Published in final edited form as:

Stat Med. 2014 November 20; 33(26): 4590–4604. doi:10.1002/sim.6257.

Variable selection in subdistribution hazard frailty models with competing risks data

Il Do Ha^a, Minjung Lee^{b,*}, Seungyoung Oh^c, Jong-Hyeon Jeong^d, Richard Sylvester^e, and Youngjo Lee^c

^a Department of Data Management, Daegu Haany University, Gyeongsan, South Korea

^b Data Science for Knowledge Creation Research Center, Seoul National University, Seoul, South Korea

^c Department of Statistics, Seoul National University, Seoul, South Korea

^d Department of Biostatistics, University of Pittsburgh, Pittsburgh, PA 15261, U.S.A

^e European Organisation for Research and Treatment of Cancer, Brussels, Belgium

Abstract

The proportional subdistribution hazards model (i.e. Fine-Gray model) has been widely used for analyzing univariate competing risks data. Recently, this model has been extended to clustered competing risks data via frailty. To the best of our knowledge, however, there has been no literature on variable selection method for such competing risks frailty models. In this paper, we propose a simple but unified procedure via a penalized h-likelihood (HL) for variable selection of fixed effects in a general class of subdistribution hazard frailty models, in which random effects may be shared or correlated. We consider three penalty functions (LASSO, SCAD and HL) in our variable selection procedure. We show that the proposed method can be easily implemented using a slight modification to existing h-likelihood estimation approaches. Numerical studies demonstrate that the proposed procedure using the HL penalty performs well, providing a higher probability of choosing the true model than LASSO and SCAD methods without losing prediction accuracy. The usefulness of the new method is illustrated using two actual data sets from multi-center clinical trials.

Keywords

competing risks; frailty models; h-likelihood penalty function; penalized h-likelihood; subdistribution hazard; variable selection

1. Introduction

In regression analysis, a challenging task is to efficiently select relevant variables from a statistical model with a large number of covariates. Recently, variable selection methods

using a penalized likelihood have been widely studied in various statistical models such as linear models, generalized linear models and Cox's proportional hazards (PH) models. The main advantage of those methods is that they select important variables and estimate the regression coefficients of the covariates, simultaneously. Such methods, for example, include the least absolute shrinkage and selection operator (referred to as LASSO [1]), smoothly clipped absolute deviation (referred to as SCAD [2]), and adaptive-LASSO [3]; for extensive reviews, see [4, 5].

In the analysis of competing risks data, variable selection methods have not been widely studied. Recently, Kuk and Varadhan [6] proposed a stepwise selection approach based on AIC, BIC and a modified BIC in the Fine-Gray model [7] without frailty term. However, classical variable selection methods including stepwise selection can be computationally intensive for a large number of covariates and often suffer from high variability [2, 3, 8]. In this paper we develop a variable selection method for the subdistribution hazard (subhazard) frailty model, which is an extension of the Fine-Gray model incorporating random effects (frailty) terms [9, 10]. It is a challenging problem because of the complex nature of the model allowing for competing risks and unobserved frailty terms. In the absence of competing risks, Fan and Li [11] proposed the penalized marginal likelihood method using the SCAD penalty function for the gamma frailty model, and Androulakis et al. [12] recently extended it to other frailty distributions such as inverse Gaussian. It is well known that the log-normal frailties are useful, particularly for modeling multi-component [13] or correlated frailties [14]. However, the marginal likelihood function of such models involves analytically intractable integrals when eliminating the frailties. The hierarchical likelihood (h-likelihood [15]) obviates the need for marginalization over the frailty distribution and provides a statistically efficient procedure in various random-effect models such as hierarchical GLMs (HGLMs [15, 16]) and correlated frailty models [14, 17].

In this paper, we propose a simple but unified penalized h-likelihood method for variable selection of fixed effects in a general class of subhazard frailty models. We consider three penalty functions, LASSO, SCAD, and h-likelihood (referred to as HL [18]). The SCAD penalty provides good properties such as oracle property, while the HL penalty is unbounded at the origin [18] and gives a very good performance in various high dimensional problems [19, 20, 21]. Note that the SCAD penalty method leads to an oracle maximum likelihood (ML) estimator, whereas the HL penalty approach gives an oracle shrinkage estimator [18]. Here, an oracle ML estimator is the ML estimator when all covariates with nonzero coefficients are known. Similarly, an oracle shrinkage estimator is the shrinkage estimator when it is known which covariates have nonzero coefficients. Fan and Peng [22] showed that a local solution of the SCAD penalty is asymptotically equivalent to an oracle ML estimator. Furthermore, Kwon et al. [23] showed that a local solution for the HL penalty is an oracle shrinkage estimator. It is well known that shrinkage estimation would be preferred for prediction [24, 25, 26]. Simulation results in Section 4 demonstrate that the HL has higher probability of choosing the true model than the LASSO and SCAD methods without losing prediction accuracy.

We show that the proposed approach can be easily implemented via a slight modification to the existing h-likelihood estimation procedures [10, 17]. Through simulation studies, we

evaluate performances of the three variable-selection methods (LASSO, SCAD, and HL). The methods are illustrated using two actual time-to-event datasets from multi-center clinical trials. The paper is organized as follows. In Section 2 we review a general class of subhazard frailty models and outline the corresponding h-likelihood. In Section 3 we discuss various penalty functions and then show how the standard h-likelihood procedure for subhazard frailty models can be easily extended to accommodate variable selection. Simulation studies and practical examples are presented in Sections 4 and 5, respectively. Finally, a brief discussion is given in Section 6. The additional simulation results are provided in Supplemental Materials.

2. Subhazard frailty models and h-likelihood

2.1. A general class of subhazard frailty models

Suppose that the data consist of censored time-to-event observations collected from q clusters (or centers). We also assume that there are L distinct event types in each cluster. For a subject j in cluster i , let T_{ij} be the time to the first event and let $\varepsilon_{ij} \in \{1, 2, \dots, L\}$ be the corresponding cause of event ($i = 1, \dots, q, j = 1, \dots, n_i, n = \sum_i n_i$). Then observable random variables become $Y_{ij} = \min(T_{ij}, C_{ij})$ and $\xi_{ij} = I(T_{ij} < C_{ij})\varepsilon_{ij}$, where C_{ij} is the independent censoring time, $\xi_{ij} \in \{0, 1, 2, \dots, L\}$ and $I(\cdot)$ is the indicator function.

The hazard function of subdistribution (subhazard function) for cause 1 is defined by [7]

$$\begin{aligned}\lambda_1^s(t) &= \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \Pr\{t \leq T_{ij} \leq t + \Delta t, \varepsilon_{ij} = 1 | T_{ij} \geq t \cup (T_{ij} < t \cap \varepsilon_{ij} \neq 1)\} \\ &= -d \log \{1 - F_1(t)\} / dt,\end{aligned}$$

which is expressed as the cumulative incidence function (CIF) $F_1(t) = P(T_{ij} \leq t, \varepsilon_{ij} = 1)$, i.e. the probability that an individual will experience a type 1 event by time t . For simplicity, we consider two event types ($L = 1, 2$), so that ξ_{ij} takes 0, 1 or 2; 1 for an event of interest, 2 for a competing event, and 0 for censoring. Fine and Gray [7] proposed the proportional subdistribution hazards model to investigate directly the effects of covariates on the CIF for an event of interest, $L = 1$. Katsahian et al. [9], Katsahian and Boudreau [27] and Christian [28] have extended the Fine-Gray model to subhazard frailty models with only one random component (i.e. random center effect) to analyze multi-center competing risks data. Recently, Ha et al. [10] proposed a general class of subhazard frailty models allowing for two random components (i.e., random center and random treatment effects) via the h-likelihood approach.

Denote v_i be an r -dimensional vector of unobserved log-frailties (random effects) associated with the i th cluster. As described in Ha et al. [29], we assume that given v_i , $(T_{ij}, \varepsilon_{ij})$ and C_{ij} ($j = 1, \dots, n_i$) are conditionally independent, and that given v_i , C_{ij} ($j = 1, \dots, n_i$) are non-informative about v_i . Suppose that we are interested in assessing the effects of covariates on the conditional CIF for cause 1 given the frailties v_i , defined by $F_1(t|v_i) = \Pr(T_{ij} \leq t, \varepsilon_{ij} = 1 | v_i)$. Following Katsahian et al. [9] and Ha et al. [10], the conditional subhazard function for cause 1 given v_i is modeled as

$$\lambda_{ij1}^s(t|v_i) = \lambda_{01}^s(t) \exp(\eta_{ij}), \quad (1)$$

where $\lambda_{01}^s(\cdot)$ is the unknown baseline subhazard function, $\eta_{ij} = x_{ij}^T \beta + z_{ij}^T v_i$ is the linear predictor for the log-hazard, and $x_{ij} = (x_{ij1}, \dots, x_{ijp})^T$ and $z_{ij} = (z_{ij1}, \dots, z_{ijr})^T$ are $p \times 1$ and $r \times 1$ covariate vectors corresponding to fixed effects $\beta = (\beta_1, \dots, \beta_p)^T$ and log-frailties v_i , respectively. Here z_{ij} is often a subset of x_{ij} [30]. Although the results of this paper can be extended to non-normal frailties (e.g. gamma frailty), for simplicity, we assume a multivariate normal distribution, $v_i \sim N_r(0, \Sigma)$, which is useful for modelling multi-component frailties [13] including multilevel (nested) structures and/or correlated frailties including negative correlation [14, 17]. Here, the covariance matrix $\Sigma = \Sigma(\theta)$ depends on a vector of unknown frailty parameters θ .

Model (1) includes some well-known models as special cases. In a multicenter medical study, let v_{i0} be a random intercept or random center effect that modifies the baseline risk for center i , and let v_{i1} be associated with the treatment effect, i.e., a random treatment effect (or random treatment-by-center interaction). If we consider $z_{ij} = 1$ and $v_i = v_{i0}$ for all i, j , the model (1) becomes the random center or shared subhazard frailty model [9, 28] with

$$\eta_{ij} = x_{ij}^T \beta + v_{i0}, \quad (2)$$

where $v_{i0} \sim N(0, \sigma_0^2)$ for all i . Model (2) can be extended as follows. Let β_1 be the main treatment effect associated with the treatment indicator x_{ij1} and let β_m be the fixed effects corresponding to covariates x_{ijm} ($m = 2, \dots, p$). Our two random components leads to a bivariate subhazard model [10, 14, 17] with

$$\eta_{ij} = v_{i0} + (\beta_1 + v_{i1}) x_{ij1} + \sum_{m=2}^p \beta_m x_{ijm}, \quad (3)$$

which is easily obtained by taking $z_{ij} = (1, x_{ij1})^T$ and $v_i = (v_{i0}, v_{i1})$ in (1). Here, to maintain the invariance of the model to parametrization of the treatment effect we allow for a general covariance structure [14, 16] between v_{i0} and v_{i1} within a cluster:

$$\Sigma \equiv \begin{pmatrix} \sigma_0^2 & \sigma_{01} \\ \sigma_{01} & \sigma_1^2 \end{pmatrix}, \quad (4)$$

where the correlation is denoted by $\rho = \sigma_{01}/(\sigma_0 \sigma_1)$. The bivariate normal model (3) with (4) is very useful for investigating heterogeneity in the baseline risk and the treatment effect across centers.

2.2. H-likelihood construction

First we outline the h-likelihood approach for the complete data case under competing risks without independent censoring mechanism. Let $t_{(k)}$ be the k th ($k = 1, \dots, D$) smallest distinct time for type 1 events among the t_{ij} 's, where t_{ij} is the observed value of T_{ij} . Let $R_{0(k)}$ be the risk set at $t_{(k)}$ [7], defined by

$$\mathbf{R}_{0(k)} = R(t_{(k)}) = \left\{ (i, j) : t_{ij} \geq t_{(k)} \quad \text{or} \quad (t_{ij} \geq t_{(k)} \quad \text{and} \quad \epsilon_{ij} \neq 1) \right\}.$$

Note that as compared to the classical Cox model, the risk set $R_{0(k)}$ comprises individuals who have not failed from any cause by $t_{(k)}$ but also those who have previously failed from competing causes. Under the model (1), since the functional form of the baseline subhazard function $\lambda_{01}^s(t)$ is unknown, following Ha et al. [10, 29], we use the following profile h-likelihood h^* with λ_{01}^s eliminated:

$$h^* = \sum_{ij} \ell_{1ij}^* + \sum_i \ell_{2i}, \quad (5)$$

where

$$\sum_{ij} \ell_{1ij}^* = \sum_{ij} \mathbf{I}(\epsilon_{ij}=1) \eta_{ij} - \sum_k d_{0(k)} \log \left\{ \sum_{(i,j) \in \mathbf{R}_{0(k)}} \exp(\eta_{ij}) \right\},$$

$\ell_{1ij}^* = \ell_{1ij}(\beta, \lambda_{01}^s; t_{ij}, \epsilon_{ij} | v_i) |_{\lambda_{01}^s = \hat{\lambda}_{01}^s}$ is the logarithm of the conditional density function for (T_{ij}, ϵ_{ij}) given v_i evaluated at $\hat{\lambda}_{01}^s$ which is the nonparametric maximum h-likelihood estimator of λ_{01}^s [10]. Here $d_{0(k)}$ is the number of type 1 events at $t_{(k)}$, and

$$\ell_{2i} = \ell_{2i}(\theta; v_i) = -\frac{1}{2} [\log \det \{2\pi \Sigma_i(\theta)\}] - \frac{1}{2} v_i^T \Sigma_i(\theta)^{-1} v_i$$

is the logarithm of the density function for v_i with parameters $\theta = (\sigma_0^2, \sigma_1^2, \sigma_{01})^T$. Note that $\sum_{ij} \ell_{1ij}^*$ in h^* of (5) does not depend on the nuisances λ_{01}^* ; thus h^* becomes the penalized partial likelihood (PPL [31]). The first term $\sum_{ij} \ell_{1ij}^*$ in (5) can be viewed as the logarithm of the partial likelihood for the Fine-Gray model given v_i .

In the case of right censoring under competing risks, Fine and Gray [7] developed a weighted score function based on the complete-data partial likelihood and used the inverse probability of censoring weighting (IPCW) technique [32]. This technique can also be applied to the first term $\sum_{ij} \ell_{1ij}^*$ in (5) as in Pintilie [33], Katsahian et al. [9], and Ha et al. [10]. Notice here that we observe $Y_{ij} = \min(T_{ij}, C_{ij})$ and $\xi_{ij} = I(T_{ij} < C_{ij})\epsilon_{ij}$, where C_{ij} is the independent censoring time. Let $R_{(k)}$ be the risk set at $y_{(k)}$, which is the k th smallest distinct event time for type 1 events among the observed values y_{ij} 's of Y_{ij} 's; it is defined by

$$R_{(k)} = R(y_{(k)}) = \left\{ (i, j) : y_{ij} \geq y_{(k)} \quad \text{or} \quad (y_{ij} \leq y_{(k)} \quad \text{and} \quad \xi_{ij} > 1) \right\}.$$

Accordingly, a weighted partial h-likelihood h_w^* ([10]) based on the IPCW is defined by

$$h_w^* = \ell_{1w}^* + \sum_i \ell_{2i}, \quad (6)$$

where

$$\ell_{1w}^* = \sum_{ij} \delta_{ij} \eta_{ij} - \sum_k d_{(k)} \log \left\{ \sum_{(i,j) \in R_{(k)}} w_{ij} \exp(\eta_{ij}) \right\},$$

with $\delta_{ij} = I(\xi_{ij} = 1)$, $d_{(k)}$ being the number of type 1 events at $y_{(k)}$,

$$w_{ij} = w_{ij}(y_{(k)}) = \frac{\hat{G}(y_{(k)})}{\hat{G}(y_{ij} \wedge y_{(k)})}$$

being the weight of a subject j in cluster i at $y_{(k)}$, and $\hat{G}_{(\cdot)}$ is the Kaplan-Meier estimate of the survival function for the censoring times. Here, $w_{ij} = 1$ as long as individuals have not failed (i.e. $y_{ij} > y_{(k)}$; the first condition of $R_{(k)}$), whereas $w_{ij} < 1$ and decreasing over time if they failed from another type (type 2) (i.e. $y_{ij} \leq y_{(k)}$ and $\xi_{ij} > 1$; the second condition of $R_{(k)}$) [33]. Note that h_w^* in (6) is an extension of the weighted log partial likelihood [6, 33, 34] for the Fine-Gray model to the subhazard frailty model (1). Accordingly, hereafter we use the estimation procedure based on h_w^* for model (1), which handles the general case allowing for the censoring data: for more details see [10].

3. Variable selection using the penalized h-likelihood

In this section, we discuss useful penalty functions for variable selection. Then we show how to extend the h-likelihood procedure of the subhazard frailty model (1) to a penalized likelihood procedure for the purpose of variable selection.

3.1. Penalty function for variable selection

We consider variable selection of fixed effects β in model (1) by maximizing a penalized profile h-likelihood h_p using $h_w^*(\beta, v, \theta)$ in (6) and a penalty; it is defined by

$$h_p(\beta, v, \theta) = h_w^* - n \sum_{j=1}^p J_\gamma(|\beta_j|), \quad (7)$$

where $J_\gamma(|\cdot|)$ is a penalty function that controls model complexity using the tuning parameter γ . Note here that no penalty was imposed on the frailty parameters θ . Typically, setting $\gamma = 0$ results in the subhazard frailty model, whereas the regression coefficient estimates $\hat{\beta}$ tend to 0 as $\gamma \rightarrow \infty$. That is, a larger value of γ tends to choose a simple model, whereas a smaller value of γ is inclined to choose a complex model [4]. A method for choosing an optimal value of γ will be discussed later.

Various penalty functions have been used in the literature on variable selection in statistical models including Cox-type PH models [2, 4, 11]. In this paper, we mainly consider the following three penalty functions, but our results can be applied to other penalty functions which are not discussed here. (i) LASSO [1]:

$$J_{\gamma}(|\beta|) = \gamma|\beta|, \quad (8)$$

(ii) SCAD [2]:

$$J'_{\gamma}(|\beta|) = \gamma I(|\beta| \leq \gamma) + \frac{(a\gamma - |\beta|)_+}{a-1} I(|\beta| > \gamma), \quad (9)$$

where $a = 3.7$ and x_+ denotes the positive part of x , i.e. x_+ is x if $x > 0$, zero otherwise. (iii) HL [18]:

$$J_{\gamma}(|\beta|) \equiv J_{(a,b)}(|\beta|) = \log \Gamma(1/b) + \frac{\log b}{b} + \frac{\beta^2}{2au(|\beta|)} + \frac{(b-2) \log u(|\beta|)}{2b} + \frac{u(|\beta|)}{b}, \quad (10)$$

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

A good penalty function should produce estimates that satisfy unbiasedness, sparsity, and continuity [2, 11]. The LASSO in (8) is the most common penalty as L_1 penalty, but it does not simultaneously satisfy these three properties. Moreover, the LASSO has been criticized on the grounds that it typically ends up selecting a model with too many variables to prevent overshrinkage of the regression coefficients [18, 35]. Fan and Li [2] showed that the SCAD in (9) satisfies all the three properties and that it can perform well as the oracle procedure in terms of selecting the correct subset model and estimating the true non-zero coefficients, simultaneously.

Lee and Oh [18] proposed a new penalty in (10), called the HL penalty, using h-likelihood within the framework of a random effect model, resulting in being unbounded at the origin: for the derivation, see Appendix of Lee and Oh [18]. The shapes of $J_{(a,b)}(|\beta|)$, at various values of $b = 0, 2$ and 30 and $a = 1$, are shown in Figure 1. The form of the penalty changes from a quadratic shape ($b = 0$) for ridge regressions to a cusped form ($b = 2$) for LASSO and then to an unbounded form ($b > 2$) at the origin. In case of $b > 2$, it allows for an infinite gain at zero. The SCAD provides oracle ML estimates (least squares estimators), whereas the HL gives oracle shrinkage estimates. When there is a multi-collinearity, shrinkage estimation is better than the ML estimation. Lee et al. [19, 20, 21] have shown its advantages of the HL approach over LASSO and SCAD methods, especially when the number of covariates is larger than the sample size (i.e. $p > n$); it actually has a better property for variable selection without losing prediction power. Since a in (10) has a greater sensitivity to change of penalty than b , we consider only a few values for b , e.g. $b = 2.1, 3, 10, 30, 50$ representing small, medium and large.

3.2. Penalized h-likelihood procedure

By maximizing the penalized h-likelihood h_p in (7), we need to screen variables and estimate their associated regression coefficients simultaneously. In other words, those

variables whose regression coefficients are estimated as zero are automatically deleted. To achieve the goal, using h_p , the estimation procedures of the fixed parameters (β, θ) and random effects v are required. First, the maximum penalized h-likelihood (MPHL) estimates of (β, v) , given frailty parameter θ , are obtained by solving the joint estimating equations of β and v :

$$\partial h_p / \partial \beta_j = \partial h_w^* / \partial \beta_j - n \sum_{j=1}^p [J_\gamma(|\beta_j|)]' = 0 \quad (11)$$

$$\text{and } \partial h_p / \partial v = \partial h_w^* / \partial v = 0. \quad (12)$$

Note that (11) is an adjusted estimating equation induced by adding the penalty term, whereas (12) is the same as the standard estimating equation without penalty. However, for the three penalty functions considered in (8)-(10), J_γ in (11) becomes non-differentiable at the origin and it does not have continuous second-order derivatives. To overcome this difficulty in solving (11), we use a local quadratic approximation (referred to as LQA [2]) to such penalty functions. That is, given an initial value of $\beta^{(0)}$ close to the true value of β , the penalty function J_γ can be locally approximated by a quadratic function as

$$[J_\gamma(|\beta_j|)]' = J_\gamma'(|\beta_j|) \text{sgn}(|\beta_j|) \approx \left\{ J_\gamma'(|\beta_j^{(0)}|) / |\beta_j^{(0)}| \right\} \beta_j \quad \text{for } \beta_j \approx \beta_j^{(0)}.$$

Then the negative Hessian matrix of β and v based on h_p can be explicitly written as a simple matrix form [36]:

$$H(h_p; \beta, v) = -\partial^2 h_p / \partial (\beta, v)^2 = \begin{pmatrix} X^T W^* X + n \Sigma_\gamma & X^T W^* Z \\ Z^T W^* X & Z^T W^* Z + U \end{pmatrix}, \quad (13)$$

where $\Sigma_\gamma = \text{diag}\{J_\gamma'(|\beta_j|)/|\beta_j|\}$. Here X and Z are $n \times p$ and $n \times q^*$ model matrices for β and v whose ij th row vectors are x_{ij}^T and z_{ij}^T , respectively, $W^* = W(\beta, v) = -\partial^2 h^* / \partial \eta^2$ is the form of the symmetric matrix given in Appendix 2 of Ha and Lee [36] and Ha et al. [10], $\eta = X\beta + Zv$ and $U = -\partial^2 \ell_2 / \partial v^2$ is a $q^* \times q^*$ matrix that takes a form of $U = \text{BD}(\Sigma^{-1}, \dots, \Sigma^{-1})$ if $v \sim N(0, \Sigma)$, where $q^* = q \times r$ and $\text{BD}(\cdot)$ denotes a block diagonal matrix.

Following Ha and Lee [36] and (13), it can be shown that given θ , the MPHL estimates of (β, v) are obtained from the following score equations:

$$\begin{pmatrix} X^T W^* X + n \Sigma_\gamma & X^T W^* Z \\ Z^T W^* X & Z^T W^* Z + U \end{pmatrix} \begin{pmatrix} \hat{\beta} \\ \hat{v} \end{pmatrix} = \begin{pmatrix} X^T w^* \\ Z^T w^* + R \end{pmatrix}. \quad (14)$$

where $w^* = W^* \eta + (\delta - \mu)$ with $\mu = \exp(\log w + \log \Lambda_{01}^s + \eta)$ and $R = Uv + (\partial \ell_2 / \partial v)$. Here w is the weight w_{ij} given in (6) and Λ_{01}^s is the baseline cumulative subhazard function. In particular, $R = 0$ if the log-frailty v follows $N(0, \Sigma)$. The score equations (14) are extensions of the existing estimation procedures. For example, under no-penalty (i.e., $\Sigma_\gamma = 0$) they

become the score equations of Ha et al. [10] for the standard subhazard frailty models. For variable selection under the Fine-Gray model [7] without frailty, they also reduce to

$$\left(X^T W^* X + n \Sigma_\gamma\right) \hat{\beta} = X^T w^*, \quad (15)$$

implying that the new equations (14) gives a special case of the penalized equation (15) for the Fine-Gray model. Notice that to avoid some numerical difficulty in solving (14), we employ $\Sigma_{\gamma, \varepsilon} = \text{diag}\{J'_\gamma(|\beta_j|)/(|\beta_j| + \varepsilon)\}$ for a small positive value of ε (e.g. $\varepsilon = 10^{-8}$), instead of Σ_γ to assure the existence of $\Sigma_{\gamma, \varepsilon}$ [18]. As long as ε is small, the diagonal elements of $\Sigma_{\gamma, \varepsilon}$ are very close to those of Σ_γ . In fact, this algorithm is identical to that of Hunter and Li [37] for improving the LQA; see also [38]. In this paper, we report $\hat{\beta} = 0$ if all five printed decimals are zero. In case of the SCAD and HL penalties, there may exist several local maximums. Thus, a good initial value is essential to obtain a proper estimate $\hat{\beta}$. In this paper, a LASSO solution will be used as the initial value for the SCAD and HL penalties.

Next, for estimation of θ , we use an adjusted profile h-likelihood $p_\tau(h_p)$ [16, 36] which eliminates (β, v) from h_p in (7), defined by

$$p_\tau(h_p) = \left[h_p - \frac{1}{2} \log \det \left\{ \frac{H(h_p; \tau)}{(2\pi)} \right\} \right] \Big|_{\tau = \hat{\tau}}, \quad (16)$$

where $\tau = (\beta^T, v^T)^T$ and $\hat{\tau} = \hat{\tau}(\theta) = \left(\hat{\beta}^T(\theta), v^T(\theta) \right)^T$. The estimates of θ are obtained by solving the score equations $p'_\tau(h_p)/\theta = 0$ as in Ha et al. [10]. Accordingly, we see that the proposed procedure is easily implemented via a slight modification to the existing h-likelihood procedures [10, 17, 36].

3.3. Standard error and selection of tuning parameter

We first show that the standard error (SE) of $\hat{\beta}$ can be obtained by computing an approximated covariance estimate of $\hat{\beta}$. For this we consider a further penalized profile h-likelihood after eliminating v in h_p of (7), defined by

$$\hat{h}_p(\beta, \theta) \equiv h_p|_{v=\hat{v}} = \hat{h} - n \sum_{j=1}^p J_\gamma(|\beta_j|),$$

where $\hat{h} = \hat{h}(\beta, \theta) = h_w^*(\beta, \theta, v)|_{v=\hat{v}}$. In frailty models, regression parameters β and frailty parameter θ are asymptotically orthogonal [15, 17, 36], so that, in estimating the covariance matrix of $\hat{\beta}$ only, the information loss caused by estimating θ is minimal. Fine and Gray [7]

proposed a robust sandwich covariance estimator to estimate $\text{cov}(\hat{\beta})$ using an empirical process theory because the martingale properties break down under the Fine-Gray model due to the use of IPCW and hence the standard asymptotic theories are not valid [10]. Thus, the SEs of $\hat{\beta}$ can be obtained from a sandwich formula [11, 39] based on $\hat{h}_p$:

$$\begin{aligned} \text{cov}(\hat{\beta}) &= H(\hat{h}_p; \beta)^{-1} \text{cov}(\partial \hat{h}_p / \partial \beta) H(\hat{h}_p; \beta)^{-1} \\ &= (H_{\beta\beta} + n\Sigma_\gamma)^{-1} H_{\beta\beta} (H_{\beta\beta} + n\Sigma_\gamma)^{-1}, \end{aligned} \quad (17)$$

where $H(\hat{h}_p; \beta) \equiv -\partial^2 \hat{h}_p / \partial \beta^2 = H_{\beta\beta} + n\Sigma_\gamma$. Here, $H_{\beta\beta} \equiv H(\hat{h}; \beta) \equiv -\partial^2 \hat{h} / \partial \beta^2$ is explicitly computed as follows:

$$\begin{aligned} H_{\beta\beta} &= \left\{ \left(-\partial^2 h_w^* / \partial \beta^2 \right) - \left(-\partial^2 h_w^* / \partial \beta \partial v \right) \left(-\partial^2 h_w^* / \partial v^2 \right) \left(-\partial^2 h_w^* / \partial v \partial \beta \right) \right\} \Big|_{v=\hat{v}} \\ &= \left\{ \left(X^T W^* X \right) - \left(X^T W^* Z \right) \left(Z^T W^* Z + U \right)^{-1} \left(Z^T W^* X \right) \right\} \Big|_{v=\hat{v}} \end{aligned}$$

since $\hat{h} / \beta = \{ (h_w^* / \beta) + (h_w^* / v)(v / \hat{v}) \} \Big|_{v=\hat{v}}$ [17, 36]. Here, we use $H_{\beta\beta}$ to estimate $\text{cov}(\hat{h}_p / \beta)$: see also Yang [40]. We investigate the performance of the proposed SE using (17) by simulation studies in the next section.

Selecting important variables, using the penalized likelihood approaches, also depends on an appropriate choice of the tuning parameters [41, 42]. For the choice of tuning parameters γ , a generalized cross-validation (GCV) statistic has been extensively used [2, 11, 12]. However, Wang et al. [41] showed that the GCV approach can not select the tuning parameters satisfactorily, with a nonignorable overfitting effect in the resulting model [4, 42]. Thus, they proposed to use a BIC-based selection criterion. In spirit of Wang et al. [41], we propose to use a BIC-type criterion based on the h-likelihood for selecting tuning parameters γ , defined by

$$BIC(\gamma) = -2p_v(h_w^*) \left(\hat{\beta}, \hat{\theta} \right) + e(\gamma) \log(n), \quad (18)$$

where $p_v(h_w^*) = [h_w^* - (1/2) \log \det\{H(h_w^*; v)/(2\pi)\}] \Big|_{v=\hat{v}}$ with $H(h_w^*; v) = -\partial^2 h_w^* / \partial v^2$ is the first-order Laplace approximation to the marginal partial likelihood $m_w^*(\beta, \theta) = \log \{ \int \exp(h_w^*) dv \}$ [17, 43] and it is evaluated at $(\hat{\beta}, \hat{\theta})$, and

$$e(\gamma) = \text{tr} \left\{ H(\hat{h}_p; \beta)^{-1} H(\hat{h}; \beta) \right\} = \text{tr} \left[\{ H_{\beta\beta} + n\Sigma_\gamma \}^{-1} H_{\beta\beta} \right]$$

is the effective number of parameters [13, 15]. Note that $\hat{\gamma} = \arg \min_\gamma \{ BIC(\gamma) \}$ is calculated using a simple grid search method as in Fan and Li [11].

In summary, in the inner loop we maximize h_p for $\tau = (\beta^T, v^T)$ (i.e., we solve (14)) and the adjusted profile h-likelihood $p_\tau(h_p)$ in (16) for θ , respectively. In the outer loop, we find γ that minimizes $BIC(\gamma)$ in (18). After convergence has occurred, we compute the estimates of the SEs for $\hat{\beta}$ using (17).

4. Simulation studies

We have conducted numerical studies, based upon 100 replications of simulated data as in Fan and Li [2, 11], Zhang and Lu [44] and Chen et al. [45], to evaluate the performance of

the proposed h-likelihood procedure for the subhazard frailty model (1). We have compared performances of the three variable-selection methods with LASSO, SCAD, and HL penalties, under the following three simulation scenarios.

Simulation I (shared subhazard frailty model with 8 covariates)

Following the scenarios considered in Fine and Gray [7] and Fan and Li [11], we have generated data from the subhazard model with a shared frailty (2). Included in the model were eight covariates $x_{ij} = (x_{ij1}, \dots, x_{ij8})^T$ and a shared random effect v_{i0} with mean 0 and variance σ_0^2 . The conditional subdistribution for a type 1 event given x_{ij} and v_{i0} is given by

$$F_1(t|x_{ij}, v_{i0}) = P(T_{ij} \leq t, \epsilon_{ij}=1|x_{ij}, v_{i0}) = 1 - \left[1 - p(1 - e^{-t})\right]^{\exp(\eta_{ij}^{(1)})},$$

where $p = P(\epsilon_{ij} = 1|x_{ij} = 0, v_{i0} = 0)$ is the proportion of type 1 events and $\eta_{ij}^{(1)} = x_{ij}^T \beta_1 + v_{i0}$. Here $\beta_1 = (\beta_{1,1}, \dots, \beta_{1,8})^T$ are regression parameters for type 1 events. Thus the conditional distribution function of T_{ij} given a type 1 event as well as x_{ij} and v_{i0} is given by

$$F(t|x_{ij}, v_{i0}, \epsilon_{ij}=1) = \frac{1 - [1 - p(1 - e^{-t})]^{\exp(\eta_{ij}^{(1)})}}{1 - (1 - p)^{\exp(\eta_{ij}^{(1)})}}. \quad (19)$$

Times to type 1 events of interest were generated from the conditional distribution function in (19) using the probability integral transformation, given x_{ij} and v_{i0} . The conditional subdistribution for type 2 events was simply obtained by taking $P(\epsilon_{ij} = 2|x_{ij}, v_{i0}) = 1 - P(\epsilon_{ij} = 1|x_{ij}, v_{i0})$ and using an exponential distribution with rate $\exp(x_{ij}^T \beta_2 + v_{i0})$ for $P(T_{ij} \leq t|\epsilon_{ij} = 2, x_{ij}, v_{i0})$, where $\beta_2 = (\beta_{2,1}, \dots, \beta_{2,8})^T$ are regression parameters for type 2 events. Thus the conditional distribution function of T_{ij} given a type 2 event as well as x_{ij} and v_{i0} is given by

$$F(t|x_{ij}, v_{i0}, \epsilon_{ij}=2) = 1 - \exp\left\{-\exp(\eta_{ij}^{(2)})t\right\}, \quad (20)$$

where $\eta_{ij}^{(2)} = x_{ij}^T \beta_2 + v_{i0}$. As before, type 2 event times (times-to-competing events) were generated from the conditional distribution function in (20) using the probability integral transformation.

The following sample sizes were considered: $n = \sum_{i=1}^q n_i$ with $n = 100, 250, 300$, and 500 , and $(q, n_i) = (50, 2), (50, 5), (150, 2)$, and $(100, 5)$. The value of p was set to 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 covariates $x_{ij} = (x_{ij1}, \dots, x_{ij8})^T$ were generated with an AR(1) structure with a correlation coefficient $\rho = 0.5$. The variance of the random effects was assumed to be $\sigma_0^2 = 0.5$. 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 30%.

For the criteria for variable selection, we have adopted the average number of zero coefficients, the probability of choosing the true model (PT), and model error (ME). Following Fan and Li [11], the ME for the subhazard model with a shared frailty (2) was defined by $ME(\hat{\beta}) = E\left\{ \exp(-x^T \hat{\beta}) - \exp(-x^T \beta) \right\}^2$. Let MRME stand for the median of ratios of ME of a selected model to that of the standard estimate under the full model. SAS/IML was used for model fitting and further computation. The simulation results are summarized in Table 1. Here, the column labeled 'C' (5 is the best) indicates the average number of regression coefficients, of the five true zeros, correctly found to zero, and 'IC' (0 is the best) indicates the average number of the three true non-zeros incorrectly found to zero.

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 size of q or n_i . In addition, the HL consistently outperforms the SCAD in terms of 'C' and 'PT', but it does so only when sample size is small as (50,2) or (50,5) in terms of MRME.

Simulation II (correlated subhazard frailty model with 8 covariates)

Next, we generated datasets from the correlated subhazard frailty model (3) with a general covariance matrix (4), where $\sigma_0^2 = \sigma_1^2 = 0.5$ and $\rho = -0.5$, giving $\sigma_{01} = -0.25$. In the multi-center study described in Section 5.2, the average number of patients per center, i.e. average of n_i , is 11.8. Thus, three cases of sample sizes were considered for our simulations:

$n = \sum_i^q n_i = 400$ and 800 with $(q, n_i) = (4, 10), (20, 20)$ and $(80, 10)$. The first component x_1 of the covariate vector $x = (x_1, x_2, \dots, x_8)^T$ was generated from a Bernoulli distribution with success probability of 0.5 in order to mimic the binary treatment covariate of the multi-center study, and the other 7 components $(x_2, \dots, x_8)$ of x were generated from the AR(1) as before. The rest of the simulation schemes were the same as those in Simulation I. The results are summarized in Table 2. The trends in Table 2 are similar to those presented in Table 1. In particular, the HL leads to smaller values of MRME than the SCAD for small sample cases $(q, n_i) = (40, 10)$ and $(20, 20)$. With a given sample size of $n = \sum_i n_i = 400$, larger values of n_i yield greater 'C' values, rather than those of q .

Simulation III (shared subhazard frailty model with 15 covariates)

We have conducted a simulation study with more covariates, i.e. 15 covariates, in the shared subhazard frailty model (2). The corresponding true parameters are $\beta = (1, 0.8, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)$, where the numbers of zeros and nonzeros are 13 and 2, respectively. The remaining simulation settings are the same as those in Simulation I. The results are summarized in Table 3. Table 3 again shows that the SCAD and HL overall work very well. Moreover, the HL is better than the SCAD in terms of 'C', 'PT', and MRME.

In Supplementary material, under three simulation scenarios I, II, and III, we computed the mean of non-zero coefficients of $\hat{\beta}$, their standard deviation (SD), and standard error (SE) which is obtained from the sandwich formula (17). Note that the SE is the average of 100

estimated standard errors for $\hat{\beta}$ and that the SD is the estimates of the true $\left\{ \text{var}(\hat{\beta}) \right\}^{1/2}$. In Supplementary Tables 1, 2, and 3, one can see 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.

5. Practical examples

5.1. Bladder-cancer data: Shared subhazard frailty model

We illustrate the proposed procedures using a dataset from a bladder cancer trial conducted by European Organization for Research and Treatment of Cancer (EORTC) [46]. We consider 396 patients with stage Ta and T1 bladder cancer from 21 centres included in the EORTC trial 30791, focusing on two competing endpoints, time to first bladder recurrence (an event of interest; type 1 event) and time to death prior to recurrence (competing event; type 2 event). Of 396 patients, 200 (50.51%) had recurrence of bladder cancer and 81 (20.45%) died prior to recurrence. 115 Patients (29.04%) who were still alive without recurrence were censored at the date of the last available follow-up. The numbers of patients per centre varied from 3 to 78, with mean 18.9 and median 14.

Nine categorical covariates (x) of interest are included in the analysis: main treatment (CHEMO; no, yes), age (< 65 years, > 65 years), sex, prior recurrent rate (PRIORREC; primary, $< 1/\text{yr}$, $> 1/\text{yr}$), number of tumors (NOTUM; single, 2-7 tumors, ≥ 8 tumors), tumor size ($< 3\text{cm}$, $\geq 3\text{cm}$), T category (Ta, T1), carcinoma in situ (no, yes), and G grade (GLOCAL; G1, G2, G3). For covariates with three categories (PRIORREC, NOTUM, and GLOCAL), we generated two indicator covariates. For example, with variable PRIORREC we coded $\text{PRIORREC1} = I(\text{PRIORREC} = < 1/\text{yr})$ and $\text{PRIORREC2} = I(\text{PRIORREC} > 1/\text{yr})$, where $I(\cdot)$ is the indicator function. Similarly, with variables NOTUM and GLOCAL, we have used the respective indicators (NOTUM1, NOTUM2) and (GLOCAL1, GLOCAL2). Thus total 12 covariates were included in the model.

We fitted the proportional subhazards model with the shared frailty (2) by using the penalized h-likelihood procedure presented in Section 3. The selected values of the tuning parameters γ by the BIC (18) were 0.012, 0.084, and $(a, b) = (0.011, 50)$ for the LASSO, SCAD, and HL, respectively. The estimates of the frailty parameter σ_0^2 for no-penalty, LASSO, SCAD, and HL are 0.106, 0.072, 0.107, and 0.088, respectively. The estimated coefficients and their standard errors for bladder cancer recurrence (i.e., type 1 event) are given in Table 4. The main covariate, CHEMO (x_1), is very significant in all the four methods: see also [17] under no penalty. The LASSO chooses nine covariates ($x_1, x_2, x_5, x_6, x_7, x_8, x_9, x_{11}, x_{12}$) out of the twelve covariates, whereas the SCAD and HL choose six ($x_1, x_5, x_6, x_7, x_{11}, x_{12}$) and seven ($x_1, x_2, x_5, x_6, x_7, x_{11}, x_{12}$) covariates, respectively. In particular, the non-zero estimates by the SCAD are overall similar to the corresponding estimates without penalty ($\gamma = 0$). As expected, the LASSO selects more covariates as compared to the SCAD and HL. Notice here that the LASSO chooses the two covariates (x_8 and x_9) which are not significant under no-penalty. This may be because the LASSO selects unimportant variables much more than the other two methods, as evident in lower 'C' values

of the LASSO in Tables 1-3. These findings indicate that the LASSO might not properly identify important variables in the shared subhazard frailty models.

5.2. Breast-cancer data: Correlated subhazard frailty model

We re-examine the data (B-14) from a multicenter breast cancer trial conducted by the National Surgical Adjuvant Breast and Bowel Project (NSABP) [47, 48]. **Two** thousand five hundred and forty six (2,546) eligible patients from 162 distinct centers were followed up for about 20 years since randomization. The aim of this analysis is to investigate the effect of treatment on local or regional recurrence. For simplicity, we consider only the old 1763 patients (i.e., age ≥ 50) in the data set. The number of patients per center varied from 1 to 114, with the mean of 11.8 and the median of 6. The patients were randomized to one of two treatment arms, tamoxifen (1413 patients) or placebo (1404 patients). Here we consider two event types. The first type is local or regional recurrence (type 1) and the second type is a new primary cancer, distance recurrence or death (type 2); only the event that occurs first is of interest in this analysis, so that the repeated event times are not considered. That is, type 1 is an event of interest (465 patients; 26.38%), type 2 is a competing event (469 patients; 26.60%), and patients with no-events were censored at the last follow-up (1200 patients; 47.02%). We studied dependence of the time to local or regional recurrence on the following ten covariates: treatment group (GROUP; placebo, tamoxifen), race (RACE; white, black, other), menopausal status (MENSE; premenopausal, perimenopausal, postmenopausal), number of nodes removed (RNOD), tumor size (TSIZE), estrogen receptor level (ER), progesterone receptor level (PR), and surgery type (lumpectomy, mastectomy). As before, we created two indicator covariates for variables RACE and MENSE (see Table 5). Four continuous covariates (RNOD, TSIZE, ER, PR) are standardized while other covariates are binary, a total of 10 covariates being included in the model.

Let v_{i0} and v_{i1} be random center and random treatment effects, respectively. We consider the following correlated subhazard frailty model (3) with v_{i0} and v_{i1} :

$$\eta_{ij} = v_{i0} + (\beta_1 + v_{i1}) x_{ij1} + \sum_{m=2}^{10} \beta_m x_{ijm},$$

where x_{ij1} is GROUP and x_{ijm} ($m = 2, \dots, 10$) are the remaining covariates. Here we assume the correlation structure (4) between v_{i0} and v_{i1} . The fitted results are as follows. The selected values of the tuning parameters γ were 0.004, 0.026, and $(a, b) = (0.001, 50)$ for the LASSO, SCAD, and HL, respectively. The frailty-parameter estimates $(\hat{\sigma}_0^2, \hat{\sigma}_1^2, \hat{\rho})$ for the no-penalty, LASSO, SCAD, and HL are (0.297, 0.116, -0.988), (0.290, 0.101, -0.996), (0.289, 0.115, -0.992), and (0.294, 0.101, -0.996), respectively. The estimated coefficients and their SEs for type 1 events are reported in Table 5. The LASSO chooses five covariates (x_1, x_6, x_7, x_8, x_9), but the SCAD and HL select four covariates (x_1, x_6, x_7, x_8). Christian [28] and Ha et al. [10] have shown that the main treatment effect (GROUP; x_1) is significant, which is also confirmed by the three methods (LASSO, SCAD, and HL). Given that unlike the ER (x_8), the PR (x_9) is not known as an important prognostic factor in breast cancer, the LASSO procedure seems more liberal in selecting important variables, in contrast to the SCAD and

HL. These results indicate that the SCAD and HL might identify important variables better in general subhazard frailty models than the LASSO.

6. Discussion

Using a penalized h-likelihood procedure, we have shown how to select important variables in general subhazard frailty models. We have demonstrated via numerical studies and data analyses that the proposed procedure with HL or SCAD penalty overall performs well. In particular, the simulation results indicate that the HL method is preferable to the SCAD method because it identifies zero and non-zero coefficients better without losing prediction accuracy. An advantage of our method can be easily implemented by a slight modification to the existing h-likelihood estimation procedure. Thus our method can be straightforwardly applied to variable selection in the cause-specific PH frailty models [28, 49].

The proposed h-likelihood framework is based on the SCAD or HL penalty. However, the SCAD method may not be directly applicable to the high dimensional case with $p > n$ [4, 19]. With the HL penalty method, an extension to such high dimensional case in competing-risks frailty models would be also an interesting topic.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors thank the European Organization for Research and Treatment of Cancer Genito-Urinary Tract Cancer Group for permission to use the data from EORTC trial 30791 for this research. This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology, Korea (No. 2010-0021165). This work was supported by the National Research Foundation of Korea (NRF) funded by the Korea government (MSIP) (No. 2011-0030037). Dr. Jeong's research was support in part by National Institute of Health (NIH) grants 5-U10-CA69974-09 and 5-U10-CA69651-11.

References

1. Tibshirani R. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B*. 1996; 58:267–288.
2. Fan J, Li R. Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*. 2001; 96:1348–1360. DOI: 10.1198/016214501753382273.
3. Zou H. The adaptive Lasso and its oracle properties. *Journal of the American Statistical Association*. 2006; 101:1418–1429. DOI: 10.1198/016214506000000735.
4. Fan J, Lv J. A selective overview of variable selection in high dimensional feature space. *Statistica Sinica*. 2010; 20:101–148. [PubMed: 21572976]
5. Tibshirani R. Regression shrinkage and selection via the lasso: a retrospective. *Journal of the Royal Statistical Society, Series B*. 2011; 73:273–282.
6. Kuk D, Varadhan R. Model selection in competing risks regression. *Statistics in Medicine*. 2013; 32:3077–3088. DOI: 10.1002/sim.5762. [PubMed: 23436643]
7. Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association*. 1999; 94:496–509. DOI: 10.1080/01621459.1999.10474144.

8. Breiman L. Heuristics of instability and stabilization in model selection. *The Annals of Statistics*. 1996; 24:2350–2383.
9. Katsahian S, Resche-Rigon M, Chevret S, Porcher R. Analysing multicentre competing risk data with a mixed proportional hazards model for the subdistribution. *Statistics in Medicine*. 2006; 25:4267–4278. DOI: 10.1002/sim.2684. [PubMed: 16960919]
10. Ha ID, Christian NJ, Jeong JH, Park J, Lee Y. Analysis of clustered competing risks data using subdistribution hazards models with multivariate frailties. *Statistical Methods in Medical Research*. 2014 DOI: 10.1177/0962280214526193.
11. Fan J, Li R. Variable selection for Cox's proportional hazards model and frailty model. *The Annals of Statistics*. 2002; 30:74–99.
12. Androulakis E, Koukouvinos C, Vonta F. Estimation and variable selection via frailty models with penalized likelihood. *Statistics in Medicine*. 2012; 31:2223–2239. DOI: 10.1002/sim.5325. [PubMed: 22419584]
13. Ha ID, Lee Y, MacKenzie G. Model selection for multi-component frailty models. *Statistics in Medicine*. 2007; 26:4790–4807. DOI: 10.1002/sim.2879. [PubMed: 17476647]
14. Rondeau V, Michiels S, Liqueur B, Pignon JP. Investigating trial and treatment heterogeneity in an individual patient data meta-analysis of survival data by means of the penalized maximum likelihood approach. *Statistics in Medicine*. 2008; 27:1894–1910. DOI: 10.1002/sim.3161. [PubMed: 18069745]
15. Lee Y, Nelder JA. Hierarchical generalized linear models (with discussion). *Journal of the Royal Statistical Society, Series B*. 1996; 58:619–678.
16. Lee, Y.; Nelder, JA.; Pawitan, Y. *Generalised Linear Models with Random Effects: Unified Analysis via h-Likelihood*. Chapman and Hall; London: 2006.
17. Ha ID, Sylvester R, Legrand C, MacKenzie G. Frailty modelling for survival data from multi-centre clinical trials. *Statistics in Medicine*. 2011; 30:2144–2159. DOI: 10.1002/sim.4250. [PubMed: 21563206]
18. Lee Y, Oh HS. A new sparse variable selection via random-effect model. *Journal of Multivariate Analysis*. 2014; 125:89–99. DOI: 10.1016/j.jmva.2013.11.016.
19. Lee D, Lee W, Lee Y, Pawitan Y. Super sparse principal component analysis for high-throughput genomic data. *BMC Bioinformatics*. 2010; 11:296. DOI: 10.1186/1471-2105-11-296. [PubMed: 20525176]
20. Lee D, Lee W, Lee Y, Pawitan Y. Sparse partial least-squares regression and its applications to high-throughput data analysis. *Chemometrics and Intelligent Laboratory Systems*. 2011; 109:1–8. DOI: 10.1016/j.chemolab.2011.07.002.
21. Lee W, Lee D, Lee Y, Pawitan Y. Sparse canonical covariance analysis for high-throughput data. *Statistical Applications in Genetics and Molecular Biology*. 2011; 10:1–24. DOI: 10.2202/1544-6115.1638.
22. Fan J, Peng H. Nonconcave penalized likelihood with a diverging number of parameters. *The Annals of Statistics*. 2004; 32:928–961.
23. Kwon S, Oh S, Lee Y. The use of random-effect models for high-dimensional variable selection problems. Revision submitted to *Scandinavian Journal of Statistics*. 2013
24. Efron B, Morris C. Data analysis using Stein's estimator and its generalizations. *Journal of the American Statistical Association*. 1975; 70:311–319.
25. Casella G. An introduction to empirical Bayes data analysis. *The American Statistician*. 1985; 39:83–87.
26. Lee Y, Nelder JA. Double hierarchical generalized linear models (with discussion). *Journal of the Royal Statistical Society, Series C*. 2006; 55:139–185. DOI: 10.1111/j.1467-9876.2006.00538.x.
27. Katsahian S, Boudreau C. Estimating and testing for center effects in competing risks. *Statistics in Medicine*. 2011; 30:1608–1617. DOI: 10.1002/sim.4132. [PubMed: 21341296]
28. Christian, NJ. PhD thesis. Department of Biostatistics, University of Pittsburgh; Pittsburgh, PA: 2011. Hierarchical likelihood inference on clustered competing risk data..
29. Ha ID, Lee Y, Song JK. Hierarchical likelihood approach for frailty models. *Biometrika*. 2001; 88:233–243. DOI: 10.1093/biomet/88.1.233.

30. Vaida F, Xu R. Proportional hazards model with random effects. *Statistics in Medicine*. 2000; 19:3309–3324. DOI: 10.1002/gepi.20043. [PubMed: 11122497]
31. Ripatti S, Palmgren J. Estimation of multivariate frailty models using penalized partial likelihood. *Biometrics*. 2000; 56:1016–1022. DOI: 10.1111/j.0006-341X.2000.01016.x. [PubMed: 11129456]
32. Robins, JM.; Rotnitzky, A. Recovery of information and adjustment for dependent censoring using surrogate markers.. In: Jewell, N.; Dietz, K.; Farewell, V., editors. *AIDS Epidemiology-Methodological Issues*. Birkhauser; Boston: 1992. p. 24-33.
33. Pintilie M. Analysing and interpreting competing risk data. *Statistics in Medicine*. 2007; 26:1360–1367. DOI: 10.1002/sim.2655. [PubMed: 16900575]
34. Ruan PK, Gray RJ. Analyses of cumulative incidence functions via non-parametric multiple imputation. *Statistics in Medicine*. 2008; 27:5709–5724. DOI: 10.1002/sim.3402. [PubMed: 18712779]
35. Radchenko P, James GM. Variable inclusion and shrinkage algorithms. *Journal of the American Statistical Association*. 2008; 103:1304–1315. DOI: 10.1198/016214508000000481.
36. Ha ID, Lee Y. Estimating frailty models via Poisson hierarchical generalized linear models. *Journal of Computational and Graphical Statistics*. 2003; 12:663–681. DOI: 10.1198/1061860032256.
37. Hunter D, Li R. Variable selection using MM algorithms. *The Annals of Statistics*. 2005; 33:1617–1642.
38. Johnson BA, Lin DY, Zeng D. Penalized estimating functions and variable selection in semiparametric regression models. *Journal of the American Statistical Association*. 2008; 103:672–680. DOI: 10.1198/016214508000000184. [PubMed: 20376193]
39. Cai J, Fan J, Li R, Zhou H. Variable selection for multivariate failure time data. *Biometrika*. 2005; 92:303–316. DOI: 10.1093/biomet/92.2.303. [PubMed: 19458784]
40. Yang, H. PhD thesis. Department of Statistics, North Carolina State University; Raleigh, NC: 2007. Variable selection procedures for generalized linear mixed models in longitudinal data Analysis..
41. Wang H, Li R, Tsai CL. Tuning parameter selectors for the smoothly clipped absolute deviation method. *Biometrika*. 2007; 94:553–568. DOI: 10.1093/biomet/asm053. [PubMed: 19343105]
42. Zhang Y, Li R, Tsai CL. Regularization parameter selections via generalized information criterion. *Journal of the American Statistical Association*. 2010; 105:312–323. DOI: 10.1198/jasa.2009.tm08013. [PubMed: 20676354]
43. Therneau TM, Grambsch PM, Pankratz VS. Penalized survival models and frailty. *Journal of Computational and Graphical Statistics*. 2003; 12:156–175. DOI: 10.1198/1061860031365.
44. Zhang HH, Lu W. Adaptive Lasso for Cox's proportional hazards model. *Biometrika*. 2007; 94:691–703. DOI: 10.1093/biomet/asm037.
45. Chen Z, Tang ML, Gao W, Shi NZ. New robust variable selection methods for linear regression models. *Scandinavian Journal of Statistics*. 2014 DOI: 10.1111/sjos.12057.
46. Sylvester R, van der Meijden APM, Oosterlinck W, Witjes J, Bouffieux C, Denis L, Newling DWW, Kurth K. 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. *European Urology*. 2006; 49:466–477. DOI: 10.1016/j.eururo.2005.12.031. [PubMed: 16442208]
47. 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 England Journal of Medicine*. 1989; 320:479–484. DOI: 10.1056/NEJM198902233200802. [PubMed: 2644532]
48. 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. *Journal of the National Cancer Institute*. 1996; 88:1529–1542. DOI: 10.1093/jnci/88.21.1529. [PubMed: 8901851]
49. Gorfine M, Hsu L. Frailty-based competing risks model for multivariate survival data. *Biometrics*. 2011; 67:415–426. DOI: 10.1111/j.1541-0420.2010.01470.x. [PubMed: 20707868]

Figure 1.
HL penalties with $b = 0, 2$ and 30 .

Table 1

(8 covariates) Simulation results using 100 replications under the shared subhazard frailty model allowing 30% censoring: proportion of type 1 events is $p = 0.5$

(q, n_i)	Method	C	IC	PT	MRME
(50,2)	LASSO	2.89	0.04	0.07	0.078
	SCAD	4.57	0.11	0.64	0.189
	HL	4.93	0.10	0.88	0.053
(50,5)	LASSO	2.63	0	0.01	1.066
	SCAD	4.70	0	0.73	0.681
	HL	4.92	0	0.92	0.309
(150,2)	LASSO	2.84	0	0.04	1.846
	SCAD	4.57	0	0.63	0.690
	HL	4.91	0	0.91	0.943
(100,5)	LASSO	2.62	0	0.02	2.040
	SCAD	4.75	0	0.80	0.666
	HL	5.00	0	1.00	0.561

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 non-zero incorrectly set to zero; PT, probability of choosing the true model; MRME, median of relative model errors

Table 2

(8 covariates) Simulation results using 100 replications under the correlated subhazard frailty model allowing 30% censoring: proportion of type 1 events is $p = 0.5$

(q, n_i)	Method	C	IC	PT	MRME
(40,10)	LASSO	1.63	0.06	0.01	1.062
	SCAD	4.52	0.06	0.65	0.677
	HL	4.82	0.03	0.82	0.394
(20,20)	LASSO	2.60	0.25	0.01	1.379
	SCAD	4.76	0.20	0.63	0.691
	HL	4.95	0.12	0.83	0.482
(80,10)	LASSO	2.69	0	0.04	1.532
	SCAD	4.65	0	0.67	0.615
	HL	4.92	0	0.89	0.585

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 non-zero incorrectly set to zero; PT, probability of choosing the true model; MRME, median of relative model errors

Table 3

(15 covariates) Simulation results using 100 replications under the shared subhazard frailty model allowing 30% censoring: proportion of type 1 events is $p = 0.5$

(q, n_i)	Method	C	IC	PT	MRME
(50,2)	LASSO	10.20	0	0.04	0.047
	SCAD	11.57	0.02	0.24	0.178
	HL	12.88	0.03	0.87	0.085
(50,5)	LASSO	10.65	0	0.07	0.181
	SCAD	12.24	0	0.52	0.157
	HL	12.85	0	0.86	0.064
(150,2)	LASSO	10.74	0	0.05	0.614
	SCAD	11.99	0	0.41	0.216
	HL	12.90	0	0.90	0.159
(100,5)	LASSO	10.89	0	0.10	1.129
	SCAD	12.33	0	0.61	0.273
	HL	12.93	0	0.93	0.271

q , No. of clusters; n_i , cluster size; HL, h-likelihood penalty function; C, average number of coefficients, of the **thirteen** true zeros, correctly set to zero; IC, average number of the **two** true non-zero incorrectly set to zero; PT, probability of choosing the true model; MRME, median of relative model errors

Table 4

Bladder cancer data: estimated coefficients and standard errors (in parentheses) in the shared subhazard frailty model for bladder cancer recurrence

Variable	No-penalty	LASSO	SCAD	HL
x_1 : CHEMO=yes	-0.933 (0.187)	-0.666 (0.166)	-0.929 (0.182)	-0.785 (0.174)
x_2 : Age > 65 years	-0.343 (0.147)	-0.214(0.120)	0 (0)	-0.218 (0.119)
x_3 : Sex=Female	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 3 cm	0.213 (0.175)	0.060 (0.094)	0 (0)	0 (0)
x_9 : TLOCC=T1	0.171 (0.173)	0.127 (0.115)	0 (0)	0 (0)
x_{10} : CIS=yes	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)

HL, h-likelihood penalty function; CHEMO, main treatment (no, yes); Age, age at diagnosis (< 65 years, > 65 years); Sex (Male, Female); PRIORREC, prior recurrent rate (primary, < 1 /yr, > 1 /yr); PRIORREC1=I(PRIORREC < 1 /yr); PRIORREC2=I(PRIORREC > 1 /yr); NOTUM, number of tumors (single, 2-7 tumors, > 8 tumors); NOTUM1=I(NOTUM = 2-7 tumors); NOTUM2=I(NOTUM > 8 tumors); TUM3CM, tumor size (< 3 cm, > 3 cm); TLOCC, T category (Ta, T1); CIS, carcinoma in situ (no, yes); GLOCAL, G grade (G1, G2, G3); GLOCAL1=I(GLOCAL=G2); GLOCAL2=I(GLOCAL=G3)

Table 5

Breast cancer data: estimated coefficients and standard errors (in parentheses) in the correlated subhazard frailty model for a type 1 event

Variable	No-penalty	LASSO	SCAD	HL
x_1 : GROUP=tamoxifen	-0.617(0.107)	-0.528 (0.097)	-0.610(0.106)	-0.521 (0.097)
x_2 : RACE1	-0.202 (0.267)	0 (0)	0 (0)	0 (0)
x_3 : RACE2	-0.165 (0.340)	0 (0)	0 (0)	0 (0)
x_4 : MENSE1	0.112(0.222)	0 (0)	0 (0)	0 (0)
x_5 : MENSE2	-0.158(0.265)	0 (0)	0 (0)	0 (0)
x_6 : RNOD	-0.139(0.051)	-0.124 (0.046)	-0.139(0.050)	-0.109 (0.044)
x_7 : TSIZE	0.272 (0.041)	0.254 (0.039)	0.266 (0.040)	0.253 (0.039)
x_8 : ER	0.077 (0.037)	0.069 (0.035)	0.022 (0.019)	0.068 (0.032)
x_9 : PR	0.058 (0.045)	0.052 (0.040)	0 (0)	0 (0)
x_{10} : SURGTYPE=mastectomy	-0.089(0.101)	0 (0)	0 (0)	0 (0)

HL, h-likelihood penalty function; GROUP, treatment group (placebo, tamoxifen); RACE, race (white, black, other); RACE1=I(RACE=white); RACE2=I(RACE=black); MENSE, menopausal status (premenopausal, perimenopausal, postmenopausal); MENSE1=I(MENSE=premenopausal); MENSE2=I(MENSE=perimenopausal); RNOD, number of nodes removed; TSIZE, tumor size (mm); ER, estrogen receptor level; PR, progesterone receptor level; SURGTYPE, surgery type (lumpectomy, mastectomy)