PREDICTING POTENTIAL SURVIVAL RATES OF KIDNEY TRANSPLANT CANDIDATES FROM DATABASES WITH EXISTING ALLOCATION POLICIES

Inbal Yahav and Galit Shmueli
R.H. Smith School of Business,
University of Maryland, College Park, MD
{iyahav,gshmueli}@rhsmith.umd.edu

Abstract

Under the current kidney allocation system in the United States, kidneys are allocated to patients primarily through a combination of tissue matching, sensitization level, and waiting time. However, due to recent trends in medicine and the shortfall of kidney supply, the current system fails to match donors and recipients well. In an effort to improve the allocation system, the United Network of Organ Sharing (UNOS) defined principle factors that would determine a new allocation policy. The most prominent factor is patients' potential remaining lifetime.

Estimating "potential remaining lifetime" is complicated for several reasons. First, the characteristics of candidates in the waitlist are different from kidney recipients, implying that the mortality of candidates does not represent the mortality that would have occurred among recipients, had they not received a transplant. Second, treatment methods of patients without a transplant have changed over the last two decades, making lifetime predictions less certain. Lastly, the lifetime model should extrapolate future survival beyond the duration of the data.

In this paper, we use a parametric Weibull Accelerated Failure Time (AFT) model to predict survival rates and show its advantage over common in terms of prediction accuracy. We also use data mining methods, and in particular, classification and regression trees, to tackle recipients' selection bias and bias caused by changes in medical treatment.

Keywords: kidney allocation, parametric survival analysis, selection bias, predictive accuracy, classification trees.

Introduction

According to the Scientific Registry of Transplant Recipients (SRTR) annual statistics, more than 85,000 candidates with kidney failure End Stage Renal Disease (ESRD) are currently waiting for transplantation in the United States. Whereas the number of annual transplants stands at approximately 15,000, the number of annual waiting list additions reaches 30,000. This imbalance between demand and supply of organs raises a need for an efficient *organ allocation policy* to determine the order in which candidates are offered

an organ, when one becomes available.

The current kidney allocation system was developed in the 1980's and 1990's. The system revolves around a set of priority points given to candidates based on tissue matching and sensitization level combined with waiting time and age. However, as the waiting list has grown and the population of candidates has aged the allocation process has become dominated by waiting time, thereby effectively becoming a "first come first transplant" system.

In an effort to improve the allocation system, the Organ Procurement and Transplantation Network (OPTN)/United Network for Organ Sharing (UNOS) Kidney Transplantation Committee evaluated a variety of different alternative approaches to kidney allocation. One of the most prominent factors that the committee defined, is the medical outcome of expected patient survival without a transplant.

Predicting patients' potential lifetime offers various statistical challenges. The main challenge is that existing data are 'contaminated' by the existing allocation policy, and hence the training data differ from the prediction set. Candidates in the waiting list are not selected for transplant at random, implying that the medical history of transplant recipients is statistically different from that of candidates who have not received a transplant. For transplant recipients, we do not know their lifetimes had they *not* received the transplant. Hence, we can only use medical history data on candidates without transplant, which may not be representative of the actual waiting list. Other challenges include waiting list sampling biases, and extrapolating lifetime predictions beyond the duration of the available data.

In this paper, we propose a data mining procedure for reducing patients' selection bias that imputes the remaining lifetime of candidates if they do not receive a transplant. This procedure yields a complete data set of waiting list lifetimes. We then present a parametric survival model for estimating candidates' survival rates, which is based on the completed dataset.

Data

We consider a dataset of waiting list registrations and transplants of kidney and simultaneous kidney-pancreas[1] that have been listed or performed in the U.S. and reported to the OPTN between October 1, 1987 and August 15, 2008. The dataset includes records on both deceased and living-donor transplants. The data were exclusively provided by UNOS.

Preliminary analysis of the data exhibits a rapid increase in patients' lifetime over the last 30 years, plausibly resulting from changes in recent medicine and dialysis treatments. Rather than incorporating these changes into our model, as suggested in [2], we consider a subset of the patients who *joined* the waiting list after January 2000. Truncating the

---

[1]Simultaneous transplantation of a kidney and pancreas is performed for those who have kidney failure as a complication of insulin-dependent diabetes mellitus (also called Type I diabetes).

data according to arrival time (as opposed to waiting list status on that year) ensures that the patients' arrivals are approximately uniformly distributed[2] over the studied interval: [2000, 2008]. Our studied subset includes over 265,000 patients, of which about 100,000 have received transplant and about 28,000 died while waiting for transplant.

In the reminder of the analysis we split the dataset of candidates without transplant into two sets: training (%60) and holdout (%40). The training set is used to build the model for imputing lifetime of kidney recipients, had they not received a transplant, as is classically done in data mining. The holdout set is later used for performance evaluation.

Correcting for Patients' Selection Bias

In this section we address the differences between the medical history of transplant recipients ("recipients") and that of candidates without transplant ("candidates"). Whereas previous literature assumes there are no significant differences between these two groups ([4, 5, 6, 8], with the exception of [3, 7]), we find that recipients and candidates differ statistically in terms of mortality-rate-related characteristics such as diabetes status ("DIAB"), need for a simultaneous kidney-pancreas transplant ("KP"), and hypertension information ("HYPERTENSION"), as illustrated in Figure 1. Although patients are not selected for transplant according to their survival rate, in practice patients with higher survival rates are likely to refuse to accept offers that are not best suited to them (as advices by their physicians). In fact, [9] states that the refusal rate of offered kidneys is as high as 45%.
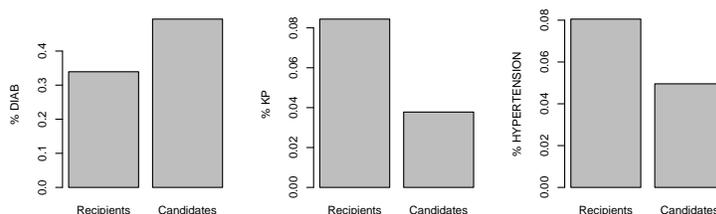

Figure 1: Comparison between medical records of recipients of kidneys and candidates (who have not received transplant).

To reduce bias when modeling the remaining lifetimes based on the data at hand, [7] suggests a conceptual procedure for imputing expected lifetimes of recipients from candidates that utilizes medical similarities *at time of kidney offer* between recipients and a single other patient. The definition of candidates' medical similarity, however, remains undefined.

We impute recipients' expected lifetime using a different approach that consists of two steps. In the first step, we impute death incidences ("DEATH") of recipients (assuming

---

[2]Under the assumption of Poisson arrival times, as evidenced from the data.

that they have not received a transplant) using a classification tree procedure that utilizes candidates medical history records *at time of joining the waiting list* (denoted "H_ARR") along with their *most recent information* (denoted "H_CURR"). Then, *conditional on death incidence*, we impute recipients' remaining lifetime ("WT"), using a regression tree. If a recipient is predicted to survive without the transplant, we complete his lifetime as the span between his joining the waiting list ("ARR_DATE") and the dataset truncation date ("END_DATE"). A schematic representation of this computation (for patient i) is provided in Equation 1, where F() is the regression tree imputation of WT.

$$\text{WT}_i = \begin{cases} \text{F(H\_ARR}_i, \text{H\_CURR}_i), & \text{if DEATH}_i=1 \\ \text{END\_DATE} - \text{ARR\_DATE}_i, & \text{if DEATH}_i=0 \end{cases} \tag{1}$$

Whereas imputing recipients' lifetime using a regression tree is straightforward (their lifetime is predicted as the mean lifetime of the respective terminal node), imputing death incidents (a binary outcome) requires an extra step of choosing a cutoff value, c (see Equation 2). For brevity, we omit the discussion on choosing the cutoff value from this abstract. Figure 2 presents a ROC curve showing sensitivity vs. 1-specificity for different choices of cutoff values between 0 and 1. In the remainder of the analysis we chose a cutoff value of 0.3, for which the total error rate is relatively low (less then 30% for each error type) and distributed evenly between the error types. Using this cutoff value, we predict 25% death incidences among recipients.

$$\text{DEATH}_i = \begin{cases} 1, & \text{if the survival rate of the respective terminal node} < c \\ 0, & \text{otherwise} \end{cases} \tag{2}$$
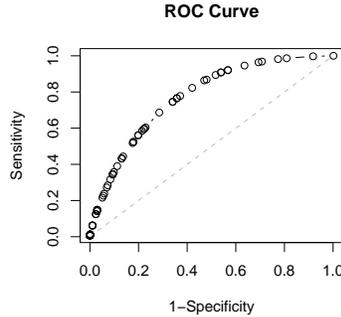

Figure 2: ROC Curve for different choices of cutoff values.

Choice of Survival Model

Although the semi-parametric Cox proportional hazards model is widely used in medical research (e.g., [8]), we use a parametric Accelerated Failure Time (AFT) model with a Weibull distribution for predicting patient survival rates. The AFT model offers two advantages that are useful in our context. First, AFT models are known to be more robust to unmeasured or neglected covariates (also referred to as *hidden heterogeneity*), compared to proportional hazards models [1]. Second, unlike the Cox model, which does

not provide extrapolation capabilities beyond the training set duration, the AFT model enables us to extrapolate survival rates beyond the studied interval (here, 8 years).

The choice of a Weibull distribution over other common survival distributions (e.g., lognormal and exponential) is based on predictive accuracy considerations. In particular, we find that an AFT model with a lognormal distribution appears incorrectly over-optimistic regarding patients' lifetime compared to an AFT with a Weibull distribution, as can be seen in Figure 3, which compares the forecast error distributions (the left skewed shape of the residual plots stems from the difference between the range of the observed lifetime [0,8] and that of the potential lifetime [0,∞)).
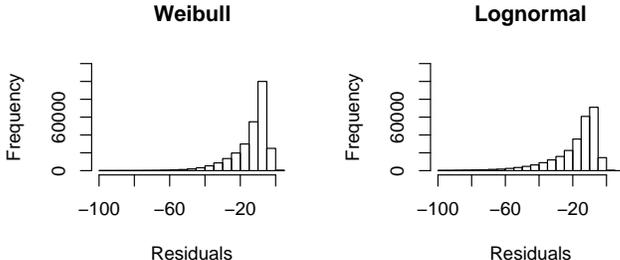
Figure 3: Comparing prediction error distributions for Weibull and Lognormal AFT models.

In Figure 4 we plot the baselines of the survival models, with and without the imputation of death and lifetime. The predictors that are included in the survival model where selected according to the approximating in [8]. Clearly, the model without the imputation is incorrectly over-pessimistic in terms of the number of death incidences.
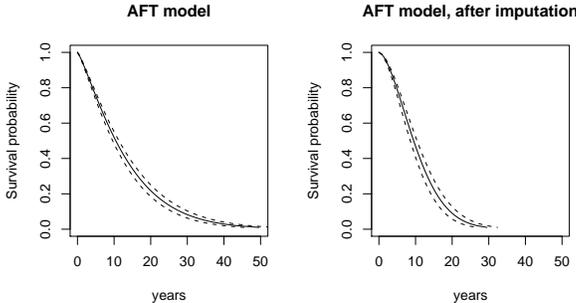
Figure 4: Survival baseline of ATF model with Weibull distribution, before and after imputation.

Discussion and Future Work

This paper presents a model for predicting survival rates of kidney transplant candidates from databases with existing allocation policies. Our approach consists of a pre-processing procedure to impute recipients' lifetime information and a parametric survival

analysis that predicts patients' survival rates. The highlights of this procedure are provided below:

Data imputation: Imputing recipients' potential lifetime (had they *not* received transplant) in two steps:

> **Step 1:** Imputing death incidences among recipients
>
> **Step 2:** Completing recipients' lifetime, conditional on the event of death

Survival analysis: Using a parametric Accelerated Failure Time (AFT) model to predict patients' survival rates based on the completed dataset

In future work we plan to test additional data imputation techniques and to evaluate them based on their *direct* predictive accuracy, that is, predicting lifetime and death incidences of candidates in holdout set, and their *indirect* predictive accuracy, or in other words, predicting survival rates from the imputed dataset. We will compare the performance of our survival model to the weighted proportional hazards regression model, proposed in [3].

References

[1] P Lambert, D Collett, A Kimber, and R Johnson. Parametric accelerated failure time models with random effects and an application to kidney transplant survival. *Statistics in medicine*, 23(20):3177–3192, 2004.

[2] EA Mauger, RA Wolfe, and FK Port. Transient effects in the Cox proportional hazards regression model. *Statistics in medicine*, 14(14):1553–1565, 1995.

[3] Q Pan and DE Schaubel. Evaluating bias correction in weighted proportional hazards regression. *Lifetime data analysis*, 15(1):120–146, 2009.

[4] DE Schaubel, MK Guidinger, SW Biggins, JD Kalbfleisch, EA Pomfret, P. Sharma, and RM Merion. Survival benefit-based deceased-donor liver allocation. *American Journal of Transplantation*, 9(4p2):970–981, 2009.

[5] SRTR. Methods for Discounting Median Lifetimes. *Working paper*. 2007.

[6] SRTR. Predicting the Life Years From Transplant (LYFT): Choosing a Metric. *Working paper*. 2007.

[7] RA Wolfe. Avoiding statistical bias in predicting the life years from transplant (LYFT). *Working paper*. 2007.

[8] RA Wolfe, KP McCullough, DE Schaubel, JD Kalbfleisch, S. Murray, MD Stegall, and AB Leichtman. Calculating life years from transplant (LYFT): Methods for kidney and kidney-pancreas candidates. *American Journal of Transplantation*, 8(4p2):997–1011, 2008.

[9] SA Zenios. Models for kidney allocation. *Operations Research and Health Care: A Handbook of Methods and Applications*, page 537, 2004.