Published in final edited form as:

Stat Methods Med Res. 2019 ; 28(10-11): 3404–3414. doi:10.1177/0962280218803654.

Bi-level variable selection for case-cohort studies with group variables

Soyoung Kim, Kwang Woo Ahn

Division of Biostatistics, Medical College of Wisconsin, Milwaukee, WI, USA

Abstract

The case-cohort design is an economical approach to estimate the effect of risk factors on the survival outcome when collecting exposure information or covariates on all patients is expensive in a large cohort study. Variables often have group structure such as categorical variables and highly correlated continuous variables. The existing literature for case-cohort data is limited to identifying non-zero variables at individual level only. In this article, we propose a bi-level variable selection method to select non-zero group and within-group variables for case-cohort data when variables have group structure. The proposed method allows the number of variables to diverge as the sample size increases. The asymptotic properties of the estimator including bi-level variable selection consistency and the asymptotic normality are shown. We also conduct simulations to compare our proposed method with some existing method and apply them to the Busselton Health data.

Keywords

Case-cohort design; efficiency; multiple diseases; survival analysis; variable selection

1 Introduction

The case-cohort design is widely used to estimate the effect of risk factors on survival outcome when measuring exposure information is costly. Case-cohort data consist of a random sample, called the subcohort, from the full cohort and all cases outside the subcohort. In other words, the expensive covariate information is collected from subjects being in the subcohort or having events of interest outside the subcohort. When studying multiple diseases of interest, the case-cohort design capable of using the same subcohort has been advocated.¹

The extensive work has been done for analyzing case-cohort data for survival outcomes. With a single disease, a pseudo-likelihood approach was proposed by Prentice¹ and Self and

Corresponding author: Soyoung Kim, Division of Biostatistics, Medical College of Wisconsin, Milwaukee, WI, USA. skim@mcw.edu.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

Supplemental Material

Supplemental material for this article is available online.

Prentice.² In order to improve efficiency, Barlow³ and Kulich and Lin⁴ proposed a robust estimator using a time-varying weight and a class of weighted estimation using all available information, respectively. When several case-cohort studies have already been conducted for multiple diseases, Kang and Cai⁵ developed a joint model with multivariate failure time. However, they did not use extra information from the other diseases when estimating the effect of risk factors for a disease of interest. Kim et al.⁶ developed a more efficient estimation method with a new weight to make full use of information from the other diseases.

In spite of the progress made in developing methods analyzing case-cohort data, studying variable selection under the case-cohort design has been limited. Recently, Ni et al.⁷ proposed a variable selection procedure using the smoothly clipped absolute deviation (SCAD) penalty to select non-zero individual variables for a single case-cohort study. In practice, variables often have group structure such as categorical variables and highly correlated continuous variables. For example, in the Busselton Health data from Cullen⁸ and Knuiman et al.,⁹ a case-cohort study was conducted to investigate the association between serum ferritin and stroke event. It includes categorical variables such as smoking status and categorized serum ferritin levels, and continuous variables. For a continuous variable x , investigators often examine the effects of $x, \dots, x^a$ on the outcome of interest, where a is a positive integer. In practice, individual variable selection methods are used to identify important variables among $x, \dots, x^a$.^{7,10} However, because they are highly correlated, one may treat $x, \dots, x^a$ as a group and apply a bi-level variable selection technique to identify non-zero group and within-group variables. Thus, bi-level selection may be more efficient than individual variable selection in selecting non-zero variables.

Many methods have been developed for variable selection with group structure under the proportional hazards model. Ma et al.¹¹ and Kim et al.¹² proposed the supervised group lasso and the group lasso method, respectively, which are limited to group variable selection. Huang et al.¹³ studied the group bridge to select variables at bi-level, that is, at group and within-group levels. However, they showed the group variable selection consistency only, not the bi-level variable selection consistency. Alternative approach by Wang and Nan¹⁴ considered a hierarchically penalized proportional hazards regression for bi-level variable selection. For the subdistribution hazards model with group variables, Ahn et al.¹⁵ developed the adaptive group bridge that has the bi-level variable selection consistency. However, to our best knowledge, there is no literature on bi-level variable selection for survival outcome under the case-cohort design.

In this article, we propose the adaptive group bridge which is capable of identifying non-zero variables at group and within-group levels for the univariate proportional hazards model under the case-cohort design. To study a disease of interest, we consider two case-cohort designs: a single case-cohort study and multiple case-cohort studies. In contrast with a single case-cohort study, multiple case-cohort studies have extra information from the other diseases. We propose to use such extra information to improve variable selection accuracy. We show the bi-level variable selection consistency and the asymptotic normality of the proposed estimator while allowing the number of variables to diverge as the sample size increases. The proposed method is applied to the Busselton Health data.

2 Model selection with adaptive group bridge under case-cohort designs

2.1 Estimation for case-cohort designs

We define notations for multiple case-cohort studies for generality. A single case-cohort study can be handled as a special case. To elaborate multiple case-cohort studies, we consider multivariate outcomes one of which is the outcome of interest for the univariate proportional hazards model.

Suppose there are n independent subjects and K diseases in the full cohort. For subject i with disease k , let T_{ik} , C_{ik} , and $X_{ik} = \min(T_{ik}, C_{ik})$ denote the failure time of interest, censoring time, and observed time, respectively. Let

$\Delta_{ik} = I(T_{ik} \leq C_{ik})$, $N_{ik}(t) = I(X_{ik} \leq t, \Delta_{ik} = 1)$, and $Y_{ik}(t) = I(X_{ik} \geq t)$ be, respectively, a failure indicator, the counting process, and an at-risk indicator, where $I(\cdot)$ is an indicator function.

Let $\mathbf{Z}_{ik}(t) = \left\{ Z_{ik1}(t), \dots, Z_{ikd_n}(t) \right\}^T$ be a $d_n \times 1$ possibly time-dependent covariate vector.

Without loss of generality, we assume that $\mathbf{Z}_{ik}(t)$ are centered and standardized. We assume that T_{ik} is independent of C_{ik} given $\mathbf{Z}_{ik}(\cdot)$. The study period is assumed to be $[0, \tau]$.

Consider the event-specific hazards model: the hazard function for subject i with disease k is assumed to be $h_{ik}\{t|\mathbf{Z}_{ik}(t)\} = h_{0k}(t)\exp\{\boldsymbol{\beta}_k^T \mathbf{Z}_{ik}(t)\}$, where $h_{0k}(t)$ is an unspecified baseline

hazard function and $\boldsymbol{\beta}_k = (\beta_1, \dots, \beta_{d_n})^T$ is an unknown parameter vector of interest.¹⁶ Let

$\boldsymbol{\beta}_0 = (\beta_{10}, \dots, \beta_{d_n 0})^T$ be the true parameter vector.

Under multiple case-cohort studies, covariate information is available for (i) a randomly selected subcohort from the full cohort and (ii) all cases from any causes outside the subcohort. The subcohort is shared by all outcomes. More specifically, fixed $\tilde{n}$ subjects are randomly selected from the full cohort for the shared subcohort. Let ξ_i be an indicator of subject i being selected into the subcohort. Each subject is selected with the same probability $\tilde{\alpha} = \text{pr}(\xi_i = 1) = \tilde{n}/n$. The data under multiple case-cohort studies consist of

$\{X_{ik}, \Delta_{ik}, \xi_i, \mathbf{Z}_{ik}(t), 0 \leq t \leq X_{ik}\}$ when $\xi_i = 1$ or $\Delta_{ik} = 1$ for $k = 1, \dots, K$ and $\{X_{ik}, \Delta_{ik}, \xi_i\}$ when $\xi_i = 0$ and $\Delta_{ik} = 0$ for $k = 1, \dots, K$.

Assume disease k is of interest. For disease k , the negative pseudo-partial likelihood is

$$\tilde{l}_k(\boldsymbol{\beta}) = - \sum_{i=1}^n \int_0^\tau \left[\boldsymbol{\beta}_k^T \mathbf{Z}_{ik}(t) - \log \sum_{j=1}^n w_{jk}(t) Y_{jk}(t) \exp\{\boldsymbol{\beta}_k^T \mathbf{Z}_{jk}(t)\} \right] dN_{ik}(t) \quad (1)$$

where $w_{jk}(t)$ is the time-varying weight function for case-cohort data. There are several weight functions.^{1,3,6,17,18} In this paper, we consider two efficient time-varying weight functions. The first weight function proposed by Kalbfleisch and Lawless¹⁷ for a single case-cohort study has the following form

$$w_{ik,1}(t) = \Delta_{ik} + (1 - \Delta_{ik})\xi_i \hat{\alpha}_k^{-1}(t) \quad (2)$$

where $\hat{\alpha}_k(t) = \sum_{i=1}^n \xi_i Y_{ik}(t) (1 - \Delta_{ik}) / \sum_{i=1}^n Y_{ik}(t) (1 - \Delta_{ik})$ which is an estimator for the subcohort selection probability representing the proportion of sampled subjects among subjects who have disease k and remain in the risk set at time t .

The weight function $w_{ik,1}(t)$ ignores the extra information from the other diseases under multiple case-cohort studies. To use all collected covariate information for subjects who have the other diseases outside the subcohort, Kim et al.⁶ proposed the following more efficient weight

$$w_{ik,2}(t) = \left\{ 1 - \prod_{k=1}^K (1 - \Delta_{ik}) \right\} + \prod_{k=1}^K (1 - \Delta_{ik}) \xi_i \tilde{\alpha}_k^{-1}(t) \quad (3)$$

where $\tilde{\alpha}_k(t) = \sum_{i=1}^n \xi_i Y_{ik}(t) \prod_{k=1}^K (1 - \Delta_{ik}) / \sum_{i=1}^n Y_{ik}(t) \prod_{k=1}^K (1 - \Delta_{ik})$ which is the proportion of sampled subjects among subjects who do not have any diseases and remain in the risk set at time t . The parameter estimator $\tilde{\beta}$ can be obtained by minimizing the negative pseudo-partial likelihood (1) using weights (2) or (3).

2.2 Adaptive group bridge

In this section, we propose the adaptive group bridge for bi-level variable selection with group variables. First, we define notations on group variables and their membership. Suppose that there are G groups and group memberships $A_1, \dots, A_G$ are defined as subsets of $\{1, \dots, d_n\}$. The cardinality of a set A is denoted by $|A|$. Groups are allowed to overlap. We define a $|A| \times 1$ vector $\beta_A = (\beta_m, m \in A)^T$ and $\beta_{A0} = (\beta_{m0}, m \in A)^T$. For individual membership, we define B_1 and B_2 such that $\beta_m = 0$ if $m \in B_1$ and $\beta_m = 0$ if $m \in B_2$. Without loss of generality, we assume $\beta_{A_g} \neq \mathbf{0}$ for $g = 1, \dots, G_1$ and $\beta_{A_g} = \mathbf{0}$ for $g = G_1 + 1, \dots, G$.

For data with group structure, we propose the following penalized pseudo-partial likelihood to obtain the estimator $\hat{\beta}$.

$$\mathcal{L}_n(\beta) = \tilde{l}_k(\beta) + \lambda_n \sum_{g=1}^G c_g \left(\sum_{m \in A_g} \frac{|\beta_m|}{|\tilde{\beta}_m|^v} \right)^\gamma \quad (4)$$

where $0 < \gamma < 1$, $\lambda_n > 0$, $v > 0$, the c_g 's are the constants to adjust for different $|A_g|$'s, and $\tilde{\beta}$ is a consistent estimator of β_0 . The estimator $\tilde{\beta}$ can be obtained from pseudo-partial likelihood (1) using weights (2) or (3). In practice, $c_g \propto |A_g|^{1-\gamma}$ is widely used.^{19,13} When v

$= 0$, the penalty term in (4) is the group bridge penalty of Huang et al.^{13,19} Using $\tilde{\beta}_m$ as the weight is similar to the adaptive lasso of Zou.²⁰ Thus, we call our proposed penalty as the adaptive group bridge penalty. In the case with $\gamma = 1/2$ and $c_g = 1$ for all g , the adaptive group bridge penalty is the same as the adaptive hierarchical penalty of Wang and Nan.¹⁴

Although Huang et al.^{13,19} proposed the group bridge for bi-level variable selection, they showed the group variable selection consistency only, not the bi-level variable selection consistency. Because the group bridge set v to 0 in (4), it equally penalizes β_m 's for $m \in A_g$ within group g . This may lead to inconsistent within-group variable selection. To overcome this limitation of the group bridge, the adaptive group bridge uses the weight $\tilde{\beta}_m$ in the penalty term like the adaptive lasso. When $\beta_{m0} = 0$, $\tilde{\beta}_m$ is close to 0 for sufficiently large n .

Thus, $1/|\tilde{\beta}_j|^v$ assigns larger penalties to zero parameters. On the other hand, $1/|\tilde{\beta}_m|^v$ converges to a non-zero constant when $\beta_{m0} \neq 0$. By doing so, putting the weight $\tilde{\beta}_m$ into the penalty term enables the adaptive group bridge to identify non-zero variables at bi-level more consistently than the group bridge.

2.3 Asymptotic properties

In this section, we study the asymptotic properties of the adaptive group bridge estimator $\hat{\beta}$ using weight (3). The asymptotic properties of the estimator using weight (2) can be similarly shown and thus their proofs are omitted in this article. We define $\mathbf{a}^{\otimes 0} = 1$, $\mathbf{a}^{\otimes 1} = \mathbf{a}$, $\mathbf{a}^{\otimes 2} = \mathbf{a}\mathbf{a}^T$, and the following notations:

$$S_k^{(d)}(\beta, t) = \frac{1}{n} \sum_{i=1}^n Y_{ik(t)} \mathbf{Z}_{ik(t)} \otimes_d \beta^T \mathbf{Z}_{ik(t)}, \quad d = 0, 1, 2$$

$$\tilde{S}_k^{(d)}(\beta, t) = \frac{1}{n} \sum_{i=1}^n w_{ik, 2(t)} Y_{ik(t)} \mathbf{Z}_{ik(t)} \otimes_d \beta^T \mathbf{Z}_{ik(t)}, \quad d = 0, 1, 2$$

$$s_k^{(d)}(\beta, t) = E\{S_k^{(d)}(\beta, t)\}, \quad d = 0, 1, 2, \quad \mathbf{e}_k(\beta, t) = s_k^{(1)}(\beta, t)/s_k^{(0)}(\beta, t)$$

$$v_k(\beta, t) = \frac{s_k^{(2)}(\beta, t)s_k^{(0)}(\beta, t) - s_k^{(1)}(\beta, t) \otimes 2}{s_k^{(0)}(\beta, t)^2}$$

$$V_k(\beta, t) = \frac{\tilde{S}_k^{(2)}(\beta, t)\tilde{S}_k^{(0)}(\beta, t) - \tilde{S}_k^{(1)}(\beta, t) \otimes 2}{\tilde{S}_k^{(0)}(\beta, t)^2}$$

$$\Omega(\beta) = \int_0^\tau v_k(\beta, t) s_k^{(0)}(\beta, t) h_{0k}(t) dt$$

$$\Gamma(\beta) = \frac{1}{n} \text{Var} \{ \partial \tilde{l}_k(\beta) / \partial \beta \}$$

We make the following assumptions:

1. For all k , $\int_0^\tau h_{0k}(t) dt < \infty$ and $P \{ Y_{ik}(t) = 1 \} > 0$ for $t \in [0, \tau]$, $i = 1, \dots, n$.
2. $|Z_{ijk}(0)| + \int_0^\tau |dZ_{ijk}(t)| < D_z < \infty$, $i = 1, \dots, n$, $j = 1, \dots, d_n$ almost surely and D_z is a constant.
3. For $d = 0, 1, 2$, there exists a neighborhood $\mathcal{B}$ of β_0 such that $s_k^{(d)}(\beta, t)$ are continuous functions and $\sup_{t \in [0, \tau], \beta \in \mathcal{B}} \|s_k^{(d)}(\beta, t) - s_k^{(d)}(\beta_0, t)\|_2 \rightarrow 0$, where $\|a\|_p$ defines the L_p norm of a .
4. For all $\beta \in \mathcal{B}$, $t \in [0, \tau]$, and $k = 1, \dots, K$, $S_k^{(1)}(\beta, t) = \partial S_k^{(0)}(\beta, t) / \partial \beta$, and $S_k^{(2)}(\beta, t) = \partial^2 S_k^{(0)}(\beta, t) / \partial \beta \partial \beta^T$,
where $S_k^{(d)}(\beta, t)$ for $d = 0, 1, 2$ are continuous functions of $\beta \in \mathcal{B}$ uniformly in $t \in [0, \tau]$ and are bounded on $\mathcal{B} \times [0, \tau]$; $s_k^{(0)}$ is bounded away from zero on $\mathcal{B} \times [0, \tau]$.
5. There exist constants C_1 and C_2 such that

$$0 < C_3 < \text{eigen}_{\min} \{ \Omega(\beta_0) \} \leq \text{eigen}_{\max} \{ \Omega(\beta_0) \} < C_4 < \infty$$

where $\text{eigen}_{\min} \{A\}$ and $\text{eigen}_{\max} \{A\}$ are minimal and maximal eigenvalues of a matrix A , respectively.

6. There exist constants C_3 and C_4 such that

$$0 < C_3 < \text{eigen}_{\min} \{ \Omega(\beta_0) \} \leq \text{eigen}_{\max} \{ \Omega(\beta_0) \} < C_4 < \infty$$

7. $\lim_{n \rightarrow \infty} \tilde{\alpha} = \alpha$, where $\tilde{\alpha} = \tilde{n}/n$ and α is a positive constant.

8. For some v_1 and v_2 such that $0 < v_1 < 1, 0 < v_2$, and $v_2/(1 - v_1) < v$,
 $\min_{j \in B_1} |\beta_{j0}(\tau)| = O_p \left\{ (d_n/n)^{v_1/2} \right\}, \max_g |A_g \cap B_1| = O \left\{ (n/d_n)^{v_2/2} \right\}$, we assume
 $\sum_{g=1}^{G_1} c_g \left\{ \left(\sum_{j \in A_g \cap B_1} |\beta_{j0}|^{1-v} \right)^{\gamma-1} \sum_{j \in A_g \cap B_1} 1/|\beta_{j0}|^v \right\} \leq M_n$ and
 $\sum_{g=1}^{G_l} c_g \left\{ \left(\sum_{j \in A_g \cap B_1} |\beta_{j0}|^{1-v} \right)^{\gamma-2} \sum_{j \in A_g \cap B_1} 1/|\beta_{j0}|^{2v} \right\} \leq M_n$, where $M_n = O_p$
 (1).
9. $\lambda_n/\sqrt{n} \rightarrow 0, \sqrt{n/d_n} \tilde{\beta}_j = O_p(1)$, and $\min(\lambda_n n^{(v-1)/2} d_n^{-(1+v)/2},$
 $\lambda_n n^{\gamma(v-1)/2} d_n^{-1+\gamma(1-v)/2}) \rightarrow \infty$.

Conditions 1–6 are standard conditions for the proportional hazards model. They are similar to Conditions A1–A3 of Cai et al.¹⁰ and Conditions 1–4 of Ni et al.⁷ They guarantee local asymptotic quadratic property of $\tilde{l}_k(\beta)$ and the existence of local minimizer of $\mathcal{L}_n(\beta)$.

Condition 7 is boundness for the subcohort selection probability. Conditions 8–9 control λ_n , the number of variables within group, and the magnitude of the true non-zero parameters within non-zero groups as $n \rightarrow \infty$. Similar conditions to Conditions 8–9 were used in Ahn et al.¹⁵

We have the following theorem.

Theorem 1. *Under Conditions 1–9, we have*

1. *Consistency: if $d_n^A/n \rightarrow 0$, $\|\hat{\beta} - \beta_0\|_2 = O_p(\sqrt{d_n/n})$.*
2. *Bi-level variable selection consistency: $P(\hat{\beta}_{B_2} = \mathbf{0}) \rightarrow 1$.*
3. *Asymptotic distribution: if $d_n^5/n \rightarrow 0$, we have*

$$n^{1/2} \mathbf{u}^T \Omega_{11}^{-1/2} \Gamma_{11} (\hat{\beta}_{B_1} - \beta_{B_1,0}) \rightarrow_d N(0, 1)$$

where $\mathbf{u}$ is a $|B_1| \times 1$ constant vector with $\|\mathbf{u}\| = 1$, and Ω_{11} and Γ_{11} are the leading $|B_1| \times |B_1|$ submatrices of $\Omega(\beta_0)$ and $\Gamma(\beta_0)$, respectively.

The Supplemental Materials include the proof of Theorem 1 and the estimators of $\Omega(\beta_0)$ and $\Gamma(\beta_0)$. Theorem 1 establishes the oracle property of the adaptive group bridge estimator. In particular, it shows that the adaptive group bridge consistently selects not only non-zero group variables, but also non-zero within-group variables. We have the following corollary:

Corollary 1. *Let $\tilde{\beta}$ be the estimator based on the weighted pseudo-partial likelihood (1) using weights (2) or (3). Under Conditions 1–7 and $d_n^A/n \rightarrow 0$, $\|\tilde{\beta} - \beta_0\|_2 = O_p(\sqrt{d_n/n})$.*

2.4 Computation

Since it is difficult to directly minimize $\mathcal{L}_n(\boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$, we formulate minimizing $\mathcal{L}_n(\boldsymbol{\beta})$ to minimizing

$$\mathcal{Q}(\boldsymbol{\beta}, \boldsymbol{\theta}) = \tilde{l}_k(\boldsymbol{\beta}) + \sum_{g=1}^G \theta_g^{1-1/\gamma} c_g^{1/\gamma} \sum_{m \in A_g} \frac{|\beta_m|}{|\tilde{\beta}_m|^v} + \zeta_n \sum_{g=1}^G \theta_g \quad (5)$$

where $\boldsymbol{\theta} = (\theta_1, \dots, \theta_G)^T$ and $\lambda_n = \zeta_n^{1-\gamma} \gamma^{-\gamma} (1-\gamma)^{\gamma-1}$. We have the following proposition:

Proposition 1. Assume that $\lambda_n = \zeta_n^{1-\gamma} \gamma^{-\gamma} (1-\gamma)^{\gamma-1}$. Then, $\hat{\boldsymbol{\beta}}$ minimizes $\mathcal{L}_n(\boldsymbol{\beta})$ if and only if $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}})$ minimizes $\mathcal{Q}(\boldsymbol{\beta}, \boldsymbol{\theta})$, where $\theta_g > 0$ and $\hat{\theta}_g > 0$ for $g = 1, \dots, G$

We can show Proposition 1 similarly to Huang et al.¹⁹ Thus, its proof is omitted.

To approximate $\tilde{l}_k(\boldsymbol{\beta})$, we define $\nabla \tilde{l}(\boldsymbol{\beta}) = \partial \tilde{l}_k(\boldsymbol{\beta}) / \partial \boldsymbol{\beta}$ and $\nabla^2 \tilde{l}(\boldsymbol{\beta}) = \partial^2 \tilde{l}_k(\boldsymbol{\beta}) / (\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T) = \mathbf{X}^T \mathbf{X}$, where $\mathbf{X}$ can be obtained by the Cholesky decomposition and $\mathbf{Y} = (\mathbf{X}^T)^{-1} \{ \nabla^2 \tilde{l}(\boldsymbol{\beta}) \boldsymbol{\beta} - \nabla \tilde{l}(\boldsymbol{\beta}) \}$. To minimize $\mathcal{Q}(\boldsymbol{\beta}, \boldsymbol{\theta})$, we minimize

$$\tilde{\mathcal{Q}}(\boldsymbol{\beta}, \boldsymbol{\theta}) = \frac{1}{2} (\mathbf{Y} - \mathbf{X} \boldsymbol{\beta})^T (\mathbf{Y} - \mathbf{X} \boldsymbol{\beta}) + \sum_{g=1}^G \theta_g^{1-1/\gamma} c_g^{1/\gamma} \sum_{m \in A_g} \frac{|\beta_m|}{|\tilde{\beta}_m|^v} + \zeta_n \sum_{g=1}^G \theta_g \quad (6)$$

Define $\boldsymbol{\beta}^{(p)}$ is the estimator at the p th step in optimization. Then, the algorithm is as follows:

Step 1 Obtain $\tilde{\boldsymbol{\beta}}$ by minimizing likelihood (1) and set it to the initial value $\boldsymbol{\beta}^{(0)}$.

Step 2 Compute $\mathbf{X}$ and $\mathbf{Y}$ using $\boldsymbol{\beta}^{(p)}$ at the p th step.

Step 3 Compute

$$\theta_g^{(p)} = c_g \left(\frac{1-\gamma}{\zeta_n^\gamma} \right)^\gamma \left(\sum_{m \in A_g} \frac{|\beta_m^{(p)}|}{|\tilde{\beta}_m|^v} \right)^\gamma, g = 1, \dots, G$$

Step 4 After plugging $\boldsymbol{\theta}^{(p)} = (\theta_1^{(p)}, \dots, \theta_G^{(p)})^T$ from Step 3 into $\tilde{\mathcal{Q}}(\boldsymbol{\beta}, \boldsymbol{\theta}^{(p)})$, minimize $\tilde{\mathcal{Q}}(\boldsymbol{\beta}, \boldsymbol{\theta}^{(p)})$ to obtain $\boldsymbol{\beta}^{(p+1)}$.

Step 5 Repeat Step 2 – Step 4 until $\|\boldsymbol{\beta}^{(p+1)} - \boldsymbol{\beta}^{(p)}\|_1 < 10^{-4}$.

To choose a tuning parameter, we can use the generalized cross validation by following Huang et al.¹³:

$$\frac{\tilde{l}_k(\hat{\beta})}{n\{1 - \hat{d}(\lambda_n)/n\}^2}$$

where $\hat{d}(\lambda_n)$ is the number of non-zero coefficients given λ_n . To estimate the covariance matrix of $\hat{\beta} \neq \mathbf{0}$, we can use a quadratic approximation as in Fan and Li²¹ and Huang et al.¹³ It can be estimated as follows:

$$\left\{ \nabla^2 \tilde{l}(\hat{\beta}) + \Upsilon(\hat{\beta}, \hat{\theta}) \right\}^{-1} \widehat{Cov} \left\{ \nabla \tilde{l}(\hat{\beta}) \right\} \left\{ \nabla^2 \tilde{l}(\hat{\beta}) + \Upsilon(\hat{\beta}, \hat{\theta}) \right\}^{-1}$$

where

$$\Upsilon(\hat{\beta}, \hat{\theta}) = \text{diag} \left\{ \sum_{m \in A_g} \hat{\theta}_m^{1-1/\gamma} c_g^{1/\gamma} \frac{\mathbf{I}(\hat{\beta}_m \neq 0)}{|\hat{\beta}_m| |\tilde{\beta}_m|^\nu}, m = 1, \dots, d_n \right\}$$

$$\hat{\theta}_m = c_g \left(\frac{1-\gamma}{\zeta_n^\gamma} \right)^\gamma \left(\sum_{m \in A_g} \frac{|\hat{\beta}_m|}{|\tilde{\beta}_m|^\nu} \right)^\gamma$$

3 Simulation

We conducted simulations to evaluate the performance of the adaptive group bridge and compare it with the group bridge from Huang et al.¹³ and SCAD of Ni et al.⁷ under two case-cohort studies. The failure time for disease 1, T_{1i} , was generated based on the proportional hazards model. To generate failure time for disease 2, T_{2i} , we used the Clayton–Cuzick model²²

$$F(t_1, t_2 | \mathbf{Z}_{i1}, \mathbf{Z}_{i2}) = \left\{ S_1(t_1; \mathbf{Z}_{i1})^{-1/\eta} + S_2(t_2; \mathbf{Z}_{i2})^{-1/\eta} - 1 \right\}^{-\eta}$$

where $\mathbf{Z}_{i1} = \mathbf{Z}_{i2} = \mathbf{Z}_i$, $S_k(t; \mathbf{Z}_i) = \Pr(T_k > t | \mathbf{Z}_{ik}) = \exp \left\{ - \int_0^t h_{0k}(u) e^{\boldsymbol{\beta}_k^T \mathbf{Z}_k} du \right\}$ is survival function,

$h_{0k}(t)$ and $\boldsymbol{\beta}_k$ ($k = 1, 2$) are the baseline hazard function and the covariate effect for disease k , respectively, and η is the association parameter between the failure times of the two diseases. The relationship between Kendall's tau τ_η and η is $\tau_\eta = 1/(2\eta + 1)$. A larger Kendall's tau represents a higher correlation between T_1 and T_2 . Value of 4 was used for η . The corresponding Kendall's tau value was approximately 0.11. Disease 1, that is, $k = 1$ was of interest.

We considered two censoring rates: 90% and 80%. The censoring distribution was generated independently from the uniform distribution. We assumed the constant baseline hazards:

$h_{01}(t) = 2$ and $h_{02}(t) = 8$. We examined two event rates: 10% and 20% for $k = 1$. The corresponding event rates for $k = 2$ were 20% and 40%, respectively. The control-to-case ratio was set to 1:2. Two sample sizes of the full cohort were considered: $n = 750$ and $n = 1500$. Let n_{case} be the expected number of cases given censoring rate. For each setting, 500 replications were conducted.

To compare the performance of the three variable selection methods, we calculated the group correction rate and the individual correction rate representing the proportion that each variable selection method correctly identified the true non-zero groups and non-zero individual variables of the underlying model, respectively. We also calculated group size (GS) and model size (MS) defined as the average number of non-zero groups and non-zero individual variables selected by each method. The ratio of mean squared errors was defined as the ratio of the mean squared error for each variable selection relative to the mean squared error of the oracle estimator

$$\frac{\sum_{i=1}^{500} \|\hat{\beta}^i - \beta_0\|_2^2}{\sum_{i=1}^{500} \|\hat{\beta}_{Oracle}^i - \beta_0\|_2^2}$$

where $\hat{\beta}_{Oracle}^i$ is the oracle estimator of β_0 at the i th iteration. The oracle estimator was obtained from the pseudo-partial likelihood (1) with either weight (2) or weight (3) assuming we already knew the true non-zero variables. Therefore, the ratio of mean squared errors closer to 1 indicates a better estimation of β_0 .

In the first simulation, we examined group variables consisting of continuous variables. When the event rate was 10% with population size 750, there were five group variables with GSs $(|A_1|, |A_2|, |A_3|, |A_4|, |A_5|)^T = (2, 3, 4, 5, 3)^T$. Groups 2 and 3 were overlapped as follows

$$(\beta_1, \dots, \beta_{15}) = \underbrace{(1.1, 0.9)}_{A_1}, \overbrace{(-1.2, 0, 0, 0, 0.1)}^{A_2}, \underbrace{0.1, 1}_{A_3}, \overbrace{(-1, 0, 0.9, 0, 0)}^{A_4}, \underbrace{0, 0}_{A_5}$$

For the event rate 20% with population size 750 or the event rate 10% with population size 1500, one more zero group with size 2, $(\beta_{16}, \beta_{17})^T = (0, 0)^T$, was added. For the event rate 20% with population size 1500, an additional zero group with size 2, $(\beta_{18}, \beta_{19})^T = (0, 0)^T$, was added. By doing so, we allowed d_n to increase as n_{case} increased. All variables within each group were generated from the multivariate normal distribution with mean 0, variance 1, and correlation 0.5. All groups other than Groups 2 and 3 were assumed to be independent. Thus, the true GS and MS were 3 and 6, respectively. For each scenario, the efficient case-cohort weight (3) and the traditional case-cohort weight (2) were examined.

Table 1 summarizes the results. Table 1 reports the group correction rate (GRC %), individual correction rate (IDC %), GS, MS, and ratio of mean square error (MSER). For all variable selection methods, higher event rates for disease 1 and larger sample sizes produced

higher group correction and individual correction rates; group and MSs closer to 3 and 6, respectively; the MSER closer to 1. The results show that (i) group correction and individual correction rates for the adaptive group bridge are higher than those of the group bridge and SCAD; (ii) the GSs and the MSs for the adaptive group bridge are closer to 3 and 6, respectively, compared to the other two methods. Moreover, all three variable selection methods using the efficient case-cohort weight (3) have better group correction and individual correction rates than those using the traditional case-cohort weight (2), in particular when the event rate for disease 2 is higher and correlation between two failure times is smaller. We also examined different event rates, different pairwise correlations between variables, and a large number of coefficients with a high pairwise correlation between variables. In all scenarios, we had similar results to Table 1. The detailed settings and results of some additional simulation studies are provided in the Supplemental Materials.

In the second simulation, we considered group structure with continuous and categorical variables. For 10% of event rate of disease 1, there were five groups: two groups consisting of categorical variables (A_1, A_2) and three groups consisting of continuous variables (A_3, A_4, A_5). The GSs were $(|A_1|, |A_2|, |A_3|, |A_4|, |A_5|) = (2, 3, 3, 4, 5)$ and overlapping groups were A_3 and A_4 as follows

$$(\beta_1, \dots, \beta_{15}) = (\underbrace{1.1, 0.9}_{A_1}, \underbrace{0, 0, 0}_{A_2}, \underbrace{-1.2, 0, 0}_{A_3}, \underbrace{0, 0}_{A_4}, \underbrace{0.1, -1, 0, 0.9, 0}_{A_5})$$

The categorical variables in GS 2 and 3 were generated from variables with three and four categories, respectively. The reference groups were set to 0. The continuous variables were generated from multivariate normal distribution with mean 0, correlation within group 0.5, correlation between different groups 0. When the event rate for disease 1 was 20% with population size 750 or the event rate for disease 1 was 10% with population size 1500, we added one more group consisting of categorical variables with size 2, $(\beta_{16}, \beta_{17})^T = (0, 0)^T$. When the event rate for disease 1 was 20% with population size 1500, one more group consisting of categorical variables with size 2 such as $(\beta_{18}, \beta_{19})^T = (0, 0)^T$ was added. Table 2 summarizes the results. For the group/individual correction rates, GS, and MS, the adaptive group bridge outperformed the group bridge and SCAD; the efficient case-cohort weight (3) correctly identified the true non-zero variables better than the traditional case-cohort weight (2) in all three methods. The mean square ratio of the adaptive group bridge is always smaller than that of the group bridge in all settings and is comparable with or less than that of SCAD.

We also examined the performance of variable selection when there were continuous variables (Z_{ij} 's) and their squared variables (Z_{ij}^2 's). The detailed settings and simulation results are presented in the Supplemental Materials. The results are similar to Table 1: the adaptive group bridge has better bi-level selection accuracy than the group bridge and SCAD. And the mean group and MSs of the adaptive group bridge are closer to their true

values compared to the other two methods. This suggests that when some of the variables are highly collinear, bi-level selection may be beneficial even if individual variable selection is of interest.

4 Data analysis

We applied the proposed method to analyze the data from the Busselton Health Study.^{8,9} The Busselton Health Study was conducted in the south-west of Western Australia and questionnaires were used every three years from 1966 to 1981 to collect general health information for adult participants. The main aims of this study were to evaluate the association between stroke and serum ferritin effect and to identify risk factors related to stroke. The population consisted of 1612 men and women aged 40–89 who participated in 1981 and were free of coronary heart disease or stroke at that time. The outcome of interest was the time to stroke event, defined as hospital admission, any procedure, or death from stroke, which had followed by 31 December 1998. The time to stroke event was considered censored if subjects did not have an event by the end of study time or lost to follow-up during the study period.

To reduce the cost and preserve the blood sample, the case-cohort study was conducted for stroke. In addition to case-cohort sample for stroke, additional serum ferritin information was obtained from another case-cohort study for coronary heart disease. Under this design, the serum ferritin was measured for all the subjects with coronary heart disease and/or stroke as well as those in the subcohort. The full cohort size and the subcohort size were 1210 and 450, respectively. There were 117 subjects and 55 subjects who had stroke events in the full cohort and the subcohort, respectively. Extra information of serum ferritin was available for 174 subjects who had only coronary heart disease outside the subcohort. The other risk factors included age, gender, body mass index (BMI), blood pressure treatment, cholesterol, triglycerides, hemoglobin and smoking status.

To identify non-zero variables, we used the adaptive group bridge with weights (2) and (3) and compared the results with those from group bridge and SCAD. For continuous variables such as age, BMI, and triglycerides, we considered the following hierarchical structure: (Group 1) each continuous variable (GS 1); (Group 2) each continuous variable and its squared value (GS 2). All variables were centered and standardized. Since triglycerides were severely skewed, log-transformed triglycerides was used before standardization.

Table 3 reports estimated coefficients and their standard errors using the group bridge, adaptive group bridge, and SCAD. The results show that SCAD selected more variables than the group bridge and the adaptive group bridge regardless of the weights. In particular, the standard errors of serum ferritin levels, the square of log-transformed triglycerides, BMI, and diabetes treatment that SCAD identified even with weight (3) were large for their estimates. When using the traditional weight (2), all three methods selected more variables than when using the efficient weight (3). More specifically, compared to the variables selected by the group bridge with weight (3), the group bridge with weight (2) selected four additional variables including ferritin tertile 2, log-transformed triglycerides, its quadratic terms, and diabetes treatment. The standard errors of those four variables' estimates were

relatively large for their estimates. On the other hand, the adaptive group bridge with weight (2) additionally selected diabetes treatment only compared to the variables selected by the adaptive group bridge with weight (3). Both the group bridge and the adaptive group bridge using the efficient weight (3) selected the same three variables: age, blood pressure treatment, and sex. Therefore, increased age, blood pressure treatment, and female were associated with stroke event.

We compared the model errors of the three methods using a fivefold cross-validation evaluation. Following Huang et al.,¹³ we considered the following model error:

$ME(\hat{\beta}) = E\left\{\exp(-\hat{\beta}^T Z) - \exp(-\beta_0^T Z)\right\}^2$. We estimated the model error for case-cohort data as follows

$$\widehat{ME}(\hat{\beta}) = \frac{1}{n} \sum_{i=1}^n \left\{ \Delta_{ik} + (1 - \Delta_{ik}) \xi_i \tilde{\alpha} \right\} \left\{ \exp(-\hat{\beta}^T Z_i) - \exp(-\beta_0^T Z_i) \right\}^2$$

For the traditional weight (2), the estimated model errors for the adaptive group bridge, the group bridge, and SCAD were 1.16, 2.35, and 11.05, respectively. When the efficient weight (3) was used, the estimated model errors for the three methods were smaller than those using the traditional weight (2): 0.18, 0.63, and 1.85 for the adaptive group bridge, the group bridge, and SCAD, respectively. For both weights, the adaptive group bridge had the smallest model error, which indicates the adaptive group bridge had a better prediction than the other two methods.

5 Discussion

We proposed the adaptive group bridge for case-cohort data and studied its asymptotic properties. The simulation studies and the Busselton Health data example showed the adaptive group bridge was superior to the group bridge and SCAD in terms of variable selection and prediction. The objective function of the proposed penalized proportional hazards model with the adaptive group bridge is non-convex. To minimize the non-convex objective function, we proposed a coordinate decent algorithm using the quadratic approximation of the pseudo-likelihood function and the optimization scheme for the adaptive L_1 penalty.

The proposed method is limited to when $d_n < n$. Studying bi-level selection when $d_n > n$ would be an important future research problem. One way to deal with this problem is screening. The two-stage selection procedure may be considered: we screen group variables in the first stage and then we make the parsimonious list of non-zero variables using the adaptive group bridge in the second stage.

Making an inference using the selected variables only may not be valid because hypotheses are generated from the data and not pre-specified. Recently post-selection inference for LASSO got much attention.^{23–25} Developing a method to conduct valid inference after model selection would be an important research problem.

In this paper, we have considered the event-specific model. The joint analysis should be conducted when an investigator compares the risk effects on different diseases. Developing a variable selection method for multivariate failure time models under multiple case-cohort studies would be another interesting future research problem.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank Professor Matthew Knuiman and the Busselton Population Medical Research Foundation for permission to use their data.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported in part by Institutional Research Grant IRG #16–183-31 from the American Cancer Society and the MCW Cancer Center, and the United States National Cancer Institute (U24CA076518).

References

1. Prentice R A case-cohort design for epidemiologic cohort studies and disease prevention trials. *Biometrika* 1986; 73: 1–11.
2. Self SG and Prentice RL. Asymptotic distribution theory and efficiency results for case-cohort studies. *Ann Statist* 1988; 34: 103–119.
3. Barlow W Robust variance estimation for the case-cohort design. *Biometrics* 1994; 50: 1064–1072. [PubMed: 7786988]
4. Kulich M and Lin DY. Improving the efficiency of relative-risk estimation in case-cohort study. *J Am Statist Assoc* 2004; 99: 832–844.
5. Kang S and Cai J. Marginal hazard model for case-cohort studies with multiple disease outcomes. *Biometrika* 2009; 96: 887–901. [PubMed: 23946547]
6. Kim S, Cai J and Lu W. More efficient estimators for case-cohort studies. *Biometrika* 2013; 100: 695–708. [PubMed: 24634519]
7. Ni A, Cai J and Zeng D. Variable selection for case-cohort studies with failure time outcome. *Biometrika* 2016; 103: 547–562. [PubMed: 28529347]
8. Cullen KJ. Mass health examinations in the Busselton population, 1996 to 1970. *Aust J Med* 1972; 2: 714–718.
9. Knuiman MW, Divitini ML, Olynyk JK, et al. Serum ferritin and cardiovascular disease: a 17-year following-up study in Busselton, Western Australia. *Am J Epidemiol* 2003; 158: 144–149. [PubMed: 12851227]
10. Cai J, Fan J, Li R, et al. Variable selection for multivariate failure time data. *Biometrika* 2005; 92: 303–316. [PubMed: 19458784]
11. Ma S, Song X and Huang J. Supervised group lasso with applications to microarray. *BMC Bioinformatics* 2007; 3: 60.
12. Kim J, Sohn I, Jung S, et al. Analysis of survival data with group lasso. *Comm Statist Simulation Comput* 2012; 41: 1593–1605.
13. Huang J, Li L, Liu Y, et al. Group selection in the cox model with a diverging number of covariates. *Statist Sin* 2014; 24: 1787–1810.
14. Wang S and Nan B. Hierarchically penalized cox regression with grouped variables. *Biometrika* 2009; 96: 307–322.
15. Ahn KW, Banerjee A, Sahr N, et al. Group and within-group variable selection for competing risks data. *Lifetime Data Analysis* 2018; 24, 407–424.

16. Cox DR. Regression models and life-tables (with discussion). *J R Statist Soc B* 1972; 34: 187–220.
17. Kalbfleisch JD and Lawless JF. Likelihood analysis of multistate models for disease incidence and mortality. *Statist Med* 1988; 7: 149–160.
18. Borgan O, Langholz B, Samuelsen SO, et al. Exposure stratified case-cohort designs. *Lifetime Data Anal* 2000; 6: 39–58. [PubMed: 10763560]
19. Huang J, Ma S, Xie H, et al. A group bridge approach for variable selection. *Biometrika* 2009; 96: 339–355. [PubMed: 20037673]
20. Zou H The adaptive lasso and its oracle properties. *J Am Statist Assoc* 2006; 101: 1418–1429.
21. Fan J and Li R. Variable selection for cox’s proportional hazards model and frailty properties. *J Am Statist Assoc* 2002; 30: 74–99.
22. Clayton D and Cuzick J. Multivariate generalizations of the proportional hazards model (with discussion). *J R Statist Soc A* 1985; 148: 82–117.
23. Lee J, Sun D, Sun Y, et al. Exact post-selection inference, with application to the lasso. *Ann Stat* 2016; 44: 907–927.
24. Lockhart R, Jonathan Taylor J, Tibshirani R, et al. A significance test for the lasso. *Ann Stat* 2014; 42: 413–468. [PubMed: 25574062]
25. Tibshirani R, Taylor J, Lockhart R, et al. Exact post-selection inference for sequential regression procedures. *J Am Statist Assoc* 2016; 111: 600–620.

Table 1.

Simulation results for group variables consisting of continuous variables.

n	Weight	$P(\beta_1, \beta_2)$	d_n	Method	GRC%	IDC%	GS	MS	MSER
750	Traditional case-cohort weight	(0.1,-)	15	AGB	27.8	14.2	4.09	9.02	2.22
				GB	33.2	5.2	3.96	9.53	2.02
				SCAD	8.8	8.2	4.27	8.67	2.43
		(0.2,-)	17	AGB	73.4	59.6	3.32	6.73	1.46
				GB	73.2	21.8	3.32	7.61	1.72
				SCAD	38.8	38.2	3.68	7.05	1.69
	Efficient case-cohort weight	(0.1,0.2)	15	AGB	54.2	32.8	3.58	7.48	1.64
				GB	52.8	12.2	3.56	8.36	1.70
				SCAD	17.6	16.6	4.06	8.01	2.07
		(0.2,0.4)	17	AGB	83.0	71.2	3.20	6.48	1.36
				GB	79.4	23.4	3.23	7.35	1.72
				SCAD	47.4	46.4	3.56	6.79	1.50
1500	Traditional case-cohort weight	(0.1,-)	17	AGB	34.4	20.0	4.12	8.68	2.02
				GB	31.2	4.4	4.18	9.85	2.00
				SCAD	12.0	11.8	4.37	8.74	2.18
		(0.2,-)	19	AGB	83.8	76.0	3.22	6.45	1.30
				GB	73.0	18.8	3.33	7.56	1.70
				SCAD	66.0	65.6	3.39	6.57	1.39
	Efficient case-cohort weight	(0.1,0.2)	17	AGB	56.6	41.2	3.69	7.67	1.75
				GB	51.4	10.8	3.69	8.54	1.84
				SCAD	34.8	33.8	3.89	7.63	1.82
		(0.2,0.4)	19	AGB	92.8	85.2	3.10	6.22	1.24
				GB	82.8	23.4	3.19	7.28	1.74
				SCAD	75.4	74.8	3.26	6.34	1.23

GRC: group correction; IDC: individual correction; GS: group size; MS: model size; MSER: mean square error ratio; AGB: adaptive group bridge; GB: group bridge; SCAD: smoothly clipped absolute deviation.

Table 2.

Simulation results for group variables consisting of continuous and categorical variables.

n	Weight	$P(1, 2)$	d_n	Method	GRC%	IDC%	GS	MS	MSER
750	Traditional case-cohort weight	(0.1,-)	15	AGB	25.2	9.8	4.08	8.88	2.06
				GB	26.8	2.0	3.95	9.46	2.06
				SCAD	10.6	8.6	4.31	9.00	2.20
		(0.2,-)	17	AGB	67.2	51.2	3.40	6.87	1.40
				GB	68.2	14.2	3.38	7.82	1.71
				SCAD	33.6	33.4	3.73	7.10	1.47
	Efficient case-cohort weight	(0.1,0.2)	15	AGB	46.0	24.0	3.69	7.75	1.65
				GB	48.8	6.0	3.61	8.52	1.76
				SCAD	17.0	15.6	4.07	8.25	1.89
		(0.2,0.4)	17	AGB	78.2	64.4	3.26	6.55	1.31
				GB	77.4	15.8	3.25	7.53	1.67
				SCAD	43.4	42.8	3.61	6.86	1.36
1500	Traditional case-cohort weight	(0.1,-)	17	AGB	25.6	12.2	4.40	9.34	1.87
				GB	21.2	1.4	4.36	10.18	1.91
				SCAD	15.0	14.6	4.45	8.90	1.88
		(0.2,-)	19	AGB	78.4	65.4	3.27	6.57	1.37
				GB	72.2	15.4	3.36	7.68	1.78
				SCAD	61.8	61.4	3.43	6.61	1.34
	Efficient case-cohort weight	(0.1,0.2)	17	AGB	46.4	28.4	3.79	7.84	1.53
				GB	46.4	5.2	3.79	8.71	1.71
				SCAD	30.0	29.4	3.98	7.76	1.53
		(0.2,0.4)	19	AGB	84.0	74.4	3.19	6.38	1.34
				GB	82.8	20.8	3.21	7.41	1.91
				SCAD	68.6	68.4	3.33	6.42	1.24

GRC: group correction; IDC: individual correction; GS: group size; MS: model size; MSER: mean square error ratio; AGB: adaptive group bridge; GB: group bridge; SCAD: smoothly clipped absolute deviation.

Table 3.

Estimated coefficients and standard errors for the Busselton Health Study Data.

Variable	Traditional case cohort weight (2)			Efficient case cohort weight (3)		
	AGB $\hat{\beta}$ (SE)	GB $\hat{\beta}$ (SE)	SCAD $\hat{\beta}$ (SE)	AGB $\hat{\beta}$ (SE)	GB $\hat{\beta}$ (SE)	SCAD $\hat{\beta}$ (SE)
Ferritin tertile (ref = 1)						
Ferritin tertile 2	0(–)	0(–)	0.04 (0.15)	0(–)	0(–)	0.11 (0.15)
Ferritin tertile 3	0(–)	0.09 (0.11)	0.14 (0.13)	0(–)	0(–)	0.11 (0.13)
Age	0.86 (0.13)	0.91 (0.13)	0.88 (0.13)	0.85 (0.13)	0.83 (0.12)	0.90 (0.13)
Age ²	0(–)	0(–)	0(–)	0(–)	0(–)	0(–)
BMI	0(–)	0(–)	0(–)	0(–)	0(–)	0.04 (0.13)
BMI ²	0(–)	0(–)	0(–)	0(–)	0(–)	0(–)
Cholesterol	0(–)	0(–)	0(–)	0(–)	0(–)	0(–)
Cholesterol ²	0(–)	0(–)	0(–)	0(–)	0(–)	0(–)
log(TR)	0(–)	–0.17 (0.25)	0(–)	0(–)	0(–)	0(–)
log(TR) ²	0(–)	0.21 (0.21)	0.03 (0.13)	0(–)	0(–)	0.05 (0.12)
Diabetes TRT	–0.17 (0.1)	–0.16 (0.1)	–0.13 (0.09)	0(–)	0(–)	–0.05 (0.08)
Blood Pressure TRT	0.28 (0.09)	0.3 (0.1)	0.31 (0.10)	0.26 (0.09)	0.25 (0.09)	0.30 (0.10)
Sex(1 = female)	–0.33 (0.11)	–0.31 (0.12)	–0.32 (0.12)	–0.27 (0.11)	–0.26 (0.11)	–0.25 (0.11)
Smoking(ref = Never)						
Former	0(–)	0(–)	0(–)	0(–)	0(–)	0(–)
Current	0(–)	0(–)	0(–)	0(–)	0(–)	0(–)

TR: triglycerides; TRT: treatment; AGB: adaptive group bridge; GB: group bridge; SCAD: smoothly clipped absolute deviation; BMI: body mass index.