Statistics
in Medicine

Research Article

Received 16 December 2012, Accepted 20 December 2013 Published online 17 January 2014 in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.6089

Methods for comparing center-specific survival outcomes using direct standardization

Kevin He*† and Douglas E. Schaubel

The evaluation of center-specific outcomes is often through survival analysis methods. Such evaluations must account for differences in the distribution of patient characteristics across centers. In the context of censored event times, it is also important that the measure chosen to evaluate centers not be influenced by imbalances in the center-specific censoring distributions. The practice of using center indicators in a hazard regression model is often invalid, inconvenient, or undesirable to carry out. We propose a semiparametric version of the standardized rate ratio (SRR) useful for the evaluation of centers with respect to a right-censored event time. The SRR for center j can be interpreted as the ratio of the expected number of deaths in the total population (if the total population were in fact subject to the center j mortality hazard) to the observed number of events. The proposed measure is not affected by differences in center-specific covariate or censoring distributions. Asymptotic properties of the proposed estimators are derived, with finite-sample properties examined through simulation studies. The proposed methods are applied to national kidney transplant data.

Keywords: center effect; Cox regression; survival analysis; standardized rate ratio; stratification

1. Introduction

In many situations, interest lies in the comparison of survival outcomes by center (e.g., treatment facility, hospital, or other entity serving as healthcare provider). Center-specific evaluations can be carried out on a regular basis (e.g., annually) to identity centers with poor performance. Alternatively, a retrospective evaluation over a longer period could be used to identify centers with exceptionally good results, with the goal of specifying best practices. In other cases, comparisons across centers may be an interesting secondary analysis; for example, in a multi-center study to evaluate the impact of a specific treatment on mortality. An accurate comparison of center-specific survival outcomes needs to account for imbalances in risk factor distributions among centers. For instance, the inclusion of high-risk patients by a given center can make that center's survival appear substandard. In addition, in the context of survival times subject to censoring, the same phenomenon can occur because of differences in center-specific censoring distributions.

In this report, we propose a semiparametric version of direct standardization, suitable for mortality comparisons by center. The proposed approach involves first fitting a Cox [1] regression model, stratified by center. The regression model is not used to directly estimate center effects, but rather to ensure that adjustment covariate effects are not confounded by center (which could occur in the absence of adjustment for center). For each center, j, the standardized rate ratio (SRR_j) is then computed as the ratio of expected to observed numbers of deaths, where 'observed' refers to the total number of deaths across all centers ($j = 1, \ldots, J$) and 'expected' represents the number of total deaths estimated to occur if in fact all centers had mortality hazard equal to that of center j. Because of the use of direct standardization, the $\{SRR_1, SRR_2, \ldots, SRR_J\}$ can be compared (and validly ordered) because the same

Department of Biostatistics, School of Public Health, University of Michigan, Ann Arbor, MI 48109-2029, U.S.A.
**Correspondence to: Kevin He, Department of Biostatistics, School of Public Health, University of Michigan, Ann Arbor, MI 48109-2029, U.S.A.*
†E-mail: kevinhe@umich.edu

covariate and censoring distribution is applied to each, that is, the total study population serves as the standard. This is in contrast to indirectly standardized measures, such as the standardized mortality ratio (SMR).

The motivating example for this study involves the evaluation of center-specific post-transplant mortality for kidney transplant patients. Examples of factors known to strongly affect the post-transplant mortality hazard include age, primary renal diagnosis, and pre-transplant time on dialysis; each of which may differ in distribution by center. Centers with mortality significantly greater than the national average may be subject to various degrees of intervention, including site visits and perhaps de-accreditation. Given the high stakes of such evaluations, it is important that the statistical methods used for identifying outlying centers be accurate.

A commonly used measure for evaluating center-specific survival is the SMR, 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. The SMR is a tool familiar to fields such as epidemiology (for example, [2–5]). The comparison of observed and expected outcomes is also commonly used for healthcare regulation (for example, [6, 7]). In the context of renal research, Wolfe *et al.* [8, 9] calculated SMRs among kidney transplant patients using mortality tables published by the United States Renal Data Systems. To relax the assumption of known standard mortality rates, Dickinson *et al.* [10] studied a semiparametric SMR based on Cox regression. Limitations of the SMR include its use of indirect standardization; each center's SMR is essentially adjusted to a different (center-specific) covariate and censoring distribution. Centers cannot be rank ordered on the basis of indirect standardization, because two centers with equal covariate-specific mortality hazards could have different SMRs, merely due to differences in their respective covariate or censoring distributions. Therefore, although SMRs are potentially useful for internal evaluation (e.g., for centers to evaluate themselves or for a governing body to evaluate this center's mortality comparing to that expected at the national level), they are less useful in the work of external evaluation (e.g., for surgeons and patients to compare center-specific results in the same region), because comparisons of center-specific results would play at least some role. We provide additional commentary on the SMR in Section 5.

Earlier, we note limitations in the SMR approaches. Naturally, this technique has its role in analysis contrasting centers. However, noting room for improvement, we propose semiparametric version of direct standardization useful for survival analysis of centers. Direct standardization is also a commonly used approach in comparisons of mortality, usually through a measure termed the standardized rate ratio (SRR) or comparative mortality figure (CMF). One can express the SRR as the ratio of expected to observed numbers of deaths in the whole study population; the numerator of the SRR represents the expected number of deaths if all patients were treated at the given center, while the denominator equals the total observed number of deaths in the study population. Breslow and Day [3] and Hazel [4] compared the SRR with SMR in the framework of person-year methods. In particular, the main drawback of SMR (with respect to comparisons across centers) is not inherent to the SRR, because the same standard population is applied to all centers. Hence, center-specific SRRs are directly comparable.

The SRR has a long history in fields such as epidemiology, and there are many settings in which direct standardization is appropriate. The Cox model has dominated applications involving regression analysis of censored data since its development. Thus, the use of Cox regression to estimate directly standardized center effects is a natural choice. The main contribution of this report is to formalize procedures for the Cox regression-based SRR, including rigorous derivation of asymptotic properties, simulation studies, and detailed comparisons to alternative approaches.

2. Methods

First, we provide the notation to be used in this article. Let T_i and C_i represent the survival and censoring time, respectively, for the ith patient, where $i = 1, \ldots, n$. Let J be the number of centers. The total number of subjects is denoted by $n = \sum_{j=1}^{J} n_j$, where n_j is the number of subjects in center j. Observation times are denoted by $X_i = T_i \wedge C_i$, with at-risk indicator $Y_i(t) = I(X_i \geq t)$, where $a \wedge b = \min\{a, b\}$ and $I(A)$ is an indicator function taking the value 1 when condition A holds and 0 otherwise. The observed death indicators are denoted by $\Delta_i = I(T_i \leq C_i)$, and the death counting process is defined as $N_i(t) = \Delta_i I(X_i \leq t)$. Let G_i denote the center for subject i and set $G_{ij} = I(G_i = j)$. Correspondingly, we set $Y_{ij}(t) = Y_i(t)G_{ij}$ and $N_{ij}(t) = N_i(t)G_{ij}$. The observed data consist of n independent vectors, $(X_i, \Delta_i, G_i, Z_i)$, where Z_i is a vector of adjustment covariates.

The assumed center-stratified Cox model can be formulated as

$$\lambda_{ij}(t) \equiv \lambda(t|Z_i, G_i = j) = \lambda_{0j}(t) \exp\left\{\beta_0^T Z_i\right\}, \tag{1}$$

where $\lambda_{0j}(t)$ is an unspecified center-specific baseline hazard function and β_0 is a parameter vector. The partial likelihood estimator [11] of β_0 is denoted by $\hat{\beta}$ and is given by the solution to $U(\beta) = 0$, where

$$U(\beta) = \sum_{j=1}^{J} \sum_{i=1}^{n} \int_0^\tau \left\{Z_{ij} - \overline{Z}_j(u;\beta)\right\} dN_{ij}(u)$$

with $\overline{Z}_j(u;\beta) = S_j^{(0)}(u;\beta)^{-1} S_j^{(1)}(u;\beta)$ and $S_j^{(d)}(u;\beta) = n^{-1} \sum_{i=1}^n Y_{ij}(u) Z_i^{\otimes d} \exp\{\beta^T Z_i\}$ for $d = 0, 1, 2$. The Breslow estimator [12] of $\Lambda_{0j}(t) = \int_0^t \lambda_{0j}(u)du$ is then given by $\hat{\Lambda}_{0j}\left(t;\hat{\beta}\right)$, where

$$\hat{\Lambda}_{0j}(t;\beta) = \frac{1}{n} \sum_{i=1}^{n} \int_0^t \frac{dN_{ij}(u)}{S_j^{(0)}(u;\beta)}.$$

2.1. *Indirect standardization: standardized mortality ratio*

We begin by introducing an alternative to the proposed measure. The SMR for center j is calculated as the ratio of observed to expected numbers of events

$$\widehat{SMR}_j(t) = \frac{O_j(t)}{E_j(t)}, \tag{2}$$

where the numerator is given by $O_j(t) = \sum_{i=1}^n N_{ij}(t)$ and the expected number of events is computed as

$$E_j(t) = \sum_{i=1}^{n} \int_0^t Y_{ij}(u) \exp\left\{\hat{\beta}^T Z_i\right\} d\hat{\Lambda}_0\left(u;\hat{\beta}\right),$$

with $\hat{\Lambda}_0\left(t;\hat{\beta}\right)$ representing an estimator of the average cumulative baseline hazard

$$\hat{\Lambda}_0(t;\hat{\beta}) = \frac{1}{n} \sum_{i=1}^{n} \int_0^t \frac{dN_i(u)}{S^{(0)}\left(u;\hat{\beta}\right)} \tag{3}$$

and $\hat{\beta}$ is based on model (1). The estimator given in (2) was developed in [13], where part of the innovation was the proposed use of the center-stratified model to estimate β_0. Intuition would suggest computing $E_j(t)$ using an unstratified model. However, estimation of β_0 in the absence of adjustment for center effects may produce a substantially biased estimate due to confounding by center or merely due to center affecting the hazard function, with or without confounding (i.e., due to the nonlinear link function).

The calculation of $\widehat{SMR}_j(t)$ involves two stages. A stratified Cox model (model (1)) is fitted in the first stage, with the population average cumulative baseline hazard, $\hat{\Lambda}_0\left(t;\hat{\beta}\right)$, and then computed in the second stage (e.g., through a unstratified Cox model with no covariates) using $\hat{\beta}' Z_i$ as an offset. At the second stage, $O_j(t)$, $E_j(t)$ are calculated.

2.2. *Direct standardization: standardized rate ratio*

On the basis of a form of indirect standardization, the SMR described in Section 2.1 can be viewed as a weighted ratio of center-specific cumulative hazards, with weight functions based on center-specific $S_j^{(0)}(t;\beta)$. These weight functions have an obvious disadvantage: they involve center-specific censoring and covariate distributions, which can differ considerably across centers. To rule out the possibility that differences among centers are due merely to different censoring and/or covariate distributions, the weight function should be specified such that differences among centers with respect to the resulting

measure are a function only of corresponding differences in center-specific hazards. Motivated by such considerations, we propose an alternative method, referred to as the SRR, which can be interpreted as a semiparametric version of direct standardization. The proposed SRR is computed, for center j, as

$$\widehat{SRR}_j(t) = \frac{\mathcal{E}_j(t)}{O(t)}, \tag{4}$$

with $O(t) = \sum_{j=1}^{J} O_j(t)$ being the total observed number of deaths across all centers and

$$\mathcal{E}_j(t) = \sum_{\ell=1}^{J}\sum_{i=1}^{n}\int_0^t Y_{i\ell}(u)\exp\left\{\hat{\beta}^T Z_i\right\} d\hat{\Lambda}_{0j}(u) \tag{5}$$

representing the expected number of total deaths if all centers had mortality hazard equal to that of center j. Similar to the SMR, the SRR is easily interpreted and is well understood by clinical investigators. The SRR also involves a ratio of observed and expected numbers of deaths. However, the 'expected' component is in the SRR's numerator, while the 'observed' count is in the denominator. With respect to interpretation, $SRR_j > 1$ indicates that center j has a greater mortality rate than the overall average. Note that, although $SRR_j(t)$ also involves the censoring and covariate distributions, the same weight function is applied across all centers, thus factoring out the impact of imbalances in center-specific censoring and covariate distributions. The proposed measures are desirable in this light, because their center-specific limiting values would differ only due to corresponding differences in center-specific hazards.

2.3. *Asymptotic properties*

We summarize the asymptotic properties of the proposed SRR with the following theorem; we outline the proof in Appendix A.

Theorem 1
Under the regularity conditions listed in Appendix A, $\widehat{SRR}_j(t)$ converges in probability to $SRR_j(t)$ uniformly in $t \in [0, \tau]$, where

$$SRR_j(t) = \frac{\sum_{\ell=1}^{J}\int_0^t s_\ell^{(0)}(u;\beta_0)d\Lambda_{0j}(u)}{\sum_{\ell=1}^{J}\int_0^t s_\ell^{(0)}(u;\beta_0)d\Lambda_{0\ell}(u)}$$

and $n^{\frac{1}{2}}\left\{\widehat{SRR}_j(t) - SRR_j(t)\right\}$ converges weakly to a zero-mean Gaussian process with variance function $\sigma_j(t) = E[\xi_{ij}(t;\beta_0)^2]$, where

$$\xi_{ij}(t;\beta) = w(t;\beta)\int_0^t \frac{s^{(0)}(u;\beta)}{s_j^{(0)}(u;\beta)} dM_{ij}(u;\beta) + w(t;\beta)\int_0^t \left\{Y_i(u)\exp(\beta^T Z_i) - s^{(0)}(u;\beta)\right\} d\Lambda_{0j}(u) \tag{6}$$

$$- w(t;\beta)^2\int_0^t s^{(0)}(u;\beta)d\Lambda_{0j}(u)\sum_{\ell=1}^{J}\int_0^t \left\{dN_{i\ell}(u) - s_\ell^{(0)}(u;\beta)d\Lambda_{0\ell}(u)\right\} \tag{7}$$

$$+ w(t;\beta)\int_0^t r_j^T(u;\beta)d\Lambda_{0j}(u)\Omega(\beta)^{-1}\int_0^\tau \{Z_{i\ell} - \overline{z}_\ell(u;\beta)\}\, dM_{ij}(u;\beta) \tag{8}$$

with

$$w(t;\beta) = \left\{\sum_{\ell=1}^{J}\int_0^t s_\ell^{(0)}(u;\beta)d\Lambda_{0\ell}(u)\right\}^{-1}$$

$$r_j(u;\beta) = s^{(1)}(u;\beta) - s^{(0)}(u;\beta)\,\overline{z}_j(u;\beta).$$

The variance function can be consistently estimated by $\hat{\sigma}_j^2(t;\hat{\beta}) = n^{-1}\sum_{i=1}^{n}\hat{\xi}_{ij}(t;\hat{\beta})^2$, with $\hat{\xi}_{ij}(t;\hat{\beta})$ obtained by replacing limiting values in $\xi_{ij}(t;\beta)$ with their empirical counterparts. With respect to this variance formula, a potentially time-saving strategy is to treat the estimators of β_0 as constants and hence ignore their variability. A justification for such simplification applies when the total study population is very large, such that $\hat{\beta}$ has little variability. The resulting variance estimator is given by

$$\xi_{ij}^R(t;\beta) = w(t;\beta)\int_0^t \frac{s^{(0)}(u;\beta)}{s_j^{(0)}(u;\beta)} dM_{ij}(u;\beta) + w(t;\beta)\int_0^t \left\{Y_i(u)\exp(\beta^T Z_i) - s^{(0)}(u;\beta)\right\} d\Lambda_{0j}(u) \tag{9}$$

$$- w(t;\beta)^2 \int_0^t s^{(0)}(u;\beta) d\Lambda_{0j}(u) \sum_{\ell=1}^{J}\int_0^t \left\{dN_{i\ell}(u) - s_\ell^{(0)}(u;\beta) d\Lambda_{0\ell}(u)\right\}, \tag{10}$$

obtained by removing the line (8) in the formula of $\xi_{ij}(t;\beta)$. The quantity $\hat{\xi}_{ij}^R\left(t;\hat{\beta}\right)$ is obtained by replacing limiting values in $\xi_{ij}^R(t;\beta)$ with their empirical counterparts.

3. Simulation

We evaluate the finite-sample properties of the estimators described in Section 2 through a series of simulation studies. Death times were generated from the Weibull model, $\lambda_{ij}(t) = \alpha_j\gamma_j t^{\gamma_j - 1}\exp(\beta^T Z_i)$ for $i = 1,\ldots,n_j$ and $j = 1,\ldots,10$, where $Z_i = (Z_{i1}, Z_{i2}, Z_{i3})^T$. We set $\boldsymbol{\beta}_0^T = (\beta_1,\beta_2,\beta_3) = (0.02, -0.5, 0.2)$. There are $J = 10$ centers. The number of subjects within each center varied under different scenarios. Censoring times were generated from either a Uniform distribution or an exponential distribution. In order to compare direct standardization with indirect standardization in the framework of semiparametric models, SMR and SRR were calculated at $t = 1, t = 2$, and $t = 3$. Each data configuration was replicated 1000 times.

3.1. Setting 1: center-independent hazards

The first simulation setting considered the case where the hazard functions are equal across centers for all $t \in [0,\tau]$. The censoring and covariate distributions were chosen to be center-independent. Specifically, censoring times were generated from a Uniform (0.5, 10) distribution; Z_{i1} followed a Bernoulli (0.5) distribution; Z_{i2} followed a logistic distribution with probability dependent on Z_{i1}; and Z_{i3} came from a Normal distribution with constant variance 25 and mean dependent on Z_{i1} and Z_{i2} (e.g., $E[Z_{i3}|Z_{i1},Z_{i2}] = 50 + 0.2\,Z_{i1} - 0.5\,Z_{i2}$). Under this setting, the hazard functions are equal across centers, such that the limiting values of SRR equal 1 for each center. Results at time $t = 3$ are displayed in Table I. For all centers, the average estimated SRR was very close to 1. The average asymptotic standard

Table I. Simulation setting 1: $\hat{SRR}_j(t)$; center-independent hazards, covariate and censoring distributions; $t = 3$.

					Theorem 1 (6)–(8)		Theorem 1 (approx) (9), (10)	
Center	(γ_j,α_j)	True	Bias	ESD	ASE	CP	ASE	CP
1	(1, 0.2)	1.000	0.001	0.154	0.153	0.96	0.153	0.96
2	(1, 0.2)	1.000	0.004	0.160	0.154	0.94	0.153	0.94
3	(1, 0.2)	1.000	−0.000	0.151	0.153	0.95	0.153	0.95
4	(1, 0.2)	1.000	−0.006	0.163	0.153	0.93	0.153	0.93
5	(1, 0.2)	1.000	−0.007	0.149	0.153	0.95	0.153	0.95
6	(1, 0.2)	1.000	−0.003	0.155	0.153	0.94	0.153	0.94
7	(1, 0.2)	1.000	0.007	0.162	0.154	0.93	0.154	0.93
8	(1, 0.2)	1.000	−0.001	0.152	0.153	0.95	0.153	0.95
9	(1, 0.2)	1.000	0.002	0.158	0.154	0.95	0.152	0.95
10	(1, 0.2)	1.000	0.002	0.157	0.153	0.94	0.153	0.94

ESD, empirical standard deviation; ASE, asymptotic standard errors; CP, coverage probability.

errors were generally close to the empirical standard deviations, while the empirical coverage probabilities (CP) were generally consistent with the nominal value. This held for both the variance estimator derived in Theorem 1 (with corresponding asymptotic expansion given by (6)–(8) and its approximation given by (9) and (10). In fact, the two sets of asymptotic standard errors are virtually indistinguishable. Results were generally consistent across the observation time distribution (data not shown).

It is worth noting that whether or not the covariate or censoring distributions were center dependent has no influence on the results in this setting. Correspondingly, similar results were found for the case of center-dependent covariate and censoring distribution.

3.2. *Setting 2: center-dependent hazards; center-independent covariate and censoring distributions*

For the second set of simulations, different values of α_j and γ_j were used, such that the hazard functions increased with increasing center number ($j = 1, \ldots, 10$). The covariate distributions were chosen to be center independent and were generated from the same distributions from setting 1. Results are provided in Table II for various sample sizes and censoring percentages. The proposed SRR appears to be approximately unbiased, with CP close to 0.95 for both the standard error based on Theorem 1 and that based on the approximation. When the center-specific sample size is small (e.g., 25), the empirical CPs were

Table II. Simulation setting 2: $\hat{SRR}_j(t)$; center-dependent hazards, center-independent covariate and censoring distributions; $t = 3$.

							Theorem 1 (6)–(8)		Theorem 1 (approx) (9), (10)	
n_j	Censoring	Center	(γ_j, α_j)	True	Bias	ESD	ASE	CP	ASE	CP
100	20%	2	(0.85, 0.08)	0.406	0.004	0.096	0.093	0.94	0.093	0.94
		4	(0.95, 0.16)	0.797	0.003	0.132	0.134	0.95	0.134	0.95
		6	(1.05, 0.24)	1.180	0.001	0.174	0.169	0.94	0.169	0.94
		8	(1.15, 0.32)	1.549	0.006	0.206	0.204	0.95	0.203	0.95
		10	(1.25, 0.4)	1.912	0.005	0.230	0.238	0.96	0.237	0.96
50	20%	2	(0.85, 0.08)	0.406	0.002	0.135	0.130	0.93	0.130	0.93
		4	(0.95, 0.16)	0.797	−0.008	0.183	0.187	0.95	0.186	0.95
		6	(1.05, 0.24)	1.180	0.007	0.252	0.238	0.94	0.238	0.94
		8	(1.15, 0.32)	1.549	0.001	0.290	0.285	0.95	0.284	0.95
		10	(1.25, 0.4)	1.912	0.026	0.344	0.338	0.95	0.337	0.94
100	40%	2	(0.85, 0.08)	0.406	0.008	0.097	0.097	0.94	0.097	0.94
		4	(0.95, 0.16)	0.797	0.003	0.139	0.138	0.95	0.138	0.95
		6	(1.05, 0.24)	1.180	0.007	0.172	0.174	0.95	0.173	0.95
		8	(1.15, 0.32)	1.549	0.014	0.214	0.208	0.94	0.208	0.94
		10	(1.25, 0.4)	1.912	0.029	0.243	0.244	0.95	0.243	0.95
50	40%	2	(0.85, 0.08)	0.406	0.009	0.137	0.136	0.94	0.135	0.94
		4	(0.95, 0.16)	0.797	0.011	0.204	0.195	0.94	0.195	0.94
		6	(1.05, 0.24)	1.180	−0.007	0.249	0.242	0.94	0.241	0.94
		8	(1.15, 0.32)	1.549	0.019	0.302	0.294	0.94	0.294	0.94
		10	(1.25, 0.4)	1.912	0.043	0.357	0.345	0.95	0.344	0.95
125	20%	2	(0.85, 0.08)	0.529	0.006	0.110	0.106	0.94	0.106	0.94
100		4	(0.95, 0.16)	1.043	−0.004	0.169	0.171	0.95	0.171	0.95
75		6	(1.05, 0.24)	1.550	−0.002	0.256	0.255	0.95	0.255	0.95
50		8	(1.15, 0.32)	2.025	−0.015	0.384	0.381	0.93	0.380	0.93
25		10	(1.25, 0.4)	2.535	−0.004	0.651	0.632	0.93	0.630	0.92
125	40%	2	(0.85, 0.08)	0.601	0.001	0.122	0.118	0.94	0.118	0.94
100		4	(0.95, 0.16)	1.118	0.007	0.190	0.192	0.96	0.192	0.96
50		6	(1.05, 0.24)	1.773	−0.001	0.375	0.359	0.93	0.358	0.93
25		8	(1.15, 0.32)	2.324	0.045	0.636	0.621	0.93	0.619	0.93
15		10	(1.25, 0.4)	2.900	−0.062	1.034	0.914	0.91	0.910	0.90

ESD, empirical standard deviation; ASE, asymptotic standard errors; CP, coverage probability.

slightly underestimated. Such results suggest that center-specific sample sizes play an important role in the proposed methods. In particular, the minimum sample size (across centers) needs to be reasonably large. Given this concern, centers of size less than 20 were eliminated from the real data analysis in Section 4. Collectively, simulation results from settings 1 and 2 indicate that the proposed method is quite accurate at and away from the null.

3.3. *Setting 3: center-dependent hazards, center-dependent censoring distribution*

As mentioned previously, estimators based on indirect standardization may be misleading if either the censoring or the covariate distributions are center-dependent. To illustrate this point, we performed simulations with the following three conditions: (i) the hazard functions were center dependent, while center $2j-1$ had exactly the same hazard function as that of center $2j$ for $j=1,\ldots,5$; (ii) the censoring distributions for centers $2j-1$ and $2j$ were substantially different. For center $2j-1$, the censoring times were generated from a uniform distribution such that the censoring mainly occurred in the later stages, while for center $2j$, the censoring times were generated from a uniform distribution for which the censoring tended to occur in the early stages; and (iii) the covariate distributions (reused from setting 2) were center independent.

We compared the SRR and SMR in Table III for setting 3. With respect to the true values, the limiting values of $\widehat{SRR}_{2j}(t)$ and $\widehat{SRR}_{2j-1}(t)$ are equal, as one would hope. This is not the case for $\widehat{SMR}_{2j}(t)$ and $\widehat{SMR}_{2j-1}(t)$, the differences being due to differences in the censoring distributions. With respect to the estimators themselves, both $\widehat{SRR}_j(t)$ and $\widehat{SMR}_j(t)$ are approximately unbiased. Note that the bias of $\widehat{SMR}_j(t)$ was calculated as the difference between it and its own limiting value.

3.4. *Setting 4: center-dependent hazards; center-dependent covariate distributions*

We also performed simulations under a setting in which the distribution of the covariate vector differed by center. The setup for hazard functions and censoring distribution was the same as in setting 2, while centers $2j$ and $2j-1$ had substantially different covariate distributions. In center $2j-1$, the covariate Z_{i1} followed a Bernoulli (0.2) distribution, Z_{i2} followed a Bernoulli (0.8) distribution, and Z_{i3} came from a Normal distribution with mean 30 and standard deviation 10. In center $2j$, Z_{i1} followed a Bernoulli (0.8)

Table III. Simulation setting 3: centers $2j$ and $2j-1$ have equal hazards but different censoring distributions; $t=3$.

Measure	Center	(γ_j, α_j)	True	Bias	ESD
$\text{SMR}_j(t)$	1	(0.8, 0.04)	**0.227**	−0.003	0.089
	2	(0.8, 0.04)	**0.211**	−0.002	0.068
	3	(0.9, 0.12)	**0.658**	0.001	0.162
	4	(0.9, 0.12)	**0.637**	0.003	0.121
	5	(1, 0.2)	1.050	0.003	0.208
	6	(1, 0.2)	1.051	−0.000	0.157
	7	(1.1, 0.3)	**1.500**	0.002	0.239
	8	(1.1, 0.3)	**1.547**	0.004	0.201
	9	(1.25, 0.4)	**1.881**	−0.002	0.279
	10	(1.25, 0.4)	**2.003**	−0.008	0.219
$\text{SRR}_j(t)$	1	(0.8, 0.04)	0.221	−0.003	0.088
	2	(0.8, 0.04)	0.221	−0.002	0.073
	3	(0.9, 0.12)	0.645	0.002	0.162
	4	(0.9, 0.12)	0.645	0.004	0.124
	5	(1, 0.2)	1.044	0.002	0.216
	6	(1, 0.2)	1.044	−0.002	0.158
	7	(1.1, 0.3)	1.530	0.001	0.277
	8	(1.1, 0.3)	1.530	0.004	0.202
	9	(1.25, 0.4)	1.996	−0.004	0.351
	10	(1.25, 0.4)	1.996	−0.008	0.230

ESD, empirical standard deviation.

Table IV. Simulation setting 4: centers $2j$ and $2j-1$ have equal hazards but different covariate distributions; $t = 3$.

Measure	Center	(γ_j, α_j)	True	Bias	ESD
$SMR_j(t)$	1	(0.8, 0.04)	**0.319**	−0.003	0.089
	2	(0.8, 0.04)	**0.331**	−0.002	0.068
	3	(0.9, 0.12)	**0.931**	0.001	0.162
	4	(0.9, 0.12)	**0.919**	0.003	0.121
	5	(1, 0.2)	**1.484**	0.003	0.208
	6	(1, 0.2)	**1.307**	−0.000	0.157
	7	(1.1, 0.3)	**2.070**	0.002	0.239
	8	(1.1, 0.3)	**1.602**	0.004	0.201
	9	(1.25, 0.4)	**2.551**	−0.002	0.279
	10	(1.25, 0.4)	**1.725**	−0.008	0.219
$SRR_j(t)$	1	(0.8, 0.04)	0.345	−0.002	0.120
	2	(0.8, 0.04)	0.345	−0.000	0.041
	3	(0.9, 0.12)	0.929	0.005	0.212
	4	(0.9, 0.12)	0.929	0.003	0.105
	5	(1, 0.2)	1.386	−0.007	0.272
	6	(1, 0.2)	1.386	−0.000	0.186
	7	(1.1, 0.3)	1.908	−0.007	0.344
	8	(1.1, 0.3)	1.908	0.013	0.330
	9	(1.25, 0.4)	2.329	−0.006	0.412
	10	(1.25, 0.4)	2.329	−0.001	0.455

ESD, empirical standard deviation.

distribution, Z_{i2} followed a Bernoulli (0.2) distribution, and Z_{i3} was derived from a Normal distribution with mean 50 and standard deviation 10.

Results based on setting 4 are given in Table IV. Trends are similar to those from setting 3 but much more pronounced. The limiting values of $\widehat{SRR}_{2j}(t)$ are equal to those of $\widehat{SRR}_{2j-1}(t)$, as one would expect. Conversely, the true values of $\widehat{SMR}_{2j}(t)$ and $\widehat{SMR}_{2j-1}(t)$ are different, with the differences being quite pronounced for $j = 3$, $j = 4$, and $j = 5$. Moreover, it appears that $SMR_7(t) > SMR_{10}(t)$, which is misleading in the sense that $SRR_7(t) < SRR_{10}(t)$.

4. Application

We applied the proposed methods to investigate the performance of transplant centers with respect to post kidney transplant survival. Data were obtained from the Scientific Registry of Transplant Recipients (SRTR) and submitted by members of the Organ Procurement and Transplantation Network. The SRTR database contains information on all wait-listed candidates, transplant recipients, and organ donors in the USA. Included in the analysis were adult patients (≥18 years of age at transplant) who underwent deceased donor kidney transplantation between January 2000 and December 2008. Adjustment covariates in this study included age, race, gender, diagnosis, donation after cardiac death, expanded criteria donor, body mass index, dialysis time, indicator of previous kidney transplant, and cold ischemia time. These variables have face validity from a clinical perspective and are based on a list of covariates used in SRTR. Transplant centers with sample size ≤20 and patients who received a living-donor transplant were eliminated from additional analysis. The final sample size was then $n = 74,088$ from $J = 217$ centers across the USA. Failure time (recorded in years) was defined as the time from transplantation to graft failure or death, whichever occurred first. Graft failure was considered to occur when the transplanted kidney ceased to function.

Stratified Cox regression was employed to model the hazard function. The indirectly standardized estimator, $\widehat{SMR}_j$, was calculated using SAS PROC PHREG with an offset. The proposed directly standardized estimator, $\widehat{SRR}_j$, was computed using SAS IML. Figure 1a represents the pairwise comparisons of the SMRs and SRRs. Figure 1b shows the standard error of these two measures. Figure 1c compares

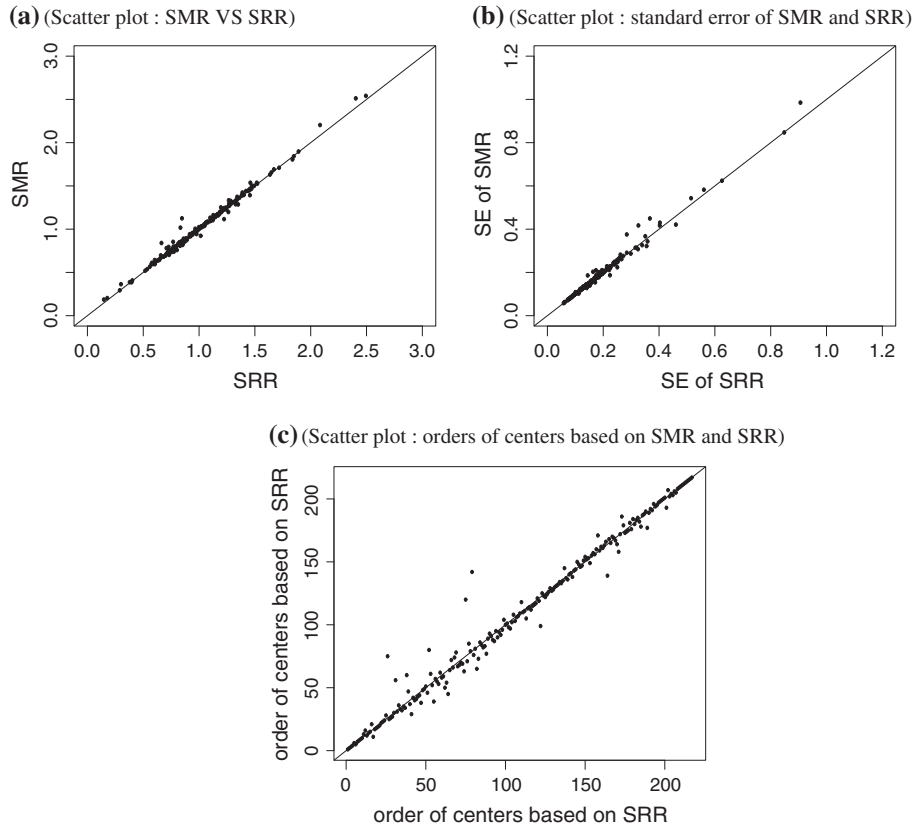

Figure 1. Evaluation of $J = 217$ kidney transplant centers. SMR, standardized mortality ratio; SRR, standardized rate ratio; SE, standard error.

the orders of centers based on SRRs and SMRs. As shown, there are some discrepancies between these two measures. We applied bootstrapped techniques to evaluate whether the change in center-specific orderings (SMR versus SRR) exceeded that attributable to only sampling variation. Specially, we calculated the distribution of center-specific SMR orderings from 100 bootstrapped samples. We then calculated the 95% confidence intervals of these orders. On the basis of the bootstrapped resamples, for 20 of 217 centers, the 95% confidence intervals based on SMR does not cover the order based on SRR from the original dataset. Among these 20 centers, 10 had SRR significantly different from the national average.

Using the asymptotic normality of the proposed estimators, we constructed the point-wise confidence intervals for SRR at $t = 5$ years (Figure 2). Center numbers are re-ordered by values of SRR. A total of 38 centers had observed number of events significantly lower than the expected calculation based on the national average hazards, while 28 centers were significantly above the expected. It is clear that the hazard functions varied among centers. Table V presents the pairwise comparison of the numbers and percentages of 'outlier' centers identified by p-values corresponding to their SMRs and SRRs (using tests of $H_0 : SMR_j = 1$ and $H_0 : SRR_j = 1$, respectively). A total of six centers changed 'memberships' based on these two measures. Specifically, two centers were flagged to be significant based on SMR but flagged to be normal based on SRR; on the other hand, four centers were flagged to be significant based on SRR but flagged to be normal based on SMR. Through fitting a sequence of logistic regression models (rotating the center indicators as the response variates), it was revealed that, for these six centers, approximately half of the adjustment covariates had distributions significantly different from the remaining centers. In addition, through fitting a Cox model using censored as the event, three of the six centers in question were significantly predictive of the censoring hazard. In summary, we do observe some differences when comparing center effects estimated through direct versus indirect standardization, and the strongest examples of such discrepancies appear to be due to differences in the center-specific covariate and censoring distributions, consistent with the concepts described earlier in this report.

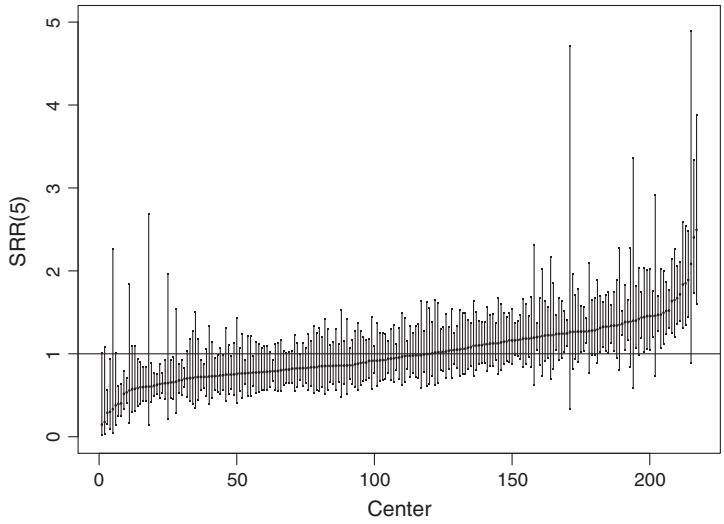

Figure 2. Evaluation of $J = 217$ kidney transplant centers: point estimates and 95% confidence interval of $\hat{SRR}_j$ at $t = 5$ years. SRR, standardized rate ratio.

Table V. Number and percentage of centers giving significant results under SMR and SRR.

	SRR		
SMR	Nonsignificant	Significant	Row sum
Nonsignificant	147 (67.8%)	4 (1.8%)	151 (69.6%)
Significant	2 (0.9%)	64 (29.5%)	66 (30.4 %)
Column sum	149 (68.7%)	68 (31.3%)	217 (100%)

SMR, standardized mortality ratio; SRR, standardized rate ratio.

5. Discussion

We propose semiparametric methods for estimating standardized rate ratios, as a means of evaluating center-specific mortality through direct standardization. Large-sample properties are derived and shown through simulation to be appropriate in finite samples. A computationally faster variance estimator is proposed for the SRR and is shown to work practically as well as the full version. Application of the methods demonstrates several significant differences among kidney transplant centers in the USA.

There is some judgement required in deciding when to use indirect standardization and when to use direct standardization. Indirectly standardized estimators, such as SMR, provide a valid approach to evaluate how does a center's mortality compare to that predicted at the population level for the kinds of patients at this center. However, it is important to emphasize that center-specific SMRs should not be compared with one another (a caution that applies to all indirectly standardized rates). The SRR, a directly standardized measure, does not share this drawback. The SRRs for two given centers will be unequal only because the center-specific mortality hazards differ; direct standardization accounts for imbalance with respect to center-specific covariate and censoring distributions. The proposed SRR shares the SMR's ease of interpretation but rectifies its key disadvantages and, hence, is a more appropriate choice in settings where mortality comparisons across centers are an objective.

The degree to which the SMR and SRR are different will depend on the application. In some settings, the two may not agree well, while in others, they may be quite similar. The only way to know with certainty if SMR and SRR are equal would be to calculate both measures, which would not be a desirable option in many cases. In settings where, for a particular center, the SMR and SRR were unequal, it would be very difficult to claim that the SMR was correct, for the several reasons documented previously. Given the high stakes of evaluations by regulatory bodies, and the fact that the credibility of such organizations depends in part on the accuracy of their evaluations, it would appear the preferred analysis is the one that is most likely to be accurate.

The proposed SRR is computed using a stratified Cox model, which makes no assumptions about the functional form of the impact of center on the hazard function. The stratification by center plays a major role. For instance, the expected number of deaths considers each patient's covariate vector, such that the regression parameter must be estimated consistently. Unless the death hazard is conditionally independent of center given the covariates, covariate effect estimates will generally be biased if based on a model with a nonlinear link function and no accounting for center. This is an issue for the SMR as well, particularly because the Cox version of the SMR has historically been computed using a common-baseline (i.e., unstratified) model. He [13] proposed modifying the SMR through stratification, leading to the quantity we denoted by SMR and used in simulations (Section 3) for comparisons with the SRR. Such properties were demonstrated empirically by He [13]; the magnitude of the bias (in the case of unstratified Cox models) increases when covariates are also center-dependent.

Random effects models are an option for contrasting centers. Moreover, a Bayesian formulation may be an attractive approach for center effect studies. The advantage for random effects model is that this approach would allow for the inclusion of even very small centers. In contrast, when the number of events for a center is small, the estimated center parameter from a fixed effect model may be unstable. However, Kalbfleish and Wolfe [14] compared the properties of a fixed effect model (FEM) and a random effect model (REM) for the purpose of profiling kidney dialysis facilities under various conditions. Essentially, the REM estimates are shrunk toward overall mean and hence reduce the reported variation of facility performance. Second, the FEM has the highest statistical power to identify exceptional facilities, for a given false positive rate; identifying such extreme facilities is usually a main objective of center evaluations. Another issue for REM is the potential confounding effects when the patient risks are correlated with center effects. These findings suggest that a simple REM method may not be good enough and more sophisticated approaches are necessary. Further discussion of such issues is provided by Ohlssen *et al.* [15], who develop a more flexible random effects model using Bayesian nonparametric methods, in order to remedy the influence of outlying centers to which basic random effects models are susceptible.

The direct standardization methods derived in this report could be extended in several useful directions. Perhaps most notably, it is often of interest to evaluate center effects in settings where the event of interest is recurrent (e.g., hospitalizations and infections). Furthermore, it would also be useful to develop methods based on direct standardization that can accommodate competing risks or dependent censoring. Direct standardization could also be applied to compare center-specific survival probability and restricted mean lifetime.

Appendix A

To derive the large-sample properties for the SRR, we impose the following regularity conditions under the stratified Cox model:

(a) $(X_i, \Delta_i, G_i, Z_i)$ are independent and identically distributed random vectors.
(b) $P(X_i \geqslant \tau) > 0$ where τ is a pre-specified time point.
(c) Z_{ik} have bounded total variation, that is, $|Z_{ik}| < \kappa$ for all $i = 1, \ldots n$ and $k = 1, \ldots, p$, where κ is a constant and Z_{ik} is the kth component of Z_i.
(d) $\int_0^\tau \lambda_{0j}(t)dt < \infty$.
(e) Continuity of the following functions:

$$s_j^{(1)}(t;\beta) = \frac{\partial}{\partial\beta} s_j^{(0)}(t;\beta),\ s_j^{(2)}(t;\beta) = \frac{\partial^2}{\partial\beta\partial\beta^T} s_j^{(0)}(t;\beta)$$

and $s_j^{(0)}(t;\beta)$, where $s_j^{(d)}(t;\beta)$ is the limiting value of $S_j^{(d)}(t;\beta)$ for $d = 0, 1, 2$, with $s_j^{(1)}(t;\beta)$ and $s_j^{(2)}(t;\beta)$ bounded and $s_j^{(0)}(t;\beta)$ bounded away from 0 for $t \in [0, \tau]$.
(f) Positive-definiteness of the matrix $\Omega_j(\beta)$:

$$\Omega_j(\beta) = \int_0^\tau v_j(t;\beta) s_j^{(0)}(t;\beta)\lambda_{0j}(t)dt,$$

$$v_j(t;\beta) = \frac{s_j^{(2)}(t;\beta)}{s_j^{(0)}(t;\beta)} - \overline{z}_j(t;\beta)^{\otimes 2},$$

where $\overline{z}_j(t;\beta) = s_j^{(0)}(t;\beta)^{-1} s_j^{(1)}(t;\beta)$ is the limiting value of $\overline{Z}_j(t;\beta)$.

(g) $P(G_{ij} = 1|Z_i) > 0$.

Condition (a) is employed in the derivation of the weak convergence. Condition (b) is a standard identifiability requirement. Condition (c) leads to the boundedness of several quantities and is applicable in most practical applications. Conditions (d) and (e) are not essential but simplify our proofs. With respect to condition (g), the selection probability given covariates is nonzero for all centers. This condition guarantees that the sample size n_j of each center goes to ∞ as the total sample size n goes to ∞.

We first show that $\hat{SRR}_j(t) \xrightarrow{P} SRR_j(t)$ uniformly for $t \in [0,\tau]$. The triangle inequality leads to

$$\begin{aligned}
&\left\| \int_0^t S_\ell^{(0)}\left(u;\hat{\beta}\right) d\hat{\Lambda}_{0j}\left(u;\hat{\beta}\right) - \int_0^t s_\ell^{(0)}(u;\beta_0) d\Lambda_{0j}(u) \right\| \\
\leq &\left\| \int_0^t S_\ell^{(0)}\left(u;\hat{\beta}\right) d\hat{\Lambda}_{0j}\left(u;\hat{\beta}\right) - \int_0^t s_\ell^{(0)}(u;\beta_0) d\hat{\Lambda}_{0j}\left(u;\hat{\beta}\right) \right\| \qquad (A.1) \\
+ &\left\| \int_0^t s_\ell^{(0)}(u;\beta_0) d\hat{\Lambda}_{0j}\left(u;\hat{\beta}\right) - \int_0^t s_\ell^{(0)}(u;\beta_0) d\Lambda_{0j}(u) \right\| \qquad (A.2)
\end{aligned}$$

To show that $(A.1) \xrightarrow{P} 0$ uniformly in t, recall that $S_\ell^{(0)}(t;\beta) = n^{-1}\sum_{i=1}^n Y_{i\ell}(t)\exp(\beta^T Z_i) = \mathbb{P}_n[I(X \geqslant t)I(G = \ell)\exp(\beta^T Z)]$, where $\mathbb{P}_n$ is the empirical measure; that is, $\mathbb{P}_n[I(X \geqslant t)I(G = \ell)\exp(\beta^T Z)] = n^{-1}\sum_{i=1}^n Y_{i\ell}(t)\exp(\beta^T Z)$. The collection of all cells, $[t,\infty)$, in the real line is a VC class of index 2 and, hence, satisfies the entropy conditions for the Glivenko–Cantelli theorem [16]. The boundedness conditions ensures that $\{I(X \geqslant t)I(G = \ell)\exp(\beta^T Z), t \in [0,\tau]\}$ belong to some Glivenko–Cantelli class; that is, $S_\ell^{(0)}(t;\hat{\beta}) \xrightarrow{a.s.} s_\ell^{(0)}(u;\hat{\beta})$ uniformly in $t \in [0,\tau]$. Next, for β, such that $\hat{\beta} \xrightarrow{P} \beta_0$ (e.g, [17, 18]), the bounded condition of $\{\hat{\Lambda}_0(t), t \in [0,\tau]\}$ and an application of the dominant convergence theorem entails that $(A.1) \xrightarrow{P} 0$ uniformly in t. Similarly, $(A.2) \xrightarrow{P} 0$ uniformly in t. We already demonstrate that $\int_0^t S_\ell^{(0)}(u;\beta_0)d\hat{\Lambda}_{0j}(u) \xrightarrow{P} \int_0^t s_\ell^{(0)}(u;\beta_0)d\Lambda_{0j}(u)$ uniformly for $t \in [0,\tau]$. Similarly, $\int_0^t S_\ell^{(0)}(u;\beta_0)d\hat{\Lambda}_{0\ell}(u) \xrightarrow{P} \int_0^t s_\ell^{(0)}(u;\beta_0)d\Lambda_{0\ell}(u)$ uniformly for $t \in [0,\tau]$. The monotonicity and boundedness conditions ensure that $\left\{\int_0^t S_\ell^{(0)}(u;\beta_0)d\hat{\Lambda}_{0\ell}(u)\right\}^{-1} \xrightarrow{P} \left\{\int_0^t s_\ell^{(0)}(u;\beta_0)d\Lambda_{0\ell}(u)\right\}^{-1}$ uniformly for $t \in [0,\tau]$. Therefore, we have that $\hat{SRR}_j(t) \xrightarrow{P} SRR_j(t)$ uniformly for $t \in [0,\tau]$.

To prove weak convergence, we use the following decomposition:

$$\begin{aligned}
&n^{\frac{1}{2}}\left\{\widehat{SRR}_j(t) - SRR_j(t)\right\} \\
= &n^{\frac{1}{2}}\left\{\frac{\int_0^t S_\ell^{(0)}(u;\hat{\beta})d\hat{\Lambda}_{0j}(u;\hat{\beta})}{\int_0^t S_\ell^{(0)}(u;\hat{\beta})d\hat{\Lambda}_{0\ell}(u)} - \frac{\int_0^t s_\ell^{(0)}(u;\beta_0)d\Lambda_{0j}(u;\hat{\beta})}{\int_0^t S_\ell^{(0)}(u;\beta_0)d\hat{\Lambda}_{0\ell}(u;\hat{\beta})}\right\} \qquad (A.3) \\
+ &n^{\frac{1}{2}}\left\{\frac{\int_0^t s_\ell^{(0)}(u;\beta_0)d\Lambda_{0j}(u)}{\int_0^t S_\ell^{(0)}(u;\hat{\beta})d\hat{\Lambda}_{0\ell}(u;\hat{\beta})} - \frac{\int_0^t s_\ell^{(0)}(u;\beta_0)d\Lambda_{0j}(u)}{\int_0^t s_\ell^{(0)}(u;\beta_0)d\Lambda_{0\ell}(u)}\right\}. \qquad (A.4)
\end{aligned}$$

First, we have that

$$\begin{aligned}
(A.3) &= w(t,\beta_0)\, n^{\frac{1}{2}}\left\{\int_0^t S^{(0)}(u;\hat{\beta})\frac{\sum_{i=1}^n dN_{ij}(u)}{nS_j^{(0)}(u;\hat{\beta})} - \int_0^t s^{(0)}(u;\beta_0)\frac{s_\ell^{(0)}(u;\beta_0)}{s^{(0)}(u;\beta_0)}d\Lambda_{0\ell}(u)\right\} + o_p(1) \\
&= w(t;\beta_0)\left[n^{-\frac{1}{2}}\sum_{i=1}^n\int_0^t \frac{s^{(0)}(u;\beta_0)}{s_j^{(0)}(u;\beta_0)}\left\{dN_{ij}(u;\beta_0) - s_j^{(0)}(u;\beta_0)d\Lambda_{0j}(u)\right\}\right. \\
&\qquad \left. + \int_0^t r_j^T(u;\beta)d\Lambda_{0j}(u)\, n^{\frac{1}{2}}(\hat{\beta}-\beta)\right\} + o_p(1),
\end{aligned}$$

where $w(t;\beta)$ and $r_j(u;\beta)$ are defined in Section 2. Note that the second equality of the aforementioned argument is obtained through the functional delta method and Lemma 19.24 of [19]. On the basis of previously established empirical process theory for the Cox model (e.g., [17]), we have that

$$n^{\frac{1}{2}}(\hat{\beta}-\beta_0)=\left\{\sum_{\ell=1}^{J}\Omega_\ell(\beta_0)^{-1}\right\}G_n\left[\sum_{\ell=1}^{J}\{\{Z-\overline{z}_\ell(X;\beta_0)\}\Delta I(G=\ell)\right.$$
$$\left.-\int_0^\tau\{Z-\overline{z}_\ell(u;\beta_0)\}\exp\left(\beta_0^TZ\right)I(X\geqslant u)I(G=\ell)d\Lambda_{0\ell}(u)\right]+o_p(1),$$

where the $o_p(1)$ is uniform in t, and $\mathbb{G}_n$ is the empirical process defined by $\mathbb{G}_nf=\sqrt{n}(\mathbb{P}_n-\mathbb{P})f$. Through the functional delta method,

$$(A.4)=-\frac{\int_0^t s^{(0)}(u;\beta_0)d\Lambda_{0j}(u)}{\left\{\sum_{\ell=1}^{J}\int_0^t s_\ell^{(0)}(u;\beta_0)d\Lambda_{0\ell}(u)\right\}^2}\sum_{\ell=1}^{J}G_n[N_\ell(t)]+o_p(1).$$

Combining (A.3) and (A.4),

$$n^{\frac{1}{2}}\left\{\widehat{SRR}_j(t)-SRR_j(t)\right\}$$
$$=G_n\left[w(t;\beta_0)\left\{\int_0^t\frac{s^{(0)}(u;\beta_0)}{s_j^{(0)}(u;\beta_0)}dN_j(u)+\int_0^t r_j^T(u;\beta_0)d\Lambda_{0j}(u)\left\{\sum_{\ell=1}^{J}\Omega_\ell(\beta_0)^{-1}\right\}\right.\right.$$
$$\left.\left.\sum_{\ell=1}^{J}\left\{\{Z-\overline{z}_\ell(X;\beta_0)\}\Delta I(G=\ell)-\int_0^\tau\{Z-\overline{z}_\ell(u;\beta_0)\}\exp\left(\beta_0^TZ\right)I(X\geqslant u)I(G=\ell)d\Lambda_{0\ell}(u)\right\}\right\}\right.$$
$$\left.-\frac{\int_0^t s^{(0)}(u;\beta_0)d\Lambda_{0j}(u)}{\left\{\sum_{\ell=1}^{J}\int_0^t s_\ell^{(0)}(u;\beta_0)d\Lambda_{0\ell}(u)\right\}^2}\Delta I(X\leqslant t)\right]+o_p(1);$$

Because the VC classes with finite index satisfy the entropy conditions for the Donsker theorem [16], $\{I(X\geqslant t),t\in[0,\tau]\}$ and $\{I(X\leqslant t),t\in[0,\tau]\}$ belong to some Donsker classes. The same holds for the bounded monotone stochastic process $\{N(t),t\in[0,\tau]\}$. Finally, the class of functions of Lipschitz transformations of Donsker classes is Donsker. Therefore, with the various bounded conditions and by applying the Donsker theorem, Theorem 1 follows.

Acknowledgements

This work was supported in part by the National Institutes of Health grant 5R01-DK070869 and a grant from the Michigan Institute for Clinical and Health Research (MICHR). The authors thank the Scientific Registry of Transplant Recipients (SRTR) for allowing us to access the organ failure database.

References

1. Cox DR. Regression models and life tables (with Discussion). *Journal of the Royal Statistical Society, Series B* 1972; **34**:187–200.
2. Berry G. The analysis of mortality by the subject-years method. *Biometrics* 1983; **39**:173–184.
3. Breslow NE, Day NE. The standardized mortality ratio. In *Biostatistics: Statistics in Biomedical, Public Health and Environmental Sciences*, Sen PK (ed.), The Bernard G. Greenberg Volume, 1985; 55–74.
4. Hazel I. Standardization methods. In *Encyclopedia of Biostatistics*, Vol. 7, 2nd ed. Wiley, 2005; 5151–5163.
5. Logan BR, Nelson GO, Klein JP. Analyzing center specific outcomes in hematopoietic cell transplantation. *Lifetime Data Analysis* 2008; **14**:389–404. DOI: 10.1007/s10985-008-9100-6.
6. Spiegelhalter D, Johnson CS, Bardsley M, Blunt I, Wood C, Grigg O. Statistical methods for healthcare regulation: rating, screening and surveillance. *Journal of the Royal Statistical Society, Series A* 2012; **175**:1–47. DOI: 10.1111/j.1467-985X.2011.01010.x.
7. Keiding N. The method of expected number of deaths, 1786-1886-1986. *International Statistical Review* 1987; **55**:1–20.
8. Wolfe RA, Gaylin DS, Port FK, Held PJ, Wood CL. Using USRDS generated mortality tables to compare local ESRD mortality rates to national rates. *Kidney International* 1992; **42**:991–996.

9. Wolfe RA. The standardized morality ratio revisited: improvements, innovations, and limitations. *American Journal of Kidney Diseases* 1994; **24**:290–297.
10. Dickinson DM, Shearon TH, O'Keefe J, Wong HH, Berg CL, Rosendale JD, Delmonico FL, Webb RL, Wolfe RA. SRTR center-specific reporting tools: posttransplant outcomes. *American Journal of Transplantation* 2006; **6**:1198–1211. DOI: 10.1111/j.1600-6143.2006.01275.x.
11. Cox DR. Partial likelihood. *Biometrika* 1975; **62**:269–276.
12. Breslow NE. Contribution to the discussion on the paper by D. R. Cox, regression and life table. *Journal of the Royal Statistical Society, Series B* 1972; **34**:216–217.
13. He K. Semi-parametric and parametric methods for the analysis of multi-center survival data. *Ph.D. Thesis*, University of Michigan, Department of Biostatistics, Ann Arbor, 2012.
14. Kalbfleish JD, Wolfe RA. On monitoring outcomes of medical provider. *Statistics in the Bioscience* 2013; **5**(2):286–302.
15. Ohlssen DI, Sharples LD, Spiegelhalter DJ. Flexible random-effects models using Bayesian semi-parametric models: application to institutional comparisons. *Statistics in Medicine* 2007; **26**(9):2088–2112. DOI: 10.1002/sim.2666.
16. van der Vaart AW, Wellner JA. *Weak Convergence and Empirical Processes*. Springer: New York, 1996.
17. Kosorok MR. *Introduction to Empirical Processes and Semiparametric Inference*. Springer Series in Statistics, 2008.
18. Andersen PK, Borgan Ø, Gill RD, Keiding N. *Statistical Models Based on Counting Processes*. Springer Series in Statistics: New York, 1993.
19. van der Vaart AW. *Asymptotic Statistics*. Cambridge, 1998.