Statistics and Its Interface Volume 16 (2023) 459–474

Quadratic upper bound algorithms for estimation under Cox model in case-cohort studies

Jieli Ding, Jiaqian Zhang, Yanqin Feng*, and Yuxuan Du

A case-cohort design is a cost-effective biased-sampling scheme in large cohort studies. Implementation of parameter estimators for case-cohort data requires numerical approaches. Using the minorization-maximization principle, which is a versatile tool for constructing optimization algorithms, we develop two quadratic-upper-bound algorithms for estimations in the Cox model under case-cohort design. The proposed algorithms are monotonic and reliably converge to the weighted estimators considered. These algorithms involve the inversion of the derived upper-bound matrix only one time in the whole process, and the upper-bound matrix is independent of both parameter and weight functions. These features make the proposed algorithms have simple update and low per-iterative cost, especially to large-dimensional problems. We conduct simulation studies and real data examples to illustrate the performance of the proposed algorithms, and compare them to Newton's method.

Keywords and phrases: Case-cohort design, Minorization maximization algorithm, Quadratic upper bounds, Cox model, Estimating equations.

1. INTRODUCTION

In epidemiologic and biomedical observational studies that relate disease occurrence to individual exposures, it is likely to follow up a large number of subjects for a long time period. Major budgets and costs typically arise from the assembling of expensive covariates. Therefore, development of a cost-effective design which can enhance the efficiency while reducing the cost is always desirable in practice. For time-to-event data with censoring, case-cohort design is one of the most widely used biased-sampling schemes. The key idea of the case-cohort design is to assemble the measurements of expensive exposures only on a subset of the entire cohort (subcohort) and all the subjects who experience the event of interest (cases). After the landmark article of Prentice (1986), extensive researches on case-cohort design have been developed, mainly along two lines, likelihood-based approaches (Self and Prentice, 1988; Chen and Lo, 1999; Kang et al., 2018; Liu et al., 2018) and estimating-equation-based approaches (Kulich and Lin, 2000; Sun et al., 2004; Cai and Zeng, 2004, 2007; Kulich and Lin, 2004; Kong et al., 2004; Lu and Tsiatis, 2006; Breslow and Wellner, 2007; Kang et al., 2013; Steingrimsson and Strawderman, 2017, etc). Recent works have interests on case-cohort studies with multivariate failure times (Kang and Cai, 2009; Kang et al., 2013; Yan et al., 2017; Kim et al, 2018; Maitra et al, 2020).

When statistical methods for parameter estimation are proposed, the numerical calculation of estimates is often involved, which is important especially in practice. Newton-Raphson algorithm is one of the most widely used numerical algorithms, which has many advantages, including quadratic convergence rate around the maximum point. However, Newton-Raphson algorithm will also encounter some problems in actual calculation, especially in large-dimensional or high-dimensional situations. Calculation of the inverse matrix of the information matrix, which is conducted in each iterative step, may have high computational cost or even fail, and the method of Newton is lack of the monotonicity under some situations. For large-dimensional or high-dimensional cases, numerous literatures have developed the methods of variable selection to achieve the purpose of dimension reduction (Tibshirani, 1996, 1997; Fan and Li, 2001, 2002; Zou, 2006; Huang et al, 2008, 2010, etc). However, most of these methods require sparsity assumption. Under the cases that data does not meet the sparsity requirement or the dimension is still large even after the variables are selected, it is desirable to develop some new algorithms.

Recently, a minorization-maximization (MM) algorithm has been applied widely. The MM algorithm is a principle for creating algorithm rather than a single algorithm, whose essential idea is to create a surrogate function with computational superiorities over the objective function in order to achieve optimization transformation (De Pierro, 1995; Becker et al., 1997; Lange et al., 2000; Hunter and Lange, 2002, 2004; Lange, 2004, 2010, etc). For survival data from simple random sampling, Böhning and Linday (1988) developed a quadratic lower bound algorithm for Cox model. Ding et al. (2015) developed a modified MM algorithm for Cox model with parameter constraints. For survival data from case-cohort design, Deng et al. (2018) studied an MM algorithm for a constrained estimator of parameter in Cox model.

In this paper, we develop two new quadratic upper-bound algorithms for implementation of estimators of regression parameter in Cox model under the case-cohort design. We first find two global upper-bound matrices on the

*Corresponding author.

observed information matrix, which is derived from a series of weighted estimating equations proposed for the parameter inference. In the spirit of the MM principle, we the establish two quadratic upper-bound (QUB) algorithms based on these upper-bound matrice and obtain the convergence of the proposed algorithms. One novelty of our QUB algorithm is that both the algorithms only need to calculate the inverse matrix of the corresponding upper-bound matrix one time in the whole process, instead of calculating the inverse matrix of the observed information matrix in each iterative step in Newton-Raphson algorithm. The other novelty is that the upper-bound matrice used in the proposed algorithms are independent of both parameter and weight functions, which makes these QUB algorithms universal for calculation of a class of weighted estimators regardless of the weight selections in case-cohort studies, with simple update and low per-iteration cost. Furthermore, simulation studies and real data examples suggest that the proposed QUB algorithms work stably and efficiently. What needs to be pointed out in particular is that the proposed QUB algorithms perform well in nonsparse or large-dimensional cases.

The rest of this paper is organized as follows. In Section 2, we introduce a general weighted estimating equation approach for parameter inference in the Cox model under case-cohort design. In Section 3, we propose two quadratic upper-bound algorithms and obtain some theoretical properties. In Section 4, we conduct simulation studies to evaluate the practical behavior of the proposed algorithms. In Section 5, we analyze two real data examples from a Wilm tumor study and a SEER breast cancer study.

2. ESTIMATIONS UNDER CASE-COHORT DESIGN

Let $\widetilde{T}$ be the potential failure time and C be the censoring time. The observed time is $T = \min(\widetilde{T},\ C)$ and the right censoring indicator is $\Delta = I(\widetilde{T} \leq C)$, where $I(\cdot)$ is the indicator function. Denote $\boldsymbol{Z}$ to be the p-vector covariates. It is assumed that $\widetilde{T}$ and C are conditionally independent given $\boldsymbol{Z}$. The most widely used model for survival data is the proportional hazards model (Cox, 1972) with the hazard function:

$$\lambda(t|\boldsymbol{Z}) = \lambda_0(t)\exp\{\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{Z}\}, \tag{1}$$

where $\lambda_0(t)$ is the unspecified baseline hazard function, and $\boldsymbol{\beta}$ is the p-vector parameter of primary interest.

Suppose that there are n independent subjects in the underlying population, and $(T_i,\ \Delta_i,\ \boldsymbol{Z}_i)$, $i = 1, \cdots, n$, are n independent copies of $(T,\ \Delta,\ \boldsymbol{Z})$. Under the case-cohort design, a subset of size $\widetilde{n}$, referred to as the subcohort, is selected from the full cohort via simple random sampling without replacement. Let ξ_i be the subcohort sampling indicator for the ith subject, taking the value 1 or 0, whether the ith subject is included in the subcohort or not. Assume $\mathrm{P}(\xi_i = 1) = \widetilde{p} = \widetilde{n}/n$, which is the probability of being sampled into the subcohort for each subject. Covariate measurements are taken only on the subcohort members and all the cases outside the subcohort. Thus, the observable information under the case-cohort design is $(T_i,\ \Delta_i,\ \xi_i,\ \boldsymbol{Z}_i)$ when $\xi_i = 1$ or $\Delta_i = 1$, and $(T_i,\ \Delta_i,\ \xi_i)$ when $\xi_i = 0$ and $\Delta_i = 0$.

If the full cohort data were available, the estimation of the true parameter $\boldsymbol{\beta}_0$ could be obtained by solving the following estimating equation,

$$\begin{aligned} &U_F(\boldsymbol{\beta}) \\ &= \sum_{i=1}^{n}\Delta_i\left[\boldsymbol{Z}_i - \frac{\sum_{j=1}^{n} I(T_j \geq T_i)\exp\{\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{Z}_j\}\boldsymbol{Z}_j}{\sum_{j=1}^{n} I(T_j \geq T_i)\exp\{\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{Z}_j\}}\right] = 0. \end{aligned} \tag{2}$$

For data from case-cohort studies, since $\boldsymbol{Z}_i$'s are not available for cohort members outside the case-cohort sample, (2) cannot be calculated. For the inference of $\boldsymbol{\beta}$ under the case-cohort design, we adopt a general weighted estimating equation,

$$\begin{aligned} &U_w(\boldsymbol{\beta}) \\ &= \sum_{i=1}^{n}\Delta_i\left[\boldsymbol{Z}_i - \frac{\sum_{j=1}^{n} w_j(T_i) I(T_j \geq T_i)\exp\{\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{Z}_j\}\boldsymbol{Z}_j}{\sum_{j=1}^{n} w_j(T_i) I(T_j \geq T_i)\exp\{\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{Z}_j\}}\right] = 0, \end{aligned} \tag{3}$$

where $w_j(t)$ is a possibly time-varying weight function which is introduced to "unbias" the biased-sampling nature of the case-cohort design. For example, motivated by inversely weighting the incomplete observations (Horvitz and Thompson, 1951; Kalbfleisch and Lawless, 1988), the weight function can take the time-independent form as,

$$w_i = \Delta_i + \frac{(1-\Delta_i)\xi_i}{\widetilde{p}}. \tag{4}$$

In particular, the weight for a case is 1 regardless of their subcohort membership and for a censored subject in the subcohort is $\widetilde{p}^{-1}$.

We can also consider a time-varying version of weight function (Barlow, 1994; Borgan et al, 2000; Kulich and Lin, 2004):

$$w_i(t) = \Delta_i + \frac{(1-\Delta_i)\xi_i}{\widehat{p}(t)}, \tag{5}$$

where

$$\widehat{p}(t) = \frac{\sum_{i=1}^{n}(1-\Delta_i)\xi_i I(T_i \geq t)}{\sum_{i=1}^{n}(1-\Delta_i) I(T_i \geq t)}.$$

This weight function is defined to be equal to 1 for the cases and to $\widehat{p}(t)^{-1}$ for the sampled censored subjects, where $\widehat{p}(t)$ is the estimator of the true sampling probability $\widetilde{p}$ and denotes the number of sampled censored subjects divided by the number of censored subjects remaining in the risk set at time t. These weighted estimating functions have also been considered in the multivariate failure time context for case-cohort studies (Kang and Cai, 2009).

Under the case-cohort design, the estimator of $\boldsymbol{\beta}$, $\widehat{\boldsymbol{\beta}}_w$ is defined as the solution to $U_w(\boldsymbol{\beta}) = 0$. Then, using the similar arguments in Kang and Cai (2009) for the case of $k = 1$, $\widehat{\boldsymbol{\beta}}_w$ can be theoretically shown to be consistent and asymtotically normal.

Since the solution to this series of weighted estimating equations $U_w(\boldsymbol{\beta}) = 0$ has no closed expression, the calculation of the estimator $\widehat{\boldsymbol{\beta}}_w$ requires numerical calculation methods. The Newton-Raphson algorithm is the most popular approach to find the solution via the iterations of the form:

$$\boldsymbol{\beta}^{(m+1)} = \boldsymbol{\beta}^{(m)} + \left[H_w(\boldsymbol{\beta}^{(m)})\right]^{-1} U_w(\boldsymbol{\beta}^{(m)}), \tag{6}$$

where $H_w(\boldsymbol{\beta}) = -\nabla_{\boldsymbol{\beta}} U_w(\boldsymbol{\beta})$ is the observed information matrix and $\nabla_{\boldsymbol{\beta}}$ denotes the partial derivative with respect to $\boldsymbol{\beta}$, and $\boldsymbol{\beta}^{(m)}$ is the mth iteration of $\boldsymbol{\beta}$ in the algorithm. However, several complications can compromise the performance of Newton-Raphson algorithm in case-cohort studies, especially in large-dimensional situations: (a) the information matrix $H_w(\boldsymbol{\beta})$ may fail to be computationally invertible; (b) calculating the information matrix $H_w(\boldsymbol{\beta})$ and solving the linear system $H_w(\boldsymbol{\beta})x = U_w(\boldsymbol{\beta})$ may be expensive or time-consuming; and (c) the Newton's method is lack of the monotonicity under some situations. Therefore, we are committed to finding some new algorithms for efficiently solving such problems.

3. QUB ALGORITHM

To derive new algorithms to solve a class of weighted estimating equations in case-cohort studies, we apply the minorization-maximization (MM) principle, which is a principle for creating algorithm rather than a single algorithm. Specifically, we consider the optimization problem $\max_{\boldsymbol{\beta} \in \mathcal{B}} l(\boldsymbol{\beta})$, where $l(\boldsymbol{\beta})$ is the objective function and $\mathcal{B}$ is the parametric space of $\boldsymbol{\beta}$. An MM principle involves minorizing $l(\boldsymbol{\beta})$ by a surrogate function $Q(\boldsymbol{\beta}|\boldsymbol{\beta}^{(m)})$ anchored at the current iteration $\boldsymbol{\beta}^{(m)}$ of a search. The surrogate function should satisfy the two properties:

$$\begin{aligned} & l(\boldsymbol{\beta}^{(m)}) = Q(\boldsymbol{\beta}^{(m)}|\boldsymbol{\beta}^{(m)}), \\ & l(\boldsymbol{\beta}) \geq Q(\boldsymbol{\beta}|\boldsymbol{\beta}^{(m)}), \qquad \boldsymbol{\beta} \neq \boldsymbol{\beta}^{(m)}. \end{aligned} \tag{7}$$

Construction of the surrogate function $Q(\boldsymbol{\beta}|\boldsymbol{\beta}^{(m)})$ constitutes the first M-step of an MM algorithm.

In the second M-step of the algorithm, we maximize the surrogate $Q(\boldsymbol{\beta}|\boldsymbol{\beta}^{(m)})$ instead of $l(\boldsymbol{\beta})$. Define $\boldsymbol{\beta}^{(m+1)} = \max_{\boldsymbol{\beta} \in \mathcal{B}} Q(\boldsymbol{\beta}|\boldsymbol{\beta}^{(m)})$. Then this action forces the following ascent property:

$$\begin{aligned} & l(\boldsymbol{\beta}^{(m+1)}) \\ & \quad = \left[l(\boldsymbol{\beta}^{(m+1)}) - Q(\boldsymbol{\beta}^{(m+1)}|\boldsymbol{\beta}^{(m)})\right] + Q(\boldsymbol{\beta}^{(m+1)}|\boldsymbol{\beta}^{(m)}) \\ & \quad \geq \left[l(\boldsymbol{\beta}^{(m)}) - Q(\boldsymbol{\beta}^{(m)}|\boldsymbol{\beta}^{(m)})\right] + Q(\boldsymbol{\beta}^{(m)}|\boldsymbol{\beta}^{(m)}) \\ & \quad = l(\boldsymbol{\beta}^{(m)}), \end{aligned}$$

which guarantees that optimizing $l(\boldsymbol{\beta})$ is equivalent to iteratively maximizing the surrogate $Q(\boldsymbol{\beta}|\boldsymbol{\beta}^{(m)})$.

Back to our problem of solving the weighted estimating equation in (3), we introduce a "working" likelihood function as follows:

$$\begin{aligned} & \ell_w(\boldsymbol{\beta}) \\ & \quad = \sum_{i=1}^{n} \Delta_i \left[\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{Z}_i - \log \sum_{j=1}^{n} w_j(T_i) I(T_j \geq T_i) \exp\{\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{Z}_j\}\right], \end{aligned} \tag{8}$$

which satisfies that $\nabla_{\boldsymbol{\beta}} \ell_w(\boldsymbol{\beta}) = U_w(\boldsymbol{\beta})$ for $\boldsymbol{\beta} \in \mathcal{B}$. Therefore, finding the solution to the equation $U_w(\boldsymbol{\beta}) = 0$ can be transferred to an optimization problem $\max_{\boldsymbol{\beta} \in \mathcal{B}} \ell_w(\boldsymbol{\beta})$. The key in devising an MM algorithm revolves around how to choose a good surrogate function, and then transfer the optimization of $\ell_w(\boldsymbol{\beta})$ to such a surrogate function.

Due to the fact that $\ell_w(\boldsymbol{\beta})$ is twice continuously differentiable, we obtain the Taylor expansion of $\ell_w(\boldsymbol{\beta})$ around $\boldsymbol{\beta}^{(m)}$ to be

$$\begin{aligned} \ell_w(\boldsymbol{\beta}) = {} & \ell_w(\boldsymbol{\beta}^{(m)}) + U_w(\boldsymbol{\beta}^{(m)})^{\mathrm{T}}(\boldsymbol{\beta} - \boldsymbol{\beta}^{(m)}) \\ & - \frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\beta}^{(m)})^{\mathrm{T}} H_w(\widetilde{\boldsymbol{\beta}})(\boldsymbol{\beta} - \boldsymbol{\beta}^{(m)}), \end{aligned}$$

where $H_w(\boldsymbol{\beta}) = -\nabla_{\boldsymbol{\beta}} U_w(\boldsymbol{\beta}) = -\nabla^2_{\boldsymbol{\beta}} \ell_w(\boldsymbol{\beta})$ and $\widetilde{\boldsymbol{\beta}}$ lies in the line between $\boldsymbol{\beta}$ and $\boldsymbol{\beta}^{(m)}$. Inspirited by the idea of quadratic approximation of Böhning and Linday (1988), we propose a minorizing function as follows:

$$\begin{aligned} Q_w(\boldsymbol{\beta} \,|\, \boldsymbol{\beta}^{(m)}) = {} & \ell_w(\boldsymbol{\beta}^{(m)}) + U_w(\boldsymbol{\beta}^{(m)})^{\mathrm{T}}(\boldsymbol{\beta} - \boldsymbol{\beta}^{(m)}) \\ & - \frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\beta}^{(m)})^{\mathrm{T}}\, \mathbf{B}\, (\boldsymbol{\beta} - \boldsymbol{\beta}^{(m)}), \end{aligned} \tag{9}$$

where $\mathbf{B}$ is some positive definite matrix not depending on $\boldsymbol{\beta}$ and satisfies that

$$H_w(\boldsymbol{\beta}) \leq \mathbf{B}, \tag{10}$$

which means $\mathbf{B} - H_w(\boldsymbol{\beta})$ is a non-negative definite matrix, for all $\boldsymbol{\beta} \in \mathcal{B}$. That is, $\mathbf{B}$ is a global upper-bound of $H_w(\boldsymbol{\beta})$. It is easy to show that the minorizing function $Q_w(\boldsymbol{\beta} \,|\, \boldsymbol{\beta}^{(m)})$ satisfies the two properties in (7). By transferring the optimization of $\ell_w(\boldsymbol{\beta})$ to iteratively maximizing the surrogate

$Q_w(\boldsymbol{\beta} \,|\, \boldsymbol{\beta}^{(m)})$, we can find the solution to $U_w(\boldsymbol{\beta}) = 0$ via the new updates:

$$\boldsymbol{\beta}^{(m+1)} = \boldsymbol{\beta}^{(m)} + \mathbf{B}^{-1} U_w(\boldsymbol{\beta}^{(m)}). \tag{11}$$

In the spirit of Böhning and Linday (1988), we call the above algorithm to be the quadratic upper bound (QUB) algorithm. We first present the convergence of such a QUB algorithm in the following theorem and give the proof in the Appendix.

Theorem 1 (Convergence of QUB Algorithm)**.** *Let $\widehat{\boldsymbol{\beta}}$ be the solution to the weighted equation $U_w(\boldsymbol{\beta}) = 0$. Suppose that $\boldsymbol{\beta}^{(0)} \in \mathcal{B}$ and $\boldsymbol{\beta}^{(m)}$ is the sequence obtained iteratively by the iterations* (11)*. Under some regularity conditions in the Appendix, $\boldsymbol{\beta}^{(m)}$ converges to $\widehat{\boldsymbol{\beta}}$ as $m \to \infty$.*

The proposed QUB algorithm has several advantages: (i) the proposed algorithm obviates the calculation of the inverse matrix of the information matrix which may be computationally singular; (ii) instead of calculating the inverse matrix of the information matrix in each iterative step, the algorithm only needs to calculate the inverse matrix of the upper-bound matrix one time in the whole process; (iii) the algorithm is monotonic according to the proof of the convergence. Although the QUB algorithm has at best a linear rate of convergence, its update is very simple and its per-iteration-cost is relatively low. These can trip the computational balance in its favor.

Under this idea, the key technique of our proposed algorithm lies in finding the upper-bound matrix $\mathbf{B}$. In the following, we find two global upper bound matrices for $H_w(\boldsymbol{\beta})$. We provide the main results below and give the proofs in the Appendix.

Theorem 2.

$$H_w(\boldsymbol{\beta}) \leq \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \Delta_i I(T_j \geq T_i) \boldsymbol{Z}_j^{\otimes 2} \triangleq \mathbf{B}_1. \tag{12}$$

where $a^{\otimes 2} = aa'$ for a vector a.

Note that the upper-bound matrix $\mathbf{B}_1$ is independent of not only the parameter $\boldsymbol{\beta}$ but also the weight function $w_i(t)$. This means that the proposed QUB algorithm can be used to solve a series of weighted estimating equations under the case-cohort design regardless of the weight selections.

Algorithm 1 QUB Algorithm for Cox Model under Case-Cohort Design

Input: $\boldsymbol{\beta}^{(0)}$, $\varepsilon > 0$.

Calculate $\mathbf{B}_1 = \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \Delta_i I(T_j \geq T_i) \boldsymbol{Z}_j^{\otimes 2}$;
while $\left\| \boldsymbol{\beta}^{(m+1)} - \boldsymbol{\beta}^{(m)} \right\| > \varepsilon$ **do**
  Suppose the current estimator $\boldsymbol{\beta}^{(m)}$ with $m \geq 0$
  Do
    Calculate $U_w(\boldsymbol{\beta}^{(m)})$;
    Update $\boldsymbol{\beta}^{(m+1)} = \boldsymbol{\beta}^{(m)} + \mathbf{B}_1^{-1} U_w(\boldsymbol{\beta}^{(m)})$;
  end
end while

Output: $\widehat{\boldsymbol{\beta}} = \boldsymbol{\beta}^{(m+1)}$.

In the process of deriving the upper bound matrix $\mathbf{B}_1$, we find out that the inequality used can be further improved to capture the invariance of location that the informative matrix processes and reduce the influence of the choices of the initial values to Algorithm 1. We provide the improved upper-bound matrix in the theorem as follows.

Theorem 3.

$$H_w(\boldsymbol{\beta}) \leq \frac{1}{2} \sum_{i=1}^{n} \Delta_i \left[\sum_{j=1}^{n} I(T_j \geq T_i) \boldsymbol{Z}_j^{\otimes 2} - \frac{1}{n_i} \left\{ \sum_{j=1}^{n} I(T_j \geq T_i) \boldsymbol{Z}_j \right\}^{\otimes 2} \right] \triangleq \mathbf{B}_2. \tag{13}$$

Algorithm 2 QUB Algorithm for Cox Model under Case-Cohort Design

Input: $\boldsymbol{\beta}^{(0)}$, $\varepsilon > 0$.

Calculate

$$\mathbf{B}_2 = \frac{1}{2} \sum_{i=1}^{n} \Delta_i \left[\sum_{j=1}^{n} I(T_j \geq T_i) \boldsymbol{Z}_j^{\otimes 2} - \frac{1}{n_i} \left\{ \sum_{j=1}^{n} I(T_j \geq T_i) \boldsymbol{Z}_j \right\}^{\otimes 2} \right];$$

while $\left\| \boldsymbol{\beta}^{(m+1)} - \boldsymbol{\beta}^{(m)} \right\| > \varepsilon$ **do**
  Suppose the current estimator $\boldsymbol{\beta}^{(m)}$ with $m \geq 0$
  Do
    Calculate $U_w(\boldsymbol{\beta}^{(m)})$;
    Update $\boldsymbol{\beta}^{(m+1)} = \boldsymbol{\beta}^{(m)} + \mathbf{B}_2^{-1} U_w(\boldsymbol{\beta}^{(m)})$;
  end
end while

Output: $\widehat{\boldsymbol{\beta}} = \boldsymbol{\beta}^{(m+1)}$.

The above results suggest that both the upper-bound matrices B_1 and B_2 can be used to establish the proposed QUB algorithm. We also compare the convergence rates of the two corresponding algorithms and show that $\mathbf{B}_2$ can further improve the algorithm rate.

Theorem 4. *The convergence rate of Algorithm 2 is faster than that of Algorithm 1.*

4. SIMULATION STUDIES

In this section, we conduct simulation studies to evaluate the practical behavior of the propose two QUB algorithms (QUB1 and QUB2) and the Newton-Raphson (NR) Algorithm. We consider two scenarios, two-dimensional data and large-dimensional data.

4.1 Study I: two-dimensional data

We consider the Cox model with two-dimensional covariates in this scenario:

$$\lambda(t|Z_1, Z_2) = \lambda_0(t) \exp\left\{\beta_1 Z_1 + \beta_2 Z_2\right\}. \tag{14}$$

We set $\beta_1 = 0$ and 0.693, and $\beta_2 = 0$ and -0.5. Z_1 is generated from a standard normal distribution, and Z_2 is generated from a Bernoulli distribution with success probability of 0.5. The baseline hazard function $\lambda_0(t)$ is set to be 1. Thus, the failure time $\widetilde{T}$ can be generated from an exponential distribution with failure rate $\exp(\beta_1 Z_1 + \beta_2 Z_2)$. The censoring time C is generated from a uniform distribution $U[0, c]$ with c being chosen to achieve the desired censoring rate $\rho = 80\%$ and 90%. For the case-cohort design, the full cohort size is set to be $N = 1000$, and a subcohort of size $\widetilde{n} = 300$ is randomly selected from the cohort without replacement. The subcohort and all the case outside the subcohort constitute the case-cohort sample.

Under each configuration, we compare the proposed estimators, $\widehat{\boldsymbol{\beta}}_{IPW}$, which is the solution to $U_w(\boldsymbol{\beta}) = 0$ using an inverse-probability weighted function in (4), and $\widehat{\boldsymbol{\beta}}_{TVW}$, which is the solution to $U_w(\boldsymbol{\beta}) = 0$ using a time-varying weighted function in (5), with two competing estimators, $\widehat{\boldsymbol{\beta}}_{Full}$, which is the solution to $U_F(\boldsymbol{\beta}) = 0$ using the full cohort sample, and $\widehat{\boldsymbol{\beta}}_{Naive}$, which is the solution to $U_F(\boldsymbol{\beta}) = 0$ using a simple random sample from the cohort with the same size as the case-cohort sample. For the calculation of these four estimators, we apply the NR, QUB1, QUB2 algorithms. The estimated biases (Biases) given by the sample means minus the true values, the sample standard deviations of the estimates (SDs), the means of the estimated standard errors (SEs), and the coverage probabilities (CPs) of 95% nominal confidence intervals are obtained from 1000 independently generated data sets. The simulation results are summarized in Table 1 and Table 2.

Under all the cases considered here, the four estimators calculated by the three algorithms for both β_1 and β_2 are all practically unbiased. The estimated standard errors provide good estimations for the sample standard deviations. The confidence intervals attain coverage close to the nominal 95% level. The proposed estimator under case-cohort design, $\widehat{\boldsymbol{\beta}}_{IPW}$ and $\widehat{\boldsymbol{\beta}}_{TVW}$ are more efficient than $\widehat{\boldsymbol{\beta}}_{Naive}$. The estimated efficiency of $\widehat{\boldsymbol{\beta}}_{IPW}$ relative to $\widehat{\boldsymbol{\beta}}_{Full}$ show that the proposed estimator of β_1 reaches about 55% of the efficiency of the full-cohort estimator when only about 28% subjects of the entire cohort are included in the case-cohort design for $\rho = 90\%$. This supports the notion that the case-cohort design can be a cost-effective alternative to the simple random sampling design in large cohort studies.

On the other hand, the estimators calculated by the QUB1 and QUB2 algorithms show almost identical results with the NR algorithm, which suggests that the proposed QUB algorithms can be good alternatives to the NR algorithm in this scenario. The proposed QUB algorithms only need to calculate the inverse matrix of B_1 and B_2 one time in the whole process instead of calculating the inverse matrix of $H_w(\boldsymbol{\beta}^{(m)})$ in each iterative step. Furthermore, both B_1 and B_2 are independent of the weight function $w_i(t)$, which makes our proposed algorithms universal for the implementation of a class of estimators under case-cohort design in practice.

4.2 Study II: large-dimensional data

In this scenario, we conduct simulation studies to assess the performance of the proposed QUB1 and QUB2 algorithms in the case of large-dimensional data, comparing with the NR algorithm. We generate data from the following Cox model:

$$\lambda(t|\boldsymbol{Z}) = \lambda_0(t) \exp\left\{\boldsymbol{\beta}^T \boldsymbol{Z}\right\}, \tag{15}$$

where $\boldsymbol{\beta} = (\beta_1, ..., \beta_p)^T$ is a p-dimensional vector. We set $\boldsymbol{\beta} = (0.5 \cdot \mathbb{1}_q^T,\ \mathbb{0}_{p-q}^T)^T$, where $\mathbb{1}_m$ and $\mathbb{0}_m$ respectively denote m-dimensional vectors of ones and zeros. Define $r = q/p$, which represents the proportion of non-zero components in $\boldsymbol{\beta}$. $\boldsymbol{Z}$ is generated from a p-dimensional multivariate normal distribution with mean $\mathbb{1}_p^T$ and covariance matrix $\Sigma = (\sigma_{ij})_{p\times p}$, where $\sigma_{ij} = 0.5^{|i-j|}, i, j = 1, ..., p$. This means we consider the case that there exists correlation among the covariates. $\lambda_0(t)$ is set to be 1. $\widetilde{T}$ is generated from an exponential distribution with mean $\left[\exp\left\{\boldsymbol{\beta}^T \boldsymbol{Z}\right\}\right]^{-1}$. C is generated from a uniform distribution $U[0, c]$ with c being chosen to achieve the censoring rate $\rho = 80\%$ and 90%. The setup for the case-cohort design is that $N = 1000$ and $\widetilde{n} = 200$ or 300.

We implement the inverse-probability-weighted estimator $\widehat{\boldsymbol{\beta}}_{IPW}$ by using the QUB1 and QUB2 algorithms and the competing NR algorithm. For the three considered algorithms, the convergence criteria used is as follow: the algorithms stop if either the normed difference of successive iterates is less than 10^{-4}, or the number of iterations exceed 3×10^4. Firstly, we present the simulation results of the parametric estimation. We set $p = 10$ and 20, $r = 0.1$, $\widetilde{n} = 300$ and $\rho = 90\%$. The true value $\boldsymbol{\beta}_0 = (0.5 \cdot \mathbb{1}_{0.1p}^T,\ \mathbb{0}_{0.9p}^T)^T$. We report the results on estimation of the first 5 components of $\boldsymbol{\beta}$ in Table 3.

Secondly, in the case of large p, the NR algorithm may suffer calculational errors from computational singularity of the observed information matrix $H_w(\boldsymbol{\beta})$ and failure of convergence. Meanwhile, the proposed QUB1 and QUB2 algorithms may encounter calculational errors caused by com-

Table 1. Simulation results on estimation of β_1 and β_2

				$\widehat{\beta}_1$				$\widehat{\beta}_2$			
(β_1, β_2)	ρ	Method	Algorithm	Bias	SD	SE	CP	Bias	SD	SE	CP
$(0.693, -0.5)$	80%	$\widehat{\boldsymbol{\beta}}_{Full}$	NR	-0.0000	0.0758	0.0747	0.947	-0.0000	0.1436	0.1457	0.953
			QUB1	0.0075	0.0724	0.0746	0.965	0.0551	0.1355	0.1439	0.947
			QUB2	0.0181	0.0719	0.0747	0.965	0.0163	0.1402	0.1449	0.958
		$\widehat{\boldsymbol{\beta}}_{Naive}$	NR	0.0015	0.1149	0.1136	0.947	0.0012	0.2257	0.2209	0.948
			QUB1	0.0022	0.1134	0.1134	0.955	0.0750	0.1802	0.2188	0.964
			QUB2	0.0135	0.1170	0.1137	0.944	0.0112	0.2167	0.2207	0.958
		$\widehat{\boldsymbol{\beta}}_{IPW}$	NR	0.0093	0.1017	0.0962	0.926	-0.0088	0.1986	0.1886	0.935
			QUB1	0.0139	0.1041	0.0962	0.920	0.0353	0.1865	0.1880	0.946
			QUB2	0.0184	0.1044	0.0964	0.916	0.0097	0.1969	0.1886	0.943
		$\widehat{\boldsymbol{\beta}}_{TVW}$	NR	0.0050	0.1070	0.1071	0.939	0.0075	0.1903	0.1946	0.949
			QUB1	0.0111	0.1114	0.0986	0.910	0.0331	0.1812	0.1724	0.918
			QUB2	0.0114	0.1119	0.0960	0.914	0.0130	0.1892	0.1831	0.929
	90%	$\widehat{\boldsymbol{\beta}}_{Full}$	NR	0.0000	0.1057	0.1032	0.940	0.0000	0.2064	0.2074	0.948
			QUB1	0.0089	0.1002	0.1034	0.946	0.0500	0.1886	0.2050	0.970
			QUB2	0.0180	0.0992	0.1035	0.949	0.0109	0.1965	0.2066	0.962
		$\widehat{\boldsymbol{\beta}}_{Naive}$	NR	0.0134	0.1721	0.1720	0.948	-0.0163	0.3549	0.3483	0.954
			QUB1	0.0033	0.1730	0.1703	0.941	0.0876	0.2641	0.3414	0.979
			QUB2	0.0079	0.1719	0.1713	0.945	0.0012	0.3332	0.3471	0.965
		$\widehat{\boldsymbol{\beta}}_{IPW}$	NR	0.0189	0.1265	0.1263	0.941	-0.0090	0.2479	0.2464	0.954
			QUB1	0.0098	0.1284	0.1253	0.941	0.0224	0.2300	0.2444	0.971
			QUB2	0.0132	0.1274	0.1256	0.942	0.0017	0.2374	0.2453	0.965
		$\widehat{\boldsymbol{\beta}}_{TVW}$	NR	0.0099	0.1323	0.1326	0.929	-0.0127	0.2511	0.2521	0.937
			QUB1	0.0149	0.1337	0.1280	0.929	0.0314	0.2307	0.2288	0.944
			QUB2	0.0150	0.1401	0.1261	0.922	0.0097	0.2511	0.2414	0.938
$(0.693, 0)$	80%	$\widehat{\boldsymbol{\beta}}_{Full}$	NR	0.0000	0.0757	0.0751	0.952	-0.0000	0.1423	0.1436	0.956
			QUB1	0.0268	0.0755	0.0749	0.930	-0.0015	0.1201	0.1430	0.980
			QUB2	0.0314	0.0750	0.0749	0.926	0.0043	0.1389	0.1432	0.954
		$\widehat{\boldsymbol{\beta}}_{Naive}$	NR	0.0047	0.1156	0.1148	0.947	-0.0019	0.2239	0.2193	0.960
			QUB1	0.0142	0.1132	0.1149	0.949	-0.0032	0.2093	0.2185	0.953
			QUB2	0.0183	0.1129	0.1150	0.947	-0.0010	0.2232	0.2187	0.944
		$\widehat{\boldsymbol{\beta}}_{IPW}$	NR	0.0072	0.0981	0.0973	0.953	0.0003	0.1902	0.1872	0.954
			QUB1	0.0186	0.0978	0.0973	0.946	0.0048	0.1859	0.1873	0.945
			QUB2	0.0227	0.0979	0.0974	0.942	0.0007	0.1974	0.1876	0.941
		$\widehat{\boldsymbol{\beta}}_{TVW}$	NR	0.0056	0.1092	0.1053	0.940	0.0041	0.1847	0.1896	0.951
			QUB1	0.0106	0.1090	0.0993	0.915	0.0194	0.1789	0.1689	0.918
			QUB2	0.0156	0.1074	0.0966	0.915	0.0215	0.1861	0.1801	0.924
	90%	$\widehat{\boldsymbol{\beta}}_{Full}$	NR	-0.0000	0.1048	0.1021	0.946	-0.0000	0.2035	0.2004	0.951
			QUB1	0.0240	0.1035	0.1021	0.943	0.0053	0.1705	0.1996	0.974
			QUB2	0.0293	0.1028	0.1022	0.942	0.0083	0.1933	0.2000	0.959
		$\widehat{\boldsymbol{\beta}}_{Naive}$	NR	0.0101	0.1251	0.1251	0.944	0.0019	0.2422	0.2391	0.947
			QUB1	0.0093	0.1741	0.1696	0.943	-0.0124	0.3069	0.3323	0.974
			QUB2	0.0128	0.1735	0.1699	0.947	-0.0138	0.3321	0.3336	0.959
		$\widehat{\boldsymbol{\beta}}_{IPW}$	NR	0.0189	0.1265	0.1263	0.941	-0.0090	0.2479	0.2464	0.954
			QUB1	0.0167	0.1301	0.1247	0.943	-0.0018	0.2278	0.2391	0.957
			QUB2	0.0204	0.1297	0.1249	0.942	-0.0042	0.2386	0.2394	0.954
		$\widehat{\boldsymbol{\beta}}_{TVW}$	NR	0.0044	0.1282	0.1314	0.934	0.0030	0.2342	0.2437	0.942
			QUB1	0.0110	0.1318	0.1260	0.938	0.0066	0.2306	0.2237	0.935
			QUB2	0.0169	0.1332	0.1249	0.923	0.0016	0.2332	0.2350	0.952

Table 2. Simulation results on estimation of β_1 and β_2

				$\widehat{\beta}_1$				$\widehat{\beta}_2$			
(β_1, β_2)	ρ	Method	Algorithm	Bias	SD	SE	CP	Bias	SD	SE	CP
$(0, -0.5)$	80%	$\widehat{\boldsymbol{\beta}}_{Full}$	NR	−0.0011	0.0726	0.0713	0.947	−0.0028	0.1411	0.1461	0.956
			QUB1	0.0100	0.0727	0.0713	0.946	0.0500	0.1370	0.1443	0.946
			QUB2	0.0219	0.0725	0.0713	0.931	0.0124	0.1428	0.1453	0.955
		$\widehat{\boldsymbol{\beta}}_{Naive}$	NR	0.0010	0.1122	0.1082	0.942	−0.0030	0.2210	0.2220	0.952
			QUB1	0.0114	0.1108	0.1075	0.944	0.0768	0.1831	0.2202	0.960
			QUB2	0.0089	0.1114	0.1076	0.947	−0.0035	0.2296	0.2224	0.944
		$\widehat{\boldsymbol{\beta}}_{IPW}$	NR	−0.0005	0.0921	0.0916	0.955	−0.0055	0.1839	0.1830	0.954
			QUB1	0.0023	0.0953	0.0910	0.942	0.0222	0.1686	0.1826	0.969
			QUB2	0.0069	0.0953	0.0911	0.945	−0.0022	0.1783	0.1830	0.957
		$\widehat{\boldsymbol{\beta}}_{TVW}$	NR	−0.0036	0.0935	0.0932	0.935	−0.0014	0.1931	0.1901	0.932
			QUB1	0.0068	0.0937	0.0890	0.939	0.0330	0.1829	0.1672	0.912
			QUB2	0.0084	0.0931	0.0870	0.935	−0.0095	0.1827	0.1786	0.948
	90%	$\widehat{\boldsymbol{\beta}}_{Full}$	NR	−0.0000	0.1027	0.1001	0.937	−0.0000	0.2032	0.2072	0.958
			QUB1	0.0127	0.1039	0.1004	0.937	0.0618	0.1846	0.2044	0.964
			QUB2	0.0233	0.1041	0.1004	0.933	0.0161	0.2053	0.2064	0.955
		$\widehat{\boldsymbol{\beta}}_{Naive}$	NR	−0.0011	0.1750	0.1639	0.934	−0.0100	0.3503	0.3468	0.957
			QUB1	0.0039	0.1657	0.1643	0.930	0.0812	0.2910	0.3409	0.974
			QUB2	0.0057	0.1676	0.1648	0.931	−0.0108	0.3585	0.3474	0.953
		$\widehat{\boldsymbol{\beta}}_{IPW}$	NR	0.0051	0.1171	0.1167	0.957	−0.0074	0.2314	0.2361	0.960
			QUB1	0.0026	0.1168	0.1167	0.942	0.0161	0.2396	0.2356	0.947
			QUB2	0.0072	0.1166	0.1168	0.945	−0.0051	0.2488	0.2363	0.938
		$\widehat{\boldsymbol{\beta}}_{TVW}$	NR	0.0040	0.1209	0.1193	0.933	−0.0024	0.2405	0.2414	0.932
			QUB1	−0.0003	0.1163	0.1153	0.942	0.0220	0.2260	0.2194	0.937
			QUB2	0.0071	0.1163	0.1140	0.947	−0.0058	0.2300	0.2315	0.946
$(0, 0)$	80%	$\widehat{\boldsymbol{\beta}}_{Full}$	NR	−0.0016	0.0720	0.0714	0.948	−0.0005	0.1392	0.1429	0.953
			QUB1	0.0244	0.0728	0.0714	0.931	0.0027	0.1276	0.1425	0.969
			QUB2	0.0294	0.0725	0.0715	0.926	0.0025	0.1426	0.1426	0.950
		$\widehat{\boldsymbol{\beta}}_{Naive}$	NR	0.0028	0.1112	0.1083	0.939	0.0043	0.2145	0.2177	0.955
			QUB1	0.0057	0.1096	0.1082	0.940	0.0058	0.2053	0.2168	0.962
			QUB2	0.0099	0.1094	0.1081	0.942	0.0050	0.2172	0.2170	0.952
		$\widehat{\boldsymbol{\beta}}_{IPW}$	NR	0.0034	0.0905	0.0909	0.953	0.0055	0.1824	0.1805	0.952
			QUB1	0.0073	0.0883	0.0909	0.957	0.0075	0.1759	0.1806	0.957
			QUB2	0.0106	0.0884	0.0909	0.953	0.0077	0.1850	0.1806	0.950
		$\widehat{\boldsymbol{\beta}}_{TVW}$	NR	0.0062	0.0914	0.0931	0.938	−0.0094	0.1804	0.1824	0.942
			QUB1	0.0077	0.0899	0.0885	0.950	−0.0041	0.1638	0.1619	0.930
			QUB2	0.0095	0.0895	0.0870	0.941	0.0029	0.1706	0.1728	0.942
	90%	$\widehat{\boldsymbol{\beta}}_{Full}$	NR	0.0017	0.1021	0.1014	0.945	0.0018	0.2078	0.2040	0.953
			QUB1	0.0258	0.1062	0.1015	0.930	0.0078	0.1867	0.2038	0.966
			QUB2	0.0321	0.1060	0.1015	0.928	0.0081	0.2062	0.2041	0.954
		$\widehat{\boldsymbol{\beta}}_{Naive}$	NR	0.0011	0.1744	0.1668	0.941	−0.0089	0.3483	0.3424	0.960
			QUB1	0.0010	0.1658	0.1663	0.956	−0.0025	0.3177	0.3409	0.972
			QUB2	0.0065	0.1664	0.1664	0.955	−0.0030	0.3451	0.3421	0.957
		$\widehat{\boldsymbol{\beta}}_{IPW}$	NR	−0.0051	0.1196	0.1168	0.941	0.0022	0.2452	0.2339	0.949
			QUB1	0.0032	0.1150	0.1174	0.959	−0.0097	0.2235	0.2331	0.956
			QUB2	0.0068	0.1148	0.1174	0.960	−0.0105	0.2330	0.2333	0.944
		$\widehat{\boldsymbol{\beta}}_{TVW}$	NR	0.0002	0.1154	0.1200	0.947	−0.0019	0.2333	0.2377	0.944
			QUB1	0.0036	0.1204	0.1156	0.935	0.0036	0.2173	0.2185	0.945
			QUB2	0.0063	0.1239	0.1135	0.918	−0.0008	0.2313	0.2284	0.943

Table 3. Simulation results on the estimation of the first 5 components of parameter

			QUB1				QUB2			
p	True value		Bias	SD	SE	CP	Bias	SD	SE	CP
10	0.5	$\widehat{\beta}_1$	0.0146	0.1549	0.1459	0.941	0.0147	0.1549	0.1459	0.941
	0	$\widehat{\beta}_2$	0.0068	0.1665	0.1616	0.941	0.0068	0.1665	0.1616	0.941
	0	$\widehat{\beta}_3$	−0.0086	0.1611	0.1607	0.949	−0.0086	0.1611	0.1607	0.949
	0	$\widehat{\beta}_4$	0.0125	0.1643	0.1609	0.945	0.0125	0.1643	0.1609	0.944
	0	$\widehat{\beta}_5$	−0.0065	0.1668	0.1606	0.939	−0.0065	0.1668	0.1606	0.939
20	0.5	$\widehat{\beta}_1$	0.0484	0.1564	0.1543	0.930	0.0484	0.1564	0.1543	0.930
	0.5	$\widehat{\beta}_2$	0.0481	0.1802	0.1720	0.929	0.0482	0.1802	0.1720	0.929
	0	$\widehat{\beta}_3$	−0.0027	0.1836	0.1712	0.927	−0.0027	0.1836	0.1712	0.927
	0	$\widehat{\beta}_4$	0.0040	0.1763	0.1702	0.943	0.0040	0.1763	0.1702	0.943
	0	$\widehat{\beta}_5$	0.0048	0.1803	0.1717	0.934	0.0048	0.1803	0.1717	0.934

NOTE: $\widetilde{n}=300,\ \rho=90\%,\ r=0.1,\ \boldsymbol{\beta}_0=(0.5\cdot\mathbb{1}^T_{0.1p},\ \mathbb{0}^T_{0.9p})^T$.

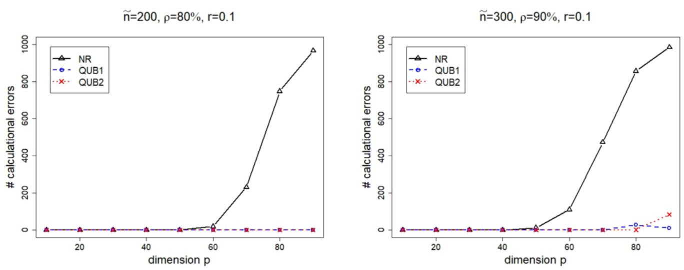

Figure 1. Calculational errors for three algorithms against dimension p.

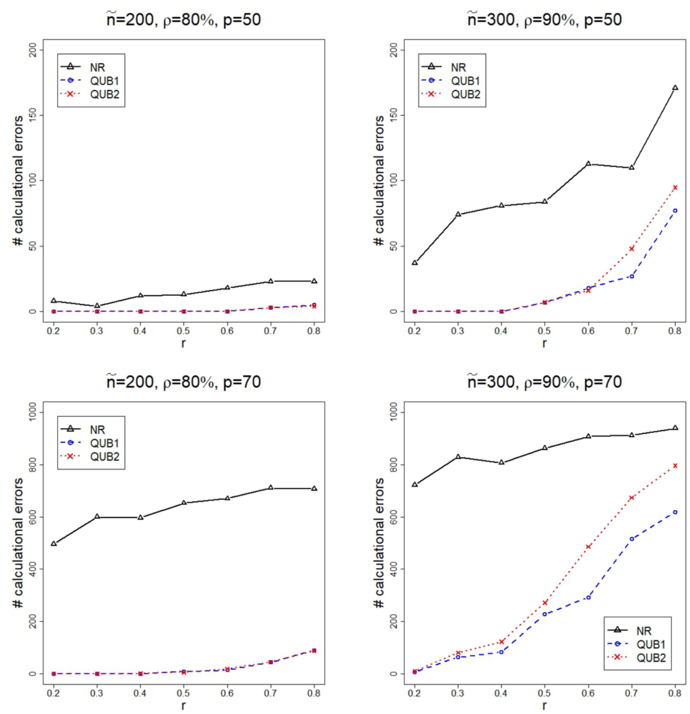

Figure 2. Calculational errors for three algorithms against non-zero proportion r.

putational non-convergence. We record the number of calculational errors for each algorithm and present the results in Figure 1 and 2. In Figure 1, we set the data dimension p to be $10, 20, ..., 90$ and fix the proportion of non-zero components in parameter to be 0.1 ($r = 0.1$). In Figure 2, we set $p = 50$ and 70, respectively, and vary r from 0.2 to 0.8.

The results suggest that the proposed QUB1 and QUB2 algorithms are more stable and efficient than the NR algorithm in large p case. As can be seen from Figure 1, the number of calculation errors for the NR algorithm raises as the dimension p increases, sharply when p is larger than 60. For example, when $\widetilde{n} = 300, \rho = 90\%, r = 0.1$ and $p = 90$, the rate of calculation errors for the NR algorithm is around 0.984. Meanwhile, the proposed two QUB algorithms perform very well. Figure 2 suggests that the numbers of calculation errors for all three algorithms increase as the proportion of non-zero components in parameter r increases. However, the proposed QUB algorithms work much better than the NR algorithm, even in non-sparsity cases.

5. REAL DATA ANALYSIS

5.1 Wilms tumor study

The National Wilms Tumor Study Group (NWTSG) has conducted a Wilms tumor study to assess the association between the tumor histology and time-to-relapse of 4028 children diagnosed with Wilms tumor, a rare kidney cancer in young children (Beckwith and Palmer, 1978). The most

Table 4. Demographics and characteristics of the Wilms tumor data

	Subcohort	Case	Case-Cohort Sample
Histology type (%)			
0 = Favorable histology	66.32 (590/668)	66.02 (377/571)	78.77 (909/1154)
1 = Unfavourable histology	11.68 (78/668)	33.98 (194/571)	21.23 (245/1154)
Disease stage (%)			
1 = Stage I	39.97 (267/668)	20.49 (117/571)	31.37 (362/1154)
2 = Stage II	25.00 (167/668)	29.07 (166/571)	26.95 (311/1154)
3 = Stage III	24.70 (165/668)	30.65 (175/571)	27.12 (313/1154)
4 = Stage IV	10.33 (69/668)	19.79 (113/571)	14.56 (168/1154)
Age at diagnosis (mean ± sd)			
	44.02 ± 32.53	51.83 ± 37.55	47.31 ± 34.75

Table 5. Results for analysis of the Wilms tumor study data

	Type			Stage			Age		
Algorithm	Est.	SE	p-value	Est.	SE	p-value	Est.	SE	p-value
NR	1.3336	0.1301	< 0.0001*	0.3438	0.0537	< 0.0001*	0.1093	0.0541	0.0435*
QUB1	1.1235	0.1332	< 0.0001*	0.2712	0.0525	< 0.0001*	0.1290	0.0516	0.0124*
QUB2	1.3242	0.1306	< 0.0001*	0.3556	0.0538	< 0.0001*	0.1256	0.0539	0.0197*

NOTE: Significant effect at 5% level.

important predictor, patients' histological type, was initially diagnosed by a local pathologist at the time of treatment and then by an experienced pathologist in the NWTSG Pathology Center. The latter assessment was much more accurate, however, quite expensive and time-consuming at the same time. Hence an example of case-cohort design was proposed by Breslow and Chatterjee (1999) for the study. Specifically, a subcohort of size 668 was randomly selected from the full cohort. The subcohort and the children who experienced relapse or death but were not included in the subcohort constituted the case-cohort sample. For illustration, we apply the proposed methods to analyze such a case-cohort example.

We consider three potential confounders. The first one is the histology type (Type), which is divided into two categories: "Unfavourable histology" (Type =1) if patient's tumor is composed of one of the rare cell types; "Favourable histology" (Type = 0), otherwise. The second one is the stage of the disease (Stage), which is classified as, only in the kidney and resected (Stage =1), spread but resected (Stage =2), residual in the abdomen or lymph nodes (Stage=3), metastatic to the liver or lung (Stage =4). We consider the age at diagnosis (Age) as the third one, which is recorded in months, and has been scaled. Table 4 provides the demographic characteristics for all the covariates considered here.

We use the following model to fit the case-cohort data:

$$\lambda(t|\boldsymbol{Z}) = \lambda_0(t)\exp\{\beta_1\text{Type} + \beta_2\text{Stage} + \beta_3\text{Age}\}. \tag{16}$$

In order to calculate the inverse-probability-weighted estimates of the regression coefficients, the proposed QUB numerical algorithms are applied and a comparison with the NR algorithm is conducted. The results are summarized in Table 5. Overall, the results from the three algorithms are consistent. Those suggest that patients with unfavorable histology type of tumor tend to have a higher risk by about 3.8 times than those with favorable type. As expected, patients with higher stage of disease or older age of diagnosis are more likely to relapse or die. Unsurprisingly, the proposed QUB algorithms are good alternatives to the NR algorithm in practice. Furthermore, we find out that the proposed QUB algorithms are less sensitive to the choice of initial values of parameters than the NR algorithm in this example.

5.2 SEER breast cancer study

The Surveillance, Epidemiology and End Results (SEER) Program provides information on cancer statistics in an effort to reduce the cancer burden among the U.S. population. The registries routinely collect data on patient demographics, primary tumor site, tumor morphology and stage at diagnosis, etc., on various types of cancer. Breast cancer has become which is one of the most common malignant tumors in the world. We obtained a dataset of 124785 female patients of breast cancer from the SEER database.

Demographic variables in the dataset include the age at diagnosis (Age), which is classified into 17 groups by every 5 years old divided into one group, starting from the age of five. The race of patient (Race), which is classified as black, white and the others. Tumor morphological variable include the type of histology (Grade), which is classifies as four types: well differentiated, moderately differentiated, poorly differentiated, and undifferentiated anaplastic. Stage at diagnosis is described by three variables according to the TNM Classification, which is proposed by American Joint Committee on Cancer (AJCC, 7th edition). The T category describes the primary tumor, T0 indicates that there is no evidence of the primary tumor, T1–T4 indicates the extent of the primary tumor, including the subcategories T1a, T1b, T1c, T1mic, T1NOS, T4a, T4b, T4c, and T4d, Tis indicates the tumors diagnosed as in situ during a clinical workup, and TX indicates that there is no information available to determine T. The N category describes the regional lymph node metastasis, N0 indicates that there is no typical evidence of metastatic activity, including the subcategories N0, N0(i+), N0(i-), N0(mol+), and N0(mol-) according to the findings for isolated tumor cells, N1–N3 indicates the extent of the regional lymph node metastasis, including the subcategories N1a, N1b, N1c, N1mi, N1NOS, N2a, N2b, N2NOS, N3a, N3b, N3c, and N3NOS, and NX indicates that there is no information available to determine N. The M category describes the distant metastasis, M0 (M=0) indicates that there is no evidence of metastasis at distant sites on clinical evaluation and M1 (M=1) indicates that there is clinical evidence or histologic confirmation of distant metastasis.

The outcome of interest is the survival months, which is subject to censoring, and the vital status is recorded as of the study cutoff, when the individual is censored, vital status is recorded as "Alive", otherwise vital status is recorded as "Dead". The censoring rate is about 84.4%. We focus on evaluating the effects of the above confounders on the survival time of breast cancer patients. To illustrate the case-cohrt design, we artificially select a subcohort of size 50000 from the full cohort through random sampling. The subcohort and all the patients who died during the trial and are not included in the subcohort consist of the case-cohort sample.

By introducing some dummy variables for the categorical variables, we fit the case-cohort data by Cox model (1) with $p = 37$. The results of the IPW method and summarized in Table 6. Overall, the results from QUB1 and QUB2 algorithms under case-cohort design are both consistent with those using the full cohort sample. Most of the covariates

Table 6. Results for analysis of the SEER breast cancer study data

	Full cohort						Case-cohort design					
	QUB1			QUB2			QUB1			QUB2		
	Est.	SE	p-value	Est.	SE	p-value	Est.	SE	p-value	Est.	SE	p-value
Age	0.7024	0.0088	< 0.0001*	0.7026	0.0088	< 0.0001*	0.3961	0.0136	< 0.0001*	0.3943	0.0136	< 0.0001*
Grade	0.3332	0.0110	< 0.0001*	0.3338	0.0110	< 0.0001*	0.3272	0.0191	< 0.0001*	0.3224	0.0192	< 0.0001*
Black	0.5247	0.0328	< 0.0001*	0.5264	0.0329	< 0.0001*	0.5222	0.0616	< 0.0001*	0.5129	0.0623	< 0.0001*
White	0.2124	0.0275	< 0.0001*	0.2136	0.0275	< 0.0001*	0.1689	0.0449	0.0002*	0.1583	0.0455	0.0005*
Tis	−0.6784	0.1853	0.0003*	−0.6774	0.1858	0.0003*	−0.9431	0.3372	0.0052*	−0.9623	0.3372	0.0043*
T0	−0.7007	0.2277	0.0021*	−0.6950	0.2278	0.0023*	−0.6078	0.3879	0.1172	−0.6363	0.3924	0.1049
T1a	−1.2821	0.0573	< 0.0001*	−1.2772	0.0574	< 0.0001*	−1.2227	0.4290	0.0044*	−1.2481	0.4318	0.0039*
T1b	−1.1125	0.0445	< 0.0001*	−1.1078	0.0447	< 0.0001*	−1.0965	0.1607	< 0.0001*	−1.1213	0.1591	< 0.0001*
T1c	−0.8579	0.0400	< 0.0001*	−0.8534	0.0402	< 0.0001*	−0.8527	0.1599	< 0.0001*	−0.8763	0.1583	< 0.0001*
T1mic	−1.3178	0.1061	< 0.0001*	−1.3130	0.1061	< 0.0001*	−1.2392	0.2030	< 0.0001*	−1.2638	0.2019	< 0.0001*
T1NOS	−0.9499	0.2064	< 0.0001*	−0.9466	0.2066	< 0.0001*	−1.1405	0.4014	0.0045*	−1.1651	0.4020	0.0038*
T2	−0.4342	0.0383	< 0.0001*	−0.4298	0.0385	< 0.0001*	−0.3932	0.1604	0.0142*	−0.4155	0.1588	0.0089*
T3	−0.1187	0.0422	0.0049*	−0.1142	0.0424	0.0070*	−0.0944	0.1626	0.5617	−0.1162	0.1610	0.4704
T4a	0.0848	0.0613	0.1671	0.0865	0.0616	0.1603	0.0663	0.2888	0.8185	0.0436	0.2868	0.8792
T4b	0.0872	0.0447	0.0513	0.0921	0.0449	0.0402*	0.1111	0.1724	0.5191	0.0907	0.1709	0.5957
T4c	0.1805	0.1080	0.0947	0.1838	0.1081	0.0892	0.2286	0.2489	0.3584	0.2095	0.2476	0.3974
T4d	0.1472	0.0555	0.0081*	0.1528	0.0556	0.0060*	0.1667	0.1707	0.3289	0.1465	0.1692	0.3865
T4NOS	−0.0918	0.1260	0.4664	−0.0914	0.1265	0.4700	−0.1979	0.3161	0.5314	−0.2181	0.3161	0.4902
N0	−0.8055	0.0426	< 0.0001*	−0.7938	0.0430	< 0.0001*	−0.8107	0.1440	< 0.0001*	−0.8585	0.1441	< 0.0001*
N0(i-)	−1.3129	0.0471	< 0.0001*	−1.3012	0.0475	< 0.0001*	−1.3190	0.1205	< 0.0001*	−1.3669	0.1207	< 0.0001*
N0(i+)	−1.2190	0.0773	< 0.0001*	−1.2073	0.0775	< 0.0001*	−1.2082	0.1719	< 0.0001*	−1.2562	0.1720	< 0.0001*
N0(mol-)	−0.8045	0.1986	0.0001*	−0.7943	0.1988	0.0001*	−0.9814	0.5219	0.0600	−1.0307	0.5210	0.0479*
N0(mol+)	−11.1182	0.0427	< 0.0001*	−11.1063	0.0430	< 0.0001*	−10.9142	0.6931	< 0.0001*	−10.9634	0.6926	< 0.0001*
N1	−0.3600	0.0443	< 0.0001*	−0.3484	0.0446	< 0.0001*	−0.3938	0.1071	0.0002*	−0.4389	0.1068	< 0.0001*
N1a	−0.9284	0.0468	< 0.0001*	−0.9166	0.0472	< 0.0001*	−0.9065	0.1372	< 0.0001*	−0.9539	0.1370	< 0.0001*
N1b	−0.5293	0.3422	0.1219	−0.5195	0.3427	0.1296	−0.7948	0.6536	0.2240	−0.8403	0.6557	0.2000
N1c	−0.8228	0.3583	0.0217*	−0.8117	0.3583	0.0235*	−0.7771	0.9956	0.4351	−0.8257	0.9996	0.4088
N1mi	−1.0847	0.0592	< 0.0001*	−1.0731	0.0595	< 0.0001*	−1.0877	0.1400	< 0.0001*	−1.1357	0.1400	< 0.0001*
N1NOS	−0.7027	0.0834	< 0.0001*	−0.6904	0.0835	< 0.0001*	−0.6430	0.1537	< 0.0001*	−0.6903	0.1534	< 0.0001*
N2a	−0.4462	0.0479	< 0.0001*	−0.4351	0.0482	< 0.0001*	−0.4657	0.1272	0.0003*	−0.5118	0.1270	0.0001*
N2b	−0.6811	0.1570	< 0.0001*	−0.6689	0.1569	< 0.0001*	−0.7195	0.2809	0.0104*	−0.7635	0.2809	0.0066*
N2NOS	−0.4449	0.0849	< 0.0001*	−0.4328	0.0850	< 0.0001*	−0.4274	0.1701	0.0120*	−0.4724	0.1702	0.0055*
N3a	−0.2235	0.0523	< 0.0001*	−0.2117	0.0526	0.0001*	−0.2496	0.1240	0.0440*	−0.2945	0.1237	0.0173*
N3b	−0.5772	0.0818	< 0.0001*	−0.5642	0.0818	< 0.0001*	−0.4828	0.1487	0.0012*	−0.5263	0.1482	0.0004*
N3c	−0.0568	0.0712	0.4248	−0.0473	0.0715	0.5076	−0.1815	0.1351	0.1791	−0.2230	0.1346	0.0977*
N3NOS	−0.4423	0.1347	0.0010*	−0.4276	0.1343	0.0015*	−0.3699	0.2537	0.1448	−0.4139	0.2543	0.1036
M	1.3663	0.0243	< 0.0001*	1.3669	0.0243	< 0.0001*	1.3049	0.0635	< 0.0001*	1.2997	0.0640	< 0.0001*

NOTE: * Significant effect at 5% level. Race of "others", "TX", and "NX" are the reference groups.

have significant effects on the survival months. Unsurprisingly, patients with worse histology type or older age at diagnosis have a higher risk of death. Black women have a higher risk than the white women and the other races. As for the TNM staging, those with no information available to determine T or N are most likely to die. The wider extent of the primary tumor or regional lymph node metastasis results in higher risk. Patients with presence of distant metastasis are more likely to die than those with absence of distant metastasis.

We can see that the proposed QUB algorithms work very well in the large-dimensional and nonsparse case. We also tried to use the Newton-Raphson algorithm but it didn't work. This suggests that the proposed QUB algorithms are superior over the Newton-Raphson algorithm when data dimension is large.

6. CONCLUSIONS

The case-cohort design is widely used as a cost-effective biased sampling mechanism in large cohort studies. In practice, the inference of parameter requires numerical calculation methods in case-cohort studies. We propose new quadratic-upper-bound (QUB) algorithms by using an MM principle for estimations in the Cox model under case-cohort design. The proposed QUB algorithms have several computational superiorities to the Newton-Raphson algorithm. Such algorithms are universal for a class of weighted estimators widely used in case-cohort studies because they are independent of choices of weight functions. The QUB algorithms are monotonic according to the proof of convergence and have simple update and low per-iterative cost by only calculating the inversion of upper-bound matrix one time in the whole process. Simulation studies suggest that the proposed QUB algorithms perform stable and efficient, especially in the situations that covariate dimension is large and data are non-sparse. Two real data examples from a Wilm tumor study and a SEER breast cancer study demonstrates the application of the proposed algorithms in practice. The results suggest that the proposed QUB algorithms can perform very well and stably in practice, especially in the large-dimensional and nonsparse cases.

It is important to theoretically study the effect of the subcohort size on the efficiency of the proposed method. When the budget is given, the searching of the optimal size of subcohort to maximize the efficiency of the proposed estimator under the case-cohort design is an interesting but also challenging work in the future. Here we build the numerical algorithms for a series of weighted estimators for case-cohort data. Future studies will extend to the development on algorithms for some likelihood-based estimators. The case-cohort design is particularly useful when the event rate is low. Studies on numerical algorithms for survival data with a low or medium censoring rate under a generalized case-cohort design (Cai and Zeng, 2007) and a failure time outcome-dependent sampling design (Ding et al., 2014; Yu et al., 2016) are interesting issues in the future.

In addition to the proposed calculation algorithm, the calculation speed is also crucial in practice. How to improve the efficiency of program execution while maintaining the simplicity of writing programs has become an important concern. *Rcpp* is an R package which enables us to write R extensions with $C++$ (Eddelbuettel and Francois, 2011; Eddelbuettel and Sanderson, 2014). *Rcpp* is a brand new computing environment with large-scale integration of existing libraries. In the future, we consider to write the program via *Rcpp* for the proposed QUB algorithm. By reconstructing existing functions in R using $C++$ and developing new functions for efficient numerical experiments, we are expected to improve the efficiency of R program execution and greatly increase the speed of the proposed algorithm.

ACKNOWLEDGEMENTS

This research is supported in part by the National Natural Science Foundation of China (11671310 to J.D., 11471252 to Y.F.).

APPENDIX: PROOFS OF THEOREMS

Asymptotic properties of $\widehat{\boldsymbol{\beta}}$ is established by Lin et al. (2000) in order to make statistical inferences later. Before we present the result of asymptotic properties, some notations need to be defined first. Let $\boldsymbol{\beta}_0$ represent the true value of $\boldsymbol{\beta}$ and denote τ to be the time when the study stops or discontinues. Then bring in an at-risk process $Y_i(t) = I(T_i \geq t)$ and a counting process $N_i(t) = \Delta_i I(T_i \leq t)$. Define $M_i(t) = N_i(t) - \int_0^t Y_i(s)\lambda_0(s)\exp\left\{\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{Z}_i\right\}ds$. Set $a^{\otimes 0} = 1$, $a^{\otimes 1} = a$ and $a^{\otimes 2} = aa^{\mathrm{T}}$ while a is a column vector. For $d = 0, 1, 2$, define $S_w^{(d)}(\boldsymbol{\beta}, t) = n^{-1}\sum_{i=1}^n \omega_i Y_i(t)\exp\left\{\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{Z}_i\right\}\boldsymbol{Z}_i^{\otimes d}$.

After all the definitions above, the following regularity conditions can be given as premises in this paper.

(C1) The space of parameter $\mathcal{B}$ is compact and convex as well.
(C2) The space of covariate, $\mathcal{Z}$, is compact.
(C3) $\widetilde{p} = \widetilde{n}/n \to p$ as $N \to \infty$ for some $p \in (0,\ 1)$.
(C4) $\int_0^\tau \lambda_0(t)dt < \infty$.
(C5) $\mathrm{P}(Y_1(t) = 1,\ \text{for any } t \in [0, \tau]) > 0$.
(C6) For $d = 0,\ 1,\ 2$,

$$\sup_{\boldsymbol{\beta}\in\mathcal{B},\, t\in[0,\tau]} \left\| S_w^{(d)}(\boldsymbol{\beta}, t) - s^{(d)}(\boldsymbol{\beta}, t)\right\| \xrightarrow{P} 0, \text{ as } n \to \infty,$$

where $\|\cdot\|$ represents Euclidean norm, $s^{(d)}(\boldsymbol{\beta}, t) = \mathrm{E}\left[Y_1(t)\exp\left\{\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{Z}_1\right\}\boldsymbol{Z}_1^{\otimes d}\right]$, and $s^{(d)}(\boldsymbol{\beta}, t)$ is both absolutely and uniformly continuous of $\boldsymbol{\beta}$ on $\mathcal{B}$ in $t \in [0, \tau]$.

(C7) The following matrix $\Sigma(\boldsymbol{\beta}_0)$ is finite and positive definite,

$$\Sigma(\boldsymbol{\beta}_0) = \int_0^\tau \left[\frac{s^{(2)}(\boldsymbol{\beta}_0, t)}{s^{(0)}(\boldsymbol{\beta}_0, t)} - \left\{ \frac{s^{(1)}(\boldsymbol{\beta}_0, t)}{s^{(0)}(\boldsymbol{\beta}_0, t)} \right\}^{\otimes 2} \right] \cdot s^{(0)}(\boldsymbol{\beta}_0, t)\lambda_0(t)dt.$$

The asymptotic properties of the proposed $\widehat{\boldsymbol{\beta}}$ can be derived by similar arguments in Kang and Cai (2009). We summarize these properties in the following lemma.

Lemma 1. *Under regularity conditions (C1)–(C7), as $n \to \infty$, $\widehat{\boldsymbol{\beta}}$ converges in probability to $\boldsymbol{\beta}_0$, i.e., and*

$$\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{d} N_p(0, \Sigma^{-1}(\boldsymbol{\beta}_0)\left\{\Sigma_1(\boldsymbol{\beta}_0) + \Sigma_2(\boldsymbol{\beta}_0)\right\}\Sigma^{-1}(\boldsymbol{\beta}_0)),$$

where

$$\Sigma_1(\boldsymbol{\beta}_0) = E\left[G_1(\boldsymbol{\beta}_0)^{\otimes 2}\right],$$
$$\Sigma_2(\boldsymbol{\beta}_0) = \frac{1-p}{p} E\left[(1-\Delta_1)\, G_1(\boldsymbol{\beta}_0)^{\otimes 2}\right],$$

and

$$G_1(\boldsymbol{\beta}_0) = \int_0^\tau \left[\boldsymbol{Z}_1 - \frac{s^{(1)}(\boldsymbol{\beta}_0, t)}{s^{(0)}(\boldsymbol{\beta}_0, t)} \right] dM_1(t).$$

Proof of Theorem 1: For the weighted quadratic approximation function $Q_w(\boldsymbol{\beta}|\boldsymbol{\beta}^{(m)})$ in (9), note that

$$\begin{aligned} &Q_w(\boldsymbol{\beta}^{(m+1)}|\boldsymbol{\beta}^{(m)}) - \ell_w(\boldsymbol{\beta}^{(m)}) \\ =&U_w(\boldsymbol{\beta}^{(m)})^{\mathrm{T}} \left\{\boldsymbol{\beta}^{(m+1)} - \boldsymbol{\beta}^{(m)}\right\} \\ &- \frac{1}{2}\left\{\boldsymbol{\beta}^{(m+1)} - \boldsymbol{\beta}^{(m)}\right\}^{\mathrm{T}} \mathbf{B} \left\{\boldsymbol{\beta}^{(m+1)} - \boldsymbol{\beta}^{(m)}\right\}, \end{aligned}$$

where $\boldsymbol{\beta}^{(m+1)} = \boldsymbol{\beta}^{(m)} + \mathbf{B}^{-1}\nabla_{\boldsymbol{\beta}}\ell_w(\boldsymbol{\beta}^{(m)})$ in (11). Let $x = U_w(\boldsymbol{\beta}^{(m)})$. Therefore, we have

$$\begin{aligned} &Q_w(\boldsymbol{\beta}^{(m+1)}|\boldsymbol{\beta}^{(m)}) - \ell_w(\boldsymbol{\beta}^{(m)}) \\ =&x^{\mathrm{T}}\mathbf{B}^{-1}x - \frac{1}{2}\left\{\mathbf{B}^{-1}x\right\}^{\mathrm{T}} \mathbf{B}\left\{\mathbf{B}^{-1}x\right\} \\ =&x^{\mathrm{T}}\mathbf{B}^{-1}x - \frac{1}{2}x^{\mathrm{T}}\mathbf{B}^{-1}x = \frac{1}{2}x^{\mathrm{T}}\mathbf{B}^{-1}x. \end{aligned}$$

Since $\mathbf{B}$ is a positive matrix, $h^{\mathrm{T}}\mathbf{B}^{-1}h > 0$, for any $h \neq 0$. Thus, $Q_w(\boldsymbol{\beta}^{(m+1)}|\boldsymbol{\beta}^{(m)}) - \ell_w(\boldsymbol{\beta}^{(m)}) = \frac{1}{2}x^{\mathrm{T}}\mathbf{B}^{-1}x \geq 0$, which means $Q_w(\boldsymbol{\beta}^{(m+1)}|\boldsymbol{\beta}^{(m)}) \geq \ell_w(\boldsymbol{\beta}^{(m)})$.

Notice that $\ell_w(\boldsymbol{\beta}^{(m+1)}) \geq Q_w(\boldsymbol{\beta}^{(m+1)}|\boldsymbol{\beta}^{(m)})$ for the reason that $H_w(\boldsymbol{\beta}) \leq \mathbf{B}$ in (10). Then, we can easily obtain the ascent property that $\ell_w(\boldsymbol{\beta}^{(m+1)}) \geq \ell_w(\boldsymbol{\beta}^{(m)})$. On the other hand, the working likelihood function $\ell_w(\boldsymbol{\beta})$ is also bounded above, illustrating that there exists a supremum of $\ell_w(\boldsymbol{\beta})$. Due to the fact that the parameter space $\mathcal{B}$ is compact and convex, the supremum can be obtained at the point of $\widehat{\boldsymbol{\beta}}$, which means that $\ell_w(\boldsymbol{\beta}^{(m)}) \leq \ell_w(\widehat{\boldsymbol{\beta}})$ for all m. According to monotone bounded theorem, we can learn that the sequence $\left\{\ell_w(\boldsymbol{\beta}^{(m)})\right\}$ is convergent to $\ell_w(\widehat{\boldsymbol{\beta}})$ monotonically, which is equivalent to the fact that the sequence $\left\{\boldsymbol{\beta}^{(m)}\right\}$ converges to $\widehat{\boldsymbol{\beta}}$. □

Proof of Theorem 2: Define $R_i := \{j|T_j \geq T_i\}$ as the risk set of T_i. We consider a more general situation where the weights change over time (Kang and Cai, 2009). The Hessian matrix of working likelihood function $\ell_w(\boldsymbol{\beta})$ can be obtained as follows:

$$H_w(\boldsymbol{\beta}) = \sum_{i=1}^n \Delta_i \left[\frac{\sum_{j\in R_i} w_j(T_i)\exp\left\{\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{Z}_j\right\}\boldsymbol{Z}_j\boldsymbol{Z}_j^{\mathrm{T}}}{\sum_{j\in R_i} w_j(T_i)\exp\left\{\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{Z}_j\right\}} - \left\{ \frac{\sum_{j\in R_i} w_j(T_i)\exp\left\{\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{Z}_j\right\}\boldsymbol{Z}_j}{\sum_{j\in R_i} w_j(T_i)\exp\left\{\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{Z}_j\right\}} \right\}^{\otimes 2} \right]. \tag{A.1}$$

Define

$$p_w^{ki}(T_i) = \frac{w_k(T_i)\exp\left\{\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{Z}_k\right\}}{\sum_{j\in R_i} w_j(T_i)\exp\left\{\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{Z}_j\right\}}, \quad k \in R_i. \tag{A.2}$$

For convenience, let $p_w^{ki} = p_w^{ki}(T_i)$. Thus, $H_w(\boldsymbol{\beta})$ can be rewritten as

$$H_w(\boldsymbol{\beta}) = \sum_{i=1}^n \Delta_i \left[\sum_{j\in R_i} \boldsymbol{Z}_j{\boldsymbol{Z}_j}^{\mathrm{T}} p_w^{ji} - \left\{ \sum_{j\in R_i} \boldsymbol{Z}_j p_w^{ji} \right\}^{\otimes 2} \right]. \tag{A.3}$$

Let $M_i = \sum_{j\in R_i} \boldsymbol{Z}_j{\boldsymbol{Z}_j}^{\mathrm{T}} p_w^{ji} - \left\{\sum_{j\in R_i} \boldsymbol{Z}_j p_w^{ji}\right\}^{\otimes 2}$. Consequently, for any $x \neq 0$, we have

$$x^{\mathrm{T}}M_i x = \sum_{j\in R_i} \left\{x^{\mathrm{T}}\boldsymbol{Z}_j\right\}^2 p_w^{ji} - \left[\sum_{j\in R_i} x^{\mathrm{T}}\boldsymbol{Z}_j p_w^{ji}\right]^2. \tag{A.4}$$

Consider Y as a discrete random variable, with support set $\left\{Y_j|Y_j = x^{\mathrm{T}}\boldsymbol{Z}_j, j \in R_i\right\}$ and the value Y_j is taken on with the probability p_w^{ji}. Therefore, we have

$$\begin{aligned} x^{\mathrm{T}}M_i x &= \sum_{j\in R_i}\left\{{Y_j}^2 p_w^{ji}\right\} - \left[\sum_{j\in R_i} Y_j p_w^{ji}\right]^2 \\ &= \mathrm{E}(Y^2) - \left[\mathrm{E}(Y)\right]^2. \end{aligned} \tag{A.5}$$

Obviously, the formula in equation (A.5) is the variance of the random variable Y. Let $min_{(i)}$ and $max_{(i)}$ represent minimum and maximum value of Y_j. Easily note that this varience of Y can be maximized when the $min_{(i)}$ and $max_{(i)}$ are both taken on with the probability 1/2. So equation (A.5) can be dominated by:

$$\begin{aligned} x^{\mathrm{T}}M_i x &\leq \frac{min_{(i)}^2 + max_{(i)}^2}{2} - \left[\frac{min_{(i)} + max(i)}{2}\right]^2 \\ &\leq \frac{min_{(i)}^2 + max_{(i)}^2}{2} \end{aligned}$$

$$\leq \frac{\sum_{j\in R_i} Y_i^2}{2} = \frac{\sum_{j\in R_i} x^{\mathrm{T}} \boldsymbol{Z}_j \boldsymbol{Z}_j^{\mathrm{T}} x}{2}.$$

Hence, $M_i \leq \sum_{j\in R_i} \boldsymbol{Z}_j \boldsymbol{Z}_j^{\mathrm{T}}/2$ and we can deduce that

$$H_w(\boldsymbol{\beta}) = \sum_{i\in R_i} \Delta_i M_i \leq \sum_{i\in R_i} \Delta_i \sum_{j\in R_i} \boldsymbol{Z}_j \boldsymbol{Z}_j^{\mathrm{T}}/2.$$

$\square$

Proof of Theorem 3: Set p a n-dimensional vector and define $D(p)$ as a diagonal matrix with matrix size n where the ith value p_i of the vector p is located in (i,i) of $D(p)$. Due to equation (A.4), this quadratic form can be rewritten as follows:

$$\begin{aligned} x^{\mathrm{T}} M_i x &= \sum_{j\in R_i} \left\{x^{\mathrm{T}} \boldsymbol{Z}_j\right\}^2 p_w^{ji} - \left[\sum_{j\in R_i} x^{\mathrm{T}} \boldsymbol{Z}_j p_w^{ji}\right]^2 \\ &= \left\{\boldsymbol{Z}^{\mathrm{T}} x\right\}^{\mathrm{T}} D(p_w^{(i)}) \left\{\boldsymbol{Z}^{\mathrm{T}} x\right\} \\ &\quad - \left\{\boldsymbol{Z}^{\mathrm{T}} x\right\}^{\mathrm{T}} p_w^{(i)} \left[\left\{\boldsymbol{Z}^{\mathrm{T}} x\right\}^{\mathrm{T}} p_w^{(i)}\right]^{\mathrm{T}} \\ &= \left\{\boldsymbol{Z}^{\mathrm{T}} x\right\}^{\mathrm{T}} \left[D(p_w^{(i)}) - p_w^{(i)} \left(p_w^{(i)}\right)^{\mathrm{T}}\right] \left\{\boldsymbol{Z}^{\mathrm{T}} x\right\}, \end{aligned} \tag{A.6}$$

where $p_w^{(i)} = \left(p_w^{1i}, p_w^{2i}, \cdots, p_w^{n_i i}\right)^{\mathrm{T}}, \boldsymbol{Z} = (\boldsymbol{Z}_1, \boldsymbol{Z}_2, \cdots, \boldsymbol{Z}_{n_i})$. n_i is the number of the risk set of $\boldsymbol{Z}_i$.

For convenience, let $\nu = \boldsymbol{Z}^{\mathrm{T}} x$ and $W(p) = D(p) - pp^{\mathrm{T}}$, which is a function of the probability vector p. Then we can obtain a concise form of equation (A.6):

$$x^{\mathrm{T}} M_i x = \nu^{\mathrm{T}} W(p_w^{(i)}) \nu. \tag{A.7}$$

From the proof of Theorem 2, we can learn that our goal is to find a upper bound of M_i without any unknown parameter $\boldsymbol{\beta}$ included. Notice that parameter $\boldsymbol{\beta}$ is only in the vector $p_w^{(i)}$, which belongs to the $n_i \times n_i$ matrix $W(p_w^{(i)})$. So we want to seek an upper bound of $W(p_w^{(i)})$, which is independent on $\boldsymbol{\beta}$. We consider such a proportional expression as follows:

$$\frac{y^{\mathrm{T}} W(p_w^{(i)}) y}{y^{\mathrm{T}} W(p^*) y}, \tag{A.8}$$

where $y \in R^{n_i}$ and $p^* = \left(\frac{1}{n_i}, \frac{1}{n_i}, \cdots, \frac{1}{n_i}\right)^{\mathrm{T}}$.

Then we will seek an upper bound of the formula above. For arbitrary $y \in R^{n_i}$ fixed, suppose that y_{max} is the maximum component and y_{min} is the minimum component of the vector y. If there exist more than one component to be maximal or minimal, choose the first maximum or minimum component we encounter in sequence to be y_{max} or y_{min}. We take the numerator into account and seek an upper bound of the numerator first. According to Theorem 2, the argument indicates the result as follows:

$$\begin{aligned} y^{\mathrm{T}} M_i y &= \sum_{j\in R_i} \left\{y^{\mathrm{T}} \boldsymbol{Z}_j\right\}^2 p_w^{ji} - \left[\sum_{j\in R_i} y^{\mathrm{T}} \boldsymbol{Z}_j p_w^{ji}\right]^2 \\ &\leq \frac{y_{max}^2 + y_{min}^2}{2} - \left[\frac{y_{max} + y_{min}}{2}\right]^2 \\ &= \frac{\left(y_{max} - y_{min}\right)^2}{4}. \end{aligned} \tag{A.9}$$

Set $\boldsymbol{Z}_j$ to be a unit vector, denoted by e_j where the ith component of the vector $\boldsymbol{Z}_j$ is 1 and the rest components are 0. Due to this condition, we can obtain that

$$\begin{aligned} M_i &= \sum_{j\in R_i} e_j e_j{}^{\mathrm{T}} p_w^{ji} - \left[\sum_{j\in R_i} e_j p_w^{ji}\right]^{\otimes 2} \\ &= \sum_{j\in R_i} diag(0, \cdots, p_w^{ji}, \cdots, 0) \\ &\quad - \left[\sum_{j\in R_i} (0, \cdots, 1, \cdots, 0)^{\mathrm{T}} p_w^{ji}\right]^{\otimes 2} \\ &= D(p_w^{(i)}) - p_w^{(i)} (p_w^{(i)})^{\mathrm{T}} = W(p_w^{(i)}). \end{aligned} \tag{A.10}$$

According to equations (A.9) and (A.10), we can deduce that, when arbitrary y from the total space is fixed,

$$y^{\mathrm{T}} W(p_w^{(i)}) y \leq \frac{\left(y_{max} - y_{min}\right)^2}{4}. \tag{A.11}$$

We then consider the denominator and search for a lower bound of the denominator. Notice that

$$\begin{aligned} y^{\mathrm{T}} W(p^*) y &= y^{\mathrm{T}} D(p^*) y - (y^{\mathrm{T}} p^*)^2 \\ &= \sum_{j\in R_i} y_j^2/n_i - (\sum_{j\in R_i} y_j/n_i)^2. \end{aligned} \tag{A.12}$$

We can regard equation (A.12) as a variance of a discrete random variable which has fixed mass $1/n_i$ at each supporting point y_j, where $j = 1, \cdots, n_i$. So our goal is equivalent to find a lower bound of the variance. Keep the maximum component y_{max} and the minimum component y_{min} fixed while consider all other components of the vector y as unknown variables of the variance function. Obviously, under this circumstances, the variance function is minimized when the mass $(n_i - 2)/n_i$ is put at the midrange. That is, for the discrete random variable possessing the minimum variance, all other components of its supporting points are equal to $(y_{max} + y_{min})/2$. Easy to calculate that the expectation of this random variable is $(y_{max} + y_{min})/2$. Hence this minimum variance can be seen as a lower bound of the denominator as follows when we fix an arbitrary y from the total space:

$$\begin{aligned} y^{\mathrm{T}} W(p^*) y &\geq \left(y_{max} - \frac{y_{max} + y_{min}}{2}\right) \frac{1}{n_i} \\ &\quad + \left(y_{min} - \frac{y_{max} + y_{min}}{2}\right) \frac{1}{n_i} \\ &= \frac{(y_{max} - y_{min})^2}{2n_i}. \end{aligned} \tag{A.13}$$

According to equations (A.11) and (A.13), we have found a upper bound of the numerator and a lower bound of the denominator of the proportion expression in (A.8) when the arbitrary y from the total space is fixed. So a upper bound of (A.8) can be obtained as follows:

$$\frac{y^{\mathrm{T}}W(p_w^{(i)})y}{y^{\mathrm{T}}W(p^*)y} \leq \frac{(y_{max}-y_{min})^2}{4}\Big/\frac{(y_{max}-y_{min})^2}{2n_i} = \frac{n_i}{2}. \tag{A.14}$$

Notice that this upper bound is independent on y, which means that the upper bound in (A.14) is a uniform and consistent bound of (A.8). So we can obtain that for all y from the total space, $y^{\mathrm{T}}W(p_w^{(i)})y \leq y^{\mathrm{T}}\frac{n_i}{2}W(p^*)y$. This is equivalent to the fact that

$$W(p_w^{(i)}) \leq \frac{n_i}{2}W(p^*). \tag{A.15}$$

Consequently, we have found a upper bound matrix $W(p^*)$, which doesn't contain any unknown parameter $\boldsymbol{\beta}$. According to (A.7) and (A.15), we can obtain that

$$\begin{aligned}
&x^{\mathrm{T}}M_ix\\
=&\nu^{\mathrm{T}}W(p_w^{(i)})\nu \leq \nu^{\mathrm{T}}\frac{n_i}{2}W(p^*)\nu\\
=&\frac{n_i}{2}\left\{\boldsymbol{Z}^{\mathrm{T}}x\right\}^{\mathrm{T}}\left[D(p^*)-p^*\left(p^*\right)^{\mathrm{T}}\right]\left\{\boldsymbol{Z}^{\mathrm{T}}x\right\}\\
=&\frac{n_i}{2}\left\{\boldsymbol{Z}^{\mathrm{T}}x\right\}^{\mathrm{T}}D(p^*)\left(\boldsymbol{Z}^{\mathrm{T}}x\right)-\frac{n_i}{2}\left\{\boldsymbol{Z}^{\mathrm{T}}x\right\}^{\mathrm{T}}p^*\left[\left\{\boldsymbol{Z}^{\mathrm{T}}x\right\}^{\mathrm{T}}p^*\right]^{\mathrm{T}}\\
=&\frac{n_i}{2}\sum_{j\in R_i}\left\{x^{\mathrm{T}}\boldsymbol{Z}_j\right\}^2\frac{1}{n_i}-\frac{n_i}{2}\left[\sum_{j\in R_i}x^{\mathrm{T}}\boldsymbol{Z}_j\frac{1}{n_i}\right]^2\\
=&\sum_{j\in R_i}x^{\mathrm{T}}\boldsymbol{Z}_j\boldsymbol{Z}_j^{\mathrm{T}}x/2-x^{\mathrm{T}}\sum_{j\in R_i}\boldsymbol{Z}_j\left\{\sum_{j\in R_i}\boldsymbol{Z}_j\right\}^{\mathrm{T}}x\Big/(2n_i)\\
=&x^{\mathrm{T}}\left[\sum_{j\in R_i}\boldsymbol{Z}_j\boldsymbol{Z}_j^{\mathrm{T}}/2-\left\{\sum_{j\in R_i}\boldsymbol{Z}_j\right\}^{\otimes 2}\Big/(2n_i)\right]x.
\end{aligned}$$

Hence

$$M_i \leq \left[\sum_{j\in R_i}\boldsymbol{Z}_j\boldsymbol{Z}_j^{\mathrm{T}}-\left\{\sum_{j\in R_i}\boldsymbol{Z}_j\right\}^{\otimes 2}\Big/n_i\right]\Big/2, \tag{A.16}$$

and we can deduce that

$$\begin{aligned}
H_w(\boldsymbol{\beta}) &= \sum_{i\in R_i}\Delta_iM_i\\
&\leq \sum_{i\in R_i}\Delta_i\left[\left\{\sum_{j\in R_i}\boldsymbol{Z}_j\right\}^{\otimes 2}\Big/n_i-\sum_{j\in R_i}\boldsymbol{Z}_j\boldsymbol{Z}_j^{\mathrm{T}}\right]\Big/2. \qquad \text{(A.17)}
\end{aligned}$$

□

Proof of Theorem 4: From the proof of theorem 1 we can learn that

$$Q_w(\boldsymbol{\beta}^{(m+1)}|\boldsymbol{\beta}^{(m)})-\ell_w(\boldsymbol{\beta}^{(m)}) = \frac{1}{2}U_w(\boldsymbol{\beta}^{(m)})^{\mathrm{T}}B_w^{-1}U_w(\boldsymbol{\beta}^{(m)}). \tag{A.18}$$

Note that the rate of convergence, measured by degree of improvement at each iterative step is monotone decreasing with the least bound matrix $\mathbf{B}$ from (A.18). And

$$\begin{aligned}
\mathbf{B}_1-\mathbf{B}_2 &= \sum_{i=1}^n\Delta_i\sum_{j\in R_i}\boldsymbol{Z}_j\boldsymbol{Z}_j^{\mathrm{T}}/2\\
&\quad-\sum_{i=1}^n\Delta_i\left[\sum_{j\in R_i}\boldsymbol{Z}_j\boldsymbol{Z}_j^{\mathrm{T}}-\left\{\sum_{j\in R_i}\boldsymbol{Z}_j\right\}^{\otimes 2}\Big/n_i\right]\Big/2\\
&= \sum_{i=1}^n\Delta_i\left[\left\{\sum_{j\in R_i}\boldsymbol{Z}_j\right\}^{\otimes 2}\Big/n_i\right]\Big/2 \geq 0.
\end{aligned}$$

As a consequence that $\mathbf{B}_1 \geq \mathbf{B}_2$, we can deduce that $r_1 \leq r_2$. □

Received 10 January 2022

REFERENCES

[1] Andersen, P. K. and Gill, R. D. (1982). Cox's regression model for counting processes: a Large sample study. *The Annals of Statistics*, **10(4)**, 1100–1120. MR0673646

[2] Andersen, P. K., Klein, J. P., Knudsen, K. M., and Palacios, R. T. (1997). Estimation of variance in Cox's regression model with gamma frailties. *Biometrics*, **53(4)**, 1475–1484. MR1614394

[3] Barlow, W. (1994). Robust variance estimation for the case-cohort design. *Biometrics*, **50(4)**, 1064–1072.

[4] Becker, M.P., Yang, I., and Lange, K. (1997). EM algorithms without missing data. *Statistical Methods in Medical Research*, **6(1)**, 38–54.

[5] Borgan, O., Langholz, B., Samuelsen, S., Goldstein, L. and Pogoda, J. (2000). Exposure stratified case-cohort designs. *Lifetime Data Analysis*, **6(1)**, 39–58. MR1767493

[6] Böhning, D. and Lindsay, B. G. (1988). Monotonicity of quadratic approximation algorithms. *Annals of the Institute of Statistical Mathematics*, **40(4)**, 641–663. MR0996690

[7] Breslow, N. E. and Chatterjee, N. (1999). Design and analysis of two-phase studies with binary outcome applied to Wilms tumor prognosis. *Journal of the Royal Statistical Society, Series C*, **48(4)**, 457–468.

[8] Breslow, N. E. and Wellner, J. A. (2007). Weighted likelihood for semiparametric models and two-phase stratified samples, with application to cox regression. *Scandinavian Journal of Statistics*, **34(1)**, 86–102. MR2325244

[9] Cai, J. and Zeng, D. (2004). Sample size/power calculation for case-cohort studies. *Biometrics*, **60(4)**, 1015–1024. MR2133554

[10] Cai, J. and Zeng, D. (2007). Power calculation for case-cohort studies with nonrare events. *Biometrics*, **63(4)**, 1288–1295, 1316. MR2414609

[11] Chen, K. and Lo, S. (1999). Case-cohort and case-control analysis with Cox's model. *Biometrika*, **86(4)**, 755–764. MR1741975

[12] Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, **34(2)**, 187–220. MR0341758

[13] Cox, D. R. (1975). Partial likelihood. *Biometrika*, **62(2)**, 269–276. MR0400509

[14] Dan Toomey. (2014). R for Data Science. Birmingham: Packet Publishing.

[15] D'Angio, G. J., Breslow, N., Bechwith, J. B., Evans, A., Baum, E., deLorimier, A., Fernbach, D., Hrabovsky, E., Jones, B., Kelalis, P., Otherson, H. B., and Tefft, M. (1989) Treatment of Wilms' tumor, results of the third national Wilms' tumor study. *Cancer*, **64**, 349–360.

[16] De Pierro, A. R. (1995). A modified expectation maximization algorithm for penalized likelihood estimation in emission tomography. *IEEE Transactions on Medical Imaging*, **14**, 132–137.

[17] Deng, L., Ding, J., Liu, Y., and Wei, C. (2018). Regression analysis for the proportional hazards model with parameter constraints under case-cohort design. *Computational Statistics and Data Analysis*, **117**, 194–206. MR3711490

[18] Ding, J., Zhou, H., Liu, Y., Cai, J., and Longnecker, M. P. (2014). Estimating effect of environmental contaminants on women's subfecundity for the MoBa study data with an outcome-dependent sampling scheme. *Biostatistics*, **15(4)**, 636–650.

[19] Ding, J., Tian G. L., and Kam C. Y. (2015). A new MM algorithm for constrained estimation in the proportional hazards model. *Computational Statistics and Data Analysis*, **84**, 135–151. MR3292802

[20] Eddelbuettel, D. and Francois, R. (2011). Rcpp: Seamless R and C++ integration. *Journal of Statistical Software, Articles*, **40**, 1–18.

[21] Eddelbuettel, D. and Sanderson, C. (2014). Rcpparmadillo: Accelerating R with high-performance C++ linear algebra. *Computational Statistics and Data Analysis*, **71**, 1054–1063. MR3132026

[22] Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, **96(456)**, 1348–1360. MR1946581

[23] Fan, J. and Li, R. (2002). Variable selection for Cox's proportional hazards model and frailty model. *The Annals of Statistics*, **30(1)**, 74–99. MR1892656

[24] Foutz, R. V. (1977). On the unique consistent solution to the likelihood equations. *Journal of the American Statistical Association*, **72(357)**, 147–148. MR0445686

[25] Green, D. M., Breslow, N. E., Bechwith, J. B., Finklestein, J. Z., Grundy, P. E., Thomas, P. R., Kim, T., Shochat, S. J., Haase, G. M., Ritchey, M. L., Kelalis, P. P., and D'Angio, G. J. (1998). Comparison between single-dose and divided-dose administration of dactinomycin and doxorubicin for patients with Wilms' tumor: a report from the national Wilms' tumor study group. *Journal of Clinical Oncology*, **16**, 237–245.

[26] Horvitz, D. G. and Thompson, D. J. (1951). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, **47**, 663–685. MR0053460

[27] Huang, J., Ma,S., and Zhang, C. H. (2008). Adaptive Lasso for sparse high-dimensional regression models. *Statistica Sinica*, **18(4)**, 1603–1618. MR2469326

[28] Huang, J., Breheny, P., Zhang, C., and Zhang, C. H. (2010). The mnet method for variable selection. *Statistica Sinica*, **26(3)**, 903–923. MR3559936

[29] Hunter, D. R. and Lange, K. (2002). Computing estimates in the proportional odds model. *Annals of the Institute of Statistical Mathematics*, **54(1)**, 155–168. MR1893548

[30] Hunter, D. R. and Lange, K. (2004). A tutorial on MM algorithms. *The American Statistician*, **58(1)**, 30–37. MR2055509

[31] Kalbfleisch, J. D. and Lawless, J. F. (1988). Likelihood analysis of multi-state models for disease incidence and mortality. *Statistics in Medicine*, **7**, 147–160.

[32] Kang, S. and Cai, J. (2009). Marginal hazard model for case-cohort studies with multiple disease outcomes. *Biometrika*, **96(4)**, 887–901. MR2767277

[33] Kang, S., Cai, J., and Chambless, L. (2013). Marginal additive hazards model for case-cohort studies with multiple disease outcomes: an application to the Atherosclerosis Risk in Communities (ARIC) study. *Biostatistics*, **14(1)**, 28–41.

[34] Kang, S., Lu, W., and Liu, M. (2016). Efficient estimation for accelerated failure time model under case-cohort and nested case-control sampling. *Biometrics*, **73(1)**, 114–123. MR3632357

[35] Kim, S., Zheng, D., and Cai, J. (2018). Analysis of multiple survival events in generalized case-cohort designs. *Biometrics*, **74(4)**, 1250–1260. MR3908143

[36] Kong, L., Cai, J., and Sen, P. K. (2004). Weighted estimating equations for semiparametric transformation models with censored data from a case-cohort design. *Biometrika*, **91(2)**, 305–319. MR2081303

[37] Kong, L. and Cai, J. (2009). Case-cohort analysis with accelerated failure time model. *Biometrics*, **65(1)**, 135–142. MR2665853

[38] Kulich, M. and Lin, D. Y. (2000). Additive hazards regression for case-cohort studies. *Biometrika*, **87(1)**, 73–87. MR1766829

[39] Kulich, M. and Lin, D. Y. (2004). Improving the efficiency of relative-risk estimation in case-cohort studies. *Journal of the American Statistical Association*, **99(467)**, 832–844. MR2090916

[40] Lange, K., Hunter, D. R., and Yang, I. (2000). Optimization transfer using surrogate objective functions (with discussion). *Journal of the Computational and Graphical Statistics*, **9**, 1–20. MR1819865

[41] Lange, K. (2004). Optimization. New York: Springer. MR2072899

[42] Lange, K. (2010). Numerical Analysis for Statisticians. New York: Springer. MR2655999

[43] Lin, D. Y. and Ying, Z. (1993). Cox regression with incomplete covariate measurements. *Journal of the American Statistical Association*, **88(424)**, 1341–1349. MR1245368

[44] Lin, D., Wei, L., Yang, I., and Ying, Z. (2000). Semiparametric regression for the mean and rate functions of recurrent events. *Journal of the Royal Statistical Society*, **62(4)**, 711–730. MR1796287

[45] Liu, D., Cai, T., Lok, A., and Zheng, Y. (2018). Nonparametric maximum likelihood estimators of time-dependent accuracy measures for survival outcome under two-stage sampling designs. (2018). *Journal of the American Statistical Association*, **113(522)**, 882–892. MR3832234

[46] Lu, W. and Tsiatis, A. A. (2006). Semiparametric transformation models for the case-cohort study. *Biometrika*, **93(1)**, 207–214. MR2277751

[47] Maitra, P., Amorim, L.D.A.F., and Cai, J. (2020). Multiplicative rates model for recurrent events in case-cohort studies. *Lifetime Data Analysis*, **26(1)**, 134–157. MR4048042

[48] Prentice, R. L. (1986). A case-cohort design for epidemiologic cohort studies and disease prevention trials. *Biometrika*, **73(1)**, 1–11.

[49] Self, S. G. and Prentice, R. L. (1988). Asymptotic distribution theory and efficiency results for case-cohort studies. *The Annals of Statistics*, **16(1)**, 64–81. MR0924857

[50] Spiekerman, C. F. and Lin, D. Y. (1988). Marginal regression models for multivariate failure time data. *Journal of the American Statistical Association*, **93(443)**, 1164–1175. MR1649210

[51] Steingrimsson, J. A. and Strawderman, R. L. (2017). Estimation in the semiparametric accelerated failure time model with missing covariates: improving efficiency through augmentation. *Journal of the American Statistical Association*, **112(519)**, 1221–1235. MR3735372

[52] Sun, J., Sun, L. and Flournoy, N. (2004). Addictive hazards model for competing risks analysis of the case-cohort design. *Communication in Statistics – Theory and Method*, **33**, 351–366. MR2045320

[53] Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society, Series B*, **58(1)**: 267–288. MR1379242

[54] Tibshirani, R. (1997). The Lasso method for variable selection in the Cox model. *Statistics in medicine*, **16**, 385–395.

[55] van der Vaart, A. W. and Wellner, J. A. (1996). Weak

Convergence and Empirical Processes. New York: Springer. MR1385671

[56] Yu, J., Liu, Y., Cai, J., Sandler, D. P., and Zhou, H. (2016). Outcome-dependent sampling design and inference for Cox's proportional hazards Model. *Journal of Statistical Planning and Inference*, **178**, 24–36. MR3531525

[57] Zou, H. (2006). The adaptive Lasso and its oracle properties. *Journal of the American Statistical Association*, **101(476)**, 1418–1429. MR2279469

Jieli Ding
School of Mathematics and Statistics
Wuhan University
Wuhan
China
E-mail address: jlding.math@whu.edu.cn

Jiaqian Zhang
School of Mathematics and Statistics
Wuhan University
Wuhan
China E-mail address: 2015301470037@whu.edu.cn

Yanqin Feng
School of Mathematics and Statistics
Wuhan University
Wuhan
China
E-mail address: yqfeng.math@whu.edu.cn

Yuxuan Du
School of Mathematics and Statistics
Wuhan University
Wuhan
China
E-mail address: dyxstat@qq.com