Journal of the American Statistical Association
Publication details, including instructions for authors and subscription information:
http://amstat.tandfonline.com/loi/uasa20

Frailty Survival Model Analysis of the National Deceased Donor Kidney Transplant Dataset Using Poisson Variance Structures

Shibao Feng, Robert A. Wolfe and Friedrich K. Port
Shibao Feng is Assistant Professor, Department of Biomathematics and Biostatistics and Department of Oncology, Georgetown University Medical Center, Washington, DC 20057 . Robert A. Wolfe is Professor, Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109. Friedrich K. Port is President, University Renal Research and Education Association, Ann Arbor, MI 48103. The authors thank John D. Kalbfleisch, Xihong Lin, and Julie C. Horrocks for their helpful suggestions, and Miles P. Finley for proofreading the manuscript. The editor, the associate editor, and two referees provided constructive comments on the earlier version of this article.
Version of record first published: 01 Jan 2012.

Frailty Survival Model Analysis of the National Deceased Donor Kidney Transplant Dataset Using Poisson Variance Structures

Shibao FENG, Robert A. WOLFE, and Friedrich K. PORT

In a recent study of transplant outcomes, donor age, cerebrovascular accident as the cause of death (CVA), renal insufficiency (serum creatinine >1.5 mg/dL), and history of hypertension have been identified as donor factors associated with elevated risk of kidney transplant failure. It is of great interest to know whether there remain other unmeasured donor factors associated with elevated risk of graft failure. In this article we study a sample of 6,024 deceased donor kidney transplants performed in 194 centers from 1995 to 2000. In addition to variation among transplant recipients, there are two other random effects: unmeasured donor and unrecorded center factors (data not available at the physician level). These two random effects are crossed, because the two kidneys from the same donor can be transplanted in different centers. Multivariate frailty models are applied to analyze the data. The likelihood functions of both parametric (e.g., with piecewise constant baseline hazard) and semiparametric multivariate frailty models are shown to be proportional to the likelihood functions of a class of mixed Poisson regression models. The penalized quasi-likelihood method is used as the numerical procedure for these mixed Poisson regression models. Thus we are able to estimate and model crossed random-effects structures for survival analysis. Although about 30% of recipient graft survival rate variation due to donor factors is explained by the measured donor characteristics, the remaining variation among donors in graft survival rate is still statistically significant, suggesting that there may be other unmeasured donor factors associated with a reduced graft survival rate. We also find significant variation of graft failure rates among transplant centers due to unrecorded center factors. Therefore, this study suggests that practice patterns at transplant centers and identification of other donor factors may merit further investigation.

KEY WORDS: End-stage renal disease; Mixed Poisson likelihood function; Multivariate survival analysis; Penalized quasi-likelihood; Semiparametric.

1. INTRODUCTION

More than 350,000 patients are receiving chronic dialysis treatment for end-stage renal disease (ESRD) kidney failure in the United States today. Although kidney transplantation was originally thought to be primarily a quality of life–enhancing treatment, patients receiving kidney transplantation have been found to have substantially longer lifetimes than patients remaining on dialysis (Wolfe et al. 1999). Most transplanted kidneys are obtained from deceased donors, although a substantial fraction are donated by living donors. Transplantation is a critical medical procedure that typically involves quick action to identify a potential donor, assemble a medical team to remove the organ, identify a candidate from among those awaiting a transplant, transport the organ, and implant the organ. The rules used to prioritize the candidates awaiting a transplant have been developed to consider both efficiency and equity. Efficiency goals include ensuring that organs are transplanted before they become unusable and that organs are allocated to candidates who will benefit the most. Equity considerations assure that the allocation system is "fair." Although the details of the rules are complicated and differ by region of the country, candidates generally are registered on a waiting list, and deceased donor kidneys are offered to candidates ordered with respect to their waiting times.

Unfortunately, there are many more patients joining the waiting list for deceased donor kidney transplantation than there are organs available each year, which has led to longer waiting times and more deaths among wait-listed candidates. The number of patients on the national kidney waiting list has increased from 22,063 in 1992 to 51,144 in 2001 (132%), whereas the number of kidney waiting list deaths has increased from 1,077 to 2,918 in those same years (171%). However, the corresponding numbers of deceased donor kidney transplant recipients has increased only moderately, from 7,202 to 8,202 (14%). Although the absolute number of deaths has increased, posttransplant death rates in the first year for patients age 18–34 have declined from 30 to 15 per 1,000 patient years between 1992 and 2001, due in large part to improvements in immunosuppressant therapy (OPTN/SRTR 2003, table 5.7; Merion 2003). One option for meeting the shortage of organs is to use organs that carried too high of a risk of failure in previous years but that have acceptable levels of risk with current transplant outcomes. Kidneys from deceased donors age 60 years or older and from those age 50–59 years with at least two of three defined conditions [cerebrovascular accident as the cause of death (CVA), renal insufficiency (serum creatinine >1.5 mg/dL), and history of hypertension] have been identified as expanded criteria donor (ECD) kidneys (Port et al. 2002), that is, kidneys whose graft loss relative rate was greater than 1.70 when compared with a low-risk group. Until recently, kidneys from such donors were often considered "marginal" or unsuitable. Ojo et al. (2001) showed that although outcomes from ECD kidneys are worse than those from standard kidneys, both life expectancy and quality of life are improved for ECD kidney recipients compared with patients remaining on dialysis treatment.

Increasing the use of higher-risk donor organs is intended to lead to outcomes that are no worse than were deemed acceptable in previous years, while potentially increasing the current

Shibao Feng is Assistant Professor, Department of Biomathematics and Biostatistics and Department of Oncology, Georgetown University Medical Center, Washington, DC 20057 (E-mail: *sf96@georgetown.edu*). Robert A. Wolfe is Professor, Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109. Friedrich K. Port is President, University Renal Research and Education Association, Ann Arbor, MI 48103. The authors thank John D. Kalbfleisch, Xihong Lin, and Julie C. Horrocks for their helpful suggestions, and Miles P. Finley for proofreading the manuscript. The editor, the associate editor, and two referees provided constructive comments on the earlier version of this article.

Journal of the American Statistical Association
September 2005, Vol. 100, No. 471, Applications and Case Studies
DOI 10.1198/016214505000000123

number of available donors dramatically. A 70% increase in failure rates compared with today's rates may lead to a post-transplant death rate of 26 per 1,000 patient years, which is lower than the risk reported for 1992. In recognition of the fact that risk varies by more than a factor of 2 among types of donors, the current transplant policy provides informed decision making on the part of potential recipients by identifying which organs are from higher-risk donors. This was recently implemented by creating a separate registration process to be added to the waiting list for these higher-risk organs. Patients on the alternate waiting list are prioritized for ECD kidneys in addition to non-ECD kidneys, which could substantially shorten a candidate's waiting time and thus reduce that candidate's probability of waiting list death. The degree of confidence in the classification of higher-risk organs depends not only on the magnitude of the differences among donors based on their measured characteristics, but also on the remaining unexplained variation among donors. Frailty (or random-effects) survival models offer direct estimates of the variation among donors in recipient death rates, even in this case with only two replications based on two kidneys per donor transplanted into two different recipients. Simulation results reported in this article document the accuracy of such analyses. It is shown here that the ECD characteristics identified by Port et al. (2002) explain nearly 30% of the variability among donors.

In addition to variation among donors, there is substantial variation among transplant centers in recipient survival. The decision to accept an organ for transplantation lies with both the transplant surgeon and the patient. When kidneys become available from a deceased donor, they are offered for specific patients according to a regional and national allocation system that considers donor and recipient matching of blood group and tissue type, level of recipient preformed antibodies, and priority based on each candidate's time on the transplant waiting list. Patients may choose in advance to be eligible for organs with reduced probability of function according to ECD characteristics (a binary variable). The transplant surgeon may turn down an offered organ because of poor quality or other factors. Such a turndown leads to reallocation to the next ranked candidate, usually at another center. A surgeon may choose to accept this offer even if the offer was previously rejected at another center. Thus center factors play an important rule, even though the offers of organs are regulated by well-defined allocation rules. Because some centers tend to use more higher-risk donors than others, it is important to account for both donor and center factors simultaneously when evaluating the magnitude of either factor. Outcomes of organ function and patient survival are remarkably complete; the Scientific Registry of Transplant Recipients (SRTR) ascertains organ failure and death through both reports from transplant centers and from matching with Social Security and Medicare files.

In ascertaining the effectiveness of the ECD criteria in classifying high-risk organs, it is of great interest to find out how much of the variation in graft survival among donor kidneys is explained by these ECD characteristics after adjusting for other important covariates, such as recipient characteristics. We also adjusted for and estimated the amount of variation in outcomes among transplant centers. In addition to adjusting the analysis for potential confounding between ECD utilization and transplant center outcomes, this also allows identification of centers with good outcomes, which could be useful for generating future clinical trial hypotheses. To address these questions, we used multivariate frailty models for the analysis, where the unobserved donor characteristics and center factors were treated as two separate frailties affecting graft survival. Note that these frailties are clustered within donors and within centers. They are also crossed, because the two kidneys from the same donor can be transplanted in different centers.

Let i denote the subjects, $i = 1, \ldots, n$. Then the multivariate frailty model is written as

$$\lambda_i(t; \mathbf{x}_i, \mathbf{z}_i, \boldsymbol{\omega}) = \lambda_0(t) e^{\mathbf{x}_i'\boldsymbol{\beta} + \mathbf{z}_i'\boldsymbol{\omega}}. \tag{1}$$

In this study, $\lambda_i(\cdot)$ denotes the hazard function for the ith individual conditional on $\mathbf{x}_i'$, $\mathbf{z}_i'$, and $\boldsymbol{\omega}$; $\lambda_0(\cdot)$ is either a parametric or nonparametric baseline hazard function; $\mathbf{x}_i'$ and $\boldsymbol{\beta}$ correspond to a p-dimensional vector of time-independent covariates and coefficients (although the analysis could be extended to model time-dependent covariates); $\mathbf{z}_i'$ is $D+F$ dimensional with D dummy variables for the donor and F dummy variables for the facility; and $\boldsymbol{\omega} = (\boldsymbol{\omega}_D, \boldsymbol{\omega}_F)$ is a $(D+F)$-dimensional vector of frailties. [$(\boldsymbol{\omega}_D, \boldsymbol{\omega}_F)$ are assumed to have independent normal distributions, $\boldsymbol{\omega}_D \sim \mathrm{N}(\mathbf{0}_{D\times 1}, \theta_D^2 \mathbf{I}_{D\times D})$ and $\boldsymbol{\omega}_F \sim \mathrm{N}(\mathbf{0}_{F\times 1}, \theta_F^2 \mathbf{I}_{F\times F})$.]

The rest of the article is organized as follows. Section 2 describes the national kidney transplant dataset. Section 3 shows the proportionality of the likelihood functions for the multivariate frailty models and a class of mixed Poisson regression models, and introduces PQL as the computational procedure. Section 4 describes the analysis results of the national kidney transplant dataset. Section 5 presents a simulation study that demonstrates the performance of the proposed procedure under two lognormal crossed frailties. Section 6 concludes with discussion of other medical applications and extensions of the method.

2. KIDNEY TRANSPLANT DATASET

Data are analyzed from the SRTR, which contains information on solid organ transplantation in the United States including kidney, heart, liver, heart and lung, lung, and pancreas transplants. We study deceased donor kidney transplants performed for patients without previous kidney or kidney and pancreas transplants between 1995 and 2000, which includes a total of 40,721 transplants, 6,178 of which were ECD transplants. The follow-up time is through October 31, 2002. Because of the complexity of the crossed random-effects structures, we are unable to analyze the whole dataset using the computer that we can access. To include more centers in the analysis to yield a solid center frailty estimate, up to 12 ECD and 12 non-ECD kidney transplants are randomly selected from each transplant center with the other organ from each donor. If we regard the unselected transplants as missing data, then these transplants can be considered to be missing at random (MAR), because missingness depends on the number of ECD and non-ECD kidney transplants performed in a center between 1995 and 2000. Large centers generally have more "missing transplants." Center size (a binary variable) is included as a covariate in all analysis models, to reduce the bias caused by this sampling scheme. Only donors donating both kidneys are included, because single-kidney donors contribute less information to the

Table 1. Descriptive Statistics for Some Selected Donor, Center, and Recipient Level Covariates

Covariates	*Average$_{(SD)}$*	*Covariates*	*Average$_{(SD)}$*
Donor age (years)	$41_{(18)}$	Recipient BMI <20	7%
Age <10	4%	20–24	23%
Age 10–39	39%	25–29	22%
Age 40–49	17%	≥ 30	16%
Age 50–59	24%	Missing	32%
Age ≥ 60	17%	Primary cause of ESRD	
Donor gender (female)	44%	Diabetes	24%
Donor race (black)	9%	Hypertension	23%
Donor hypertension	28%	GN	25%
Donor cause of death (stroke)	51%	Other	28%
Terminal creatinine >1.5 mg/dL	13%	0 DR mismatch	37%
Center size (>300 operations)	20%	1 DR mismatch	42%
Recipient age (years)	$49_{(14)}$	2 DR mismatches	22%
Recipient gender (female)	39%	CIT ≤ 24 hours	69%
Recipient race (black)	26%	>24 hours	23%
Time on dialysis (years)	$3_{(2)}$	Missing	8%

evaluation of variation among donors. Most patients are treated in large transplant centers, so small centers (i.e., centers performing fewer than three ECD kidney transplants or five kidney transplants altogether) are excluded from the analysis. In fact, although this strategy eliminates 24% of centers, it eliminates fewer than 4% of transplants. This results in 6,024 deceased donor kidney transplants distributed in 194 centers, with 949 of these 3,012 donors being ECDs. The mean number of kidney transplants selected from each center is 31, with a minimum of 12 and a maximum of 74, whereas the mean number of ECD kidney transplants is 10, with a minimum of 3 and a maximum of 26. As a result of the constraint to include both kidneys from the same donor, most centers have more than 24 kidney transplants in the analysis. Failure time (recorded in days) is defined as the time from transplantation to allograft failure or death, whichever occurred first. No recipients were lost during the follow-up, so censoring occurred only at the end of the study. The censoring rate is 72%.

The fixed-effects covariates used in the analysis are the same as those identified by Port et al. (2002), who established the ECD criteria that are now in use throughout the United States. The four ECD factors: donor age [0–9, 10–39 (baseline), 40–49, 50–59, ≥ 60 years], history of hypertension, impaired renal function (terminal serum creatinine >1.5 mg/dL), and cause of death (stroke vs. other) and four other donor factors: sex (female vs. male), race (black vs. nonblack), history of diabetes, and year of donation [95 (baseline), 96, 97, 98, 99, and 00] are analyzed. One center-level covariate, center size (>300 vs. 0–300 kidney transplants) is also included in the models. The unobserved donor characteristics and center factors are treated as two crossed frailties. Recipient factors are also included in the models, to more precisely evaluate the graft outcome by donor and center effects. All models contain the following recipient covariates: recipient age, sex (female vs. male), race (black vs. nonblack), ethnicity (hispanic vs. nonhispanic), panel-reactive antibody (PRA) level [0–9 (baseline), 10–79, and >80%], prior time on ESRD therapy [<1, 1–3 (baseline), and >3 years], pretransplantation blood transfusions [yes, no (baseline), and missing], body mass index (BMI) [<20, 20–24 (baseline), 25–29, and ≥ 30], cause of ESRD [diabetes, hypertension, glomerulonephritis (GN) (baseline), or other], cold ischemia time (CIT) [≤ 24 (baseline), 24 hours, and missing], number of antigen A mismatches [0 (baseline), 1, and 2], number of antigen B mismatches [0 (baseline), 1, and 2], and number of antigen DR mismatches [0 (baseline), 1, and 2]. Table 1 gives the descriptive statistics for some of these donor, center, and recipient covariates.

3. POISSON VARIANCE STRUCTURE LIKELIHOODS

There has already been some work done on the extension of survival analysis regression techniques from shared to multivariate frailty survival data (e.g., Sastry 1997; Vaida and Xu 2000; Ripatti and Palmgren 2000; Parner 1998). For crossed random-effects structures, a multilevel logistic model has been developed to examine discrete hazard survival data (Ecochard and Clayton 1998). However, there are few methods that can model the crossed random-effects structure in the kidney transplant dataset.

In survival analysis, baseline hazard function is often modeled nonparametrically, with the mass being distributed only at the observed failure times. For independent failure time data, this procedure yields the Cox regression estimates (Breslow 1972). Moreover, for the cell means of contingency tables with Poisson data, loglinear models are exactly equivalent to log hazard models for survival data with an independent and noninformative censoring mechanism, a piecewise constant hazard, and categorical covariates (Laird and Olivier 1981). Lawless and Zhan (1998) further proposed applying mixed Poisson regression models to analyze interval-grouped recurrent-event data. In this article we show that under the first two conditions (Laird and Olivier 1981), the conditional likelihood of a multivariate frailty model is proportional to the conditional likelihood of a multivariate mixed Poisson regression model. Note that in frailty models, the independence and noninformative properties of the censoring mechanism are conditional on the frailties (Andersen, Borgan, Gill, and Keiding 1993).

We use the notation described in Section 1, and let t_i be the follow-up time for the ith subject and let δ_i be the event indicator, $\delta_i = 1$ if the ith subject has an event and 0 if censored. The study period is divided into m intervals, and the baseline hazard is assumed to be constant on each of these intervals,

$$\lambda_0(t) = \lambda_k, \qquad t \in \Omega_k = (t_{k-1}, t_k], \qquad k = 1, \ldots, m.$$

The fitting procedure for the baseline hazard function can be considered nonparametric when each t_k corresponds to an observed failure time and m is the total number of distinct observed failure times in the study. We let t_{ik} be the total time for subject i in Ω_k and let δ_{ik} be the event indicator for subject i in Ω_k. Under independent and noninformative censoring, for the kth interval, the contribution of the ith subject to the conditional likelihood is

$$L_{ik}(\lambda_k, \boldsymbol{\beta}; \boldsymbol{\omega}) = \left(\lambda_k e^{\mathbf{x}_i'\boldsymbol{\beta}+\mathbf{z}_i'\boldsymbol{\omega}}\right)^{\delta_{ik}} e^{-\lambda_k t_{ik} e^{\mathbf{x}_i'\boldsymbol{\beta}+\mathbf{z}_i'\boldsymbol{\omega}}} \propto \left(\lambda_k t_{ik} e^{\mathbf{x}_i'\boldsymbol{\beta}+\mathbf{z}_i'\boldsymbol{\omega}}\right)^{\delta_{ik}} e^{-\lambda_k t_{ik} e^{\mathbf{x}_i'\boldsymbol{\beta}+\mathbf{z}_i'\boldsymbol{\omega}}}$$

for $i = 1, \ldots, I$ and $k = 1, \ldots, m$. Thus δ_{ik} has a Poisson likelihood with mean $\lambda_k t_{ik} e^{\mathbf{x}_i'\boldsymbol{\beta}+\mathbf{z}_i'\boldsymbol{\omega}}$ if the true baseline hazard is constant on Ω_k, that is, the length of Ω_k is sufficiently short. The full conditional likelihood given $\boldsymbol{\omega}$ is then

$$L(\lambda_1, \ldots, \lambda_m, \boldsymbol{\beta}; \boldsymbol{\omega}) = \prod_{i=1}^{n}\prod_{k=1}^{m}\left(\lambda_k e^{\mathbf{x}_i'\boldsymbol{\beta}+\mathbf{z}_i'\boldsymbol{\omega}}\right)^{\delta_{ik}} e^{-\lambda_k t_{ik} e^{\mathbf{x}_i'\boldsymbol{\beta}+\mathbf{z}_i'\boldsymbol{\omega}}} \propto \prod_{k=1}^{m}\prod_{i=1}^{n}\left(\lambda_k t_{ik} e^{\mathbf{x}_i'\boldsymbol{\beta}+\mathbf{z}_i'\boldsymbol{\omega}}\right)^{\delta_{ik}} e^{-\lambda_k t_{ik} e^{\mathbf{x}_i'\boldsymbol{\beta}+\mathbf{z}_i'\boldsymbol{\omega}}},$$

which is proportional to the conditional likelihood of a Poisson regression model with parameters $(\lambda_1, \ldots, \lambda_m, \boldsymbol{\beta})$ and offsets t_{ik} in the special case where all Poisson counts are equal to 0 or 1. Thus the observed likelihood from a multivariate frailty model is proportional to the observed likelihood from a mixed Poisson regression model,

$$L_{\text{obs}}(\lambda_1, \ldots, \lambda_m, \boldsymbol{\beta}, \theta_D, \theta_F) \propto \frac{1}{\theta_D^D \theta_F^F}\int_{\mathbb{R}^{D+F}}\prod_{k=1}^{m}\prod_{i=1}^{n}\left(\lambda_k t_{ik} e^{\mathbf{x}_i'\boldsymbol{\beta}+\mathbf{z}_i'\boldsymbol{\omega}}\right)^{\delta_{ik}} e^{-\lambda_k t_{ik} e^{\mathbf{x}_i'\boldsymbol{\beta}+\mathbf{z}_i'\boldsymbol{\omega}}} \times e^{-1/2(\boldsymbol{\omega}_D'\boldsymbol{\omega}_D/\theta_D^2+\boldsymbol{\omega}_F'\boldsymbol{\omega}_F/\theta_F^2)}\, d\boldsymbol{\omega}.$$

Therefore, for likelihood-based inference, one can use mixed Poisson regression models instead of multivariate frailty models, where time-dependent covariates and left-truncation data can be easily modeled. Recently, Ma, Krewski, and Burnett (2003) also used a Poisson approach to modeling nested frailties, but their approach depends on an orthodox best linear unbiased predictor (BLUP) approach for the random effects.

The observed likelihood of the mixed Poisson regression models proposed here is proportional to the observed likelihood of parametric (with piecewise constant baseline hazard) or semiparametric multivariate survival frailty models. Thus, the estimators from the mixed Poisson regression models follow the asymptotic properties of the maximum likelihood estimators (MLEs) from the piecewise exponential frailty models, which are consistent and asymptotically normal. The estimators are also asymptotically equivalent to the nonparametric MLEs (NPMLEs) from the semiparametric multivariate frailty models when the baseline hazard function is discretized with one distinct event time per interval. Under regularity conditions, Parner (1998) proved the consistency, asymptotic normality, and efficiency of the NPMLEs from multivariate gamma frailty models. The asymptotic properties of the NPMLEs from other frailty models, such as multivariate lognormal frailty models, are not yet established.

To yield the full maximum likelihood inference based on the marginal likelihood of a mixed Poisson regression model, numerical integrations are required to calculate the log-likelihood, score equations, and information matrix. For some complicated problems, such as the crossed random effects in the kidney transplant data, the dimension of integrals may be irreducibly high, which is currently intractable. To address this limitation, different methods have been proposed to approximate the high-dimensional integrals. For example, the penalized quasi-likelihood (PQL) method (Breslow and Clayton 1993) is a simple and general technique for approximate estimation in generalized linear mixed models (GLMMs). We concentrate on the lognormal frailty models in this article, because existing statistical software, such as SAS macro Glimmix, can be applied to make inference on these models.

Feng and Wolfe (2005) found that the proposed nonparametric Poisson PQL method for lognormal frailty models is equivalent to the penalized partial likelihood (PPL) approach (Ripatti and Palmgren 2000). However, in their implementation of PPL, instead of substituting the second derivative of the penalized full likelihood into the approximate profile likelihood for θ, the second derivative of the PPL is used, and the resulting estimates for both θ and their standard errors (SEs) are underestimated. The proposed PQL method yields more accurate estimates for these parameters by using the full likelihood (Feng and Wolfe 2005).

4. EVALUATION OF EXPANDED CRITERIA DONOR CHARACTERISTICS

Multivariate frailty models (1) are applied to the kidney transplant dataset, where the unmeasured donor characteristics and the unrecorded center factors are modeled as two separate lognormal frailties. These two frailties are crossed because of the 3,012 donors donating both kidneys in the dataset, 2,060 had their two kidneys transplanted in different centers. To demonstrate the utility of these random-effects survival models, results are shown for a variety of models, including one with no random effects. The baseline hazard function is assumed to be constant for the following 10 time intervals: day 1–day 30, day 31–day 90, day 91–day 270, three 270-day intervals, three 300-day intervals, and one 579-day interval. The interval length is increased gradually as the condition of the recipients becomes more stable. The piecewise exponential model is quite general and can approximate various shapes of the baseline hazard function. We choose the model because of its simplicity. Fixed-effects and frailty estimates from models with unspecified baseline hazard function are often very similar to the estimates from models with piecewise baseline hazard function when 8–10 intervals are assumed for the study period (Lawless and Zhan 1998). For this dataset, most of the fixed-effects estimates from the piecewise exponential model ignoring donor- and center-level frailties change only at the third decimal place, when compared with the estimates from a Cox model that ignores the frailties. Models are fitted with and without ECD characteristics. We determine whether there are further unmeasured ECD characteristics associated with reduced graft survival by comparing the estimated donor frailty variances between these two models and testing whether the donor

Table 2. Parameter Estimates From the Kidney Transplant Dataset (N = 6,024 recipients)

	Estimate$_{(SE)}$		
		Mixed Poisson model	
Variables	Model 1 (Cox)	Model 2 (without donor characteristics)	Model 3 (with ECD characteristics)
Recipient age (10 years)	$44_{(20)}{}^{a}$	$71_{(20)}{}^{b}$	$42_{(22)}$
Recipient female	$-79_{(52)}$	$-96_{(54)}$	$-99_{(54)}$
Recipient black	$203_{(61)}{}^{b}$	$217_{(65)}{}^{b}$	$220_{(65)}{}^{b}$
Time on dialysis			
≤1 year	$-182_{(65)}{}^{b}$	$-189_{(68)}{}^{b}$	$-178_{(68)}{}^{b}$
>3 years	$26_{(58)}$	$42_{(61)}$	$30_{(61)}$
Recipient BMI			
<20	$178_{(108)}$	$140_{(113)}$	$161_{(112)}$
25–29	$57_{(74)}$	$35_{(77)}$	$54_{(77)}$
≥30	$213_{(80)}{}^{b}$	$174_{(83)}{}^{a}$	$199_{(83)}{}^{b}$
Cause of ESRD			
Diabetes	$223_{(71)}{}^{b}$	$227_{(75)}{}^{b}$	$226_{(74)}{}^{b}$
Hypertension	$128_{(74)}$	$126_{(78)}$	$124_{(78)}$
Other	$29_{(70)}$	$29_{(74)}$	$30_{(73)}$
1 DR mismatch	$192_{(62)}{}^{b}$	$200_{(66)}{}^{b}$	$195_{(65)}{}^{b}$
2 DR mismatch	$295_{(75)}{}^{b}$	$293_{(79)}{}^{b}$	$316_{(79)}{}^{b}$
CIT (>24 hrs)	$73_{(57)}$	$120_{(61)}{}^{a}$	$90_{(61)}$
Donor age			
50–59 years	$277_{(80)}{}^{b}$		$295_{(85)}{}^{b}$
≥60 years	$588_{(78)}{}^{b}$		$619_{(83)}{}^{b}$
Donor hypertension	$106_{(60)}$		$140_{(63)}{}^{a}$
Donor CVA	$130_{(61)}{}^{a}$		$122_{(65)}$
High creatinine (>1.5 mg/dL)	$167_{(69)}{}^{a}$		$170_{(74)}{}^{a}$
Center size (>300)	$-143_{(64)}{}^{a}$	$-145_{(92)}$	$-147_{(92)}$
Donor variance ($n = 3{,}012$)		$192_{(66)}{}^{b}$	$137_{(63)}{}^{a}$
Center variance ($n = 194$)		$99_{(23)}{}^{b}$	$103_{(23)}{}^{b}$

NOTE: All entries are multiplied by 1,000.
[a]For p value <.05.
[b]For p value <.01.

frailty could be omitted in the full model. The estimated center frailty variance also summarizes the varying center performance due to unrecorded transplant center factors.

Four survival models are fitted, where transplant year, center size, and recipient characteristic covariates are adjusted for in all models. These models include a simple Cox regression model ignoring frailties but with all donor characteristics (model 1), a piecewise exponential crossed lognormal frailty model without donor characteristics (model 2), a piecewise exponential crossed lognormal frailty model with only the four ECD characteristics (model 3), and a piecewise exponential crossed lognormal frailty model that includes all eight donor characteristics (model 4). Some of the estimates from the first three models are tabulated in Table 2. Estimates from model 4 are very similar to those from model 3 and hence are omitted from Table 2.

In model 2, the donor-level frailty variance parameter is significantly different from 0 and estimated to be .192, with SE .066 and p value .002. After adding ECD characteristic covariates into model 3, the donor-level frailty variance estimate decreases by 30% to .137, with SE .063 and p value .015. Therefore, we conclude that, after adjusting for other covariates, the ECD characteristics explain a substantial portion of the graft survival outcome variation due to donor kidney among recipients. However, there may exist further unmeasured donor characteristics that are associated with varying risk of graft loss.

We further perform a stepwise frailty model regression analysis to apportion the 30% of variation among ECD characteristics. Three ECD characteristics—donor age, CVA, and history of hypertension—are significantly associated with each other, whereas donor renal insufficiency is uncorrelated with any of them. Two more models are fitted: a piecewise exponential crossed lognormal frailty model with only donor age (model 5), because donor age is the most important predictor, and a piecewise exponential lognormal frailty model with donor age, CVA, and history of hypertension (model 6). Transplant year, center size, and recipient characteristic covariates are adjusted for in both models. The donor-level frailty variance estimate decreases from .192 in model 2 to .143 in model 5, suggesting that in terms of recipient graft survival rate, donor age explains 26% of the variation due to donor kidneys. In model 6, the donor-level frailty variance estimate further decreases to .138, implying that an additional 3% of the variation due to donor kidneys is explained by donor CVA and history of hypertension. Finally, donor renal insufficiency explains less than 1% of the variation due to donor kidneys when the foregoing three ECD factors are adjusted for in model 3.

Figure 1 plots the estimated density functions of the standard donor and ECD random-effects estimates from models 2 and 3 by the density(·) function of the software S–PLUS. These estimates of the random effects are empirical Bayes estimates that shrink toward mean 0 and are optimal for individual random effects (Shen and Louis 1998). The mean of the ECD random-effects estimates is higher than the mean of the non-ECD random-effects estimate in Figure 1(a), but both means are equal 0 after adjusting for the ECD factors in model 3. The estimated distribution functions are less variable in Figure 1(b), as indicated by the smaller standard deviations. Figure 1 gives a level of assurance in applying the ECD criteria to classify high-risk organs.

Multimodality is apparent in both figures, and the modes correspond to the random-effects estimates of standard donors and ECDs with zero, one, and two recipient kidney failures or deaths. Although both figures suggest a nonnormal distribution, the distributional assumption of normality is made less important by simulation results with mixture distributions by Lai and Shih (2003), which suggest that the estimates of random-effects parameters, such as the donor frailty variance parameter in our example, are asymptotically consistent even though the mixing distribution is far from the assumed normal parametric family.

The ECD factors are highly significant as predictors of outcome in model 3. Kidneys from the 50–59 and 60 and older age groups are estimated to be 1.34 [$p = .001$; 95% confidence interval (CI) = 1.14–1.59] and 1.86 ($p < .001$; 95% CI = 1.58–2.19) times more likely to fail when compared with kidneys from the 10–39 age group. The estimated relative risks for CVA, renal insufficiency, and hypertension are 1.13 ($p = .063$; 95% CI = .99–1.28), 1.19 ($p = .021$; 95% CI = 1.03–1.37), and 1.15 ($p = .028$; 95% CI = 1.02–1.30) when compared with donors without these characteristics. Three more donor characteristic covariates—donor sex, race, and history of diabetes—are added in model 4. However, none of them is significant, and

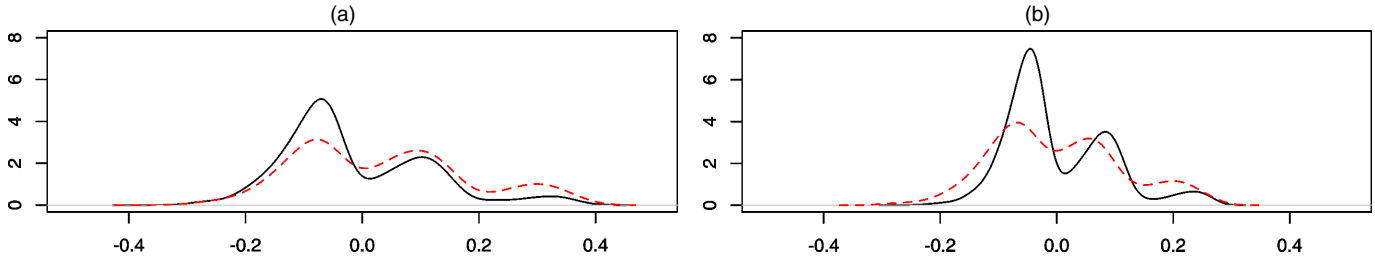

Figure 1. Estimated Density Functions of the BLUP Estimates of the Standard Donor Random Effects Between 2,063 Donors and the ECD Random Effects Between 949 Donors. (a) Estimates from model 2 (—— standard mean = −.02, SD = .13; - - - ECD mean = .03, SD = .15); (b) estimates from model 3 (—— standard mean = 0, SD = .09; - - - ECD mean = 0, SD = .11). The mean of the estimated random effects is higher for the ECD donors in (a), whereas in (b), after adjusting for the ECD factors, the means are both 0. Standard deviations are smaller in (b).

there is little change in the other parameter estimates. Hence we conclude that with respect to graft failure rate, the four ECD characteristics give an effective summary of variation in this dataset.

The random-effects models yield different estimates for the fixed effects in these models when compared with the fixed-effects estimates in the Cox model that does not include donor- and center-level frailties (Henderson and Oman 1999). The parameter estimates from the Cox model have a recipient population–level interpretation, whereas those from the frailty models have a recipient specific–level interpretation. However, the SE estimates of the regression coefficients from the Cox model ignoring donor and center frailty are invalid when these frailties are present. Compared with the fixed-effect estimates in the Cox model, the donor hypertension estimate is increased by 32%, and the SE of center size is increased by 44% after introducing donor and center frailties in model 3. Recipient sex and cold ischemia time coefficient estimates are also increased, by 25% and 23%. There is a slight increase for most of the other fixed-effects estimates and their corresponding SEs in model 3. Because ECD recipients are older (mean, 51.5 years) than non-ECD recipients (mean, 47.2 years), the estimated recipient age coefficient is reduced by about 40% when ECD status is accounted for by including donor factors in the model (model 2 vs. models 1 and 3).

The center-level frailty variance estimates are highly significant and remain unchanged in models 2 and 3, suggesting the existence of substantial remaining unexplained variation among centers. The estimated frailties for the centers provide an evaluation of the magnitude of unrecorded center factors. The log relative risk (RR) estimated for frailty one standard deviation above the mean in model 3 is .32, corresponding to a 38% higher mortality rate (RR = 1.38). The estimates for center frailties are also empirical Bayes estimates that shrink toward the center population mean 0. Figure 2 shows the scatterplot of the estimated log RRs of unobserved center factors vs. the proportion of ECD kidney transplants of those 194 centers. We note that these two variables are not associated with each other, suggesting that the different tendencies of using high-risk organs among transplant centers are not the sources of the unobserved center variation. We also note that there are substantial differences among unobserved transplant center effects estimates. Ten transplant centers have log RR estimates $>.4$, whereas three centers report log RR estimates $<-.4$. The centers with the largest and smallest RR estimates in the dataset are not small centers. The numbers of transplants performed in these two centers are 35 and 48, whereas the median of the number of transplants per center is 31.

5. SIMULATION RESULTS

We now report simulation results that assess the bias and variance of estimates from mixed Poisson PQL regression models for multivariate crossed lognormal frailty survival data. Efficiency comparisons of parametric versus unspecified baseline models are also shown.

The dataset contains 200 observations, of which 15% are independent and noninformative right-censored, and 2 crossed frailties (call them donor and center). The donor frailty involves 100 clusters of size 2, and the center frailty involves 40 clusters of size 5. Given the frailties, failure time observations are generated with conditional hazard rate for $i = 1, \ldots, 200$ by

$$\lambda_i(t; \mathbf{x}_i, \mathbf{z}_i, \boldsymbol{\omega}_1, \boldsymbol{\omega}_2) = .1e^{\mathbf{x}_i'\boldsymbol{\beta}+\mathbf{z}_{i1}'\boldsymbol{\omega}_1+\mathbf{z}_{i2}'\boldsymbol{\omega}_2}, \tag{2}$$

where $\boldsymbol{\omega}_1$ and $\boldsymbol{\omega}_2$ are 100×1 and 40×1 vectors of iid normal variables with mean 0 and variance 1 and .5, and $\mathbf{z}_{i1}'$ and $\mathbf{z}_{i2}'$ are the ith row of the design matrix $\mathbf{Z}_1$ and $\mathbf{Z}_2$, which are randomly generated under the constraints of cluster size. The fixed covariates are $\mathbf{x}_i = (x_{i1}, x_{i2})'$, where $x_{i1} \sim \mathrm{N}(0, 1)$ is a transplant recipient–level covariate and x_{i2} is a donor-level covariate, with $x_{i2} = 1$ for half of the donors and 0 for the other half, and $\boldsymbol{\beta} = (\beta_1, \beta_2)' = (1, -.5)'$. To use the proposed nonparametric PQL approach for inference, we convert the failure time data into a contingency table form. For each time period between two consecutive time points where either an event or censoring happens, one interval is generated that includes the later time point, and a constant hazard rate is assumed on that interval. For the time periods with censoring occurring only at the later point, the hazard rate is assumed to be 0, which speeds up the procedure. Parametric models are also applied to the datasets. Estimates of fixed and random effects from mixed Poisson regression models with constant baseline hazards are calculated to evaluate the efficiency gain when the baseline hazard function is correctly specified. The simulation is performed 500 times.

Table 3 presents the simulation results. Both the recipient-level covariate β_1 and its SE are slightly underestimated, whereas the estimates for the donor-level covariate β_2 and its SE are nearly unbiased. There is some underestimation of the

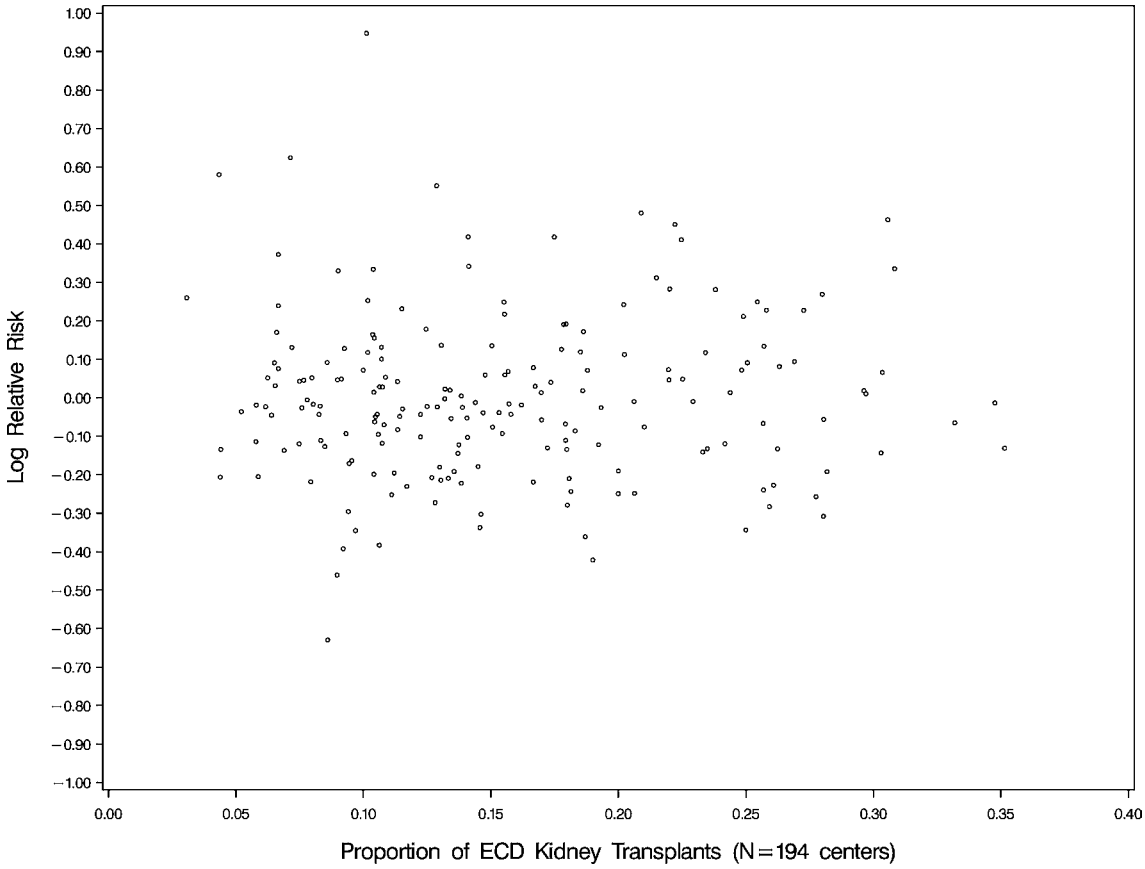

Figure 2. Estimated Adjusted Log Relative Risk versus Proportion of ECD Kidney Transplants for 194 Kidney Transplant Centers. The plot shows no association between these two variables. In addition, among the adjusted center log relative risk estimates, four of them are $>.5$, while one is $<-.5$. Further investigation for the practice patterns at transplant centers may help identify good patterns that would benefit transplant recipients.

two cluster variance estimates, whereas the SE for $\hat{\theta}_1$ is overestimated. The covering rates for β_1 and θ_2 are below the nominal level of 95%. Both the parameter estimates and CI covering rates are improved when we specify constant baseline hazard functions for the models. The estimates for β_1 and θ_2 are closer to the true values, whereas the CI covering rates for β_1 and θ_2 are improved by 1.6% and 1.4%. The estimated relative efficiencies of the parameters from the unspecified versus constant baseline hazard function model are about .7 for β_1, 1 for β_2, .5 for θ_1, and .8 for θ_2.

Table 3. Simulation Results From Crossed Balanced Lognormal Frailty Models

Parameters	*True value*	*Mean*	*Estimated SE*	*Empirical SE*	*Covering rate*
NP PQL					
β_1	1.0	.976	.123	.143	91.0%
β_2	−.5	−.481	.269	.270	94.2%
θ_1	1.0	.970	.416	.358	94.2%
θ_2[a]	.5	.476	.248	.236	91.6%
Exponential PQL					
β_1	1.0	.996	.113	.119	92.6%
β_2	−.5	−.492	.273	.269	95.0%
θ_1	1.0	1.030	.288	.254	97.2%
θ_2[b]	.5	.495	.216	.213	93.0%

NOTE: The donor frailty effect involves 100 clusters each with 2 subjects, whereas the center frailty involves 40 clusters each with 5 subjects.

[a] Three iterations report zero variance.

[b] Two iterations report zero variance.

6. DISCUSSION

Our analysis shows that the four simple ECD factors explain a substantial fraction (nearly 30%) of the variation among transplant recipient survival outcomes due to donor kidneys, which gives a level of assurance that the ECD waiting list does effectively distinguish many high-risk donor organs to candidates. The 30% of variation is further apportioned among ECD factors. Donor age appears to be the most important predictor, which explains about 26% of the variation, whereas the other three factors explain an additional 4% of the variation. There may also remain further unmeasured donor factors associated with elevated risk of graft failure. Therefore, identification of other donor factors may merit further investigation. Moreover, the analysis has also allowed us to identify variations in outcomes by center while adjusting for recipient- and donor-related factors. We find significant differences in transplant recipient survival outcomes among centers, which raises awareness of the need to study center-level factors for the goal of improving outcomes for transplant recipients.

Our method extends the practical application of random-effects models to a very general framework for survival analysis, which can include crossed random effects into RR models. PQL has been chosen as the computational procedure because

of its simplicity. Although approximations and ad hoc adjustments are used in PQL, the mixed Poisson procedure still yields good estimates for datasets with crossed clusters and for a mean number of 1.7 and 4.3 events per cluster in our simulation studies.

One current limitation in our study is that instead of using the whole dataset, we select a random sample from the dataset to reduce the computation burden. We expect that this limitation will no longer exist with the advance of information technology in the future. We have also carried out extensive sensitivity analyses to examine how the results differ with varying sampling schemes. The analysis results based on a random sample are quite similar to the results presented in this article, except that the center frailty estimate (SE) is smaller, .032 (.018) (random samples) versus .103 (.023) (oversample large facilities as in this article). The random donor sample includes many small centers, which treat few patients. There is more variation among large centers, where most patients get transplants. We report the results for large centers to show that outcomes at large centers do vary and to also show that they can be ranked with regard to outcomes. Results for small centers are less interesting because small centers vary less and they are harder to rank because of their small sample size. One referee suggests exploring whether center factor, which includes both random and fixed factors, is an effect modifier on the impact of ECD factors on survival. However, to the best of our knowledge, we are not aware of an existing methodology that can examine this issue. This interesting topic will be addressed in our future research.

The current inference methods ignore the variability of θ in the estimation, which may affect the precision of the fixed-effects parameter estimates. A possible solution is to apply bootstrap methods to construct the empirical sampling distribution, which makes the estimating procedure more time-consuming. Although the procedure yields an almost unbiased SE estimate for the frailty variance, simulations (not shown here) suggest that the CI based on the estimate may not be valid until the number of clusters becomes large in the dataset. When the number of clusters is not sufficiently large, one possible alternative for obtaining inference on frailty variance is to adopt resampling methods.

Lognormal frailty is the only frailty that can be easily fitted by the mixed Poisson procedure using SAS software. Although it is a reasonable approximation for many random effects, we may wish to use other frailty distributions under certain conditions. For example, gamma frailty models might be more appropriate when some clusters have exceptionally low risk. Nonparametric frailty models would provide another extension of the method.

[Received March 2003. Revised November 2004.]

REFERENCES

Andersen, P. K., Borgan, O., Gill, R. D., and Keiding, N. (1993), *Statistical Models Based on Counting Processes*, New York: Springer-Verlag.

Breslow, N. E. (1972), "Discussion of Professor Cox's Paper," *Journal of the Royal Statistical Society*, Ser. B, 34, 216–217.

Breslow, N. E., and Clayton, D. G. (1993), "Approximate Inference in Generalized Linear Mixed Models," *Journal of the American Statistical Association*, 88, 9–25.

Ecochard, R., and Clayton, D. G. (1998), "Multi-Level Modelling of Conception in Artificial Insemination by Donor," *Statistics in Medicine*, 17, 1137–1156.

Feng, S. B., and Wolfe, R. A. (2005), "Estimating Organ Procurement Organization and Transplant Center Effects on Patient Waiting Time to Deceased Donor Kidney Transplant Using Lognormal Frailty Models," unpublished manuscript.

Henderson, R., and Oman, P. (1999), "Effect of Frailty on Marginal Regression Estimates in Survival Analysis," *Journal of the Royal Statistical Society*, Ser. B, 61, 367–379.

Lai, T. L., and Shih, M. C. (2003), "A Hybrid Estimator in Nonlinear and Generalised Linear Mixed-Effects Models," *Biometrika*, 90, 859–879.

Laird, N., and Olivier, D. (1981), "Covariance Analysis of Censored Survival Data Using Log-Linear Analysis Techniques," *Journal of the American Statistical Association*, 76, 231–240.

Lawless, J. F., and Zhan, M. (1998), "Analysis of Interval-Grouped Recurrent-Event Data Using Piecewise Constant Rate Functions," *Canadian Journal of Statistics*, 26, 549–565.

Ma, R. J., Krewski, D., and Burnett, R. T. (2003), "Random-Effects Cox Models: A Poisson Modelling Approach," *Biometrika*, 90, 157–169.

Merion, R. M. (2003), "Preface: SRTR Report on the State of Transplantation," *American Journal of Transplantation*, 3 (Suppl 4), 5–6.

Ojo, A. O., Hanson, J. A., Meier-Kriesche, H., Okechukwu, C. N., Wolfe, R. A., Leichtman, A. B., Agodoa, L. Y., Kaplan, B., and Port, F. K. (2001), "Survival in Recipients of Marginal Cadaveric Donor Kidneys Compared With Other Recipients and Wait-Listed Transplant Candidates," *Journal of the American Society of Nephrology*, 12, 589–597.

OPTN/SRTR (2003), "2002 Annual Report of the U.S. Organ Procurement and Transplantation Network and the Scientific Registry of Transplant Recipients: Transplant Data 1992–2001," Department of Health and Human Services, Health Resources and Services Administration, Office of Special Programs, Division of Transplantation.

Parner, E. (1998), "Asymptotic Theory for the Correlated Gamma-Frailty Model," *The Annals of Statistics*, 26, 183–214.

Port, F. K., Bragg-Gresham, J. L., Metzger, R. A., Dykstra, D. M., Gillespie, B. W., Young, E. W., Delmonico, F. L., Wynn, J. J., Merion, R. M., Wolfe, R. A., and Held, P. J. (2002), "Donor Characteristics Associated With Reduced Graft Survival: An Approach to Expanding the Pool of Kidney Donors," *Transplantation*, 74, 1281–1286.

Ripatti, S., and Palmgren, J. (2000), "Estimation of Multivariate Frailty Models Using Penalized Partial Likelihood," *Biometrics*, 56, 1016–1022.

Sastry, N. (1997), "A Nested Frailty Model for Survival Data, With an Application to the Study of Child Survival in Northeast Brazil," *Journal of the American Statistical Association*, 92, 426–435.

Shen, W., and Louis, T. A. (1998), "Triple-Goal Estimates in Two-Stage Hierarchical Models," *Journal of the Royal Statistical Society*, Ser. B, 60, 455–471.

Vaida, F., and Xu, R. (2000), "Proportional Hazards Model With Random Effects," *Statistics in Medicine*, 19, 3309–3324.

Wolfe, R. A., Ashby, V. B., Milford, E. L., Ojo, A. O., Ettenger, R. E., Agodoa, L. Y. C., Held, P. J., and Port, F. K. (1999), "Patient Survival for Waitlisted Dialysis versus Cadaveric Renal Transplant Patients in the United States," *New England Journal of Medicine*, 341, 1725–1730.