RESEARCH ARTICLE

WILEY Statistics in Medicine

Computationally efficient inference for center effects based on restricted mean survival time

Xin Wang[1,2] | Yingchao Zhong[1] | Purna Mukhopadhyay[3] | Douglas E. Schaubel[1,4]

[1]Department of Biostatistics, University of Michigan, Ann Arbor, Michigan

[2]Vertex Pharmaceuticals, Boston, Massachusetts

[3]Arbor Research Collaborative for Health, Ann Arbor, Michigan

[4]Department of Biostatistics, Epidemiology, and Informatics, University of Pennsylvania, Philadelphia, Pennsylvania

Correspondence
Douglas E. Schaubel, Department of Biostatistics, Epidemiology, and Informatics University of Pennsylvania Blockley Hall Philadelphia, PA 19104.
Email: douglas.schaubel@pennmedicine.upenn.edu

Restricted mean survival time (RMST) has gained increased attention in biostatistical and clinical studies. Directly modeling RMST (as opposed to modeling then transforming the hazard function) is appealing computationally and in terms of interpreting covariate effects. We propose computationally convenient methods for evaluating center effects based on RMST. A multiplicative model for the RMST is assumed. Estimation proceeds through an algorithm analogous to stratification, which permits the evaluation of thousands of centers. We derive the asymptotic properties of the proposed estimators and evaluate finite sample performance through simulation. We demonstrate that considerable decreases in computational burden are achievable through the proposed methods, in terms of both storage requirements and run time. The methods are applied to evaluate more than 5000 US dialysis facilities using data from a national end-stage renal disease registry.

KEYWORDS
censored data, center effect, facility profiling, failure time, restricted mean survival time

1 | INTRODUCTION

Restricted mean survival time (RMST) is often of great clinical interest in practice and is gaining increased attention among biostatisticians. There are now several existing methods to model RMST, with the methods distinguished by their estimation approaches and assumptions on the censoring mechanism.[1-8] Compared to proceeding indirectly by transforming other pivotal functions into RMST, directly modeling RMST has the appeal of straightforward parameter interpretation and computational convenience.

We propose computationally convenient methods for evaluating center effects based on RMST. A multiplicative model for the RMST is assumed. In the interest of robustness and flexibility, we avoid making any distributional assumptions on the underlying survival time. Estimation is based on generalized estimating equations (GEE). Numerical techniques analogous to stratification permit the simultaneous evaluation of thousands of center effects.

Distinguishing features of the proposed (relative to existing related) methods include the ability to conveniently estimate a very large number (eg, thousands) of covariate-adjusted center effects. Our methods are motivated by an analysis of the national end-stage renal disease (ESRD) database, with our objectives being to (a) estimate center effects for each of the >5000 dialysis facilities in the United States and (b) estimate the effects (on the RMST scale) of many patient-level covariates. As will be described, the numerical techniques we propose for parameter estimation in this context are quite different than those in previous related works.[6,8] Specifically, we propose a two-stage approach that involves estimating a regression parameter corresponding to the adjustment covariates at the first stage, with the center effects then estimated at the second stage. Note that neither the work of Tian et al[6] nor the work of Wang and Schaubel[8] were developed for

facility profiling. Their application in this setting would be through a one-stage approach that estimated all parameters simultaneously and, hence, could be computationally prohibitive.

Although parameter estimation for the RMST model is generally convenient, computational difficulties may arise if the dimension of the covariate vector is quite large. Standard software packages (eg, R and SAS) typically handle data sets with tens of thousands of subjects better than they handle several hundred covariates. Examples include clustered data or data with a high-dimensional covariate. In this report, the terms "cluster," "facility," and "center" could be used interchangeably. In the interest of concreteness, we use the term "center" hereafter, in part due to its connection with our motivating example. A conventional way to adjust for fixed facility effects in a regression model is to code potentially very large number of center indicators; this can introduce a high covariate dimension, in turn, greatly increasing the computational burden and, in some cases, preventing model convergence.

The case we consider in this report involves data characterized by a large number of centers and evaluated in an environment where fixed center effects are desired.[9] The data set that motivates our proposed methods consists of ESRD patients receiving hemodialysis in the United States. The study population has more than 1 million patients from more than 5000 dialysis facilities. Computational demands are underscored by the requirement of 50 GB of memory under conventional methods. We are interested in evaluating the effect on RMST of some variables historically reported to be important prognostic factors, including age, race, gender, underlying diagnosis, and comorbidity information. However, facility has frequently been reported to be a strong predictor of ESRD patient survival. Moreover, it is generally acknowledged that the prognostic factor distributions may differ markedly by facility. Such considerations are not unique to ESRD; in the presence of data on subjects treated at different centers, great potential for bias exists if the regression model serving as the basis for covariate adjustment does not adjust for facility.[10]

It should be noted that the purpose of this report is not to contrast fixed versus random effects as methods for adjusting for center. Strong cases can be made in either direction, such that the choice of one over the other depends largely on the data structure at hand and analytic objectives. In our case, covariate effects are of interest, but of equal (or more) interest are the facility effects. Such results are often of great interest to various stakeholders (eg, regulatory bodies, oversight committees, insurers). In this particular framework, one reason to prefer fixed versus random facility effects is that a key assumption of the random effects model (ie, the independence of facility effect and patient characteristics) is unlikely to hold. Fixed cluster effects are arguably more appropriate when we suspect the individual covariates are correlated with facility effects. In addition, fixed effect methods have been demonstrated to yield less biased estimators with smaller mean squared error when the true facility effect is far from the average across facilities.[9,11]

In the context of ESRD data, conventional methods would involve simultaneous estimation of the parameter vector corresponding to the covariate vector and the parameters corresponding to the $\approx$ 5000 facility indicators. As will be demonstrated, the proposed techniques separate the estimation of center specific baseline RMST from the estimation of covariate effects. As detailed in Section 2, the techniques we develop can exploit standard software for implementation, a desirable property for practitioners. The proposed methods yield much faster run times relative to those typically employed in the generalized linear models (GLMs) setting. In particular, we dissect the model structure and connect it to the estimation procedure of the stratified proportional hazards model.[12,13]

The novelty of the methods proposed in this report is primarily from two perspectives. First, to the best of our knowledge, no previous work has proposed methods for estimating center effects through a direct model of RMST. Second, no previous report has addressed computational issues likely to arise for RMST modeling in the presence of large data sets. We focus on center effects (as a frequently occurring instance of high-dimensional categorical covariates) and propose techniques that greatly reduce storage space and computing time and, hence, are amenable to very large databases.

The remainder of this report is organized as follows. In Section 2, we formulate the data structure, describe the proposed methods, and then describe the proposed estimation procedures. Large sample properties are derived in Section 3, with numerical studies conducted in Section 4 in order to assess the accuracy of the proposed procedures in finite samples. We illustrate our methods in Section 5 through application to the aforedescribed ESRD database. Discussion and possible future directions are presented in Section 6. R code for carrying out the proposed methods is provided in the Supporting Information document corresponding to this article.

2 | PROPOSED METHODS

We now describe the proposed methods, beginning with the requisite notation.

2.1 | Notation and assumptions

Let i denote the ith patient ($i = 1, \ldots, n$) and let g_i denote this patient's center, where $g_i = 1, \ldots, J$ and J is usually a relatively large number (eg, $J = 1000$). To simplify the notation, we create a vector $\boldsymbol{G}_i = (G_{i1}, \ldots, G_{iJ})'$, where $G_{ij} = I(g_i = j)$ and $I(\cdot)$ represents an indicator variable taking the value 1 when its argument is true and 0 otherwise. That is, $\boldsymbol{G}_i$ will have all elements equal to 0 except the g_ith element (which equals 1). Baseline covariates are denoted by $\boldsymbol{Z}_i$, a vector of length p. Let D_i denote the time of death, which is subject to right censoring time C_i. Due to the occurrence of censoring, we observe the minimum of death and censoring time, $X_i = D_i \wedge C_i$, and hence, we define the death indicator $\Delta_i^D = I(D_i \leq C_i)$. Suppose L is the pre-specified truncation time point of interest (eg, 5 years, 10 years, etc). We then define the truncated survival time and its corresponding indicator by $Y_i = D_i \wedge L$ and $\Delta_i^Y = I(D_i \wedge L \leq C_i)$, respectively. Our observed data are then $\mathcal{O} = \{\mathcal{O}_i; \ i = 1, \ldots, n\}$, where $\mathcal{O}_i = \{\boldsymbol{Z}_i, \boldsymbol{G}_i, X_i, Y_i, \Delta_i^Y, \Delta_i^D\}$.

We are interested in average survival up to time L, and the model of interest can be represented in a general sense by $\mu_{ij} = E(D_i \wedge L | \boldsymbol{Z}_i, g_i = j)$. Since our intention is to develop a useful tool to evaluate survival based on the information available at the time origin, we model the RMST as a function of baseline (time 0) covariates. We assume the following model for the L-year RMST:

$$\mu_{ij} = \mu_{0j} \exp\left\{\boldsymbol{\beta}_0' \boldsymbol{Z}_i\right\}, \tag{1}$$

where $\boldsymbol{\beta}_0 = (\beta_{01}, \ldots, \beta_{0p})'$ is the covariate effect of interest, and $\boldsymbol{\mu}_0 = (\mu_{01}, \ldots, \mu_{0J})'$ is the center-specific baseline RMST. Model (1) has the same structure as a GLM with a log link. However, note that the variance structure is unspecified. The model is equivalent to a model with centers represented by J indicator variables; ie, $\exp\{\boldsymbol{\beta}_0' \boldsymbol{Z}_i + \boldsymbol{G}_i' \log(\boldsymbol{\mu}_0)\}$. For the data structure of interest in this report, J is usually a large number, such that fitting the model requires careful consideration to avoid computational difficulties. In order to avoid estimating the J center effects simultaneously, we propose a two-stage procedure that allows us to separately estimate $\boldsymbol{\beta}_0$ and $\boldsymbol{\mu}_0$. Note that the dependence of μ_{ij} on L is suppressed throughout this article.

2.2 | Censoring models

In the absence of censoring, $E[G_{ij}\boldsymbol{Z}_i\{D_i \wedge L - \mu_{0j} \exp(\boldsymbol{\beta}_0'\boldsymbol{Z}_i)\}] = \boldsymbol{0}$. This can serve as the basis for constructing estimating equations, but requires modification in the presence of censoring. To accommodate censoring, we employ a variant of Inverse Probability of Censoring Weighting (IPCW).[14,15] In our context, IPCW reweights the uncensored ($D_i \wedge L$) values, such that the weighted data represent the ($D_i \wedge L$) distribution of the target population. We allow the censoring distribution to depend on the baseline covariates and to differ across centers. Note that covariate-dependent censoring is quite common. For example, consider an observational study, conducted during a fixed calendar period, with staggered entry. In this case, subjects who enter later in the observation window would have a different censoring distribution than those who enter earlier. Covariate-dependent censoring would result if survival time also depended on calendar time (eg, survival improved, due to therapeutic advances).

Denote the hazard function for censoring time C_i by $\lambda_{ij}^C(t)$ for patient i from center j; specifically, $\lambda_{ij}^C(t) = \lim_{h\to 0} h^{-1} P(X_i \in [t, t+h), \Delta_i^D = 0 | X_i \geq t, \boldsymbol{Z}_i, g_i = j)$. Denote the corresponding cumulative hazard by $\Lambda_{ij}^C(t) = \int_0^t \lambda_{ij}^C(u) du$. Defining the IPCW weight by $W_{ij} = \exp\{\Lambda_{ij}^C(Y_i)\}$, it can be shown that

$$E\left[G_{ij}\Delta_i^Y W_{ij} \boldsymbol{Z}_i \left(Y_i - \mu_{0j} \exp\left\{\boldsymbol{\beta}_0' \boldsymbol{Z}_i\right\}\right)\right] = \boldsymbol{0}. \tag{2}$$

In practice, $\lambda_{ij}^C(t)$ is rarely known and needs to be estimated from the observed data. For this purpose, we assume the following Cox model for censoring:

$$\lambda_{ij}^C(t) = \lambda_{0j}^C(t) \exp\left\{\boldsymbol{\theta}_0' \boldsymbol{Z}_i\right\}, \tag{3}$$

which can be fitted by reversing the "event/censoring" roles of D and C.

The use of Cox regression is well established in the context of IPCW, and the above censoring assumption can easily accommodate both covariate-independent and covariate-dependent censoring. After estimating $\widehat{\boldsymbol{\theta}}$ and $\widehat{\Lambda}_{0j}^C(t)$ through standard partial likelihood[16] and Breslow[17] estimators, respectively, we can estimate the IPCW weights, $W_{ij} = P(C_i > Y_i | \boldsymbol{Z}_i, G_i = j, Y_i)^{-1}$, by $\widehat{W}_{ij} = \exp\{\exp(\widehat{\boldsymbol{\theta}}' \boldsymbol{Z}_i)\widehat{\Lambda}_{0j}^C(Y_i)\}$. We then substitute the estimated weights $\widehat{\boldsymbol{W}} = (\widehat{W}_1, \ldots, \widehat{W}_n)'$ in place of their corresponding true values $\boldsymbol{W} = (W_1, \ldots, W_n)'$.

Several notes regarding model (3) are in order at this juncture. First, note that dependence of the C distribution on center is accommodated through the center-stratified baseline hazard. Second, generally, it would be preferable to use all data to estimate model (3). However, doing so assumes proportionality across $(0, \tau]$, which is more restrictive than a

proportionality assumption across $(0, L]$. Therefore, if proportionality is violated and if $L \ll \tau$ (eg, $L = 1$ year, $\tau = 5$ years), then it may be preferable to artificially censor subjects at L for the purposes of estimating model (3). We return to the discussion of the censoring distribution in Section 6.

2.3 | Estimating equations

Based on the zero-mean property given in (2), we construct the following estimating equations:

$$\sum_{j=1}^{J} \sum_{i=1}^{n} G_{ij} W_{ij} \Delta_i^Y Z_i (Y_i - \mu_{ij}) = \mathbf{0}$$

$$\sum_{i=1}^{n} G_{ij} W_{ij} \Delta_i^Y (Y_i - \mu_{ij}) = 0, \quad j = 1, \dots, J.$$

Substituting $\widehat{W}$ for W, we can estimate β_0 and μ_0 from following $p + J$ working estimating equations:

$$\sum_{j=1}^{J} \sum_{i=1}^{n} G_{ij} \widehat{W}_{ij} \Delta_i^Y Z_i \left(Y_i - \mu_{0j} \exp \left\{ (\beta_0' Z_i \right\} \right) = \mathbf{0}, \tag{4}$$

$$\sum_{i=1}^{n} G_{ij} \widehat{W}_{ij} \Delta_i^Y \left(Y_i - \mu_{0j} \exp \left\{ \beta_0' Z_i \right\} \right) = 0, \quad j = 1, \dots, J. \tag{5}$$

Solving (4) and (5) simultaneously implies simultaneous estimation of $p + J$ parameters, which is subject to numerical instability when J is quite large. Instead, we propose estimating β_0 first through iteration and then estimating μ_0 through J separate closed-form expressions. Along these lines, we define

$$S_j^{(k)}(\beta, W) = \frac{\sum_{i=1}^{n} G_{ij} W_{ij} \Delta_i^Y \exp\{\beta' Z_i\} Z_i^{\otimes k}}{\sum_{i=1}^{n} G_{ij}}, \tag{6}$$

$$\overline{S}_j(\beta, W) = \frac{S_j^{(1)}(\beta, W)}{S_j^{(0)}(\beta, W)}, \tag{7}$$

for $j = 1, \dots, J$ and $k = 0, 1, 2$. Note that, for a vector a, we define $a^{\otimes 0} = 1$, $a^{\otimes 1} = a$, and $a^{\otimes 2} = a' a$. Using (6) and (7), we can rewrite estimating equations (4)-(5) as follows:

$$\sum_{j=1}^{J} \sum_{i=1}^{n} G_{ij} \{ Z_i - \overline{S}_j(\beta, \widehat{W}) \} \widehat{W}_{ij} \Delta_i^Y Y_i = \mathbf{0}, \tag{8}$$

$$\mu_{0j} = \frac{\sum_{i=1}^{n} G_{ij} \widehat{W}_{ij} \Delta_i^Y Y_i}{\sum_{i=1}^{n} G_{ij} \widehat{W}_{ij} \Delta_i^Y \exp\{\beta' Z_i\}}, \quad j = 1, \dots, J. \tag{9}$$

The algebra underlying the equivalence of (4)-(5) and (8)-(9) is provided in Section 2.4. Note that (8) is free of the center-specific parameters μ_0 and that (9) is a closed-form calculation of μ_0, allowing us to separately estimate β_0 and μ_0.

2.4 | Equivalence of systems (8)-(9) and (4)-(5)

Solving (5) yields

$$\mu_{0j} = \frac{\sum_{i=1}^{n} G_{ij} \widehat{W}_{ij} \Delta_i^Y Y_i}{\sum_{i=1}^{n} G_{ij} \widehat{W}_{ij} \Delta_i^Y \exp\{\beta' Z_i\}}, j = 1, \dots, J,$$

which, for fixed $\boldsymbol{\beta}$, is equal to (9). Then, substituting this expression for μ_{0j} into (4) yields

$$
\begin{aligned}
&= \sum_{j=1}^{J} \sum_{i=1}^{n} G_{ij} \widehat{W}_{ij} \Delta_i^Y Y_i \boldsymbol{Z}_i - \sum_{j=1}^{J} \sum_{i=1}^{n} G_{ij} \widehat{W}_{ij} \Delta_i^Y \exp\{\boldsymbol{\beta}' \boldsymbol{Z}_i\} \boldsymbol{Z}_i \frac{\sum_{k=1}^{n} G_{kj} \widehat{W}_{kj} \Delta_k^Y Y_k}{\sum_{k=1}^{n} G_{kj} \widehat{W}_{kj} \Delta_k^Y \exp\{\boldsymbol{\beta}' \boldsymbol{Z}_k\}} \\
&= \sum_{j=1}^{J} \sum_{i=1}^{n} G_{ij} \widehat{W}_{ij} \Delta_i^Y Y_i \boldsymbol{S}_i - \sum_{j=1}^{J} \sum_{k=1}^{n} G_{kj} \widehat{W}_{kj} \Delta_k^Y Y_k \frac{\boldsymbol{S}_j^{(1)}(\boldsymbol{\beta}, \widehat{\boldsymbol{W}})}{S_j^{(0)}(\boldsymbol{\beta}, \widehat{\boldsymbol{W}})} \\
&= \sum_{j=1}^{J} \sum_{i=1}^{n} G_{ij} \widehat{W}_{ij} \Delta_i^Y Y_i \left\{ \boldsymbol{S}_i - \bar{\boldsymbol{s}}_j(\boldsymbol{\beta}, \widehat{\boldsymbol{W}}) \right\},
\end{aligned}
$$

which is equal to (8).

2.5 | Fitting the proposed model using Cox regression software

Exploiting the connections between the proposed estimating equations and proportional hazards regression, we now describe an algorithm for estimating parameters from model (1) using Cox regression software. The steps are as follows.

(i) Estimate the censoring hazard $\widehat{\Lambda}_{ij}^C(t)$ from model (3) by unweighted partial likelihood and the Breslow estimator. Construct IPCW weights by $\widehat{W}_{ij} = \exp\{\widehat{\Lambda}_{ij}^C(Y_i)\}$ for $i = 1, \ldots, n$.

(ii) Create a data set where each subject i has a row containing $\{\boldsymbol{Z}_i, g_i, \widehat{W}_{ij}^\dagger, \Delta_i^1, X_i^1, \text{Off}_i\}$, where $\widehat{W}_{ij}^\dagger = \widehat{W}_{ij} \Delta_i^Y Y_i$, $\Delta_i^1 = 1$, $X_i^1 = 1$ and $\text{Off}_i = -\log(Y_i)$.

(iii) Fit a stratified inverse-weighted Cox model to the data set created in Step (ii), with g_i serving as strata, covariate $\boldsymbol{Z}_i$, weight $\widehat{W}_{ij}^\dagger$, token "follow-up time" X_i^1 (set to 1 for all i) and "death indicator" Δ_i^1 (also set to 1 for all i), and offset Off_i. Note that ties should be handled by Breslow option.

We justify this algorithm algebraically in Appendix, although some immediate comments are in order. The inverse weighting in Step (i) is implied by Section 2. The use of a "token" follow-up time (equal for all subjects) and death indicator reconcile the software's expectation of an integration over time with the property that (8) and (9) invoke one contribution per subject (hence, with no associated integral). Note that the follow-up time could be any nonnegative number, so long as it is set equal for all subjects. Step (iii) can be implemented by several statistical software packages. It has been tested by the authors in R and SAS.

The algorithm is quite fast, even for very large data sets, owing to the stratification. The resulting coefficient and baseline hazard serve as our proposed estimators $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\mu}}$.

Note that the procedure described by Steps (i), (ii), and (iii) above treats the $\widehat{W}_{ij}$ as fixed. We outline a justification for this shortcut in Section 3, between Theorem 1 and Theorem 2.

2.6 | Center effects

Note that $\{\mu_{01}, \ldots \mu_{0J}\}$ represent the center-specific baseline RMST and, analogous to center-specific intercepts, do not represent center-specific contrasts. For settings where contrasts between centers are of interest, we propose the following rescaling:

$$
\eta_j = \frac{\mu_{0j}}{\boldsymbol{w}' \boldsymbol{\mu}_0} \quad j = 1, \ldots, J,
$$

where $\boldsymbol{w} = (w_1, \ldots, w_J)'$ is a pre-specified weight vector with $\boldsymbol{w}'\boldsymbol{1} = 1$. An example of $\boldsymbol{w}$ would be $\boldsymbol{w} = (1, \ldots, 1)'/J$, ie, equal weight across all J centers. The rescaled $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_J)$ represents a vector of covariate-adjusted contrasts, with element j contrasting center j with the overall weighted average center. Note that the weighted average of the contrasts is equal to 1 (ie, $\boldsymbol{w}'\widehat{\boldsymbol{\eta}} = 1$), which is a desirable property for interpretation purposes.

3 | ASYMPTOTIC PROPERTIES

Regularity conditions are as specified in the Supplementary Information document. These conditions can be relaxed at the expense of additional technical development.

Our main asymptotic results are summarized in the following three theorems, with proofs for each presented in Appendix.

Theorem 1. *Under regularity conditions (a)-(g), as $n \to \infty$, $\widehat{\boldsymbol{\beta}}$ converges in probability to $\boldsymbol{\beta}_0$ and $n^{1/2}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ converges to a zero-mean Normal with variance $\boldsymbol{A}(\boldsymbol{\beta}_0)^{-1}\boldsymbol{B}(\boldsymbol{\beta}_0)\boldsymbol{A}(\boldsymbol{\beta}_0)^{-1}$ with $\boldsymbol{A}(\boldsymbol{\beta}_0)$ and $\boldsymbol{B}(\boldsymbol{\beta}_0)$ defined as follows:*

$$\boldsymbol{A}(\boldsymbol{\beta}) = \sum_{j=1}^{J} E\left\{ G_{ij}\Delta_i^Y W_{ij}Y_i\left(\frac{\boldsymbol{s}_j^{(2)}(\boldsymbol{\beta})}{s_j^{(0)}(\boldsymbol{\beta})} - \bar{\boldsymbol{s}}_j(\boldsymbol{\beta})^{\otimes 2} \right) \right\} \quad \boldsymbol{B}(\boldsymbol{\beta},\boldsymbol{W}) = E\left[\sum_{j=1}^{J} G_{ij}\left\{ \boldsymbol{b}_{ij}(\boldsymbol{\beta},\boldsymbol{W}) \right\}^{\otimes 2} \right],$$

where

$$\boldsymbol{b}_{ij}(\boldsymbol{\beta},\boldsymbol{W}) = \left\{ \boldsymbol{Z}_i - \bar{\boldsymbol{S}}_j(\boldsymbol{\beta},\boldsymbol{W}) \right\} W_{ij}\Delta_i^Y\left(Y_i - \mu_{ij} \right) + \boldsymbol{K}(\boldsymbol{\beta},\boldsymbol{\theta},\boldsymbol{W})\boldsymbol{\Theta}(\boldsymbol{\theta})^{-1}\boldsymbol{U}_{ij}(\boldsymbol{\theta})$$

$$+ \int_0^L \frac{\boldsymbol{H}_j(u;\boldsymbol{\beta},\boldsymbol{\theta},\boldsymbol{W})}{r_j^{(0)}(u;\boldsymbol{\theta})} dM_{ij}^C(u),$$

$$\boldsymbol{K}(\boldsymbol{\beta},\boldsymbol{\theta},\boldsymbol{W}) = \sum_{j=1}^{J} E\left\{ G_{ij}\boldsymbol{\epsilon}_i(\boldsymbol{\beta},\boldsymbol{W})\boldsymbol{D}_i(\boldsymbol{\theta})' \right\},$$

$$\boldsymbol{U}_{ij}(\boldsymbol{\theta}) = \sum_{j=1}^{J} G_{ij}\int_0^{\tau}\left\{ \boldsymbol{Z}_i - \bar{\boldsymbol{r}}_j(u;\boldsymbol{\theta}) \right\} dM_{ij}^C(u),$$

$$\boldsymbol{H}_j(t;\boldsymbol{\beta},\boldsymbol{\theta},\boldsymbol{W}) = E\left[\boldsymbol{\epsilon}_i(\boldsymbol{\beta},\boldsymbol{W})\exp\{\boldsymbol{\theta}'\boldsymbol{Z}_i\}R_i(t) \right],$$

with $M_{ij}^C(t) = G_{ij}\{N_i^C(t) - \int_0^t R_i(u)\exp(\boldsymbol{\theta}'\boldsymbol{Z}_i)\lambda_{0j}^C(u)du\}$ as the censoring martingale and where $N_i^C(t) = I(X_i \leq t, \Delta_i^D = 0)$ represents the censoring counting process.

The consistency of $\widehat{\boldsymbol{\beta}}$ holds by the Inverse Function Theorem,[18] while the proof of asymptotic normality follows through the combination of various Taylor series expansions and the Cramer-Wold Theorem. This sandwich variance with $\boldsymbol{B}$ as the middle matrix treats IPCW weights as estimated from the data, which well reflects the reality. However, the calculation of this variance could be complicated. A useful short cut involves replacing the middle matrix $\boldsymbol{B}$ with

$$\boldsymbol{B}^*(\boldsymbol{\beta},\boldsymbol{W}) = E\left[\sum_{j=1}^{J} G_{ij}\left\{ \boldsymbol{b}_{ij}^*(\boldsymbol{\beta},\boldsymbol{W}) \right\}^{\otimes 2} \right],$$

where $\boldsymbol{b}_{ij}^*(\boldsymbol{\beta},\boldsymbol{W}) = \{\boldsymbol{Z}_i - \bar{\boldsymbol{S}}_j(\boldsymbol{\beta},\boldsymbol{W})\}W_{ij}\Delta_i^Y(Y_i - \mu_{ij})$ is the first and primary component of the original $\boldsymbol{b}_{ij}(\boldsymbol{\beta},\boldsymbol{W})$. This short cut treats the IPCW weights as fixed rather than estimated. Although it does not fully reflect the actual estimating procedure, this short cut is much easier to calculate and should serve as a useful substitute for the more complicated variance estimator implied by Theorem 1, particularly since the primary source of variation is still captured. We evaluate this variance estimator through simulation in Section 4.

Theorem 2. *Under regularity conditions (a)-(g), as $n \to \infty$, $\widehat{\boldsymbol{\mu}}_0$ converges in probability to $\boldsymbol{\mu}_0$ and $n^{1/2}(\widehat{\boldsymbol{\mu}}_0 - \boldsymbol{\mu}_0)$ converges to a zero-mean Normal with variance $\boldsymbol{V}_\mu$, where*

$$\boldsymbol{V}_\mu = E\left[\left\{ \begin{pmatrix} \rho_1^{-1}s_1^{(0)}(\boldsymbol{\beta})^{-1}G_{ij}W_{i1}\Delta_i^Y(Y_i - \mu_{ij}) \\ \vdots \\ \rho_J^{-1}s_J^{(0)}(\boldsymbol{\beta})^{-1}G_{ij}W_{iJ}\Delta_i^Y(Y_i - \mu_{ij}) \end{pmatrix} - \sum_{j=1}^{J} G_{ij}\begin{pmatrix} \mu_{01} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \mu_{0J} \end{pmatrix}\begin{pmatrix} \bar{\boldsymbol{s}}_1(\boldsymbol{\beta})' \\ \vdots \\ \bar{\boldsymbol{s}}_J(\boldsymbol{\beta})' \end{pmatrix}\boldsymbol{b}_{ij}(\boldsymbol{\beta},\boldsymbol{W}) \right\}^{\otimes 2} \right],$$

where $\rho_j = P(g_i = j)$, a marginal probability which can be estimated by $\widehat{\rho}_j = n_j/n$.

Theorem 3. *Under regularity conditions (a)-(g), as $n \to \infty$, $\widehat{\boldsymbol{\eta}}$ converges in probability to $\boldsymbol{\eta}_0$ and $n^{1/2}(\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0)$ converges to a zero-mean Normal with variance $\boldsymbol{V}_\eta$, where*

$$\boldsymbol{V}_\eta = \left(\boldsymbol{\mu}_0' \boldsymbol{w}\right)^{-4} \left(\boldsymbol{\mu}_0' \boldsymbol{w} \boldsymbol{I}_J - \boldsymbol{\mu}_0 \boldsymbol{w}'\right) \boldsymbol{V}_\mu \left(\boldsymbol{\mu}_0' \boldsymbol{w} \boldsymbol{I}_J - \boldsymbol{w} \boldsymbol{\mu}_0'\right),$$

and with $\boldsymbol{I}_J$ denoting a $J \times J$ identity matrix.

The proofs of Theorem 2 and Theorem 3 proceed by applying the Delta Method to the results of Theorem 1.

4 | SIMULATION STUDY

In this section, we first evaluate the finite sample properties of the proposed methods. We then compare the proposed method to a conventional one-stage method of estimating the center effects and regression parameter simultaneously. The conventional methods, which create 0/1 indicator variates for each center, then solve estimating equations (4)-(5) simultaneously; this can be implemented in R using the package geepack.[8]

We generated the number of patients across $J = 50$ centers from a multinomial distribution with equal weights $1/J$ and the total sample size n. Four total sample sizes are tested: $n = 2500$, $n = 5000$, $n = 10\,000$, and $n = 20\,000$. Death times were generated from an Exponential with mean $\{\mu_{0j}^\dagger \exp(\beta_1^\dagger Z_{1i} + \beta_2^\dagger Z_{2i})\}^{-1}$, where Z_{1i} and Z_{2i} each follow independent Normal(0, 1) distributions. We set $\beta_1^\dagger = 0.5$, $\beta_2^\dagger = 1$ and let $\mu_{01}^\dagger, \ldots, \mu_{0J}^\dagger$ range from 0.158 to 0.550 with an equal increment. The true parameter values corresponding to $E(D_i \wedge L | \boldsymbol{Z}_i, g_i)$ were determined empirically based on a sample size of 10 million. The censoring time also followed an Exponential distribution with hazard $\lambda_{0j}^C \exp(\theta_1 Z_{1i} + \theta_2 Z_{3i})$, where two censoring patterns are tested, resulting in $\approx 15\%$ and $\approx 30\%$ censoring. The first censoring pattern uses $\boldsymbol{\theta} = (0.4, 0.1)'$ and with $\lambda_{01}^C, \ldots, \lambda_{0J}^C$ ranging from 0.0108 to 0.05 with an equal increment. The second censoring pattern uses $\boldsymbol{\theta} = (0.5, -0.5)'$ and sets $\lambda_{01}^C, \ldots, \lambda_{0J}^C$ to range from 0.712 to 0.810 with an equal increment. The specification that centers with higher (lower) death rate have higher (lower) censoring probability serves to balance the percent censored across centers. The performance of the proposed methods is evaluated at two different truncation points: $L = 1.8$ and $L = 5.4$, which represent approximately the 50th and 75th percentiles, respectively, of the death time distribution.

4.1 | Finite-sample properties of the proposed methods

For brevity, we present the simulation results for $L = 1.8$ and $L = 5.4$. As shown in Table 1, the magnitude of the bias decreases generally as sample size increases. Average standard error (ASE) is calculated using the aforementioned shortcut (which treats the estimated $\widehat{\boldsymbol{W}}$ as known) and is on average very close to empirical standard deviation (ESD). The coverage probability (CP) corresponding to ASE is quite close to 95%, except in a few scenarios under the heavier censoring setup. We omit the simulation results corresponding to the standard error estimator derived from Theorem 1, since the results are very similar to those presented for the shortcut formula.

Figure 1 show plots of the ASE versus ESD with respect to the rescaled center effects, $\widehat{\boldsymbol{\eta}}$, under light and heavy censoring for $L = 1.8$ and $L = 5.4$. As sample size increases, the ASEs more closely approximate the corresponding ESDs. Note that the ASE tends to overestimate the ESD; this is likely due to having treated the estimate weights as fixed.[19]

4.2 | Comparison between proposed and conventional methods

In the Supporting Information document, we present results for the conventional method with respect to the estimated regression parameter and rescaled center effects. Performance is very similar to the proposed methods. As shown in Figure 2, the conventional method of estimating the rescaled center effects has very similar performance to the proposed methods.

To illustrate the difference in the run time between our proposed methods and a standard weighted GLM approach (which would simultaneously estimate the covariate and center parameters), we choose $L = 1.8$ under the first censoring setup aforedescribed. Run times are presented for $J = 25, 50, 100, 200, 400, 600, 800, 1000$ and on average 50, 60, 70, 80, 90, 100 patients per center, respectively. Each run time is calculated using the average across 10 replicates. Relative to the proposed methods, the conventional method runs much slower and result in approximately a 10-fold to 3000-fold increase in run time required to estimate the model. This is depicted in Figure 2, when number of centers ranges from $J = 50$ to $J = 500$. The savings in computation time offered by our proposed algorithm increases rapidly with increasing J and also with increasing average number of patients per center.

TABLE 1 Simulation results: $L = 1.8$ and $L = 5.4$ under light and heavy censoring

L	Censoring	Parameter	n	BIAS	ESD	ASE	CP (%)
1.8	≈15%	$\beta_1 = -0.132$	2500	0.002	0.010	0.009	94
			5000	0.001	0.007	0.007	93
			10 000	0.001	0.005	0.005	95
		$\beta_2 = -0.264$	2500	0.002	0.010	0.010	93
			5000	0.001	0.007	0.007	95
			10 000	0.000	0.005	0.005	94
	≈ 30%	$\beta_1 = -0.132$	2500	0.004	0.010	0.010	93
			5000	0.002	0.008	0.007	93
			10 000	0.001	0.005	0.005	96
		$\beta_2 = -0.264$	2500	0.002	0.010	0.010	94
			5000	0.001	0.007	0.007	95
			10 000	0.001	0.005	0.005	96
5.4	≈ 15%	$\beta_1 = -0.227$	2500	0.004	0.014	0.014	95
			5000	0.002	0.010	0.010	94
			10 000	0.001	0.007	0.007	95
		$\beta_2 = -0.456$	2500	0.004	0.015	0.014	92
			5000	0.003	0.010	0.010	95
			10 000	0.001	0.007	0.007	94
	≈ 30%	$\beta_1 = -0.227$	2500	0.016	0.018	0.018	86
			5000	0.010	0.013	0.013	90
			10 000	0.005	0.009	0.010	92
		$\beta_2 = -0.456$	2500	0.006	0.017	0.016	93
			5000	0.003	0.012	0.012	94
			10 000	0.002	0.008	0.008	95

Abbreviations: ASE, average standard error; CP, coverage probability; ESD, empirical standard deviation.
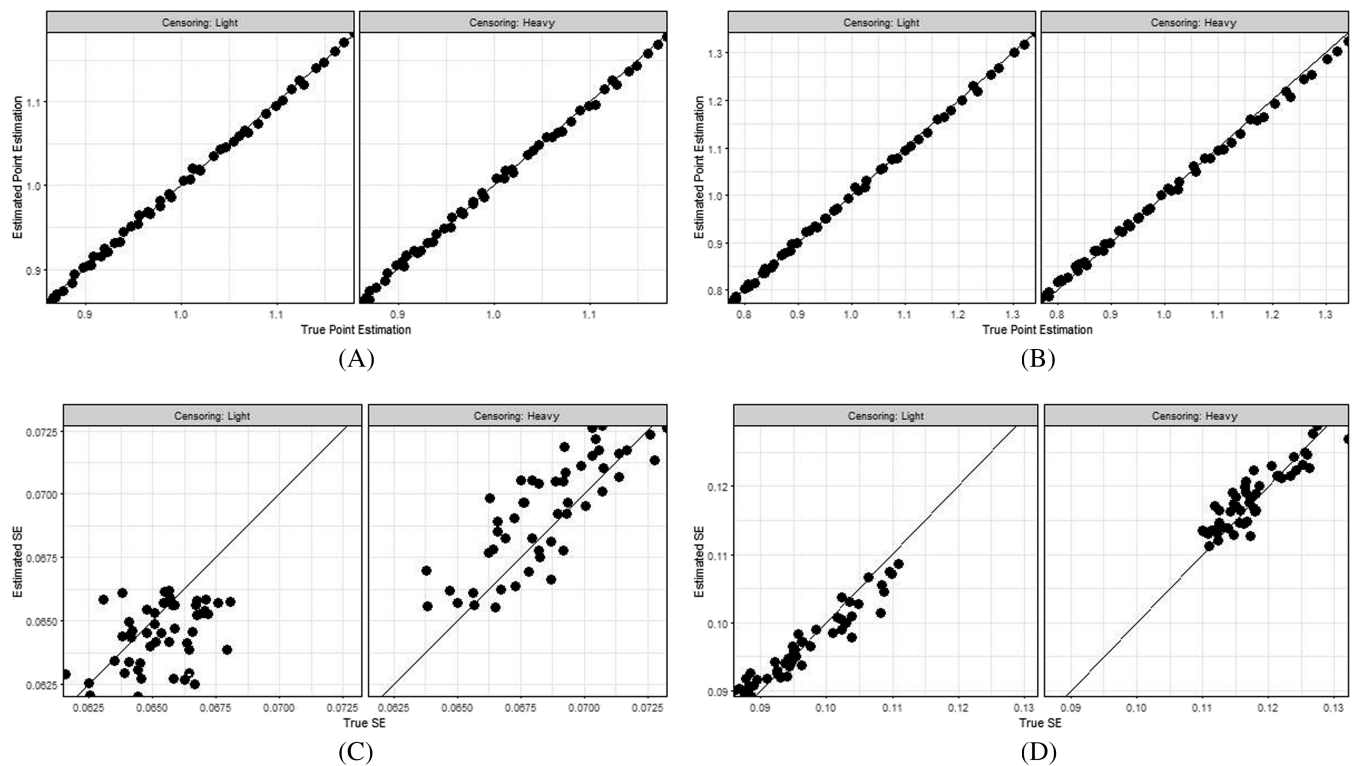

FIGURE 1 True and estimated values and standard deviation of $\hat{\eta}$ for $L = 1.8$ and $L = 5.4$. A, Estimation for $L = 1.8$; B, Estimation for $L = 5.4$; C, Standard error for $L = 1.8$; D, Standard error for $L = 5.4$
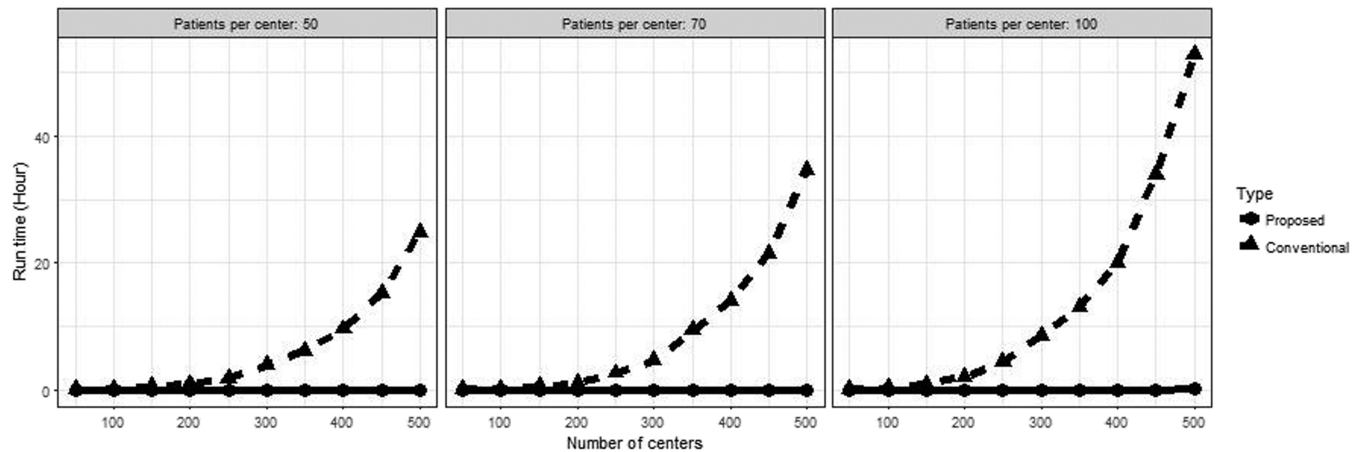
FIGURE 2 Computation time for our proposed and conventional methods with different *J*'s and number of patients per center

Another disadvantage of using conventional methods to fit model (1) is that such an approach requires large storage to create the center indicators for a large data set. For example, the data set that motivated our methods (with > 5000 centers and >1 million patients) requires R to allocate $\approx$ 50 GB to create all the center indicators using the common data types, and about 10 GB if special packages are used (eg, *sparseMatrix*). In contrast, fitting model (1) through our proposed methods does not require the creation of center indicators, which greatly reduces the storage requirements.

In the Supporting Information document, we present simulation results for an additional setup where 20 of the centers ($j = 1, \dots, 10, 41, \dots, 50$) have 25 subjects, with the remaining centers having 50 subjects. Residual bias is increased for centers with 25 subjects. We do not endorse the proposed methods for centers with < 25 subjects.

5 | ANALYSIS OF US DIALYSIS FACILITIES

We applied the proposed methods to analyze survival times for ESRD patients, using data obtained from the United States Renal Data System (USRDS). We included all patients initiating renal replacement therapy (RRT) on hemodialysis in the United States between January 1, 2004, and December 31, 2014. We excluded patients with a prior kidney transplant and patients aged <18 at the time of RRT initiation. For each patient, follow-up started at the date of RRT initiation and continued until the earliest of the following four events: death, transplantation, loss to follow-up, or 12/31/2014. The event of primary interest is death. We have $n = 1\,061\,403$ patients from $J = 5301$ ESRD facilities. Approximately, 64% of patients are observed to die. We chose $L = 5$ years as the truncation point. Out of $n = 1\,061\,403$ patients, 55% were observed to die before L, 27% were censored before L, and 18% were truncated at L.

Prognostic factors historically reported as being important and, hence, included in our analysis include the following: calendar year of RRT initiation (centered at 2004), age at RRT initiation (centered at 50 years and scaled by 5), gender, race (Caucasian, Asian, Black, and Other), ethnicity (Hispanic or not), primary renal diagnosis (glomerulonephritis (GN), diabetes, hypertension, and others), and 10 binary indicators of comorbidity conditions: cancer, diabetes, athlerosclerotic heart disease (ASHD), congestive heart failure (CHF), chronic obstructive pulmonary disease (COPD), cerebrovascular accident (CVA), peripheral vascular disease (PVD), elicit drug use, smoking status (current/former, non), and alcohol consumption. The RMST model of mortality includes the aforementioned covariates as predictors and $J = 5301$ ESRD Network facilities as centers. The Cox model for censoring includes the same set of covariates and is stratified by center. Estimated coefficients for the RMST model are displayed in Table 2.

The center effect is evaluated by both center-specific RMST μ_j and rescaled η_j. Figure 3 shows the histogram of the $J = 5301$ center-specific $\hat{\mu}_j$'s, the majority of which lie between 3.5 and 5. Each μ_{0j} can be interpreted as the estimated 5-year RMST for a patient from facility j with all covariates equal to 0. As implied by Table 2, such a "reference" patient would be age 50, initiate dialysis in 2004, be a non-Hispanic male Caucasian, have a GN as a primary renal diagnosis, and no comorbid conditions.

TABLE 2 Analysis of USRDS data: estimated covariate effects on RMST (L = 5 years)

Covariate	$\hat{\beta}$	SE	p	$\exp\hat{\beta}$
(Age-50)/5 (Years)	−0.057	0	< 0.001	0.945
Initiation year-2004	−0.041	0	< 0.001	0.960
Gender (vs male)				
female	−0.003	0.002	0.060	0.997
Ethnicity (vs non-Hispanic)				
Hispanic	0.139	0.003	< 0.001	1.149
Race (vs Caucasian)				
Asian	0.147	0.004	< 0.001	1.158
Black	0.11	0.002	< 0.001	1.116
Other	−0.041	0.008	< 0.001	0.960
Primary renal diagnosis (vs GN)				
Diabetes	−0.025	0.003	< 0.001	0.975
Hypertention	−0.016	0.003	< 0.001	0.984
Other	−0.117	0.004	< 0.001	0.890
Comorbidity (vs no)				
ASHD	0.009	0.002	< 0.001	1.010
Cancer	−0.196	0.004	< 0.001	0.822
CHF	−0.15	0.002	< 0.001	0.861
COPD	−0.141	0.003	< 0.001	0.868
CVA	−0.08	0.003	< 0.001	0.923
Diabetes	−0.022	0.003	< 0.001	1.022
Drug use	−0.096	0.007	< 0.001	0.908
PVD	−0.102	0.003	< 0.001	0.903
Tobacco use	−0.011	0.003	< 0.001	0.989
Alcohol use	−0.132	0.007	< 0.001	0.876

Abbreviations: ASHD, atherosclerotic heart disease; CHF, congestive heart failure; COPD, chronic obstructive pulmonary disease; CVA, cerebrovascular accident; PVD, peripheral vascular disease; RMST, restricted mean survival time; SE, standard error; USRDS, United States Renal Data System.
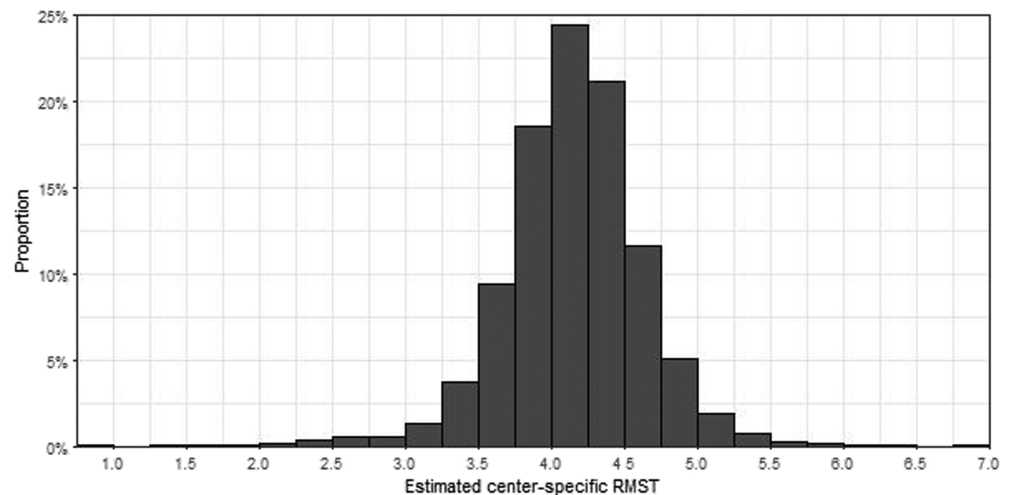

FIGURE 3 Histogram of estimated J = 5301 center-specific restricted mean survival time (RMST) μ_j's

Figure 4 displays the point and interval estimates (95% confidence level) of rescaled η_j's. A total of 656 (12%) of facilities are significantly below average 5-year RMST, while 582 (11%) are significantly above average. There were 4063 (77%) facilities that were not significantly different from the average 5-year RMST.

It took R approximately 11.33 minutes to calculate the IPCW weights, 2.65 minutes to estimate our proposed methods, and then another 4.65 hours to calculate the standard error for $\hat{\mu}$. However, it requires R to allocate about 50 GB memory to create the data needed for conventional methods, which is impossible for most of the local computers. Thus, in this particular example, storage considerations alone preclude a meaningful comparison between the proposed and conventional GLM procedures with respect to run times.
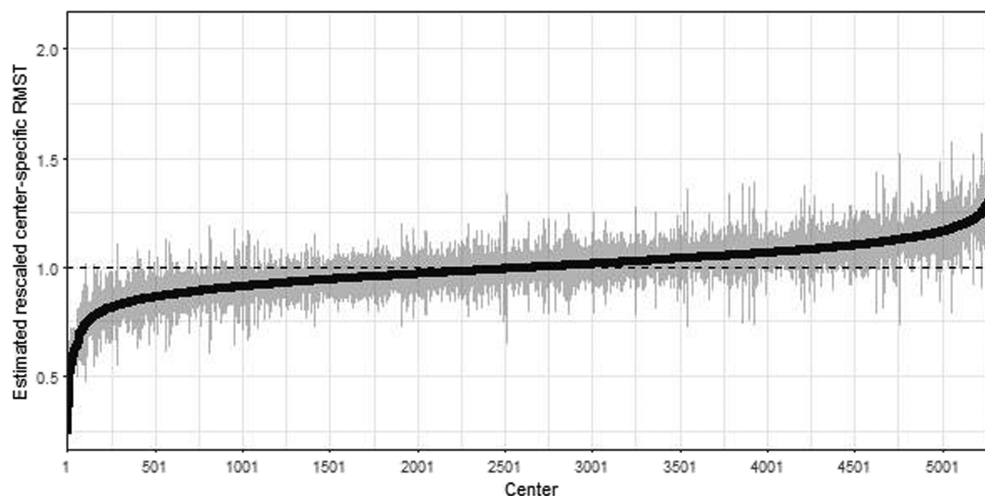
FIGURE 4 Point estimator and confidence interval of $J = 5301$ rescaled η_j's. RMST, restricted mean survival time

6 | DISCUSSION

In this report, we have developed a computationally attractive way to carry out facility profiling in terms of RMST. The proposed methods accommodate the estimation of fixed center effects through a normalized center effect measure. The methods are applicable large data numbers of both subjects and centers, as demonstrated through both simulation studies and the analysis of a large registry database. Computational advantages include great reductions in storage requirements and run times relative to conventional methods. We have demonstrated that our proposed methods have good finite-sample performance.

We applied our methods to ESRD data and out of $J = 5301$ ESRD facilities detected about 12% facilities significantly below and 11% significantly above average in terms of 5-year RMST. This proportion might be inflated, which suggests perhaps employing an empirical null method.[9,11,20] We expect that, under an empirical null, considerably fewer centers would be flagged as significantly different from average.

The parameter choice of the weights depends on the research objective of the analysis. Our data analysis uses equal weight across all the center so that the resulted weighted average is not dominated by the large centers. This way, it is easier to detect the centers with unusual performance. Some other reasonable choices include center-size proportional weights, resulting into the national average; this choice carries more interpretability but is driven by large centers.

It should be noted that the identification of μ_{0j} in model (1) requires that at least some subjects in center j have potential follow-up time $\geq L$; ie, $P(C_i > L | G_i = j, Z_i) > 0$. If not, then μ_{0j} cannot be estimated. More broadly, μ_{0j} unlikely to be estimated with meaningful precision unless ≥ 5 subjects have potential follow-up time $\geq L$. The choice of L is always an issue in the direct modeling of RMST; the issue is more pervasive in the context of facility profiling.

The methods developed in this article require the assumed RMST model to be correctly specified. Moreover, the computational techniques really quite heavily on the multiplicative structure in model (1). Analogous developments for other classes of RMST models may also be possible.

DATA AVAILABILITY STATEMENT

In keeping with their Data Use Agreement, the authors are not permitted to supply the USRDS data sets used in Section 5.

ORCID

Douglas E. Schaubel 🔟 https://orcid.org/0000-0002-9792-4474

REFERENCES

1. Karrison T. Restricted mean life with adjustment for covariates. *J Am Stat Assoc*. 1987;82:1169-1176.
2. Chen P, Tsiatis A. Causal inference on the difference of the restricted mean lifetime between two groups. *Biometrics*. 2001;57:1030-1038.
3. Andersen P, Hansen M, Klein J. Regression analysis of restricted mean survival time based on pseudo-observations. *Lifetime Data Anal*. 2004;10:335-350.
4. Andersen P, Perme MP. Pseudo-observations in survival analysis. *Stat Methods Med Res*. 2009;19:71-99.
5. Zhang M, Schaubel D. Estimating differences in restricted mean lifetime using observational data subject to dependent censoring. *Biometrics*. 2011;67:740-749.
6. Tian L, Zhao L, Wei L. Predicting the restricted mean event time with the subject's baseline covariates in survival analysis. *Biostatistics*. 2014;15:222-233.
7. Zhao L, Claggett B, Tian L, et al. On the restricted mean survival time curve in survival analysis. *Biometrics*. 2016;72:215-221.
8. Wang X, Schaubel D. Modeling restricted mean survival time under general censoring mechanisms. *Lifetime Data Anal*. 2018;24:176-199.
9. Kalbfleisch J, Wolfe R. On monitoring outcomes of medical providers. *Stat Biosci*. 2013;5:286-302.
10. He K, Schaubel D. Standardized mortality ratio for evaluating center-specific mortality: assessment and alternative. *Stat Biosci*. 2015;7:296-321.
11. He K, Kalbfleisch J, Li Y, Li Y. Evaluating hospital readmission rates in dialysis facilities; adjusting for hospital effects. *Lifetime Data Anal*. 2013;19:490-512.
12. Cox DR. Regression models and life tables (with discussion). *J R Stat Soc Ser B Methodol*. 1972;34:187-200.
13. Boudreau C, Lawless J. Survival analysis based on the proportional hazards model and survey data. *Can J Stat*. 2006;34:203-216.
14. Robins JM, Rotnitzky A. Recovery of information and adjustment for dependent censoring using surrogate markers. In: Jewell NP, Dietz K, Farewell VT, eds. *AIDS Epidemiology: Methodological Issues*. Boston, MA: Birkhäuser; 1992:297-331.
15. Robins J, Finkelstein D. Correcting for noncompliance and dependent censoring in an AIDS clinical trial with inverse probability of censoring weighted (IPCW) log-rank tests. *Biometrics*. 2000;56:779-788.
16. Cox DR. Partial likelihood. *Biometrika*. 1975;62:269-275.
17. Breslow N. Contribution to the discussion of paper by D.R. Cox. *J R Stat Soc Ser B*. 1972;34:216-217.
18. Foutz R. On the unique consistent solution to the likelihood equations. *J Am Stat Assoc*. 1977;72:147-148.
19. Lunceford J, Davidian M. Stratification and weighting via the propensity score in estimation of treatment effects: a comparative study. *Statist Med*. 2004;23:2937-2960.
20. Efron B. Large-scale simultaneous hypothesis testing: the choice of a null hypothesis. *J Am Stat Assoc*. 2004;99:96-104.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

APPENDIX

JUSTIFICATION OF ALGORITHM IN SECTION 2.5

Here, we introduce some additional notation, for the sole purpose of establishing the requisite connections between (8)-(9) and the Cox score equations. Along these lines, consider the following stratified Cox model:

$$\lambda_{ij}^{\dagger}(t) = \lambda_{0j}^{\dagger}(t) \exp\{\gamma' Z_i\},$$

for the death hazard of a patient $i \in \{1, \ldots, n\}$ from cluster $j \in \{1, \ldots, J\}$ with baseline covariate Z_i. Naturally, we are not assuming that this model holds; we only introduce it since Cox regression software is designed to fit models with this structure. We set $N_i^{\dagger}(t)$ and $R_i^{\dagger}(t)$ as the counting process for death and at-risk indicator, respectively. As implemented by,

for example, R and SAS, γ and $\lambda_{0j}^{\dagger}$ can be estimated from the estimating equations given below with weights $W_i^{\dagger}(t) = 1$. A variant of the standard estimating equation is well developed by weighting with $W_i^{\dagger}(t)$ as IPCW weights.[5]

Our goal is to coerce the software (eg, *coxph* in R, *phreg* in SAS) to fit model (1) by solving estimating equations (8)-(9). As such, we now establish connections between (8)-(9) and the IPCW version of the Cox score equations

$$\sum_{j=1}^{J} \sum_{i=1}^{n} \int_0^{\tau} G_{ij} \widehat{W}_{ij}^{\dagger}(u) \left\{ \mathbf{Z}_i - \overline{\mathbf{S}}_j^{\dagger}(u; \gamma, \mathbf{W}^{\dagger}) \right\} dN_i^{\dagger}(u) = \mathbf{0}, \tag{A1}$$

$$\int_0^{t} \frac{\sum_{i=1}^{n} G_{ij} W_{ij}^{\dagger}(u) dN_i^{\dagger}(u)}{\sum_{i=1}^{n} G_{ij} W_{ij}^{\dagger}(u) \exp\{\gamma' \mathbf{Z}_i\} R_i^{\dagger}(u)} = \Lambda_{0j}^{\dagger}(t), j = 1, \ldots, J, \tag{A2}$$

where, for $k = 0, 1, 2,$

$$\mathbf{S}_j^{(k)\dagger}(t; \gamma, \mathbf{W}^{\dagger}) = \frac{\sum_{i=1}^{n} G_{ij} W_{ij}^{\dagger}(t) \exp\{\gamma' \mathbf{Z}_i\} R_i^{\dagger}(t) \mathbf{Z}_i^{\otimes k}}{\sum_{i=1}^{n} G_{ij} W_{ij}^{\dagger}(t) \exp\{\gamma' \mathbf{Z}_i\} R_i^{\dagger}(t)}, \tag{A3}$$

$$\overline{\mathbf{S}}_j^{\dagger}(t; \gamma, \mathbf{W}^{\dagger}) = \frac{\mathbf{S}_j^{(1)\dagger}(t; \gamma, \mathbf{W}^{\dagger})}{\mathbf{S}_j^{(0)\dagger}(t; \gamma, \mathbf{W}^{\dagger})}. \tag{A4}$$

First, remove the integral signs from (A1)-(A2), such that only the increment at time u is considered. Next, replace $\widehat{W}_{ij}^{\dagger}(u)$ with $\widehat{W}_{ij} \Delta_i^Y Y_i$, then set $R_i^{\dagger}(u) = 1$ and $dN_i^{\dagger}(u) = 1$. By this point, it is clear that u is arbitrary and can be set to any positive integer (ie, the same integer for all i). To conform with the software, we can set u equal to any positive number; here, we set $u = 1$. Then, (A1)-(A2) are equivalent to our (8)-(9) after adding an offset $-\log(Y_i)$ to the linear predictor.

Combining the above information implies that our proposed model can be fitted using standard Cox regression software, with the data set augmented such that (a) observation time set to 1 for each subject, (b) $\widehat{W}_{ij} \Delta_i^Y Y_i$ used for a weight, (c) $-\log(Y_i)$ used for an offset, and (d) center serve as strata.