HHS Public Access

Author manuscript

Stat Med. Author manuscript; available in PMC 2024 September 10.

Published in final edited form as:

Stat Med. 2023 September 10; 42(20): 3685–3698. doi:10.1002/sim.9825.

Healthcare Center Clustering for Cox's Proportional Hazards Model by Fusion Penalty

Lili Liu[1,6], **Kevin He**[2], **Di Wang**[2], **Shujie Ma**[3], **Annie Qu**[4], **Lu Lin**[5], **J. Philip Miller**[1], **Lei Liu**[1]

[1]Division of Biostatistics, Washington University in St. Louis, St. Louis, U.S.A.

[2]Department of Biostatistics, University of Michigan, Ann Arbor, U.S.A

[3]Department of Statistics, University of California, Riverside, California, U.S.A

[4]Department of Statistics, University of California, Irvine, California, U.S.A

[5]Zhongtai Securities Institute for Financial Studies, Shandong University, Jinan, China

[6]Research Center for Mathematics and Interdisciplinary Sciences, Shandong, University, Qingdao, China

Summary

There has been growing research interest in developing methodology to evaluate healthcare centers' performance with respect to patient outcomes. Conventional assessments can be conducted using fixed or random effects models, as seen in provider profiling. We propose a new method, using fusion penalty to cluster healthcare centers with respect to a survival outcome. Without any priori knowledge of the grouping information, the new method provides a desirable data-driven approach for automatically clustering healthcare centers into distinct groups based on their performance. An efficient alternating direction method of multipliers algorithm is developed to implement the proposed method. The validity of our approach is demonstrated through simulation studies, and its practical application is illustrated by analyzing data from the national kidney transplant registry.

Keywords

Latent class; Fixed effects; Random effects; Provider profiling; Fusion penalty

1 | INTRODUCTION

Assessing the comparative performance of healthcare centers (e.g., hospitals, nursing homes, transplant centers, or dialysis facilities) has attracted significant interest over the past decades. The objective is to outline and compare the performance of these centers in order to facilitate improvements through accountability and feedback. This information can aid

*Correspondence: Lei Liu, Division of Biostatistics, Washington University in St. Louis, St. Louis, U.S.A., lei.liu@wustl.edu.

SUPPLEMENTARY MATERIALS

The supplementary material for this article is available online.

individuals in selecting the most suitable healthcare facility and also enable stakeholders and payers to identify areas requiring enhancements.

Our motivating example is the national kidney transplant registry data collected by the U.S. Organ Procurement and Transplantation Network (OPTN). Our goal is to evaluate transplant centers based on their five-year post-transplant graft survival rates. For patients with end-stage renal disease, kidney transplantation provides the best opportunity for survival. The Scientific Registry of Organ Recipients (SRTR) commonly utilizes the five-year post-transplant graft survival metric, defined as the time until either death or graft failure within five years following transplantation, for regulatory monitoring of transplant centers[1,2]. Consequently, we will employ the five-year post-transplant graft survival to assess the quality of care provided by transplant centers.

Traditionally, profiling methods have been developed to evaluate the quality of care provided by various healthcare centers, using multiple patient outcome quality measures, such as readmission, mortality, and hospitalization. Existing transplant center profiling approaches typically employ inference-based procedures and generate a three-tier system, indicating whether centers perform worse than expected, as expected, or better than expected. Random effects and fixed effects models are two prevalent analytical methods used in profiling[3,4,5,6]. However, both models have their drawbacks. For random effects models, healthcare centers on the tails of the distribution tend to have small sample sizes, leading to substantially shrunk estimates towards the population mean[7,8]. This may result in reduced sensitivity when classifying healthcare centers in the tail areas, causing the majority of healthcare centers to be classified as expected, despite noticeable heterogeneity[9]. Additionally, misspecification of the random effects distribution can pose challenges in both estimation and inference. In contrast, fixed effects models suffer from a loss of efficiency due to a large number of parameters. Moreover, the simultaneous testing of the null hypothesis for extensive healthcare center effects is computationally demanding.

In order to offer more comprehensive ratings for kidney transplantation, the SRTR has implemented a five-tier rating system[10,11] that indicates whether a transplant center performs better than expected, somewhat better than expected, as expected, somewhat worse than expected, or worse than expected. However, concerns arise regarding the selection of appropriate cutoffs to categorize transplant centers into distinct groups. Furthermore, the decision regarding the total number of tiers is arbitrary.

To tackle the aforementioned challenges, we introduce a new fused effects model[12] designed to automatically identify homogeneous groups of healthcare centers without requiring a priori classification knowledge. We employ Cox's proportional hazards model[13] with fusion penalty[14] to cluster transplant centers based on the post-transplant graft survival outcome. Unlike random or fixed effects models, this new method offers a data-driven approach that does not rely on inference tests of statistical significance. Our model can also investigate risk factors associated with post-transplant graft survival. Our method can be considered as an alternative of the latent class model, where we use fusion penalty to "classify" providers into different latent groups.

We employ a local quadratic approximation to the partial likelihood and optimize the penalized partial likelihood with the fusion penalty. Prioritizing clustering accuracy, we opt for the smoothly clipped absolute deviation (SCAD)[15] penalty function over the LASSO penalty[16]. Compared to the LASSO, the SCAD penalty is nearly unbiased in identifying groups and enforces a sparser solution more aggressively[17]. The alternating direction method of multipliers (ADMM) algorithm can be utilized to implement the estimation, ensuring rapid convergence[18]. Due to the information loss during computation, we perform refitting by maximizing the log partial likelihood with the grouped data to obtain accurate parameter estimates.

The remainder of this paper is structured as follows: in Section 2, we outline the penalized Cox's regression model with fusion penalty for clustering healthcare centers. Section 3 evaluates the performance of our approach through Monte Carlo simulation studies. In Section 4, we demonstrate the proposed method using the kidney transplant data as a practical example. Finally, we summarize our methodology and discuss potential future directions in Section 5.

2 | METHODS

2. 1 | Model

We begin by introducing the required notations to formulate our model. For subject $j = 1, \ldots, n_i$ from healthcare center $i = 1, \ldots, m$, we have data in the format $(y_{ij}, \mathbf{x}_{ij}, \delta_{ij})$, where the observed time y_{ij} is the minimum of the censoring time c_{ij} and the event time $t_{ij}, \delta_{ij} = I(t_{ij} \leq c_{ij})$ is the censoring indicator, and $\mathbf{x}_{ij}$ is a $p \times 1$ vector of predictors. The Cox's proportional hazards model is

$$h_{ij}(t \mid \mathbf{x}_{ij}) = h_0(t)\exp\left(a_i + \mathbf{x}_{ij}^{\top}\boldsymbol{\beta}\right), \tag{2.1}$$

where $h_{ij}(t)$ is the hazard for patient j of center i at time t, $h_0(t)$ is a baseline hazard function, a_i is the center-specific effect, and $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_p)^{\top}$ is the vector of covariate coefficients. A constraint $\sum_{i=1}^{m} a_i = 0$ ensures that all parameters are identifiable. For estimation and inference, we often rely on the partial likelihood, where the unspecified baseline hazards can be canceled out. Patients are assumed to be independent within each healthcare center, i.e., we assume independence of y_{ij} given $\mathbf{x}_{ij}$ and a_i. The partial likelihood for model (2.1) can be written as

$$L\left(\boldsymbol{a}, \boldsymbol{\beta}\right) = \prod_{i=1}^{m}\prod_{j=1}^{n_i}\left[\frac{\exp\left(a_i + \mathbf{x}_{ij}^{\top}\boldsymbol{\beta}\right)}{\sum_{i'=1}^{m}\sum_{j'=1}^{n_{i'}} I(y_{i'j'} \geq y_{ij})\exp\left(a_{i'} + \mathbf{x}_{i'j'}^{\top}\boldsymbol{\beta}\right)}\right]^{\delta_{ij}}, \tag{2.2}$$

where $\boldsymbol{a} = (a_1, \cdots, a_m)^{\top}$. We assume that a_i belongs to one of K groups $\mathcal{G}_1, \cdots, \mathcal{G}_K$, which are mutually exclusive partitions of $\{1, \cdots, m\}$; and the number of groups is much smaller than that of centers, i.e., $K \ll m$. Moreover, the number of groups and the group membership are unknown in advance.

We utilize the fused SCAD penalty to identify homogeneous center performance and then fuse them as shared parameters to classify groups of healthcare centers. Incorporating the fusion penalty into the partial likelihood (2.2) results in the following optimization problem:

$$\left(\widehat{\boldsymbol{a}}, \widehat{\boldsymbol{\beta}}\right) = \arg\min\left[-\ell\left(\boldsymbol{a}, \boldsymbol{\beta}\right) + \sum_{1 \le i < k \le m} p_\gamma(|a_i - a_k|, \lambda)\right], \tag{2.3}$$

where $\ell(\boldsymbol{a}, \boldsymbol{\beta})$ is the log of the partial likelihood, and $p_\gamma(t, \lambda)$ is the SCAD penalty function[15] defined as

$$p_\gamma\left(t, \lambda\right) = \lambda \int_0^{|t|} \min\{1, (\gamma - x/\lambda)_+/(\gamma - 1)\} dx,$$

with $(x)_+ = x$ if $x > 0$ and = 0 otherwise, $\lambda \ge 0$ is a tuning parameter, $\gamma \ge 0$ is a parameter that controls the concavity of the penalty functions. Following Ma et al.,[17] we treat γ as a fixed constant.

2. 2 | Estimation procedure

Note that the penalty function $p_\gamma(|a_i - a_k|, \lambda)$ cannot be written in the form of addition of separate terms of $p_\gamma(|a_i|, \lambda)$ and $p_\gamma(|a_k|, \lambda)$ as in LASSO. Here, we introduce a new set of parameters $\theta_{ik} = a_i - a_k$, which are equivalent to the pairwise differences of healthcare center effects. Consequently, the above minimization problem (2.3) can be transformed into the following constraint optimization problem:

$$F_0\left(\boldsymbol{a}, \boldsymbol{\beta}, \boldsymbol{\theta}\right) = -\ell\left(a, \boldsymbol{\beta}\right) + \sum_{i < k} p_\gamma(|\theta_{ik}|, \lambda) \text{ subject to } a_i - a_k - \theta_{ik} = 0, \tag{2.4}$$

where $\boldsymbol{\theta} = \{\theta_{ik}, i < k\}^\top$. As a result, the alternating direction method of multipliers (ADMM) algorithm can be used to identify the groups in the objective function (2.4). The ADMM algorithm combines the strengths of dual decomposition and augmented Lagrangian methods for constrained optimization. The estimates of the parameters are obtained by the augmented Lagrangian

$$F\left(\boldsymbol{a}, \boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{\upsilon}\right) = F_0\left(\boldsymbol{a}, \boldsymbol{\beta}, \boldsymbol{\theta}\right) + \sum_{i < k} \upsilon_{ik}(\theta_{ik} - a_i + a_k) + \frac{\vartheta}{2} \sum_{i < k} (\theta_{ik} - a_i + a_k)^2, \tag{2.5}$$

where $\boldsymbol{\upsilon} = \{\upsilon_{ik}, i < k\}^\top$ are Lagrange multipliers, ϑ is the penalty parameter. We use ADMM to iteratively compute the estimates of $(\boldsymbol{a}, \boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{\upsilon})$. For given $\boldsymbol{\theta}^{(s)}, \boldsymbol{\upsilon}^{(s)}$ at step s, the iterations can be specified as follows:

$$\left(\boldsymbol{a}^{(s+1)}, \boldsymbol{\beta}^{(s+1)}\right) = \arg\min_{\boldsymbol{a}, \boldsymbol{\beta}} F\left(\boldsymbol{a}, \boldsymbol{\beta}, \boldsymbol{\theta}^{(s)}, \boldsymbol{\upsilon}^{(s)}\right), \tag{2.6}$$

$$\boldsymbol{\theta}^{(s+1)} = \arg\min_{\boldsymbol{\theta}} F\left(\boldsymbol{a}^{(s+1)}, \boldsymbol{\beta}^{(s+1)}, \boldsymbol{\theta}, \boldsymbol{\upsilon}^{(s)}\right), \tag{2.7}$$

$$\upsilon_{ik}^{(s+1)} = \upsilon_{ik}^{(s)} + \vartheta\left(a_i^{(s+1)} - a_k^{(s+1)} - \theta_{ik}^{(s+1)}\right). \tag{2.8}$$

To update $\boldsymbol{a}$ and $\boldsymbol{\beta}$, minimizing (2.6) is equivalent to minimizing

$$f\left(\boldsymbol{a}, \boldsymbol{\beta}\right) = -\ \ell\left(\boldsymbol{a}, \boldsymbol{\beta}\right) + \frac{\vartheta}{2}\sum_{i<k}\left(a_i - a_k - \theta_{ik}^{(s)} + \vartheta^{-1}\upsilon_{ik}^{(s)}\right)^2 + C, \tag{2.9}$$

where $\boldsymbol{C}$ is the constant independent of $\boldsymbol{a}$ and $\boldsymbol{\beta}$. To solve the optimization problem (2.9), we approximate the nonlinear logpartial likelihood function using a two-term Taylor series expansion. At each iteration, we solve a reweighted least squares problem 19. Specifically, let $\mathbf{y} = \left(\mathbf{y}_1^\top, \cdots, \mathbf{y}_m^\top\right)^\top$ with $\mathbf{y}_i = \left(y_{i1}, \cdots, y_{in_i}\right)^\top, \mathbf{X} = \left(\mathbf{X}_1^\top, \cdots, \mathbf{X}_m^\top\right)^\top$ with $\mathbf{X}_i = \left(\mathbf{x}_{i1}, \cdots, \mathbf{x}_{in_i}\right)^\top$, $\mathbf{A} = \mathrm{diag}(1_{n_1}, \cdots, 1_{n_m})$ with $1_{n_i} = (1, \cdots, 1)^\top$ being the vector with n_i ones, and $\boldsymbol{\eta} = \mathbf{A}\boldsymbol{a} + \mathbf{X}\boldsymbol{\beta}$. Let $\ell'(\boldsymbol{\eta})$, $\ell''(\boldsymbol{\eta})$ denote the gradient and Hessian of the log-partial likelihood with respect to $\boldsymbol{\eta}$, respectively. The log-partial likelihood $\ell(\boldsymbol{a}, \boldsymbol{\beta})$ can be approximated by the following quadratic form (see the supplementary material)

$$\frac{1}{2}\left(\mathbf{z} - \mathbf{A}\boldsymbol{a} - \mathbf{X}\boldsymbol{\beta}\right)^\top \ell''\left(\boldsymbol{\eta}\right)\left(\mathbf{z} - \mathbf{A}\boldsymbol{a} - \mathbf{X}\boldsymbol{\beta}\right)$$

where $\mathbf{z} = \boldsymbol{\eta} - \ \ell''(\boldsymbol{\eta})^{-1}\ \ell'(\boldsymbol{\eta})$ with

$$\frac{\partial\, \ell\,(\boldsymbol{\eta})}{\partial \eta_{ij}} = \delta_{ij} - \exp(\eta_{ij})\sum_{i'j' \in D_{ij}} \frac{1}{\sum_{uv \in R_{i'j'}} \exp(\eta_{uv})},$$
$$\frac{\partial^2\, \ell\,\left(\boldsymbol{\eta}\right)}{\partial \eta_{ij}^2} = -\exp(\eta_{ij})\sum_{i'j' \in D_{ij}} \frac{1}{\sum_{uv \in R_{i'j'}} \exp(\eta_{uv})} + \exp(2\eta_{ij})\sum_{i'j' \in D_{ij}} \frac{1}{\left\{\sum_{uv \in R_{i'j'}} \exp(\eta_{uv})\right\}^2},$$

where D_{ij} is the set of indices $i'j'$ with $t_{i'j'} < y_{ij}$ (the times for which observation of the i th subject in the j th center is still at risk) and $R_{i'j'}$ is the set of indices uv with $y_{uv} \geq t_{i'j'}$ (those at risk at time $t_{i'j'}$). Since $\ell''(\boldsymbol{\eta})$ is a full matrix, it requires computation of $O\left(N^2\right)$ elements, we instead replace $\ell''(\boldsymbol{\eta})$ by a diagonal matrix with the same diagonal elements as $\ell''(\boldsymbol{\eta})$.[20,21,19] This substitution works well because the diagonal elements of $\ell''(\boldsymbol{\eta})$ are much larger than the off-diagonal elements. Denote $-\ \ell''(\boldsymbol{\eta})$ by $\mathbf{W}$. $\mathbf{W}$ and $\mathbf{z}$ are computed based on $\boldsymbol{a}^{(s)}$ and $\boldsymbol{\beta}^{(s)}$ at iteration s. Equation (2.9) can be rewritten as

$$f\left(\boldsymbol{a}, \boldsymbol{\beta}\right) = \frac{1}{2}\left(\mathbf{z} - \mathbf{A}\boldsymbol{a} - \mathbf{X}\boldsymbol{\beta}\right)^\top \mathbf{W}\left(\mathbf{z} - \mathbf{A}\boldsymbol{a} - \mathbf{X}\boldsymbol{\beta}\right) + \frac{\vartheta}{2}\| \mathbf{A}\boldsymbol{a} - \boldsymbol{\theta}^{(s)} + \vartheta^{-1}\boldsymbol{\upsilon}^{(s)} \|^2 + C, \tag{2.10}$$

where $\mathbf{\Delta} = \{(\mathbf{e}_i - \mathbf{e}_j), i < j\}^\top$ with $\mathbf{e}_i$ being a $m \times 1$ vector whose i th element is 1 and the remaining elements are 0. Thus, for given $\boldsymbol{\theta}^{(s)}, \boldsymbol{\upsilon}^{(s)}$ at the sth step, we set the derivatives $\partial f(\boldsymbol{\alpha}, \boldsymbol{\beta})/\partial \boldsymbol{\alpha} = 0$ and $\partial f(\boldsymbol{\alpha}, \boldsymbol{\beta})/\partial \boldsymbol{\beta} = 0$ to obtain the following updates $\boldsymbol{\alpha}^{(s+1)}$ and $\boldsymbol{\beta}^{(s+1)}$:

$$\boldsymbol{\alpha}^{(s+1)} = \left(\mathbf{A}^\top \mathbf{Q}_x \mathbf{A} + \vartheta \mathbf{\Delta}^\top \mathbf{\Delta}\right)^{-1} \left[\mathbf{A}^\top \mathbf{Q}_x \mathbf{z} + \vartheta \mathbf{\Delta}^\top \left(\boldsymbol{\theta}^{(s)} - \vartheta^{-1} \boldsymbol{\upsilon}^{(s)}\right)\right] \tag{2.11}$$

where $\mathbf{Q}_x = \mathbf{W} - \mathbf{W}\mathbf{X}\left(\mathbf{X}^\top \mathbf{X}\right)^{-1} \mathbf{X}^\top \mathbf{W}$. We let $\boldsymbol{\alpha}^{(s+1)} = \boldsymbol{\alpha}^{(s+1)} - \text{mean}\left(\boldsymbol{\alpha}^{(s+1)}\right)$ to guarantee that the estimate $\boldsymbol{\alpha}^{(s+1)}$ satisfies the constraint $\sum_{i=1}^{m} a_i = 0$, and

$$\boldsymbol{\beta}^{(s+1)} = \left(\mathbf{X}^\top \mathbf{W} \mathbf{X}\right)^{-1} \mathbf{X}^\top \mathbf{W}\left(\mathbf{z} - \mathbf{A}\boldsymbol{\alpha}^{(s+1)}\right). \tag{2.12}$$

To update θ, we minimize the function

$$\frac{\vartheta}{2}\left(\theta_{ik} - \pi_{ik}^{(s)}\right)^2 + \sum_{i<k} p_\gamma(|\theta_{ik}|, \lambda)$$

where $\pi_{ik}^{(s)} = a_i^{(s)} - a_k^{(s)} + \vartheta^{-1} \upsilon_{ik}^{(s)}$. It is worth noting that by using the concave penalties, the objective function $F(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{\upsilon})$ is no longer a convex function. However, it is convex with respect to each θ_{ik} when $\gamma > 1/\vartheta + 1$ for the SCAD penalty; consequently, given $(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\upsilon})$, the minimizer of $F(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{\upsilon})$ with respect to θ_{ik} is unique and has a closed-form solution as follows:

$$\theta_{ik}^{(s+1)} = \begin{cases} S\left(\pi_{ik}^{(s)}, \lambda/\vartheta\right) & \left|\pi_{ik}^{(s)}\right| \le \lambda + \lambda/\vartheta, \\ \dfrac{S\left(\pi_{ik}^{(s)}, \gamma\lambda/\left((\gamma-1)\vartheta\right)\right)}{1 - 1/\left((\gamma-1)\vartheta\right)} & \lambda + \lambda/\vartheta < \left|\pi_{ik}^{(s)}\right| \le \lambda\gamma, \\ \pi_{ik}^{(s)} & \left|\pi_{ik}^{(s)}\right| > \lambda\gamma, \end{cases} \tag{2.13}$$

where $S(x, t) = x(1 - t/|x|)_+$ is a groupwise soft thresholding operator. Finally, the Lagrange multiplier υ_{ik} is updated by (2.8). This process is conducted iteratively until the convergence over a grid of values for λ. The iterative algorithm terminates when primal residuals $\mathbf{r}_{\text{primal}}^{(s+1)} = \mathbf{\Delta}\boldsymbol{\alpha}^{(s+1)} - \boldsymbol{\theta}^{(s+1)}$ and dual residuals $\mathbf{r}_{\text{dual}}^{(s+1)} = \vartheta \mathbf{\Delta}^\top\left(\boldsymbol{\theta}^{(s+1)} - \boldsymbol{\theta}^{(s)}\right)$ are close to zero, i.e., $\| \mathbf{r}_{\text{primal}}^{(s+1)} \| < \epsilon_{\text{primal}}^{(s+1)}$ and $\| \mathbf{r}_{\text{dual}}^{(s+1)} \| < \epsilon_{\text{dual}}^{(s+1)}$, where $\epsilon_{\text{dual}}^{(s+1)}$ and $\epsilon_{\text{primal}}^{(s+1)}$ are specified as suggested by Boyd et al.[18] It is important to find appropriate initial values for the ADMM algorithm. In this paper, the initial values $\boldsymbol{\alpha}^{(0)}$ are obtained from the fixed effects Cox model, then we set $\theta_{ik}^{(0)} = a_i^{(0)} - a_k^{(0)}$ and $\boldsymbol{\upsilon}^{(0)} = 0$. If $\hat{\theta}_{ik} = 0$, a_i and a_k are classified into the same group. As a result, we obtain $\hat{K}$ estimated groups $\hat{\mathcal{G}}_1, \cdots, \hat{\mathcal{G}}_{\hat{K}}$, and set α_k to be the common value of a_i 's from the k th group.

Due to the approximation in the computation, Equation (2.10) does not fully capture parameter information, and the iteration may also lead to a loss in efficiency. Once the groups have been identified, we conduct a refitting step to estimate $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ by maximizing the following log-partial likelihood with the grouped information:

$$\ell\left(\boldsymbol{\alpha},\boldsymbol{\beta}\right)=\sum_{k=1}^{\widehat{K}}\sum_{i=\hat{G}_k}\sum_{j=1}^{n_i}\delta_{ij}\left\{\alpha_k+\mathbf{x}_{ij}^{\top}\boldsymbol{\beta}-\log\left[\sum_{i'j'\in R_{ij},\,i'\in\hat{\mathcal{G}}_k}\exp\left(\alpha_k+\mathbf{x}_{i'j'}^{\top}\boldsymbol{\beta}\right)\right]\right\},\tag{2.14}$$

where $\boldsymbol{\alpha}=(\alpha_1,\cdots,\alpha_{\hat{K}})^{\top}$, and $\boldsymbol{R}_{ij}$ is the set of indices $i'j'$ with $y_{i'j'}\geq t_{ij}$ (those at risk at time t_{ij}). Here rather than the constraint $\sum_{i=1}^{\hat{K}}\alpha_k=0$ (as for a_i in Section 2.1), we assume that $\alpha_{\hat{k}}=0$ to simplify the interpretation, where $\hat{k}=\arg\min_{k=1,\cdots,\hat{K}}|\alpha_k|$. The above procedure can be conveniently implemented using R function coxph, which also provides standard errors and p-values of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$.

After obtaining a path of solutions, it becomes essential to choose an optimal tuning parameter λ by minimizing the modified Bayesian Information Criterion (BIC) using a grid search.[22]. The BIC is given by

$$\mathrm{BIC}(\lambda)=-2\,\ell\left(\hat{\boldsymbol{a}}(\lambda),\hat{\boldsymbol{\beta}}(\lambda)\right)+C_N\left(\widehat{K}(\lambda)+p\right)\log N,$$

where $\hat{\boldsymbol{a}}(\lambda),\hat{\boldsymbol{\beta}}(\lambda)$ and $\widehat{\boldsymbol{K}}(\lambda)$ are the estimates of $\boldsymbol{a},\boldsymbol{\beta}$ and K at given λ, respectively. $\ell\left(\hat{\boldsymbol{a}}(\lambda),\hat{\boldsymbol{\beta}}(\lambda)\right)$ is the log-partial likelihood evaluated at $\hat{\boldsymbol{a}}(\lambda)$ and $\hat{\boldsymbol{\beta}}(\lambda)$, p is the dimension of the parameter $\boldsymbol{\beta}$, and C_N is a positive number depending on the total number of observations $N=\sum_{i=1}^{m}n_i$. If $C_N=1$, the modified BIC reduces to the traditional BIC[23] Following Wang et al[22], $C_N=\log(\log(N+p))$.

We summarize our method in Algorithm 1 below.

Algorithm 1 ADMM algorithm for healthcare center clustering

Require: Initialize $\boldsymbol{\theta}^{(0)}$ and $\boldsymbol{v}^{(0)}$.

for s = 0,1,2,⋯**do**

Compute $\boldsymbol{a}^{(s+1)}$ using (2.11)

Compute $\boldsymbol{\beta}^{(s+1)}$ using (2.12)

Compute $\boldsymbol{\theta}^{(s+1)}$ using (2.13)

Compute $\boldsymbol{v}^{(s+1)}$ using (2.8)

if the convergence criterion is met, **then**

Stop and denote the last iteration by $\left(\hat{\boldsymbol{a}}(\lambda),\hat{\boldsymbol{\beta}}(\lambda)\right)$,

else

$s=s+1$.

end if

end for

After identifying the groups, estimate $\left(\widehat{\boldsymbol{\alpha}},\widehat{\boldsymbol{\beta}}\right)$ using (2.14).

Ensure: Output

3 | SIMULATION

In this section, we evaluate the finite sample performance of the proposed methods through simulation studies. Conventionally, in Model (2.1), the healthcare center effect a_i is treated as either a random effect or a fixed effect. We compare our fused effects model with SCAD penalty (SCAD) to the random effects model (RE) and fixed effects model (FE). Specifically, we examine the efficiency of estimation and the accuracy of healthcare center classification. We consider two different numbers of centers, $m = 50$ and 100, and obtain all simulation results via 100 replicates.

Example 1. To assess the relative performance of different models, we mimic the censoring rate observed in the real-life application of kidney transplant centers in Section 4. We assume that among the m center effects a_i 's, 10% are set to - 1, representing healthcare centers performing "better than expected," 10% are set to 1, indicating centers performing "worse than expected," and the remaining 80% are set to 0, representing centers performing "as expected." The survival time is generated from the Weibull distribution with scale and shape parameters of 2 and 3, respectively. The censoring time is generated from a uniform distribution $U(0,1)$ to achieve approximately a 70% censoring rate, which is close to the real data. The covariates $x_{ij} = (x_{i1}, x_{i2})^\top$ are generated from the multivariate normal distribution with mean 0, variance 1 and an exchangeable correlation $\rho = 0.2$. The coefficient $\boldsymbol{\beta} = (2,2)^\top$. We consider the number of patients in each center $n_i \sim$ Uniform (50,100).

The Rand Index (RI)[24] is employed to assess the level of agreement between the estimated partitions and the true partitions. Each pair of observations a_i and a_k falls to one of four categories: (i) true positive (TP) where a_i and a_k from the same group are assigned to the same cluster; (ii) true negative (TN) where a_i and a_k from different groups are assigned to different clusters; (iii) false negative (FN) where a_i and a_k from different groups are assigned to the same cluster; (iv) false positive (FP) where a_i and a_k from the same group are assigned to different clusters. The Rand Index is given by

$$\mathrm{RI} = \frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{TP}+\mathrm{FP}+\mathrm{TN}+\mathrm{FN}} = \frac{\mathrm{TP}+\mathrm{TN}}{\binom{N}{2}}.$$

Intuitively, TP and TN represent agreement between the true group and the estimated cluster, while FP and FN indicate disagreement between the true group and the estimated cluster. RI ranges from 0 to 1, with a larger value indicating a higher degree of agreement.

The top panel of Table 1 presents the mean, median, and standard deviation (SD) of the estimated number of groups $\widehat{K}$, the percentage of $\hat{K}$ equal to the true number of groups (per), and the RI for evaluating the classification accuracy. As expected, the RI and the percentages of correctly classifying centers in comparison to the reference are very close to

1 and improve as m increases. To visually display the distribution of $\widehat{K}$, the histograms of $\widehat{K}$ are depicted in Figure 1

Let $\hat{\alpha}_1, \hat{\alpha}_2$ and $\hat{\alpha}_3$ be the average estimates of center effects from "better than expected", "as expected", and "worse than expected" groups, with true values being −1,0 and 1, respectively. The top panel of Table 2 presents the bias, the standard deviation (SD), the standard error (SE), and the coverage probability of 95% confidence intervals (CP) of the estimators $\hat{\alpha}_1$ and $\hat{\alpha}_3$. The estimator $\hat{\alpha}_2$ serves as a reference, such as the national norm. The Oracle estimators are obtained with a priori knowledge of the true grouping information. As not all replications are clustered into three groups by SCAD, we only use the replications with the estimated number of groups equal to three to compute the bias, SD, SE, and CP of $\hat{\alpha}_1$ and $\hat{\alpha}_3$. For the Oracle, the measures are calculated based on all 100 replications. From Table 2. We notice that our method performs very closely to the Oracle, as it can accurately recover the group structure. Evidently, our estimators $\hat{\alpha}_1$ and $\hat{\alpha}_3$ align well with the corresponding true values on average for all cases. The inference is also adequately precise, with a strong correspondence between SD and SE, and the coverage probabilities are near the nominal level of 0.95. All the original estimates of α_1, α_2, and α_3 without the refitting step are reported in Table S1 of the supplementary material.

We compute the mean squared error (MSE) of $\hat{\boldsymbol{a}}$ using the formula $\sum_{i=1}^{m}(\hat{a}_i - a_i)^2/m$ for each replicated dataset. The top panel of Table 3 presents the MSE of the estimator $\hat{\boldsymbol{a}}$, the bias, the standard deviation (SD), the standard error (SE), and the coverage probability of 95% confidence intervals (CP) for the estimators $\hat{\beta}_1$ and $\hat{\beta}_2$. It is worth mentioning that, in contrast to Table 2, which only displays results for replications with an estimated group count of three, Table 3 includes results for all replicates, regardless of whether the estimated number of groups is three or not. The MSEs of $\hat{\boldsymbol{a}}$ from our fused effects model are much smaller than those from the random effects and fixed effects models in all cases. As a result, our method not only accurately classifies centers but also obtains precise estimates of these centers. To graphically visualize the numerical results of Table 3, the boxplots of the MSE of $\hat{\boldsymbol{a}}$ are depicted in Figure 2 For the estimators $\hat{\boldsymbol{\beta}}$, our method and the random effects model exhibit satisfactory performance, while the fixed effects model yields larger bias.

Example 2. In this example, we generate the covariates $x_{ij} = (x_{i1}, x_{i2})^\top$ from the multivariate normal distribution with mean 0, variance 1, and an exchangeable correlation $\rho = 0.2$. The survival time is generated from the Weibull distribution with scale and shape parameters being 2 and 3, respectively. The censoring time is generated from a uniform distribution $U(0,1)$ to achieve about a 70% censoring rate. We set the coefficient $\boldsymbol{\beta} = (0.1,0.1)^\top$, and the number of patients in each center $n_i \sim$ Uniform (50, 100). To demonstrate the robustness of our method, a_i is simulated from the standard normal distribution $N(0,1)$. Consequently, the random effects model is the correct model.

The grouping results of $\hat{a}_i$ using our method are presented in the second panel of Table 1 The median of the estimated number of groups $\hat{K}$ is 8 with $m = 50$ and 9 with $m = 100$. The fusion penalty generally tends to select fewer groups for $m = 50$ and more groups for $m = 100$. As

the number of parameters increases with m, the median and the standard deviation of $\widehat{K}$ also grow. To graphically visualize the distribution of $\widehat{K}$, the histograms of $\widehat{K}$ are shown in Figure 1

The second panel of Table 3 presents the mean squared error (MSE) of the estimator $\hat{\boldsymbol{a}}$, the bias, the standard deviation (SD), the standard error (SE), and the coverage probability of 95% confidence intervals (CP) for the estimators $\widehat{\beta}_1$ and $\widehat{\beta}_2$. From Table 3. we observe that the MSE values of $\hat{\boldsymbol{a}}$ using our method are smaller than those in the fixed effects model. Moreover, our method outperforms the random effects model when the number of centers is small ($m = 50$), likely due to the shrinkage of the predicted values of a_i 's in the random effects model. As the number of centers increases, the random effects model performs the best, but our method's performance is only slightly worse than the random effects model and still much better than the fixed effects model. All methods exhibit similar performance in estimating $\widehat{\beta}_1$ and $\widehat{\beta}_2$.

Example 3. In this example, we evaluate the performance of the proposed model with four groups. The covariates $x_{ij}, \boldsymbol{\beta}, n_i,\ h_0(t)$, and the censoring time are generated from the same distributions as in Example 1. We assume that among the m center effects a_i 's, 10% are set to 1 and 10% are set to 2, indicating centers performing at different levels of "worse than expected"; 20% are set to −1.5, signifying centers performing "better than expected"; the remaining 60% are set to 0, indicating centers performing "as expected".

The grouping results of $\hat{a}_i$ using our method are presented in the bottom panel of Table 1, where the medians of $\hat{K}$ over the 100 replicates are 4, which is the true number of subgroups, and the mean values are very close to 4. Furthermore, the RI values and the percentage of correctly selecting the number of subgroups approach 1. Therefore, our methods perform well in cases with an even number of groups. To graphically visualize the distribution of $\widehat{K}$, the histograms of $\widehat{K}$ are shown in Figure 1.

Let $\widehat{\alpha}_1, \widehat{\alpha}_2, \widehat{\alpha}_3$, and $\widehat{\alpha}_4$ be the average estimates for the a_i 's from four groups where the true values are −1.5,0,1, and 2, respectively. The bottom panel of Table 2 presents the bias, the standard deviation (SD), the standard error (SE), and the coverage probability of 95% confidence intervals (CP) for the estimators $\widehat{\alpha}_1, \widehat{\alpha}_3$, and $\widehat{\alpha}_4$. The estimator $\widehat{\alpha}_2$, which has the smallest absolute value, is set as the reference. From the bottom panel of Table 2. we can see that the means of $\widehat{\alpha}_1, \widehat{\alpha}_3$, and $\widehat{\alpha}_4$ are close to the true values and the Oracle estimators. We also observe that the SDs of $\widehat{\alpha}_1, \widehat{\alpha}_3$, and $\widehat{\alpha}_4$ are close to the corresponding SEs, resulting in valid coverage probabilities.

The bottom panel of Table 3 reports the MSE of the estimator $\hat{\boldsymbol{a}}$, the bias, the standard deviation (SD), the standard error (SE), and the coverage probability of 95% confidence intervals (CP) for the estimators $\hat{\beta}_1$ and $\hat{\beta}_2$. From the bottom panel of Table 3 we observe that the MSE values of $\hat{a}$ using SCAD are smaller than those of the random effects and fixed effects models. For the estimators $\widehat{\boldsymbol{\beta}}$, our method and the random effects model exhibit satisfactory performance, while the fixed effects model produces larger bias. These results indicate that the proposed method performs well with an even number of groups.

4 | APPLICATION

To demonstrate the proposed methods, we conduct an evaluation of kidney transplant centers using the national kidney transplant registry data obtained from the U.S. Organ Procurement and Transplantation Network (https://optn.transplant.hrsa.gov/data/). We limit the study cohort to adult kidney transplant recipients (age ≥ 18) who received a transplant between January 1, 2007, and December 31, 2007. The analysis cohort includes 4198 patients from 60 centers, with the number of patients per center ranging from 50 to 100. In our dataset, out of 4198 patients who received their kidney transplant in 2007, 1137 (27.1%) either died or experienced graft failure within five years after receiving a kidney transplant. All others were censored at five years of follow-up. In our survival analysis, the failure time (termed "graft survival") t_{ij} is defined as the duration (in years) from transplantation to graft failure or death, whichever occurred first. The censoring time c_{ij} is at the end of the five-year period after the transplantation. Thus, the observed time to event is $y_{ij} = \min(t_{ij}, c_{ij})$, with a 72.9% censoring rate. We apply our method to investigate the risk factors of five-year graft survival using Cox's proportional hazards model and assess the performance of centers concerning their graft outcomes.

The study cohort included 15 baseline characteristics of donors and recipients. The characteristics of the study population are as follows: Time on end-stage renal disease (ESRD, reference: <1 years), donor age (reference: 30–45 years old), donor gender (male = 1, female=0), donor body mass index (BMI, reference: normal), donor race (reference: white), donor history of hypertension (DON-HTN, yes=1, no=0), donor meeting expanded criteria (DON-EC, yes=1, no=0), recipient gender (male = 1, female=0), recipient race (reference: white), recipient insulin dependent diabetes (REC-DIAB Type I, yes=1, no=0), recipient non-insulin dependent diabetes (REC-DIAB Type II, yes=1, no=0), recipient age at transplant (REC-AGE, reference: 50–60), recipient body mass index (reference: normal), recipient previous kidney transplant (REC-PREV-KI, yes=1, no=0), recipient total cold ischemia time (REC-COLD-ISCH, > 20 hours = 1,< 20 hours =0).

We use the Akaike Information Criteria (AIC) to assess the performance of three methods. A smaller value of AIC indicates better performance. When treating the center effect a_i as a random intercept, we obtain an AIC of 18501.43. For the fixed effects model, the AIC value is 18535.64. If a_i is estimated by our fused effects model, the 60 centers are classified into 3 groups with an AIC of 18451.24. Our method leads to a significant improvement in model fitting.

Table 4 reports the estimate (Est.), standard error (SE), and p-value of $\hat{\boldsymbol{\beta}}$ for testing the significance of the coefficients by our fused effects model with the SCAD penalty (SCAD), the fixed effects model (FE) and the random effects model (RE). Longer time spent in end-stage renal disease, donor age over 60, black donor, donor history of hypertension, black recipient, recipient aged 60 or older at transplant, recipient with non-insulin dependent diabetes, and donor with low body mass index all have a significantly worse effect on five-year graft survival according to all three methods, with p-values less than 0.05. Asian recipients, on the other hand, tended to experience better survival outcomes.

Next we utilize the standardized mortality ratio (SMR) as an evaluation measure to assess center-specific survival, defined as the ratio of the observed number of deaths at a given center to the number expected if the center had mortality equal to the population average[25,4]. An SMR greater or smaller than 1 indicates that the center's observed mortality ratio is under-performing or over-performing relative to the population norm, respectively. Let $\hat{\alpha}_k$ be the common value of estimators $\hat{a}_i$'s in group $\widehat{\mathcal{G}_k}$ with $k = 1,2,3$. Table 5 reports the estimates $\hat{\alpha}_1$ and $\hat{\alpha}_3$ relative to a reference $\hat{\alpha}_2$ and the number of elements (num.) in each group, standard error (SE), p-value and SMR by our fused effects method. We identify 75.0% (45) of centers as the reference, 10.0% (6) of centers better than the reference and 15.0% (9) of centers worse than the reference. The estimated SMR of better and worse centers are 0.574 and 1.636, respectively.

In our national kidney transplant dataset, we only have a center-specific factor - the numbers of patients within a center (denoted by center size n_i). Since center-level confounder, e.g., n_i is not identifiable with the center-specific effect a_i, we cannot include n_i directly in our model. Along the lines of He et al.[4], we instead fit the model $n_i = \beta\hat{a}_i + \epsilon_i$, where $\hat{a}_i$ is the estimated center specific effect. The p-value is 0.915, suggesting a minimal correlation between the center-specific effects and the center size. This observation is further supported by the boxplots in Figure 3.

5 | DISCUSSION

In this study, we proposed an innovative fusion method to assess healthcare centers concerning survival outcomes. Our method proves more efficient than the fixed effects model by using fewer parameters for healthcare centers. Additionally, it outperforms random effects models in identifying healthcare centers with better classification accuracy and lower bias. Through simulation studies, we demonstrated that our method surpasses existing approaches in performance. Our model can be thought as an alternative way to latent class models, as we use fusion penalty to classify different healthcare centers into latent subgroups.

There are several potential extensions for our method. First, when the number of centers is very large, fitting the pairwise fusion penalized Cox model directly becomes computationally demanding. In such cases, we could employ the divide-andconquer strategy[26], which typically involves dividing the full sample into multiple subsets, solving the optimization problem for each subset, and combining the subset-specific estimates into a single estimate. Second, our method can be applied to cluster regression coefficients in the Cox's proportional hazards model when parameters are partially heterogeneous across subgroups. For instance, grouping treatment heterogeneity can enable the provision of precise medical treatments to diverse patient subgroups[27.17].

Finally, it is of interest to consider more sophisticated hierarchical data, e.g., patients clustered within practitioners which are clustered within practices. For this case, a three-level Cox's proportional hazards model with two fusion effects can be proposed. Let x_{ijk} denote the covariates for patient k nested within doctor j in center i, the three-level Cox model can be written as

$$h_{ijk}(t \mid \mathbf{x}_{ijk}) = h_0(t)\exp(a_i + b_{ij} + \mathbf{x}_{ijk}^{\top}\boldsymbol{\beta}),$$

where a_i is the center-specific effect, b_{ij} is the practitioner-specific effect, $h_{ijk}(t)$ is the hazard for patient k treated by practitioner j in center i at time t, and $h_0(t)$ is a baseline hazard function. We can cluster a_i and b_{ij} by using two fusion penalties. Estimation and inference of this model is of future interest.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENT

This research is partly supported by NIH grants R21 EY031884, R21 EY033518, R21 AG063370, UL1 TR002345, China Postdoctoral Science Foundation 2021M691952, Natural Science Foundation of Shandong Province ZR2021QA024.

DATA AVAILABILITY STATEMENT

Restrictions apply to the availability of these data, which were used under license for this study.

References

1. Wynn JJ, Distant DA, Pirsch JD, et al. Kidney and pancreas transplantation. American Journal of Transplantation 2004; 4: 72–80.
2. Wang JH, Skeans MA, Israni AK. Current status of kidney transplant outcomes: dying to survive. Advances in chronic kidney disease 2016; 23(5): 281–286. [PubMed: 27742381]
3. Normand ST, Shahian DM. Statistical and clinical aspects of hospital outcomes profiling. Statistical Science 2007; 22: 206–226.
4. He K, Kalbfleisch JD, Li Y, Li Y. Evaluating hospital readmission rates in dialysis facilities; adjusting for hospital effects. Lifetime Data Analysis 2013; 19: 490–512. [PubMed: 23709309]
5. Estes JP, Chen Y, Sentürk D, et al. Profiling dialysis facilities for adverse recurrent events. Statistics in Medicine 2020; 39: 1374–1389. [PubMed: 31997372]
6. Kalbfleisch JD, Wolfe RA. On monitoring outcomes of medical providers. Statistics in Biosciences 2013; 5: 286–302.
7. Racz M, Sedransk J. Bayesian and frequentist methods for provider profiling using risk-adjusted assessments of medical outcomes. Journal of the American Statistics Association 2010; 105: 48–58.
8. Ohlssen D, Sharples L, Spiegelhalter D. A hierarchical modelling framework for identifying unusual performance in health care providers. Journal of the Royal Statistical Society A 2007; 170: 865–890.
9. Salkowski N, Snyder J, Zaun D, Leighton T, Israni A, Kasiske BL. Bayesian methods for assessing transplant program performance. American Journal of Transplantation 2014; 14(6): 1271–1276. [PubMed: 24787026]
10. Van Pilsum Rasmussen SE, Thomas AG, Garonzik-Wang J, et al. Reported effects of the Scientific Registry of Transplant Recipients 5-tier rating system on US transplant centers: results of a national survey. Transplant International 2018; 31(10): 1135–1143. [PubMed: 29802802]
11. Wey A, Salkowski N, Kasiske BL, Israni AK, Snyder JJ. A five-tier system for improving the categorization of transplant program performance. Health services research 2018; 53(3): 1979–1991. [PubMed: 28608369]

12. Liu L, Gordon M, Miller J, et al. Capturing Heterogeneity in Repeated Measures Data by Fusion Penalty. Statistics in Medicine 2021; 40: 1901–1916. [PubMed: 33517583]
13. Cox DR. Regression Models and Life Tables. Journal of the Royal Statistical Society B 1972; 34: 187–220.
14. Tibshirani R, Saunders M, Rosset S, Zhu J, Knight K. Sparsity and smoothness via the fused lasso. Journal of the Royal Statistical Society B 2005; 67: 91–108.
15. Fan J, Li R. Variable selection via nonconcave penalized likelihood and its oracle properties. Journal of the American Statistical Association 2001; 96: 1348–1360.
16. Tibshirani R The Lasso method for variable selection in the Cox model. Statistics in Medicine 1997; 16: 385–395. [PubMed: 9044528]
17. Ma S, Huang J. A concave pairwise fusion approach to subgroup analysis. Journal of the American Statistical Association 2017; 112: 410–423.
18. Boyd S, Parikh N, Chu E, Peleato B, Eckstein J. Distributed optimization and statistical learning via the alternating direction method of multipliers. Foundations and Trends in Machine Learning 2011; 3: 1–122.
19. Hastie T, Tibshirani R. Generalized Additive Models. Chapman and Hall, London.. 1990.
20. Gui J, Li H. Penalized Cox Regression Analysis in the High-Dimensional and Low Sample Size Settings, with Applications to Microarray Gene Expression Data. Bioinformatics 2005; 25: 3001–3008.
21. Simon N, Friedman J, Hastie T, Tibshirani R. Regularization Paths for Cox's Proportional Hazards Model via Coordinate Descent. Journal of Statistical Software 2011; 39: 1–13.
22. Wang H, Li B, Leng C. Shrinkage tuning parameter selection with a diverging number of parameters. Journal of Royal Statistical Society, Series B 2009; 71: 671–683.
23. Schwarz C Estimating the dimension of a model. The Annals of Statistics 1978; 6: 461–464.
24. Rand WM. Objective criteria for the evaluation of clustering methods. Journal of the American Statistics Association 1971; 66: 846–850.
25. He K, Schaubel DE. Methods for comparing center-specific survival outcomes using direct standardization. Statistics in Medicine 2014; 33(12): 2048–2061. [PubMed: 24436222]
26. Wang Y, Hong C, Palmer N, et al. A fast divide-and-conquer sparse Cox regression. Biostatistics 2021; 22: 381–401. [PubMed: 31545341]
27. Liu L, Lin L. Subgroup analysis for heterogeneous additive partially linear models and its application to car sales data. Computational Statistics and Data Analysis 2019; 138: 239–259.

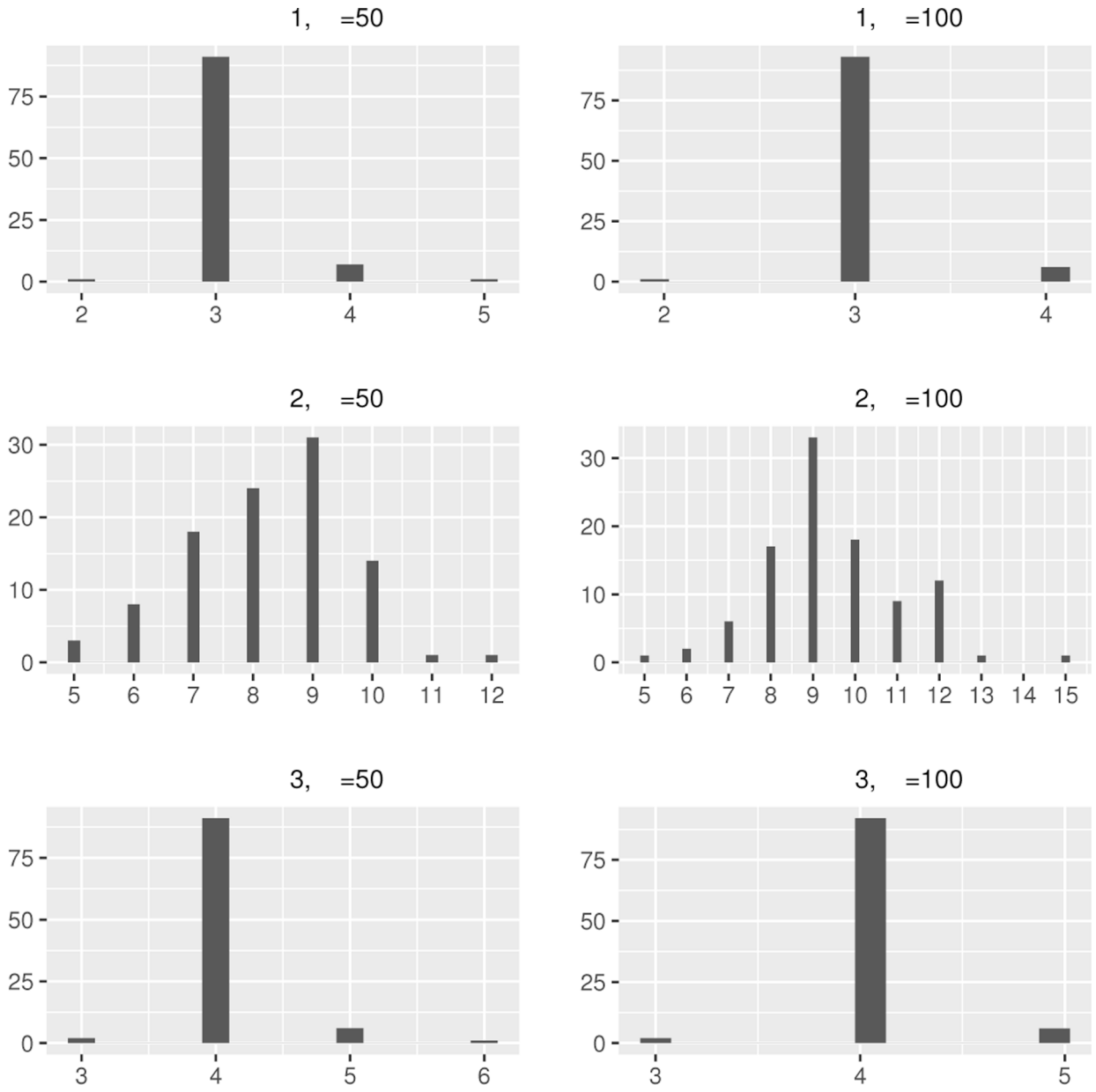

FIGURE 1.
The histograms of $\widehat{K}$ in Example 1–3.

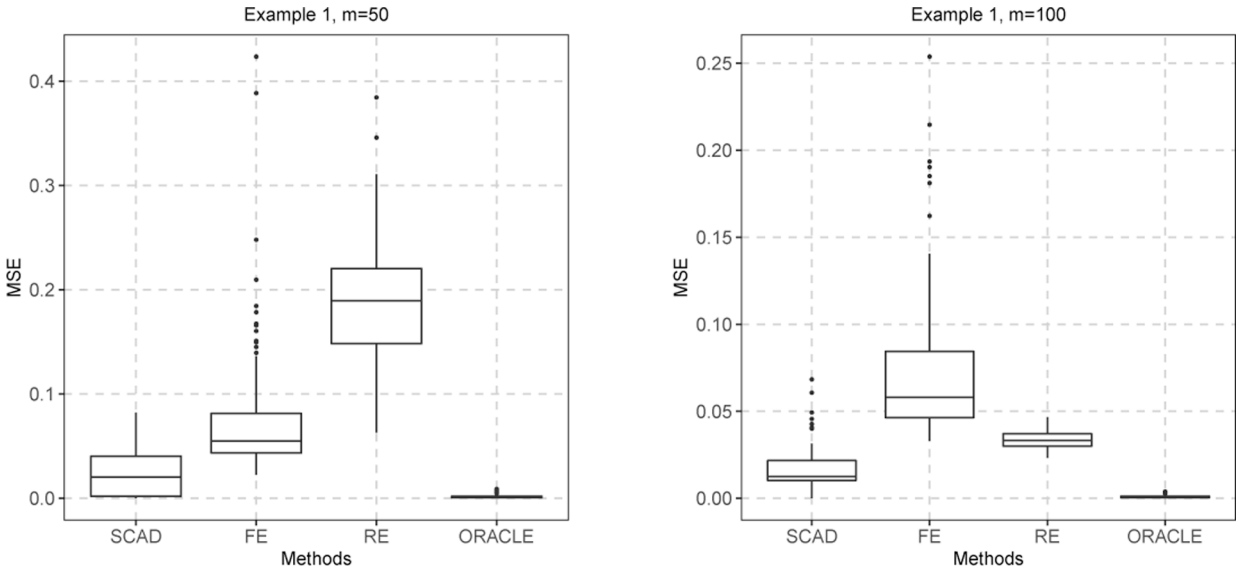

FIGURE 2.
The boxplots of the MSE of $\hat{\boldsymbol{a}}$ with $m = 50$ and $m = 100$ in Example 1.

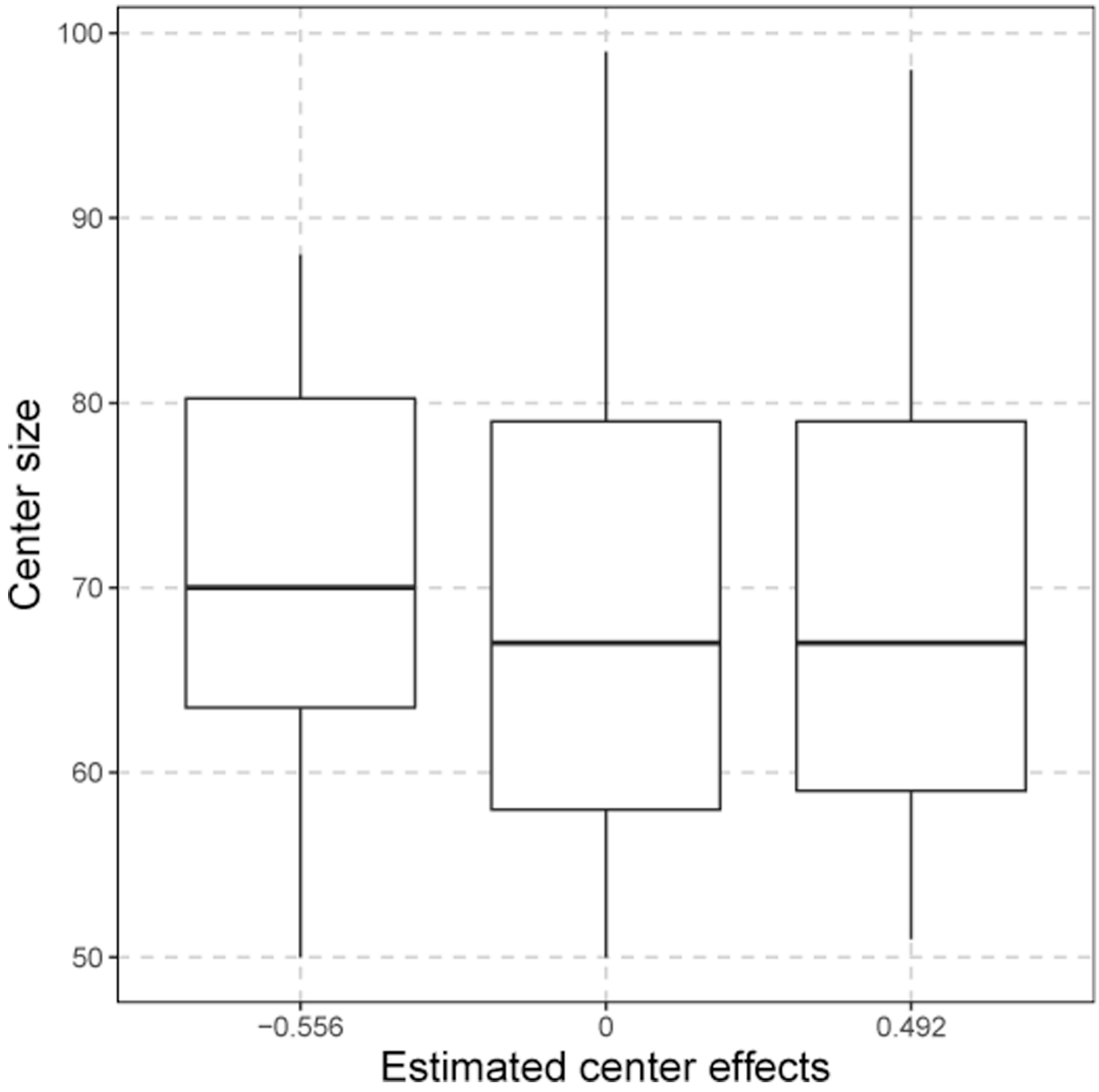

Figure 3.
The boxplots of center size from three groups in the national kidney transplant study.

TABLE 1

The mean, median, and standard deviation (SD) of $\widehat{K}$, the percentage (per) of $\widehat{K}$ equal to the true number of subgroups, and the Rand Index (RI) value by our method with $m = 50,100$ in Examples 1 to 3, respectively.

	m	mean	median	SD	per	RI
Example 1	50	3.08	3.00	0.339	0.91	0.960
	100	3.05	3.00	0.261	0.93	0.969
Example 2	50	8.23	8.00	1.362		
	100	9.43	9.00	1.653		
Example 3	50	4.06	4.00	0.343	0.91	0.963
	100	4.04	4.00	0.281	0.92	0.970

TABLE 2

The bias, standard deviation (SD), standard error (SE), and the coverage probability of 95% confidence intervals (CP) of the estimators in Example 1 and Example 3.

			m=50				m=100			
	m	Method	Bias	SD	SE	CP	Bias	SD	SE	CP
Example 1	$\hat{\alpha}_1$	SCAD	0.014	0.090	0.113	98.9	0.008	0.070	0.078	97.8
		Oracle	0.009	0.094	0.105	96.0	0.018	0.078	0.074	94.0
	$\hat{\alpha}_3$	SCAD	0.028	0.094	0.086	91.5	0.010	0.055	0.061	97.8
		Oracle	0.006	0.094	0.084	92.0	0.006	0.055	0.060	98.0
Example 3	$\hat{\alpha}_1$	SCAD	0.009	0.086	0.087	93.4	0.004	0.063	0.062	97.4
		Oracle	0.004	0.082	0.087	95.0	0.002	0.061	0.062	98.0
	$\hat{\alpha}_3$	SCAD	0.005	0.069	0.089	97.8	0.019	0.058	0.062	94.8
		Oracle	0.004	0.084	0.086	96.0	0.004	0.059	0.061	94.0
	$\hat{\alpha}_4$	SCAD	0.006	0.088	0.085	92.3	0.008	0.057	0.060	93.5
		Oracle	0.009	0.087	0.085	93.0	0.006	0.055	0.060	96.0

TABLE 3

The bias, standard deviation (SD), standard error (SE), and the coverage probability of 95% confidence intervals (CP) of the estimators in Examples 1 to 3.

			$\hat{a}$	$\hat{\beta}_1$				$\hat{\beta}_2$			
		Method	**MSE**	**Bias**	**SD**	**SE**	**CP**	**Bias**	**SD**	**SE**	**CP**
Example 1	50	SCAD	0.022	0.011	0.043	0.045	95.0	0.012	0.043	0.045	95.0
		FE	0.078	0.029	0.045	0.047	91.0	0.026	0.045	0.047	94.0
		RE	0.189	0.005	0.044	0.046	97.0	0.008	0.044	0.046	95.0
		Oracle	0.002	0.003	0.043	0.046	97.0	0.005	0.043	0.046	94.0
	100	SCAD	0.017	0.002	0.034	0.032	93.0	0.004	0.033	0.032	94.0
		FE	0.074	0.034	0.032	0.033	88.0	0.032	0.034	0.033	87.0
		RE	0.034	0.001	0.032	0.033	95.0	0.001	0.033	0.033	94.0
		Oracle	0.001	0.002	0.032	0.032	95.0	0.001	0.033	0.032	93.0
Example 2	50	SCAD	0.321	0.001	0.032	0.030	93.0	0.002	0.031	0.030	94.0
		FE	1.183	0.002	0.033	0.030	93.0	0.002	0.031	0.030	94.0
		RE	1.810	4e-4	0.032	0.030	94.0	0.003	0.031	0.030	94.0
	100	SCAD	0.242	0.002	0.019	0.021	97.0	0.002	0.022	0.021	93.0
		FE	1.080	0.001	0.020	0.021	98.0	0.003	0.022	0.021	93.0
		RE	0.060	0.001	0.019	0.021	98.0	0.001	0.022	0.021	93.0
Example 3	50	SCAD	0.016	0.004	0.047	0.046	94.0	0.007	0.044	0.046	97.0
		FE	0.079	0.031	0.046	0.047	90.0	0.027	0.045	0.047	92.0
		RE	0.847	0.003	0.045	0.046	96.0	0.006	0.044	0.046	96.0
		Oracle	0.003	9e-5	0.044	0.046	96.0	0.003	0.042	0.046	97.0
	100	SCAD	0.012	0.001	0.033	0.032	94.0	1e-4	0.032	0.032	92.0
		FE	0.077	0.035	0.032	0.033	88.0	0.034	0.033	0.032	89.0
		RE	0.039	0.003	0.032	0.033	94.0	0.001	0.033	0.033	92.0
		Oracle	0.001	0.003	0.032	0.032	94.0	0.002	0.033	0.032	92.0

TABLE 4

The estimates (Est.), standard error (SE), and P-value of $\widehat{\boldsymbol{\beta}}$ in the national kidney transplant study.

Variable	Fused			FE			RE		
	Est	SE	P-value	Est	SE	P-value	Est	SE	P-value
Time on ESRD									
< 1 years	Ref								
1–5 years	0.093	0.041	0.022	0.100	0.041	0.015	0.086	0.041	0.035
> 5 years	0.150	0.040	0.000	0.142	0.042	0.001	0.143	0.041	0.001
Donor age									
< 15	−0.063	0.044	0.149	−0.073	0.044	0.098	−0.069	0.044	0.114
15–30	−0.066	0.042	0.113	−0.067	0.042	0.112	−0.075	0.042	0.073
30–45	Ref								
45–60	0.073	0.040	0.068	0.076	0.040	0.061	0.075	0.040	0.062
> 60	0.113	0.043	0.009	0.115	0.044	0.009	0.114	0.043	0.009
Donor race									
White	Ref								
Black	0.066	0.028	0.021	0.072	0.029	0.015	0.071	0.029	0.013
Asian	0.001	0.030	0.987	0.004	0.031	0.908	0.001	0.030	0.969
DON-HTN	0.090	0.035	0.011	0.088	0.036	0.014	0.087	0.035	0.014
DON-EC	−0.003	0.045	0.953	−0.002	0.046	0.973	−0.003	0.046	0.947
Recipient gender	0.048	0.030	0.115	0.053	0.031	0.086	0.053	0.031	0.083
Recipient race									
White	Ref								
Black	0.079	0.030	0.009	0.090	0.034	0.008	0.091	0.032	0.004
Asian	−0.100	0.037	0.007	−0.105	0.038	0.005	−0.099	0.037	0.008
Recipient BMI									
Normal	Ref								
Under	0.054	0.029	0.068	0.050	0.030	0.093	0.054	0.030	0.070
Over	0.005	0.037	0.896	0.003	0.038	0.934	−0.001	0.037	0.971
Obesity	0.026	0.037	0.484	0.024	0.038	0.530	0.027	0.037	0.472
REC-PREV-KI	0.046	0.030	0.127	0.038	0.031	0.220	0.053	0.030	0.084
REC-COLD-ISCH	0.006	0.030	0.837	−0.004	0.033	0.895	0.014	0.031	0.662
REC-AGE									
18–35	0.050	0.034	0.150	0.055	0.035	0.116	0.045	0.035	0.189
35–50	−0.029	0.038	0.439	−0.025	0.039	0.520	−0.029	0.038	0.449
50–60	Ref								
60–70	0.112	0.034	0.001	0.110	0.035	0.001	0.114	0.034	0.001
> 70	0.082	0.030	0.006	0.081	0.030	0.007	0.095	0.030	0.001
REC-DIAB									
Type I	0.056	0.029	0.054	0.061	0.030	0.042	0.057	0.030	0.055
Type II	0.075	0.030	0.012	0.071	0.031	0.021	0.075	0.030	0.014
Donor gender	0.005	0.030	0.872	0.007	0.030	0.818	0.005	0.030	0.863

Variable	Fused			FE			RE		
	Est	SE	P-value	Est	SE	P-value	Est	SE	P-value
Donor BMI									
Normal	Ref								
Under	0.104	0.037	0.005	0.108	0.037	0.004	0.103	0.037	0.005
Over	0.037	0.035	0.292	0.036	0.035	0.300	0.034	0.035	0.329
Obesity	−0.008	0.035	0.815	−0.009	0.035	0.798	−0.003	0.035	0.927

TABLE 5

Results of $\widehat{\alpha}_i$ in the national kidney transplant study.

	$\widehat{\alpha}_1$	$\widehat{\alpha}_2$	$\widehat{\alpha}_3$
Est	−0.556	ref	0.492
num	6	45	9
SE	0.129	ref	0.073
P-value	<0.001	ref	<0.001
SMR	0.574	1.000	1.636