Biometrics 74, 289–299
March 2018

DOI: 10.1111/biom.12739

Evaluating Center Performance in the Competing Risks Setting: Application to Outcomes of Wait-Listed End-Stage Renal Disease Patients

Sai H. Dharmarajan,[1,*] Douglas E. Schaubel,[1,] and Rajiv Saran[2]**

[1]Department of Biostatistics, University of Michigan, Ann Arbor, Michigan, U.S.A.
[2]Department of Internal Medicine, University of Michigan, Ann Arbor, Michigan, U.S.A.
[*]*email:* shdharma@umich.edu
[**]*email:* deschau@umich.edu

Summary. It is often of interest to compare centers or healthcare providers on quality of care delivered. We consider the setting where evaluation of center performance on multiple competing events is of interest. We propose estimating center effects through cause-specific proportional hazards frailty models that allow correlation among a centers cause-specific effects. Estimation of our model proceeds via penalized partial likelihood and is implemented in R. To evaluate center performance, we also propose a directly standardized excess cumulative incidence (ECI) measure. Therefore, based on our proposed methods, practitioners can evaluate centers either through the cause-specific hazards or the cumulative incidence functions. We demonstrate, through simulations, the advantages of the proposed methods to detect outlying centers, by comparing the proposed methods and existing methods which assume uncorrelated random center effects. In addition, we develop a Correlation Score Test to test the null hypothesis that the competing event processes within a center are correlated. Using data from the Scientific Registry of Transplant Recipients, we apply our method to evaluate the performance of Organ Procurement Organizations on two competing risks: (i) receipt of a kidney transplant and (ii) death on the wait-list.

Key words: Cause-specific hazards; Center Effects; Competing Risks; Correlation Score Test; Cumulative Incidence; Kidney Transplantation.

1. Introduction

The availability of electronic health records and the demand for value-driven healthcare have led to greatly increased interest in the methods for evaluation of center performance (Ash et al., 2012). For continuous or binary outcomes, center effects are usually estimated as either fixed or random effects models. Evaluation of center performance is then generally carried out by comparing these estimated risk-adjusted center effects to some fixed quantity, or the average center effect, or by using graphical checks (Spiegelhalter et al., 2012).

The proposed methods are motivated by the end-stage renal disease (ESRD) setting. There are thousands more patients in need of transplantation than there are donor kidneys. As a result, medically suitable ESRD patients are placed on a waiting list. For example, in 2015, there were 98,956 patients on the kidney waiting list at year-end, but only 11,594 deceased-donor kidney transplants (Hart et al., 2016). In the United States, there are 58 wait-lists, each administered by an Organ Procurement Organization (OPO). Our objective here is to evaluate OPOs with respect to (i) kidney transplantation and (ii) pre-transplant death (competing risks) among wait-listed patients.

While there has been extensive research conducted into establishing methods for institutional comparisons with respect to binary and continuous outcomes, apart from a few recent studies, time-to-event outcomes have received considerably less attention. He and Schaubel (2014a) assessed the standardized mortality ratio (SMR) measure based on the Cox model and developed an alternative based on stratification. In another study, He and Schaubel (2014b) developed a direct standardized measure of center performance.

Oftentimes in clinical and epidemiological settings, there is more than one competing outcome of interest. In such cases, there are two approaches to conceptualize the event times for the competing risks. The first approach assumes that, for every patient, a latent event time (Gail, 1975; Crowder, 2001) exists for each outcome and only the minimum of these (Cox, 1959) is observed. Under this conceptualization, latent event times must act independently in order for marginal quantities (e.g., cause- or event-specific survival function) to be identifiable. A second approach, adopted in our report, assumes that only one event time, pertaining to the cause of failure, exists for each subject (Kalbfleisch and Prentice, 2002). Data from such settings can now be analyzed through the analysis of cause-specific hazards (Prentice and Kalbfleisch, 1978; Kalbfleisch and Prentice, 2002).

With competing risks data, a comparison of centers with respect to all-cause mortality has the potential to obscure important findings by averaging of dissimilar results (Van Rompaye et al., 2010). An analysis by cause has the

potential to yield more interpretable and insightful conclusions (Putter et al., 2007). Fan and Schaubel (2016) proposed, as a center performance measure, the difference between the estimated cumulative incidence of transplant for patients at a given center and the average of the estimated cumulative incidences. Based on similar techniques, Van Rompaye, Erikson, and Goetghebeur (2015) developed an "excess cause-specific cumulative incidence" (ECI). For indirectly standardized measures, center performance is evaluated at the patient mix or covariate distribution of each center. Although useful for internal benchmarking, directly standardized measures are preferred for comparisons across centers (Varewyck et al., 2014). Note that random center effects may be preferable to fixed effects in the presence of small center sizes (Ohlssen et al., 2006; Ash et al., 2012; Kalbfleisch and Wolfe, 2013).

Most existing methods for clustered competing risks model the within-cluster dependence through a random effect, and concentrate on a single risk (or separate models for each risk) (Katsahian and Boudreau, 2011; Do Ha et al., 2014). In contrast, we propose a class of frailty models which allow a centers cause-specific random effects be correlated. This approach utilizes the additional information available in the form of correlation between cause-specific random effects within a center.

In this article, we develop a directly standardized ECI measure to contrast center performance on competing outcomes. We utilize an easily implementable penalized partial likelihood method (Ripatti and Palmgreen, 2000). Note that Gorfine and Hsu (2011) and Gorfine et al. (2014) also developed frailty models for correlated event times within-cluster. However, an Expectation-Maximization (EM) algorithm was used which requires numerical integration at each E-step. In comparison, our estimation procedure does not require any numerical integration and is implemented through a single call to **coxme** function of the **coxme** package (Therneau, 2009).

If competing events are indeed uncorrelated, fitting separate models is appropriate and easier than the proposed methods. Therefore, we also develop a convenient score test for the presence of correlation between competing risks within-center. The score test does not require fitting the joint model and, thus, provides an a priori checks the appropriateness of using separate cause-specific models, in lieu of the proposed methods.

2. Proposed Methods

2.1. *Model and Likelihood*

There are J centers or clusters, with each center j having n_j members $(j = 1, \ldots, J)$ so that there are $\sum_{j=1}^{J} n_j = n$ individuals in the entire sample. For each subject $i (i = 1, \ldots, n_j)$ in center j, let T_{ij}^0 and C_{ij} denote the failure time and the censoring time, respectively, and let $\boldsymbol{X}_{ij}$ be a vector of time-independent covariates. The observed event time is then defined as $T_{ij} = \min(T_{ij}^0, C_{ij})$. Each subject fails due to one of K causes, we use Δ_{ij} $(\Delta_{ij} \in \{0, \ldots, K\})$ to indicate the cause of the observed failure for subject i in center j, with $\Delta_{ij} = 0$ if $T_{ij}^0 > C_{ij}$. The observed data consist of $\{T_{ij}, \Delta_{ij}, \boldsymbol{X}_{ij}, A_{ij}\}$ for $i = 1, \ldots, n_j$ and $(j = 1, \ldots, J)$, where $A_{ij} = 1$ if subject i belongs to center j and 0 otherwise.

Additionally, we define a vector of center-specific random effects or frailties, for the jth center, $\boldsymbol{\gamma}_j = (\gamma_{j1}, \ldots, \gamma_{jK})^T$, given which the event times for all subjects within that center are assumed to be conditionally independent. Thus, the cause-specific hazard function for cause k, for the subject i in the center j, is given by:

$$\lambda_{ijk}(t|\boldsymbol{X}_{ij}, \gamma_{jk}) = \lim_{h \downarrow 0} \frac{1}{h} Pr(t \leq T_{ij}^0 < t + h, \Delta_{ij} = k | T_{ij}^0 \geq t, \boldsymbol{X}_{ij}, \gamma_{jk})$$

and is assumed to be following the proportional hazards model:

$$\lambda_{ijk}(t|\boldsymbol{X}_{ij}, \gamma_{jk}) = \lambda_{0k}(t) \exp\{\boldsymbol{\beta}_k^T \boldsymbol{X}_{ij} + \gamma_{jk}\} \tag{1}$$

for $k = 1, \ldots, K$ where $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_k$ and $\lambda_{01}, \ldots, \lambda_{0k}$ are cause-specific regression coefficients and cause-specific baseline hazards respectively. Here, we assume that the vector of covariates $\boldsymbol{X}_{ij}$ is the same for all causes, but it can be replaced by cause-specific vectors of covariate s $\boldsymbol{X}_{ijk}$. The center-specific random effects imply a correlation between the cause-specific hazards across subjects within a center. Further, by assuming that the center-specific random effect vectors arise from a multivariate normal distribution with mean zero and covariance matrix $\boldsymbol{V}_j$, that is, $\boldsymbol{\gamma}_j \sim MVN(\boldsymbol{0}, \boldsymbol{V}_j)$, our model allows for the association of different cause-specific hazards across individuals within a center. It is important to note that our model implies that the cause-specific hazards for different causes may be correlated across individuals within a center and not that the cause-specific event times within each individual are correlated. Indeed, as we do not adopt the latent failure time paradigm, our model is agnostic about the existence of different cause-specific event times within each individual.

We focus on the case of $K = 2$ competing causes, and allow for center-specific random effects for the two different causes to be negatively associated, that is, $\mathrm{Corr}(\gamma_{j1}, \gamma_{j2}) \leq 0$. To this end, we reformulate the cause-specific hazards in equation (1) as

$$\lambda_{ij1}(t|\boldsymbol{X}_{ij}, b_j^0, b_j^1) = \lambda_{01}(t) \exp\{\boldsymbol{\beta}_1^T \boldsymbol{X}_{ij} + b_j^1 + b_j^0\} \tag{2}$$

$$\lambda_{ij2}(t|\boldsymbol{X}_{ij}, b_j^0, b_j^2) = \lambda_{02}(t) \exp\{\boldsymbol{\beta}_1^T \boldsymbol{X}_{ij} + b_j^2 - b_j^0\} \tag{3}$$

where $b_j^1 + b_j^0 = \gamma_{j1}$ and $b_j^2 - b_j^0 = \gamma_{j2}$. We have decomposed a center's cause-specific random-effect into two independent components: a shared random-effect, b_j^0, acting in opposite directions on the hazards of the two different risks, and a cause-specific random effect component b_j^k. This implies that $\mathrm{Cov}(\gamma_{j1}, \gamma_{j2}) = -\mathrm{Var}(b_j^0)$. We further assume that jointly $\boldsymbol{b}_j = (b_j^0, b_j^1, b_j^2) \sim p(\boldsymbol{b}_j; \boldsymbol{D}_j) = MVN(\boldsymbol{0}, \boldsymbol{D}_j(\boldsymbol{\theta}_j))$, where $\boldsymbol{D}_j(\boldsymbol{\theta}_j)$ is a diagonal covariance matrix with unknown parameters denoted by the vector $\boldsymbol{\theta}_j$.

We now construct the likelihood function for the model implied in equation (1) in terms of the parameters $(\lambda_0(t), \boldsymbol{\beta}_k^T, \boldsymbol{\theta}_j)$. Note that, for any given subject, $\lambda_{ij}(t|\boldsymbol{X}_{ij}, \boldsymbol{b}_j) = \sum_{k=1}^{K} \lambda_{ijk}(t|\boldsymbol{X}_{ij}, \boldsymbol{b}_j)$. Thus, the cause-specific densities can be

represented as $f_{ijk}(t|\boldsymbol{X}_{ij}, \boldsymbol{b}_j) = \lambda_{ijk}(t|\boldsymbol{X}_{ij}, \boldsymbol{b}_j)S_{ij}(t|\boldsymbol{X}_{ij}, \boldsymbol{b}_j)$ for $k = \{1, \ldots, K\}$, where $S_{ij}(t|\boldsymbol{X}_{ij}, \boldsymbol{b}_j) = \exp\{-\sum_{k=1}^{K} \lambda_{ijk}(t|\boldsymbol{X}_{ij}, \boldsymbol{b}_j)\}$. Hence, the likelihood function can be written in terms of cause-specific hazard functions. Let the at-risk indicator for subject i in center j be given by $Y_{ij}(t) = I(T_{ij} \geq t)$. Using the notation given in Section 2.1, we write the likelihood for subjects in center j as:

$$\begin{aligned} L_j = \int \prod_{i=1}^{n_j} \prod_{k=1}^{K} &\{\lambda_{0k}(T_{ij}) \exp\{\boldsymbol{\beta}_k^T \boldsymbol{X}_{ij} + \boldsymbol{b}_j^T \boldsymbol{Z}_{ijk}\}\}^{I(\Delta_{ij}=k)} \\ &\times \left[\exp\left(-\int_0^t Y_{ij}(u)\lambda_{0k}(u) \exp\{\boldsymbol{\beta}_k^T \boldsymbol{X}_{ij} + \boldsymbol{b}_j^T \boldsymbol{Z}_{ijk}\} du\right)\right] \\ &\times p(\boldsymbol{b}_j; \boldsymbol{D}_j(\boldsymbol{\theta}_j)) d\boldsymbol{b}_j \end{aligned} \tag{4}$$

where the integral sign represents the unobserved frailties given by $\boldsymbol{b}_j$ being integrated out and $\boldsymbol{Z}_{ijk}$ are design vectors setup to obtain the cause-specific hazard models in equations (2) and (3). Specifically, if subject i is in center j then $\boldsymbol{Z}_{ij1} = (1, 1, 0)$ and $\boldsymbol{Z}_{ij2} = (-1, 0, 1)$, and if subject i does not belong to center j then $\boldsymbol{Z}_{ij1} = \boldsymbol{Z}_{ij2} = (0, 0, 0)$. It is important to note that for the construction of the above likelihood, we assumed the following: (1) Conditional on $\{\boldsymbol{X}_{ij}, \boldsymbol{Z}_{ijk}, \boldsymbol{b}_j\}$, the event times and censoring times are independent and the censoring times are non-informative for $\{\boldsymbol{\beta}_k, \lambda_{0k}, k = 1, 2\}$, (2) $\boldsymbol{X}_{ij}$ and $\boldsymbol{b}_j$ are independent.

2.2. *Estimation*

It follows from equation (4) above that the overall likelihood of the data is given by:

$$\begin{aligned} L = \int \prod_{j=1}^{J} \prod_{i=1}^{n_j} \prod_{k=1}^{K} &\{\lambda_{0k}(T_{ij}) \exp\{\boldsymbol{\beta}_k^T \boldsymbol{X}_{ij} + \boldsymbol{b}_j^T \boldsymbol{Z}_{ijk}\}\}^{I(\Delta_{ij}=k)} \\ &\times [\exp(-\Lambda_{0k}(T_{ij}) \exp\{\boldsymbol{\beta}_k^T \boldsymbol{X}_{ij} + \boldsymbol{b}_j^T \boldsymbol{Z}_{ijk}\})] \\ &\times p(\boldsymbol{b}_j; \boldsymbol{D}_j(\boldsymbol{\theta}_j)) d\boldsymbol{b}_j, \end{aligned} \tag{5}$$

where $\Lambda_{0k}(T_{ij}) \exp\{\boldsymbol{\beta}_k^T \boldsymbol{X}_{ij} + \boldsymbol{b}_j^T \boldsymbol{Z}_{ijk}\} = \int_0^{T_{ij}} Y_{ij}(u)\lambda_{0k}(u) \exp\{\boldsymbol{\beta}_k^T \boldsymbol{X}_{ij} + \boldsymbol{b}_j^T \boldsymbol{Z}_{ijk}\} du$.

Let $\boldsymbol{b} = \{\boldsymbol{b}_1{}^T, \ldots, \boldsymbol{b}_J{}^T\}^T$ be a vector of all random-effects, obtained by stacking the center-specific vectors of random effects $\boldsymbol{b}_j, j = 1, \ldots, J$. Correspondingly, we define $p(\boldsymbol{b}; \boldsymbol{D}(\boldsymbol{\theta})) = MVN(\boldsymbol{0}, \boldsymbol{D}(\boldsymbol{\theta}))$ such that $\boldsymbol{D}(\boldsymbol{\theta})$ is a block-diagonal covariance matrix composed of blocks formed by $\boldsymbol{D}_j(\boldsymbol{\theta}_j)$. We further assume that $\boldsymbol{\theta}_j = \boldsymbol{\theta}_{j'} = (\theta_0, \theta_1, \theta_2)$; that is, the center-specific random effect vectors, $\boldsymbol{b}_j$ are i.i.d with $\mathrm{Var}(b_j^l) = \theta_l, l = \{0, 1, 2\}$.

The integrand in equation (5) above can be viewed as the full likelihood of the data under our model, composed of the conditional likelihood of the data given random effects $\boldsymbol{b}$, multiplied by the likelihood of the random effects. Taking the log, we define:

$$\begin{aligned} l_{full} = l_{cond} + l_{\boldsymbol{b}} = &\sum_{j=1}^{J} \sum_{i=1}^{n_j} \sum_{k=1}^{K} I(\Delta_{ij} = k)\{\log(\lambda_{0k}(T_{ij})) \\ &+ \boldsymbol{\beta}_k^T \boldsymbol{X}_{ij} + \boldsymbol{b}_j{}^T \boldsymbol{Z}_{ijk}\} - \sum_{j=1}^{J} \sum_{i=1}^{n_j} \sum_{k=1}^{K} \Lambda_{0k}(T_{ij}) + \boldsymbol{\beta}_k^T \boldsymbol{X}_{ij} + \boldsymbol{b}_j{}^T \boldsymbol{Z}_{ijk} \\ &+ \log|\boldsymbol{D}|^{-\frac{1}{2}} - \frac{1}{2}\boldsymbol{b}^T \boldsymbol{D}^{-1} \boldsymbol{b} \end{aligned} \tag{6}$$

The above equation is a penalized log-likelihood for the observed data. As in Ripatti and Palmgren (2000), treating $\boldsymbol{b}$ as a fixed effect and using profile likelihood to estimate $\Lambda_{0k}(t)$ parameters, then plugging back the resulting Breslow (1974) estimator $\hat{\Lambda}_{0k}(t)$ into equation (6) yields the following penalized partial log-likelihood (PPLL):

$$\begin{aligned} l_{ppll} = l_1 + l_{\boldsymbol{b}} = &\sum_{j=1}^{J} \sum_{i=1}^{n_j} \sum_{k=1}^{K} I(\Delta_{ij} = k)\{\boldsymbol{\beta}_k^T \boldsymbol{X}_{ij} + \boldsymbol{b}_j{}^T \boldsymbol{Z}_{ijk} \\ &- \log \sum_{r=1}^{J} \sum_{q=1}^{n_j} Y_{qr}(T_{ij}) \exp\{\boldsymbol{\beta}_k^T \boldsymbol{X}_{qr} + \boldsymbol{b}_r{}^T \boldsymbol{Z}_{qrk}\}\} \\ &+ \log|\boldsymbol{D}|^{-\frac{1}{2}} - \frac{1}{2}\boldsymbol{b}^T \boldsymbol{D}^{-1} \boldsymbol{b}. \end{aligned} \tag{7}$$

As recommended in Ripatti and Palmgren (2000), we suggest obtaining the estimates of $((\boldsymbol{\beta}_k, \boldsymbol{b}), k = \{1, 2\})$ as solutions to the PPLL. To estimate $\boldsymbol{\theta}$, we need to integrate out $\boldsymbol{b}$. As in Breslow and Clayton (1993), we use a Laplace saddle point approximation to the integration of penalized partial likelihood $L_{PPLL} = \exp(l_{ppll})$, with respect to $d\boldsymbol{b}$. Doing so, we obtain an expression for the log of the integrated likelihood as:

$$\begin{aligned} l_{INT} &= -\frac{1}{2}\log|\boldsymbol{D}| - \frac{1}{2}\log|K"(\hat{\boldsymbol{b}})| - K(\hat{\boldsymbol{b}}) \\ K(\hat{\boldsymbol{b}}) &= \sum_{j=1}^{J} \sum_{i=1}^{n_j} \sum_{k=1}^{K} I(\Delta_{ij} = k)\{\boldsymbol{\beta}_k^T \boldsymbol{X}_{ij} + \hat{\boldsymbol{b}}_j^T \boldsymbol{Z}_{ijk} \\ &\quad - \log \sum_{r=1}^{J} \sum_{q=1}^{n_j} Y_{qr}(T_{ij}) \exp\{\boldsymbol{\beta}_k^T \boldsymbol{X}_{qr} + \hat{\boldsymbol{b}}_r^T \boldsymbol{Z}_{qrk}\}\} \\ &\quad + \log|\boldsymbol{D}|^{-\frac{1}{2}} - \frac{1}{2}\hat{\boldsymbol{b}}^T \boldsymbol{D}^{-1} \hat{\boldsymbol{b}} \end{aligned}$$

and $\hat{\boldsymbol{b}}$ denotes the solution to the partial derivatives of $K(\boldsymbol{b})$ with respect to $\boldsymbol{b}$, that is, $\hat{\boldsymbol{b}}$ solves:

$$\begin{aligned} K'(\boldsymbol{b}) = &\sum_{j=1}^{J} \sum_{i=1}^{n_j} \sum_{k=1}^{K} I(\Delta_{ij} = k) \\ &\times \left[\boldsymbol{Z}_{ijk} - \frac{\sum_{r=1}^{J} \sum_{q=1}^{n_j} Y_{qr}(T_{ij}) \boldsymbol{Z}_{qrk} \exp\{\boldsymbol{\beta}_k \boldsymbol{X}_{qr} + \boldsymbol{b}_r{}^T \boldsymbol{Z}_{qrk}\}}{\sum_{r=1}^{J} \sum_{q=1}^{n_j} Y_{qr}(T_{ij}) \exp\{\boldsymbol{\beta}_k \boldsymbol{X}_{qr} + \boldsymbol{b}_r{}^T \boldsymbol{Z}_{qrk}\}}\right] \\ &- \boldsymbol{D}^{-1} \boldsymbol{b} = 0. \end{aligned} \tag{8}$$

The quantity $K"(\hat{\boldsymbol{b}})$ is the set of second partial derivatives of $K(\boldsymbol{b})$ at $\hat{\boldsymbol{b}}$. $K"(\hat{\boldsymbol{b}})$ is also the second partial derivative of l_{PPLL}, evaluated at $\hat{\boldsymbol{b}}$. If we define $\boldsymbol{H}$ as the matrix of second derivatives or Hessian of the PPLL with respect to $(\boldsymbol{\beta}, \boldsymbol{b})$, such that:

$$\boldsymbol{H} = \begin{bmatrix} \boldsymbol{H}_{11} & \boldsymbol{H}_{12} \\ \boldsymbol{H}_{21} & \boldsymbol{H}_{22} \end{bmatrix} = -\mathcal{I}(\boldsymbol{\beta}, \boldsymbol{b}) + \begin{bmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{D}^{-1} \end{bmatrix}$$

where $\mathcal{I}(\boldsymbol{\beta}, \boldsymbol{b}) = -\partial^2 l_1 / \partial(\boldsymbol{\beta}, \boldsymbol{b}) \partial(\boldsymbol{\beta}, \boldsymbol{b})'$, then $\boldsymbol{H}(\boldsymbol{\beta}, \hat{\boldsymbol{b}})_{22} = K"(\hat{\boldsymbol{b}})$. We then have:

$$l_{INT} \approx l_1(\boldsymbol{\beta}, \hat{\boldsymbol{b}}) + l_{\boldsymbol{b}}(\boldsymbol{\theta}, \hat{\boldsymbol{b}}) - \frac{1}{2} \log |\boldsymbol{H}(\boldsymbol{\beta}, \hat{\boldsymbol{b}})_{22}| \tag{9}$$

As demonstrated by Ripatti and Palmgren (2000), ignoring the last term on the right hand side of equation (9) while estimating $(\boldsymbol{\beta}, \boldsymbol{b})$ leads to very little loss of information. This corresponds to using the PPLL to estimate $(\boldsymbol{\beta}, \boldsymbol{b})$ via a Newton-Raphson algorithm. We have the following estimating equation for $\boldsymbol{\beta}$:

$$\partial l_{PPLL} / \partial \boldsymbol{\beta} = \sum_{j=1}^{J} \sum_{i=1}^{n_j} \sum_{k=1}^{K} I(\Delta_{ij} = k) \times \left[\boldsymbol{X}_{ij} - \frac{\sum_{j=1}^{J} \sum_{q=1}^{n_j} Y_{qr}(T_{ij}) \boldsymbol{X}_{qr} \exp\{\boldsymbol{\beta}_k \boldsymbol{X}_{qr} + \boldsymbol{b}_r{}^T \boldsymbol{Z}_{qrk}\}}{\sum_{j=1}^{J} \sum_{q=1}^{n_j} Y_{qr}(T_{ij}) \exp\{\boldsymbol{\beta}_k \boldsymbol{X}_{qr} + \boldsymbol{b}_r{}^T \boldsymbol{Z}_{qrk}\}} \right] = 0 \tag{10}$$

The estimating equation for $\boldsymbol{b}$ is similarly obtained by setting $\partial l_{PPLL} / \partial \boldsymbol{b}$ to zero, and is identical to equation (8). Thus, equation (8), required for the saddle point Laplace approximation, is automatically satisfied when PPLL is used to estimate $\boldsymbol{b}$. To estimate $\boldsymbol{D}(\boldsymbol{\theta})$, we plug the estimated values $(\hat{\boldsymbol{\beta}})$ into equation (9) and solve for $\boldsymbol{\theta}$ that maximizes l_{INT}. This gives us the following estimating equation:

$$-\frac{1}{2} \left[tr\left(\boldsymbol{D}^{-1} \frac{\partial \boldsymbol{D}}{\partial \boldsymbol{\theta}} \right) + tr\left(\boldsymbol{H}_{22}^{-1} \frac{\partial \boldsymbol{D}^{-1}}{\partial \boldsymbol{\theta}} \right) - \hat{\boldsymbol{b}}^T \boldsymbol{D}^{-1} \frac{\partial \boldsymbol{D}}{\partial \boldsymbol{\theta}} \boldsymbol{D}^{-1} \hat{\boldsymbol{b}} \right] = 0 \tag{11}$$

For a diagonal covariance matrix, as in our case, we obtain the following solution:

$$\hat{\theta}_l = \frac{(\hat{\boldsymbol{b}}^l)^T (\hat{\boldsymbol{b}}^l) + tr(\boldsymbol{H}_{22}^l (\hat{\boldsymbol{b}}^l)^{-1})}{J}, l = \{0, 1, 2\} \tag{12}$$

where $\hat{\boldsymbol{b}}^l = \{\hat{\boldsymbol{b}}_1^l, \ldots, \hat{\boldsymbol{b}}_J^l\}$ and $\boldsymbol{H}_{22}^l(\hat{\boldsymbol{b}}^l)$ is the sub-matrix corresponding to $\hat{\boldsymbol{b}}^l$ terms. The proposed estimation algorithm begins with an initial guess of $\boldsymbol{\theta}$, then alternates between using the PPLL to estimate $(\boldsymbol{\beta}, \boldsymbol{b})$ as listed above and using equation (12) to update $\boldsymbol{\theta}$ until convergence. As suggested by Gray (1992), the variance of $(\hat{\boldsymbol{\beta}}^T, \hat{\boldsymbol{b}}^T)^T$ is obtained as:

$$\hat{V}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{b}}) = \boldsymbol{H}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{b}})^{-1} \mathcal{I}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{b}}) \boldsymbol{H}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{b}})^{-1} \tag{13}$$

To obtain the asymptotic distribution for $(\hat{\boldsymbol{\beta}}, \hat{\mathbf{b}}, \hat{\lambda}_{0k}(s))$, we assumed that the increments $\hat{\lambda}_{0k}(s)$ are independent of $(\hat{\boldsymbol{\beta}}, \hat{\mathbf{b}})$. Under this assumption, we estimated the variance of $\hat{\lambda}_{0k}(s)$ via a non-parametric bootstrap approach where the values of $(\hat{\boldsymbol{\beta}}, \hat{\mathbf{b}})$ were treated as fixed by setting $\mathbf{X}\hat{\boldsymbol{\beta}} + \hat{\mathbf{b}}$ as an offset in the linear predictor of the instantaneous hazard. Thus, our desired asymptotic variance-covariance matrix for $(\hat{\boldsymbol{\beta}}, \hat{\mathbf{b}}, \hat{\lambda}_{0k}(s))$ was obtained using equation (13) to estimate the variance of $(\hat{\boldsymbol{\beta}}, \hat{\mathbf{b}})$ and a non-parametric bootstrap approach to estimate the variance of $\hat{\lambda}_{0k}(s)$. In doing so, we assume independence between $\hat{\lambda}_{0k}(s)$ and $(\hat{\boldsymbol{\beta}}, \hat{\mathbf{b}})$. Our simulation studies suggest this to be a safe assumption. In reality, the increments of $\hat{\lambda}_{0k}(s)$ and $(\hat{\boldsymbol{\beta}}, \hat{\mathbf{b}})$ may be weakly correlated. However, with increasing sample size one would expect this correlation to get weaker and have a negligible impact on the standard errors of estimates. Then, ignoring this correlation in return for substantial gains in computational efficiency seems appropriate. It should also be noted that, while using the Laplace approximation to the marginal log-likelihood leads to little loss of information, it might result in a slight underestimation of standard errors of fixed and random effect parameters if the cluster sizes are very small, as demonstrated in Ripatti and Palmgren (2000).

2.3. *Center Effect Measures: Cumulative Incidence*

We define the cumulative incidence function (CIF) of cause k for subject i at center j as:

$$F_{ijk}(t) = P(T_i^0 \leq t, \Delta_i = k | A_i = j, \boldsymbol{X_{ij}}), \tag{14}$$

the probability that an individual i in center j experiences a cause k event by time t. To evaluate the performance of center j with respect to type k events, we first define the average risk of events of type k at that center as $F_{jk}(t) = E_{\boldsymbol{X}}[F_{ijk}(t)]$, which is estimated as:

$$\hat{F}_{jk}(t) = \hat{E}_{\boldsymbol{X}}[F_{ijk}(t)] = \frac{\sum_{i=1}^{n} F_{ijk}(t)}{n} \tag{15}$$

Note that the above equation can be interpreted as potential risk for event k, at time t, that would be observed if the entire study population was treated at center j, assuming there are no unmeasured confounders. To compare the performance of center j to that of other centers we difference this potential risk with the average of such potential risks across all the centers. We call this measure the excess cumulative incidence. This is denoted as $\delta_{jk}(t) = F_{jk}(t) - E_A[F_{jk}(t)]$ and estimated as:

$$\hat{\delta}_{jk}(t) = \hat{F}_{jk}(t) - \frac{\sum_{q=1}^{J} \hat{F}_{qk}(t)}{J} \tag{16}$$

2.4. *Estimating Center Effects*

We estimate cumulative incidence functions, defined in equation (14) using the cause-specific hazards estimated from section 2.2. We note that the cause-specific CIF for cause

k, individual i at center j can be written as:

$$F_{ijk}(t) = \int_0^t S_{ij}(s)\lambda_{ijk}(s)ds, \tag{17}$$

for which an estimate $\hat{F}_{ijk}(t)$ is then obtained by plugging into equation (17) the following estimated quantities:

$$\hat{\lambda}_{ijk}(s) = \hat{\lambda}_{0k}(s)\exp(\hat{\boldsymbol{\beta}}_k \boldsymbol{X}_{ij} + \hat{b}_j \boldsymbol{Z}_{ijk});$$
$$\hat{S}_{ij}(s) = \exp\left\{ -\sum_{k=1}^{2} \hat{\Lambda}_{0k}(s)\exp(\hat{\boldsymbol{\beta}}_k \boldsymbol{X}_{ij} + \hat{\boldsymbol{b}}_j \boldsymbol{Z}_{ijk}) \right\}$$

where $\hat{\boldsymbol{\beta}}_k$, $\hat{\boldsymbol{b}}_j$ are estimates obtained as detailed in Section 2.2, and $\hat{\Lambda}_{0k}(t) = \int_0^t \hat{\lambda}_{0k}(s)ds$ is the cumulative cause-specific baseline hazard function obtained by integrating the Breslow–Aalen (Breslow 1974) estimate of the cause-specific baseline hazard function. Estimates of $F_{jk}(t)$ and the excess cumulative incidence at center j, $\hat{\delta}_{jk}(t)$, are subsequently obtained by plugging $\hat{F}_{ijk}(t)$ into equations (14) and (16), respectively.

To obtain the variance of the cause-specific cumulative incidence and excess cumulative incidence functions, we apply a parametric bootstrap approach. Specifically, we re-sample the estimated parameters $\hat{\boldsymbol{\beta}}_k$, $\hat{\boldsymbol{b}}_j$ and $\hat{\lambda}_{0k}(s)$ from their estimated asymptotic distributions to obtain bootstrapped estimates of the cumulative incidence functions. The variance of $\hat{F}_{jk}(t)$ and $\hat{\delta}_{jk}(t)$ are estimated as variance of the corresponding bootstrapped estimates.

3. Score Test of Correlation of Cause-Specific Hazards

As mentioned in Section 2.1, equation (1), the cause-specific hazard function for cause k, for the ith subject in center j, is assumed to follow:

$$\lambda_{ijk}(t|\boldsymbol{X}_i, \gamma_{jk}) = \lambda_{0k}(t)\exp\{\boldsymbol{\beta}_k^T \boldsymbol{X}_i + \gamma_{jk}\}$$

Thus, the likelihood for the observed data in center j is:

$$L_j = \int \prod_{i=1}^{n_j}\prod_{k=1}^{K} \{\lambda_{0k}(t_i)\exp\{\boldsymbol{\beta}_k^T \boldsymbol{X}_i + \gamma_{jk}\}\}^{\Delta_{ik}(t)} \times \left[\exp\left(-\int_0^\tau Y_i(u)\lambda_{0k}(u)\exp\{\boldsymbol{\beta}_k^T \boldsymbol{X}_i + \gamma_{jk}\}dt \right)\right] p(\boldsymbol{\gamma}_j; \boldsymbol{V}(\boldsymbol{\theta}))d\boldsymbol{\gamma}_j \tag{18}$$

To develop a score test of the correlation of cause-specific hazards within centers, we consider a special case of the model in equation (1) when only $K = 2$ causes are present. Assume that the center-specific random effects or frailty for cause 2 and cause 1 differ by a multiplicative constant, that is, $\gamma_{j2} = \omega\gamma_{j1}$, implying the following specification for the cause-specific hazards:

$$\lambda_{ij1}(t|\boldsymbol{X}_i) = \lambda_{01}(t)\exp\{\boldsymbol{\beta}_1^T \boldsymbol{X}_i + \gamma_{j1}\};$$
$$\lambda_{ij2}(t|\boldsymbol{X}_i) = \lambda_{02}(t)\exp\{\boldsymbol{\beta}_1^T \boldsymbol{X}_i + \omega\gamma_{j1}\} \tag{19}$$

The presence of a correlation between the cause-specific hazards within centers is then assessed by testing $H_0 : \omega = 0$. When $\omega = 0$, there is little evidence for a linear relationship between center-specific random effects for causes 1 and 2. Conversely, even if the center-specific random effects are not perfectly correlated as implied by the specification in (19) but have a dependence of the form specified in model (1) we would expect to reject the test of $H_0 : \omega = 0$ in favor of $H_a : \omega \neq 0$. This is because, in case of any non-zero correlation between the center-specific random effects, the specification in (19) with some $\omega \neq 0$ should provide a better fit to the observed data than that with $\omega = 0$. Thus, we propose to test for the presence of correlation between cause-sepcific hazards in model (1), that is, $H_0 : Cov(\gamma_{j1}, \gamma_{j2}) = 0$, using the specification in (19) and testing $H_0 : \omega = 0$. Under the joint model for the cause-specific hazards in (19), likelihood for observed data in center j is given by:

$$\begin{aligned} L_j = \int \prod_{i=1}^{n_j} &\{\lambda_{01}(t_i) \\ &\times \exp\{\boldsymbol{\beta}_1^T \boldsymbol{X}_i + \gamma_{j1}\}\}^{\Delta_{i1}(t)} \{\lambda_{02}(t_i)\exp\{\boldsymbol{\beta}_2^T \boldsymbol{X}_i + \omega\gamma_{j1}\}\}^{\Delta_{i2}(t)} \\ &\times \left[\exp\left(-\int_0^\tau Y_i(u)\lambda_{01}(u)\exp\{\boldsymbol{\beta}_1^T \boldsymbol{X}_i + \gamma_{j1}\}du \right)\right] \\ &\times \left[\exp\left(-\int_0^\tau Y_i(u)\lambda_{02}(u)\exp\{\boldsymbol{\beta}_2^T \boldsymbol{X}_i + \omega\gamma_{j1}\}du \right)\right] \\ &\times p(\gamma_{j1}; \theta)d\gamma_{j1}. \end{aligned} \tag{20}$$

The marginal log-likelihood for the observed data at all centers is then given by:

$$\log l(\omega, \boldsymbol{\beta}_k, \lambda_{0k}) = \sum_{j=1}^{J}\sum_{i=1}^{n_j} \left(\sum_{k=1}^{2} \Delta_{ik}(t)\{\log\lambda_{0k}(t_i) + \{\boldsymbol{\beta}_k^T \boldsymbol{X}_i\}\} \right) + \log \int K_j(z_j, t)\, p(z_j; \theta)dz_j$$

where $z_j = \log\gamma_{j1}$, and

$$\begin{aligned} K_j(z, t) = z_j^{\sum_{i=1}^{n_j} N_{i1}(t-) + \omega N_{i2}(t-)} &\times \exp\left\{ -z_j \left(\int_0^t Y_i(u)\lambda_{02}(u)\exp\{\boldsymbol{\beta}_2^T \boldsymbol{X}_i\}du \right) \right. \\ &\left. - z_j^{\omega} \left(\int_0^t Y_i(u)\lambda_{02}(u)\exp\{\boldsymbol{\beta}_2^T \boldsymbol{X}_i\}du \right) \right\}. \end{aligned}$$

3.1. *Correlation Score Test*

Using the above formulation, the score test for correlation of the two cause-specific hazards tests $H_0 : \omega = 0$. The score function is:

$$U_\omega(\omega, \boldsymbol{\beta}_k, \lambda_{0k}) = \sum_j \frac{\int \{\sum_{i=1}^{n_j} N_{i2}(t_i) - (\int_0^t Y_i(u)\lambda_{02}(u)\exp\{\boldsymbol{\beta}_2^T \boldsymbol{X}_i\}du)\} z_j^\omega \log z_j K_j(z_j) p(z_j;\theta) dz_j}{\int K_j(z_j) p(z_j;\theta) dz_j}$$

Setting $\omega = 0$ and replacing $\boldsymbol{\beta}_k, \lambda_{0k}$ and θ with their estimates when $\omega = 0$, we have:

$$U_\omega(\omega, \boldsymbol{\beta}_k, \lambda_{0k}) = \sum_j \frac{\int \{\sum_{i=1}^{n_j} N_{i2}(t_i) - (\int_0^t Y_i(u)\hat{\lambda}_{02}(u)\exp\hat{\boldsymbol{\beta}}_2^T \boldsymbol{X}_i\}du)\} \log z_j \hat{K}_j(z_j) p(z_j;\hat{\theta}) dz_j}{\int \hat{K}_j(z_j) p(z_j;\theta) dz_j} = \sum_j \widehat{M}_{2j} . \widehat{\log z_j}$$

$\widehat{M}_{2j}$ is an estimate of the $\{\sum_{i=1}^{n_j} N_{i2}(t_i) - (\int_0^t Y_i(u)\lambda_{02}(u) \times \exp\{\boldsymbol{\beta}_2^T \boldsymbol{X}_i\}du)\}$, the sum of the martingale residuals for cause 2 at center j; and $\widehat{\log z_j} = E[\log z_j | O_j]$, that is, the posterior expectation of the log frailties given the observed data in center j, O_j. If the frailties z_j are assumed to follow a log normal distribution, there is no closed form expression for $\widehat{\log z_j}$, however, we can use the estimates $\hat{\gamma}_{j1}$ obtained by maximizing the penalized partial log-likelihood for cause 1. Balan et al. (2016) note that the test of $H_0 : \omega = 0$ can be carried out by testing if $\widehat{M}_{2j}$ and $\widehat{\log z_j}$ are correlated. Thus, the correlation score test (CST) tests if there is a linear dependency between $\widehat{M}_{2j}$ and $\widehat{\log z_j}$ and uses the regular t statistic from linear regression as the test statistic, $t = r\sqrt{(J-2)/(1-r^2)}$. Under $H_0 : \omega = 0$, asymptotically, t follows a t distribution with J–2 degrees of freedom.

4. Simulation Studies

In the first (of two) set of simulations, we evaluated the fixed effect parameter estimators, variance components of the random effects, and Correlation Score Test. There were $K = 2$ competing risks, and $J = 50$ or $J = 100$ centers (configurations 1 and 2, respectively). The center-specific random effects γ_{j1}, γ_{j2} followed a mean zero multivariate normal (MVN) distribution with variance components $\boldsymbol{\sigma}_j = (\sigma_1^2, \sigma_2^2, \rho_{12}) = (0.25, 0.25, -0.5)$. Using the re-parameterization described in Sections 2.1 and 2.2, this corresponds to the center-specific random effects vector $\boldsymbol{b}_j = (b_j^0, b_j^1, b_j^2)$ being generated from a MVN with mean zero and diagonal covariance matrix $\boldsymbol{D}$ with elements $\boldsymbol{\theta}_j = (\theta_0, \theta_1, \theta_2) = (0.125, 0.125, 0.125)$. The sample size within each center was fixed at $n_j = 20$ or $n_j = 50$ for different sub-configurations. In addition, we considered a single $N(0,1)$ covariate X_i with regression coefficients $\beta_1 = 0.5$ and $\beta_2 = 1.25$ for causes $k = 1$ and $k = 2$, respectively. Given $\beta_k, \boldsymbol{\gamma}_j$ and the covariate X_i, we generated a failure time T_i^0 for each subject within center j from an exponential distribution with rate parameter $\mu = \sum_{k=1}^2 \mu_k = \sum_{k=1}^2 \exp(\beta_k X_i + \gamma_{jk})$. We assigned a cause of failure for subject i in center j given a failure at time t using $Pr(\Delta_i = k | T_i^0 = t) = \mu_k/\mu$. Finally, all censoring occurred at time $\tau = 0.4$ in all configurations.

As shown in Table 1, the proposed method performs very well in estimating the parameters of interest. Also in Table 1, we present results of simulations where the center-specific random effects γ_{j1}, γ_{j2} were generated from a mean zero MVN with $\boldsymbol{\sigma}_j = (\sigma_1^2, \sigma_2^2, \rho_{12}) = (0.25, 0.25, 0)$, in order to assess the loss in efficiency due to unnecessarily estimating a correlation parameter when the true random effects are not correlated.

In Table 2, we evaluate the proposed CST and a likelihood ratio test (LRT) of the correlation between cause-specific hazards, via $H_0 : \rho = 0$. For each (J, n_j) configuration, the Type 1 error rate was calculated as the mean number of times H_0 when the random effects were generated from a mean zero MVN with $\boldsymbol{\sigma}_j = (0.25, 0.25, 0)$. Similarly, the Power was the mean number of rejections when the random effects were generated from a mean zero MVN with $\boldsymbol{\sigma}_j = (0.25, 0.25, -0.5)$. The CST seems to do almost as well as the LRT, attaining a type I error rate closer to the nominal 0.05 and achieving nearly as much power. More importantly, the CST is carried out in much less computation time, since it does not require fitting the full model.

In the second simulation study, we evaluated our estimators of the center-specific random effects $\{\gamma_{j1}, \gamma_{j2}\}$. Again, $K = 2$, $J = 50$, and $X_i \sim N(0,1)$ with regression coefficients $\beta_1 = 0.5$ and $\beta_2 = 1.25$ for $k = 1$ and $k = 2$, respectively. Of the 50 centers, we fixed the value of the random effects for center j' and allowed the random effects for the remaining 49 centers to come from a mean 0 MVN with $\boldsymbol{\sigma}_j = (\sigma_1^2, \sigma_2^2, \rho_{12}) = (0.25, 0.25, -0.5)$. The sample size for each of these 49 centers, n_j, $j \neq j'$ was set equal to the random draw from a $N(100, 40^2)$ variate bounded at 20. Given $\beta_k, \boldsymbol{\gamma}_j$ and X_i, we generated T_i^0 from an exponential distribution with rate parameter $\mu_i = \sum_{k=1}^2 \mu_{ik}$, where $\mu_{ik} = \exp(\beta_k X_i + \gamma_{jk})$, and assigned a cause of failure using $Pr(\Delta_i = k | T_i^0 = t) = \mu_{ik}/\mu$. Censoring again occurred at time $\tau = 0.4$.

We studied the performance of our estimators at different values of the random effects $\{\gamma_{j'1}, \gamma_{j'2}\}$ and at different $n_{j'}$ values. We compared the proposed method to an approach that fits separate frailty models for each k and therefore ignores the correlation between the center-specific random effects. As shown in Table 3, the proposed method produces center effect estimates with smaller mean square error, regardless of the center size and effect.

An expanded version of Table 3 is available in the Web Appendix (see Web Table S1). While both methods produce shrinkage, leveraging information on the correlation structure of the center-specific random effects leads to estimates with reduced shrinkage and higher rates of coverage. These gains

Table 1
Estimating regression coefficients and variance components: results from 500 simulated datasets

J	n_j		True value	Bias	ESD	CP	True value	Bias	ESD	CP
50	20	β_1	0.5	0.007	0.075	0.946	0.5	0.000	0.075	0.954
		β_2	1.25	0.002	0.072	0.950	1.25	0.002	0.074	0.942
		θ_1	0.125	−0.003	0.068	–	0	0.022	0.036	–
		θ_2	0.125	−0.001	0.088	–	0.125	−0.027	0.095	–
		θ_3	0.125	0.005	0.087	–	0.125	−0.021	0.089	–
50	50	β_1	0.5	−0.001	0.043	0.962	0.5	−0.001	0.043	0.962
		β_2	1.25	0.000	0.044	0.954	1.25	−0.004	0.046	0.944
		θ_1	0.125	−0.002	0.051	–	0	0.020	0.027	–
		θ_2	0.125	0.003	0.066	–	0.125	−0.020	0.073	–
		θ_3	0.125	−0.004	0.057	–	0.125	−0.026	0.069	–
100	20	β_1	0.5	0.003	0.050	0.960	0.5	0.000	0.051	0.960
		β_2	1.25	0.001	0.051	0.946	1.25	0.001	0.053	0.942
		θ_1	0.125	−0.005	0.053	–	0	0.017	0.029	–
		θ_2	0.125	0.003	0.066	–	0.125	−0.019	0.074	–
		θ_3	0.125	−0.001	0.064	–	0.125	−0.021	0.065	–
100	50	β_1	0.5	0.001	0.032	0.942	0.5	−0.001	0.033	0.944
		β_2	1.25	0.000	0.031	0.952	1.25	0.002	0.030	0.964
		θ_1	0.125	0.002	0.037	–	0	0.015	0.023	–
		θ_2	0.125	−0.001	0.043	–	0.125	−0.017	0.053	–
		θ_3	0.125	0.000	0.041	–	0.125	−0.012	0.049	–

in bias and coverage become more pronounced with decreasing sample sizes, and as the true values of the center effects deviate from the mean of the random effect distribution.

To examine our proposed excess cumulative incidence (ECI) center effect measure, we conducted simulations where the center-specific effects $\{\gamma_{j1}, \gamma_{j2}\}$ were known for all centers. We set $J = 50$, with n_j set equal to the maximum of 20 and a $N(100, 40^2)$ variate. Center-specific effects $\{\gamma_{j1}, \gamma_{j2}\}$ were each fixed at one realization from a MVN with mean 0 and $\boldsymbol{\sigma}_j = (\sigma_1^2, \sigma_2^2, \rho_{12}) = (0.25, 0.25, -0.5)$; theses were then treated as true center effects. We set $X_i \sim N(0, 1)$, with $\beta_1 = 0.5$ and $\beta_2 = 1.25$ for causes 1 and 2, respectively. Failure times and causes were then generated as presented earlier. Censoring was again at $\tau = 0.4$. The true ECI for each center was calculated at $t = 0.3$. In Table 4, we compare the proposed method with fitting separate cause-specific Cox frailty models. In terms of mean squared error of the ECI estimates, the proposed method generally out-performs the separate-models approach. A striking example, from Table 4, is the ECI estimates for Center $j = 23$, whose true ECI values for cause 1 and cause 2 are at opposite extremes.

5. Application

We applied the proposed methods to evaluate Organ Procurement Organizations (OPOs) with respect to two competing risks: (i) deceased-donor kidney transplantation (ii) death (prior to transplantation). We use data from the Scientific Registry of Transplant Recipients (SRTR). The SRTR data system includes data on all donor, wait-listed candidates, and transplant recipients in the United States, submitted by the members of the Organ Procurement and Transplantation Network (OPTN), and has been described elsewhere. The Health Resources and Services Administration (HRSA), U.S. Department of Health and Human Services provides oversight to the activities of the OPTN and SRTR contractors.

The study cohort included patients wait-listed between January 1, 2010 and April 30, 2010. Patients were followed

Table 2
Power and type I error of proposed correlation score test (CST), and likelihood ratio (LR) tests. The null hypothesis is no correlation between cause-specific hazards within center: results from 500 simulated datasets.

Number of centers	Subjects per center	Type I error		Power	
(J)	(n_j)	LRT	CST	LRT	CST
50	20	0.006	0.032	0.416	0.358
50	50	0.028	0.026	0.782	0.692
100	20	0.022	0.048	0.710	0.654
100	20	0.034	0.036	0.982	0.960

Table 3
Estimating center-specific effects: results from 500 simulations

			Proposed method				Ignornig correaltion of random effects
$n_{j'}$		True value	Bias	ESD	ASE	CP	Relative MSE
20	$\gamma_{j'1}$	0.0	−0.019	0.231	0.322	0.988	1.113
	$\gamma_{j'2}$	0.0	−0.015	0.239	0.305	0.980	1.084
	$\gamma_{j'1}$	0.5	−0.175	0.241	0.297	0.970	1.232
	$\gamma_{j'2}$	−0.5	0.168	0.249	0.327	0.972	1.248
	$\gamma_{j'1}$	1.0	−0.276	0.244	0.276	0.870	1.252
	$\gamma_{j'2}$	−1.0	0.397	0.244	0.354	0.838	1.700
40	$\gamma_{j'1}$	0.0	−0.007	0.222	0.263	0.988	1.093
	$\gamma_{j'2}$	0.0	−0.015	0.209	0.243	0.986	1.041
	$\gamma_{j'1}$	0.5	−0.097	0.208	0.232	0.946	1.203
	$\gamma_{j'2}$	−0.5	0.108	0.214	0.271	0.974	1.274
	$\gamma_{j'1}$	1.0	−0.141	0.203	0.209	0.916	1.242
	$\gamma_{j'2}$	−1.0	0.268	0.221	0.306	0.914	1.799
60	$\gamma_{j'1}$	0.0	−0.010	0.202	0.231	0.968	1.098
	$\gamma_{j'2}$	0.0	−0.005	0.187	0.210	0.976	1.074
	$\gamma_{j'1}$	0.5	−0.066	0.195	0.200	0.962	1.142
	$\gamma_{j'2}$	−0.5	0.070	0.209	0.240	0.958	1.148
	$\gamma_{j'1}$	1.0	−0.097	0.168	0.178	0.948	1.227
	$\gamma_{j'2}$	−1.0	0.219	0.219	0.276	0.890	1.666
80	$\gamma_{j'1}$	0.0	−0.018	0.181	0.210	0.988	1.095
	$\gamma_{j'2}$	0.0	−0.020	0.179	0.191	0.964	1.080
	$\gamma_{j'1}$	0.5	−0.071	0.180	0.180	0.954	1.170
	$\gamma_{j'2}$	−0.5	0.066	0.180	0.218	0.970	1.153
	$\gamma_{j'1}$	1.0	−0.105	0.161	0.161	0.894	1.206
	$\gamma_{j'2}$	−1.0	0.193	0.201	0.256	0.918	1.554
100	$\gamma_{j'1}$	0.0	−0.020	0.170	0.194	0.976	1.095
	$\gamma_{j'2}$	0.0	−0.015	0.157	0.176	0.978	1.065
	$\gamma_{j'1}$	0.5	−0.076	0.15	0.167	0.964	1.171
	$\gamma_{j'2}$	−0.5	0.059	0.197	0.203	0.932	1.123
	$\gamma_{j'1}$	1.0	−0.095	0.141	0.149	0.928	1.218
	$\gamma_{j'2}$	−1.0	0.155	0.196	0.242	0.944	1.494

from the date of listing until the earliest of receipt of a kidney transplant, death, removal from wait-list, or the end of the observation period, December 31, 2012. Using the proposed methods, we compared OPOs across the United States with respect to the cumulative incidence of receiving a deceased-donor transplant and the cumulative incidence of death prior to transplantation. The time point we chose was two years post wait-listing, an appropriate time horizon based on previous related analyses (e.g.,Fan and Schaubel, 2016). Patients receiving a living donor transplant were treated as independently censored, which is appropriate from the perspective that living-donor transplantation depends on many factors related to a patient's specific circumstances and largely independent of OPO. Note that living-donor transplantation was not a cause of our interest, rendering unappealing its inclusion as a separate cause.

Our study population included $n = 11,759$ patients across $J = 58$ OPOs across the United States A total of 2408 patients (20.5%) received a deceased-donor kidney transplant, while 1114 (9.5%) died first. We adjusted for the following patient-level covariates: age at listing, race, sex, body mass index, primary renal diagnosis, panel reactive antibody level and blood type. Owing to the large dimension of the covariate vector, we used a two-stage approach, as done in Kalbfleisch and Wolfe (2013), to obtain the risk-adjusted center effects (see also He and Schaubel, 2014b). Specifically, we estimated the patient-level covariates at the first stage by fitting a Cox model stratified by OPO. At the second stage, we estimated the cause-specific OPO effects by fitting the proposed model, using the patient-level linear predictor from the first stage as an offset. The estimated variance components are given by $\hat{\boldsymbol{\sigma}}_j = (\hat{\sigma}_1^2, \hat{\sigma}_2^2, \hat{\rho}_{12}) = (0.619, 0.031, 0.210)$. The estimated correlation was determined to be statistically significant, with the CST yielding a p-value of 0.021.

Figure 1 displays the estimated OPO-specific ECI's at 2 years post-listing, along with 95% confidence intervals. The ECIs of transplantation ranged from −0.120 to 0.404, and the ECIs of death ranged from −0.126 to 0.115. For a given OPO, a high ECI for transplantation and a low ECI for death represent good performance. We classified OPOs as low- or high-outliers based on the 95% confidence intervals.

Table 4
Estimating excess cumulative incidence: results from 500 simulation

			Proposed method				Ignornig correaltion of random effects
Cause	Center	True value	Bias	ESD	ASE	CP	Relative MSE
1	14	−0.170	0.029	0.022	0.027	0.850	1.018
	16	−0.096	0.024	0.029	0.034	0.926	0.750
	17	−0.181	0.003	0.017	0.023	0.990	3.432
	38	−0.138	0.017	0.024	0.029	0.958	1.226
	1	−0.179	0.023	0.020	0.026	0.918	1.462
	36	0.006	−0.009	0.038	0.038	0.932	0.950
	4	−0.033	0.001	0.034	0.037	0.948	1.010
	49	−0.070	0.018	0.029	0.034	0.944	0.777
	32	−0.047	0.010	0.031	0.035	0.960	0.969
	34	0.005	0.001	0.035	0.039	0.948	1.028
	23	0.344	−0.022	0.045	0.047	0.934	1.279
	19	0.118	−0.013	0.043	0.045	0.904	0.939
	13	0.142	−0.015	0.044	0.046	0.942	1.036
	15	0.127	−0.007	0.043	0.044	0.938	1.082
	18	0.210	−0.022	0.047	0.048	0.904	0.988
2	26	−0.222	0.020	0.019	0.025	0.932	2.265
	25	−0.140	0.014	0.027	0.029	0.936	1.196
	20	−0.137	0.013	0.025	0.030	0.952	1.209
	23	−0.199	0.014	0.021	0.025	0.950	2.122
	5	−0.078	0.006	0.030	0.032	0.950	1.056
	29	−0.017	0.004	0.034	0.033	0.922	0.954
	11	−0.020	0.002	0.031	0.034	0.954	0.987
	45	−0.043	0.006	0.032	0.033	0.934	0.965
	34	−0.058	0.007	0.029	0.033	0.952	0.993
	9	−0.009	0.001	0.031	0.033	0.946	1.016
	41	0.203	−0.016	0.037	0.037	0.928	1.065
	40	0.158	−0.012	0.035	0.036	0.934	1.076
	31	0.157	−0.016	0.038	0.038	0.900	0.969
	17	0.371	−0.020	0.037	0.034	0.902	1.327
	14	0.111	−0.009	0.033	0.035	0.926	1.245

We compared the proposed method to a method that ignores the correlation between the cause-specific center effects with respect to outlier classification (Web Table S2). While the two methods produced nearly identical classifications of OPOs based on the incidence of transplant, the proposed method classified 6 more OPOs as outliers than fitting separate frailty models by cause. This is a consequence of the reduction in shrinkage in the ECI estimates by the proposed method, due to leveraging the information on the correlation structure.

6. Discussion

In this report, we develop methods for evaluating center performance in the competing risks setting. We propose estimating center effects through cause-specific proportional hazards frailty models that allow correlation among a centers cause-specific hazards. We also propose a score test to test for the presence of correlation between a center's cause-specific hazards.

In our application, the cause-specific center effects do not seem to be strongly correlated. In scenarios where the correlation between cause-specific center effects is on the higher side, as maybe the case, for example, if there exists an unmeasured covariate influencing both outcomes, using the proposed method instead of currently available methods may produce a larger change in classification of centers than seen here. Since fitting the proposed model may be computationally cumbersome, we recommend first using the proposed CST, to determine if the proposed model is warranted (the alternative being cause-specific frailty models).

To ease computational burden while adjusting for case-mix in our application, we use a two-stage approach. In the first stage, we fit a model stratified by OPO to estimate the regression parameters associated with a large number of patient characteristics. In the second stage, we used the estimated regression parameters as an offset in the linear predictor of the instantaneous hazard in a random-effects model. Note that, following this two-stage approach has the added benefit of avoiding problems due to confounding between the patient-level covariates and the OPO-specific random-effects. As mentioned in Section 2.1, correlation between covariates and random-effects is a violation of our model assumption

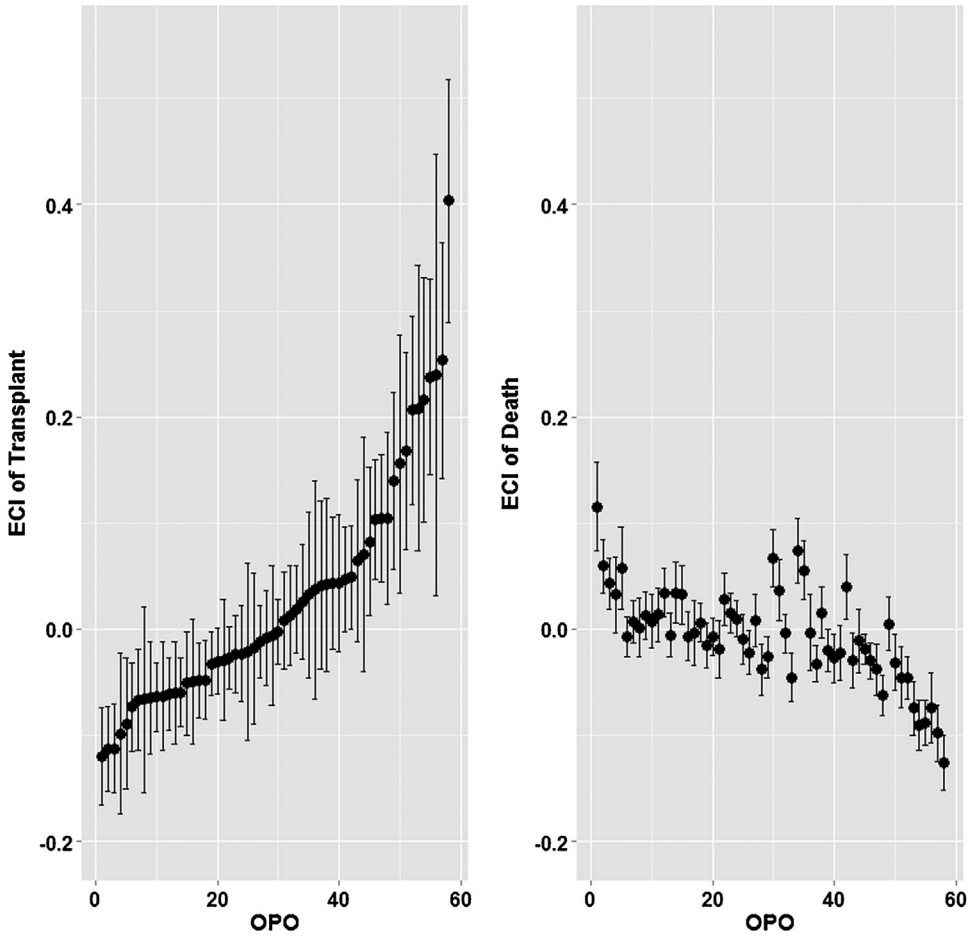

Figure 1. Analysis of scientific registry of transplant recipients (SRTR) data: caterpillar plots of excess cause-specific cumulative Incidence of death and kidney transplantation for 58 organ procurement organizations.

which may lead to biased estimates of center effects. However, using the above mentioned two-stage approach seems to rectify this issue. This is because, in the second stage, our random effects are estimated given $\mathbf{X}\hat{\boldsymbol{\beta}}$, where $\hat{\boldsymbol{\beta}}$ is estimated from the stratified model. This ensures that an unbiased estimate of $\hat{\boldsymbol{\beta}}$ is used while estimating the random effects. The random effects then estimated represent an estimate of variation between centers after all the within center variation has been accounted for accurately. It is possible that the random-effects may still be correlated with center-level averages of the covariates $\mathbf{X}$, and that this variation could further be partitioned into variation due to differences in center-level averages of the covariates $\mathbf{X}$ and other remaining variation between centers. The question of adjusting further for between-center differences while using a random-effects model may be a policy decision. An alternative, one-stage, approach to account for confounding by patient-level covariates is to use the between-method decomposition of covariates as suggested by Sjölander et al. (2013), where center-level averages of the covariates $\mathbf{X}$ are adjusted for.

7. Supplementary Materials

Web Appendix A, referenced in Section 4, and a web supplement containing R code and an example data file is available with this article at the *Biometrics* website on Wiley Online Library.

Acknowledgements

The authors thank the Associate Editor and Referee for their thoughtful suggestions which led to substantial improvement of the manuscript. This work was supported in part by National Institutes of Health Grant R01-DK070869. The data reported here have been supplied by the Minneapolis Medical Research Foundation (MMRF) 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 U.S. Government.

References

Ash, A. S., Fienberg, S. E., Louis, T. A., Normand, S. T., Stukel, T. A., and Utts, J. (2012). Statistical issues in assessing hospital performance. White paper, Committee of Presidents of Statistical Societies.

Balan, T., Boonk, S. E., Vermeer, M. H., and Putter, H. (2016). Score test for association between recurrent events and a terminal event. *Statistics in Medicine* **35**, 3037–3048.

Breslow, N. E. (1974). Covariance analysis of censored survival data. *Biometrics* **30**, 89–99.

Breslow, N. E. and Clayton, D. G. (1993). Approximate inference in generalized linear models. *Journal of the American Statistical Association* **88**, 9–25.

Cox, D. R. (1959). The analysis of exponentially distributed lifetimes with two types of failure. *Journal of Royal Statistical Society, Series B* **21**, 411–421.

Crowder, M. J. (2001). *Classical Competing Risks*. London: Chapman and Hall/CRC.

Do Ha, I., Christian, N. J., Jeong, J. H., Park, J., and Lee, Y. (2014). Analysis of clustered competing risks data using subdistribution hazard models with multivariate frailties. *Statistical methods in medical research*. Published Online. `https://doi.org/10.1177/0962280214526193`.

Fan, L. and Schaubel, D. E. (2016). Comparing center-specific cumulative incidence functions. *Lifetime Data Analysis* **22**, 1–21.

Gail, M. H. (1975). A review and critique of some models used in competing risk analysis. *Biometrics* **31**, 209–222.

Gorfine, M. and Hsu, L. (2011). FrailtyBased Competing Risks Model for Multivariate Survival Data. *Biometrics* **67**, 415–426.

Gorfine, M., Hsu, L., Zucker, D. M., and Parmigiani, G. (2014). Calibrated predictions for multivariate competing risks models. *Lifetime Data Analysis* **20**, 234–251.

Gray, R. J. (1992). Flexible methods for analyzing survival data using splines, with applications to breast cancer prognoses. *Journal of the American Statistical Association* **87**, 942–951.

Hart, A., Smith, J. M., Skeans, M. A., Gustafson, S. K., Stewart, D. E., Cherikh, W. S., et al. (2016). OPTN/SRTR Annual Data Report 2014: Kidney. *American Journal of Transplant* **16** (Suppl 2), 11–46.

He, K., and Schaubel, D. E. (2014a). Methods for comparing centerspecific survival outcomes using direct standardization. *Statistics in Medicine* **33**, 2048–2061.

He, K. and Schaubel, D. E. (2014b). Standardized Mortality Ratio for Evaluating Center-Specific Mortality: Assessment and Alternative. *Statistics in Biosciences* **7**, 1–26.

Kalbfleisch, J. D. and Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data*, 2nd Edition. New York: Wiley.

Kalbfleisch, J. D. and Wolfe, R. (2013). On monitoring outcomes of medical providers. *Statistics in Biosciences* **5**, 286–302.

Katsahian, S. and Boudreau, C. (2011). Estimating and testing for center effects in competing risks. *Statistics in Medicine* **30**, 1608–1617.

Ohlssen, D. I., Sharples, L. D., and Spiegelhalter, D. J. (2006). A hierarchical modelling framework for identifying unusual performance in health care providers. *Journal of the Royal Statistical Society, Series A* **170**, 865–890.

Prentice, R. L., Kalbfleisch, J. D., Peterson, A. V., Flournoy, V., Farewell, V. T., and Breslow, N. E. (1978). The analysis of failure times in the presence of competing risks. *Biometrics* **34**, 541–554.

Putter, H., Fiocco, M., and Geskus, R. (2007). Tutorial in biostatistics: competing risks and multi-state models. *Statistics in Medicine* **26**, 2389–2430.

Ripatti, S. and Palmgren, J. (2000). Estimation of multivariate frailtymodels using penalized partial likelihood. *Biometrics* **56**, 1016–1022.

Sjölander, A., Lichtenstein, P., Larsson, H, Pawitan, Y. (2013). Between-within models for survival analysis. *Statistics in Medicine* **32**, 3067–3076.

Spiegelhalter, D., Sherlaw-Johnson, C., Bardsley, M., Blunt, I., Wood, C., and Grigg, O. (2012). Statistical methods for healthcare regulation: Rating, screening and surveillance. *Journal of Royal Statistical Society, Series A* **175**, 1–47.

Therneau, T. (2015). Package 'coxme'. Mixed Effects Cox Models. R Package version 2.2–5.

VanRompaye, B., Goetghebeur, E., and Jaffar, S. (2010). Design and testing for clinical trials faced with misclassified causes of death. *Biostatistics* **11**, 546–558.

Van Rompaye, B., Eriksson, M., and Goetghebeur, E. (2015). Evaluating hospital performance based on excess causespecific incidence. *Statistics in Medicine* **34**, 1334–1350.

Varewyck, M., Goetghebeur, E., Eriksson, M., and Vansteelandt, S. (2014). On shrinkage and model extrapolation in the evaluation of clinical center performance. *Biostatistics* **15**, 651–664.

Zhao, L., Shi, J., Shearon, T. H., and Li, Y. (2015). A Dirichlet process mixture model for survival outcome data: Assessing nationwide kidney transplant centers. *Statistics in Medicine* **34**, 1404–1416.

Received March 2016. *Revised May* 2017. *Accepted May* 2017.