DEEP LEARNING SURVIVAL ANALYSIS FOR CLINICAL DECISION SUPPORT IN DECEASED DONOR KIDNEY TRANSPLANTATION

BY

PAUL RUALES

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Computer Science
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2019

Urbana, Illinois

Adviser:

Professor Sanmi Koyejo

ABSTRACT

In deceased donor kidney transplantation, the decision to accept or decline an offer relies on a clinicians intuition and ability to digest complex information in order to maximize patient survival. Risks affecting patient survival post-KT must be balanced with the risks of remaining on the waitlist. These risks include mortality, graft failure, and becoming too sick to transplant. The allocation system today takes these risk into account by way of the KDPI and EPTS scores. While these scores are discriminative of patient survival they were built with an assumption of independence between risks and very few donor-recipient variables. Deep learning survival analysis can effectively handle competing risks and learn complex relationships between many more donor-recipient variables. We used DeepHit to assess the risk benefit associated with accepting a kidney offer or remaining on the waitlist. Our models achieved comparable, if not better performance in certain tasks, with other high performing models in the literature and revealed that decoupling competing risks led to increased clinical information gain. We show that comprehensively modeling competing risks using machine learning can achieve more granular, meaningful clinical risk analysis enabling more effective decision making in deceased donor kidney transplantation.

To all those that have shaped the person I am today.

ACKNOWLEDGMENTS

I would like to express my great appreciation to my advisor, Professor Sanmi Koyejo, for providing mentorship and guidance for the past two years. He always gave me direction and encouragement while I explored my interests at UIUC. I leave UIUC with greater direction in life than I have ever had thanks to his patience and wisdom. Additionally, I would like to thank the other students in his research group. I came in with a background in a different discipline and they all gave me valuable insights at every lab meeting into their work and broader computer science concepts that I will forever leverage as a multi-disciplinary thinker.

I would also like to thank my friends old and new for their support and company throughout my journey. Whenever times were stressful I knew I could detach with them on the basketball court or the Rocket League field. In particular, I would like to thank Eric Pahl, CTO at OmniLife and good friend, for collaborating with me and providing me with funding to join him and everyone else at OmniLife in improving the kidney transplantation process. Here's to the several hour long conversations we would have discussing the future of medicine and philosophizing on life's beauty. I wish him and his company the best of luck moving forward.

Most importantly, I would like to thank my family for always being there for me when I need them and giving me the confidence to pursue my aspirations. Lastly, I would like to recognize my two dogs, Duke and Benji, who kept me sane during very long hours of work in front of my computer.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION

Organ transplantation is a complex process requiring effective coordination and communication amongst a highly distributed network of providers within a short window of time in order for a successful transplantation to occur (Figure 1.1). Over 80% of patients awaiting organ transplants are waiting for a kidney. Kidney transplantation (KT) is a clinically effective and cost-effective treatment for patients suffering from end-stage renal disease (ESRD) [1, 2, 3]. More than $34B of the annual Medicare budget is devoted to paying for the treatment of patients who require dialysis while awaiting a KT. Transplanting this group could save $100,000 per year per patient amounting to $46.8B annually [4] and a gain of 6.7 discounted quality-adjusted life years (QALYs) for every dialysis patient who receives a kidney. Despite the obvious clinical and cost-effectiveness advantages of kidney transplantation, there has been a decades-long rise in the kidney discard rate from 5.1% in 1988 to 19.2% in 2015 [5]. Studies have shown that a portion of these kidney could have provided a survival benefit to waitlisted patients and that behavioral factors, such as increased risk aversion, could play a part in the rising discard rate [5, 6, 7]. On the behavioral side, there is also a demonstrated reduction in acceptance of kidneys on the weekends v.s. the weekdays [8]. Ultimately, a kidney accept or decline decision comes down to a complex estimation of risk done by sole physicians relying on experience and intuition. Clinical decision support (CDS) may significantly increase access, increase quality, and reduce the cost of kidney transplantation.

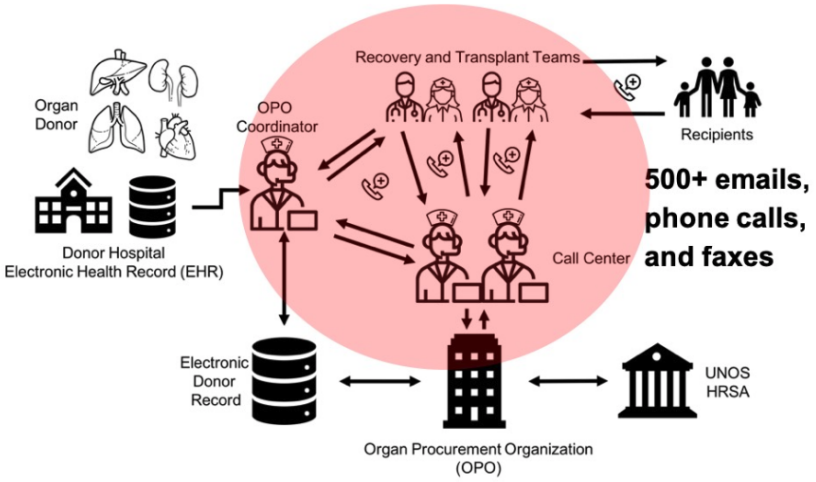

Figure 1.1: Complex process of organ allocation and procurement.

Similar to reasoning done by Wey et al.[9], the ultimate question is: does this offer max-

imize graft and patient survival for the recipient? If it does, the offer should of course be accepted. It's important to note that there isn't one single donor profile that maximizes those two variables for every recipient. Aside from the baseline compatibility requirements (e.g. size, blood type, and other medical factors), other factors such as geography and local organ supply come into play. Therefore, an accept or decline decision comes down to 3 factors:

1. Survival should the offer be accepted.
2. Survival should the offer be declined and the patient remain on the waiting list (WL).
3. Survival opportunity sometime in the future in the form of a better offer.

This thesis attempts to robustly addresses factors 1 and 2 using state of the art deep learning techniques in survival analysis with competing risk. Factor 3 is left for future work and will build on work done by Wey et al. and Bertsimas et al [9, 10].

The rest of the thesis is organized as follows. Chapter 2 dives into related work and provides a good overview on survival analysis within kidney transplantation. Chapter 3 explains our approach and describes the Scientific Registry of Transplant Recipients (SRTR) data set we used. Chapter 4 shows our results and attempts to demonstrate the clinical validity using CDS at the time of offer. Chapter 5 concludes this thesis and ventures into future work.

CHAPTER 2: RELATED WORK

Kidney transplantation as explained in the introduction is a very complex system involving multiple patients vying for a perfectly matched kidney. There are multiple data sources in play from the allocation system, the individual patients, and the organ procurement system. Machine learning techniques offer a more efficient way at processing this large amount of data and the complex interactions between them with the ultimate goal of reducing waste, transplanting more patients, and maximizing patient survival. The following sections will describe the following work in relation to our work: (1) how survival analysis is used in kidney transplantation, (2) current allocation policies using machine learning, (3) the SRTR efforts in providing analytic tools for patient and physician use, and (4) previous applications of machine learning for kidney offer evaluation.

2.1 SURVIVAL ANALYSIS IN KIDNEY TRANSPLANTATION

Survival analysis is used extensively in medicine to analyze time-to-event data such as from a cohort study or clinical trial. In the case of post-KT survival, two events are prevalent: death and graft failure. Survival analysis allows one to find casual relationships, predict temporal risk, or analyze non-parametric based patient survival. Additionally, this type of analysis allows for censoring of individuals. For example, if in the study period a patient did not exhibit any outcome of interest they are considered to be right-censored. Left-censorship occurs when a birth event is not seen.

Common methods used in KT are the non-parametric Kaplan-Meier (KM) method [11], the Cox proportional hazard model [12], and the Fine and Gray model [13]. The KM method estimates the probability of survival across a period of time using only time and censoring data. A Cox model would be used in evaluating casual relationships between predictive covariates and survival. It consists of a baseline hazard which represents the hazard of an individual with baseline covariates. This baseline is affected by each covariate, therefore allowing one to study the effect of individual covariates on survival. The Fine and Gray model is a more advanced technique providing sub-distribution hazards in the presence of competing risks. In addition the cumulative incidence function (CIF) was developed to overcome some of the shortcomings of the KM method [14]. It effectively shows the cumulative probability over time due to an event in the presence of competing events.

A proper example of survival analysis in kidney transplant was done by Sapir-Pichhadze et al. in their study of WL kidney transplant candidates [15] with competing events (death,

transplant or removal from WL for any other reason). They compared competing event analysis using a cause-specific Cox model with the Fine and Gray model. Interestingly, they demonstrated that appropriately handling competing risks leads to more refined predictions in the presence of competing risks. This thesis takes care to handle competing risks not only for WL outcomes but also in post-KT outcomes.

2.2 ALLOCATION POLICIES

The US allocation system for kidney's recently began factoring in survival benefit in 2015. Survival benefit is considered by using a candidates estimated post transplant survival (EPTS) and a donors kidney donor profile index (KDPI). These scores were developed using Cox regression. The EPTS score considers the following four factors: candidate time on dialysis, current diagnosis of diabetes, prior solid organ transplants, and candidate age.[1] The KDPI score considers a donors age, ethnicity, creatinine levels, history of hypertension, history of diabetes, cause of death, height, weight, donor type (deceased or living), and HCV status.[2] Lower KDPI scores have been shown to be associated with higher rates of survival. For example, The Organ Procurement and Transplantation Network states that the half-life of a graft from a donor with KDPI of 0%-20% is expected to be around 11 years post-KT and decreases to about 9 and 5.5 years with KDPI's of 21%-85% and 86%+, respectively [16]. Similar expectations of survival with EPTS have been found. To maximize survival benefit good recipients are matched with good donors using KDPI and EPTS. Candidates with EPTS scores of 20% or less are matched with kidneys from donors with KDPI scores of 20% or less before candidates at local, regional, and national levels of distribution [17].

KDPI and EPTS represent proper first steps towards a more efficient allocation system. However, more robust machine learning methods have shown better discriminative performance [18]. It is important to note that the variables used in these metrics are more or less guaranteed to be on hand at the time of transplant. To maintain usability in practice these variables should be included along with other ubiquitous variables.

[1] https://optn.transplant.hrsa.gov/resources/allocation-calculators/epts-calculator/
[2] https://optn.transplant.hrsa.gov/resources/allocation-calculators/kdpi-calculator/

2.3 SRTR ANALYSIS TOOLS

The SRTR currently publishes risk adjustment models models for both post-KT outcomes and WL outcomes among others.[3,4] For Post-KT outcomes they include 1 and 3 year graft and patient survival. For WL outcomes they include transplant rate, deceased donor transplant rate and WL mortality. By building seperate models for every event they are employing cause-specific Cox and Poisson survival models. This approach has been found to be the preferred method for evaluating causal relationships in the presence of competing risk, but not for predicting the actual risk of the outcome [15]. Indeed the SRTR website only lists model coefficients allowing a transplant professional to examine the effect of a predictive variable on the outcome of interest; leaving room for more patient-specific risk models to be developed. Additionally, usability in practice is limited due to the amount of manual processing and interpretation needed from a users perspective.

2.4 MACHINE LEARNING FOR KIDNEY OFFER EVALUATION

In kidney transplantation alone, machine learning has been used to predict chronic allograft rejection, delayed graft function (DGF), and allograft survival [19, 20, 21, 22, 23].

Decruyenaere et al. [19] compared logistic regression with more advanced machine learning methods for predicting DGF after kidney transplantation. DGF means a patient must go back on dialysis within the first week after transplantation. They concluded that a linear SVM is most appropriate for this task, not finding any significant performance boost using random forests (RF), stochastic gradient boosting (SGB), and decision trees (DT) among others. However, their dataset was relatively small (under 500 samples) and they only had access to 55 total variables, 20 of which were used in training. Furthermore, the DGF outcome was only observed in 12.5% of patients.

Krikov et al. [20] built tree-based models to predict the probabiliy of kidney allograft survival at 1, 3, 5, 7, and 10 years post-KT. They had relatively good success with high AUC scores of 0.63, 0.64, 0.71, 0.82 and 0.90 for their respective prediction timelines. However, their models converted the outcome to a binary indicator and potentially missed information relating to competing outcomes such as death without graft failure and graft failure without death.

Li et al. [23] used a bayesian network to predict graft rejection and survival period. Their graft rejection model performed relatively well with 97.8% accuracy, however accuracy was

[3]https://www.srtr.org/reports-tools/risk-adjustment-models-posttransplant-outcomes/

[4]https://www.srtr.org/reports-tools/risk-adjustment-models-waiting-list/

68.2% for the survival period task. Their results indicated that a bayesian network was ill suited for the former task.

These studies only looked at post-KT events, effectively only addressing factor 1. Studies that included WL survival have looked at predicting WL mortality and other events pertaining to removal from the WL [9, 18]. Both of these studies addressed factor 1 & 2 and used survival analysis techniques. Wey et al. also addressed factor 3 by incorporating survival estimates after declining an offer and receiving a subsequent living or decreased donor offer. Neither of these studies utilized deep learning to handle survival analysis in the presence of competing risks nor effectively handled bias in their performance evaluation [24].

CHAPTER 3: APPROACH

3.1 DATASET

There are multiple large, comprehensive national kidney transplantation datasets available [25]. Data curated by the Organ Procurement and Transplantation Network (OPTN), described in Table 3.1, serves as the basis for the most widely used datasets including the United Network for Organ Sharing (UNOS), the Scientific Registry of Transplant Recipients (SRTR), and the United States Renal Data System (USRDS) datasets. We used data obtained from the SRTR. Additional data for the ascertainment of graft failure and death came from the Centers for Medicare and Medicaid Services (CMS); cancer ascertainment from the Surveillance, Epidemiology, and End Results program (SEER); and additional death ascertainment from the National Death Index (NDI).

Table 3.1: The OPTN collects most data via one of three forms described in the table. These forms provide the basis for many of the larger datasets used by researchers today.

OPTN Form	Types of Data Collected
Transplant Candidate Registration (TCR)	Candidate demographic data, clinical information and history, organ-specific information at the time of listing.
Transplant Recipient Registration (TRR)	Pre-transplant clinical data, infectious disease status, data on the transplant procedure, post-transplant clinical data, information on immunosuppressive medications from the initial transplant admission.
Transplant Recipient Follow-up (TRF)	Vital status, cause of death if applicable, graft status patient education and employment status, and clinical information at each visit following a transplant.
Deceased Donor Registration (DDR) and Living Donor Registration (LDR)	Donor demographics, comorbidities, infectious disease status, and cause of death (for deceased donors) or postoperative clinical information (for live donors) at the time of organ donation.

3.2 STUDY POPULATION

The recipient cohort included all adult (aged 18 and over) KT recipients between January 1, 2005 and December 31, 2016 (n = 114,261). Multi-organ recipients were excluded. Patients who had received a prior transplant and only deceased donor KT recipients were included. The waiting list cohort included all adult patients who were activated on the KT waiting list between January 1, 2005 and December 31, 2016 (n = 217,278). Multi-organ candidates were excluded. If a candidate didn't have a removal date they were excluded and multiple registrations were consolidated.

3.3 TRANSPLANT SURVIVAL

Transplant survival was defined as the time from KT to death or graft rejection, censoring for the end of the study. We treated graft failure as a competing event to death and noted the first event that occurred as the primary outcome for that patient. To maintain usability in practice we limited the features to those used by the SRTR postransplant models (described in related work) and 12 other variables found to influence post-KT survival in the previous studies. A sample was excluded if it was missing 20% or more of it's variables. A total of 25 features were utilized. Missing continuous variables were replaced with their median, while missing categorical variables were labelled as missing. We used one-hot encoding and all variables were normalized using the standard score.

3.4 WAITLIST SURVIVAL

Waitlist (WL) survival was defined as the time from activation on the waitlist to 3 competing risks and censoring defined as:

1. death,
2. KT,
3. too sick to transplant,
4. or censorship for the end of the study or removal from the waitlist for any other reason.

Only features that had at least 90% coverage were kept, unless those features were deemed clinically relevant in previous literature and in the SRTR models. Missing continuous variables were replaced with their median, while missing categorical variables were labelled as

missing. Bae et al. showed that the inclusion of time-varying variables that affect survival, including those used in this study, are sufficient to assume that the association between these features and survival are consistent regardless of when an offer is received [18]. We used one-hot encoding and all variables were normalized using the standard score.

3.5 DEEPHIT

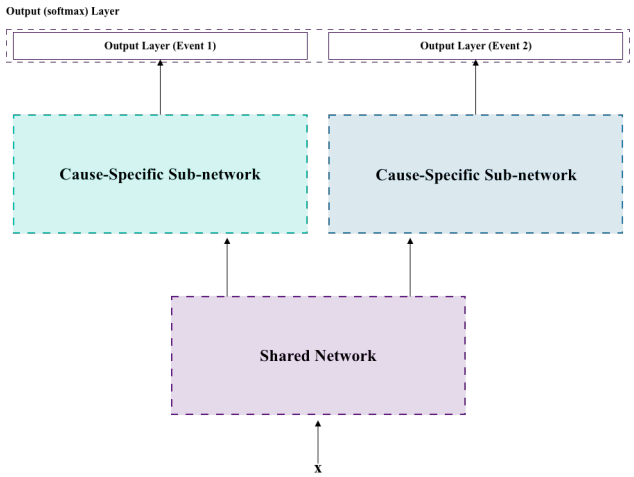

Figure 3.1: DeepHit Architecture.

We use a modified version of DeepHit developed by Lee et al. [26] based on code available on github.[1] This deep learning architecture (Figure 3.1) has been shown to enhance performance in the presence of competing risks. The first network is shared between events and is followed by cause-specific sub-networks for each competing event. Two loss functions ($\mathcal{L}_{total} = \mathcal{L}_1 + \mathcal{L}_2$) are utilized:

$$\mathcal{L}_1 = -\sum_{i=1}^{N}[\mathbb{1}(k^{(i)} \neq \varnothing) \cdot log(y^{(i)}_{k^{(i)},s^{(i)}}) + \mathbb{1}(k^{(i)} \neq \varnothing) \cdot log(1 - \sum_{k=1}^{K} \hat{F}_k(s^{(i)}|x^{(i)}))] \tag{3.1}$$

$$\mathcal{L}_2 = \sum_{k=1}^{K} \alpha_k \cdot \sum_{i \neq j} A_{k,i,j} \cdot \eta(\hat{F}_k(s^{(i)}|x^{(i)}), \hat{F}_k(s^{(i)}|x^{(j)})) \tag{3.2}$$

Where,

- $\hat{F}_{k^*}$ is the estimated CIF for event k^* at time s^*, $\hat{F}_{k^*}(s^*|x^*) = \sum_{m=0}^{s^*} y^*_{k,m}$,

[1] https://github.com/chl8856/DeepHit

- $A_{k,i,j}$ represents acceptable pairs (i,j) that experience event k at a different time and can thus be compared, $A_{k,i,j} \triangleq \mathbb{1}(k^{(i)} = k, s^{(i)} < s(j))$,
- α_k is chosen to trade off ranking losses of the k-th event,
- $\eta(x, y)$ is a convex loss function, $\eta(x, y) = exp(\frac{-(x-y)}{\sigma})$.

$\mathcal{L}_1$ is the log-likelihood of the joint distribution of the first hitting time and event that takes into account censoring. $\mathcal{L}_2$ learns the general representation for the joint distribution of the first hitting time and events using ranking loss with cause-specific CIFs helping the network to perform better on time intervals where there are a large number of death events.

3.6 MEASURES OF PERFORMANCE & CALIBRATION

We used a time dependent measure of performance based on the widely used Concordance index (c-index) in survival analysis. The ordinary c-index [27] represents the fraction of all patients that were ordered correctly out of all patients that could of been ordered. It doesn't take into account the change in risk over time and is biased if the censoring distribution is influenced by the input variables [24]. For example, in the case of KT-survival the chance of censoring over the study period is dependent on medical factors at the time of transplant, such as KDPI and EPTS. In essence, a more risky transplant corresponds to a higher chance of reduced survival and thus lower chance of censorship in any survival study. The time-dependent c-index (C(t)) adjusted for censoring [28] and is defined by:

$$C(t) = \mathbb{P}(R(t|\mathbf{x}_i) > R(t|\mathbf{x}_j)|\Delta_i = 1, T_i \leq t, T_i < T_j). \tag{3.3}$$

To measure calibration we used the Brier Score (BS) [28]:

$$BS(t) = \mathbb{E}[1(T_i \leq t) - R(t|\mathbf{x}_i))^2]. \tag{3.4}$$

3.7 RISK BENEFIT

Risk benefit of KT compared to staying on the WL was calculated by taking the difference between risks. Only the following risk differences were calculated:

1. KT death vs WL death,
2. KT death vs becoming too sick,

3. KT graft failure vs WL death,

4. KT graft failure vs becoming too sick.

The ultimate trade-off comes down to KT death vs WL death. The other measures, represent more granular trade-offs. For example, suppose a patient has a lower risk of death by taking a KT but a higher risk compared to becoming too sick. In this case, the patient could stay on the WL and wait for a better kidney since their chances of being removed from the WL are lower than death post-KT. This is one of the benefits of modeling competing risks.

CHAPTER 4: RESULTS

4.1 POPULATION CHARACTERISTICS

In the recipient cohort (n = 114,261), 33.9% were African American, the median age was 54 years old, 27.3% had diabetes as the primary diagnosis, median time on dialysis was 3.7 years, and 14.5% had previous transplants. The recipients received organs from donors of whom 14.3% were African American, the median age was 41, 7.2% had a history of diabetes, 27.8% had a history of hypertension and 33.2% donated after cerebrovascular/stroke death. The median KDPI and EPTS was 44 and 41, respectively. In the candidate group (n = 217,278), the median age was 55, 31.6% were African American, they had a little bit higher rates of hypertension (23.9%) and diabetes (36.1%) compared to those in the candidate group, 26.1% had a functional status of 80%, and the median EPTS score was 36 (Table 4.1).

Table 4.1: Population Characteristics.

	Recipient Cohort(n = 114,261)	**Candidate Cohort(n = 217,278)**
Recipient/Candidate Factors		
Age at transplant, y	54 (44-63)	
Age at waitlisting, y		55 (45-63)
Race, %		
White	41.2%	44.9%
African American	33.9%	31.6%
Hispanic/Latino	16.3%	16.5%
Other/multiracial	8.6%	7.0%
Primary Diagnosis, %		
Diabetes	27.3%	36.1%
Hypertension	25.7%	23.9%
Glomerular	11.4%	8.6%
Cystic	7.6%	6.2%
Congenital	0.7%	0.5%
Other	26.8%	24.0%
Missing	0.5%	0.6%
Total Albumin, g/dL	3.9 (3.6-4.2)	3.9 (3.6-4.2)
Education Level, %		
High School	42.3%	41.6%
College/Technical School	22.7%	22.9%
Undergraduate Degree	13.9%	14.1%
Unknown	8.4%	7.5%
Peripheral Vascular Disease	6.5%	7.5%
Time on Dialysis, y	3.7 (2.0-5.9)	0.0 (0.0-1.2)
Primary Insurance, %		
Medicare FFS	51.5%	37.6%
Private	22.2%	37%

Table 4.1 Continued

Medicare & Choice	19.4%	14.9%
EPTS, %	41 (18-71)	36 (17-58)
Previous Transplant, %	14.5%	
Time on the Waitlist, y	2.1 (0.8-3.6)	
Cold Ischemic Time, hrs	16.5 (11.4-22.3)	
Functional Status, %		
80%[1]		26.1%
90%[2]		23.5%
70%[3]		19.4%
100%[4]		12.1%
Donor Factors		
Age, y	41 (26-52)	
Race, %		
White	82.2%	
African American	14.3%	
Other/Multiracial	3.5%	
Weight, kg	78.8 (66-93)	
Height, cm	171.0 (163.0-179.0)	
Diabetes, %	7.2%	
Hypertension, %	27.8%	
Serum Creatinine, mg/dL	0.95 (0.7-1.3)	
Hepatitis C virus infection, %	2.7%	
Cause of Death, %		
Head Trauma	36.0%	
Cerebrovascular/Stroke	33.2%	
Anoxia	27.5%	
Shared Transplant, %	25.8%	
KDPI, %	44 (21-68)	

Table 4.2: Competing Risks Summary.

	Percent	**Event/Censoring Time (y)**
Post-KT Survival		
Censored	73.2%	4.5 (2.0-7.7)
Graft Failure	13.9%	2.4 (0.5-4.8)
Death	12.8%	3.0 (0.9-5.4)
WL Survival		
KT	50.0%	2.6 (1.2-4.4)
Death	20.1%	2.5 (1.2-4.2)
Censored	17.1%	3.1 (1.7-5.1)
Too sick to transplant	12.7%	3.4 (1.9-5.3)

Values are median(IQR) for event/censoring time.

In the recipient cohort, 13.9% experienced graft failure while 12.8% died. The median

[1]Normal Activity with effort: some symptoms of disease
[2]Able to carry on normal activity: minor symptoms of disease
[3]Cares for self: unable to carry on normal activity or active work
[4]Normal, no complaints, no evidence of disease

(IQR) times until graft failure or death was 2.4 (0.5-4.8) years and 3.0 (0.9-5.4) years, respectively. In the candidate cohort, 50.0% of patients received a KT, 20.1% died on the waiting list, and 12.7% were removed from the waiting list after becoming too sick to transplant. The median (IQR) times until the events were 2.6 (1.2-4.4) years, 2.5 (1.2-4.2) years, and 3.4 (1.9-5.3) years, respectively (Table 4.2). Figures 4.1 and 4.2 show the 1-KM and CIF curves for each event. The CIF for post-kT outcomes was almost equivalent to the 1-KM curve, albeit the 1-KM curve consistently has higher risk. Meanwhile, the CIF for WL outcomes was drastically more accurate than the 1-KM curve showing the probability of KT being greater than 3x as likely to occur compared to death or becoming too sick to transplant. Interestingly, death and graft failure post-KT had relatively equivalent risk.

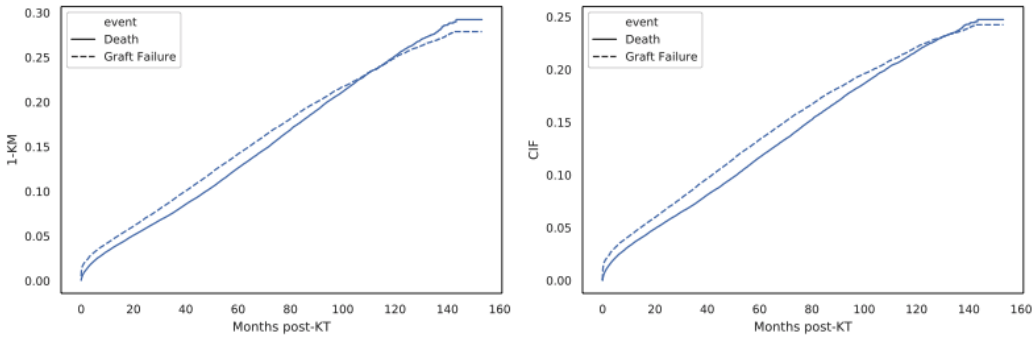

Figure 4.1: Distribution function 1-KM and cumulative incidence function of competing events post-KT.

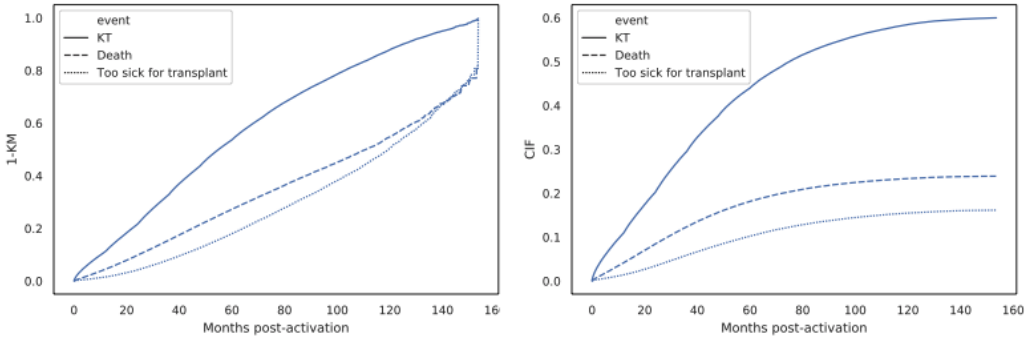

Figure 4.2: Distribution function 1-KM and cumulative incidence function of competing events on the waiting list.

4.2 POST-KT SURVIVAL

For post-KT survival the DeepHit model achieved a 5-year c-cindex of 0.70 and 0.64 for death and graft failure, respectively. In comparison, KDRI is known to have a c-index of 0.62, while several other methods achieve similar performance to this model going up to 0.724 [29, 30]. LYFTs reported performance for graft survival 0.61 [31]. The 5-year brier score was 0.46 for both death and graft failure (Figure 4.3).

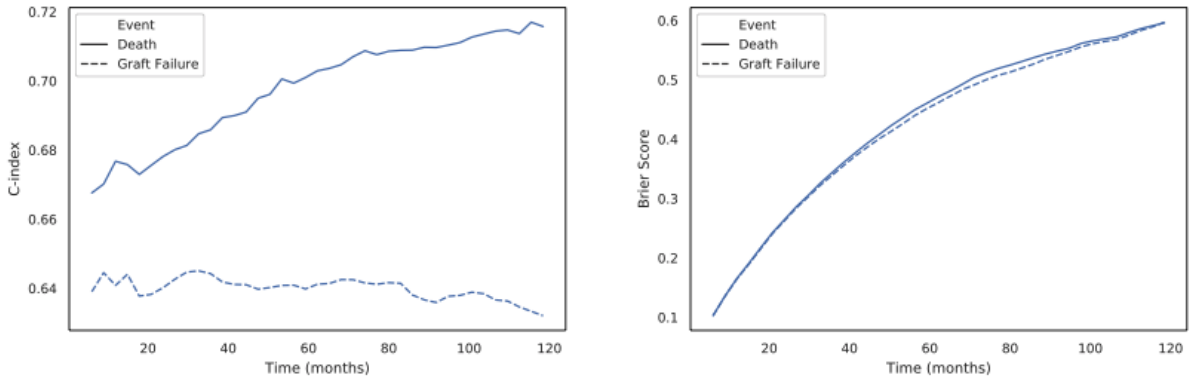

Figure 4.3: Time dependent C-index and Brier Score for the KT survival model.

The risk of graft failure was slightly greater than death at all time points. 5- and 10-year risk of death was 11.5% and 13.1%, respectively. While the 5- and 10-year risk of graft failure was 19.3% and 20.1%, respectively. Interestingly, patients that died had greater accelerated risk of death at all time points compared to patients that experienced graft failure while the opposite was observed in patients that experienced graft failure. 5- and 10-year risk of death was 16.3% and 27.1% for patients that died compared to 11.1% and 18.4% for patients that experienced graft failure. 5- and 10-year risk for graft failure was 13.1% and 20.6% for patients that died compared to 15.3% and 22.8% for patients that experienced graft failure (Figure 4.4).

EPTS showed a strong correlation with 3-, 5-, and 10-year risk of death, while showing relatively no correlation with 3-, 5- and 10-year risk of graft failure. KDPI on the other hand showed a stronger correlation with 3-, 5-, and 10-year risk of graft failure compared to death (Figures 4.5 and 4.6). Clinically this means that EPTS is more indicative of recipient condition, while KDPI is more indicative of graft condition in the future. This is consistent with the fact that EPTS and KDPI are derived from recipient and donor variables, respectively. Patients with low EPTS ($\leq$25th percentile) showed a much lower risk of death compared to patients with a high EPTS ($\geq$75th percentile). Interestingly, the risk of graft

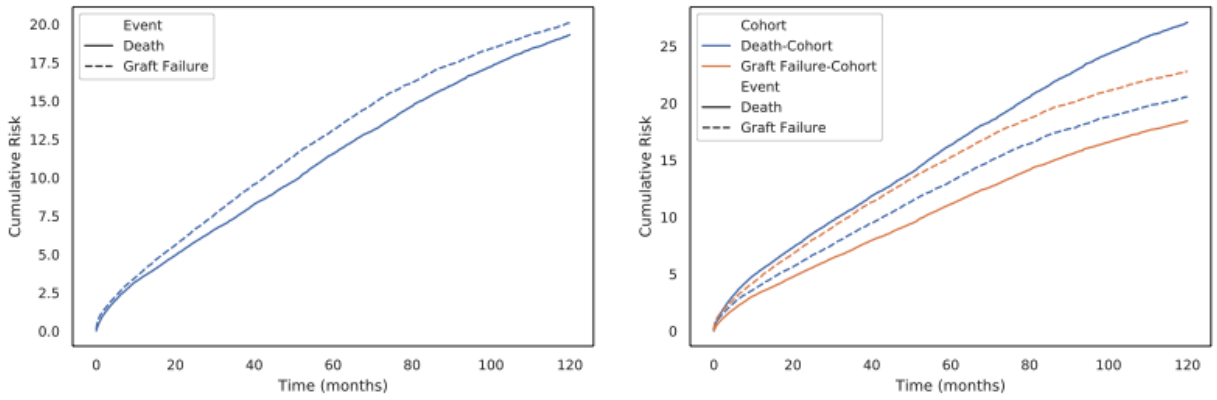

Figure 4.4: Median risk for all patients in the test set and for all patients in the test set seperated by event observed.

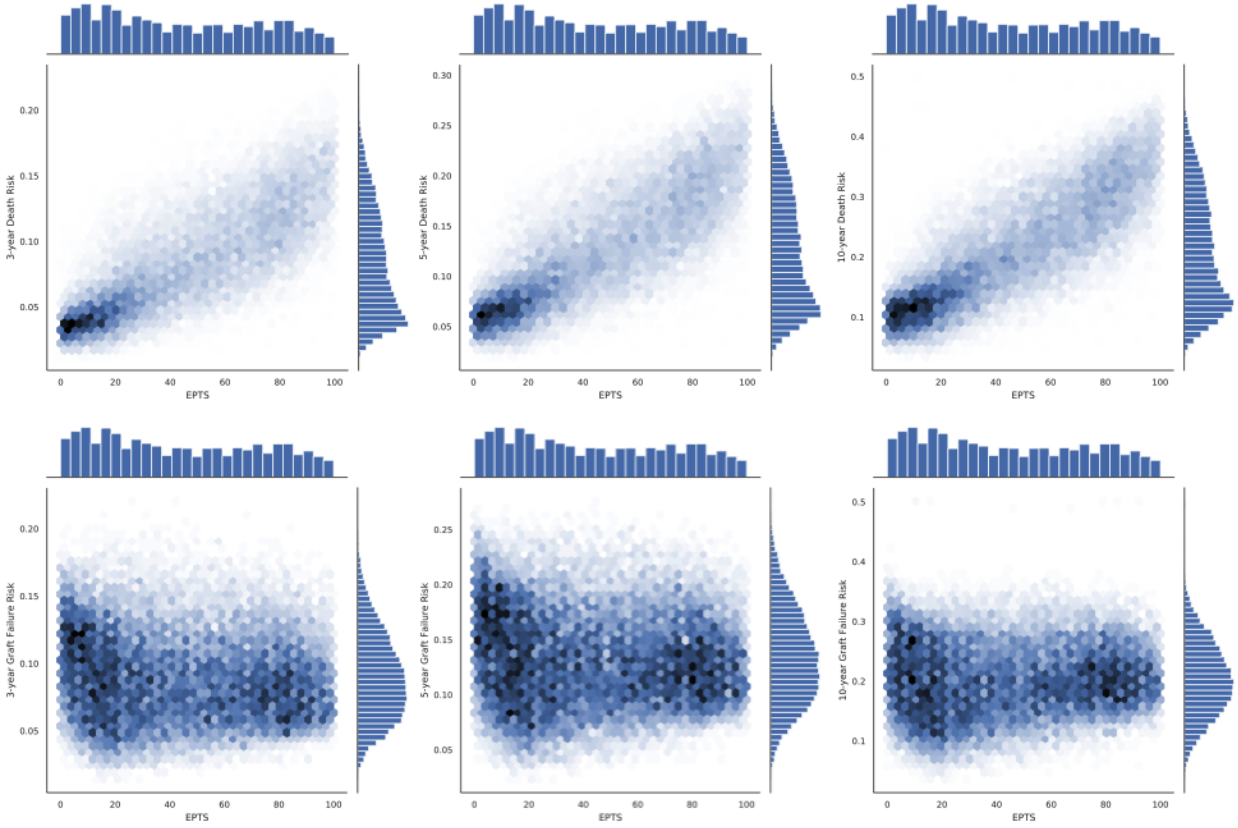

Figure 4.5: 3-, 5-, and 10-year post-KT risk and EPTS at the time of KT. Death is shown on the top row and graft failure on the bottom with their respective years post-KT.

failure was relatively the same with low EPTS patients having marginally higher risk for graft failure (<2%) from 20 to 90 months post-KT. Patients with low KDPI (≤25th percentile)

showed a much lower risk for graft failure and death post-KT (Figure 4.7).

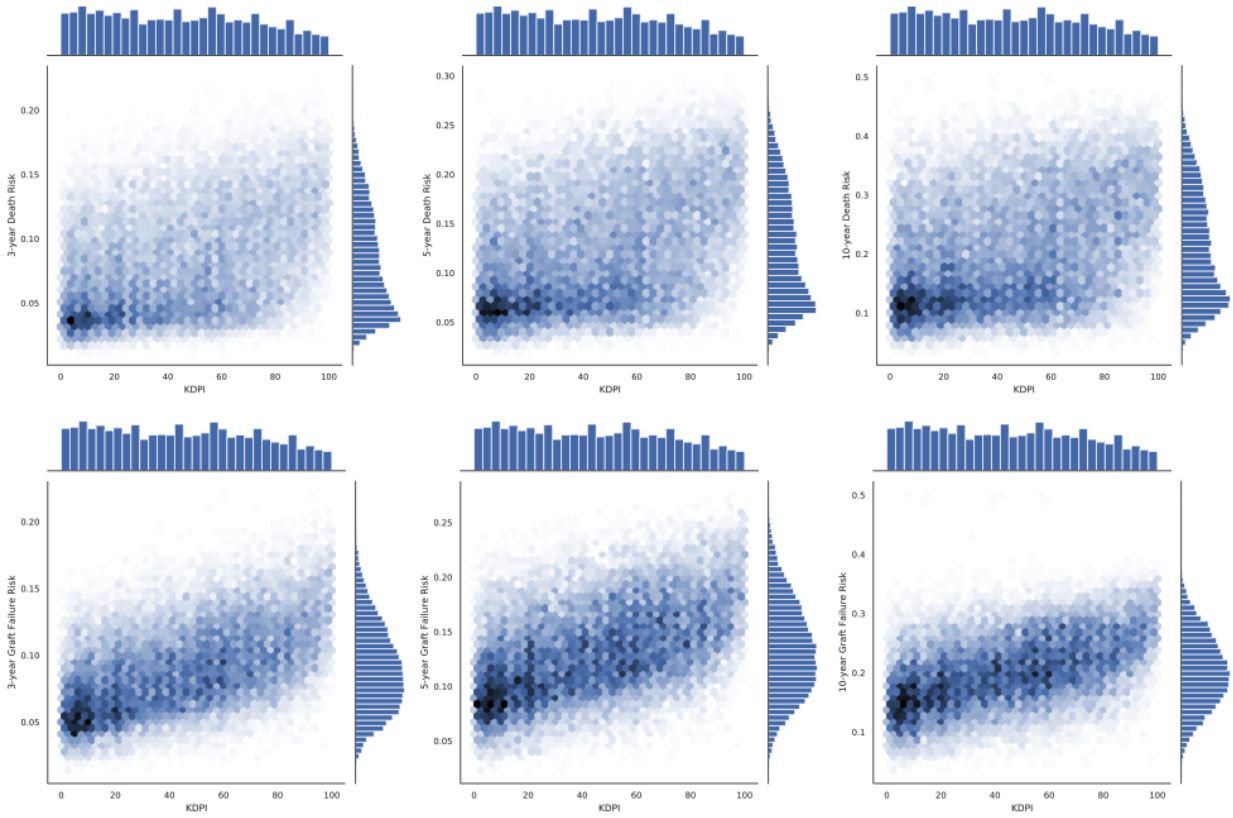

Figure 4.6: 3-, 5-, and 10-year post-KT risk and KDPI at the time of KT. Death is shown on the top row and graft failure on the bottom with their respective years post-KT.

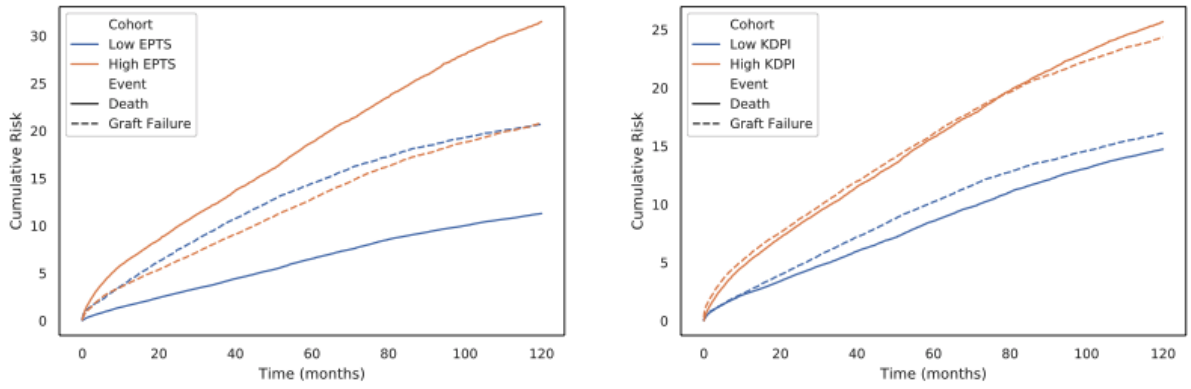

Figure 4.7: Median risk for all patients in the test set seperated by EPTS and KDPI (25th and 75th percentiles) observed.

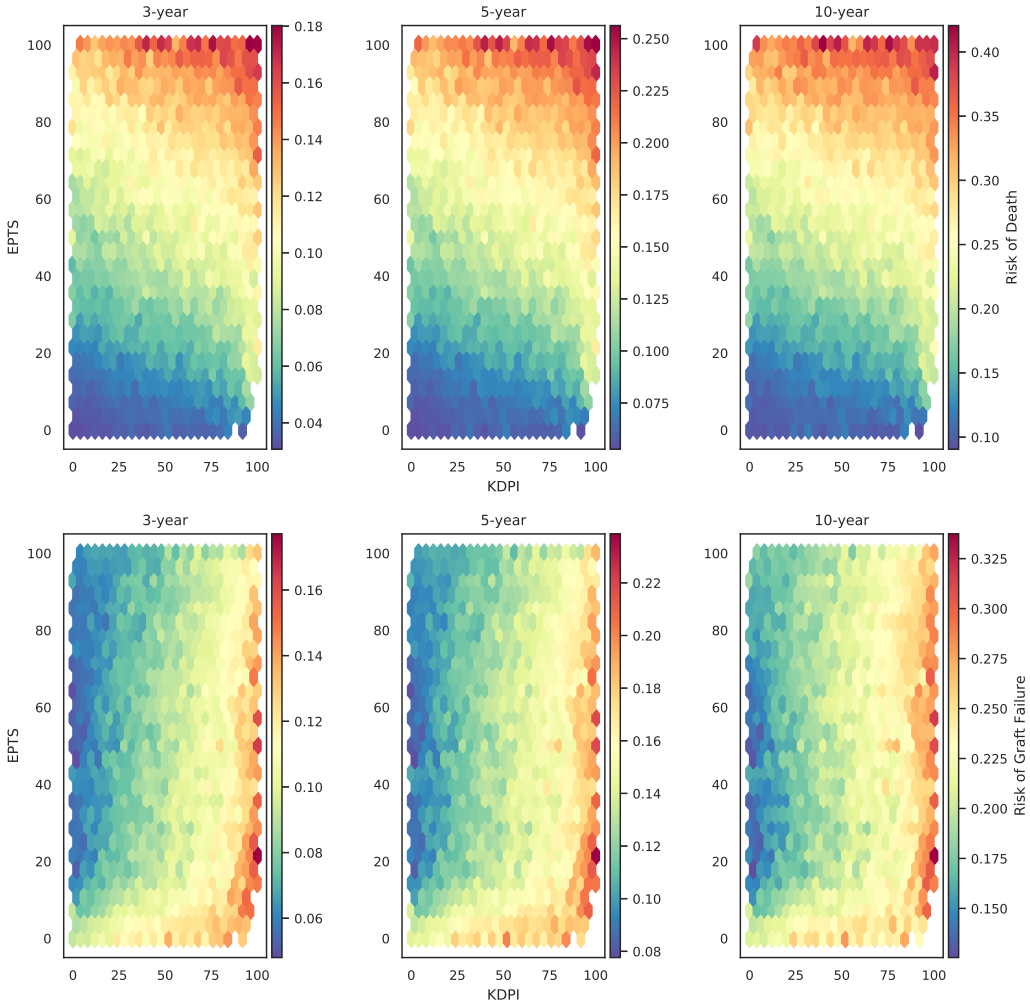

Figure 4.8: EPTS, KDPI and Risk at 3-, 5-, and 10-year post-KT risk for death and graft failure observed in the test set.

Combining EPTS and KDPI showed a linear trend with risk of death, wherein a low EPTS recipient receiving a low KDPI kidney showed much lower risk of death at 3-, 5- and 10- years post-KT, compared to the opposite. At 10-years post-KT the risk difference for death was up to 30% as KDPI and EPTS increase. The risk of graft failure on the other hand showed a strong trend with KDPI alone, corroborating with results previously discussed. At 10-years post-KT the risk difference graft failure was up to 20% as KDPI increased. Interestingly, very low EPTS (<10% EPTS) recipients were found to have higher risks of graft failure with KDPI donors up to 75%, however this EPTS score range was outside the IQR and the support for this edge case is likely very low (Figure 4.8).

4.3 WAITLIST SURVIVAL

For WL survival the DeepHit model achieved a 3-year c-cindex of 0.75, 0.72, and 0.77 for KT, death, and becoming too sick, respectively. In comparison, the LYFT score has a reported c-index of 0.68 (0-15 years) [31]. The 3-year brier score was 0.34, 0.39 and 0.46 for KT, death, and becoming too sick, respectively (Figure 4.9).

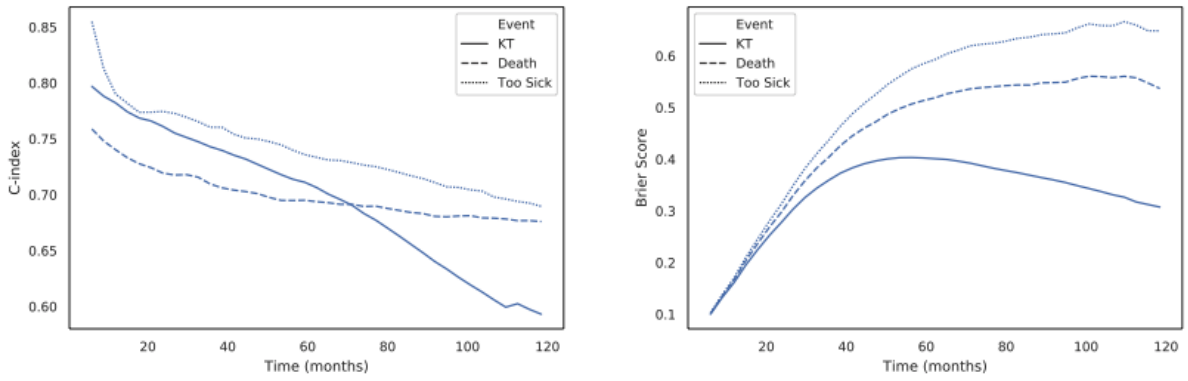

Figure 4.9: Time dependent C-index and Brier Score for the WL set.

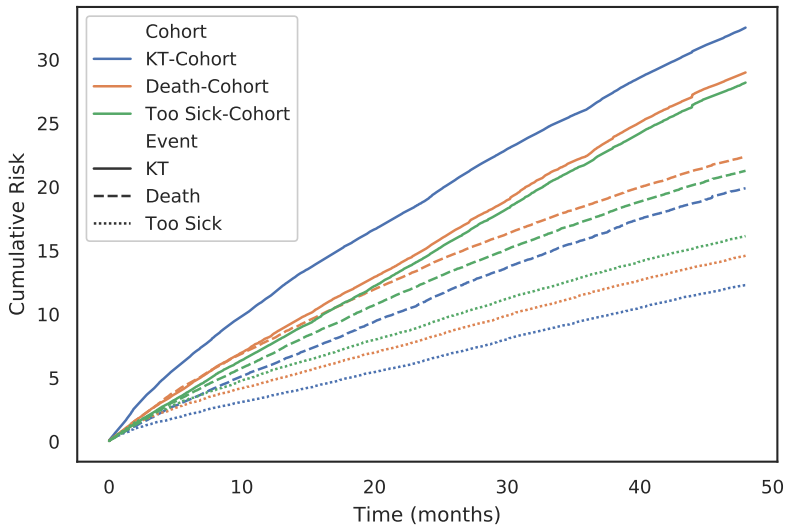

Figure 4.10: Median risk for all patients in the test set seperated by event cohort.

Patients that died or became too sick on the WL had greater risk for death or becoming too sick respectively. Patients that had a KT had a much greater chance of receiving a KT than any other event. Interestingly, in the near term (<12 months) the risk of death for

patients that died was greater than KT. On a longer time scale the chance of receiving a KT is greater than any other event. Clinically it makes sense that if a patient survives long enough he/she will most likely have a KT. For patients that underwent a KT, the median risk of KT, death, and becoming too sick was 19.0%, 11.1% and 6.3%, respectively at 2 years, and 32.4%, 19.9%, and 12.2% respectively at 4 years. For patients that died on the WL, the median risk of KT, death, and becoming too sick was 15.2%, 13.8% and 8.1%, respectively at 2 years, and 29.0%, 22.4%, and 14.6% respectively at 4 years. For patients that become to sick to transplant, the median risk of KT, death, and becoming too sick was 14.6%, 12.6% and 9.2%, respectively at 2 years, and 28.2%, 21.2%, and 16.1% respectively at 4 years (Figure 4.10). Modifying a patient to have a higher EPTS (>80) raised their risk of death and becoming too sick drastically. Again the pattern of risk of death being greater than KT in the short term emerges (Figure 4.11).

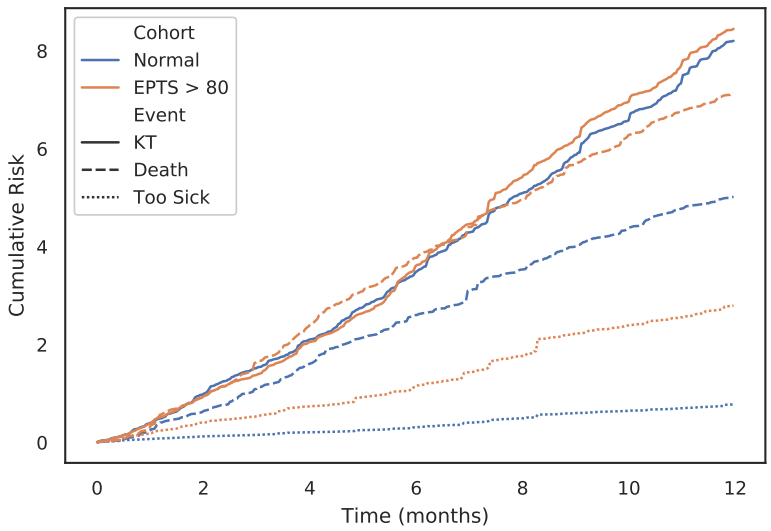

Figure 4.11: Risk for a patient that was transplanted with modified EPTS.

4.4 DECISION ANALYSIS

1-, 3-, and 5-year risk benefit of KT for graft failure vs death on the WL showed that higher EPTS patients benefit the most from a KT while very low EPTS patients can safely remain on the WL. This is also true in comparing graft failure with becoming too sick to transplant. The risk benefit at 3-years ranges from -5% to +12.5% for death on WL and -10% to +10% for becoming too sick to transplant. 1-, 3-, and 5-year risk benefit of KT for

death post-KT vs death on the WL showed that essentially every patient would do better with a KT. This is important because it corroborates the pre-emptive KT strategy put forth by clinicians, however only 20% of KTs are performed pre-emptively.[5] Ultimately, 3-year risk benefit ranged from 0% to +8% going up to +10% at 5-years post-KT. Even for low EPTS patients their risk of death is improved with KT, while their risk of graft failure is not, meaning the patient would improve their survival simply by continually undergoing a KT each time their graft fails rather than remaining on dialysis. Similar findings between death post-KT vs becoming too sick were found. although the risk benefits were comparatively smaller ranging from -6% to +4% at 5 years (Figure 4.12).

[5]https://www.mayoclinic.org/tests-procedures/pre-emptive-kidney-transplant/pyc-20384830

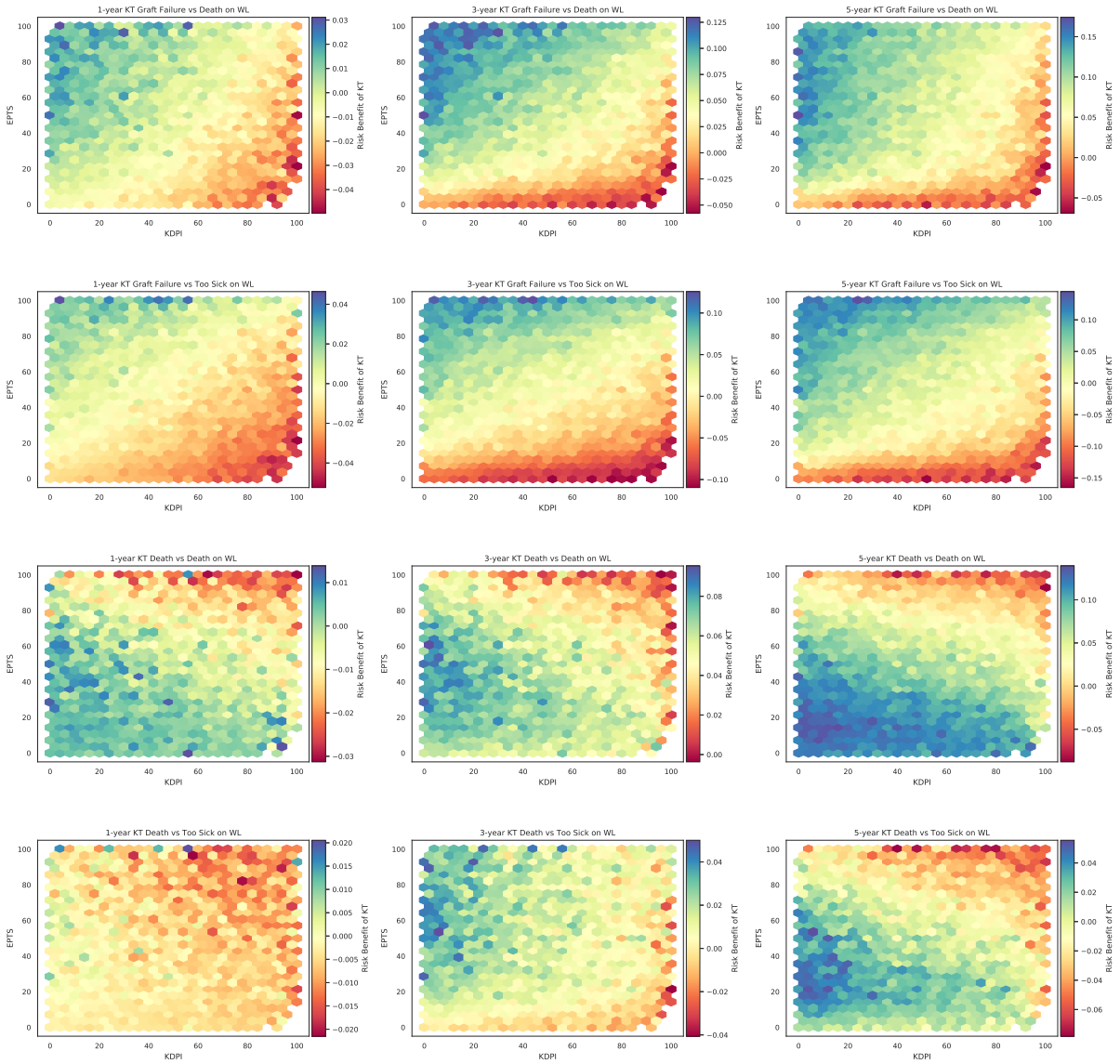

Figure 4.12: EPTS, KDPI and risk benefit of KT at 1-, 3-, and 5-years.

CHAPTER 5: CONCLUSION AND FUTURE WORK

5.1 CLINICAL SIGNIFICANCE

Ultimately, this hypothetical CDS system achieved comparable if not better performance, while also considering a more robust competing event framework. In particular, the treatment of death and graft failure as competing events proved to capture valuable clinical information affecting factor 1 (survival with KT). The WL model proved to delineate patients with higher EPTS scores and patients that died or became to sick on the waitlist, effectively addressing factor 2 (survival rejecting KT). While this model includes KT as a WL outcome it doesn't utilize any location data, other than a candidates permanent state. To effectively address factor 3 (survival opportunity in the future), more granular data such as local organ supply and geographical allocation factors are needed. However, it was shown that for the majority of patients a KT with any KDPI improved their chances of survival suggesting that a pre-emtpive transplant is a feasible strategy for patients with ESRD. In this case the only factor to consider is maximizing graft survival. High KDPI kidneys seemed to have a marginal effect on 3-year patient survival post-KT vs remaining on the waitlist, event though it does significantly increase the risk of death and graft failure post-KT. Meanwhile, low KDPI kidneys seem to have the greatest effect on patients with mid tier EPTS. Bea et al. had similar conclusions, however they did not consider graft survival as a competing event to death [18]. For graft failure, in low EPTS patients the risk benefit was lower with higher KDPI kidneys suggesting that waiting for a better kidney would reduce the chance of re-transplantation, whereas high EPTS patients would still receive a positive risk benefit with higher KDPI kidneys. In practice this would suggest that low EPTS patients can wait longer in order to improve graft survival and patient survival , while high EPTS patients would benefit by being transplanted as soon as possible while having almost no effect on patient survival.

In the future, this CDS system could be simplified further by adding another model to analyze the probability distribution of the competing events and reducing the final prediction to an accept or decline decision. It was shown that for patients that died their risk of death was greater than any other risk in the short term, suggesting that this pattern would be highly indicative of a high risk patient that could potentially benefit from any organ offer, while low risk patients seemed to have a higher chance of KT at all time points suggesting they have time to wait until graft survival is maximized.

5.2 IMPLICATIONS FOR CDS SYSTEMS IN PRODUCTION

CDS systems still have a ways to go before widespread adoption. Luckily, the attention in this field is great with work being done by even the largest companies, such as Google[1] and Microsoft[2]. An interesting example in kidney transplant is OmniLife, a Small Business Innovation Research (SBIR) funded (among other private investment) startup working in the transplant space with a unique forward thinking solution for kidney transplant offer and patient management. During a successful SBIR phase I project, OmniLife proved the feasibility of improving donor management and coordination using a streamlined communication application, TXP Chat. The key TXP Chat innovation is in its user-focused design and application specific awareness to the transplant continuum. TXP Chat replaces the slow, point-to-point call centers reliant on phones, faxes and paper forms (Figure 1.1) with more advanced technology enabling communication across institutional boundaries (Figure 5.1). TXP Chat allowed clinicians to communicate more effectively and with higher user interface satisfaction, thus facilitating more efficient decision making.

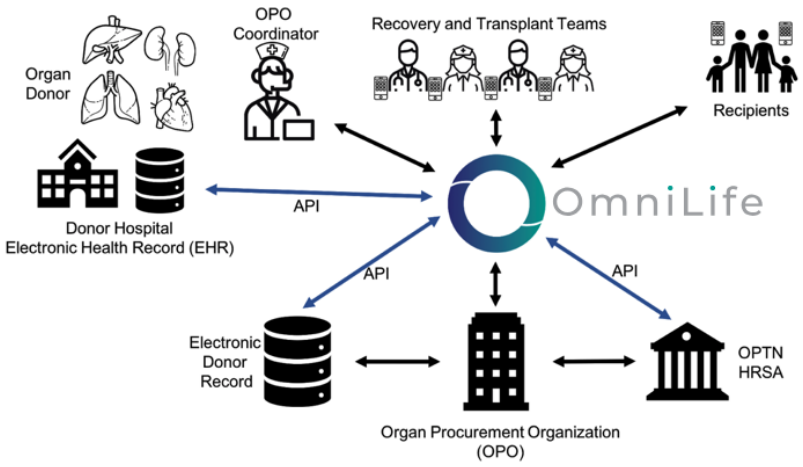

Figure 5.1: OmniLife Simplifies the complex communication network and allows rapid exchange of information at the time of offer. A clinical decision support system on their system would be accessible by all parties involved on a transplant case.

In a proposed SBIR phase II research direction, TXP Chat will be enhanced with the development of a clinical decision support (CDS) system called Ask Alan that will be capable of delivering predictive insights at the time of offer from powerful machine learning models (Figure 5.2). Ask Alans predictive capabilities will be improved upon by further integrating

[1]https://ai.google/healthcare/
[2]https://www.microsoft.com/en-us/research/research-area/medical-health-genomics/

the system with powerful data sources inherent in the transplant professionals workflow, such as the electronic medical record (EMR) (Figure 5.3).

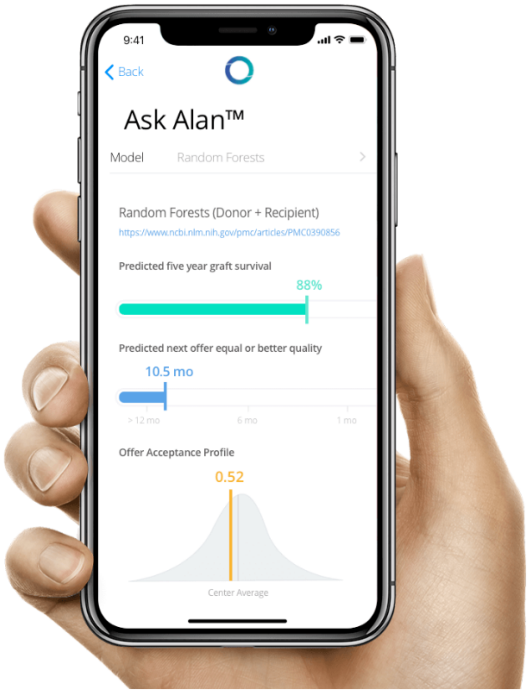

Figure 5.2: The user interface for a clinical decision support system supporting multiple clinical predictive models.

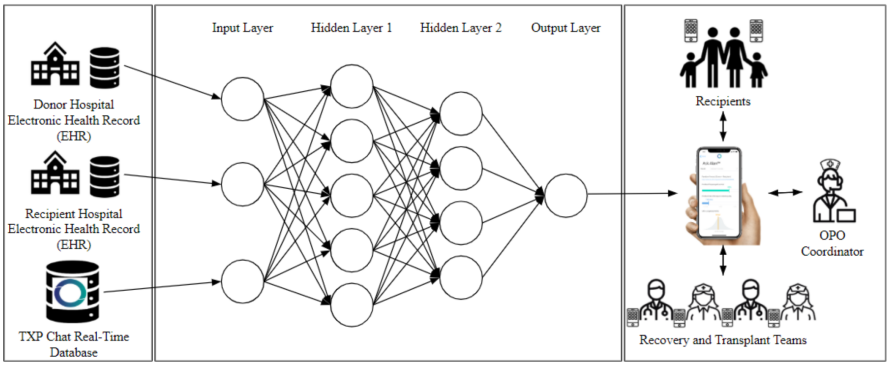

Figure 5.3: Integration with local EMR and Omnilife databases would allow for more granular and possibly accurate predictions.

The current opinion among clinicians in organ transplantation is that predictive models trained with national datasets cannot ensure transportability to local environments [32]. This is a very valid argument and attempts have been made to alleviate this issue [33, 34].

However, powerful local databases exist within every hospital and even in patient smartphones. In kidney transplantation, technology like TXP Chat represent a unique, secure access point for both predictive models and local data sources, not only removing barriers for researchers but also for hospitals and medical institutions who might not have the proper software infrastructure to support CDS systems across institutional boundaries. Other related solutions exist such as doc.ai[3] utilizing blockchain technology to democratize medical research. In any case, for the majority of models published by researchers to have clinical meaningfulness they need to demonstrate reproduciblity in local environments, and local environments need to have the ability to implement and support the model in production.

Besides the infrastructure considerations necessary to support a system like this in production, another area that should be explored is in the realm of Human Computer Interaction (HCI). Representing complex predictions in a manner that can confidently influence clinical decision making is not a trivial task [35]. While statistically one can ensure the accuracy of a prediction, one cannot so easily ensure interpretability especially in a rapid paced clinical environment where patient lives are at stake. Normally, to measure the effectiveness of a new software feature a company can employ A/B testing, but Narayanan et al. found that in machine learning these types of studies don't lead to generalized conclusions about what properties are most essential in certain contexts [35]. Therefore, more careful studies need to be done with clinicians to generate guidelines as to what properties lead to increased trust and usability, aside from performance. As noted in [36], "designing an algorithm for implementation in the clinic involves matter well beyond algorithmic performance" and they conclude that interpretability is the first towards trust. In the context of kidney transplantation, this work was able to show the clinical interpretability of predictions made by the deep learning model and any attempt to use these predictions in clinical practice would likely require similar interpretability with every organ offer.

[3]https://doc.ai/

REFERENCES

[1] R. P. Pauly, "Survival comparison between intensive hemodialysis and transplantation in the context of the existing literature surrounding nocturnal and short-daily hemodialysis," *Nephrology Dialysis Transplantation*, vol. 28, no. 1, pp. 44–47, 2012.

[2] T. S. Purnell, P. Auguste, D. C. Crews, J. Lamprea-Montealegre, T. Olufade, R. Greer, P. Ephraim, J. Sheu, D. Kostecki, N. R. Powe et al., "Comparison of life participation activities among adults treated by hemodialysis, peritoneal dialysis, and kidney transplantation: a systematic review," *American Journal of Kidney Diseases*, vol. 62, no. 5, pp. 953–973, 2013.

[3] K. D. Yoo, C. T. Kim, M.-H. Kim, J. Noh, G. Kim, H. Kim, J. N. An, J. Y. Park, H. Cho, K. H. Kim et al., "Superior outcomes of kidney transplantation compared with dialysis: An optimal matched analysis of a national population-based cohort study between 2005 and 2008 in korea," *Medicine*, vol. 95, no. 33, 2016.

[4] U. S. R. D. System, "2015 usrds annual data report: Epidemiology of kidney disease in the united states," 2015.

[5] D. E. Stewart, V. C. Garcia, J. D. Rosendale, D. K. Klassen, and B. J. Carrico, "Diagnosing the decades-long rise in the deceased donor kidney discard rate in the united states," *Transplantation*, vol. 101, no. 3, pp. 575–587, 2017.

[6] A. B. Massie, X. Luo, E. Chow, J. Alejo, N. M. Desai, and D. Segev, "Survival benefit of primary deceased donor transplantation with high-kdpi kidneys," *American Journal of Transplantation*, vol. 14, no. 10, pp. 2310–2316, 2014.

[7] S. Mohan, M. C. Chiles, R. E. Patzer, S. O. Pastan, S. A. Husain, D. J. Carpenter, G. K. Dube, R. J. Crew, L. E. Ratner, and D. J. Cohen, "Factors leading to the discard of deceased donor kidneys in the united states," *Kidney international*, vol. 94, no. 1, pp. 187–198, 2018.

[8] S. Mohan, K. Foley, M. C. Chiles, G. K. Dube, R. E. Patzer, S. O. Pastan, R. J. Crew, D. J. Cohen, and L. E. Ratner, "The weekend effect alters the procurement and discard rates of deceased donor kidneys in the united states," *Kidney international*, vol. 90, no. 1, pp. 157–163, 2016.

[9] A. Wey, N. Salkowski, W. K. Kremers, C. R. Schaffhausen, B. L. Kasiske, A. K. Israni, and J. J. Snyder, "A kidney offer acceptance decision tool to inform the decision to accept an offer or wait for a better kidney," *American Journal of Transplantation*, vol. 18, no. 4, pp. 897–906, 2018.

[10] D. Bertsimas, J. Kung, N. Trichakis, D. Wojciechowski, and P. A. Vagefi, "Accept or decline? an analytics-based decision tool for kidney offer evaluation," *Transplantation*, vol. 101, no. 12, pp. 2898–2904, 2017.

[11] E. L. Kaplan and P. Meier, "Nonparametric estimation from incomplete observations," *Journal of the American statistical association*, vol. 53, no. 282, pp. 457–481, 1958.

[12] D. R. Cox, "Regression models and life-tables," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 34, no. 2, pp. 187–202, 1972.

[13] J. P. Fine and R. J. Gray, "A proportional hazards model for the subdistribution of a competing risk," *Journal of the American statistical association*, vol. 94, no. 446, pp. 496–509, 1999.

[14] R. L. Prentice, J. D. Kalbfleisch, A. V. Peterson Jr, N. Flournoy, V. T. Farewell, and N. E. Breslow, "The analysis of failure times in the presence of competing risks," *Biometrics*, pp. 541–554, 1978.

[15] R. Sapir-Pichhadze, M. Pintilie, K. Tinckam, A. Laupacis, A. Logan, J. Beyene, and S. Kim, "Survival analysis in the presence of competing risks: the example of waitlisted kidney transplant candidates," *American Journal of Transplantation*, vol. 16, no. 7, pp. 1958–1966, 2016.

[16] "Organ procurement and transplantation network." [Online]. Available: https://optn.transplant.hrsa.gov/resources/guidance/kidney-donor-profile-index-kdpi-guide-for-clinicians/

[17] "Organ procurement and transplantation network." [Online]. Available: https://optn.transplant.hrsa.gov/resources/allocation-calculators/epts-calculator/

[18] S. Bae, A. B. Massie, A. G. Thomas, G. Bahn, X. Luo, K. R. Jackson, S. E. Ottmann, D. C. Brennan, N. M. Desai, J. Coresh et al., "Who can tolerate a marginal kidney? predicting survival after deceased donor kidney transplant by donor–recipient combination," *American Journal of Transplantation*, vol. 19, no. 2, pp. 425–433, 2019.

[19] A. Decruyenaere, P. Decruyenaere, P. Peeters, F. Vermassen, T. Dhaene, and I. Couckuyt, "Prediction of delayed graft function after kidney transplantation: comparison between logistic regression and machine learning methods," *BMC medical informatics and decision making*, vol. 15, no. 1, p. 83, 2015.

[20] S. Krikov, A. Khan, B. C. Baird, L. L. Barenbaum, A. Leviatov, J. K. Koford, and A. S. Goldfarb-Rumyantzev, "Predicting kidney transplant survival using tree-based modeling," *Asaio Journal*, vol. 53, no. 5, pp. 592–600, 2007.

[21] M. Fouad, M. M. A. Ellatif, M. Hagag, and A. Akl, "Prediction of long term living donor kidney graft outcome: Comparison between rule based decision tree and linear regression," *IJARCST.*, vol. 3, pp. 185–192, 2015.

[22] N. Petrovsky, S. K. Tam, V. Brusic, G. Russ, L. Socha, and V. B. Bajic, "Use of artificial neural networks in improving renal transplantation outcomes," *Graft*, vol. 5, no. 1, p. 6, 2002.

[23] J. Li, G. Serpen, S. Selman, M. Franchetti, M. Riesen, and C. Schneider, "Bayes net classifiers for prediction of renal graft status and survival period," *World Academy of Science, Engineering and Technology*, vol. 39, 2010.

[24] T. A. Gerds, M. W. Kattan, M. Schumacher, and C. Yu, "Estimating a time-dependent concordance index for survival prediction models with covariate dependent censoring," *Statistics in Medicine*, vol. 32, no. 13, pp. 2173–2184, 2013.

[25] A. B. Massie, L. Kuricka, and D. L. Segev, "Big data in organ transplantation: registries and administrative claims," *American Journal of Transplantation*, vol. 14, no. 8, pp. 1723–1730, 2014.

[26] C. Lee, W. R. Zame, J. Yoon, and M. van der Schaar, "Deephit: A deep learning approach to survival analysis with competing risks," in *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[27] F. E. Harrell, R. M. Califf, D. B. Pryor, K. L. Lee, and R. A. Rosati, "Evaluating the yield of medical tests," *Jama*, vol. 247, no. 18, pp. 2543–2546, 1982.

[28] C. Lee, W. R. Zame, A. M. Alaa, and M. van der Schaar, "Temporal quilting for survival analysis."

[29] Y. Zhong, D. E. Schaubel, J. D. Kalbfleisch, V. B. Ashby, P. S. Rao, and R. S. Sung, "Reevaluation of the kidney donor risk index (kdri)." *Transplantation*, 2018.

[30] E. Mark, D. Goldsman, B. Gurbaxani, P. Keskinocak, and J. Sokol, "Using machine learning and an ensemble of methods to predict kidney transplant survival," *PloS one*, vol. 14, no. 1, p. e0209068, 2019.

[31] R. Wolfe, K. McCullough, and A. Leichtman, "Predictability of survival models for waiting list and transplant patients: calculating lyft," *American Journal of Transplantation*, vol. 9, no. 7, pp. 1523–1527, 2009.

[32] C. to the C4 Article (Appendix 1), "Current opinions in organ allocation," *American Journal of Transplantation*, vol. 18, no. 11, pp. 2625–2634, 2018.

[33] A. Subbaswamy, P. Schulam, and S. Saria, "Learning predictive models that transport," *arXiv preprint arXiv:1812.04597*, 2018.

[34] M. Sugiyama, M. Krauledat, and K.-R. MÃžller, "Covariate shift adaptation by importance weighted cross validation," *Journal of Machine Learning Research*, vol. 8, no. May, pp. 985–1005, 2007.

[35] M. Narayanan, E. Chen, J. He, B. Kim, S. Gershman, and F. Doshi-Velez, "How do humans understand explanations from machine learning systems? an evaluation of the human-interpretability of explanation," *CoRR*, vol. abs/1802.00682, 2018. [Online]. Available: http://arxiv.org/abs/1802.00682

[36] "Towards trustable machine learning," *Nature Biomedical Engineering*, vol. 2, no. 10, pp. 709–710, Oct 2018.