

# A Sequential Stratification Method for Estimating the Effect of a Time-Dependent Experimental Treatment in Observational Studies

Douglas E. Schaebel,<sup>1,\*</sup> Robert A. Wolfe,<sup>1</sup> and Friedrich K. Port<sup>2</sup>

<sup>1</sup>Department of Biostatistics, University of Michigan, Ann Arbor, Michigan 48109-2029, U.S.A.

<sup>2</sup>University Renal Research and Education Association, Ann Arbor, Michigan 48103, U.S.A.

\**email*: deschau@umich.edu

**SUMMARY.** Survival analysis is often used to compare experimental and conventional treatments. In observational studies, the therapy may change during follow-up and such crossovers can be summarized by time-dependent covariates. Given the ever-increasing donor organ shortage, higher-risk kidneys from expanded criterion donors (ECD) are being transplanted. Transplant candidates can choose whether to accept an ECD organ (experimental therapy), or to remain on dialysis and wait for a possible non-ECD transplant later (conventional therapy). A three-group time-dependent analysis of such data involves estimating parameters corresponding to two time-dependent indicator covariates representing ECD transplant and non-ECD transplant, each compared to remaining on dialysis on the waitlist. However, the ECD hazard ratio estimated by this time-dependent analysis fails to account for the fact that patients who forego an ECD transplant are not destined to remain on dialysis forever, but could subsequently receive a non-ECD transplant. We propose a novel method of estimating the survival benefit of ECD transplantation relative to conventional therapy (waitlist with possible subsequent non-ECD transplant). Compared to the time-dependent analysis, the proposed method more accurately characterizes the data structure and yields a more direct estimate of the relative outcome with an ECD transplant.

**KEY WORDS:** Cohort study; Failure time data; Matching; Proportional hazards model; Risk set; Survival analysis.

## 1. Introduction

End-stage renal disease (ESRD; also known as chronic kidney failure) is increasing in many countries worldwide and has become a major public health issue due to its associated mortality and health care cost. Patients with renal failure must receive either dialysis or a kidney transplant in order to remain alive. Medically suitable patients are placed on a waiting list in order to receive a cadaveric renal transplant. Generally, mortality rates are significantly lower post-transplant compared to those on the waitlist (WL) (Wolfe et al., 1999; Rabbat et al., 2000). Due to the ever-increasing shortfall in availability of donor organs, patients are electing to receive a cadaveric organ from an expanded criterion donor (ECD) with increasing frequency. Port et al. (2002) formally quantified the term “ECD” to apply to deceased donors in whom transplanted kidneys are associated with more than a 70% increase in transplant failure, compared to a non-ECD kidney. Comparisons between ECD and non-ECD transplant outcomes are useful, but a more relevant comparison is between accepting an ECD transplant and “conventional therapy,” which is to remain on the WL with the potential to subsequently receive a non-ECD transplant.

The Cox (1972) proportional hazards model provides an extremely flexible way to make covariate-adjusted mortal-

ity comparisons among therapies when a patient’s mode of therapy may change over time. The model is widely used for survival analysis and is readily accepted by clinicians. As applied to the current setting, where there are three mutually exclusive “states” (WL dialysis, ECD transplant, and non-ECD transplant), time-dependent indicator covariates could be set up for the two transplant states and a proportional hazards analysis could be carried out in a straightforward manner. The estimated hazard ratio (HR) for ECD would reflect the mortality contrast between an ECD transplant and dialysis on the WL. In interpreting the ECD/WL hazard ratio, patients currently on the WL could determine the reduction in mortality hazard which would apply at that point in time if they received an ECD transplant compared to remaining on the waitlist.

Although informative, this “three-group time-dependent” (T3) analysis does not address the choice faced by the candidate. The question for the patient is not “Would I be better off with an ECD transplant than being on dialysis?” but, rather, “Would I be better off accepting an ECD organ, given that, if I do not accept it, I could subsequently be offered a non-ECD organ?” The “experimental” group is ECD transplant. However, the “conventional therapy” group is not “waitlist,” but, rather, “refusing the ECD organ and possibly receiving

a non-ECD kidney in the future.” The ECD/WL hazard ratio, as estimated by the T3 method, applies while the patient remains on the WL; should the patient subsequently receive a non-ECD transplant, a different HR would apply. The ECD/WL parameter, as estimated by the standard time-dependent method, does not pertain to the contrast between the experimental and conventional therapy.

On the surface, the problem appears to be reduced to a two-group time-dependent problem, with candidates crossing over from the standard therapy group to the ECD group. Analysis of these data with a single time-dependent covariate would appear to address the comparison of interest. However, two complications are described below that suggest that such a “two-group time-dependent” (T2) analysis will yield biased results.

One complication involves removal from the waitlist. Patients who become too ill to undergo transplantation are often removed from the WL. How should the follow-up time for such candidates be included in the analysis? Since unlisted dialysis patients are not transplant candidates, it is inappropriate to include them in the comparison group for transplant recipients. It would seem appropriate to censor patients at the time of removal from the WL, since they are no longer eligible to receive either an ECD or a non-ECD transplant. However, removal can be considered an intermediate event on the pathway from the currently administered dialysis therapy to the event of interest, mortality. With an ideal surveillance of patient condition, all patients might, at some point prior to death, be deemed unsuitable transplant candidates and removed from the WL. If such removals were censored, then WL mortality would be estimated to be 0, due to the informative censoring. It appears that both options, censoring at removal or including follow-up after removal, lack face validity.

Another complication arises from the inclusion of non-ECD transplant recipients in the standard therapy group. A naive two-group analysis might include the follow-up for all non-ECD transplant recipients in the standard therapy group, including that of pre-existing non-ECD transplants. However, a candidate who is offered an ECD organ has the option of waiting for a future non-ECD transplant, but does not have the option of going back in time to receive a non-ECD organ prior to the offer of the ECD organ. Pre-existing non-ECD transplants should not be included in the standard therapy group.

We propose a “sequential stratification” method which, unlike the T3 analysis, yields treatment parameters that have an interpretation that addresses the choice faced by the candidate and which, unlike the T2 analysis, avoids the two complications described in the previous two paragraphs. In order to address the complications described above, the method employs matching to create a series of strata for each ECD transplant, which includes only those patients who are eligible for transplant and which also includes follow-up times after removal from the WL. The proposed method uses removal and prior transplant information from the data structure in order to avoid the two biases described above. Under the proposed method, prior removals from the WL and non-ECD transplants are appropriately excluded from the stratum while subsequent follow-up after removal from the WL is included.

Using the proposed sequential stratification method, a significant survival benefit of ECD kidney transplantation is demonstrated, relative to remaining on the WL and possibly subsequently receiving a non-ECD organ.

The remainder of this article is organized as follows. In Section 2, we set up the requisite notation and formalize the problem. The standard (T3) time-dependent proportional hazards analysis and the T2 method are described in Section 3. In Section 4, we re-examine the organ failure/transplant data structure, then propose our alternative method of analysis in Section 5. The survival benefit of ECD transplant is assessed using the proposed method in Section 6, with concluding remarks provided in Section 7.

## 2. Notation on Data Sources

We first present notation for latent variables, which might not be observed for all subjects. Let  $D_i$  represent the time of death for patient  $i$ , for  $i = 1, \dots, n$ . Censoring time due to end of study or loss to follow-up is denoted by  $C_i$ . Since living-donor organs are not of interest in this analysis, patients are censored if and when they receive a transplant from a living donor. The time of cadaveric transplant is denoted by  $T_i = T_i^E \wedge T_i^{\bar{E}}$ , where  $T_i^E$  is the time of ECD transplant,  $T_i^{\bar{E}}$  is the time of non-ECD transplant, and  $a \wedge b = \min\{a, b\}$ . The time at which a patient is removed from the WL is denoted by  $R_i$ .

Notation for the observed data is given here. The death indicator is defined as  $\Delta_i = I(D_i < C_i)$ , with corresponding counting process  $N_i(t) = I(X_i \leq t)\Delta_i$ , where  $X_i = C_i \wedge D_i$  is the observation time for subject  $i$ . The covariate vector is given by  $\mathbf{Z}_i(t)$ . The stratum, described later, for patient  $i$  is denoted by  $s_i$ . We make the familiar independent censoring assumption. That is, we assume that  $C_i$  is conditionally independent of  $D_i$ , given  $\mathbf{Z}_i(t)$  and  $s_i$ ; specifically,

$$\begin{aligned} \lim_{\delta \downarrow 0} P(t \leq D_i < t + \delta \mid D_i > t, C_i > t, \mathbf{Z}_i(t), s_i) \\ = \lim_{\delta \downarrow 0} P(t \leq D_i < t + \delta \mid D_i > t, \mathbf{Z}_i(t), s_i). \end{aligned}$$

As such, since living-donor transplants are absorbed into  $C_i$ , we assume that the receipt of a living-donor transplant does not depend on the mortality hazard the patient would have faced had they remained on the WL.

Data were collected by the Organ Procurement and Transplantation Network and obtained from the Scientific Registry for Transplant Recipients (SRTR). The study population included patients ( $n = 97,619$ ) initially waitlisted in the United States, aged  $\geq 18$  years, between January 1, 1998 and December 31, 2002. The observation period concluded at the end of 2002. The time origin ( $t = 0$ ) for a given patient was the date that patient was initially placed on the kidney transplant waiting list. Stratification was by age (single year) and region, while gender, race, and underlying cause of renal disease were represented in the covariate vector. Age is well known to be a strong predictor of mortality among transplant patients. In addition, separate waitlists are maintained at the 60 organ procurement organizations (OPOs) in the United States, which are defined geographically (by state, approximately). Socioeconomic status, general baseline health, and other factors which could affect mortality are known to differ

geographically. Hence, careful adjustment of age and OPO is required in any meaningful comparison between therapies.

**3. Standard Time-Dependent Methods**

In this section, we describe two existing methods of analysis, beginning with the T3 analysis and followed by the T2 analysis. The Cox (1972) proportional hazards model is assumed.

*3.1 Three-Group Time-Dependent Analysis*

We now describe the T3 analysis of the evaluation of the mortality risk of expanded criterion donor transplant relative to dialysis on the WL. The proportional hazards model is given by

$$\lambda_i(t | s_i = j; \beta_0) = \lambda_{0j}(t) \exp\{\beta_0^T \mathbf{Z}_i(t)\},$$

where  $s_i$  is the stratum for patient  $i$ ,  $\lambda_{0j}(t)$  is an unspecified stratum-specific baseline hazard, and  $\mathbf{Z}_i(t)$  is a vector of time-dependent covariates that include  $I(T_i^E \leq t)$ ,  $I(T_i^E \leq t)$ , and terms for the adjustment covariates, which included gender, race, and disease leading to renal failure.

At time  $t$ , each patient who is alive and uncensored is classified into one of three mutually exclusive time-dependent “treatments”—WL, ECD transplant, or non-ECD transplant—by  $I(T_i^E \leq t)$  and  $I(T_i^E \leq t)$ . The coefficient of  $I(T_i^E \leq t)$ , the indicator for ECD transplant, is denoted by  $\beta_0^E$ , which is a component of  $\beta_0$ . The regression parameter,  $\beta_0$ , is estimated through partial likelihood (Cox, 1975) as the solution to the estimating equation  $\mathbf{U}_\beta(\beta) = \mathbf{0}$ , where  $\mathbf{0}$  is the vector of 0’s with dimension equal to that of  $\beta$ ,

$$\mathbf{U}_\beta(\beta) = \sum_{i=1}^n \sum_{j=1}^J \int_0^\tau \{\mathbf{Z}_i(s) - \bar{\mathbf{Z}}_j(s; \beta)\} Y_{ij}(s) dN_i(s),$$

with  $Y_{ij}(s) = I(X_i > s, s_i = j)$  and

$$\bar{\mathbf{Z}}_j(s; \beta) = \frac{\sum_{i=1}^n Y_{ij}(s) \mathbf{Z}_i(s) \exp\{\beta^T \mathbf{Z}_i(t)\}}{\sum_{i=1}^n Y_{ij}(s) \exp\{\beta^T \mathbf{Z}_i(t)\}},$$

with  $\tau$  typically set to  $\max\{X_1, \dots, X_n\}$  such that all observed deaths contribute to the analysis. Notice that  $R_i$  is ignored in this analysis, essentially.

Results based on this approach are presented in Table 1. It is shown that ECD transplant recipients have a covariate-adjusted mortality HR of 0.83, relative to waitlisted patients on dialysis ( $p < 0.0001$ ), with 95% confidence interval (0.76, 0.91).

This three-group analysis gives an informative description of the contrast between ECD transplant and WL mortality.

**Table 1**  
Hazard ratios estimated by three-group time-dependent (T3) method

Group	$\exp\{\hat{\beta}\}$	(95% CI)	$p$ -value
WL	1	–	–
ECD transplant	0.83	(0.76, 0.91)	<0.0001
Non-ECD transplant	0.55	(0.52, 0.58)	<0.0001

At any time  $t$ , a patient with an ECD transplant has a mortality hazard which is 83% that of a comparable waitlisted patient at time  $t$ , reflecting, for a patient on the WL, the reduction in mortality they could expect to reap upon ECD transplantation.

Renal failure patients must receive either dialysis or a kidney transplant; hence dialysis and transplantation are mutually exclusive and exhaustive states. If we were interested in comparing WL and transplantation (be it ECD or non-ECD), then defining the covariate  $I(T_i \leq t)$  and using this standard time-dependent analysis address the question of interest, as applied by Wolfe et al. (1999). However, the question is different when interest lies in one of the two transplant types (i.e., ECD or non-ECD). Patients can forego an ECD, with the objective of subsequently receiving a non-ECD organ.

The T3 analysis compares all three therapies and includes a comparison of mortality after transplantation with an ECD organ to mortality while remaining on the WL. However, the comparison of interest is between the “experimental therapy” (i.e., ECD transplant) and “conventional therapy” (foregoing an ECD transplant and remaining on the WL, with the opportunity to subsequently receive a non-ECD organ). There is no treatment group in the T3 analysis that corresponds to conventional therapy; thus, the three-group time-dependent analysis does not estimate the comparison of interest.

*3.2 Two-Group Time-Dependent Analysis*

In the T3 analysis described in the preceding subsection, patients on the WL have their WL mortality censored upon receipt of a transplant. This censoring of WL mortality can be considered independent censoring, since patients are ordered on the WL primarily based on waiting time, rather than upon patient condition. Transplanted subjects contribute two follow-up subintervals to the analysis; the first subinterval spans the time from waitlisting to transplant (ECD or non-ECD), while the second spans the time from transplant to the end of follow-up. Since conventional therapy includes both the waitlist interval and the interval after a non-ECD transplant, it seems natural to not censor a WL patient’s follow-up upon non-ECD transplantation. We refer to this method of analysis as the two-group time-dependent analysis (T2). Again using a proportional hazards model, estimation proceeds as in the T3 analysis, but with  $I(T_i^E \leq t)$  removed from  $\mathbf{Z}_i(t)$ . The parameter vector now is denoted by  $\gamma_0$ , with  $\gamma_0^E$  denoting the component of  $\gamma_0$  corresponding to  $I(T_i^E \leq t)$ .

Results from the modified time-dependent analysis are presented in Table 2.

The ECD and “WL + non-ECD” groups have virtually equal mortality hazard, with  $\exp\{\hat{\gamma}_E\} = 1.01$  (0.93, 1.10) and  $p = 0.80$ . Taking these results at face value, there appears to be no benefit of opting for an ECD transplant. As indicated

**Table 2**  
Hazard ratios estimated by two-group time-dependent (T2) method

Group	$\exp\{\hat{\gamma}\}$	(95% CI)	$p$ -value
WL + non-ECD	1	–	–
ECD transplant	1.01	(0.03, 1.10)	0.80

by the T3 analysis, mortality is highest for the WL group and lowest for the non-ECD group; so the result from the T2 analysis, which combines these together into the conventional therapy group, is not surprising.

The conventional therapy group in the T2 analysis includes all non-ECD transplant during follow-up. At any moment of time, the conventional therapy group in the T2 analysis already includes a substantial fraction of existing non-ECD transplant recipients. However, the offer of an ECD transplant at time  $t$  is conditional upon being in the WL group ( $I\{T_i \leq t\} = 0$ ), so the conventional therapy group in the T2 analysis does not represent an option available to the ECD candidate at the time an ECD kidney is offered.

In addition, it should be noted that neither the T3 nor the T2 analyses explicitly deal with removal from the WL. In the next two paragraphs, we argue that removal from the waitlist should be used to determine entry into the conventional therapy group, but should not be used to censor follow-up after entry into the WL group.

First, we discuss entry into the conventional therapy group. Not all patients on dialysis are waitlisted, and an analysis by Wolfe et al. (1999) revealed higher mortality risk for non-WL, relative to WL, dialysis patients. Mortality comparisons should be based on groups who satisfy similar eligibility criteria. In determining the appropriate comparison group for cadaveric kidney transplant patients, it is important to consider only WL patients, since dialysis patients who are not on the WL are not eligible to receive a cadaveric organ. Similarly, follow-up for patients who have been previously removed from the WL should not be included in the comparison group for ECD transplants.

Second, we discuss censoring after removal from the WL group. If a transplant fails, patients are not censored from the transplant group. Rather, an intent-to-treat approach is adopted, wherein potential graft failure is considered to be an assumed risk of transplantation. Similarly, WL patients should not be censored upon removal from the WL. In the vast majority of cases, patients are removed from the WL because, in the opinion of their physician, they have become too ill or weak to withstand surgery. Thus, if the organ allocation system operated perfectly, observed WL mortality would be 0, since patients would be too sick to undergo transplantation in the moments before they die.

In summary, the T2 analysis better captures the possibility of future non-ECD transplantation in the specification of conventional therapy, relative to the T3 analysis. However, the T2 approach fails in two ways in its attempt to define conventional therapy. First, it includes pre-existing non-ECD transplants. Second, the method of handling removals from the WL lacks face validity. Since parameters based on either the standard time-dependent analysis (T3) or the simple modification thereof (T2) do not have the desired interpretations, a new method of analysis is required.

#### 4. The Ideal Study and the ECD Transplant Experiment

Ideally, to measure the survival benefit associated with expanded criterion donor transplantation, a randomized controlled trial would be conducted. Each time an ECD kidney was procured, one patient from the WL would be randomly

selected to receive the transplant, with the rest remaining on the WL. Such a trial will never occur, due to ethical and logistical considerations. Notwithstanding, it is useful to consider the data structure generated by this ideal experiment.

In reality, patients are not randomized to receive or not receive ECD organs. However, each ECD transplant can be considered to be one replicate of an experiment; one patient receives the transplant, while the rest remain on the WL, with the option of subsequent non-ECD transplant (i.e., conventional therapy). An intent-to-treat approach is appropriate for the analysis of this experiment. Thus, patients remain in the conventional therapy group if they receive a non-ECD transplant, since potential non-ECD transplantation is one component of conventional therapy. Also, patients who had been removed from the WL prior to the time of the experiment-generating transplant are not eligible for (are excluded from) the control group, since they are not eligible for an ECD transplant. However, patients would not be censored from the control group at removal from the WL after the experiment begins. Following the intent-to-treat concept starting at the time of entry in each experiment, baseline covariates for each experiment are defined by the current value of time-dependent covariates, at the time each experiment starts. As in standard time-dependent crossover analyses, the conventional therapy patients are censored from the conventional therapy group and enter another ECD experiment at the time of ECD transplantation.

These considerations suggest an alternative method of analysis, which we now propose.

#### 5. Proposed Method: Sequential Stratification

In this section, we formalize the analysis of the ECD transplant “experiments” suggested in Section 4. An experiment is said to begin each time an ECD transplant occurs. It is convenient to order the ECD transplant times, first to last, by  $t_{(1)} < t_{(2)} < \dots < t_{(n_E)}$ , where  $n_E = \sum_{i=1}^n I(T_i^E < X_i)$  is the total number of observed ECD transplants. Correspondingly, we order the strata from which the ECD transplants arose as  $s_{(1)}, \dots, s_{(n_E)}$ . At the time of the  $\ell$ th ECD transplant, the “experimental” group consists of the patient receiving the ECD transplant. The control group consists of all patients who are comparable to the experimental patient and were eligible to receive the ECD transplant at time  $t = t_{(\ell)}$  but did not. “Comparability” requires that the control patient be from the same stratum as the ECD patient; that is,  $s_i = s_{(\ell)}$ , for subject  $i$ . “Eligibility,” for patient  $i$ , requires that patient  $i$  has not received a transplant, has not been removed, and is still under observation at time  $t = t_{(\ell)}$ . We formalize these criteria by defining the experiment entry indicators for the  $\ell$ th “experiment”:

$$e_{i(\ell)} = I\{T_i^E \geq t_{(\ell)}, T_i^{\bar{E}} \wedge X_i \wedge R_i > t_{(\ell)}, s_i = s_{(\ell)}\},$$

which equals 1 for patients eligible for the  $\ell$ th experiment and 0 otherwise. Although patients are required to not have been transplanted or removed from the WL at time  $t_{(\ell)}$  to enter the  $\ell$ th experiment, patients are not censored from the  $\ell$ th experiment if they are removed or if they receive a non-ECD transplant, since both of these events are part of conventional therapy. If they receive an ECD organ during the  $\ell$ th experiment, they are censored from that experiment and their

ECD transplant would initiate a separate experiment. Correspondingly, the risk set indicator for subject  $i$  can be written as  $Y_{i(\ell)}(s) = Y_i(s)I\{T_i^E > s\}^{I\{T_i \neq t_{(\ell)}\}}$ , with respect to the  $\ell$ th experiment, where  $Y_i(s) = I\{X_i \geq s\}$ .

The hazard function for subject  $i$  in this  $\ell$ th experiment is given by

$$\lambda_{i(\ell)}(t; \boldsymbol{\theta}_0) = \lambda_{0(\ell)}(t) \exp\{\boldsymbol{\theta}_0^T \mathbf{Z}_{i(\ell)}\},$$

where  $\boldsymbol{\theta}_0$  is a vector of unknown parameters corresponding to the covariate,  $\mathbf{Z}_{i(\ell)} = \mathbf{Z}_i(t_{(\ell)})$ , and

$$\lambda_{0(\ell)}(t) = \lim_{\delta \downarrow 0} \frac{1}{\delta} P\{t \leq D_i < t + \delta \mid D_i > t, T_i^E \geq t_{(\ell)}, T_i^E \wedge X_i \wedge R_i > t_{(\ell)}, \mathbf{Z}_{i(\ell)} = \mathbf{0}\}$$

is the experiment and stratum-specific baseline hazard, which corresponds to the hazard function for the conditional random variate,  $(D_i \mid T_i \wedge R_i > t_{(\ell)})$ . We denote  $\theta_E$  as the element of interest from  $\boldsymbol{\theta}_0$ ; that is, the element which corresponds to  $I\{T_i^E = t_{(\ell)}\}$ .

We now derive the estimating equation for  $\boldsymbol{\theta}_0$ . To begin, we define the error term,  $M_{i(\ell)}(t; \boldsymbol{\theta}) = e_{i(\ell)} \int_{t_{(\ell)}}^t Y_{i(\ell)}(s) \{dN_i(s) - \lambda_{i(\ell)}(s; \boldsymbol{\theta}) ds\}$ , which is a martingale with respect to the filtration  $\mathcal{F}_{i(\ell)}(t) = \sigma\{e_{i(\ell)}, N_i(s-), Y_{i(\ell)}(s), \mathbf{Z}_{i(\ell)}; s \in [t_{(\ell)}, t]\}$ . The fact that  $E[M_{i(\ell)}(t; \boldsymbol{\theta}_0) \mid \mathcal{F}_{i(\ell)}(t)] = 0$  leads to the following unbiased estimating equation corresponding to the  $\ell$ th experiment:

$$\sum_{i=1}^n \int_{t_{(\ell)}}^{\tau} e_{i(\ell)} \{\mathbf{Z}_{i(\ell)} - \bar{\mathbf{Z}}_{(\ell)}(s; \boldsymbol{\theta})\} dM_{i(\ell)}(s; \boldsymbol{\theta}), \tag{1}$$

where

$$\bar{\mathbf{Z}}_{(\ell)}(s; \boldsymbol{\theta}) = \frac{\sum_{i=1}^n e_{i(\ell)} Y_{i(\ell)}(s) \mathbf{Z}_{i(\ell)} \exp\{\boldsymbol{\theta}^T \mathbf{Z}_{i(\ell)}\}}{\sum_{i=1}^n e_{i(\ell)} Y_{i(\ell)}(s) \exp\{\boldsymbol{\theta}^T \mathbf{Z}_{i(\ell)}\}}.$$

The estimating function in (1) can be derived using standard partial likelihood (Cox, 1975) theory and can be easily shown to have mean zero at  $\boldsymbol{\theta} = \boldsymbol{\theta}_0$  through standard martingale results. It can also be shown that (1) is equivalent to

$$\sum_{i=1}^n \int_{t_{(\ell)}}^{\tau} e_{i(\ell)} \{\mathbf{Z}_{i(\ell)} - \bar{\mathbf{Z}}_{(\ell)}(s; \boldsymbol{\theta})\} Y_{i(\ell)}(s) dN_i(s), \tag{2}$$

simply through the definitions of  $\bar{\mathbf{Z}}_{(\ell)}(s; \boldsymbol{\theta})$  and  $dM_{i(\ell)}(s; \boldsymbol{\theta})$ . Aggregating across all strata, we propose to estimate  $\boldsymbol{\theta}_0$  by  $\hat{\boldsymbol{\theta}}$ , the solution to  $\mathbf{U}_{\boldsymbol{\theta}}(\boldsymbol{\theta}) = \mathbf{0}$ , where

$$\mathbf{U}_{\boldsymbol{\theta}}(\boldsymbol{\theta}) = \sum_{\ell=1}^{n_E} \sum_{i=1}^n \int_{t_{(\ell)}}^{\tau} e_{i(\ell)} \{\mathbf{Z}_{i(\ell)} - \bar{\mathbf{Z}}_{(\ell)}(s; \boldsymbol{\theta})\} Y_{i(\ell)}(s) dN_i(s). \tag{3}$$

Essentially, (3) corresponds to a stratified Cox model, with strata defined by ECD transplant times,  $\{t_{(\ell)}; \ell = 1, \dots, n_E\}$ , and covariates used to define  $s_i$ . Thus,  $\mathbf{U}_{\boldsymbol{\theta}}(\boldsymbol{\theta}) = \mathbf{0}$  can be solved using standard proportional hazards regression software, such as *phreg* in SAS and *coxph* in R/Splus. At each ECD transplant time, strata are set up. Consider the  $\ell$ th-ordered

ECD transplant. Using the (start, stop, event indicator) notation, the patient receiving the  $\ell$ th ECD organ (say, patient  $j$ ) contributes  $(t_{(\ell)}, X_j, \Delta_j)$ ; patients with  $s_i = s_j (i \neq j)$  with  $R_i \wedge X_i \wedge T_i > t_{(\ell)}$  are included in the stratum and contribute  $(t_{(\ell)}, X_i \wedge T_i^E, \Delta_i I\{D_i < T_i^E\})$ .

The consistency of  $\hat{\boldsymbol{\theta}}$  can be proven using straightforward extensions of techniques presented in Anderson and Gill (1982). Proof of the asymptotic normality of  $n^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$  is more complicated and beyond the scope of the current report. Speaking generally, the conditions underlying the derivations of existing robust techniques, such as those developed by Biliias, Gu, and Ying (1997) and Wei, Lin, and Weissfeld (1989) and applied by several authors (e.g., Lin et al., 2000), do not hold. One complicating feature is that  $\bar{\mathbf{Z}}_{(\ell)}(s; \boldsymbol{\theta}_0)$  need not converge to a limit,  $\bar{\mathbf{z}}_{(\ell)}(s; \boldsymbol{\theta}_0)$  for any time,  $s$ , at least in a continuous time framework. A second difficulty involves determining the contribution, asymptotically, of each individual, in the presence of an estimating equation which is essentially a double summation across all subjects.

Consequently, we propose estimating the distribution of  $\hat{\boldsymbol{\theta}}$  empirically through the bootstrap (Efron, 1982). That is, individual patients are randomly resampled  $B$  times, where  $B$  is a large number. Let  $\hat{\boldsymbol{\theta}}_b$  represent the estimator of  $\boldsymbol{\theta}_0$  based on the  $b$ th resample. The covariance matrix of  $\hat{\boldsymbol{\theta}}$  is estimated by

$$\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\theta}} = (B - 1)^{-1} \sum_{b=1}^B (\hat{\boldsymbol{\theta}}_b - \hat{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}}_b - \hat{\boldsymbol{\theta}})^T.$$

Hypothesis tests of the form  $H_0: \boldsymbol{\theta}_0 = \mathbf{0}$  versus  $H_1: \boldsymbol{\theta}_0 \neq \mathbf{0}$  could be based on the Wald test, with  $\hat{\boldsymbol{\theta}}^T \hat{\boldsymbol{\Sigma}}_{\boldsymbol{\theta}}^{-1} \hat{\boldsymbol{\theta}}$  assumed to follow a chi-square distribution with degrees of freedom equal to number of elements in  $\boldsymbol{\theta}_0$ . This Wald test assumes an asymptotic normal distribution for  $n^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ . If diagnostic plots (e.g., histograms, q/q plots) of the  $\{\hat{\boldsymbol{\theta}}_{k,b}; b = 1, \dots, B\}$  display great departure from normality, confidence intervals and hypothesis tests could be based on the bootstrapped percentiles, although this would require choosing a much larger value of  $B$  (e.g.,  $B = 10,000$ ).

**6. Application of Sequential Stratification Method to Renal Failure Data**

As stated in Section 5, for the experiment generated by the  $\ell$ th ECD transplant, the proposed method involves estimating multiplicative effects on the hazard function for  $(D_i \mid e_{i(\ell)} = 1)$ . Thus, each experiment only consists of one ECD transplant patient. We assumed that  $\lambda_{0(k)}(t) \approx \lambda_{0(\ell)}(t)$  when  $t_{(k)} \approx t_{(\ell)}$ . As such, prior to applying the sequential stratification method, we grouped strata with the same age and organ procurement organization by the week of ECD transplant. In addition to  $I\{T_i^E \leq t\}$ , components of  $\mathbf{Z}_{i(\ell)}$  included terms for gender, race, and cause of end-stage renal disease, as in the T3 and T2 analyses.

Results of the sequential stratification analysis are presented in Table 3. The hazard ratio for ECD relative to standard therapy is estimated at HR = 0.89 (0.81, 0.96). Therefore, based on the proposed method, ECD transplantation is associated with a significant ( $p = 0.02$ ) reduction in mortality, relative to remaining on the WL with the possibility of receiving a non-ECD transplant in the future.

**Table 3**  
Hazard ratios estimated by proposed sequential stratification method

Group	$\exp\{\hat{\theta}_j\}$	(95% CI)	$p$ -value
Conventional therapy	1	–	–
ECD transplant	0.89	(0.81, 0.96)	0.02

Results in Table 3 are based on 500 bootstrap replications, with the confidence intervals and  $p$ -values obtained through the normal approximation. For each bootstrap replicate, it took approximately 15 minutes to compute  $\hat{\beta}_b$ . All programming was done using SAS (v8.2).

Model diagnostics mostly centered around (i) the normality assumption, upon which inference was based and (ii) the assumption that  $\theta_E$  was constant across  $t_1, \dots, t_{n_E}$ . As indicated by the plots of the  $\hat{\theta}_{E,b}$ 's (Figure 1), the normal approximation appears to be quite reasonable.

The proposed method assumes that the ECD regression parameter is equal across several conditional mortality hazards, with the conditionality depending on the time of ECD transplant. This is a rather strong assumption that may fail frequently in practice. Fortunately, it can be evaluated empirically through interaction terms. For example, in our analysis, to examine the validity of this assumption, we fitted a model using the proposed method which allowed for separate ECD parameters, denoted by  $\theta_E^1, \theta_E^2, \theta_E^3$ , and  $\theta_E^{4+}$ , pertaining to ECD transplants occurring in years 1, 2, 3, and 4+ of follow-up, respectively. Based on this supplemen-

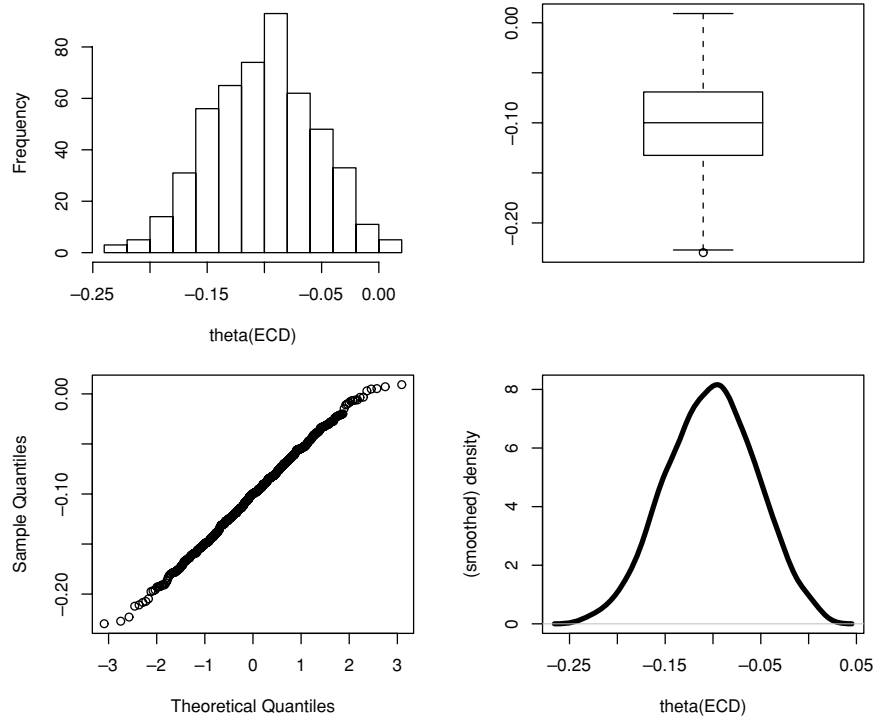
tal model, the hazard ratios were estimated at  $\exp\{\hat{\theta}_E^1\} = 0.91$  (0.81, 1.02) for year 1,  $\exp\{\hat{\theta}_E^2\} = 0.85$  (0.73, 0.99) in year 2,  $\exp\{\hat{\theta}_E^3\} = 0.92$  (0.73, 1.15) in year 3, and  $\exp\{\hat{\theta}_E^{4+}\} = 0.89$  (0.63, 1.26) in years 4–5. The random scatter of these estimated hazard ratios around the overall value,  $HR = 0.89$ , supports the assumption that the ECD hazard ratio is independent of time of transplant.

**7. Concluding Remarks**

We propose a sequential stratification method of survival analysis, useful for estimating the effect of a time-dependent treatment in the presence of other time-dependent treatments. We applied the proposed method to compare expanded criterion donor kidney transplantation versus conventional renal replacement therapy, the latter being defined as not receiving an ECD organ, with the possibility of future non-ECD transplant. The method features a natural characterization of the kidney failure/transplant data structure.

Although we discuss the proposed method mostly in the context of renal failure, the setting which motivated its development, the method has wide applicability to observational studies where time to a terminal event is of interest and the treatments being compared are time dependent.

Our analytical objective was to determine whether there is a benefit, with respect to patient survival, of accepting an ECD kidney, compared to the conventional idea of not accepting an ECD organ and remaining on the WL in the hope of obtaining a non-ECD transplant. The standard T3 three-group time-dependent analysis uses a three-level time-dependent covariate, with mutually exclusive categories: WL, ECD, and non-ECD. Using this approach, ECD is shown to be



**Figure 1.** Histogram, box/whisker, q/q plot, and smoothed density of  $\hat{\theta}_b^E$ , based on  $B = 500$  bootstrap resamples.

superior to the WL. However, this analysis does not address the research question of interest, since the “WL” group does not represent conventional therapy, as conventional therapy includes the possibility of a future non-ECD transplant.

A modification to the T3 method, T2, involves not censoring the WL follow-up of patients who receive a non-ECD kidney and, hence, reduces the number of comparison groups to two. This modified time-dependent method essentially combines the WL and non-ECD experience. When applied to the renal failure data, there appeared to be virtually no difference in mortality between ECD and “WL and non-ECD.” The ECD regression parameter estimated by this method is difficult to interpret because prior non-ECD transplants and removals from the waitlist are handled inappropriately.

The proposed method is targeted at observational data, with the objective of reconstructing the parameter estimate for an experimental treatment that would have been estimated by a randomized controlled trial. Although covariate imbalances between the experimental and conventional therapy groups can be accommodated through covariate adjustment and stratification, such adjustments may be incomplete. For example, patients who choose to receive an ECD organ may exhibit certain unmeasured characteristics which distinguish them from patients who opt for conventional therapy. Therefore, due to the observational nature of the data, unmeasured patient selection factors may bias the estimated regression parameter. Of course, clinical trials are not subject to this sort of limitation. It is possible that a propensity score approach (Rosenbaum and Rubin, 1983) may be an alternative to, or an enhancement of, the proposed method.

In the analysis of organ failure data, careful consideration of the censoring mechanisms is usually warranted. Patients who are transplanted are not censored; with transplantation coded as a time-dependent covariate, they would continue to contribute to the analysis as “transplanted.” But, transplanted patients are censored with respect to their waitlist mortality. In the context of kidney transplantation, such censoring would be largely independent since deceased-donor kidneys are allocated primarily based on waiting time. However, for chronic liver failure patients, dependent censoring would be a major concern since (under the new allocation scheme which took effect in early 2002), patients are sequenced on the WL, essentially, in decreasing order of WL mortality risk. Therefore waitlisted patients with the greatest transplant rate are also those with the greatest WL mortality risk. Extension of the proposed methods to accommodate such dependent censoring is currently underway.

Results based on the proposed sequential stratification method indicate a significant 11% reduction in mortality associated with ECD kidney transplantation. Patients deciding whether to accept an ECD organ would want to consider the experience of similar patients faced with the same decision. The sequential stratification analysis indicates that patients waitlisted between 1998 and 2002 who accepted an ECD transplant demonstrated a significant 11% decrease in mortality compared to patients who received conventional therapy. Patients and clinicians should consider these findings carefully, particularly given the ever-increasing shortage of donor organs.

Future work will involve further examination of the large-sample properties of the regression parameter estimator for the proposed method. Although arguments to demonstrate the regression parameter’s consistency are straightforward, the asymptotic distribution is difficult to derive. As such, the bootstrap was employed in the ECD analysis. Due to our current reliance on the bootstrap for inference, the size of the data set ( $n = 97,619$ ) limits the depth of the analysis. For now, we describe the benefit of ECD transplantation in the overall sense, that is, averaged across patient subgroups. It is possible that HR of  $\exp\{\hat{\theta}_E\} = 0.89$  under- or overestimates the effect of ECD kidney transplantation for certain patient subgroups. Future analyses, to be reported in the medical literature, will estimate the ECD survival benefit separately for each major demographic and clinical cross-classification.

#### ACKNOWLEDGEMENTS

The authors thank the Organ Procurement and Transplantation Network (OPTN) and the programmers and analysts at the Kidney Epidemiology and Cost Center at the University of Michigan and the University Renal Research and Education Association. We also thank the editor, associate editor, and reviewers, whose constructive comments led to substantial improvement of the original version of the manuscript. This research was supported in part by National Institutes of Health grant R01 DK-70869 (DES). The Scientific Registry of Transplant Recipients (SRTR) is funded by contract number 231-00-0116 from the Health Resources and Services Administration (HRSA), U.S. Department of Health and Human Services. The views expressed herein are those of the authors and not necessarily those of the U.S. Government. This study was approved by HRSA’s SRTR project officer. HRSA has determined that this study satisfies the criteria for the IRB exemption described in the “Public Benefit and Service Program” provisions of 45 CFR 46.101(b)(5) and HRSA Circular 03.

#### REFERENCES

- Andersen, P. K. and Gill, R. D. (1982). Cox’s regression model for counting processes: A large sample study. *Annals of Statistics* **10**, 1100–1120.
- Bilias, Y., Gu, M., and Ying, Z. (1997). Towards a general asymptotic theory for the Cox model with staggered entry. *Annals of Statistics* **25**, 662–682.
- Cox, D. R. (1972). Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society, Series B* **34**, 187–220.
- Cox, D. R. (1975). Partial likelihood. *Biometrika* **62**, 262–276.
- Efron, B. (1982). *The Jackknife, the Bootstrap and Other Resampling Plans*. Philadelphia: SIAM.
- Lin, D. Y., Wei, L. J., Yang, I., and Ying, Z. (2000). Semi-parametric regression for the mean and rate functions of recurrent events. *Journal of the Royal Statistical Society, Series B* **62**, 711–730.
- 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.
- Rabbat, C. G., Thorpe, K. E., Russell, J. D., and Churchill, D. N. (2000). Comparison of mortality risk for dialysis patients and cadaveric first renal transplant recipients in Ontario, Canada. *Journal of the American Society of Nephrology* **11**, 917–922.
- Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika* **70**, 41–55.
- Wei, L. J., Lin, D. Y., and Weissfeld, L. (1989). Regression analysis of multivariate incomplete failure time data by modelling marginal distributions. *Journal of the American Statistical Association* **84**, 1065–1073.
- 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). Comparison of mortality in all patients on dialysis, patients on dialysis awaiting transplantation, and recipients of a first cadaveric transplant. *New England Journal of Medicine* **341**, 1725–1730.

*Received August 2004. Revised October 2005.*

*Accepted October 2005.*