

## ORIGINAL ARTICLE

# Piecewise exponential models with time-varying effects: Estimating mortality after listing for solid organ transplant

Andrew Wey<sup>1</sup>  | Nicholas Salkowski<sup>1</sup> | Walter Kremers<sup>2</sup> | Yoon Son Ahn<sup>1</sup> | Jon Snyder<sup>1</sup>

<sup>1</sup>Scientific Registry of Transplant Recipients, Hennepin Healthcare Research Institute, Minneapolis, MN, USA

<sup>2</sup>Division of Biomedical Statistics and Informatics, Mayo Clinic, Rochester, MN, USA

**Correspondence**

Andrew Wey, Hennepin Healthcare Research Institute, Minneapolis, MN, USA.  
Email: awey@cdrg.org

**Present address**

914 South 8th Street, Suite S4.100, Minneapolis, MN 55404, USA

Patient mortality after listing for a solid organ transplant is a relevant, patient-centric metric, but risk factors for patient mortality after listing present severe non-proportional hazards. We propose piecewise exponential models (PEMs) with time-varying effects to account for the non-proportional hazards, and we use the LASSO to minimize the risk of overfitting. We consider two parameterizations of a PEM: The first model has an overall effect in addition to the time-varying effects (PEM-TID), whereas the second model has only time-varying effects (PEM-TD). Because the LASSO can shrink every time-varying effect to 0, risk factors in the PEM-TID model can have proportional effects during follow-up. In contrast, covariates in the PEM-TD model must have different or no effects during follow-up. These characteristics were illustrated for patients listed for liver transplant. The PEM-TID model had similar or better predictive performance than the PEM-TD model, and both were better than the Cox proportional hazards model. Thus, PEMs with time-varying effects can improve predictive performance for patient mortality after listing for a solid organ transplant.

**KEYWORDS**

proportional hazards assumption, piecewise exponential model, survival analysis

## 1 | INTRODUCTION

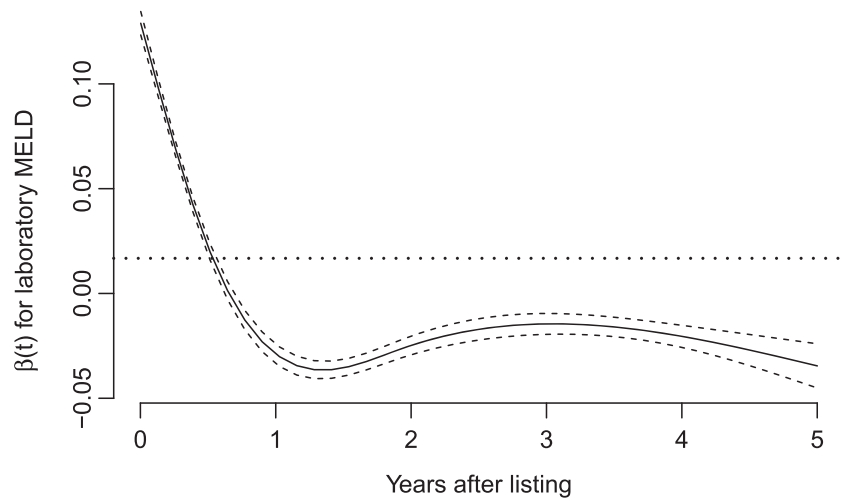
The Scientific Registry of Transplant Recipients (SRTR) is responsible for public reporting of transplant hospital outcomes. Public reporting traditionally focuses on posttransplant outcomes, for example, patient mortality 1 year after transplant (Wey et al., 2019). However, access to transplant can have a larger impact on patient survival after listing due to substantial organ shortages, especially in kidney and liver transplantation (Wey et al., 2019). Thus, SRTR began placing more public reporting emphasis on risk-adjusted transplant rates in addition to posttransplant outcomes (Kasiske et al., 2019). However, patient mortality after listing may better describe the patient experience at a transplant hospital than separately reporting transplant rates and posttransplant outcomes.

Severe non-proportional hazards complicate evaluations of patient mortality after listing. For example, the model for end-stage liver disease (MELD) score prioritizes deceased donor livers for patients with high waitlist mortality rates, but the MELD score is not a strong predictor of posttransplant outcomes. For this reason, MELD has a strong association with the hazard of mortality initially after listing but the association attenuates as more patients undergo transplant; Figure 1 presents the smoothed Scheinfeld residuals for the non-proportional effect of MELD. Risk-adjustment models that account for non-proportional hazards may be more accurate for patient mortality after listing.

Multiple modelling approaches exist that do not rely on the proportional hazards assumption. Accelerated failure time models including, for example, censored quantile regression require fundamentally different distributional assumptions (Portnoy, 2003). Alternatively, non-parametric survival methods, for example, random survival forests, remove most distributional assumptions (Ishwaran, Kogalur, Blackstone, & Lauer, 2008). However, practical issues substantially limit the utility of these alternatives. Quantiles of interest may not be defined in the presence of censoring, whereas non-parametric approaches require large sample sizes, are difficult to interpret, and require extreme computational resources. SRTR has a limited amount of time to estimate risk-adjustment models for public reporting, and the models must be accessible to the broader transplant community. Thus, non-parametric approaches are likely not appropriate for public reporting.

In addition to the practical difficulties with alternative approaches, the transplant community is familiar with the proportional hazards framework. For over a decade, a two-step procedure estimated the hazard ratios for both transplant rates and posttransplant outcomes at

**FIGURE 1** The smoothed Scheonfeld residuals for the effect of laboratory MELD. The effect had strong non-proportionality effects early during follow-up before flattening after 1 year. The dotted horizontal line was the estimated effect in a Cox proportional hazards model. The Scheonfeld residuals were determined from a univariate Cox proportional hazards model with a linear effect for laboratory MELD for patients in the period-prevalent cohort between July 1, 2015, and June 30, 2017



a hospital: (a) a proportional hazards model would estimate the expected number of events at the hospital, and (b) the hazard ratio would then be estimated by the observed and expected number of events. The sequential but separate estimation ensured that the hazard ratios depended only on the patients at the hospital and the risk-adjustment model. Although not ideal, the approach ensured that public reports were tractable and transparent for less technical audiences. Thus, the proportional hazards framework was generalized to handle non-proportional hazards.

We propose piecewise exponential models (PEMs) with time-varying effects for estimating patient mortality after listing (Lehr & Schemper, 2007; Gron, Gerds, & Andersen, 2016). Similar to the Cox proportional hazards model, PEMs model the conditional hazard function using a proportional hazard framework with a constant but different baseline hazards within a priori defined intervals. The time-varying effects weaken the proportional hazards assumption from “same effect over entire follow-up” to “same effect within an interval of follow-up,” which should better approximate the non-proportional hazards in patient mortality after listing. We prefer PEMs with time-varying effects over mixtures of different proportional hazards models due to easier interpretability of covariate effects; for example, a covariate has only a single effect at any given time (Blackstone, Naftel, & Turner Jr. 1986; Blackstone et al., 2018). However, PEMs with time-varying effects present two problems: (a) selecting the intervals with different effects and (b) overfitting due to the inherent flexibility of the model.

There are very few recommendations for selecting intervals in PEMs. Argyropoulos and Unruh (2015) used Gauss Lobatto quadrature to select cutpoints. This approach has two important limitations. First, the cutpoints are not selected with respect to the potential shape of the non-proportional hazards and, therefore, must have a sufficient number of intervals to ensure adequate flexibility. Second, mathematical functions determine the interval cutpoints and, therefore, lack a natural interpretation and are less accessible. To address these issues, we propose a simple heuristic procedure that selects cutpoints with a natural interpretation based on the covariates with the most severe violations of the proportional hazards assumption.

To prevent overfitting, Bender, Groll, and Scheipl (2018) used generalized additive models (GAMs) with smoothing splines, although the approach does not automatically remove unnecessary time-varying effects. Rodriguez-Girondo, Kneib, Cadarso-Suarez, and Abu-Assi (2013) used a double penalized GAM to remove unnecessary smoothing splines for overall and time-varying effects. However, the approach does not remove covariates from the model and, therefore, may require a second variable selection step to remove covariates with no effect. We propose the least absolute shrinkage and selection operator (LASSO) to simultaneously remove both time-varying and overall covariate effects (Tibshirani, 1996; 1997). The resulting model obtains parsimony in three ways: (a) It can remove time-varying effects for covariates with proportional hazards, (b) it can remove unnecessary non-linear effects for continuous covariates, and (c) it can remove all effects for a covariate not associated with patient mortality after listing. The proposed method is shown to possess better predicted error in patient mortality after listing than the traditional Cox proportional hazards model and alternative PEM parameterizations.

Section 2 discusses the data structure for estimating patient mortality after listing for transplant. Section 3 introduces PEMs and their estimation. Section 4 applies PEMs with time-varying effects to liver transplantation. Section 5 investigates the predicted error of the proposed models for adult patients waiting for solid organ transplants in the United States. Lastly, Section 6 discusses limitations and future directions for research.

## 2 | DATA STRUCTURE IN SOLID ORGAN TRANSPLANTATION

Throughout the article, random variables and observed variables are distinguished by capital and lower case letters, respectively. Let the random variable  $T^*$  denote the distribution of patient survival after listing, which may depend on  $p$  patient characteristics at the time of listing, denoted by  $x$ . As discussed in Section 3, we model the distribution of  $T^*$  through the conditional hazard function, denoted  $h(t|x)$ .

The primary goal of evaluating transplant hospitals motivates (a) a period-prevalent cohort, (b) setting the unit of analysis to listings rather than to patients, and (c) selecting the first listing for patients with multiple listings at the same hospital. Listings are a natural unit of analysis because

registration on the waiting list signifies that the hospital intends to treat the patient. Listings also correspond to the beginning of follow-up and align with the entry of data by transplant hospitals. The rest of Section 2 describes the impact of these issues on the resulting data structure. Further, Section 2.3 specifically addresses the exclusion of a time-varying risk factor for patients who undergo transplant.

## 2.1 | Cohort definition

A period-prevalent cohort uses data from all listings at risk at some point during the evaluation window. Let the start and end dates for the evaluation window be  $C_S$  and  $C_E$ , respectively. Inclusion of a listing depends on the listing date (i.e., when the corresponding patient joined the waiting list) and the death date, which are defined as  $W$  and  $T$ , respectively. The distribution of patient mortality after listing is  $T^* = T - w$ , where the lower case of  $w$  emphasizes that the listing date is known. Finally, listings are administratively right-censored  $\tau$  years after listing because (a) patients listed, for example, 10 years prior are less relevant to patients currently joining the waiting list and (b) transplant hospitals have less influence on outcomes further from transplant. Listings are left-truncated if the listing date occurred before the start of the evaluation window, but the patient was alive sometime during the evaluation window ( $W < C_S$  and  $T > C_S$ ). Section 5 discusses the selection of  $\tau$  for the different types of solid organ transplantation.

## 2.2 | Multiple listings

Many patients list at multiple hospitals due to the severe shortage of deceased donor organs and high geographic variability in donor supply. Additionally, patients can relist at the same hospital if a prior transplant fails or functions poorly. Although the outcomes of different listings for the same patient are correlated, each hospital signified an intent-to-treat by listing the patient, and listing at multiple hospitals likely improves long-term survival due to a higher probability of undergoing transplant. For this reason, no adjustment is made for patients listed at multiple hospitals, because hospitals could reduce their mortality after listing by, for example, educating their patients on the benefits of listing at multiple hospitals. In contrast, relisting at the same hospital typically indicates a failed prior transplant. For this reason, follow-up should continue for the original listing with the failed prior transplant. Thus, only the first listing during the evaluation window for patients at a hospital was included.

## 2.3 | Observed data

Due to individual patients listing at multiple hospitals, the unit of analysis is listings rather than patients. The observed data include listings for patients alive at the start of the evaluation window ( $t_i \geq C_S$ ) and listed before the end of the evaluation window ( $w_i \leq C_E$ ). Additionally, listings are followed for mortality only during the evaluation window and are, therefore, potentially subject to left-truncation and right-censoring. Patients listed during the window begin follow-up at listing. Otherwise, follow-up begins at the start of the window:  $s_i^* = \max\{w_i, C_S\} - w_i$ , where  $s_i^*$  is the left-truncation time for the  $i$ th listing. Because patients are referenced with the Social Security Death Master File, they are censored for the end of the cohort window ( $C_E$ ) or surviving  $\tau$  years after listing. This censoring procedure is likely random because it occurs independently of the patient. The observed follow-up time for the  $i$ th listing is  $y_i^* = \min(C_E, w_i + \tau, t_i) - w_i$ , where  $t_i$  is the date of death. Listings for patients who die during the cohort window are identified through the event indicator:  $\delta_i = I\{y_i^* = t_i^*\}$ , where  $I\{B\}$  is an indicator function that is 1 if  $B$  is true and 0 otherwise. Thus, for each listing, we observe  $\{s_i^*, y_i^*, \delta_i, x_i\}$  for  $i = 1, \dots, n$ .

It may be surprising that the observed data do not include a time-varying covariate for whether a patient underwent transplant. However, this is an intent-to-treat analysis with the primary goal of summarizing the effect of transplant hospitals on patient survival after listing. Thus, the model should not include time-varying covariates related to transplant hospital care. Because substantial variability in the likelihood of undergoing transplant exists between transplant hospitals, hospital effects for patient mortality after listing should identify such differences to the extent that it improves patient survival. For this reason, the model should not include time-varying covariates for transplant or other clinical indicators (e.g., hospitalization status) and instead should include only patient covariates at listing.

## 3 | PIECEWISE EXPONENTIAL MODELS

We propose two parameterizations of piecewise exponential models (PEMs) with time-varying effects to account for non-proportional hazards. PEMs model the conditional hazard function of the random variable  $T^*$  but assume a constant baseline hazard and covariate effects in mutually distant follow-up intervals. The first parameterization is

$$h(t|\mathbf{x}) = c_k \times \exp\{\mathbf{x}^T \boldsymbol{\beta} + \mathbf{x}^T \boldsymbol{\beta}_k\}, \quad \text{when } t \in (I_{k-1}, I_k], \quad (1)$$

where  $c_k$  is the baseline hazard for interval  $k$ ,  $I_k$  for  $k = 0, \dots, m$  are the partition points that define each interval,  $\boldsymbol{\beta}$  is the overall covariate effect and is constant over time, and  $\boldsymbol{\beta}_k$  is the deviation of the covariate effect for interval  $k$  from the overall effect. This parameterization is denoted throughout as PEM-TID because the model has “time-independent and -dependent” (TID) effects. The second parameterization does not allow a constant effect over time and, therefore, implicitly assumes that each covariate has a different effect for each interval; this parameterization is denoted throughout as PEM-TD to indicate that all effects are “time-dependent” (TD). Specifically,

$$h(t|\mathbf{x}) = c_k \times \exp\{\mathbf{x}^T \boldsymbol{\beta}_k\}, \quad \text{when } t \in (I_{k-1}, I_k]. \quad (2)$$

The PEM-TID and PEM-TD models span the same model space, but Section 4 illustrates the considerably different fits caused by the LASSO.

As described in Section 1, two important issues arise in the estimation of PEMs. First, the cutpoints (i.e.,  $l_k$  for  $k = 1, \dots, m-1$ ) must be sufficiently narrow to identify significant non-proportional hazards and changes in the baseline hazard but sufficiently wide to allow enough events for accurate estimation. We propose a heuristic procedure in Section 3.1 that balances both issues with the accessibility of the model to the transplant community, which is important for public reporting. The second issue is the significant risk of overfitting due to the additional flexibility of time-varying effects (Lehr & Schemper, 2007). In Section 3.2, we propose estimating the PEM-TID and PEM-TD models with the LASSO to minimize the risk of overfitting.

### 3.1 | Selection of $l_k$

The selection of cutpoints for PEMs must balance (a) minimizing the distance between cutpoints to ensure that non-proportionality is sufficiently identified and (b) ensuring that the intervals are sufficiently wide to accurately estimate covariate effects. We propose a simple heuristic process for selecting cutpoint locations rather than integrating the cutpoint selection into the estimation procedure. The heuristic process helps ensure that the models are relatively interpretable to the transplant community. The specific steps are (a) estimate a Cox proportional hazards model that includes each covariate of interest; (b) determine the worst violations of the proportional hazards assumption through visual inspection of the smoothed Schoenfeld residuals for each covariate on the original time-scale (Grambsch & Therneau, 1994); and (c) using the worst violations of the proportional hazards assumption as a reference, select cutpoints that (1) are close together during periods of follow-up with rapidly changing effects, (2) identify a sufficient number of events in each interval, and (3) allow for a natural interpretation.

Although a natural interpretation is not relevant to predictive performance, models for public reporting of transplant programme outcomes must be transparent and accessible for less technical audiences. For example, a cutpoint at 71 days after listing is less accessible because it does not easily translate into a different unit of time, for example, weeks or months. In contrast, a cutpoint at 90 days after listing is more appropriate because it is approximately 3 months after listing.

### 3.2 | Estimation of PEM-TID and PEM-TD Models

The PEM-TID and PEM-TD models are estimated with the LASSO through the well-known relationship between piecewise exponential models and Poisson models; that is, the likelihood of a PEM is proportional to a Poisson model with an offset equal to the natural log of days at risk during an interval (Gron et al., 2016). The days at risk for listing  $i$  during interval  $k$  is  $R_{i,k}$ ; see Supporting Information for a formal definition. The log-likelihood for the PEM-TID model is then

$$l_n(c, \alpha, \beta, \beta_1, \dots, \beta_k) = \sum_{i=1}^n \sum_{k=1}^m \left\{ \delta_{i,k} \times [\log(c) + \alpha_k + \mathbf{x}_i^T \beta + \mathbf{x}_i^T \beta_k] - c \times R_{i,k} \times \exp(\alpha_k + \mathbf{x}_i^T \beta + \mathbf{x}_i^T \beta_k) \right\} + \lambda \times \left\{ \sum_{l=1}^p |\beta^{(l)}| + \sum_{k=1}^m |\beta_k^{(l)}| + \sum_{k=1}^m |\alpha_k| \right\}, \quad (3)$$

where  $\delta_{i,k} = 1$  if  $\delta_i = 1$  and  $y_i^* \in (l_{k-1}, l_k]$ , else  $\delta_{i,k} = 0$  (i.e., the event indicator for interval  $k$ ),  $c$  is the overall baseline hazard,  $\alpha_k$  is the deviation from the overall baseline hazard for interval  $k$  on the log-scale,  $\lambda$  is the penalty parameter for the LASSO, and  $\beta^{(l)}$  and  $\beta_k^{(l)}$  are the overall and  $k$ th interval effects for the  $l$ th covariate, respectively. The PEM-TID model requires estimating  $c$ ,  $\alpha$ ,  $\beta$ , and  $\beta_k$ , for  $k = 1, \dots, m$ , whereas  $\lambda$  is typically chosen through cross-validation.

The PEM-TD model can estimate each interval separately, and the log-likelihood for interval  $k$  is

$$l_{n,k}(\beta_k, c_k) = \sum_{i=1}^n \left\{ \delta_{i,k} \times [\log(c_k) + \mathbf{x}_i^T \beta_k] - c_k \times R_{i,k} \times \exp(\mathbf{x}_i^T \beta_k) \right\} + \lambda_k \times \sum_{l=1}^p |\beta_k^{(l)}|, \quad (4)$$

where the subscript on  $\lambda_k$  emphasizes that each interval can be estimated separately and that such an approach requires selecting a different penalty for each interval. Similar to the PEM-TID model,  $\lambda_k$  is typically chosen through cross-validation, and the cross-validation procedure minimizes the predicted error separately for each interval.

Importantly, the selection of a separate  $\lambda_k$  for each interval creates a substantial disparity in the size of the design matrices between the PEM-TID and PEM-TD models. Specifically, the PEM-TID design matrix has  $p \times (1 + m) + m$  columns:  $p$  overall covariate effects,  $p \times m$  interval-specific covariate effects, and  $m$  interval-specific terms for the baseline hazard. In contrast, the PEM-TD design matrix for a single interval has only  $p$  interval-specific covariate effects. Additionally, the PEM-TID design matrix typically has significantly more rows than the PEM-TD design matrix. Thus, as illustrated in Section 5, the PEM-TD model usually has substantially lower computational costs (memory and time) than the PEM-TID model.

For estimation, the log-likelihood functions for the PEM-TID and PEM-TD models are, importantly, proportional to Poisson likelihoods with an offset equal to the natural log of days at risk during interval  $k$  (i.e.,  $\log(R_{i,k})$ ). Due to left truncation and right censoring, listings may be at risk only during a few intervals and/or later intervals. Data preparation requires creating a separate row for each interval during which a listing is at risk and including the appropriately defined covariate values, after which each model is estimable with software for fitting Poisson models with a LASSO penalty. In R, the `glmnet` package can estimate each model, although the cross-validation should assign listings rather than rows to each fold, which is not the default behaviour of the package (Friedman, Hastie, & Tibshirani, 2010).

In the PEM-TID model, if the LASSO shrinks every time-varying effect for a given covariate to 0, then the overall effect for the covariate is proportional over time. That is, if  $\beta_k = 0$  for  $k = 1, \dots, m$ , then the PEM-TID model is equivalent to a proportional hazards model. This aspect of the PEM-TID model could improve predictive performance, especially compared with the PEM-TD model, because many covariates will likely have proportional hazards despite the severe non-proportionality of other covariates, for example, MELD. In contrast, the PEM-TD model must re-estimate covariates with proportional effects for each interval, and this may lead to worse predictive performance due to higher variability in the fitting procedure. However, the separate estimation of PEM-TD models substantially reduces the computational costs and may therefore be more feasible in some situations.

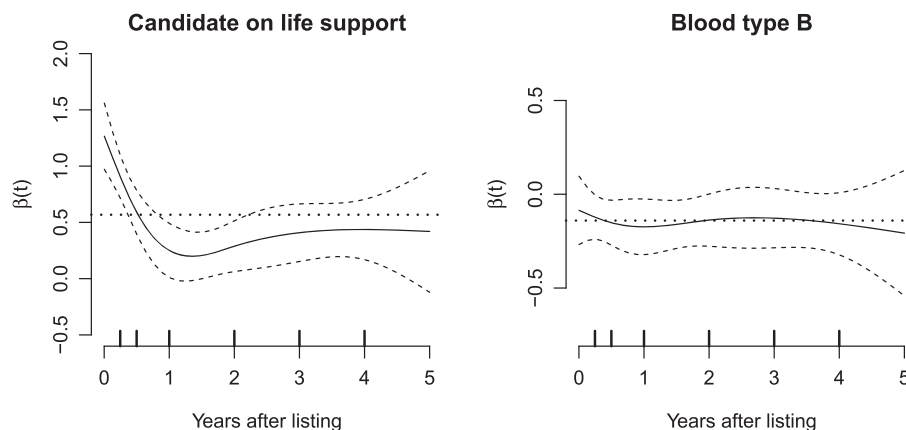
## 4 | APPLICATION TO LIVER TRANSPLANTATION

We illustrate the PEM-TID and PEM-TD models for estimating patient mortality after listing in liver transplantation. This study used SRTR data. The SRTR data system includes data on all donors, waitlisted candidates, and transplant recipients in the United States, submitted by members of the Organ Procurement and Transplantation Network (OPTN), and has been described elsewhere (Leppke et al., 2013). The Health Resources and Services Administration, U.S. Department of Health and Human Services, provides oversight of the activities of the OPTN and SRTR contractors. The period-prevalent cohort started on July 1, 2015, and ended on June 30, 2017. Data were retrieved from the August 2018 SRTR Standard Analytical File.

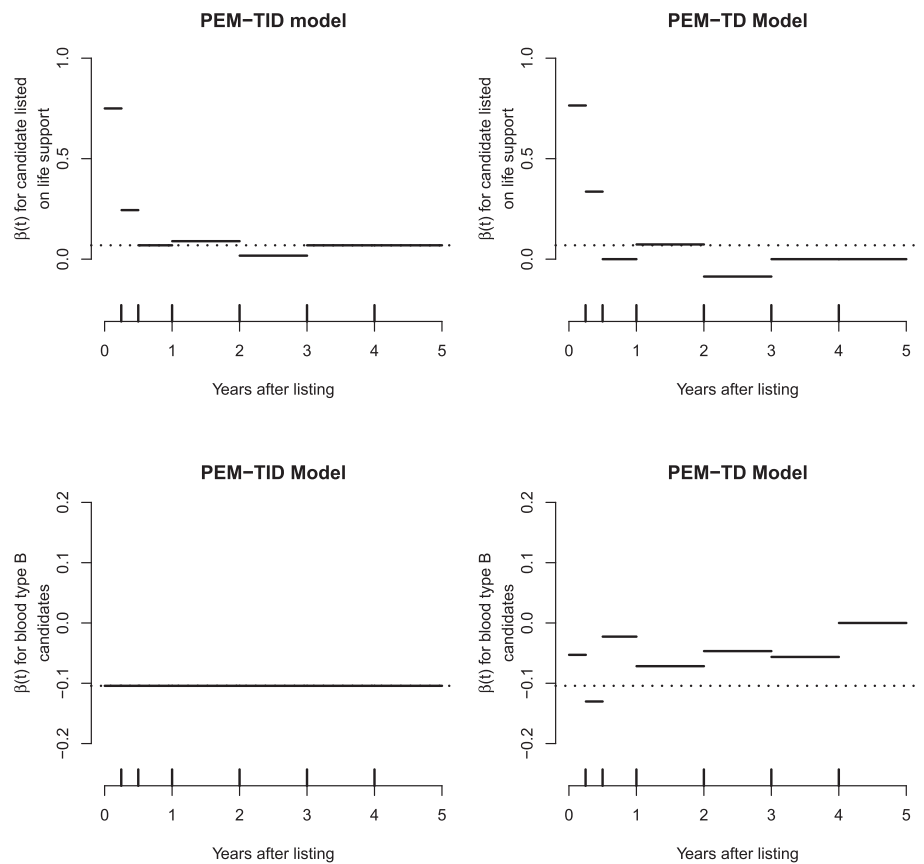
The model for patient mortality after listing included most covariates available at listing; see Supporting Information for a complete list. Left- and right-hand linear splines estimated the effect of continuous covariates. For example, if patient age had a knot at 40 years, then the right- and left-hand splines would be, respectively,  $(\text{age} - 40)_+ = \max(0, \text{age} - 40)$  and  $(40 - \text{age})_- = \max(0, 40 - \text{age})$ . The `pretty` function in R determined the location of knots. The knots were therefore evenly spaced across the covariate distribution and had a natural interpretation, which should improve the accessibility of the model to the transplant community. Multiple imputation handled missing data and was implemented by the `mice` package in R (van Buuren & Groothuis-Oudshoorn, 2011). The multiple imputation included each covariate in the model (see Supporting Information), the natural log of days at risk, and the event indicator.

As noted in Section 3.1, the first step for estimating the PEM-TID and PEM-TD models in liver transplantation is identifying the worst violations of the proportional hazards assumption. Figure 2 presents the smoothed Schoenfeld residuals for candidates listed on life support and blood type B candidates from an adjusted Cox proportional hazards model. The effect for candidates listed on life support significantly violated the proportional hazards assumption. The hazard of mortality was high immediately after listing, but decreased rapidly until about 1 year after listing; after 1 year, the effect slowly increased and then became constant after approximately 3 years. The strong non-proportional hazards immediately after listing was representative of other major violations of the proportional hazards assumption. The intervals for the PEM-TID and PEM-TD models should therefore be narrow immediately after listing due to the significant non-proportional hazards but wider further from listing due to more gradual changes. Importantly, the number of events was the highest immediately after listing, which should allow narrower intervals during the period. For these reasons, the first year after listing was split into three intervals: 0–3 months, 3–6 months, and 6–12 months. The intervals occurred every year after the first year.

Figure 3 presents the estimated time-varying effects for candidates listed on life support (top row) and blood type B candidates (bottom row) for the PEM-TID and PEM-TD models (left and right columns, respectively). The PEM-TID and PEM-TD models showed similar estimated effects for candidates on life support: a rapidly decreasing hazard immediately after listing with a plateau 1 year after listing. Thus, in this case, both approaches identified similar effects of a covariate with significant violations of proportional hazards. In contrast, the estimated effects for blood type B candidates differed. The PEM-TID model showed only an overall effect because the time-varying components were shrunk to 0 by the LASSO, whereas the PEM-TD model typically showed non-zero and different effects for each interval. In other words, the PEM-TD model estimated a time-varying effect for blood type B candidates despite little evidence of non-proportional hazards.



**FIGURE 2** The smoothed Schoenfeld residuals for the effect of candidates listed on life support and blood type B candidates in liver transplantation. The residuals were retrieved from an adjusted Cox proportional hazards model (see Supporting Information for a complete list of covariates). The dotted horizontal lines were the estimated effects in a Cox proportional hazards model



**FIGURE 3** The estimated effects for candidates listed on life support (top row) and blood type B candidates (bottom row) in liver transplantation for the PEM-TID (left column) and PEM-TD models (right column). The dotted line in each figure was the overall estimated effect from the PEM-TID model

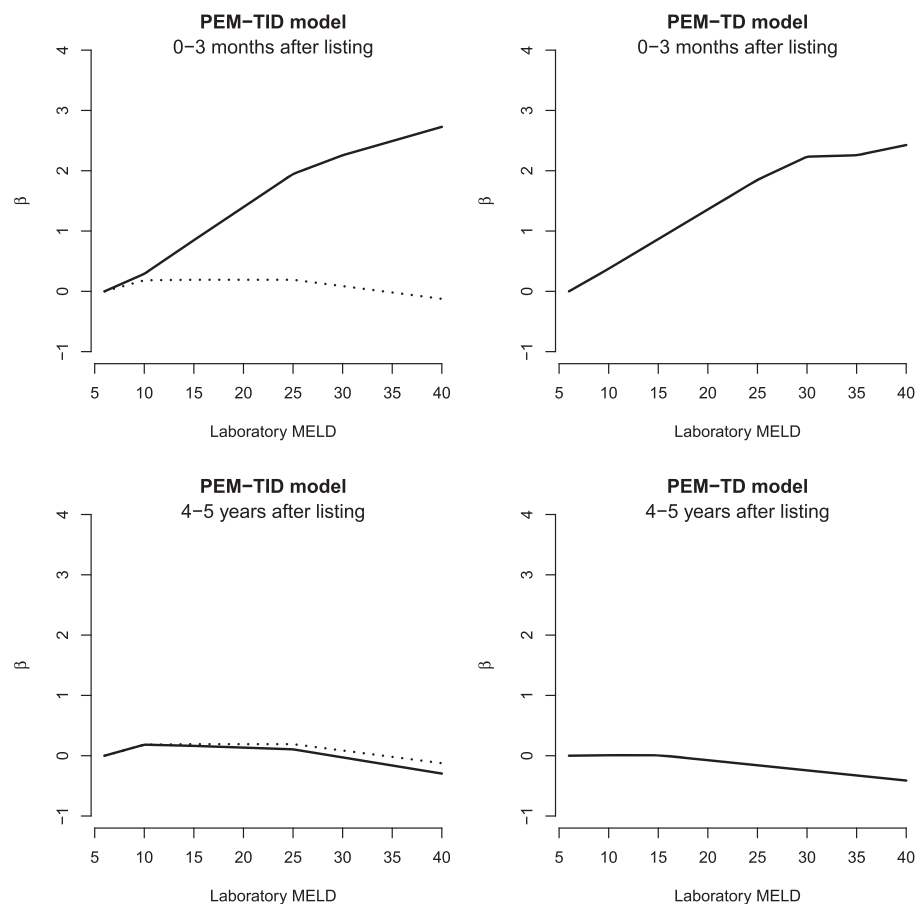
Figure 4 presents the estimated effects for the laboratory MELD score in the PEM-TID and PEM-TD models (left and right columns, respectively) for 0–3 months after listing and 4–5 years after listing (top and bottom rows, respectively). The PEM-TID and PEM-TD models showed similar estimated effects for each time period: a strong effect for 0–3 months after listing and a negligible effect 4–5 years after listing. The overall effect in the PEM-TID model was relatively flat over the range of MELD scores and, therefore, different from the effect for 0–3 months after listing but similar to the effect for 4–5 years after listing.

## 5 | ASSESSMENT OF PREDICTED ERROR

The predictive performance of the PEM-TID and PEM-TD models was compared with the traditional Cox proportional hazards model, a PEM model with the LASSO but no time-varying effects (denoted as PEM-C), and an unpenalized PEM model with time-varying effects (PEM-U). Predicted error was estimated separately for kidney, liver, lung, and heart transplant. Similar to Section 4, the period-prevalent cohort was from July 1, 2015, to June 30, 2017. The penalized PEMs modelled the effect of continuous covariates with left- and right-hand linear splines, and interactions were not considered. The Cox model used penalized splines for continuous covariates rather than linear splines, and the PEM-U model used the first interval as a reference level and included linear effects only for continuous covariates. Data were retrieved from the August 2018 SRTR Standard Analytical File, and Supporting Information lists the covariates included in each model.

The time-dependent C-statistic measured the predictive accuracy of each model. It specifically measures the concordance between observed event status and the predicted survival probability at a prespecified time (Blanche, Kattan, & Gerds, 2019). In contrast, the traditional C-statistic measures the concordance between the predicted risk over the evaluation window and the event times rather than event status. The C-statistic was estimated 1 and 2 years after the start of follow-up for each listing and was evaluated from the start of follow-up for each listing and *not* the time patients initially joined the waiting list. Importantly, 10-fold cross-validation estimated each metric of predicted error, and each fold followed the entire process; for example, the LASSO used a separate cross-validation procedure to select the penalty term for each fold (Simon, Radmacher, Dobbin, & McShane, 2003). Lastly, the number of minutes required to estimate a single model was tracked to illustrate the computational resources of each model. The C-statistic and computational time were averaged over the 10 iterations of multiple imputation.

The cutpoints for the PEM models were determined using the process outlined in Section 3.1. For kidney transplantation, the cutpoints were every year. The cutpoints for liver transplantation were the same as described in Section 4: 0–3 months, 3–6 months, 6–12 months, and then every year after the first year. The cutpoints for lung and heart transplantation were the same: 0–6 months, 6 months to 2 years, 2–3.5 years, and 3.5–5 years. Fewer lung and heart transplant patients necessitated the wider intervals. Finally, the maximum follow-up for each organ (i.e.,  $\tau$  from Section 2) was 5 years after listing or the number of years until at least 90% of candidates were no longer on the waiting list, whichever was



**FIGURE 4** The estimated effects for laboratory MELD for 0–3 months after listing (top row) and 4–5 years after listing (bottom row) for the PEM-TID (left column) and PEM-TD models (right column). The dotted lines were the overall estimated effect for laboratory MELD in the PEM-TID model

longest. The maximum follow-up for liver, lung, and heart transplantation was 5 years ( $\tau = 5$ ) and for kidney transplantation 7 years. Supporting Information presents more information on the number of listings, deaths, selection of  $\tau$ , and cutpoint selection for each organ.

Table 1 presents the time-dependent C-statistic for each model. The C-statistic for the PEM-TID model was typically the best (i.e., highest), although it was slightly lower than that for the PEM-TD model in heart transplantation. For example, the PEM-TID model correctly classified up to 2–3% more listings than the Cox model in liver transplantation and 5–6% more listings in heart transplantations. The PEM-TD model typically had the second best C-statistic. Additionally, the difference in the C-statistics for the PEM-TID and PEM-TD models was usually small. The C-statistic for the PEM-U model was slightly lower than the PEM-TID and PEM-TD models, indicating that the LASSO penalty and/or linear splines improved the C-statistic. Lastly, the PEM-C model had the uniformly worst (i.e., lowest) C-statistic compared with the other models.

A 1–2% difference in the C-statistic may seem small. Yet a 2% difference implies, for example, that an absolute increase of 2% more listings who died 1 year after the start of follow-up were correctly assigned higher risk than listings who did not die. The better classification was achieved by changing only the modelling framework rather than, for example, adding additional covariates. Further, the differences in the C-statistic are similar to the originally reported differences between random survival forests and an unpenalized Cox proportional hazards model (Ishwaran et al., 2008).

Table 2 presents the average number of minutes required for model estimation over the 10 iterations of multiple imputation. The models were clearly ranked from least to most computationally intensive: Cox (fastest), PEM-U, PEM-C, PEM-TD, and PEM-TID (slowest). The PEM-TID

Year	Model	Kidney	Liver	Lung	Heart
1	Cox	0.723	0.696	0.622	0.638
	PEM-C	0.719	0.650	0.614	0.622
	PEM-U	0.735	0.723	0.626	0.690
	PEM-TID	<b>0.739</b>	<b>0.729</b>	<b>0.644</b>	0.697
	PEM-TD	0.737	0.728	0.638	<b>0.698</b>
2	Cox	0.720	0.701	0.639	0.659
	PEM-C	0.717	0.626	0.586	0.611
	PEM-U	0.719	0.715	0.601	0.702
	PEM-TID	<b>0.722</b>	<b>0.718</b>	<b>0.640</b>	0.708
	PEM-TD	0.720	0.716	0.627	<b>0.709</b>

**TABLE 1** The C-statistics for the different models of patient mortality after listing at 1 and 2 years after the start of follow-up

Note. The bold values are the highest (i.e., best) C-statistics for each comparison.

**TABLE 2** The average number of minutes to estimate a single model over 10 iterations of multiple imputation

Model	Kidney	Liver	Lung	Heart
Cox	9.0	3.1	1.2	0.9
PEM-C	28.5	16.2	6.5	6.4
PEM-U	24.3	8.2	0.4	0.5
PEM-TID	234.5	149.3	44.8	34.9
PEM-TD	56.6	43.3	17.7	13.4

Note. The models were fit on the same virtual desktop with 8 CPUs (2.2 GHz) and 64 gigabytes of RAM.

model took significantly longer to estimate than the other models. Although the estimation time of the PEM-TID model is likely feasible for public reporting in solid-organ transplantation, it also requires significantly more memory than the alternative models due to the larger design matrix. These limitations must be considered within each computing environment, especially because the performance of the PEM-TD was usually similar.

## 6 | CONCLUSION

We propose PEMs with time-varying effects to handle the non-proportional hazards of mortality after listing for solid organ transplant. The PEM-TID parameterization, in particular, allows the LASSO to automatically remove effects for covariates without an association but also remove time-varying effects for covariates with proportional hazards. Due to the additional flexibility, PEMs with time-varying effects usually had better predicted error in transplantation than the traditional Cox proportional hazards model (see Section 5). Thus, PEM-TID and PEM-TD models are viable approaches for public reporting of patient mortality after listing and potentially other situations with non-proportional hazards.

Our primary goal is the public reporting of transplant hospital effects on patient mortality after listing. The models for patient mortality after listing aim for the best adjustment of patient risk factors. Otherwise, mortality after listing may appear worse at hospitals that list sicker patients. SRTR usually estimates hospital effects after estimating risk-adjustment models (Salkowski et al., 2014). This approach allows hospitals to better understand the effect of individual patients on evaluations, because the contribution of individual patients depends only on their characteristics and the given risk-adjustment model. Integrating the hospital effects into the risk-adjustment model through, for example, generalized linear mixed models, may improve estimation of the hospital effects. However, hospital effects would be less tractable because they would depend on other hospitals in addition to their patients. The potential advantages of such an approach are worth investigating, especially regarding the accuracy of estimated hospital effects.

It is easy to imagine that the effects of PEMs would change slowly across sequential intervals; that is, the effect of a given interval would be similar to the effects of temporally adjacent intervals. More technically, the difference in effects between adjacent intervals would likely be smaller than non-adjacent intervals, that is,  $|\beta_k - \beta_{k+1}| < |\beta_k - \beta_{k+2}|$ . Yet the proposed PEM-TID model does not incorporate information from adjacent intervals but instead estimates the interval effects separately. A possible extension of the PEM-TID model could penalize the differences between effects of adjacent intervals rather than the absolute effects. Although Gray (1992) and Bryce et al. (2018) use a ridge penalty on the differences of adjacent intervals, similar LASSO-type penalties are a more recent development; see, for example, Tibshirani and Taylor (2011).

The underlying assumptions of the PEM-TID and PEM-TD models on the hazard function are likely not satisfied in practice; for example, the baseline hazard function and the time-varying effects are likely not piecewise constant functions. Instead, the PEM framework is a simplified but convenient representation of the underlying conditional hazard function with better predictive performance than the Cox proportional hazards model. Because the assumptions are likely not true, we did not investigate the asymptotic properties of the PEM-TID and PEM-TD models, although previous research illustrated the consistency of the LASSO in, for example, linear regression (Knight & Fu, 2000).

## ACKNOWLEDGEMENTS

This work was conducted under the auspices of the Hennepin Healthcare Research Institute, contractor for the Scientific Registry of Transplant Recipients, as a deliverable under contract number HSH250201500009C (U.S. Department of Health and Human Services, Health Resources and Services Administration, Healthcare Systems Bureau, Division of Transplantation). The U.S. Government (and other acting on its behalf) retains a paid-up, nonexclusive, irrevocable, worldwide license for all works produced under the SRTR contract, and to reproduce them, prepare derivative works, distribute copies to the public, and perform publicly and display publicly, by or on behalf of the Government. The views expressed herein are those of the authors and not necessarily those of the US Government.

The authors thank SRTR colleague Nan Booth, MSW, MPH, ELS, for manuscript editing.

## DATA AVAILABILITY STATEMENT

This study used SRTR data. The SRTR data system includes data on all donors, waitlisted candidates, and transplant recipients in the United States, submitted by members of the Organ Procurement and Transplantation Network (OPTN), and has been described elsewhere (Leppke et al., 2013). The Health Resources and Services Administration, U.S. Department of Health and Human Services, provides oversight of the activities



of the OPTN and SRTR contractors. The period-prevalent cohort started July 1, 2015, and ended June 30, 2017. Data were retrieved from the August 2018 SRTR Standard Analytical File.

## ORCID

Andrew Wey  <https://orcid.org/0000-0002-2584-3018>

## REFERENCES

- Argyropoulos, C., & Unruh, M. L. (2015). Analysis of time to event outcomes in randomized controlled trials by generalized additive models. *PLoS ONE*, *10*(4), 1–33.
- Bender, A., Groll, A., & Scheipl, F. (2018). A generalized additive model approach to time-to-event analysis. *Statistical Modelling*, *18*, 299–321.
- Blackstone, E. H., Naftel, D. C., & Turner Jr., M. E. (1986). The decomposition of time-varying hazard into phases, each incorporating a separate stream of concomitant information. *Journal of the American Statistical Association*, *81*(395), 615–624.
- Blackstone, E. H., Rajeswaran, J., Cruz, V. B., Hsich, E. M., Koprivanac, M., Smedira, N. G., ..., & Starling, R. C. (2018). Continuously updated estimation of heart transplant waitlist mortality. *Journal of the American College of Cardiology*, *72*(6), 650–659.
- Blanche, P., Kattan, M., & Gerds, T. (2019). The c-index is not proper for the evaluation of *t*-year predicted risks. *Biostatistics*, *20*, 347–357.
- Bryce, C. L., Chang, C. C. H., Ren, Y., Yabes, J., Zenarosa, G., Iyer, A., ..., & Roberts, M. S. (2018). Using time-varying models to estimate posttransplant survival in pediatric liver transplant recipients. *PLoS ONE*, *12*(5), 1–21.
- Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, *33*, 1–22.
- Grambsch, P. M., & Therneau, T. M. (1994). Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika*, *81*, 515–526.
- Gray, R. J. (1992). Flexible methods for analyzing survival data using splines, with applications to breast cancer prognosis. *Journal of the American Statistical Association*, *87*(420), 942–951.
- Gron, R., Gerds, T. A., & Andersen, P. K. (2016). Misspecified poisson regression models for large-scale registry data: inference for 'large *n* and small *p*'. *Statistics in Medicine*, *35*, 1117–1129.
- Ishwaran, H., Kogalur, U. B., Blackstone, E. H., & Lauer, M. S. (2008). Random survival forests. *Annals of Applied Statistics*, *2*, 841–860.
- Kasiske, B. L., Wey, A., Salkowski, N., Zaun, D., Schaffhausen, C. R., Israni, A. K., & Snyder, J. J. (2019). Seeking new answers to old questions about public reporting of transplant program performance in the United States. *American Journal of Transplantation*, *19*(2), 317–323.
- Knight, K., & Fu, W. (2000). Asymptotics for lasso-type estimators. *The Annals of Statistics*, *28*, 1356–1378.
- Lehr, S., & Schemper, M. (2007). Parsimonious analysis of time-dependent effects in the Cox model. *Statistics in Medicine*, *26*, 2686–2698.
- Leppke, S., Leighton, T., Zaun, D., Chen, S.-C., Skeans, M., Israni, A., ..., & Kasiske, B. (2013). Scientific registry of transplant recipients: Collecting, analyzing, and reporting data on transplantation in the united states. *Transplantation Reviews*, *27*, 50–56.
- Portnoy, S. (2003). Censored regression quantiles. *Journal of the American Statistical Association*, *98*, 1001–1012.
- Rodriguez-Gironde, M., Kneib, T., Cadarso-Suarez, C., & Abu-Assi, E. (2013). Model building in nonproportional hazard regression. *Statistics in Medicine*, *32*(32), 5301–5314.
- Salkowski, N., Snyder, J. J., Zaun, D. A., Leighton, T., Israni, A. K., & Kasiske, B. L. (2014). Bayesian methods for assessing transplant program performance. *American Journal of Transplantation*, *14*, 1271–1276.
- Simon, R., Radmacher, M., Dobbin, K., & McShane, L. (2003). Pitfalls in the use of dna microarray data for diagnostic and prognostic classification. *Journal of the National Cancer Institute*, *95*(1), 14–18.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B*, *58*, 267–288.
- Tibshirani, R. (1997). The lasso method for variable selection in the Cox model. *Statistics in Medicine*, *16*, 385–395.
- Tibshirani, R. J., & Taylor, J. (2011). The solution path of the generalized lasso. *The Annals of Statistics*, *39*(3), 1335–1371.
- van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, *45*(3), 1–67.
- Wey, A., Gustafson, S., Salkowski, N., Kasiske, B. L., Skeans, M., Schaffhausen, C., ..., & Snyder, J. J. (2019). Association of pretransplant and posttransplant program ratings with candidate mortality after listing. *American Journal of Transplantation*, *19*(2), 399–406.
- Wey, A., Salkowski, N., Kasiske, B. L., Skeans, M., Schaffhausen, C., Gustafson, S., ..., & Snyder, J. J. (2019). Comparing scientific registry of transplant recipients posttransplant program-specific outcome ratings at listing with subsequent recipient outcomes after transplant. *American Journal of Transplantation*, *19*(2), 391–398.

## SUPPORTING INFORMATION

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

**How to cite this article:** Wey A, Salkowski N, Kremers W, Ahn YS, Snyder J. Piecewise exponential models with time-varying effects: Estimating mortality after listing for solid-organ transplant. *Stat*. 2020;9:e264. <https://doi.org/10.1002/sta4.264>