Received: 23 October 2017 | Revised: 22 July 2018 | Accepted: 3 December 2018

DOI: 10.1002/sim.8076

RESEARCH ARTICLE

WILEY Statistics in Medicine

Evaluating center-specific long-term outcomes through differences in mean survival time: Analysis of national kidney transplant data

Kevin He[1,2] | Valarie B. Ashby[2] | Douglas E. Schaubel[1,2]

[1]Department of Biostatistics, School of Public Health, University of Michigan, Ann Arbor, Michigan

[2]Kidney Epidemiology and Cost Center, School of Public Health, University of Michigan, Ann Arbor, Michigan

Correspondence
Kevin He, School of Public Health, University of Michigan, 1415 Washington Heights, Ann Arbor, MI 48109.
Email: kevinhe@umich.edu

Funding information
National Institutes of Health, Grant/Award Number: RO1 DK070869

Center-specific survival outcomes of kidney transplant recipients are an important quality measure, with several challenges. Existing methods based on restricted mean lifetime tend to focus on short- and medium-term clinical outcomes and may fail to capture long-term effects associated with quality of follow-up care. In this report, we propose methods that combine a lognormal frailty model and piecewise exponential baseline rates to compare the mean survival time across centers. The proposed methods allow for the consistent estimation of mean survival time as opposed to restricted mean lifetime and, within this context, permits more accurate profiling of long-term center-specific outcomes. Asymptotic properties of the proposed estimators are derived, and finite-sample properties are examined through simulation. The proposed methods are then applied to national kidney transplant data. The novelty of the proposed techniques arises from several angles. We utilize mean survival, in contrast to the most previous works that considered the restricted mean. Few previous studies have used the integrated survival function as a basis for center effects. Few provider profiling methods use a random effects model to estimate fixed center effects.

KEYWORDS

center effect, kidney transplant, lognormal random effect, mean survival time

1 | INTRODUCTION

Multicenter studies become very popular in clinical science. Of particular interest is the determination of which centers have significantly better or significantly worse outcomes. For example, the methods proposed here are motivated by the need to profile center-specific long-term kidney transplant outcomes in the United States. Kidney transplantation has a sophisticated and well-established program for center-specific outcomes reporting and quality assurance. The data from this program are used by transplant centers for quality improvement, by payers and regulators to achieve quality assurance, and by referring physicians and patients to identify the appropriate center for treatment.

Abbreviations: OPTN, organ procurementand transplantation network; SRTR, scientific registry of transplant recipients

Conceptually, mean survival time is a meaningful measure by which to evaluate transplant centers. However, in the presence of censoring, mean survival time may not be well estimated, since the estimated survival function need not drop to zero. To address this concern, a commonly used alternative measure is the restricted mean lifetime, interpreted as the expected number of time units survived, out of an upper limit. To incorporate covariates, Karrison[1] compared the restricted mean lifetime between two groups using a piecewise exponential model. Zucker[2] extended Karrison's approach based on the stratified Cox model, in which the relationship between the groups and the hazard function is left arbitrary. To account for imbalances in prognostic factors by treatment status, Chen and Tsiatis[3] proposed estimators for the average causal effect (ACE) on restricted mean lifetime. Zhang and Schaubel[4] extended these methods to accommodate dependent censoring. Due to the use of the restriction time, the restricted mean lifetime is a useful measure for short- and medium-term clinical outcomes. However, patients who receive a kidney transplant have substantially longer survival, with the majority of graft losses and patient death occurring in the long term.[5] Thus, the resulting measures for short- and medium-term outcomes may fail to capture the long-term effects that reveal the overall quality of care provided by the center. In our motivating example, the longest-term outcome available in existing center-specific reports is 3-year survival.[6] By ignoring subsequent events, the loss of information may be substantial. Long-term center-specific profiling may offer an important insight into variations in processes and intensity of care and the optimization of kidney transplantation. Given these concerns, mean lifetime may be a more appropriate measure than restricted mean lifetime in the context of kidney transplantation.

In this report, we develop novel methods to profile center-specific measures of transplant utility. The remainder of this paper is organized as follows. The data sources and study population are described in the next section. In Section 3, we summarize some important issues in the comparison of long-term center-specific outcomes. In Section 4, we describe our proposed model and a method to estimate center-specific differences in mean survival time. Finite-sample properties are examined in Section 5 through simulations. The proposed methods are applied to national kidney transplant data in Section 6. This paper concludes with a discussion in Section 7.

2 | DATA SOURCE AND STUDY POPULATION

Data were obtained from the Scientific Registry of Transplant Recipients (SRTR). The SRTR data system includes data on all donors, waitlisted candidates, and transplant recipients in the United States, submitted by the members of the Organ Procurement and Transplantation Network. The Health Resources and Services Administration, US Department of Health and Human Services provides oversight to the activities of the Organ Procurement and Transplantation Network and SRTR contractors.

Included in the analysis were adult patients ($\geq$ 18 years of age at transplant) who underwent deceased donor kidney transplantation between January 1990 and December 2008. Transplant centers with a sample size smaller than 10 were eliminated from additional analysis. The final sample size was $n = 146\,248$ from $J = 282$ centers across the United States. The mean age at transplant was 48, and the mean donor age was 36. The median survival time over the study period was 10.8 years across centers. Graft failure was considered to occur when the transplanted kidney ceased to function. Failure time was defined as the time from transplantation to graft failure or death, whichever occurred first.

3 | PRELIMINARIES

In this section, we summarize the computation challenges in the analysis of long-term center-specific outcomes for kidney transplant centers. We then discuss the reasons we chose the specific model.

3.1 | Piecewise exponential baseline rates

One challenge for estimating mean survival time is that it may not be well estimated in the presence of censoring. To address this concern, we propose a method based on piecewise exponential baseline rates with a lognormal random effect. The main reason that we select a parametric hazard model instead of the Cox model is that the latter allows for inference only on the $(0, \tau]$ interval, where τ is the maximum observation time. In other words, in the Cox model, the baseline hazard function is completely unknown, and estimation and inference on the baseline may not be stable when the risk set is small. Another advantage of piecewise exponential baseline rates is that we can use the link between mixed models

and the lognormal random effects model. Holford[7] and Laird and Olivier[8] noted that the piecewise exponential model is equivalent to a Poisson loglinear regression model with a count of either 0 or 1 for each combination of individual and interval, where the death indicator is the response and the log of exposure time enters as an offset. Such connections lead to both computational and theoretical simplification. For instance, the existing generalized linear mixed model software, such as PROC GLIMMIX in SAS, can be applied to estimate the parameters. This is especially important for our motivating example, since estimating center effects can be a computationally intense task in nationwide studies with large numbers of patients and centers. Finally, the choice of this parametric baseline hazard function leads to a fully parametric marginal likelihood, such that we can make use of maximum likelihood techniques to estimate the parameters. In contrast, for Cox models with lognormal random effects, the asymptotic properties of the penalized partial likelihood approach are not yet well established.

3.2 | Lognormal random effects

Another important consideration in our study is that the comparison of center-specific outcomes should be based on risk-adjusted models that account for center-specific heterogeneity with respect to the types of patients treated. In our motivating example, variations may exist among centers in both baseline risk and effectiveness of follow-up therapy. In other words, just as it is reasonable to think that baseline survival may differ across centers, it is possible that center-specific covariate effects may also be unequal. Evaluations of such variations are important to determine how a particular therapy should be administered. This concern can be addressed using a random effects model, in which subjects in a center share some common effects, with center-specific effects treated as a sample from a specific probability distribution. As pointed out by a Reviewer, it should be noted that the incorporation of random effects does not control for unmeasured center- or individual-level confounders, but only corrects the variance estimator such that it accounts for within-center correlations.[9]

A wide variety of random effects models have been proposed in survival analysis. Among them, the gamma frailty model[10-12] and the lognormal random effects model[13-15] are the most extensively studied approaches. As interactions between covariate and random effects can be readily implemented within the lognormal random effects model, this approach is applied in our study.

4 | ESTIMATION

Let T_i and C_i represent the survival and censoring time, respectively, for the ith patient. Observation times are denoted by $X_i = T_i \wedge C_i$, where $a \wedge b = \min\{a, b\}$. Correspondingly, we set the observed death indicator to $\Delta_i = I(T_i \leq C_i)$. Let J be the number of centers, with the total number of subjects denoted by $n = \sum_{j=1}^{J} n_j$, where n_j is the number of subjects in center j. Each subject is characterized by a time-constant covariate vector, $\mathbf{Z}_i$. Correspondingly, let $\boldsymbol{\beta}$ be the fixed effects coefficient vector for $\mathbf{Z}_i$. Let G_i denote the center for subject i, ie, $G_i = j$ means that subject i belongs to center j. Let $a_i(t)$ be a time-dependent covariate for attained age, with β_a being the fixed effects coefficient for attained age. Let $\mathbf{b} = (\mathbf{b}_0^T, \mathbf{b}_1^T)^T$ be a vector of random effects, where $\mathbf{b}_0^T = (b_{01}, \ldots, b_{0J})$ and $\mathbf{b}_1^T = (b_{11}, \ldots, b_{1J})$ are vectors containing random intercepts and random slopes, respectively. We assume that $\mathbf{b}_0 \backsim N(\mathbf{0}_{J\times 1}, \theta_0 \mathbf{I}_{J\times J})$ and $\mathbf{b}_1 \backsim N(\mathbf{0}_{J\times 1}, \theta_1 \mathbf{I}_{J\times J})$.

Our proposed estimation procedure involves two stages. In Stage 1, we obtain parameter estimates under the assumed model. In Stage 2, we then compute subject- and center-specific fitted survival functions. Finally, we estimate center-specific mean survival time averaging across the marginal covariate distribution using fitted values based on parameters obtained from Stage 1. Specifics are described in the following subsections.

4.1 | Stage 1: Model and parameter estimation

To approximate the baseline hazard by piecewise constants, we divide the observed time period into K follow-up time intervals with cut points $0 = t_0 < t_1 < \cdots < t_K = \infty$. Within each of the intervals, the baseline hazard function, $\lambda_0(t)$, is assumed to be

$$\lambda_0(t) = \lambda_k, \quad t \in \Omega_k \equiv [t_{k-1}, t_k), \quad k = 1, \ldots, K.$$

Thus, the baseline hazard is a constant within each interval but is allowed to vary across intervals. As shown by Lawless and Zhang,[16] models using a piecewise constant baseline hazard with suitable intervals often yield accurate estimates for fixed effects and random effects parameters.

One advantage of the piecewise exponential model is that it can easily accommodate time-varying covariates provided that they change values only at interval boundaries. In our study, the only time-varying covariate is attained age. To accommodate the change in attained age, we further split each of the time intervals Ω_k into multiple subintervals Ω_{kq}, where $q = 1, \ldots, k_q$ and k_q is the number of subintervals within Ω_k. The values of attained age are then defined as the age at the left boundary of each subinterval. The boundaries of subintervals are chosen as either the cut points of the original interval or the time scales corresponding to an integer value of attained age. Note that, in this way, each subinterval gets its own measure of exposure and its own death indicator, but all subintervals are tagged as belonging to the same interval; hence, they get the same baseline hazard. This procedure may be best explained by way of an example. For individual i with age at transplant equal to 18.3 and with 3-year follow-up time, we can split the particular intervals $[0, 1)$ and $[1, 3)$ hazard functions into subinterval $[0, 0.7)$, $[0.7, 1)$, $[1, 1.7)$, $[1.7, 2.7)$, $[1.7, 2.7)$, and $[2.7, 3)$, and the corresponding values of attained age are then recorded as a step function: $a_i(t) = 18.3, 19, 19.3, 20, 21$, and 21.3 for each subinterval.

The assumed proportional hazard frailty model is formulated as

$$\lambda_{ij}(t) = \lambda_0(t) \exp\left\{\mathbf{Z}_i^T \boldsymbol{\beta} + b_{0j} + a_i(t)(\beta_a + b_{1j})\right\}, \tag{1}$$

where $\lambda_{ij}(t) \equiv \lambda(t|\mathbf{Z}_i, a_i(t), G_i = j, \mathbf{b}_j)$ denotes the hazard function for subject i conditional on $\mathbf{Z}_i, a_i(t), G_i = j$, and $\mathbf{b}_j = (b_{0j}, b_{1j})^T$.

The main reason we consider time-dependent age (ie, attained age) instead of time-constant age (ie, age at time 0; in our setting, age at transplant) is that age is often the most important risk factor for long-term studies, underscoring the importance of separating the contributions of attained age and follow-up time in analyzing patient outcomes. This is particularly relevant in long-term follow-up studies where there may be considerable changes in the mortality pattern with increasing age. For example, in the context of our motivating example, preliminary analysis indicated that the trends associated with follow-up time diminish considerably as follow-up time increases. In contrast, the effect of attained age plays a more important role as age increases (details will be provided in Section 6). Since one of the concepts motivating our work is that trends over time may differ by center, it makes sense to allow for the effect of attained age to differ by center. We therefore allow attained age, $a_i(t)$, to have both fixed and random effects in model (1).

The conditional likelihood is given by

$$L(\lambda_1, \ldots, \lambda_K, \boldsymbol{\beta}; \mathbf{b}_j) \propto \Pi_{i=1}^{n_j} \Pi_{k=1}^{K} \Pi_{q=1}^{K_q} \left\{\lambda_k t_{ikq} \exp\left(\mathbf{Z}_i^T \boldsymbol{\beta} + b_{0j} + a_{ikq}(\beta_a + b_{1j})\right)\right\}^{\Delta_{ikl}} \exp\left\{-\lambda_k t_{ikq} \exp\left(\mathbf{Z}_i^T \boldsymbol{\beta} + b_{0j} + a_{ikq}(\beta_a + b_{1j})\right)\right\},$$

where t_{ikq} is the total time for subject i in subinterval Ω_{kq} and Δ_{ikq} is the event indicator for subject i in subinterval Ω_{kq}. Assuming that the J centers are independent, the conditional likelihood across all centers is given by

$$L(\lambda_1, \ldots, \lambda_K, \boldsymbol{\beta}, \mathbf{b}) = \prod_{j=1}^{J} L(\lambda_1, \ldots, \lambda_K, \boldsymbol{\beta}, \mathbf{b}_j).$$

Then, the marginal likelihood has the form

$$L(\lambda_1, \ldots, \lambda_K, \boldsymbol{\beta}, \boldsymbol{\theta}) = \prod_{j=1}^{J} \int \exp\{g(\lambda_1, \ldots, \lambda_K, \boldsymbol{\beta}, \boldsymbol{\theta})\} d\mathbf{b}_j,$$

where $\boldsymbol{\theta} = (\theta_0, \theta_1)$ and

$$g(\lambda_1, \ldots, \lambda_K, \boldsymbol{\beta}, \boldsymbol{\theta}) = \log L(\lambda_1, \ldots, \lambda_K, \boldsymbol{\beta}, \mathbf{b}) + \log f(\mathbf{b}_j|\boldsymbol{\theta}),$$

with $f(\mathbf{b}_j|\boldsymbol{\theta})$ being the bivariate normal density. The objective function for numerical minimization is twice the negative of the corresponding log-likelihood approximation. For given $\widehat{\boldsymbol{\theta}}$, the estimating equations come from setting the derivative of $\log L(\lambda_1, \ldots, \lambda_K, \boldsymbol{\beta}, \widehat{\boldsymbol{\theta}})$, with respect to the parameters $(\lambda_1, \ldots, \lambda_K, \boldsymbol{\beta})$, equal to zero. To simplify the computations, we use a Laplace approximation to approximate each such integral with a function that has a closed form. Specifically, the Laplace approximation to the marginal log-likelihood is as follows:

$$\prod_{j=1}^{J} \left\{ n_j g(\lambda_1, \ldots, \lambda_K, \boldsymbol{\beta}, \widehat{\boldsymbol{\theta}}) + \log(2\pi) - \frac{1}{2} \log\left|-n_j g''(\lambda_1, \ldots, \lambda_K, \boldsymbol{\beta}, \widehat{\boldsymbol{\theta}})\right| \right\},$$

where

$$g''(\boldsymbol{\beta}, \lambda_1, \ldots, \lambda_K, \widehat{\boldsymbol{\theta}}) = \left.\frac{\partial^2 g(\boldsymbol{\beta}, \lambda_1, \ldots, \lambda_K, \boldsymbol{\theta})}{\partial\theta\partial\theta'}\right|_{\widehat{\theta}},$$

and θ satisfies the first-order condition, ie,

$$\frac{\partial g(\boldsymbol{\beta}, \lambda_1, \ldots, \lambda_K, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = 0,$$

for given values of $\boldsymbol{\beta}$ and λ. Compared with other estimation procedures, such as the pseudo-likelihood estimator, the Laplace estimators have been demonstrated to exhibit better asymptotic behavior and less small-sample bias than pseudo-likelihood estimators. Further details can be found in the work of Feng et al.[17]

4.2 | Stage 2: Difference in mean survival time

It is well known that patient characteristics may vary significantly across centers and can have a substantial impact on the expected center-specific survival outcomes. An accurate estimation of center effects must account for potential covariate imbalances across centers. To address this concern, we use a technique based on direct standardization.

Direct standardization is a commonly used approach in the comparison of center-specific outcomes.[18] One can express a directly standardized measure as the difference or ratio of expected-to-observed outcomes in the total study population; in our case, we use the difference. The first term (which is center specific) represents the expected outcomes if all patients were treated at the given center; the second term is equal to the total observed outcomes in the study population (eg, national average). This is somewhat analogous to the ACE from the causal inference literature. For the ACE, the study population is evaluated under two conditions: all patients treated versus all patients untreated. In our case, we contrast "all patients subject to center j practices" versus "all patients subject to national average practices." Because the same standard population is applied to all centers, directly standardized measures are not affected by differences in center-specific covariate distributions and, hence, are directly comparable.

Similarly, to factor out the impact of imbalances in center-specific covariate distributions, we define the center effect for center j with respect to mean survival time as follows:

$$\delta_j = E(T_i|G_i = j) - E(T_i),$$

with an estimator for δ_j given by

$$\hat{\delta}_j = \hat{E}(T_i|G_i = j) - \hat{E}(T_i). \tag{2}$$

The first term in (2), ie,

$$\hat{E}(T_i|G_i = j) = n^{-1}\sum_{\ell=1}^{J}\sum_{i=1}^{n_\ell}\int_0^\infty \hat{S}_{i(j)}(u)du,$$

has the interpretation of mean survival time for the population, under the hypothetical scenario where all subjects were treated in center j. Here, $S_{i(j)}(u)$ corresponds to the hypothetical survival of a randomly chosen individual in our population if, possibly contrary to fact, this subject is in center j, ie,

$$\hat{S}_{i(j)}(u) = \exp\left\{-\sum_{k=1}^{K}\sum_{q=1}^{k_q}\hat{\lambda}_k(t_{ikq} \wedge u)\exp\left\{\mathbf{Z}_i^T\hat{\boldsymbol{\beta}} + \hat{b}_{0j} + a_{ikq}(\hat{\beta}_a + \hat{b}_{1j})\right\}\right\}.$$

In contrast, the second term, ie,

$$\hat{E}(T_i) = n^{-1}\sum_{\ell=1}^{J}\sum_{i=1}^{n_\ell}\int_0^\infty \hat{S}_{i\ell}(u)du,$$

is the estimated mean survival time for the population, where $\hat{S}_{i\ell}(u)$ is the estimated survival for a subject i in center ℓ, ie,

$$\hat{S}_{i\ell}(u) = \exp\left\{-\sum_{k=1}^{K}\sum_{q=1}^{k_q}\hat{\lambda}_k(t_{ikq} \wedge u)\exp\left\{\mathbf{Z}_i^T\hat{\boldsymbol{\beta}} + \hat{b}_{0\ell} + a_{ikq}(\hat{\beta}_a + \hat{b}_{1\ell})\right\}\right\}.$$

Note that δ_j is the expected survival if all patients had transplants at center j, minus the expected survival for the population. With respect to interpretation, $\delta_j > 0$ indicates that center j has a greater mean survival than the overall average. The delta method is used to obtain the variance estimator and confidence intervals. Thus, the proposed methods can be carried out as an internal evaluation of center-specific long-term mortality (eg, for centers to evaluate themselves or for a governing body to evaluate this center's mortality, comparing to that expected at the national level).

The variance estimator for $\hat{\delta}_j$ can be obtained using the delta method. In particular, $\hat{\delta}_j$ is a function of fixed effects and random effects, eg, $\hat{\delta}_j = f_j(\hat{\boldsymbol{\psi}})$, where $\hat{\boldsymbol{\psi}} = (\hat{\lambda}_1, \ldots, \hat{\lambda}_K, \hat{\boldsymbol{\beta}}, \hat{\mathbf{b}})$. Let $\widehat{\mathrm{Cov}}(\hat{\boldsymbol{\psi}})$ be the covariance matrix for $\hat{\boldsymbol{\psi}}$. Then, the variance for $\hat{\delta}_j$ can be estimated by

$$\widehat{\mathrm{Var}}(\hat{\delta}_j) = f_j'(\hat{\boldsymbol{\psi}})^T \widehat{\mathrm{Cov}}(\hat{\boldsymbol{\psi}}) f_j'(\hat{\boldsymbol{\psi}}),$$

where $f_j'(\hat{\boldsymbol{\psi}})$ is the corresponding vector of derivatives. Similarly, confidence intervals of $\hat{\delta}_j$ can be obtained. Centers with confidence intervals completely above 0 will be identified as significantly better than the national average, whereas centers with confidence intervals completely below 0 will be identified as significantly worse than the national average. The decision rule used in our analysis is consistent with rules frequently used in practice. For each center j, we essentially test $H_j : \delta_j = 0$ against a two-sided alternative with a 0.05 significance level. The test is based on a limiting $N(0,1)$, such that rejection of H_j is equivalent to the 95% confidence interval not overlapping 0. As such, the decision rule is a deterministic (binary) function of a z score, for which the variability is appropriately accounted.

Finally, on the basis of the definition of mean survival time, the integral of survival function should go to ∞. However, in practice, it is more reasonable to calculate this integral until the patient achieves a certain maximum age. We will provide more details in Section 6 about how we choose this upper limit for our analysis.

5 | SIMULATION STUDY

The finite-sample properties of the estimators described above were evaluated through a series of simulation studies. We considered $J = 10, J = 25, J = 50$, and $J = 200$ centers. Death times were generated from a piecewise exponential model. To mimic the motivating example, we divided the whole follow-up period into the following intervals: [0, 1], (1, 2], (2, 3], (3, 5], (5, 10], and (10, ∞]. For the kth interval, the hazard function for subject i in center j was given by formula (1), where Z_i followed a Bernoulli distribution. We set $\beta_0 = -0.1$; $a_i(t) = a_{0i} + t$, where a_{0i} came from a Normal distribution with constant variance and center-dependent means, and t was the follow-up time. Random effects b_{0j} and b_{1j} ($j = 1, \ldots, J$) were generated from independent Normal distributions $N(0, \theta_0)$ and $N(0, \theta_1)$, respectively, with $\theta_0 = 0.25$ and $\theta_1 = 0.005$. We varied the sample size for each center as $n_j = 25, 50, 100$. The censoring percentages were approximately 20%. Each data configuration was replicated 500 times.

To fit the model and obtain the parameter estimators in Stage 1, we implemented the Laplace approximation using the SAS procedure GLIMMIX, with the option METHOD=LAPLACE for the required integral approximation. The mean survival time in Stage 2 is a numerical integration of scalar functions in one dimension over connected finite intervals. We programmed this estimation process using the SAS function QUAD. We calculated the mean survival time for each observation until the age 85 and computed the 95% confidence intervals.

In Table 1, we list the results for the proposed measures. The ASE estimators were sufficiently accurate, and the CP was generally consistent with the nominal value of 0.95. The results for setting with $J = 10$ or $J = 25$ centers are not shown. In general, for fixed J, as n_j increase, the CP increases. On the other hand, for fixed n_j, as J increase, the CP also increases. Overall, the simulation results indicated that the method is performing reasonably well.

Finally, we consider the robustness of the lognormal distribution assumption. Random effects b_{0j} and b_{1j} ($j = 1, \ldots, J$) were generated from independent gamma(2, 0.2) and gamma(2, 0.01) distributions, respectively. The last part of Table 2 shows that the proposed method with a mis-specified model still provides acceptable estimates for most of the centers. However, for centers with the largest difference mean survival time, the bias in the estimation is more pronounced, and robustness is an issue.

6 | DATA APPLICATION

Adjustment covariates in this study included age, race, gender, primary renal diagnosis, donation after cardiac death, expanded criteria donor, body mass index, years on dialysis prior to transplant, indicator for repeat kidney transplantation, and cold ischemic time. To separately examine the effects of follow-up time and attained age, we first fitted two

TABLE 1 Simulation results: performance of center effect estimators; $\bar{\delta}$ = average of δ_j over $j = 1, \ldots, J$

Distribution	J, n_j	Parameters	TRUE	BIAS	ESD	ASE	CP
Lognormal	$J = 50, n_j = 100$	δ_1	−5.273	−0.108	0.212	0.234	0.93
		δ_5	−4.180	0.074	0.296	0.322	0.98
		δ_{15}	−2.533	0.146	0.473	0.439	0.93
		δ_{25}	−0.819	0.020	0.613	0.565	0.93
		δ_{35}	1.219	−0.049	0.782	0.737	0.94
		δ_{45}	5.046	−0.164	1.112	1.045	0.94
		δ_{50}	11.284	−0.252	1.396	1.372	0.94
		$\bar{\delta}$	0.000	−0.000	0.740	0.636	0.94
	$J = 50, n_j = 50$	δ_1	−5.225	−0.084	0.330	0.333	0.94
		δ_5	−4.110	0.081	0.426	0.455	0.97
		δ_{15}	−2.500	0.168	0.616	0.615	0.96
		δ_{25}	−0.866	0.088	0.786	0.784	0.96
		δ_{35}	1.193	−0.052	1.099	1.012	0.93
		δ_{45}	5.021	−0.255	1.456	1.422	0.96
		δ_{50}	11.025	−0.261	1.806	1.862	0.96
		$\bar{\delta}$	−0.003	0.003	1.006	0.878	0.95
	$J = 50, n_j = 25$	δ_1	−5.168	−0.036	0.449	0.482	0.96
		δ_5	−4.040	0.131	0.590	0.649	0.98
		δ_{15}	−2.479	0.239	0.872	0.859	0.95
		δ_{25}	−0.797	0.035	1.164	1.069	0.94
		δ_{35}	1.232	−0.104	1.481	1.373	0.93
		δ_{45}	4.485	−0.257	1.974	1.892	0.94
		δ_{50}	10.510	−0.258	2.535	2.463	0.96
		$\bar{\delta}$	0.000	−0.001	1.384	1.197	0.94
	$J = 200, n_j = 25$	δ_1	−5.725	−0.284	0.416	0.363	0.94
		δ_{20}	−3.845	−0.031	0.305	0.638	0.93
		δ_{60}	−2.270	0.025	0.920	1.044	0.95
		δ_{100}	−0.745	0.033	1.125	1.308	0.96
		δ_{140}	1.131	0.119	1.459	1.586	0.93
		δ_{180}	4.898	−0.086	1.964	2.198	0.96
		δ_{200}	12.534	0.038	2.598	3.411	0.92
		$\bar{\delta}$	0.001	−0.001	1.240	1.566	0.93
Gamma	$J = 50, n_j = 50$	δ_1	−2.907	−0.167	0.287	0.344	0.98
		δ_5	−2.228	−0.002	0.387	0.408	0.94
		δ_{15}	−1.341	0.071	0.521	0.492	0.94
		δ_{25}	−0.586	0.141	0.503	0.566	0.94
		δ_{35}	0.334	0.136	0.682	0.785	0.92
		δ_{45}	2.936	−0.199	0.989	1.005	0.91
		δ_{50}	8.032	−0.258	1.506	1.967	0.85
		$\bar{\delta}$	0.004	−0.004	0.645	0.758	0.94

Abbreviation: ASE, average causal effect; CP, coverage probability; ESE, empirical standard error.

preliminary piecewise exponential models. In the first model, the effect of attained age was approximated as piecewise linear (Figure 1A). In the second model, attained age was treated as the time axis, and follow-up time was considered as a categorical covariate. The corresponding effect of follow-up time is shown in Figure 1B, which suggests that the hazards associated with follow-up time stabilized as follow-up time increased. This finding motivates us to assume that the hazard functions remain constant after the maximum follow-up time. However, this choice involves extrapolation outside of observation range and is an arbitrary choice for our particular data set. In a more general situation, we may use a linear function for extrapolation. Another preliminary analysis was to choose a reasonable upper limit for calculating the mean survival time. The maximum observed value for attained age in our data was 96. However, analysis based on person-year indicated that only a few observations had an attained age larger than 85 (< 0.005%). Therefore, we calculated the mean survival time for each patient until they reached age 85.

We next performed a lognormal frailty model with a piecewise exponential baseline rate. Specifically, we chose six intervals for time. A random effect for intercept was selected for the heterogeneity in baseline risk, and a random effect

TABLE 2 Regression parameters of transplant data: lognormal frailty model with a piecewise exponential baseline rate

Variable	Coefficient	Standard Error	*p* Value
Sex: Male	0.068	0.009	< 0.001
Donation after cardiac death	0.047	0.039	0.224
BMI: (18.5, 25]	−0.042	0.011	< 0.001
BMI: (25, 30]	0.029	0.014	0.043
BMI: > 30	0.126	0.021	< 0.001
PRD: Polycystic kidney disease	−0.258	0.028	< 0.001
PRD: Diabetes	0.591	0.021	< 0.001
PRD: Hypertension	0.178	0.022	< 0.001
PRD: Other diagnosis	0.130	0.020	< 0.001
Race: American African	0.043	0.010	< 0.001
Race: Hispanic	−0.248	0.016	< 0.001
Race: Asian	−0.460	0.026	< 0.001
Race: Other	−0.179	0.038	< 0.001
Cold ischemic time	0.003	0.0005	< 0.001
Dialysis time	0.025	0.002	< 0.001
Expanded criteria donor	0.172	0.015	< 0.001
Donor age	0.007	0.0003	< 0.001
Repeat kidney transplant	−0.010	0.016	0.550

Abbreviations: BMI, body mass index; PRD, primary renal diagnosis (disease leading to ESRD); ESRD, end-stage renal disease.

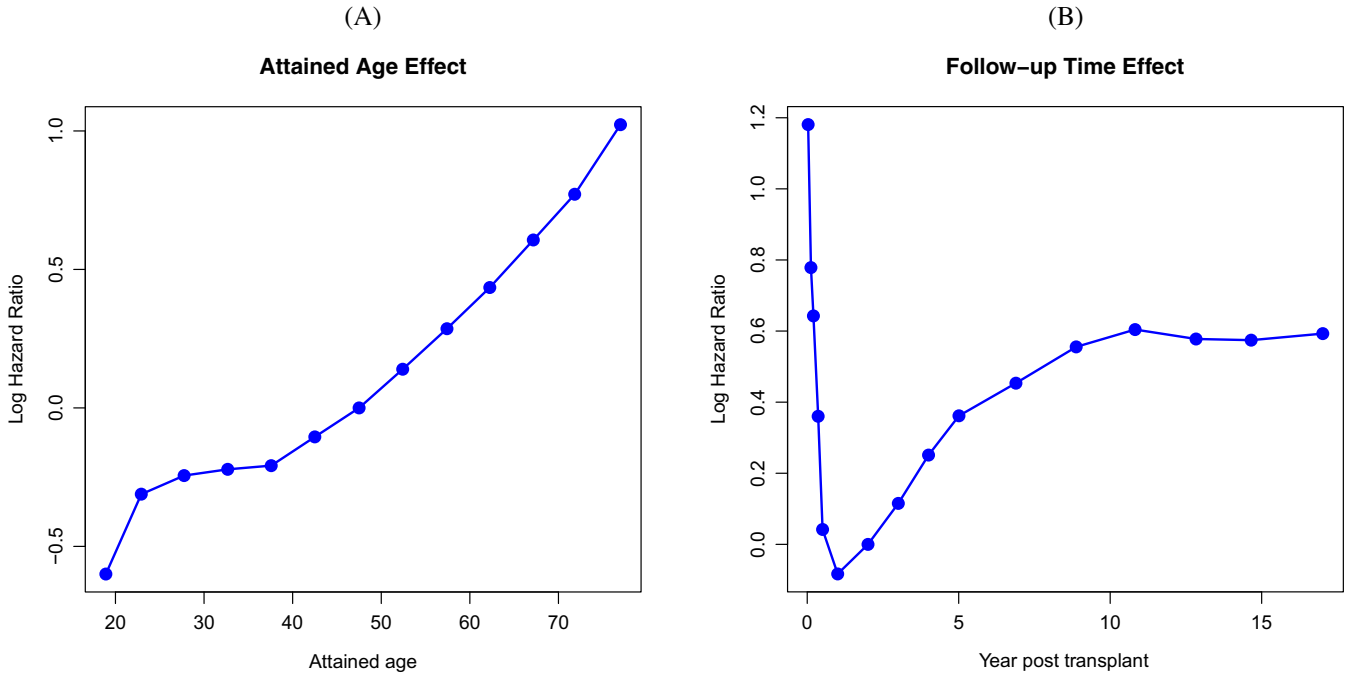

FIGURE 1 Effects of (A) attained age and (B) follow-up time [Colour figure can be viewed at wileyonlinelibrary.com]

for attained age was estimated for the heterogeneous effect of age at transplant and follow-up time. We then estimated the mean survival time across centers. We also constructed the standard error estimator for the difference of mean survival time based on the delta method described in Section 4.2.

We ordered centers based on difference in mean survival time, and we provided plots of confidence intervals in Figure 2. Centers with confidence intervals higher than 0 were identified as significantly better than the national average, whereas

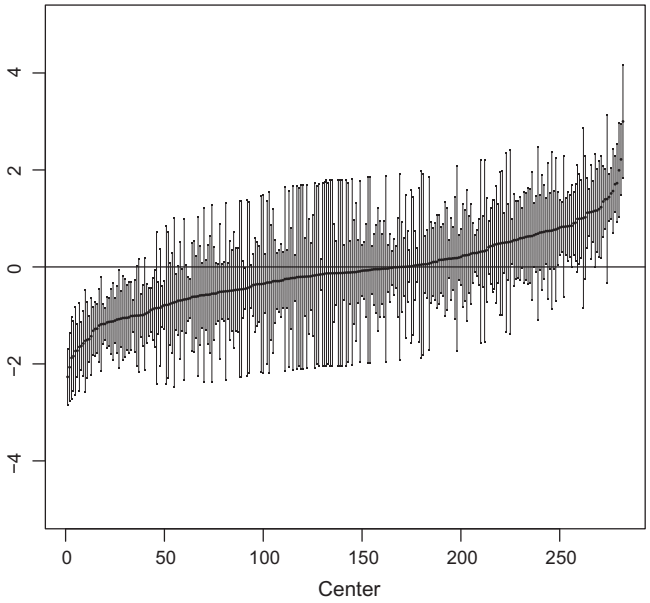

FIGURE 2 Analysis of the Scientific Registry of Transplant Recipients data: point estimates $\hat{\delta}_j$ ($j = 1, \ldots, 282$) and 95% confidence intervals

centers with confidence intervals lower than 0 were identified as significantly worse than the national average. A total of 50 centers had mean survival times significantly lower than expected, whereas 29 centers were significantly above the average.

As stated in Section 4.2, since both $\widehat{\delta}_j$ and $\widehat{\delta}_k$ (for centers j and k, respectively) are obtained by averaging over the same covariate distribution (ie, that of the total study population), one could validly compare $\widehat{\delta}_j$ and $\widehat{\delta}_k$, eg, by testing $H_{jk} : \delta_j = \delta_k$. In our analysis of the SRTR data, we took the perspective of a regulatory board, which would typically be more interested in determining which centers were different from the overall average, rather than directly comparing centers to each other.

7 | DISCUSSION

In this report, we combine a lognormal random effects model and piecewise exponential baseline hazard to compare mean survival time across centers. The Laplace approximation for integration is applied to obtain maximum likelihood estimations in a computationally tractable fashion. Our contribution over the literature can be summarized in the following aspects. First, the proposed method allows for valid estimation of long-term center-specific outcomes, in contrast to the most previous works that focused on short- or medium-term outcomes. Second, we propose a direct standardized measure based on the integrated survival function.

The proposed method is motivated by monitoring long-term center-specific outcomes. Surgeons and patients can compare center-specific results in a particular region as an external evaluation. Alternatively, center-specific evaluations can be carried out as an internal evaluation for centers to evaluate themselves or for a governing body to evaluate this center's survival, comparing to that expected at the national level. Centers with survival significantly worse than the national average may be subject to various degrees of intervention, including site visits and perhaps de-accreditation. Moreover, special care can be dedicated to such centers to improve their long-term outcomes.

ACKNOWLEDGEMENTS

This work was supported, in part, by National Institutes of Health grant R01-DK070869 and a grant from the Michigan Institute for Clinical and Health Research. The authors wish to thank two anonymous reviewers for their thoughtful comments. They also thank Kirsten Herold at the University of Michigan School of Public Health Writing Lab for her

helpful suggestions. The data reported here have been supplied by the Minneapolis Medical Research Foundation as the contractor for the Scientific Registry of Transplant Recipients (SRTR). The interpretation and reporting of these data are the responsibility of the authors and in no way should be seen as an official policy of or interpretation by the SRTR or the United States Government.

FINANCIAL DISCLOSURE

None reported.

CONFLICT OF INTEREST

The authors declare no potential conflict of interests.

ORCID

Kevin He https://orcid.org/0000-0002-8354-426X
Douglas E. Schaubel https://orcid.org/0000-0002-9792-4474

REFERENCES

1. Karrison T. Restricted mean lifetime with adjustment for covariates. *J Am Stat Assoc*. 1987;82:1169-1176.
2. Zucker DM. Restricted mean life with covariates: modification and extension of a useful survival analysis method. *J Am Stat Assoc*. 1998;93:702-709.
3. Chen PY, Tsiatis AA. Causal inference on the difference of the restricted mean lifetime between two groups. *Biometrics*. 2001;57:1030-1038.
4. Zhang M, Schaubel D. Estimating differences in restricted mean lifetime using observational data subject to dependent censoring. *Biometrics*. 2011;67:740-749.
5. Wolfe RA, Ashby VB, Milford EL, et al. Patient survival for waitlisted dialysis versus cadaveric renal transplant patients in the United States. *N Engl J Med*. 1999;341:1725-1730.
6. Organ Procurement and Transplantation Network and Scientific Registry of Transplant Recipients. Patient survival for waitlisted dialysis versus cadaveric renal transplant patients in the United States. In: OPTN/SRTR 2010 Annual Data Report. Technical Report LA-UR-02-525. Rockville, MD: Department of Health and Human Services, Health Resources and Services Administration, Healthcare Systems Bureau, Division of Transplantation; 2011.
7. Holford TR. The analysis of rates and survivorship using log-linear model. *Biometrics*. 1980;36:299-306.
8. Laird NM, Olivier D. Covariance analysis of censored survival data using log-linear analysis techniques. *J Am Stat Assoc*. 1981;76:231-240.
9. Song PXK. *Correlated Data Analysis: Modeling, Analytics, and Applications*. New York, NY: Springer Science+Business Media; 2007. *Springer Series in Statistics*.
10. Clayton DG. A model for association in bivariate life table and its application in epidemiological studies of familiar tendency in chronic disease incidence. *Biometrika*. 1978;65:141-151.
11. Clayton DG, Cuzick J. Multivariate generalization of the proportional hazards model (with discussion). *J R Stat Soc A Stat Soc*. 1985;148:82-117.
12. Klein JP. Semiparametric estimation of random effects using the Cox model based on the EM algorithm. *Biometrics*. 1992;48:795-806.
13. McGilchrist CA. REML estimation for survival models with frailty. *Biometrics*. 1993;49:221-225.
14. McGilchrist CA, Aisbett CW. Regression with frailty in survival analysis. *Biometrics*. 1991;47:461-466.
15. McGilchrist CA, Ohashi Y. Investigating centre effects in a multi-centre clinical trial of superficial bladder cancer. *Statist Med*. 1999;18:1961-1971.
16. Lawless JF, Zhang M. Analysis of interval-grouped recurrent-event data using piecewise constant rate functions. *Can J Stat*. 1998;26:549-565.
17. Feng S, Wolfe RA, Port FK. Frailty survival model analysis of the national deceased donor kidney transplant dataset using Poisson variance structures. *J Am Stat Assoc*. 2002;100:728-735.
18. He K, Schaubel DE. Methods for estimating center effects in survival analysis using direct standardization. *Statist Med*. 2014;33:2048-2061.

SUPPORTING INFORMATION

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

The following supporting information is available as part of the online article: **Table S1.** Simulation results: Performance of fixed effect and random effect parameters.

How to cite this article: He K, Ashby VB, Schaubel DE. Evaluating center-specific long-term outcomes through differences in mean survival time: Analysis of national kidney transplant data. *Statistics in Medicine*. 2019;38:1957–1967. https://doi.org/10.1002/sim.8076