0)	0 (0)	0 (0)
	x_9: TLOCC	−0.174 (0.267)	0 (0)	0 (0)	0 (0)
	x_{10}: CIS	0.473 (0.497)	0 (0)	0 (0)	0 (0)
	x_{11}: GLOCAL1	0.273 (0.260)	0.010 (0.008)	0 (0)	0 (0)
	x_{12}: GLOCAL2	−0.135 (0.470)	0 (0)	0 (0)	0 (0)

is a conditional partial log-likelihood which is the first and second terms of the right-hand side of (7), and $\mathrm{df}_c = \mathrm{trace}(H_v^{-1}H_v^*)$ is an effective degree of freedom adjustment for estimating the fixed and random effects, computed by the Hessian matrix $H_v = -\partial^2 h_v/\partial(\beta, v)^2$ in (14) and $H_v^* = -\partial^2 \ell_p/\partial(\beta, v)^2$. Note that the smaller value of cAIC indicates a better model. With this dataset, under no-penalty, the log-normal frailty model gave cAIC= 2780.22 and the gamma frailty model had cAIC= 2780.51, leading that the cAIC selects the log-normal model even if cAIC of the log-normal is slightly smaller than that of the gamma model. The resulting cAIC values from three variable selection methods under the log-normal model are 2774.10, 2768.69, and 2771.71 for LASSO, SCAD, and HL, respectively. Thus, we observe that the values of cAIC are smaller in the models with penalty than in models without penalty, leading that the cAIC detects a parsimonious model as a better model. Hence, we may choose the log-normal frailty model with SCAD penalty as a final model parsimoniously, even if the SCAD and HL select the same variables as shown in Table 3.

Furthermore, we fitted the cause-specific hazard models to conduct the variable selection under the generalized gamma frailty[38] which includes the gamma, log-normal, and Weibull distributions as special cases.[38-40] The estimation procedure and the variable selection results using the bladder cancer data are given in Appendix B of the Supplementary Materials. We again found that the estimation results including cAIC (in Table S3 of the Supplementary Materials) are similar to the results of the gamma frailty in Table S2. Because cause-specific models under the three frailty distributions (in Tables 3, S2, and S3) give small frailty variances, the analysis results are similar. Thus, in this data set, the choice of frailty distribution is difficult.

TABLE 4 Estimated regression coefficients (standard errors) and estimated dispersion parameters in the subdistribution hazard log-normal frailty model for two types of events in the multi-center bladder cancer data

Event	Variable	No-penalty	LASSO	SCAD	HL
Type 1	x_1: CHEMO	−0.933 (0.187)	−0.666(0.166)	−0.929 (0.182)	−0.785 (0.174)
	x_2: Age	−0.343 (0.147)	−0.214 (0.120)	0 (0)	−0.218 (0.119)
	x_3: Sex	0.058 (0.208)	0 (0)	0 (0)	0 (0)
	x_4: PRIORREC1	0.276 (0.249)	0 (0	0 (0)	0 (0)
	x_5: PRIORREC2	0.514 (0.200)	0.327 (0.149)	0.395 (0.180)	0.294 (0.150)
	x_6: NOTUM1	0.713 (0.168)	0.494 (0.139)	0.688 (0.164)	0.593 (0.150)
	x_7: NOTUM2	1.307 (0.283)	0.816 (0.229)	1.293 (0.272)	1.051 (0.249)
	x_8: TUM3CM	0.213 (0.175)	0.060 (0.094)	0 (0)	0 (0)
	x_9: TLOCC	0.171 (0.173)	0.127 (0.115)	0 (0)	0 (0)
	x_{10}: CIS	0.266 (0.278)	0 (0)	0 (0)	0 (0)
	x_{11}: GLOCAL1	0.474 (0.165)	0.250 (0.126)	0.491 (0.159)	0.384 (0.137)
	x_{12}: GLOCAL2	0.808 (0.274)	0.347 (0.189)	0.910 (0.250)	0.610 (0.222)
Type 2	x_1: CHEMO	0.733 (0.378)	0.203 (0.184)	0 (0)	0 (0)
	x_2: Age	1.049 (0.262)	0.737 (0.213)	1.013 (0.258)	0.787 (0.220)
	x_3: Sex	−0.435 (0.340)	0 (0)	0 (0)	0 (0)
	x_4: PRIORREC1	0.014 (0.399)	0 (0)	0 (0)	0 (0)
	x_5: PRIORREC2	0.052 (0.321)	0 (0)	0 (0)	0 (0)
	x_6: NOTUM1	−0.771 (0.251)	−0.379 (0.184)	−0.581 (0.231)	−0.343 (0.177)
	x_7: NOTUM2	−1.140 (0.481)	−0.110 (0.151)	0 (0)	0 (0)
	x_8: TUM3CM	−0.165 (0.259)	0 (0)	0 (0)	0 (0)
	x_9: TLOCC	−0.141 (0.251)	0 (0)	0 (0)	0 (0)
	x_{10}: CIS	0.202 (0.439)	0 (0)	0 (0)	0 (0)
	x_{11}: GLOCAL1	−0.110 (0.243)	0 (0)	0 (0)	0 (0)
	x_{12}: GLOCAL2	−0.461 (0.466)	0 (0)	0 (0)	0 (0)

On the other hand, Ha et al[14] also used the same dataset for the variable selection, as shown in Table 4, for recurrence (Type 1) using the subdistribution hazard (subhazard) frailty model[6] with a shared log-normal frailty. The cause-specific hazard and subhazard models have different interpretations for covariate effects because both regression models are formalized under different types of hazard functions.[7,41,42] That is, the subhazard model directly associates covariate effects with the cumulative probability of a specific cause of events over time, that is, the cumulative incidence function (CIF), whereas the cause-specific hazard model associates the covariate effects with the cause-specific hazard function. In Table 4, we added the results of variable selection for death (Type 2) using the same subhazard frailty model. Here, we find that the trends of variable selection in the subhazard frailty model for Types 1 and 2 are overall similar to those in the cause-specific frailty model in Table 3. For example, in both models, the three methods (LASSO, SCAD, and HL) all select the CHEMO (ie, main covariate) for Type 1, and the two methods (SCAD and HL) do not it for Type 2. We also see that in cause-specific hazard frailty model, the CHEMO significantly reduces the cause-specific hazard of Type 1, but that in subhazard frailty model, it significantly reduces the subdistribution hazard of Type 1 (ie, lower CIF).

It is interesting to compare different regression models for competing risks outcomes. In this article, we compare the cause-specific hazard (CSH) and subhazard (SH) frailty models using an extended Brier score. The definition (C2) and estimator (C3) of the proposed Brier score are given in Appendix C of Supplementary Materials. Following Gerds and Schumacher,[43] we use two Brier score methods based on the Kaplan-Meier (KM) estimator and Cox's PH model for estimating the censoring distribution: see Appendix C for more detailed explanations. Table 5 shows the results based on the integrated Brier score (IBS) in Equation (C7) of Appendix C. Note that the smaller value of IBS indicates a better

TABLE 5 The results of an extended IBS (integrated Brier score) for model selection of the cause-specific hazard (CSH) and subdistribution hazard (SH) log-normal frailty models for two types of events in the multi-center bladder cancer data

		No-penalty		LASSO		SCAD		HL	
Censoring	**Event**	**CSH**	**SH**	**CSH**	**SH**	**CSH**	**SH**	**CSH**	**SH**
KM	Type 1	6.966	7.079	7.633	6.972	7.198	6.922	7.579	6.927
	Type 2	3.015	3.590	2.995	3.555	2.991	3.587	2.991	3.560
Cox	Type 1	7.601	7.896	8.414	7.640	7.172	6.895	7.556	7.681
	Type 2	2.832	3.702	2.931	3.584	2.904	3.593	2.928	3.552

Note: Censoring, the estimation method of censoring distribution; KM and Cox indicate the IBS based on the Kaplan-Meier estimator and Cox's PH model, respectively.

model. The IBS based on the KM method under no-penalty selects the CSH model for both Types 1 and 2 outcomes, whereas IBSs under LASSO, SCAD, and HL penalties select the SH model for Type 1 outcome and the CSH model for Type 2 outcome. The IBSs based on the Cox PH model under no-penalty and HL penalty select the CSH model for Types 1 and 2 outcomes, whereas those under LASSO, and SCAD penalties select the SH model for Type 1 outcome and the CSH model for Type 2 outcome. We prefer to use the CSH model for both Types 1 and 2 outcomes based on IBS under the HL penalty. However, further research on the model selection would be of interest.

5.2 | Breast cancer data: Three types of events

We consider a breast cancer data set (B-14) from a multicenter clinical trial, conducted by the National Surgical Adjuvant Breast and Bowel Project (NSABP).[44,45] The aim of this study is to investigate the effect of tamoxifen against placebo following surgery in patients who had negative axillary lymph nodes and estrogen receptor positive breast cancer. We use a high risk subset of patients from the B-14 study, with tumor size greater than 2.5 cm as in the analysis of Christian et al.[5] In this subset, there were 731 women with follow-up (371 placebo and 360 tamoxifen) who were eligible for the study. The median age for women on either placebo or treatment was 55 years. A series of multiple types of treatment failure were possible; local, regional, or distant recurrence of the original cancer as well as a new second primary cancer or death because patients were followed as long as they did not withdraw their consents.

Thus, we consider the three types of failures:

- Type 1: A local or regional recurrence,
- Type 2: A new second primary cancer in the contralateral breast,
- Type 3: A distant recurrence, other new second primary cancer or death.

Among all 840 observations including multiple observations from 731 patients, the number of Types 1, 2, and 3 events was 113 (13.45%), 64 (7.62%), and 388 (46.19%), respectively, and the number of no events (censoring) until the last follow-up was 275 (32.74%). Here, about 57% of the 95 patients who had multiple events experienced both Type 1 and Type 3 events and about 20% had Type 2 and Type 3 events.

We assume that these three types of events compete against each other because once a recurrence or second primary event occurs, non-protocol therapies are often administered after the event, which would prohibit an accurate assessment of the effect of the treatment solely on that particular event type under consideration. Here, the interest of covariates were age (x_1) and treatment (x_2 is 1 for tamoxifen and 0 for placebo) which is a main covariate in this trial.

We also consider a cause-specific hazard frailty model with age and treatment under the three events above. For this purpose, the cause-specific frailty model with (2) and (3) is extended to a joint model with a common log-frailty v_i, composed of three interlinked cause-specific submodels. That is, the three event times T_{ij1}, T_{ij2}, and T_{ij3}, respectively, follow the conditional cause-specific PH model:

$$\text{(i)}\ \lambda_{ij1}(t|v_i) = \lambda_{01}(t)\exp(x_{ij}^T\beta_1 + v_i),$$

TABLE 6 Estimated regression coefficients (standard errors) and estimated dispersion parameters in the cause-specific hazard log-normal frailty model for three types of events in the breast cancer data

Event	Variable	No-penalty	LASSO	SCAD	HL
Type 1	x_1: Age	−0.015 (0.010)	−0.015 (0.010)	−0.005 (0.003)	0 (0)
	x_2: Treatment	−0.744 (0.226)	−0.170 (0.059)	−0.593 (0.208)	−0.354 (0.123)
Type 2	x_1: Age	−0.002 (0.013)	0 (0)	0 (0)	0 (0)
	x_2: Treatment	−0.158 (0.266)	0 (0)	0 (0)	0 (0)
Type 3	x_1: Age	0.016 (0.008)	0.016 (0.008)	0.012 (0.005)	0.018 (0.006)
	x_2: Treatment	−0.289 (0.165)	0 (0)	0(0)	0 (0)

$$\text{(ii)}\ \lambda_{ij2}(t|v_i) = \lambda_{02}(t)\exp(x_{ij}^T\beta_2 + \gamma_1 v_i),$$
$$\text{(iii)}\ \lambda_{ij3}(t|v_i) = \lambda_{03}(t)\exp(x_{ij}^T\beta_3 + \gamma_2 v_i).$$

Here, γ_1 and γ_2 are real-valued dispersion parameters to represent associations among the three submodels above via a shared log-frailty v_i. That is, γ_1 [γ_2] represents association between submodels (i) and (ii) [(i) and (iii)], respectively.[7,20] For simplicity, we use log-normal frailty distribution for v_i, that is, $v_i \sim N(0, \alpha)$. Then the three cause-specific models above ($k = 1, 2, 3$) are straightforwardly fitted using the penalized h-likelihood procedure with (12) and (16), by extending the partition matrices in (12) to

$$\mathbf{X} = \begin{pmatrix} X & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & X & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & X \end{pmatrix}, \quad \mathbf{Z} = \begin{pmatrix} Z \\ \gamma_1 Z \\ \gamma_2 Z \end{pmatrix}, \quad \text{and} \quad \mathbf{W}^* = \begin{pmatrix} W_1^* & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & W_3^* & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & W_3^* \end{pmatrix}$$

with $\mathbf{w}^* = (w_1^{*T}, w_2^{*T}, w_3^{*T})^T$.

The estimated coefficients and their standard errors for the three event types are summarized in Table 6, respectively. First, the estimation results under no penalty are as follows. The two estimated association parameters $\hat{\gamma}_1 = 0.820$ [$\hat{\gamma}_2 = 1.169$] show a positive correlation between Type 1 and Type 2 [Type 1 and Type 3], respectively. The results indicate that patients who experienced a local or regional recurrence will also be at a greater risk for developing a second primary cancer in the contralateral breast as well as any of Type 3 events. The estimated variance of the frailty is $\hat{\alpha} = 1.954$, indicating a fairly heterogeneous group of patients. Among covariates of three events, x_2 in Type 1 and x_1 in Type 3 are significant by t-value (=Estimate/SE). In particular, we find that tamoxifen (x_2) significantly reduces only the risk of a local or regional recurrence (ie, Type 1) as compared to patients who receive placebo.

Next, we study variable selection by LASSO, SCAD, and HL methods. The selected values of the tuning parameters ξ based on BIC in (19) are 0.012, 0.160, and 0.004 for the LASSO, SCAD, and HL, respectively. The estimates of dispersion parameters $\theta = (\gamma_1, \gamma_2, \alpha)$ for LASSO, SCAD, and HL are similar, with (0.807, 1.162, 1.935), (0.809, 1.145, 1.871), and (0.797, 1.148, 1.951), respectively. The estimated coefficients and their standard errors for Types 1, 2, and 3 under the three penalties are also given in Table 6. The main covariate (tamoxifen; x_2) in Type 1 and age (x_1) in Type 3 are significant in all the three variable selection methods (ie, LASSO, SCAD, and HL). The LASSO and SCAD select age (x_1) in Type 1 which are not significant under no-penalty, while HL does not. Here, HL provides the most pasimonious model.

6 | DISCUSSION

We have shown that the penalized h-likelihood method is useful for selecting jointly important variables of fixed effects corresponding to two or more types of events in the cause-specific hazard frailty models. The proposed variable selection method can be easily implemented by a slight modification to the existing h-likelihood estimation procedures for standard frailty models. We have also demonstrated via simulation studies and two practical data sets that the proposed procedure with SCAD or HL penalties works well overall. In particular, we have found by the simulation that HL provides a higher probability of choosing the true model than LASSO and SCAD methods without losing prediction accuracy.

Accordingly, we recommend using the HL method for the variable selection in cause-specific competing risks frailty models,

The main advantage of the proposed method based on the h-likelihood avoids intractable integrations over frailties in computing marginal likelihood, with a nonparametric estimation for high-dimensional baseline hazards.[7,46] Thus, our method provides efficient procedure for variable selection in various types of frailty models including semi-competing risks frailty models[46] and joint models[7] for repeated measures and time-to-event data. Both cause-specific hazards frailty model and subhazard frailty model can be used for analyzing the clustered survival data. However, both modeling approaches are different even if they could empirically give similar results;[7,17,47] in particular, the former can take into account the correlation between events of interest and competing events via frailties, while the latter does not by assuming that the frailty effects on both types of events are independent.[5,7] Thus, the cause-specific frailty model would be more appropriate when a dependency between different types of events or informative censoring is present. The subhazard model is useful for direct statistical inference about the CIF of the particular event type of interest. According to the results of analysis of bladder cancer data, the variable selection for Type 1 (recurrence) is identical in the two models, but that for Type 2 (death) is slightly different. However, the two models have different interpretations for covariate effects because they are defined under different types of hazard functions. For example, in both models the three methods (LASSO, SCAD, and HL) all select the CHEMO (ie, main covariate) for Type 1. Here, in cause-specific hazard frailty model the CHEMO significantly reduces the cause-specific hazard of Type 1, whereas in subhazard frailty model it significantly reduces the subdistribution hazard of recurrence (ie, lower CIF). The study on model choice would be of interest for future research.

Even if the fixed effect estimates are somewhat insensitive to the choice of frailty distribution, commonly used log-normal or gamma frailty distribution may be simple. It would be useful to consider a wider class of frailty distributions when the frailty distribution is in suspicion; for instance, we can consider the generalized gamma distribution,[38] which contains the gamma, log-normal, and Weibull distributions as special cases. For the selection of frailty distributions, we used the cAIC which is an interesting future research for further justification, including extended cause-specific hazard frailty models as in Section 5.2.

When there are two causes ($K = 2$) in the competing-risks setting of Section 2, the proposed variable selection framework is similar to the variable selection framework of joint frailty model for recurrent events and a terminal event by Han et al[48] in that both procedures provide the variable selection in the joint model with a shared frailty term for two different types of events. Note that our method uses a nonparametric baseline hazard, whereas Han et al's[48] method uses piecewise constant baseline hazard. Extensions of variable selection to cause-specific models with correlated frailties[4,5] or high dimensional case[17,24] having $p > n$ would be also an interesting further work.

ACKNOWLEDGEMENTS

The research of Il Do Ha was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (No. NRF-2017R1E1A1A03070747). The research of Youngjo Lee was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2019R1A2C1002408).

DATA AVAILABILITY STATEMENT

The first data set in Section 5.1 is available in data (bladder) of the **frailtyHL** R Package and also at Reference 7. The second data set in Section 5.2 is not publicly available due to confidentiality restrictions.

ORCID

Il Do Ha https://orcid.org/0000-0001-8434-9506
Youngjo Lee https://orcid.org/0000-0001-9820-6434

REFERENCES

1. Prentice R, Kalbfleisch J-D, Peterson A-V, Flournoy N, Farewell V-T, Breslow N-E. The analysis of failure times in the presence of competing risks. *Biometrics*. 1978;34(4):541-554.
2. Fine J-P, Gray R-J. A proportional hazards model for the subdistribution of a competing risk. *J Am Stat Assoc*. 1999;94(446):496-509.
3. Katsahian S, Boudreau C. Estimating and testing for center effects in competing risks. *Stat Med*. 2011;30(13):1608-1617.
4. Gorfine M, Hsu L. Frailty-based competing risks model for multivariate survival data. *Biometrics*. 2011;67(2):415-426.
5. Christian N-J, Ha I-D, Jeong J-H. Hierarchical likelihood inference on clustered competing risks data. *Stat Med*. 2016;35(2):251-267.
6. Ha I-D, Christian N-J, Jeong J-H, Park J, Lee Y. Analysis of clustered competing risks data using subdistribution hazard models with multivariate frailties. *Stat Methods Med Res*. 2016;25(6):2488-2505.

7. Ha I-D, Jeong J-H, Lee Y. *Statistical Modelling of Survival Data with Random Effects: H-Likelihood Approach*. Singapore: Springer; 2017.
8. Lai X, Yau K-K-W, Liu L. Competing risk model with bivariate random effects for clustered survival data. *Comput Stat Data Anal*. 2017;112:215-223.
9. Fan J, Li R. Variable selection via nonconcave penalized likelihood and its oracle properties. *J Am Stat Assoc*. 2001;96(456):1348-1360.
10. Lee Y, Nelder J-A. Hierarchical generalized linear models (with discussion). *J Roy Stat Soc B*. 1996;58(4):619-678.
11. Lee Y, Nelder J-A, Pawitan Y. *Generalised Linear Models with Random Effects: Unified Analysis Via h-Likelihood*. 2nd ed. London, UK: Chapman & Hall; 2017.
12. Ha I-D, Pan J, Oh S, Lee Y. Variable selection in general frailty models using penalized h-likelihood. *J Comput Graph Stat*. 2014;23(4):1044-1060.
13. Tibshirani R. Regression shrinkage and selection via the Lasso. *J Roy Stat Soc B*. 1996;58(1):267-288.
14. Ha I-D, Lee M, Oh S, Jeong J-H, Sylvester R, Lee Y. Variable selection in subdistribution hazard frailty models with competing risks data. *Stat Med*. 2014;33(26):4590-4604.
15. Fu Z, Parikh C-R, Zhou B. Penalized variable selection in competing risks regression. *Lifetime Data Anal*. 2017;23:353-376.
16. Ahn K-W, Banerjee A, Sahr N, Kim S. Group and within-group variable selection for competing risks data. *Lifetime Data Anal*. 2018;24:407-424.
17. Hou J, Paravati A, Hou J, Xu R, Murphy J. High-dimensional variable selection and prediction under competing risks with application to SEER-medicare linked data. *Stat Med*. 2018;37(24):3486-3502.
18. Lee Y, Oh H-S. A new sparse variable selection via random-effect model. *J Multivar Anal*. 2014;125:89-99.
19. Ha I-D, Lee Y, Song J-K. Hierarchical likelihood approach for frailty models. *Biometrika*. 2001;88(1):233-243.
20. Huang X, Wolfe R. A frailty model for informative censoring. *Biometrics*. 2002;58(3):510-520.
21. Liu L, Wolfe R-A, Huang X. Shared frailty models for recurrent events and a terminal event. *Biometrics*. 2004;60(3):747-756.
22. Breslow N-E. Discussion on Professor Cox's paper. *J Roy Stat Soc B*. 1972;34(2):216-217.
23. Wang H, Li R, Tsai C-L. uning parameter selectors for the smoothly clipped absolute deviation method. *Biometrika*. 2007;94(3):553-568.
24. Fan J, Lv J. A selective overview of variable selection in high dimensional feature space. *Stat Sin*. 2010;20(1):101-148.
25. Radchenko P, James G-M. Variable inclusion and shrinkage algorithms. *J Am Stat Assoc*. 2008;103(483):1304-1315.
26. Kwon S, Oh S, Lee Y. The use of random-effect models for high-dimensional variable selection models. *Comput Stat Data Anal*. 2016;103:401-412.
27. Ng T, Lee W, Lee Y. Change-point estimators with true identification property. *Bernoulli*. 2018;24(1):616-660.
28. Fan J, Li R. Variable selection for Cox's proportional hazards model and frailty model. *Ann Stat*. 2002;30(1):74-99.
29. Beyersmann J, Latouche A, Buchholz A, Schumacher M. Simulating competing risks data in survival analysis. *Stat Med*. 2009;28(6):956-971.
30. Pickles A, Crouchley R. A comparison of frailty models for multivariate survival data. *Stat Med*. 1995;14(13):1447-1461.
31. Ha I-D, Lee Y. Estimating frailty models via Poisson hierarchical generalized linearmodels. *J Comput Graph Stat*. 2003;12(3):663-681.
32. Ha I-D, Sylvester R, Legrand C, MacKenzie G. Frailty modelling for survival data from multi-centre clinical trials. *Stat Med*. 2011;30(17):2144-2159.
33. Gasparini A, Clements M-S, Abrams K-R, Crowther M-J. Impact of model misspecification in shared frailty survival models. *Stat Med*. 2019;38(23):4477-4502.
34. Park E, Kwon S, Kwon J, Sylvester R, Ha I-D. Penalized h-liklihood approach for variable selection in AFT random-effect models. *Stat Neerl*. 2020;74(1):52-71.
35. Sylvester R, van der Meijden APM, Oosterlinck W, et al. Predicting recurrence and progression in individual patients with stage Ta T1 bladder cancer using EORTC risk tables: a combined analysis of 2596 patients from seven EORTC trials. *Eur Urol*. 2006;49(3):466-477.
36. Vaida F, Blanchard S. Conditional Akaike information for mixed-effects models. *Biometrika*. 2005;92(2):351-370.
37. Donohue M, Overholser D, Xu R, Vaida F. Conditional Akaike information under generalized linear and proportional hazards mixed models. *Biometrika*. 2011;98(3):685-700.
38. Balakrishnan N, Peng Y. Generalized gamma frailty model. *Stat Med*. 2006;25(16):2797-2816.
39. Majakwara J, Pal S. On some inferential issues for the destructive COM-Poisson-generalized gamma regression cure rate model. *Commun Stat Simul Comput*. 2019;48(10):3118-3142.
40. Pal S, Yu H, Loucks Z-D, Harris I-M. Illustration of the exibility of generalized gamma distribution in modeling right censored survival data: analysis of two cancer datasets. *Ann Data Sci*. 2020;7:77-90.
41. Austin PC, Lee DS, Fine JP. Introduction to the analysis of survival data in the presence of competing risks. *Circulation*. 2016;133(6):601-609.
42. Austin PC, Fine JP. Practical recommendations for reporting Fine-Gray model analyses for competing risk data. *Stat Med*. 2017;36(27):4391-4400.
43. Gerds T, Schumacher M. Consistent estimation of the expected Brier score in general survival models with right-censored event times. *Biom J*. 2006;48(6):1029-1040.
44. Fisher B, Costantino J, Redmond C, et al. A randomized clinical trial evaluating tamoxifen in the treatment of patients with node-negative breast cancer who have estrogen receptor-positive tumors. *New Engl J Med*. 1989;320(8):479-484.
45. Fisher B, Dignam J, Bryant J, et al. Five versus more than five years of tamoxifen therapy for breast cancer patients with negative lymph nodes and estrogen receptor- positive tumors. *J Natl Cancer Inst*. 1996;88(21):1529-1542.

46. Ha I-D, Xiang L, Peng M, Jeong J-H, Lee Y. Frailty modelling approaches for semi-competing risks data. *Lifetime Data Anal*. 2020;26(1):109-133.
47. Geskus RB. *Data Analysis with Competing Risks and Intermediate States*. Taylor & Francis Group: Boca Raton, FL; 2016.
48. Han D, Su X, Sun L, Zhang Z, Liu L. Variable selection in joint frailty models of recurrent and terminal events. *Biometrics*. 2020;76(4):1330-1339.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Rakhmawati TW, Ha ID, Lee H, Lee Y. Penalized variable selection for cause-specific hazard frailty models with clustered competing-risks data. *Statistics in Medicine*. 2021;40(29):6541-6557. https://doi.org/10.1002/sim.9197