Title

Debiased lasso for stratified Cox models with application to the national kidney transplant data.

Journal

The Annals of Applied Statistics, 17(4)

ISSN

1932-6157

Authors

Xia, Lu

Nan, Bin

Li, Yi

Publication Date

2023-12-01

DOI

10.1214/23-aoas1775

Peer reviewed

HHS Public Access

Author manuscript

Ann Appl Stat. Author manuscript; available in PMC 2024 December 01.

Published in final edited form as:

Ann Appl Stat. 2023 December ; 17(4): 3550–3569. doi:10.1214/23-aoas1775.

Debiased lasso for stratified Cox models with application to the national kidney transplant data

Lu Xia[1], **Bin Nan**[2], **Yi Li**[3]

[1]Department of Biostatistics, University of Washington

[2]Department of Statistics, University of California, Irvine

[3]Department of Biostatistics, University of Michigan

Abstract

The Scientific Registry of Transplant Recipients (SRTR) system has become a rich resource for understanding the complex mechanisms of graft failure after kidney transplant, a crucial step for allocating organs effectively and implementing appropriate care. As transplant centers that treated patients might strongly confound graft failures, Cox models stratified by centers can eliminate their confounding effects. Also, since recipient age is a proven non-modifiable risk factor, a common practice is to fit models separately by recipient age groups. The moderate sample sizes, relative to the number of covariates, in some age groups may lead to biased maximum stratified partial likelihood estimates and unreliable confidence intervals even when samples still outnumber covariates. To draw reliable inference on a comprehensive list of risk factors measured from both donors and recipients in SRTR, we propose a de-biased lasso approach via quadratic programming for fitting stratified Cox models. We establish asymptotic properties and verify via simulations that our method produces consistent estimates and confidence intervals with nominal coverage probabilities. Accounting for nearly 100 confounders in SRTR, the de-biased method detects that the graft failure hazard nonlinearly increases with donor's age among all recipient age groups, and that organs from older donors more adversely impact the younger recipients. Our method also delineates the associations between graft failure and many risk factors such as recipients' primary diagnoses (e.g. polycystic disease, glomerular disease, and diabetes) and donor-recipient mismatches for human leukocyte antigen loci across recipient age groups. These results may inform the refinement of donor-recipient matching criteria for stakeholders.

Keywords

Confidence intervals; diverging number of covariates; end-stage renal disease; graft failure free survival; statistical inference

xialu@uw.edu .

SUPPLEMENTARY MATERIAL

Supplement to "De-biased lasso for stratified Cox models with application to the national kidney transplant data"
The Supplementary Material includes technical lemmas and proofs of Theorems 3.1 and 3.4.

R code for "De-biased lasso for stratified Cox models with application to the national kidney transplant data"
The R code document includes the implementation of the proposed de-biased lasso method via quadratic programming with simulated examples, and introduction to the main functions.

1. Introduction.

For patients with end-stage renal disease, one of the most lethal and prevalent diseases in the U.S. (Saran et al., 2020), successful renal transplantation is effective for improving quality of life and prolonging survival (Wolfe et al., 1999; Kostro et al., 2016; Ju et al., 2019). The success of kidney transplantation hinges upon various factors related to the quality of transplant operations, the quality of donated kidneys, and the physical conditions of recipients (Rodger, 2012; Legendre, Canaud and Martinez, 2014), and it is crucial to evaluate and understand how these risk factors impact on renal graft failure in order to increase the chance of success (Hamidi et al., 2016; Legendre, Canaud and Martinez, 2014). With the scarcity of organs and an increasing number of waitlisted candidates (Bastani, 2015), the results can inform more efficient strategies for kidney allocation (Rao and Ojo, 2009; Smith et al., 2012) as well as evidence-based post-transplant care (Baker et al., 2017). Therefore, how to quantify the impacts of important factors associated with prognosis, particularly renal graft failure, remains to be a central question in kidney transplantation. The Scientific Registry of Transplant Recipients (SRTR) system, a federally funded organization that keeps records of transplant information from recipients and donors, has become a rich resource for studying post-kidney transplantation prognosis (Dickinson et al., 2008).

Leveraging the SRTR data, one can develop a valid tool for characterizing the influences of risk factors on graft failure, a key step towards post-transplant prognosis. Most previous studies, which focused only on a small number of factors, i.e. kidney diagnosis, recipient age, recipient race, recipient gender, number of human leukocyte antigen (HLA) mismatches, donor age, donor race, donor gender, serum creatinine level and cold ischemia time (Alexander, Bennett and Breen, 1994), might have pre-excluded other important factors and not fully captured the complex mechanisms governing graft failure. The SRTR data contain comprehensive information on recipients and donors, such as recipient primary insurance and employment, procedure type, infection of multiple viruses, history of transplant, transfusion and drug abuse, and pre-transplant comorbidities. The data provide a unique opportunity for assessing the associations between graft failure and an extended list of variables simultaneously, which may reduce confounding (Wang, 2011). Specifically, since donor age is a major criterion for donor-recipient matching (Kasiske and Snyder, 2002; Rao et al., 2009; Veroux et al., 2012), the data enable us to examine its effect on graft failure by adjusting for confounders, including pre-existing comorbidities.

There are several statistical challenges. On the one hand, as recipients received care in various transplant centers, the center-specific effects may confound the covariate effects of interest. This motivates us to consider Cox models stratified by transplant centers, a commonly used model in the relevant context without the need to explicitly model the potentially time-varying center effects (He et al., 2021). On the other hand, recipient age is a strong risk factor and there may exist complicated interactions between recipients' age and other characteristics (Keith et al., 2004). For ease of interpretation and by convention (Morales et al., 2012; Faravardeh et al., 2013), we have opted to divide our analyzable patient population (the adult recipients with kidneys transplanted during 2000 and 2001) into [18,45], (45,60] and 60+ years old groups (Table 1), and fit models separately for

these three groups. Allowing model parameters to be age group-specific, we have avoided parametrically modeling the interactions between recipient age and the other risk factors. When the number of covariates is relatively large (94 in our data) compared to, though still less than, the sample size (for example, 1448 patients with 1013 events in the 60+ years old recipient group), the conventional maximum stratified partial likelihood estimation (MSPLE) may yield untrustworthy point estimates, confidence intervals and hypothesis testing results, as illustrated in our simulations.

For proper inferences, we consider an asymptotic framework with a diverging number of covariates, wherein the number of covariates, though smaller than the sample size, can increase with the sample size (He and Shao, 2000; Wang, 2011). Lasso provides a very popular tool for simultaneous variable selection and estimation with high-dimensional covariates (Tibshirani, 1997). For unstratified Cox models, Huang et al. (2013) and Kong and Nan (2014) presented the oracle inequalities for the lasso estimator. However, with penalization, lasso estimates are biased towards zero (van de Geer et al., 2014), and they do not possess regular limiting distributions even under linear regression with a fixed number of covariates (Fu and Knight, 2000). Conditional inference based on the selected model is invalid, either, due to the failure to account for uncertainty in model selection. Hence, lasso cannot be directly applied to draw statistical inference. There is literature on inference for unstratified Cox proportional hazards models under the related asymptotic framework. For example, Fang, Ning and Liu (2017) proposed decorrelated score tests for a low-dimensional component in the regression parameters, and Kong et al. (2021), Yu, Bradic and Samworth (2021) and Xia, Nan and Li (2022) proposed to correct asymptotic biases of the lasso estimator following the framework of van de Geer et al. (2014), Zhang and Zhang (2014) and Javanmard and Montanari (2014) that were originated from high-dimensional linear or generalized linear models. For Cox models, all of these methods, except Xia, Nan and Li (2022) which considered the "large n, diverging p" scenario, assumed sparsity on the inverse information matrix. This sparse matrix assumption, however, may not hold for models beyond linear regression, leading to insufficient bias correction and under-covered confidence intervals. Moreover, as these methods were not designed for modeling stratified data, they are not directly applicable to the analysis of the SRTR data. To our knowledge, the current literature lacks inferential methods with theoretical rigor for stratified Cox models with a diverging number of covariates.

We propose a de-biased lasso approach for Cox models stratified by transplant centers, which solves a series of quadratic programming problems to estimate the inverse information matrix, and corrects the biases from the lasso estimator for valid statistical inference. Our asymptotic results enable us to draw inference on any linear combinations of model parameters, including the low-dimensional targets in Fang, Ning and Liu (2017) and Kong et al. (2021) as special cases and fundamentally deviating from the stepwise regression adopted by Rao et al. (2009). When the number of covariates is relatively large compared to the sample size, our approach yields less biased estimates and more properly covered confidence intervals than MSPLE as well as the methods of Fang, Ning and Liu (2017); Kong et al. (2021); Yu, Bradic and Samworth (2021) adapted to the stratified

setting. Therefore, it is well-suited for analyzing the SRTR data, especially among the oldest recipient group that has the smallest sample size.

Applications of our method to the SRTR data have generated reliable estimation and inference results for the effects of an expanded list of donor and recipient factors. We find that receiving kidneys from older donors is associated with an increased hazard of graft failure after adjusting for many confounding factors, and that the dependence on donors' age is non-linear. The results may inform more comprehensive assessments of post-transplant prognosis and kidney allocation.

The article is organized as follows. We introduce the proposed de-biased lasso approach in Section 2 and establish the asymptotic results in Section 3, which form the basis of inference for the SRTR data. We conduct simulations in Section 4 and demonstrate that our method outperforms MSPLE in bias correction and confidence interval coverage. In Section 5, we analyze the SRTR data by using the proposed de-biased approach. Finally, we provide a few concluding remarks in Section 6j, the detailed list of covariates considered in the analysis of SRTR data in Appendix A, and regularity conditions in Appendix B. Technical details and proofs are deferred to the Supplementary Material.

2. De-biased lasso for stratified Cox models via quadratic programming.

We apply stratified Cox models to evaluate the impacts of risk factors on post-transplant graft failure using the SRTR data. For each recipient age group defined in the first row of Table 1, let K be the total number of transplant centers, and n_k be the number of recipients in the k-th transplant center, $k = 1, \cdots, K$. With i indexing recipients within the k-th transplant center, let T_{ki} denote the graft failure free survival time, i.e. the time from transplantation to graft failure or death, whichever comes first [a common endpoint in transplantation (Kasiske et al., 2011)], X_{ki} be a p-dimensional covariate vector, and C_{ki} be the censoring time. We assume random censoring, that is, T_{ki} and C_{ki} are independent given X_{ki}. In the SRTR data, $p = 94$ and X_{ki} includes risk factors from both donors and recipients, such as gender, ABO blood type, history of diabetes and duration, angina/coronary artery disease, symptomatic peripheral vascular disease, drug treated systemic hypertension, drug treated COPD, and mismatch for each HLA locus between donors and recipients; see a full list of covariates in Appendix A. Let $\delta_{ki} = 1(T_{ki} \leq C_{ki})$ be the event indicator and $Y_{ki} = \min(T_{ki}, C_{ki})$ be the observed time. With a center-specific baseline hazard function $\lambda_{0k}(t)$, a stratified Cox model for T_{ki} stipulates that its conditional hazard at t given X_{ki} is

$$\lambda_{ki}(t \mid X_{ki}) = \lambda_{0k}(t)\exp\left\{X_{ki}^T\beta^0\right\},$$

where $\beta^0 = \left(\beta_1^0, \ldots, \beta_p^0\right)^T \in \mathbb{R}^p$ is the vector of common regression coefficients across all centers. It is reasonable to assume that the true regression coefficients β^0 are the same across strata (Kalbfleisch and Prentice, 2002), while the center effects, though not of primary interest here, are accounted for via different baseline hazards $\lambda_{0k}(t)$'s.

2.1. Estimation method.

The MSPLE of β minimizes the following *negative* log stratified partial likelihood function

$$\ell(\beta) = -\frac{1}{N}\sum_{k=1}^{K}\sum_{i=1}^{n_k}\left[\beta^T X_{ki} - \log\left\{\frac{1}{n_k}\sum_{j=1}^{n_k} 1(Y_{kj} \geq Y_{ki})\exp\left(\beta^T X_{kj}\right)\right\}\right]\delta_{ki}, \tag{2.1}$$

where $N = \sum_{k=1}^{K} n_k$. In SRTR, the number of risk factors, though smaller than the sample size, is fairly large. In this case, our numerical examination shows that MSPLEs are biased and their confidence intervals do not yield nominal coverage. We consider a de-biased approach that has been shown to yield valid inference in linear regression (van de Geer et al., 2014; Zhang and Zhang, 2014; Javanmard and Montanari, 2014). Here we assume that $p < N$ but grows with N, which falls into the "large N, diverging p" framework. We extend the debiased lasso to accommodate stratified Cox models.

For a vector $x = (x_1, \ldots, x_p)^T \in \mathbb{R}^p$, define $x^{\otimes 0} = 1,\ x^{\otimes 1} = x$ and $x^{\otimes 2} = xx^T$. Let $\dot{\ell}(\beta)$ and $\ddot{\ell}(\beta)$ be the first and the second order derivatives of $\ell(\beta)$ with respect to β, i.e.

$$\dot{\ell}(\beta) = -\frac{1}{N}\sum_{k=1}^{K}\sum_{i=1}^{n_k}\left\{X_{ki} - \frac{\hat{\mu}_{1k}(Y_{ki};\ \beta)}{\hat{\mu}_{0k}(Y_{ki};\ \beta)}\right\}\delta_{ki},$$

$$\ddot{\ell}(\beta) = \frac{1}{N}\sum_{k=1}^{K}\sum_{i=1}^{n_k}\left\{\frac{\hat{\mu}_{2k}(Y_{ki};\ \beta)}{\hat{\mu}_{0k}(Y_{ki};\ \beta)} - \left[\frac{\hat{\mu}_{1k}(Y_{ki};\ \beta)}{\hat{\mu}_{0k}(Y_{ki};\ \beta)}\right]^{\otimes 2}\right\}\delta_{ki},$$

where $\hat{\mu}_{rk}(t;\beta) = n_k^{-1}\sum_{j=1}^{n_k} 1(Y_{kj} \geq t)X_{kj}^{\otimes r}\exp\{X_{kj}^T\beta\},\ r = 0,\ 1,\ 2$. The lasso estimate, $\hat{\beta}$, minimizes the penalized negative log stratified partial likelihood,

$$\hat{\beta} = \text{argmin}_{\beta\in\mathbb{R}^p}\{\ \ell(\beta) + \lambda\|\beta\|_1\}, \tag{2.2}$$

where $\lambda > 0$ is a tuning parameter that encourages sparse solutions. Here, $\|x\|_q = \left(\sum_{j=1}^{p}|x_j|^q\right)^{1/q}$ is the ℓ_q-norm for $x \in \mathbb{R}^p,\ q \geq 1$.

As $\hat{\beta}$ is typically biased, we can obtain the de-biased lasso estimator by a Taylor expansion of $\dot{\ell}(\beta^0)$ around $\hat{\beta}$. To proceed, let $\widehat{M}$ be a $p \times p$ matrix and $\widehat{M}_j$ its jth row. Pre-multiplying $\widehat{M}_j$ on both sides of the Taylor expansion and collecting terms, we have the following equality for the jth component of β:

$$\hat{\beta}_j - \beta_j^0 + \overbrace{\left(-\widehat{M}_j\,\dot{\ell}(\hat{\beta})\right)}^{I_j} + \overbrace{\left(-\widehat{M}_j\Delta\right)}^{II_j} + \overbrace{\left(\widehat{M}_j\,\ddot{\ell}(\hat{\beta}) - e_j^T\right)\left(\hat{\beta} - \beta^0\right)}^{III_j} = -\widehat{M}_j\,\dot{\ell}(\beta^0), \tag{2.3}$$

where the remainder $\Delta \in \mathbb{R}^p$ in II_j can be shown asymptotically negligible given the convergence rate of the lasso estimator $\hat{\beta}$, and so is III_j if $\widehat{M}_j\,\ddot{\ell}(\hat{\beta}) - e_j^T$ converges to zero

with certain rate that will be discussed later in Section 3. Hence, the de-biased lasso estimator corrects the bias of $\hat{\beta}_j$ with a one-step update of

$$\hat{b}_j = \hat{\beta}_j - \widehat{\Theta}_j \dot{\ell}(\hat{\beta}), \tag{2.4}$$

which replaces $\widehat{M}_j$ in (2.3) with the j-th row of $\widehat{\Theta}$, an estimate of the inverse information matrix Θ_{β^0}, and $-\widehat{\Theta}_j \dot{\ell}(\hat{\beta})$ is the bias correction term to $\hat{\beta}_j$. Here, Θ_{β^0} is the inverse of the population version of $\widehat{\Sigma}$ given in the following (2.6); see the explicit definition of Θ_{β^0} underneath (3.1). Denote by $\hat{b} = (\hat{b}_1, \ldots, \hat{b}_p)^T$ the vector of the de-biased lasso estimates, and, for compactness, write (2.4) in a matrix form

$$\hat{b} = \hat{\beta} - \widehat{\Theta}\, \dot{\ell}(\hat{\beta}). \tag{2.5}$$

Unlike $\hat{\beta}$, the de-biased estimator $\hat{b}$ in (2.5) is no longer sparse. Motivated by Javanmard and Montanari (2014) on high-dimensional inference in linear regression, we propose to obtain $\widehat{\Theta}$ by solving a series of quadratic programming problems. First, we compute

$$\widehat{\Sigma} = \frac{1}{N} \sum_{k=1}^{K} \sum_{i=1}^{n_k} \delta_{ki} \left[X_{ki} - \hat{\eta}_k(Y_{ki}; \hat{\beta}) \right]^{\otimes 2}, \tag{2.6}$$

where $\hat{\eta}_k(t;\beta) = \hat{\mu}_{1k}(t;\beta)/\hat{\mu}_{0k}(t;\beta)$ is the vector of weighted average covariates. We use $\widehat{\Sigma}$, in lieu of $\ddot{\ell}(\hat{\beta})$, for ease of proving theoretical properties. Indeed, as shown in the Supplementary Material, $\| \widehat{\Sigma} - \ddot{\ell}(\hat{\beta}) \|_\infty \xrightarrow{p} 0$ with a desirable rate under the conditions in Section 3. Next, for each $j = 1, \ldots, p$, we solve a quadratic programming problem

$$\min_{m \in \mathbb{R}^p} \left\{ m^T \widehat{\Sigma} m : \| \widehat{\Sigma} m - e_j \|_\infty \le \gamma \right\}, \tag{2.7}$$

where $\gamma > 0$ is a tuning parameter that is different from the lasso tuning parameter λ, e_j is a unit directional vector with only the jth element being one, and $\| \cdot \|_\infty$ is the matrix max norm, i.e. $\| A \|_\infty = \max_{i,j} |A_{ij}|$ for a real matrix A. Denote by $m^{(j)}$ the column vector of solution to (2.7). We obtain a $p \times p$ matrix $\widehat{\Theta} = \left(m^{(1)}, \ldots, m^{(p)} \right)^T$.

The constraint $\| \widehat{\Sigma} m - e_j \|_\infty \le \gamma$ in (2.7) controls deviations of the de-biased estimates from the lasso estimates. In an extreme case of $\gamma = 1$, an admissible solution is $m = 0$, and therefore there is no bias correction in the de-biased estimator; in another extreme case of $\gamma = 0$, $m^{(j)}$ is the jth column of $\widehat{\Sigma}^{-1}$. We implement (2.7) by using R solve.QP(), which can be programmed in parallel for large p. We name the method *de-biased lasso via quadratic programming* (hereafter, DBL-QP).

2.2. Tuning parameter selection.

For the DBL-QP method, the lasso tuning parameter λ can be selected via 5-fold cross-validation as in Simon et al. (2011). The selection of γ is crucial as, for example, Figure 1

reveals that γ should be selected within a specific range (shaded in figures) to achieve the most desirable bias correction and confidence interval coverage probability. It also shows the large bias and poor coverage resulting from MSPLE. Inappropriate tuning can yield even more biased estimates with poorer coverage than MSPLE. Results of lasso and oracle estimates are also provided as references, where oracle estimates are obtained from the reduced model that only contains truly nonzero coefficients.

Intuitively, γ should be chosen near zero, resulting in a de-biased estimator with estimates of large coefficients close to oracle estimates. We do not recommend evaluating crossvalidation criteria by plugging in the de-biased estimates because of accumulative estimation errors. We opt for a hard-thresholding approach that more effectively removes noise from the de-biased estimates: we retain the de-biased lasso estimate for β_j^0 only if the null hypothesis $\beta_j^0 = 0$ is rejected; otherwise, we set it to zero (shown in Algorithm 1). The set $\widehat{A}$ in Step 2.2 of Algorithm 1 is expected to estimate well the set of truly associated variables. Specifically, we set $\widehat{A}$ to be the index set of variables whose Wald statistic $\sqrt{N}\left|\hat{b}_j\right|/\widehat{\Theta}_{jj}^{1/2} > z_{\alpha/(2p)}$, where $z_{\alpha/(2p)}$ is the upper $\{\alpha/(2p)\}$-th quantile of the standard normal distribution. The cutoff is determined by Theorem 3.1 and Bonferroni correction for multiple testing. When implementing cross-validation, we can either take stratum as the sampling unit and randomly split strata, or randomly split observations within each stratum, to form training and testing subsets. We find the former improves stability of tuning parameter selection when there are a number of small-sized strata.

Algorithm 1

Selection of the tuning parameter 7 using cross-validation

Step 1 Pre-determine a grid of points for γ in [0,1], denoted as $\gamma^{(g)},\ g = 1, \cdots, G$, and set each $cv_g = 0$.

Step 2 Randomly assign the K strata into M folds, leaving one fold for testing and the others for training. Set $q = 1$.

Step 2.1 While $q \leq M$, use the qth training set to compute the de-biased lasso estimator with $\gamma^{(g)},\ g = 1, \cdots, G$, denoted as $\hat{b}^{(gq)}$, and define the active set $\widehat{A}^{(gq)}$.

Step 2.2 Define the thresholded de-biased lasso estimator $\hat{b}_{thres}^{(gq)} = \hat{b}^{(gq)} \cdot 1(j \in \widehat{A}^{(gq)})$, i.e. setting components of $\hat{b}(gq)$ outside the active set $\widehat{A}(gq)$ to 0.

Step 2.3 Compute the negative log partial likelihood on the qth testing set $\ell^{(q)}\ (\hat{b}_{thres}^{(gq)})$.

Step 2.4 Set $cv_g \leftarrow cv_g + N^{(q)}\ \ell^{(q)}\ (\hat{b}_{thres}^{(gq)})$, for $g = 1, \cdots, G$, where $N^{(q)}$ is the total number of observations in the qth testing set.

Step 2.5 Set $q \leftarrow q + 1$ and go to Step 2.1.

Step 3 Let $\hat{g} = \arg\min_g cv_g$. The final output tuning parameter value is $\gamma^{(\hat{g})}$.

3. Valid statistical inference based on the de-biased lasso estimator.

This section presents asymptotic results, which lay the groundwork for using the de-biased lasso estimator described in Section 2 to infer on the risk factors of graft failure in the SRTR analysis. The pertaining large sample framework posits that the number of strata K is fixed, the smallest stratum size $n_{min} = \min_{1 \leq k \leq K} n_k \to \infty$, and $n_k/N \to r_k > 0$ as $n_{min} \to \infty,\ k = 1, \cdots, K$.

This framework conforms to the real world setting of our concern, where the number of transplant centers nationwide is finite, and the number of patients or transplant events in each center increases over the years. We provide regularity conditions and their discussion in Appendix B, and present all the proofs in the Supplementary Material.

Let $\mu_{rk}(t;\beta) = E\left[1(Y_{k1} \geq t)X_{k1}^{\otimes r}\exp\left\{X_{k1}^T\beta\right\}\right]$ be the limit of $\widehat{\mu}_{rk}(t;\beta)$, $r = 0,\ 1,\ 2,\ k = 1,\cdots,K$. Then the limit of the weighted covariate process for $\widehat{\eta}_k(t;\beta) = \widehat{\mu}_{1k}(t;\beta)/\widehat{\mu}_{0k}(t;\beta)$ becomes $\eta_{k0}(t;\beta) = \mu_{1k}(t;\beta)/\mu_{0k}(t;\beta)$. Let

$$\Sigma_{\beta^0,k} = E[\left\{X_{ki} - \eta_{k0}\left(Y_{ki};\beta^0\right)\right\}^{\otimes 2}\delta_{ki}]$$

be the information matrix for the k-th stratum, $k = 1,\cdots,K$. The overall information matrix across all strata then becomes the weighted average of the stratum-specific information matrices,

$$\Sigma_{\beta^0} = \sum_{k=1}^{K} r_k\Sigma_{\beta^0,k}\,. \tag{3.1}$$

The inverse information matrix is $\Theta_{\beta^0} = \Sigma_{\beta^0}^{-1}$, which is to be approximated by $\widehat{\Theta}$ obtained in Section 2.1.

The following theorem establishes the asymptotic normality of any linear combination of the estimated regression parameters, $c^T\widehat{b}$ for some loading vector $c \in \mathbb{R}^p$, obtained by the proposed DBL-QP method. For an $m \times r$ matrix $A = (a_{ij})$, define the ℓ_1-induced matrix norm $\|A\|_{1,1} = \max_{1\leq j\leq r}\sum_{i=1}^{m}|a_{ij}|$. For two positive sequences $\{a_n\}$ and $\{b_n\}$, we write $a_n \asymp b_n$ if there exist two constants C and C' such that $0 < C \leq a_n/b_n \leq C' < \infty$. Let s_0 be the number of nonzero elements of β^0.

Theorem 3.1.

Assume that the tuning parameters λ and γ satisfy $\lambda \asymp \sqrt{\log(p)/n_{min}}$ *and* $\gamma \asymp \|\Theta_{\beta^0}\|_{1,1}\{\max_{1\leq k\leq K}|n_k/N - r_k| + s_0\lambda\}$, and that $\|\Theta_{\beta^0}\|_{1,1}^2\{\max_{1\leq k\leq K}|n_k/N - r_k| + s_0\lambda$ *as* $\}p\sqrt{\log(p)} \to 0$ $n_{\min} \to \infty$. *Under* Assumptions B.1–B.5 *given in Appendix B*, for any $c \in \mathbb{R}^p$ such that $\|c\|_2 = 1,\ \|c\|_1 \leq a_*$ with $a_* < \infty$ *being an absolute positive constant, and* $\left\{c^T\Theta_{\beta^0}c\right\}^{-1} = \mathcal{O}(1)$, *we have*

$$\frac{\sqrt{N}c^T(\widehat{b} - \beta^0)}{(c^T\widehat{\Theta}c)^{\frac{1}{2}}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,\ 1)\,.$$

Note that, instead of listing it as a regularity condition in Appendix B, we assume $\{c^T\Theta_{\beta^0}c\}^{-1} = \mathcal{O}(1)$ in the above theorem because the vector c is also defined here. A similar condition is assumed in van de Geer et al. (2014) [Theorem 3.3 (vi)] which is weaker

than uniformly bounding the maximum eigenvalue of Σ_{β^0}. The hypothesis testing with $H_0: c^T\beta^0 - a_0 = 0$ versus $H_1: c^T\beta^0 - a_0 \neq 0$ for some constants $c \in \mathbb{R}^p$ and a_0 entails various applications. For example, by setting $a_0 = 0$ and c to be a basis vector with only one element being 1 and all the others 0 , we can draw inference on any covariate in the presence of all the other covariates. In particular, we will draw inference on the pairwise differences in graft failure risk among donor age groups, e.g. between (10, 20] and (20, 30] (the reference level) years old, and among patients with different primary kidney diagnoses (diabetes is the reference level); see Section 5. Given an appropriately chosen c and with $T = \sqrt{N}(c^T\hat{b} - a_0)/(c^T\widehat{\Theta}c)^{1/2}$, we construct a two-sided test function

$$\phi(T) = \begin{cases} 1 & \text{if } |T| > z_{\frac{\alpha}{2}} \\ 0 & \text{if } |T| \leq z_{\frac{\alpha}{2}} \end{cases},$$

where $z_{\alpha/2}$ is the upper $(\alpha/2)$-th quantile of the standard normal distribution. Corollary 3.2 provides the asymptotic type I error and power of the test $\phi(T)$, and Corollary 3.3 formalizes the construction of level α confidence intervals for $c^T\beta^0$ which ensures the nominal coverage probability asymptotically.

Corollary 3.2.

Under the conditions specified in Theorem 3.1, $P(\phi(T) = 1 \mid H_0) \to \alpha$ as $n_{\min} \to \infty$. Moreover, under $H_1: a_0 - c^T\beta^0 \neq 0$, $P(\phi(T) = 1 \mid H_1) \to 1$.

Corollary 3.3.

Suppose that the conditions in Theorem 3.1 hold. Construct the random confidence interval $\mathcal{R}(\alpha) = \left[c^T\hat{b} - z_{\alpha/2}(c^T\widehat{\Theta}c/N)^{1/2}, c^T\hat{b} + z_{\alpha/2}(c^T\widehat{\Theta}c/N)^{1/2}\right]$. Then $P\left(c^T\beta^0 \in \mathcal{R}(\alpha)\right) \to 1 - \alpha$ as $n_{min} \to \infty$, where the probability is taken under the true β^0.

Our asymptotic results facilitate simultaneous inference on multiple contrasts in the context of post-transplant renal graft failure. For example, an important question to address is whether donor age is associated with graft failure. With categorized donor age in our data analysis, simultaneous comparisons among the seven categories, e.g. ≤ 10,(10,20],(20,30], (30,40],(40,50],(50,60] and 60+, naturally form multiple null contrasts. These contrasts can be formulated by $J\beta^0$, where J is an $m \times p$ matrix, and m represents the number of linear combinations or contrasts. The following theorem and corollary summarize the results for inference on multiple contrasts, $J\beta^0$. See an application of the asymptotic results to the SRTR data with $(m, p) = (6,94)$ in Section 5.

Theorem 3.4.

Suppose that J is an $m \times p$ matrix with $rank(J) = m$, $\| J \|_{\infty,\infty} = \mathcal{O}(1)$ and $J\Theta_{\beta^0}J^T \to F$, where F is a nonrandom $m \times m$ positive definite matrix. Assume that the tuning parameters λ and γ satisfy $\lambda \asymp \sqrt{\log(p)/n_{min}}$ and $\gamma \asymp \| \Theta_{\beta^0} \|_{1,1}\{\max_{1 \leq k \leq K} | n_k/N - r_k | + s_0\lambda\}$, *and that*

$\| \Theta_{\beta^0} \|_{1,1}^2 \{\max_{1 \le k \le K} |n_k/N - r_k| + s_0\lambda\} p\sqrt{\log(p)} \to 0$ as $n_{min} \to \infty$. *Under* Assumptions B.1–B.3, B.5 and B.6 *given in* Appendix B*, we have*

$$\sqrt{N} J(\hat{b} - \beta^0) \xrightarrow{\mathscr{D}} \mathcal{N}_m(0, F).$$

Here, $\| A \|_{\infty,\infty} = \max_{1 \le i \le m} \sum_{j=1}^{r} |a_{ij}|$ is the ℓ_∞-induced matrix norm for an $m \times r$ matrix $A = (a_{ij})$. The theorem implies the following corollary, which constructs test statistics and multi-dimensional confidence regions with proper asymptotic type I error rates and nominal coverage probabilities.

Corollary 3.5.

Suppose the conditions in Theorem 3.4 hold. For an $m \times p$ matrix J as specified in Theorem 3.4, and under $H_0: J\beta^0 = a^0 \in \mathbb{R}^m$,

$$T' = N(J\hat{b} - a^0)^T \widehat{F}^{-1} (J\hat{b} - a^0) \xrightarrow{\mathscr{D}} \chi_m^2,$$

where $\widehat{F} = J\widehat{\Theta}J^T$. Moreover, for an $\alpha \in (0,\ 1),$ define the random set $\mathscr{R}'(\alpha) = \{a \in \mathbb{R}^m : N(J\hat{b} - a)^T \widehat{F}^{-1} (J\hat{b} - a) < \chi_{m,\alpha}^2\},$ where $\chi_{m,\alpha}^2$ is the upper α-th quantile of χ_m^2. Then $P\left(J\beta^0 \in \mathscr{R}'\left(\alpha\right)\right) \to 1 - \alpha$ as $n_{min} \to \infty,$ where the probability is taken under the true β^0.

4. Simulation study.

We conduct simulations to examine the finite sample performance of the proposed DBL-QP approach in correcting estimation biases and maintaining nominal coverage probabilities of confidence intervals. For comparisons, we also perform MSPLE, the oracle estimation, and the three inference methods ["Nodewise" for Kong et al. (2021), "CLIME" for Yu, Bradic and Samworth (2021), and "Decor" for Fang, Ning and Liu (2017)] that are adapted to stratified Cox models. The following scenarios pertain to four combinations of (K, n_k, p), where $K,\ n_k$ and p are the number of strata, stratum-specific sample size and the number of covariates, respectively. Specifically, Scenarios 1–3 refer to $(K, n_k, p) = (10,\ 100,\ 10), (10,\ 100,\ 100)$, and $(5,\ 200,\ 100)$, respectively. In Scenario 4, $K = 40,\ p = 100,\ n_k$'s are simulated from a Poisson distribution with mean 40 and then fixed in all of the replications. This scenario mimics the situation of the recipient group aged over 60, the smallest group in the SRTR data.

Covariates X_{ki} are simulated from $N_p(0, \Sigma_x)$ and truncated at ±3, where Σ_x has an AR(1) structure with the (i, j)-th entry being $0.5^{|i-j|}$. The true regression parameters β^0 are sparse. Its first element β_1^0 varies from 0 to 2 by an increment of 0.2, four additional elements are assigned values of 1, 1, 0.3 and 0.3 with their positions randomly generated and then fixed for all of the simulations, and all other elements are zero. The underlying survival times T_{ki} are simulated from an exponential distribution with hazard $\lambda(t \mid X_{ki}) = \lambda_{0k} \exp\left\{X_{ki}^T \beta^0\right\}$, where

λ_{0k} are generated from Uniform(0.5,1) and then fixed throughout. As in Fang, Ning and Liu (2017) and Fan and Li (2002), the censoring times C_{ki}'s are simulated independently from an exponential distribution with hazard $\lambda_c(t \mid X_{ki}) = 0.2\lambda_{0k}\exp\left\{X_{ki}^T\beta^0\right\}$, resulting in an overall censoring rate around 20%.

For the lasso estimator, we use 5-fold within-stratum cross-validation to select λ. In Scenarios 1–3 with small numbers of strata, each stratum serves as a cross-validation fold for the selection of γ; in Scenario 4 with 40 strata, we perform 10-fold cross-validation as described in Algorithm 1 and randomly assign 4 strata to each fold. For each parameter configuration, we simulate 100 datasets, based on which we compare estimation biases of β_1^0, 95% confidence interval coverage probabilities, model-based standard errors, and empirical standard errors across the six methods.

Figure 2 shows that, in Scenario 1 that features a small number of covariates ($p = 10$), all six methods perform well and similarly; in Scenarios 2–4 with a relatively large number of covariates ($p = 100$), which is close to the number of covariates in the real data we will analyze, our proposed DBL-QP estimator well corrects the biases of the lasso estimates and maintains good confidence interval coverage (excluding the practically impossible "Oracle" estimator), but MSPLE, Nodewise, Decor and CLIME all present larger biases compared to DBL-QP as β_1^0 increases from 0 to 2. CLIME, Nodewise and MSPLE have worse confidence interval coverage in general. As de-biased lasso methods, CLIME and Nodewise produce much smaller model-based standard error estimates, which also contribute to their poor covarage probabilities. This is likely due to that both methods (CLIME and Nodewise) use penalized estimators for inverse information matrix estimation, and such penalization induces biases towards zero.

To recapitulate, the proposed DBL-QP provides less biased estimates and better confidence interval coverage than the conventional MSPLE and three other competitors (Nodewise, Decor and CLIME adapted to the stratified setup) when the sample size is moderate relative to the number of covariates, although all methods give almost identical results when p is rather small. Hence, when $p < N$, our proposed DBL-QP approach is at least as good as all the other methods, and should be recommended for use.

5. Analysis of the SRTR kidney transplant data.

The SRTR data set features 94 covariates from both donors and recipients, and the number of covariates is seen as relatively large for some recipient groups. With its reliable performance as demonstrated in simulations, we apply our DBL-QP approach to analyze the SRTR data, while using MSPLE as a benchmark. The outcome is graft failure free survival, the time from transplant to graft failure or death, whichever comes first. Our primary goal is to investigate the joint associations of these covariates with graft failure for three recipient groups defined in Table 1 separately. By simultaneously considering all available donor and recipient covariates, we aim to account for confounding and provide asymptotically valid inference for the covariate effects, which differs from post hoc inference that only focuses on a smaller set of covariates selected by stepwise selection. The effect of donor age, in

the presence of other risk factors, is worth investigating, as the debatable “one-size-fit-all” practice of donor-recipient age matching unfortunately is not suited for the benefit of transplantation (Keith et al., 2004; Veroux et al., 2012; Dayoub et al., 2018).

5.1. Data details.

Included in our analysis are 9,195 recipients who received kidney-only transplants from deceased donors, had no prior solid organ transplants, and were at least 18 years old at the time of transplantation during 2000 and 2001. We focus on those with these same cohort years in order to eliminate the cohort effect. Moreover, this group of patients had longer follow-up than those from the later cohort years. See Appendix A for a full list of included variables in the analysis. In the three receipts’ age groups, respectively, the sample sizes are 3388, 4359 and 1448, the censoring rates are 53.1%, 46.5% and 30.0%, the median numbers of patients within each transplant center are 32, 31 and 27, and the restricted mean survival times by 13 years are 9.1, 8.6 and 7.1 years. To select the tuning parameters, we implement 5-fold cross-validation by randomly selecting one fifth of transplant centers without replacement as testing data and the rest as training data.

5.2. Results.

We begin with examining the overall effect of donors’ age on graft failure and testing the null hypothesis that, within each recipient group and after adjusting for the other risk factors, all the donor age groups, i.e. ≤ 10,(10,20],(20,30],(30,40],(40,50],(50,60] and 60+, have the same risk of graft failure. Based on Theorem 3.4 and Corollary 3.5, with $(m,\ p) = (6,\ 94)$, we perform tests for the null contrasts, and the obtained statistics significantly reject the null hypotheses for all three recipient groups (within recipients aged 18–45: $\chi^2 = 40.4,\ \mathrm{df} = 6$, p-value=$3.9 \times 10^{-7}$; recipients aged 45–60: $\chi^2 = 34.5,\ \mathrm{df} = 6$, p-value=$5.3 \times 10^{-6}$; recipients aged over 60: $\chi^2 = 14.2,\ \mathrm{df} = 6$, p-value=$2.8 \times 10^{-2}$). Indeed, Figure 3, which depicts the risk-adjusted effect of donors’ age across the three recipient age groups, shows a general trend of increasing hazards for those receiving kidneys from older donors, likely due to renal aging. The estimates and confidence intervals obtained by our proposed DBL-QP differ from those obtained by MSPLE, and the differences are the most obvious in the 60+ year recipient group, which has the smallest sample size. As presented in our simulations, MSPLE may produce biased estimates with improper confidence intervals, especially when the sample size is relatively small.

On the other hand, the proposed DBL-QP method may shed new light into the aging effect, which seems to be non-linear with respect to donors’ age. First, using the results of Theorem 3.1 and Corollary 3.3, our tests detect no significant differences in hazards between those receiving kidneys from donors aged under 10 or (10,20] and (20,30] (reference level) years old, within all the three recipient age groups. Second, significantly increased hazards are observed as early as when donors’ age reached 30–40, as compared to the reference level of (20,30], in the 18–45 years old recipient group, with an estimated hazard ratio (HR) of 1.16 (95% CI: 1.01–1.34, p-value=4.1×10^{-2}). In contrast, there are no significant differences between receiving organs from (30,40] years old donors and the reference level of (20,30], among the 45–60 years old recipients (HR= 0.96, 95% CI: 0.85–1.09, p-value=5.1×10^{-1}) and the 60+ years old recipients (HR=1.07, 95% CI: 0.88–1.30, p-value=5.0×10^{-1}).

Third, kidneys from 60+ years old donors confer the highest hazards, with the estimated risk-adjusted HRs (compared to the reference level (20,30]) being 1.83 (95% CI: 1.48–2.28, p-value=4.3×10^{-8}), 1.40 (95% CI: 1.21–1.61, p-value=4.1×10^{-6}) and 1.37 (95% CI: 1.14–1.63, p-value=5.2×10^{-4}) among the three recipient age groups respectively. This means that, compared to the older recipients, recipients of 18–45 years old tend to experience a greater hazard of graft failure when receiving kidneys from donors over 60 years old. Caution needs to be exercised when allocating kidneys from older donors to young patients (Lim et al., 2010; Kaboré et al., 2017; Dayoub et al., 2018).

Our method also delineates the associations of clinical indicators with graft failure, provides more reliable inference, and compares the relative strengths across recipient age groups. By naively applying lasso, 64, 44 and 27 covariates are selected with non-zero coefficients in the 18–45, 45–60, and 60+ years old recipient groups, respectively. In contrast, the proposed DBL-QP identifies 22, 22 and 14 significant covariates in these three recipient groups, respectively, from rigorous hypothesis tests with size 0.05 based on the asymptotic distribution. Figure 4 shows the estimated coefficients and their 95% confidence intervals for covariates that are significant at level 0.05 in at least one recipient group. We highlight several noteworthy results.

First, recipients' primary kidney diagnosis plays a critical role in kidney graft failure (Wolfe, 1991). Compared to recipients with primary diagnosis of diabetes (the reference level), those with polycystic kidneys (variable 2 in Figure 4) have a reduced risk of graft failure, with highly significant lower HRs of 0.54 (95% CI: 0.42–0.70, p-value=3.6×10^{-6}), 0.65 (95% CI: 0.57–0.75, p-value=4.4×10^{-9}) and 0.74 (95% CI: 0.60–0.92, p-value=5.3×10^{-3}) for the three age groups respectively. Compared to diabetes, primary diagnosis of glomerular disease (variable 26 in Figure 4) is significantly associated with a reduced risk of graft failure only in the 60+ years old recipient group (HR=0.79, 95% CI: 0.66–0.96, p-value=1.4×10^{-2}), and primary diagnosis of hypertensive nephrosclerosis (variable 29 in Figure 4) is significantly associated with a higher hazard of graft failure only in the 45–60 years old recipient group (HR=1.12, 95% CI: 1.01–1.23, p-value=2.5×10^{-2}).

Second, since diabetes is the most prevalent among end-stage renal patients (Kovesdy, Park and Kalantar-Zadeh, 2010), we code recipients' diabetic status at transplant as non-diabetic (reference level), diabetic for 0–20 years (variable 13 in Figure 4), and 20+ years (variable 3 in Figure 4). Our stratified analysis reveals that diabetics is a stronger risk factor for young recipients aged between 18 and 45 years old than for older recipients, regardless of duration of diabetes.

Third, instead of using the total number of mismatches as done in the literature, we consider the number of mismatches separately for each HLA locus for more precisely pinpointing the effects of mismatching loci. Our results reveal that the HLA-DR mismatches (variable 9 in Figure 4) are more strongly associated with graft failure than the HLA-A (variable 18 in Figure 4) and HLA-B mismatches (non-significant in any recipient group), which are consistent with a meta-analysis based on 500,000 recipients (Shi et al., 2018).

Finally, to study the granular impact of recipient age on graft failure (Karim et al., 2014), we treat recipient age (divided by 10) as a continuous variable (variable 4 in Figure 4) in the model within each recipient age group. Interestingly, we find that increasing age is associated with a higher hazard in the two older recipient groups (HR=1.31, 95% CI: 1.19–1.44, p-value=1.3×10^{-8}, for recipients aged 45–60; HR=1.22, 95% CI: 1.07–1.40, p-value=3.6×10^{-3}, for recipients aged 60+), but with a lower hazard of graft failure in the 18–45 recipient age group (HR=0.89, 95% CI: 0.83–0.95, p-value=5.2×10^{-4}). This is likely because that younger patients generally had poorer adherence to treatment, resulting in higher risks of graft loss (Kaboré et al., 2017). The results also reinforce the necessity of separating analyses for different recipient age groups.

As a side note, we compare DBL-QP and MSPLE in the estimated coefficients and standard errors. Figure 5 shows that in the 45–60 age group with the largest number of subjects, the point estimates obtained by the two methods almost coincide with each other, whereas in the 60+ age group with the smallest sample size, MSPLE tends to have larger absolute estimates than the de-biased lasso. Moreover, the standard errors estimated by MSPLE are likely to be larger than those by our method across all the age groups. These observations agree with the results of our simulations (Scenarios 2–4), which show that MSPLE yields large biases in estimated coefficients and standard errors, especially when the sample size is relatively small, whereas our proposed DBL-QP method draws more valid inferences by maintaining proper type I errors and coverage probabilities.

6. Concluding remarks.

The work is motivated by an urgent call of better understanding the complex mechanisms behind post-kidney transplant graft failure. Our modeling framework is Cox models stratified by transplant centers, due to their strong confounding effects on graft failure. To adjust for confounders to the extent possible, we have included an extended list of 94 covariates from recipients and donors, which has not been done in the literature. A particular scientific question to address is the debatable donor-recipient age matching criterion in kidney transplantation. Fitting separate models by recipient age enables direct assessments of the donor age effects in different recipient age groups, which differs from using donor-recipient age difference as in Ferrari et al. (2011). Specifically, we have followed a common practice of fitting separate models in age groups of 18–45, 45–60 and 60+ years. The commonly used MSPLE yielded biased estimates and unreliable inference in some smaller age groups, though the samples outnumbered the covariates. In particular, the 60+ years recipient group had only 1448 recipients in 43 different transplant centers, and MSPLE yielded more dramatic estimates for those donor age effects of over 30 years old (Figure 3). Our simulation results also confirmed such a problematic phenomenon. Therefore, a statistical method that can guarantee reliable estimates and valid inference is much needed for delineating the associations of interest with graft failure when the number of covariates is relatively large in stratified Cox models.

Inspired by the de-biased lasso method for linear regression (Javanmard and Montanari, 2014), we have developed a de-biased lasso approach via quadratic programming for stratified Cox models. Despite progress made in high-dimensional inference for Cox

models, virtually no work has considered stratified settings, theoretically or empirically. We have shown that in the "large N, diverging p" scenario, our approach possesses desirable asymptotic properties and finite-sample performance, and is more suitable for the analysis of the SRTR data than the competing methods illustrated in our simulation studies. Computationally, based on a previous work on Cox models without stratification (Xia, Nan and Li, 2022), for the estimation of Θ_{β^0}, the computational speed using <mono_space> solve.QP </mono_space> in R was much faster than that using the R packages <mono_space> clime </mono_space> or </mono_space> flare </mono_space> adopted by Yu, Bradic and Samworth (2021).

Applications of our method to the SRTR data generated new biological findings. After categorizing donors' age and controlling for other risk factors listed in Appendix A, we find that organs from older donors are associated with an increased hazard of graft failure and that the dependence on donors' age is non-linear: within the youngest recipient group (18–45 years), significant differences from the reference donor age category (20–30 years) were detected as early as when donors reached 30–40 years old, whereas significant differences were detected only when donors reached 50–60 or 60+ years within the two older recipient groups, respectively; in other words, receiving kidneys from older or younger donors, such as 60+ versus 20–30 years, presented larger differences than in the other two recipient groups. These results, which were not reported in the literature, may provide new empirical evidence to aid stake-holders, such as patients, families, physicians and policy makers, in decisions on donor-recipient age matching.

A few technical points are noteworthy. First, our work deals with the "large N, diverging p" scenario, as embedded in the motivating data, and approximates Θ_{β^0} via quadratic programming without positing any sparsity conditions on Θ_{β^0}. This distinguishes from the related literature (Fang, Ning and Liu, 2017; Yu, Bradic and Samworth, 2021; Kong et al., 2021) in the "large p, small N" scenario that relies upon sparsity conditions on the numbers of non-zero elements in the rows of Θ_{β^0}, which are hardly discussed in depth and may not hold nor have explicit interpretations for Cox models. For example, when the rows of Θ_{β^0} are not sparse, our dimension requirement for p is less stringent than in Yu, Bradic and Samworth (2021), by a factor of $\sqrt{\log(Np)}$. Moreover, when $p > N$, the several de-biased methods aforementioned may not yield reliable inference results, as empirically Θ_{β^0} cannot be estimated well, and biases in the lasso estimator are often not sufficiently corrected for in this scenario for Cox models. New approaches, such as sample-splitting approaches (Fei and Li, 2021) that bypass the estimation of Θ_{β^0}, can be consulted.

Second, tuning parameter selection is critical in high-dimensional inference. Our proposed method deploys a single tuning parameter γ for de-biasing the estimates of all β_j's. This is a computationally feasible and commonly adopted strategy, presenting a satisfactory performance in our numerical studies, and can be extended to adapt to the variability of individual coefficient estimation. For example, one may consider the following estimation procedure for the jth row of $\widehat{\Theta}$ along the line of adaptive CLIME (Cai, Liu and Zhou, 2016):

$$\min_{m} \{ m^T \widehat{\Sigma} m : (|\widehat{\Sigma} m - e_j|_k) \leq \gamma_{jk}, k = 1, \ldots, p \} .$$

Here, γ_{jk}'s are supposed to be adaptively estimated through a carefully designed procedure. However, the design of such an appropriate procedure requires complicated theoretical analysis in Cox models, unstratified or stratified, to determine the desirable rates of γ_{jk}'s, among other tasks. Given that such complexity is beyond the scope of this paper, we will not pursue this route here in details but will leave it for future research.

Third, though primarily focusing on the associations between the risk factors and survival (through Theorem 3.1), the proposed method can be used for patient risk scoring and conditional survival probability estimation. For example, the de-biased estimates may be plugged into the Breslow's estimator (Kalbfleisch and Prentice, 2002) for stratum-specific baseline hazards. The conditional survival probability estimation may not go beyond the time point τ due to censoring.

Lastly, we use Cox models stratified by transplant centers to account for but avoid explicitly modeling the center effects. Alternatively, random effects models can be used for clustered survival data analysis; for example, Vaida and Xu (2000) generalized the usual frailty model to allow multivariate random effects. However, in a random effects model, the distribution of random effects needs to be specified, and the coefficients only have conditional interpretations, given a cluster. We may pursue this elsewhere.

We have implemented the proposed DBL-QP method with cross-validation in R and Rcpp, which is available both in the Supplementary Material and online at https://github.com/luxia-bios/StratifiedCoxInference/ with simulated examples.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements.

This work was supported in part by NIH Grants (R01AG056764, R01AG075107, R01CA249096 and U01CA209414) and NSF Grant (DMS-1915711).

APPENDIX A: SRTR DATA

The SRTR dataset analyzed in this article can be accessed by applying through the OPTN website https://optn.transplant.hrsa.gov. The interpretation and reporting of the SRTR data results are solely the responsibility of the authors and should not be viewed as official opinions of the SRTR or the United States Government.

The 94 covariates, including dummy variables, are derived from the following factors. Donor factors include: ABO blood type, age, cytomegalovirus antibody, hepatitis C virus antibody, cause of death, cardiac arrest since event leading to declaration of death, serum creatinine, medication given to donor (DDAVP, dopamine and dobutamine), gender, height,

history of cancer, cigarette smoking, history of drug abuse, hypertension, diabetes, inotropic support, inotropic agents at time of incision, non-heart beating donor, local or shared organ transplant, race, and weight. Recipient factors include: ABO blood type, history of diabetes and duration, angina/coronary artery disease, symptomatic peripheral vascular disease, drug treated systemic hypertension, drug treated COPD, gender (and previous pregnancies for females), sensitization (whether peak and/or current panel-reactive antibodies exceed 20%), previous malignancy, peptic ulcer disease, symptomatic cerebrovascular disease, race, total serum albumin, age at transplant, number of HLA mismatches (A, B and DR), cytomegalovirus status, total cold ischemic time, primary kidney diagnoses, pre-transplant dialysis and duration, the Epstein–Barr virus serology status, employment status, hepatitis B virus status, hepatitis C virus status, height, pre-implantation kidney biopsy, pre-transplant blood transfusions, transplant procedure type, warm ischemic time and weight.

APPENDIX B: REGULARITY CONDITIONS

Assumptions B.1–B.5 below ensure that Theorem 3.1 hold.

Assumption B.1.

Covariates are almost surely uniformly bounded, i.e. $\| X_{ki} \|_\infty \leq M$ for some positive constant $M < \infty$ for all k and i.

Assumption B.2.

$\left| X_{ki}^T \beta^0 \right| \leq M_1$ uniformly for all k and i with some positive constant $M_1 < \infty$ almost surely.

Assumption B.3.

The follow-up time stops at a finite time point $\tau > 0$, with probability $\pi_0 = \min_k P(Y_{ki} \geq \tau) > 0$.

Assumption B.4.

For any $t \in [0, \tau]$,

$$\frac{c^T \Theta_{\beta^*}}{c^T \Theta_{\beta^*} c} \left[\sum_{k=1}^{K} r_k \int_0^t \left\{ \mu_{2k}\left(u;\ \beta^0\right) - \frac{\mu_{1k}\left(u;\ \beta^0\right) \mu_{1k}\left(u;\ \beta^0\right)^T}{\mu_{0k}\left(u;\ \beta^0\right)} \right\} \lambda_{0k}(u) du \right] \Theta_{\beta^*} c \to v(t;\ c)$$

as $n \to \infty$ for some function $v(t; c) > 0$ of t that also depends on the choice of c.

Assumption B.5.

There exists a constant $\epsilon_0 > 0$ such that $\lambda_{\min}(\Sigma_{\beta^0}) \geq \epsilon_0$, where $\lambda_{\min}(\cdot)$ is the smallest eigenvalue of a matrix.

For inference on multiple linear combinations or contrasts as described in Theorem 3.4, Assumption B.4 needs to be replaced with the following Assumption B.6, which is a multivariate version of Assumption B.4.

Assumption B.6.

For any $\omega \in \mathbb{R}^m$ and any $t \in [0, \tau]$,

$$\frac{\omega^T J \Theta_{\beta^0}}{\omega^T J \Theta_{\beta^0} J^T \omega} \left[\sum_{k=1}^{K} r_k \int_0^t \left\{ \mu_{2k}\left(u;\ \beta^0\right) - \frac{\mu_{1k}\left(u;\ \beta^0\right) \mu_{1k}\left(u;\ \beta^0\right)^T}{\mu_{0k}\left(u;\ \beta^0\right)} \right\} d\Lambda_{0k}(u) \right] \Theta_{\beta^0} J^T \omega$$

converges to $v'(t; \omega, J)$ as $n \to \infty$, for some function $v'(t; \omega, J) > 0$ of t, that also depends on the choice of ω and x.

It is common in the literature of high-dimensional inference to assume bounded covariates as in Assumption B.1. Fang, Ning and Liu (2017) and Kong et al. (2021) also posed Assumption B.2 for Cox models, i.e. uniform boundedness on the multiplicative hazard. Under Assumption B.1, Assumption B.2 can be implied by the bounded overall signal strength $\| \beta^0 \|_1$. Assumption B.3 is a common assumption in survival analysis (Andersen and Gill, 1982). Assumption B.4 and its multivariate version, Assumption B.6, ensure the convergence of the variation process, which is key in applying the martingale central limit theorem. They are less stringent comparing to the boundedness assumption on $\| \Theta_{\beta^0} X_{ki} \|_\infty$ that is equivalent to the assumptions for statistical inference in van de Geer et al. (2014) on high-dimensional generalized linear models and in Fang, Ning and Liu (2017) on high-dimensional Cox models. The boundedness of the smallest eigenvalue of Σ_{β^0} in Assumption B.5 is common in inference for high-dimensional models (van de Geer et al., 2014; Kong et al., 2021). Since we focus on random designs, unlike Huang et al. (2013), Yu, Bradic and Samworth (2021) and Fang, Ning and Liu (2017), we do not directly assume the compatibility condition on $\ddot{\ell}\left(\beta^0\right)$; instead, we impose Assumption B.5 on the population-level matrix Σ_{β^0}, which leads to the compatibility condition for a given data set with probability going to one.

REFERENCES

ALEXANDER JW, BENNETT LE and BREEN TJ (1994). Effect of donor age on outcome of kidney transplantation. A two-year analysis of transplants reported to the United Network for Organ Sharing Registry. Transplantation 57 871–876. [PubMed: 8154034]

ANDERSEN PK and GILL RD (1982). Cox's regression model for counting processes: A large sample study. Ann. Statist. 10 1100–1120.

BAKER RJ, MARK PB, PATEL RK, STEVENS KK and PALMER N (2017). Renal association clinical practice guideline in post-operative care in the kidney transplant recipient. BMC Nephrol. 18 174. [PubMed: 28571571]

BASTANI B (2015). The worsening transplant organ shortage in USA; desperate times demand innovative solutions. J. Nephropathol. 4 105–109. [PubMed: 26457256]

CAI TT, LIU W and ZHOU HH (2016). Estimating sparse precision matrix: Optimal rates of convergence and adaptive estimation. Ann. Statist. 44 455–488.

DAYOUB JC, CORTESE F, ANŽIČ A, GRUM T and DE MAGALHÃES JP (2018). The effects of donor age on organ transplants: A review and implications for aging research. Exp. Gerontol. 110 230–240. [PubMed: 29935294]

DICKINSON DM, ARRINGTON CJ, FANT G, LEVINE GN, SCHAUBEL DE, PRUETT TL, ROBERTS MS and WOLFE RA (2008). SRTR program-specific reports on outcomes: A guide for the new reader. Am. J. Transplant. 8 1012–1026. [PubMed: 18336703]

FAN J and LI R (2002). Variable selection for Cox's proportional hazards model and frailty model. Ann. Statist 30 74–99.

FANG EX, NING Y and LIU H (2017). Testing and confidence intervals for high dimensional proportional hazards models. J. R. Stat. Soc. Ser. B. Stat. Methodol. 79 1415–1437.

FARAVARDEH A, EICKHOFF M, JACKSON S, SPONG R, KUKLA A, ISSA N, MATAS AJ and IBRAHIM HN (2013). Predictors of graft failure and death in elderly kidney transplant recipients. Transplantation 96 1089–1096. [PubMed: 24056622]

FEI Z and LI Y (2021). Estimation and inference for high dimensional generalized linear models: A splitting and smoothing approach. J. Mach. Learn. Res. 22 1–32.

FERRARI P, LIM W, DENT H and MCDONALD SP. (2011). Effect of donor-recipient age difference on graft function and survival in live-donor kidney transplantation. Nephrol. Dial. Transplant. 26 702–708. [PubMed: 20601369]

FU W and KNIGHT K (2000). Asymptotics for lasso-type estimators. Ann. Statist. 28 1356–1378.

HAMIDI O, POOROLAJAL J, FARHADIAN M and TAPAK L (2016). Identifying important risk factors for survival in kidney graft failure patients using random survival forests. Iran. J. Public Health 45 27–33. [PubMed: 27057518]

HE X and SHAO Q-M (2000). On parameters of increasing dimensions. J. Multivariate Anal. 73 120–135.

HE K, ZHU J, KANG J and LI Y (2021). Stratified Cox models with time-varying effects for national kidney transplant patients: A new blockwise steepest ascent method. Biometrics. in press. DOI: 10.1111/biom.13473.

HUANG J, SUN T, YING Z, YU Y and ZHANG C-H (2013). Oracle inequalities for the lasso in the Cox model. Ann. Statist. 41 1142–1165.

JAVANMARD A and MONTANARI A (2014). Confidence intervals and hypothesis testing for high-dimensional regression. J. Mach. Learn. Res. 15 2869–2909.

JU A, CHOW BY, RALPH AF, HOWELL M, JOSEPHSON MA, AHN C, BUTT Z, DOBBELS F, FOWLER K, JOWSEY-GREGOIRE S, JHA V, LOCKE JE, TAN JC, TAYLOR Q, RUTHERFORD C, CRAIG JC and TONG A (2019). Patient-reported outcome measures for life participation in kidney transplantation: A systematic review. Am. J. Transplant. 19 2306–2317. [PubMed: 30664327]

KABORÉ R, COUCHOUD C, MACHER M-A, SALOMON R, RANCHIN B, LAHOCHE A, ROUSSEY-KESLER G, GARAIX F, DECRAMER S, PIETREMENT C et al. (2017). Age-dependent risk of graft failure in young kidney transplant recipients. Transplantation 101 1327–1335. [PubMed: 27482961]

KALBFLEISCH JD and PRENTICE RL (2002). The statistical analysis of failure time data. Hoboken: John Wiley & Sons.

KARIM A, FARRUGIA D, CHESHIRE J, MAHBOOB S, BEGAJ I, RAY D and SHARIF A (2014). Recipient age and risk for mortality after kidney transplantation in England. Transplantation 97 832–838. [PubMed: 24342978]

KASISKE BL and SNYDER J (2002). Matching older kidneys with older patients does not improve allograft survival. J. Am. Soc. Nephrol. 13 1067–1072. [PubMed: 11912268]

KASISKE BL, ISRANI AK, SNYDER JJ, SKEANS MA and PATIENT OUTCOMES IN RENAL TRANSPLANTATION (PORT) INVESTIGATORS (2011). The relationship between kidney function and long-term graft survival after kidney transplant. Am. J. Kidney Dis. 57 466–475. [PubMed: 21257243]

KEITH DS, DEMATTOS A, GOLCONDA M, PRATHER J and NORMAN D (2004). Effect of donor recipient age match on survival after first deceased donor renal transplantation. J. Am. Soc. Nephrol. 15 1086–1091. [PubMed: 15034113]

KONG S and NAN B (2014). Non-asymptotic oracle inequalities for the high-dimensional Cox regression via Lasso. Statist. Sinica 24 25–42.

KONG S, YU Z, ZHANG X and CHENG G (2021). High-dimensional robust inference for Cox regression models using desparsified Lasso. Scand. J. Statist. 48 1068–1095.

KOSTRO JZ, HELLMANN A, KOBIELA J, SKÓRA I, LICHODZIEJEWSKA-NIEMIERKO M, DEBSKA-ŚLIZIEŃ A and ŚLEDZIŃSKI Z (2016). Quality of life after kidney transplantation: A prospective study. Transplant. Proc. 48 50–54. [PubMed: 26915842]

KOVESDY CP, PARK JC and KALANTAR-ZADEH K (2010). Glycemic control and burnt-out diabetes in ESRD. In Semin. Dial. 23 148–156. Wiley Online Library.

LEGENDRE C, CANAUD G and MARTINEZ F (2014). Factors influencing long-term outcome after kidney transplantation. Transpl. Int. 27 19–27. [PubMed: 24138291]

LIM WH, CHANG S, CHADBAN S, CAMPBELL S, DENT H, RUSS GR and MCDONALD SP (2010). Donor-recipient age matching improves years of graft function in deceased-donor kidney transplantation. Nephrol. Dial. Transplant. 25 3082–3089. [PubMed: 20736266]

MORALES JM, MARCÉN R, DEL CASTILLO D, ANDRES A, GONZALEZ-MOLINA M, OPPENHEIMER F, SERÓN D, GIL-VERNET S, LAMPREAVE I, GAINZA FJ, VALDÉS F, CABELLO M, ANAYA F, ESCUIN F, ARIAS M, PALLARDÓ L and BUSTAMANTE J (2012). Risk factors for graft loss and mortality after renal transplantation according to recipient age: a prospective multicentre study. Nephrol. Dial. Transplant. 27 iv39–iv46. [PubMed: 23258810]

RAO PS and OJO A (2009). The alphabet soup of kidney transplantation: SCD, DCD, ECD – fundamentals for the practicing nephrologist. Clin. J. Am. Soc. Nephrol. 4 1827–1831. [PubMed: 19808229]

RAO PS, SCHAUBEL DE, GUIDINGER MK, ANDREONI KA, WOLFE RA, MERION RM, PORT FK and SUNG RS (2009). A comprehensive risk quantification score for deceased donor kidneys: the kidney donor risk index. Transplantation 88 231–236. [PubMed: 19623019]

RODGER RSC (2012). Approach to the management of end-stage renal disease. Clin. Med. 12 472–475.

SARAN R, ROBINSON B, ABBOTT KC, BRAGG-GRESHAM J, CHEN X, GIPSON D, GU H, HIRTH RA, HUTTON D, JIN Y et al. (2020). US Renal Data System 2019 Annual Data Report: Epidemiology of Kidney Disease in the United States. Am. J. Kidney Dis. 75 (suppl 1) Svi–Svii.

SHI X, LV J, HAN W, ZHONG X, XIE X, SU B and DING J (2018). What is the impact of human leukocyte antigen mismatching on graft survival and mortality in renal transplantation? A meta-analysis of 23 cohort studies involving 486,608 recipients. BMC Nephrol. 19 116. [PubMed: 29776389]

SIMON N, FRIEDMAN J, HASTIE T and TIBSHIRANI R (2011). Regularization paths for Cox's proportional hazards model via coordinate descent. J. Stat. Softw. 39 1–13.

SMITH J, BIGGINS S, HASELBY D, KIM W, WEDD J, LAMB K, THOMPSON B, SEGEV D, GUSTAFSON S, KANDASWAMY R, STOCK P, MATAS A, SAMANA C, SLEEMAN E, STEWART D, HARPER A, EDWARDS E, SNYDER J, KASISKE B and ISRANI A (2012). Kidney, pancreas and liver allocation and distribution in the United States. Am. J. Transplant. 12 3191–3212. [PubMed: 23157207]

TIBSHIRANI R (1997). The lasso method for variable selection in the Cox model. Stat. Med. 16 385–395. [PubMed: 9044528]

VAIDA F and XU R (2000). Proportional hazards model with random effects. Stat. Med. 19 3309–3324. [PubMed: 11122497]

VAN DE GEER S, BÜHLMANN P, RITOV Y and DEZEURE R (2014). On asymptotically optimal confidence regions and tests for high-dimensional models. Ann. Statist. 42 1166–1202.

VEROUX M, GROSSO G, CORONA D, MISTRETTA A, GIAQUINTA A, GIUFFRIDA G, SINAGRA N and VEROUX P (2012). Age is an important predictor of kidney transplantation outcome. Nephrol. Dial. Transplant. 27 1663–1671. [PubMed: 21926404]

WANG L (2011). GEE analysis of clustered binary data with diverging number of covariates. Ann. Statist. 39 389–417.

WOLFE RA (1991). Survival analysis methods for the end-stage renal disease (ESRD) program of Medicare. In Rettig RA and Levinsky NG (eds), Kidney Failure and the Federal Government 353–400. National Academies Press.

WOLFE RA, ASHBY VB, MILFORD EL, OJO AO, ETTENGER RE, AGODOA LY, HELD PJ and PORT FK (1999). Comparison of mortality in all patients on dialysis, patients on dialysis awaiting transplantation, and recipients of a first cadaveric transplant. New Engl. J. Med. 341 1725–1730. [PubMed: 10580071]

XIA L, NAN B and LI Y (2022). Statistical inference for Cox proportional hazards models with a diverging number of covariates. Scand. J. Statist. in press. DOI: 10.1111/sjos.12595.

YU Y, BRADIC J and SAMWORTH RJ (2021). Confidence intervals for high-dimensional Cox models. Statist. Sinica 31 243–267.

ZHANG C-H and ZHANG SS (2014). Confidence intervals for low dimensional parameters in high dimensional linear models. J. R. Stat. Soc. Ser. B. Stat. Methodol. 76 217–242.

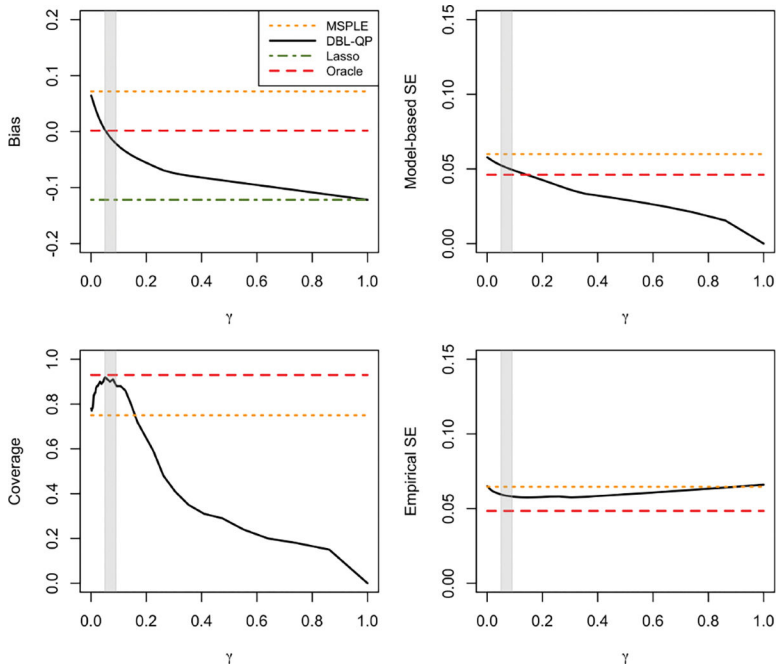

FIG 1.

The impact of choices of γ on the averages of biases, empirical coverage probabilities, model-based and empirical standard errors, based on 100 simulations with $K = 5$ *strata and* $n_k = 200$ in each stratum, $p = 100$ covariates simulated from a multivariate normal distribution with mean zero and an AR(1) covariance matrix ($\rho = 0.5$) *and truncated at ±3. Survival times are simulated with a hazard* $\lambda_{0k} \exp\{X_{ki}^T \beta^0\}$, *where* λ_{0k} *are constants generated from Uniform* (0.1,0.5), *and four nonzero coefficients in* β^0 *take 1, 1, 0.3 and 0.3, respectively. Censoring times are independently simulated from Uniform* (1,30).

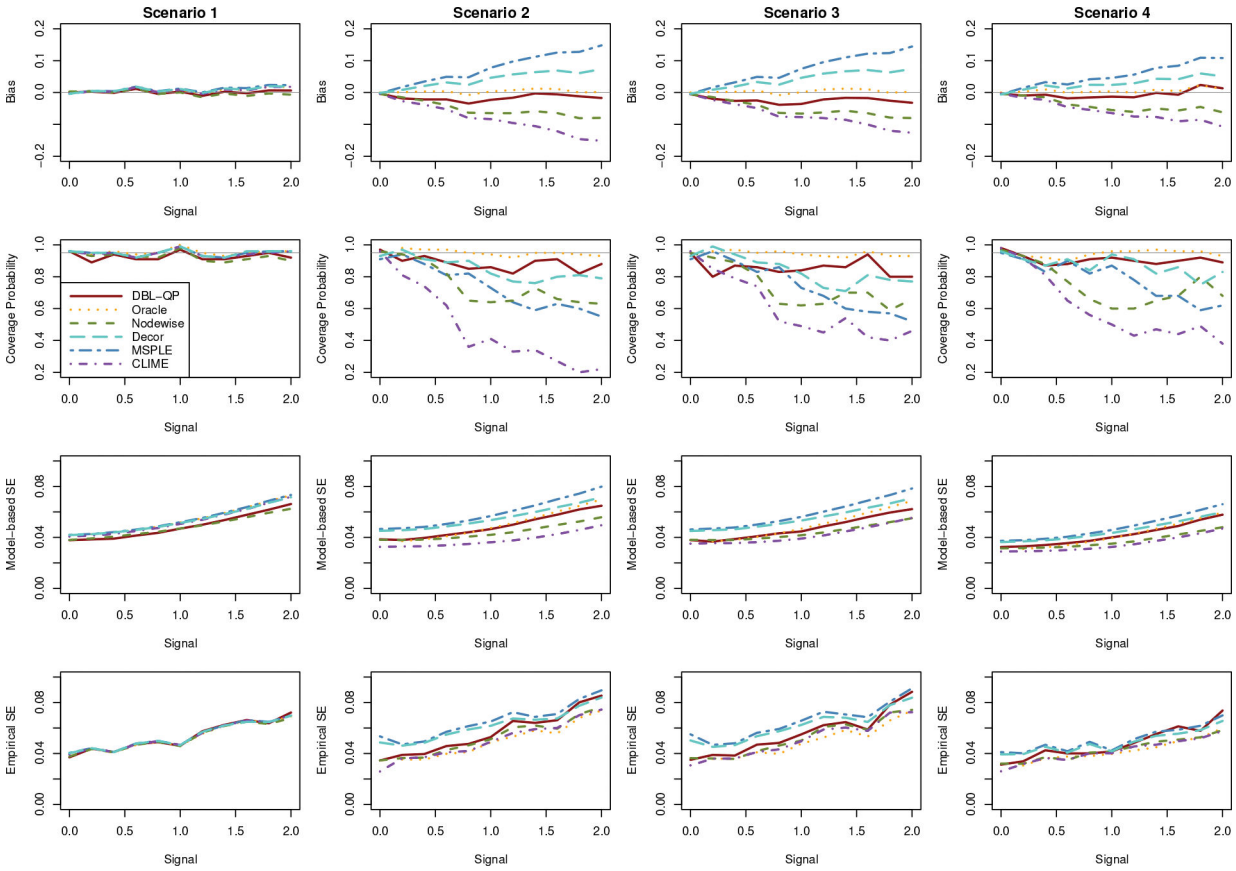

FIG 2.

Estimation bias, 95% coverage, model-based standard error, and empirical standard error for β_1^0 of six different methods. The horizontal lines in the first two rows are references to 0 for bias and 95% for coverage probability, respectively.

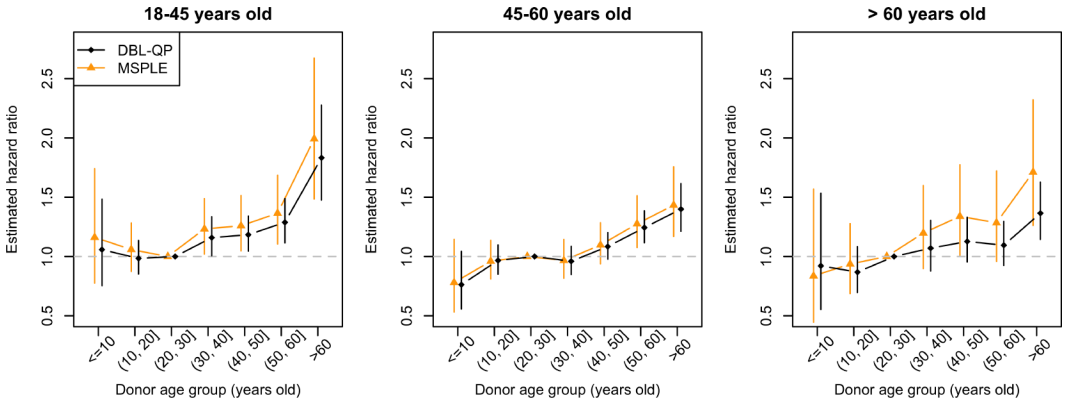

FIG 3.

Estimated hazard ratios and the corresponding 95% confidence intervals of different donor age categories with reference to the (20,30] donor age category, after adjusting for all other variables, in three recipient groups.

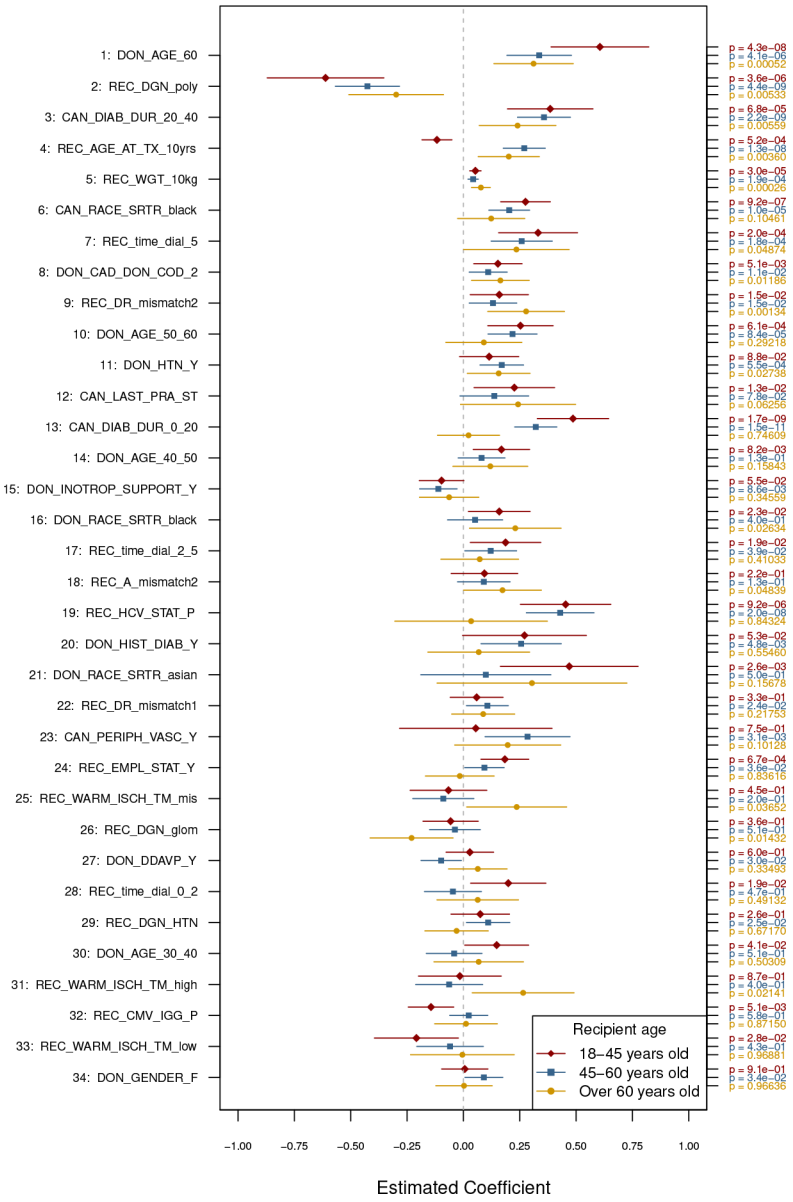

FIG 4.

Estimated regression coefficients in the stratified Cox models using the proposed DBL-QP method, and the corresponding 95% confidence intervals, presented by recipient age group. The covariates included are significant at level 0.05 in at least one recipient group, after adjusting for all other covariates.

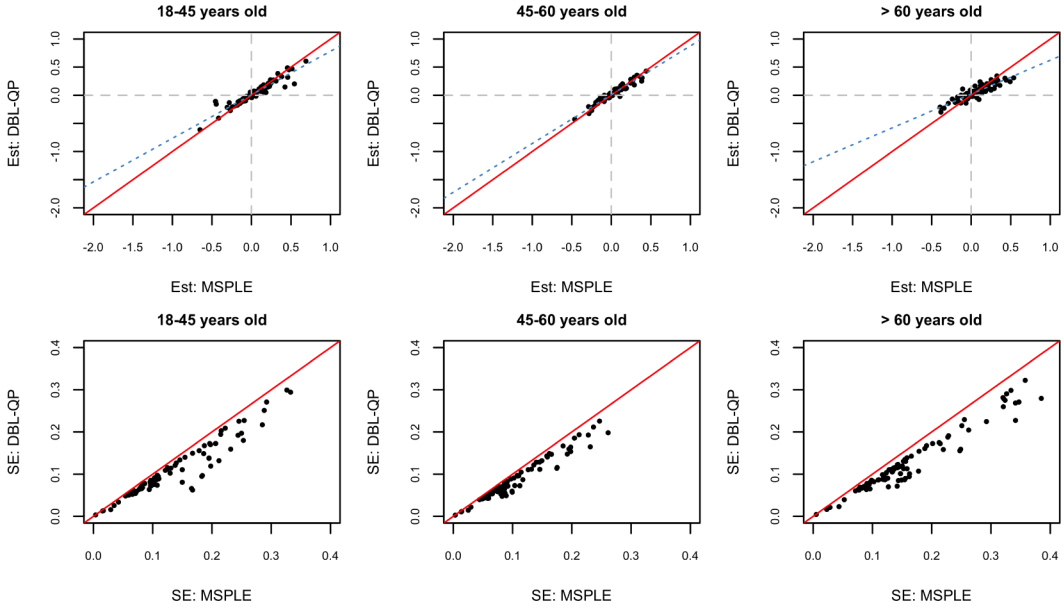

FIG 5.

Comparison between the coefficient estimates (top) and the model-based standard errors (bottom) by the de-biased lasso (DBL-QP) and the maximum stratified partial likelihood estimation (MSPLE) in three recipient age groups. The solid and the dashed lines are 45-degree and zero-value reference lines, respectively; and the dotted lines represent the fitted linear regression of the DBL-QP estimates on the MSPLE estimates.

TABLE 1

Study population characteristics by recipient age group

Recipient age group	**[18,45]**	**(45,60]**	**60+**
Variable	**Mean (SD) / Count (%)**		
# Centers	84(−)	107 (−)	43 (−)
# Patients	3388 (100%)	4359 (100%)	1448 (100%)
# Events	1588 (46.9%)	2334 (53.5%)	1013 (70.0%)
Recipient age	35.7 (7.0)	53.0 (4.2)	66.6 (4.3)
Donor age (years)			
≤ 10	276 (8.1%)	223 (5.1%)	61 (4.2%)
(10,20]	580 (17.1%)	611 (14.0%)	137 (9.5%)
(20,30]	633 (18.7%)	683 (15.7%)	179 (12.4%)
(30,40]	505 (14.9%)	599 (13.7%)	174(12.0%)
(40,50]	753 (22.2%)	947 (21.7%)	256 (17.7%)
(50,60]	498 (14.7%)	893 (20.5%)	318 (22.0%)
60+	143 (4.2%)	403 (9.2%)	323 (22.3%)
Recipient gender			
Male	1997 (58.9%)	2671 (61.3%)	913 (63.1%)
Female	1391 (41.1%)	1688 (38.7%)	535 (36.9%)
Donor gender			
Male	2039 (60.2%)	2563 (58.8%)	803 (55.5%)
Female	1349 (39.8%)	1796 (41.2%)	645 (44.5%)