EFFECTS OF LONGITUDINAL PROFILES OF PEDIATRIC END-STAGE LIVER DISEASE (PELD) SCORES ON POST-TRANSPLANT SURVIVAL

by

Zhaowen Sun

B.S in Nursing, Wuhan University, Wuhan, China, 2010

Submitted to the Graduate Faculty of

the Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Master of Science

University of Pittsburgh

2012

UNIVERSITY OF PITTSBURGH

GRADUATE SCHOOL OF PUBLIC HEALTH

This thesis was presented

by

Zhaowen Sun

It was defended on

August 9, 2012

and approved by

Thesis advisor:Chung-Chou H. Chang, Ph.D, Associate Professor, Departments of Medicine and Biostatistics, School of Medicine and Graduate School of Public Health, University of Pittsburgh

Francis Pike, Ph.D, Assistant Professor, Departments of Critical Care Medicine and Biostatistics, School of Medicine and Graduate School of Public Health, University of Pittsburgh

Ada O. Youk, Ph.D, Assistant Professor, Departments of Biostatistics, Epidemiology and Clinical & Translational Medicine, Graduate School of Public Health and School of Medicine, University of Pittsburgh

EFFECTS OF LONGITUGINAL PROFILES OF PEDIATRIC END-STAGE LIVER DISEASE (PELD) SCORES ON POST-TRANSPLANT SURVIVAL

Zhaowen Sun, M.S.

University of Pittsburgh, 2012

Liver transplant has long been the ultimate treatment for patients with end-stage liver disease. For pediatric patients under 12 years old, the Pediatric End-Stage Liver Disease (PELD) scoring system has been developed and implemented for almost 10 years as an estimation of 90-day mortality pre-transplantation. PELD score has been used as an appropriate tool to discriminate patients when allocating organs, yet as a potential predictor of post-transplant graft failure, its effect has been somewhat inconsistent. In previous studies, researchers tended to use single measured PELD score at the time of listing or at transplant in predicting graft survival, while in clinical practice PELD scores are often calculated multiple times between listing and transplant to monitor patients' disease status. To make the most use of the information of the repeated PELD measures, we propose to employ latent group-based trajectory models to explore the underlying distinct patterns of the PELD scores and then integrate such longitudinal profiles into survival analysis to examine its effects as a predictor of graft survival.

Public health significance: This study will provide both the medical and public health communities with important information that will facilitate their evidence-based practice concerning pediatric patients in need of liver transplant. Researchers will get to see the big picture regarding this health issue since national data have been used to derive the results. They will have a better idea of patients' disease status and prognosis and will be able to make treatment or intervention decisions with more confidence.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

This work can never be done without the guidance of my advisor, Dr. Chung-Chou H. Chang, who led me into this fascinating field. I also wish to thank Drs. Francis Pike and Ada O. Youk for being on my thesis committee and Ms. Yi Ren for her work regarding data cleaning. As always, to my family and friends, it’s with their non-stopping love and support that I’ve been able to come this far.

1.0 INTRODUCTION

Liver transplantation has long been the life-saving and the ultimate treatment method in patients with end-stage liver diseases. [1] Unfortunately the available organ source has always been greatly outnumbered by patients in need of transplant and among which pediatric liver recipients are greatly outnumbered by adult recipients (roughly 1 to 17). [2] To address the imbalance of resources and needs in liver transplant, numerous organ allocation systems have been implemented over the years, such as the Child-Turcotte-Pugh (CTP) score. [3] Starting in February 27, 2002, the Pediatric End-Stage Liver Disease (PELD) scoring system has been developed and promoted by the United Network for Organ Sharing (UNOS) in an attempt to create a more objective and effective organ allocation paradigm. Unlike the CTP score, which divides patients into different status categories and assigns different priorities on the waiting list, the PELD score is a continuous scale of liver disease severity and provides the assessment of 3-month pre-transplant mortality in pediatric patients under 12 years old. The PELD score is calculated using patient's age, serum total albumin, serum total bilirubin, international normalized ratio (INR) for prothrombin time along with presence/absence of growth failure: [4]

PELD score = 0.436 (Age <1 years old) + 0.687 log(albumin g/dL) + 0.480 log(total bilirubin mg/dL) + 1.87 log(INR) + 0.667 (growth failure).

The main goal of the PELD system is to optimize transplant benefit, which means maximal benefit (usually measured by additional life time gained from transplant) and minimal harm (measured by mortality on the waiting list). As a preoperative measure, PELD score is an indicator of pre-transplant mortality by its definition yet its potential as a predictor of post-transplant outcomes such as graft survival, length of hospital stay (LOS), length of intensive care unit stay (LOS ICU), and post-transplant complications has not been fully realized. Several studies have been carried out to explore the relationship between the PELD score and post-transplant survival where mixed results were seen.

Cowles et al. (2008) reviewed 71 pediatric biliary atresia cases that received liver transplant between January 1998 and December 2006 and showed a 14% graft failure rate in 53 kids with PELD score greater or equal to 10 versus a graft failure rate of 0% in 18 kids with PELD score less than 10. PELD score can be used as an informative predictor of graft failure and a PELD score of 10 can act as a potential cutoff. [5]

Farmer et al. (2009) studied 122 children with 159 liver transplants from 1984 to 2008. Their analysis showed that without considering other covariates, PELD score over 25 was associated with the worst 5-year and 10-year graft survival. After adjusting for other covariates, the higher PELD score was still significantly associated with worse graft survival along with other measures like creatinine clearance. [6] Their conclusions were supported by Oh et al. (2010) in a retrospective, single-centered study with 113 children where a PELD score over 25 was significantly associated with graft loss. [7]

On the other hand, null results have also been detected by other researchers. Caroll et al. (2003), after analyzing 67 pediatric liver transplant cases in Chicago dated from August 1997 to February 2002, concluded that PELD score was only associated with post-transplant LOS but not

associated with 1-year post-transplant mortality or ICU LOS. [8] A similar conclusion was reached by Freeman et al. (2005) where they did not detect significant 1-year graft survival results across different PELD strata. [9]

It should also be noted that in past studies, investigators tended to use single measured PELD score (at time of listing or at the time of transplant) as the predictor of the outcomes of interest. While in the study carried out by Bourdeux et al. (2005), they took both the PELD score at pre-transplant assessment and the PELD score at transplant and calculated a change score from the two. Their results showed that neither PELD measure taken individually was significantly associated with post-transplant survival whereas a higher change score in PELD is associated with lower post-transplant survival. [10]

In this study, we propose to investigate whether there is any association between the trajectory of the PELD scores and post-transplant survival. For one patient, the trajectory of PELD scores actually reflects the change in his or her disease status overtime and the latent trends, or the patterns in PELD scores overtime may contain additional information in predicting patients' post-transplant survival. We aim to explore the latent patterns of the PELD scores and their significance in predicting post-transplant survival. We also compared the prediction ability of the listing PELD, PELD at the time of transplant, change in PELD score from listing to transplant, and the trajectories of the PELD scores on the post-transplant survival.

In Section 2, we describe the data source that is used for this study and introduce the latent group-based trajectory model. We will describe how we used this method to identify different trajectories of the PELD scores among transplant recipients. In Section 3, we show the results of our findings and conclude in Section 4.

2.0 METHODS

2.1 DATA

The data used were derived from the United Network for Organ Sharing (UNOS) Standard Transplant Analysis and Research (STAR) data file for the liver transplant waiting list. Our analysis was restricted to pediatric liver transplant recipients, who were 12 years old or younger at the time of listing, who started on the waiting list on or after February 27, 2002 and received a liver transplant on or before June, 2010. The dataset does not contain patients who eventually received a multi-organ transplant. Patients who started or ended with a Model for the End-Stage Liver Disease (MELD) score measure(s) were excluded because the calculations and the ranges of PELD and MELD scores are different. This ended to 4,445 patients.

All the patients had PELD score at the time of listing and at the time of transplant. The number of PELD score measurements in between the time of listing and the time of transplant varied from 1 to 80. In the trajectory analysis of repeated PELD measurements, we only kept the measurements within 90 days prior to transplant. We further excluded patients who had less than 2 repeated PELD measurements. The final analytic dataset contains 1,745 transplant recipients with 9,237 total PELD measurements (Figure 1).

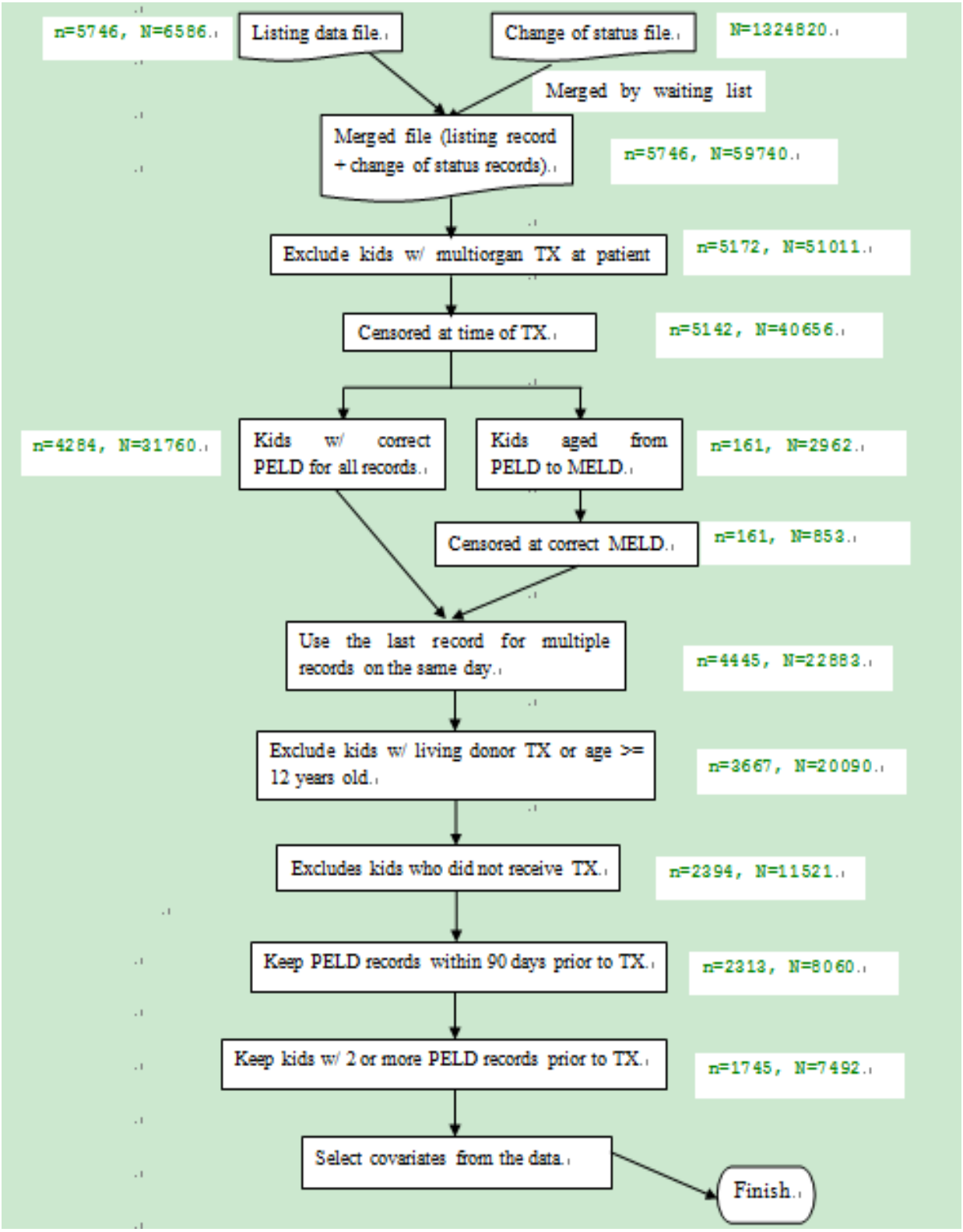


Figure 1 Flowchart of UNOS PELD data set

In addition to the PELD scores, covariates under consideration in the analyses included the demographic information of recipients and donors (age, gender, race/ethnicity, and ABO blood type), recipient's post-transplant outcomes (death, re-transplant, or re-transplant-free survival), and recipient's lab measures at the time of listing (albumin level, total bilirubin level, international normalized ratio [INR], and serum creatinine level).

2.2 STATISTICAL ANALYSIS

2.2.1 Latent group based trajectory analysis

Latent group-based trajectory analysis has been widely used in areas including developmental/abnormal psychology, sociology, and criminology to describe and monitor an individual's psychological status over time and to predict an individual's possibility of relapsing back into criminal behaviors. [11] It uses a mixture model approach to categorize subjects with similar longitudinal patterns into the same group. The group-based modeling approach assumes that the population is composed of a mixture of distinct behavior groups and can be defined by its respective trajectory models. [12] In medicine, a population may not be literally made up of several distinct groups, yet there may be some specific patterns or important cutoff or ranges that can contribute to decision making in clinical practice once identified. We adopted this approach to categorize liver transplant recipients into several distinct PELD trajectory pattern groups.

Let $Y_i = \{y_{i1}, y_{i2}, y_{i3}, \ldots, y_{iT}\}$ represent the longitudinal sequence of measurements on an individual i over T time points and let $P(Y_i = y_i)$ denote the probability of observing y_i. The group-based trajectory model assumes that the population is composed of a mixture of J underlying trajectory groups such that:

$$P(Yi) = \sum_{j} \pi_j \, P^j \, (Y_i)$$

,

where $P^j\,(Y_i)$ denotes the unconditional probability of observing $Y_i = y_i$ given subject *i* is in group *j*; and π_j is the group membership probability, probability of subject *i* belongs to group *j*. Since group membership is neither arbitrarily assigned nor directly observed, group membership probabilities for group *j*, π_j, *j* =1, …., *J*, are parameters of interest and are estimated by a multinomial logit function:

$$\pi_j = e^{\theta_j} \Big/ \sum_{1}^{J} e^{\theta_j},$$

where θ_j is normalized to zero thus each estimation of π_j will fall between 0 and 1. The form of $P^{jt}\,(Y_{it})$ is decided based on the type of data to be analyzed, in this study, the PELD score measurement is assumed to follow a normal distribution and the linkage between time from listing to transplant and PELD score is established by means of a latent variable:

$$y_{it}^{*j} = \beta_0^j + \beta_1^j time_{it} + \beta_2^j time_{it}^2 + \beta_3^j time_{it}^3 + \varepsilon_{it}.$$

The current software, SAS Proc TRAJ will allow us to fit a model with up to a cubic term, where ε_{it} is a disturbance assumed to be normally distributed with a mean of 0 and a fixed but unknown variance of σ^2.[13]

To select the best trajectory model, we first needed to decide the optimal number of distinct trajectory groups and then to determine the appropriate functional form for each trajectory. Maximum likelihood was used for the estimation of model parameters and Wald tests were performed to test whether trajectory parameters and generalized logit parameters differ

significantly from zero. The Bayesian information criterion (BIC) was used in model selection. For a given model, the BIC was calculated as follows:

$$\mathbf{BIC} = \log(L) - \mathbf{0.5} * \mathbf{log}(n) * (k),$$

where L is the value of the maximum likelihood corresponding to the model; n is the sample size; and k is the number of parameters under estimation.

We fit the trajectory model starting with a basic one-group model, then the two-group model, and so forth until a universal maximum likelihood had been reached. The final number of distinct trajectory groups was determined by the Bayes factor B_{ij}, which was defined by Schwartz and Kass, Wasserman with the form:

$$\mathbf{Bayes\ Factor} \approx \mathbf{e}^{BIC_i - BIC_j}.$$

The Bayes factor measures the odds of each of the two competing models being the correct model. It is computed as the ratio of the probability of *i* being the correct model to *j* being the correct model. For example, a Bayes factor of 10 implies that model *i* is 10 times more likely than *j*. The interpretation of the Bayes factor is summarized in Table 1.2. [11]

Besides the statistical evidence provided using the methods above, we also incorporated clinical knowledge and kept the rule of parsimony in mind in model selection. SAS Proc TRAJ was used to carry out the group-based modeling. The package was first developed by Dr. Bobby Jones from the Carnegie Mellon University in 2001 and the last updated in December 2010. This package and references can be downloaded from the site http://www.andrew.cmu.edu/user/bjones/index.html.[14]

2.2.2 Survival analysis of graft failure rates

Our outcome of interest, post-transplant graft failure was defined as the time from transplant to either death or re-transplant. Our goal was to predict this outcome using a specific trajectory pattern of the PELD scores.

We used the Kaplan-Meier method to estimate the unadjusted function of graft failure for all and for each PELD trajectory group. The log-log survival plot and the Grambch-Therneau test were used to test the proportional hazards (PH) assumption. If the PH assumption was not violated, a log-rank test was used to compare the functions of graft failure among different PELD trajectory groups. If the PH assumption was violated, weighted log-rank tests (Peto-Peto, Tarone-Ware, and generalized Wilcoxon test) were performed to compare the functions.

We compared the ability in predicting graft failure for the following 3 predictors: the PELD score trajectory patterns, the PELD score at the time of listing, and the PELD score at the time of transplant. Three Cox PH regression models for each of the predictors were fit. We used the Akaike information criterion (AIC) and the c-statistic to evaluate the prediction performance. All above-mentioned analyses were redone for each of the six disease diagnoses: biliary atresia, acute liver failure, autoimmune disorder, cancer, metabolic disorder, and other liver disease. All survival-related analyses were carried out using the package of STATA version 11.

Data used in this study, taken from the UNOS STAR registry, are possibly the best dataset available thus far since it contains data across the United States after the PELD scoring system has been implemented until June 2010. The only thing that might impair the statistical power is the results of predicting graft failure using PELD trajectory patterns. It is highly probable that the sizes of trajectory groups will vary greatly and a sparse group will arise. When subjects fall into PELD trajectory groups with a vastly uneven sample sizes and then is carried

over to the subsequent survival analysis of graft failure, such imbalance and sparseness will possibly lead to inflated standard error of an estimate thus resulting in inaccurate coefficient estimates. This is an additional criterion we considered in choosing the “best” number of PELD trajectory patterns in the first part of this study.

3.0 RESULTS

3.1 Longitudinal Profiles of the PELD Scores

PELD scores in the data set ranged from -11 to 77. The best trajectory pattern was selected with an effort to balance the complexity of the model and optimal likelihood. An eight latent-group trajectory model was chosen based on the BIC factors (Table 1, 2) [15]. However, this eight-group model contains a sparse group with only 11 subjects, which may affect the power of subsequent survival analyses. It is also worth noting that in many situations; statistically significantly differences among patients might not be clinically significant. To address the above concerns, we chose the seven-group trajectory model as the longitudinal profiles of the PELD scores and carried them over into the subsequent survival analysis phase.

Except for intercept terms, all of the coefficients of the linear and quadratic terms in the trajectory model were positive with a considerably small absolute value. This indicates that there was only a slightly increasing trend in the PELD scores over the 90-day period prior to transplantation. The final seven-group trajectory pattern had orders of 0, 2, 1, 1, 1, 2, and 2 (0=constant, 1=linear, 2=quadratic). Group 1 had the lowest PELD scores overtime; group 2 has the second lowest PELD scores over time, and so forth. Group1 had a constant trajectory pattern with a mean around -3. Group 2, although had a quadratic pattern, seemed to be clustered around 0 with significant but minimal slope and quadratic coefficients. Groups 3, 4, and 5 all had linear trajectory patterns with mean PELD scores of 12, 20, and 26, respectively. The two groups with the highest PELD scores over time (groups 6 and 7) appeared to have much more fluctuating trajectories (Table 3 and Figure 2).

Table 1 Summary of group-based trajectory analysis (n=1,745)

# of groups	BIC (N= 9237)	Orders	Group Membership (%)	Group Membership (n)
1	-37755.60	2	100	1,745
2	-33660.14	0	41.52	724
		1	58.48	1,021
3	-32225.85	3	33.63	587
		1	38.43	671
		1	27.94	487
4	-31530.11	0	30.72	536
		2	24.82	433
		1	30.93	540
		2	13.52	236
5	-31186.15	2	23.99	418
		0	15.95	279
		2	28.72	501
		1	24.71	431
		2	6.63	116
6	-31012.44	0	23.74	414
		1	16.01	279
		2	28.05	489
		1	24.11	421
		1	7.60	133
		2	0.49	9

Table 1 Summary of group-based trajectory analysis (n=1,745), continued

7	-30753.25	0	15.50	270
		2	16.65	291
		1	16.53	288
		1	21.70	379
		1	18.96	331
		2	8.98	157
		2	1.68	29
8	-30591.80	0	15.58	272
		2	16.57	289
		1	16.46	287
		1	21.48	375
		1	18.32	320
		2	6.77	118
		2	4.20	73
		2	0.62	11
9	-30660.04	0	15.60	273
		2	16.60	290
		0	16.10	281
		2	0.63	12
		1	21.65	378
		1	18.42	322
		2	0.14	3
		1	8.98	157
		1	1.58	28

Table 2 Interpretation of Bayes factors

Bayes Factor (B_{ij})	Interpretation
$B_{ij} < 1/10$	Strong evidence for model j
$1/10 < B_{ij} < 1/3$	Moderate evidence for model j
$1/3 < B_{ij} < 1$	Weak evidence for model j
$1 < B_{ij} < 3$	Weak evidence for model i
$3 < B_{ij} < 10$	Moderate evidence for model i
$B_{ij} > 10$	Strong evidence for model i

Table 3 Maximum likelihood estimates for the seven PELD trajectory groups

Group	Parameter	Estimate	Standard error	Prob > |T|
1	intercept	-6.73	0.23	<0.0001
2	intercept	1.08	0.33	0.0012
	linear	0.05	0.02	0.0051
	quadratic	0.00051	0.0002	0.0328
3	intercept	11.82	0.285	<0.0001
	linear	0.037	0.006	<0.0001
4	intercept	20.15	0.266	<0.0001
	linear	0.048	0.0054	<0.0001
5	intercept	25.76	0.26	<0.0001
	linear	0.068	0.0063	<0.0001
6	intercept	37.51	0.46	<0.0001
	linear	0.25	0.033	<0.0001
	quadratic	0.001	0.0004	0.0066
7	intercept	48.5	0.645	<0.0001
	linear	0.24	0.055	<0.0001
	quadratic	0.0015	0.00075d	0.0474

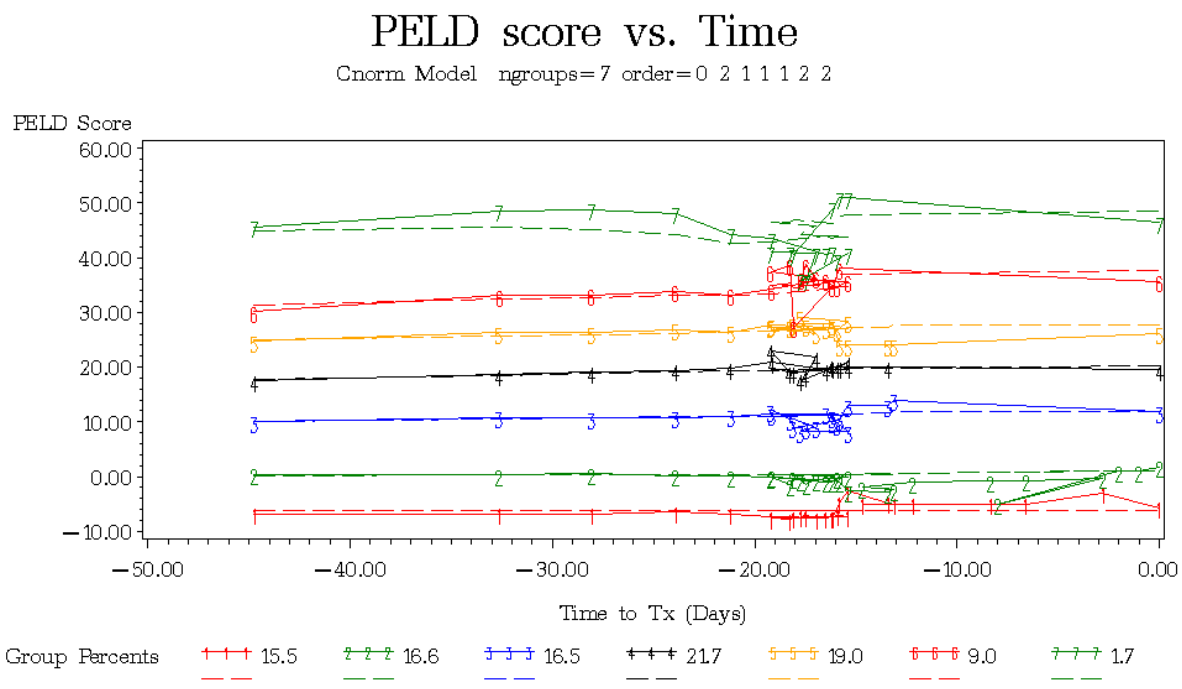


Figure 2 Plots for the seven groups PELD trajectories

3.2 Overall Graft Survival for All Patients

Among the 1,705 pediatric patients with non-missing graft survival information, 313 (18.36%) of them died or underwent re-transplantation with the majority of these events happening within 1-year of transplantation. The 1-year graft failure rate was 13.4% and the 5-year graft failure rate was about 20%. The median follow-up time for all patients was 721 days (about 2 years) and the longest follow-up time was about 8 years (Figure 3).

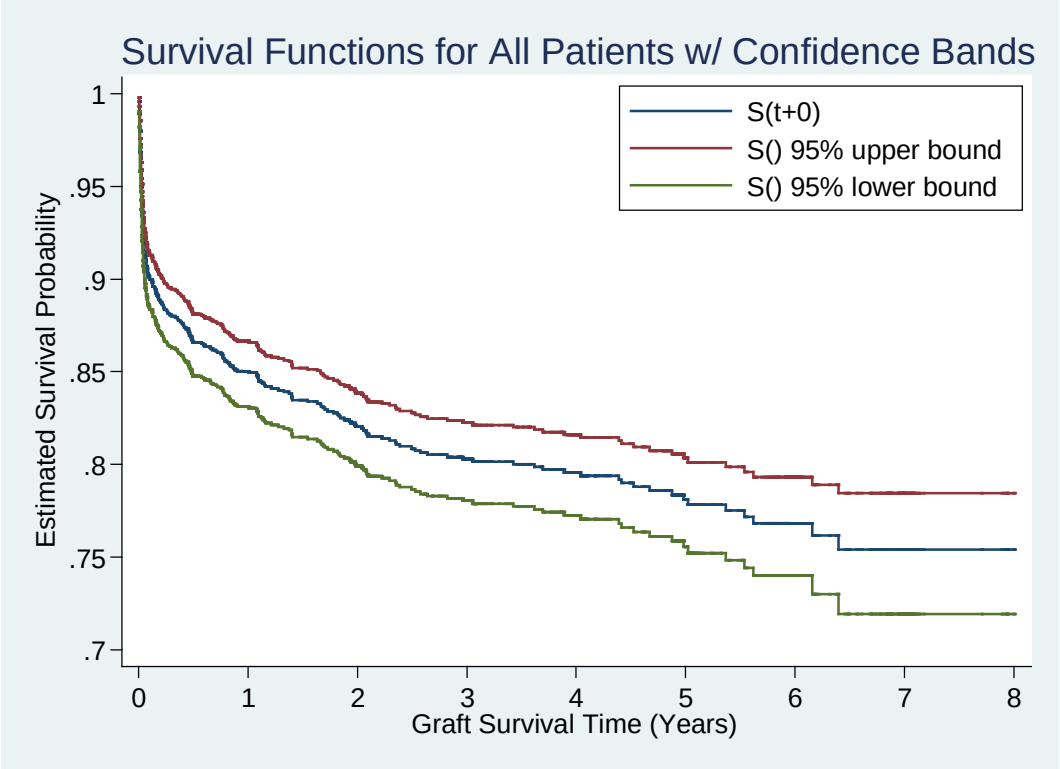


Figure 3 Overall Graft Survival Functions with Confidence Interval

3.3 Graft Survival for Patients with Different Trajectory Patterns

At first glance, patients in group 2 and 3 had the lowest graft failure rates (15.8% and 16.2% respectively) and patients in group 5 and 6 had the highest graft failure rates (23.6% and 20.9%; hazard ratios 1.30 & 1.18, using Group 1 as reference) (Table 4). Based on the results of log-rank test, the survival experiences among trajectory groups did not significantly differ from each other. Because the sequence of trajectory groups was decided by the actual values of the PELD score, a test of trend was carried out to assess whether there was a significant trend for increasing graft failure rate with increasing PELD scores, the p-value associated with the test of trend was still not statistically significant (Table 5).

Table 4 Graft failure rates across different PELD trajectory groups (N=1,705)

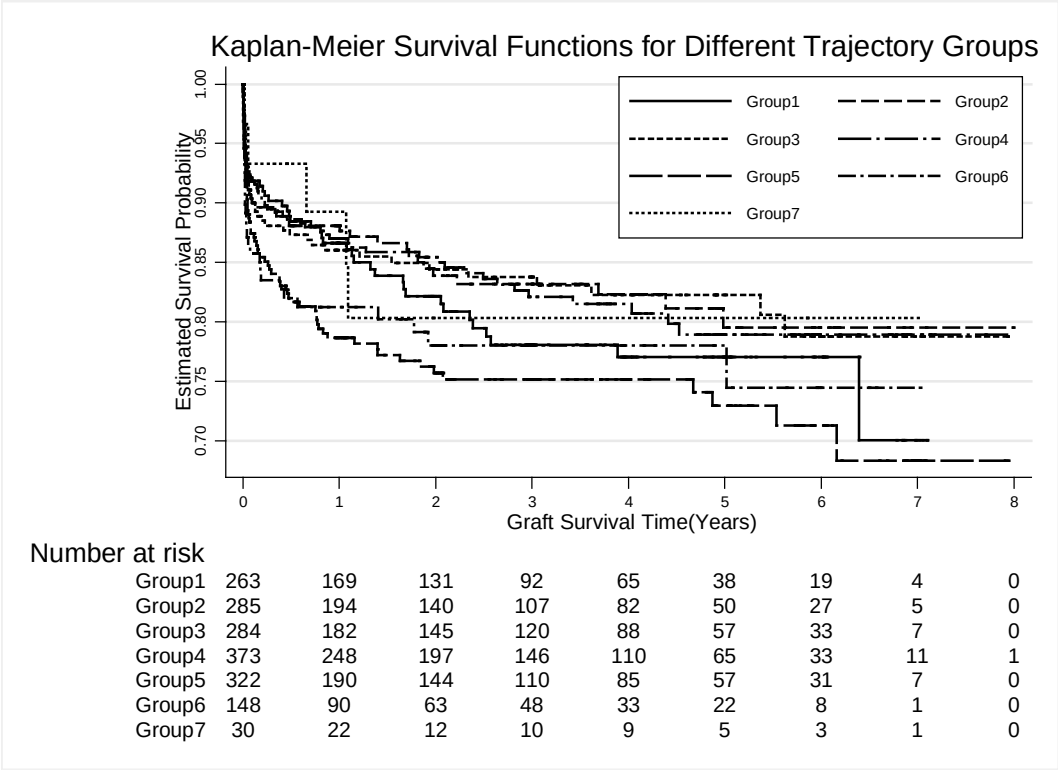
Group	N	%	Average Listing PELD	Average Transplant PELD	Median Follow-up time (Days)	Range of Follow-up Time (min, max)	Censoring (n, %)
1	163	15.43	-7	-5.94	721	2592 (1, 2593)	214, 81.4
2	285	16.72	0.46	1.73	739	2920 (1, 2921)	240, 84.2
3	84	16.66	9.73	11.75	766	2902 (1, 2903)	238, 83.8
4	73	21.88	16.49	19.54	481.5	2923 (1, 2924)	312, 83.7
5	22	18.89	22.92	26.31	642	2899 (1, 2900)	246, 76.4
6	48	8.68	29.53	35.93	695	2569 (1, 2570)	117, 79.1
7	10	1.76	45	46.33	723.5	2573(1, 2574)	25, 83.3

Table 5 Log-rank tests of graft survival functions for different PELD trajectory groups, listing PELD and transplant PELD (n=1,705)

Test	Test of Equality		Test of Trend	
	Chi2	p-value	Chi2(1)	p-value
PELD Trajectory Groups	10.58	0.1021	2.84	0.0918
Listing PELD	11.37	0.0776	--	--
Transplant PELD	10.63	0.1003	--	--

It should be noted that group 1, the group with the lowest PELD scores, did not end up with the highest graft survival rate, which in a way, contradicts the common notion that the lower the PELD score, the better the patient's prognosis. However, patients in group 7, the group with

the highest PELD scores, turned out to have better graft survival than patients in group 5 and 6. Such abnormal outcomes may be the hint that there are some other variables that are complimentary to the PELD-score trajectory patterns in explaining patients' graft survival experiences (Figure 4).



Number at risk

Group1	263	169	131	92	65	38	19	4	0
Group2	285	194	140	107	82	50	27	5	0
Group3	284	182	145	120	88	57	33	7	0
Group4	373	248	197	146	110	65	33	11	1
Group5	322	190	144	110	85	57	31	7	0
Group6	148	90	63	48	33	22	8	1	0
Group7	30	22	12	10	9	5	3	1	0

Figure 4 Graft Survival Functions for Different PELD Trajectory Groups

3.4 Graft Survival for Patients with Different Diagnoses

Etiology has long been regarded as an important predictor of the prognosis for patients with liver transplant. In this sample, we divided patients' primary diagnoses into 6 main groups. About

50% of all patients suffered from biliary atresia. The second most common etiology was metabolic disorder. Other diagnoses included acute liver failure, autoimmune disorder, cancer, and miscellaneous (Table 6).

Table 6 Graft failure rates across different diagnosis groups (N=1,705)

Diagnosis	n	%	Average listing PELD	Average final PELD	Median follow-up time (Days)	Range of follow-up time (min, max)	Censoring (n, %)
110 Acute liver failure	165	9.68	27.42	24.57	695	2571 (1, 2572)	133, 80.6
206 Auto-immune disorder	64	3.75	12.55	13.78	703.5	2917 (7, 2924)	55, 85.9
207 Cancer	164	9.62	3.29	0.92	647	2591 (1, 2592)	118, 72.0
208 Metabolic disorder	245	14.37	3.54	6.12	742	2919 (2, 2921)	203, 82.9
210 Other	203	11.91	11.92	13	721	2898 (1, 2899)	166, 81.8
211 Biliary atresia	864	20.67	14.16	18.3	724	2888 (1, 2889)	717, 83.0

Unlike what we have seen for the PELD-score trajectory groups, cancer patients had higher graft failure rate compared to patients with other diagnoses (graft failure rate = 28% and the median follow-up time = 642 days). For other diagnoses, the graft failure rates ranged from 14.1% to 19.4% (mean = 17.3%) and the median follow-up time excluding cancer patients was 722 days (Table 7). The Kaplan-Meier survival curves for all groups excluding cancer patients seem to separate slightly during the first two years post-transplantation and then joined each other closely afterwards. We can also see that patients with acute liver failure had the second

worst outcomes: there was a sudden drop in survival probability during the starting period of follow-up and the survival probability remained stable afterwards. Patients with autoimmune disorder and biliary atresia had relatively lower graft failure rates (Figure 5).

Table 7 Tests of equality of survival functions for different diagnosis groups (N=1,705)

Test	Test of Equality	
	Chi2(5)	p-value
Breslow	8.71	0.1210
Peto-Peto	11.26	0.0465
Tarone-Ware	10.02	0.0746

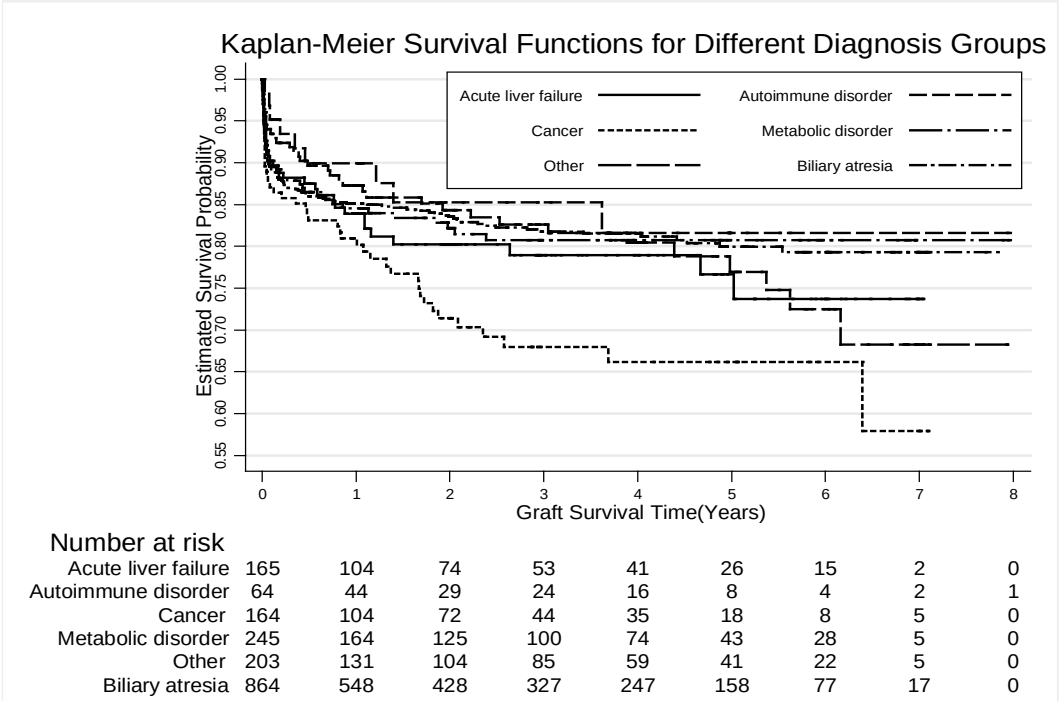


Figure 5 Graft Survival Functions for Different Diagnosis Groups

3.5 Trajectory Groups vs. Diagnosis Groups

As trajectory groups and the diagnosis groups are the two main predictors explored in this project, it was essential to explore the cross-distribution of these two. The results may provide some clue to give a better explanation of our findings so far (Table 8).

Table 8 Cross distribution of diagnosis vs. trajectory groups (N=1,705)

Group	Diagnosis						Total
	Acute Liver Failure	Autoimmune Disorder	Cancer	Metabolic Liver Disease	Other	Biliary Atresia	
1	1	8	94	89	33	38	70
2	4	14	57	70	54	86	91
3	10	10	6	33	40	185	288
4	34	13	2	15	20	289	379
5	76	11	4	20	27	184	331
6	34	7	0	13	18	76	157
7	6	1	1	5	11	6	29
Total	165	64	164	245	203	864	1,705

**Pearson chi2(30) = 805.2603 Pr < 0.001

More than 57% of cancer patients turned out to fall into trajectory group 1, which had the lowest PELD scores but not the lowest graft failure rate as we found. Another diagnosis group that was disproportionally distributed among the PELD-score trajectory groups was acute liver failure. It was not hard to find out that cancer and acute liver failure were the two diagnoses whose survival curves stood out among all the diagnosis groups. To better analyze the effect of

PELD-score trajectory groups on graft survival with least noise as possible, we analyzed the data again with cancer patients excluded from the data set. The characteristics of cancer patients were reported and discussed separately.

3.6 Graft Survival Excluding Cancer Patients

After removing the cancer patients from the data set, groups 5 and 6 had significantly higher graft failure rates while group 1 and 2 had the lowest graft failure rates. The trajectory pattern became a significant predictor of graft survival experience (p=0.0263). Besides, all the tests of equality of survival functions as well as the tests for trend were highly significant. The influence of the PELD-score trajectory tends to be more obvious with the absence of cancer patients (Table 9, 10, and Figure 6).

Table 9 Graft failure rates across PELD trajectory groups excluding liver cancer patients (N=1,541)

Group	n	%	Average Listing PELD	Average Transplant PELD	Median Follow-up time (Days)	Range of Follow-up Time (min, max)	Censoring (n, %)
1	169	10.97	6.73	5.70	745	2592 (1, 2593)	146, 86.4
2	228	14.80	0.80	1.35	730.5	2920 (1, 2921)	198, 86.8
3	278	18.04	9.79	11.74	744.5	2902 (1, 2903)	234, 84.2
4	371	24.08	16.47	19.53	780	2923 (1, 2924)	311, 83.8
5	318	20.64	22.88	26.31	510.5	2899 (1, 2900)	244, 76.7
6	148	9.60	29.53	35.93	668	2569 (1, 2570)	117, 79.1
7	29	1.88	45.13	46.51	683	2573(1, 2574)	24, 82.7

Table 10 Log-rank tests of graft survival functions for different PELD trajectory groups, listing PELD and transplant PELD excluding liver cancer patients (n= 1,541)

Test	Test of Equality		Test of Trend	
	Chi2	p-value	Chi2(1)	p-value
PELD Trajectory Groups	14.92	0.0209	9.80	0.0017
Listing PELD	15.27	0.0183	---	---
Transplant PELD	14.91	0.0209	---	---

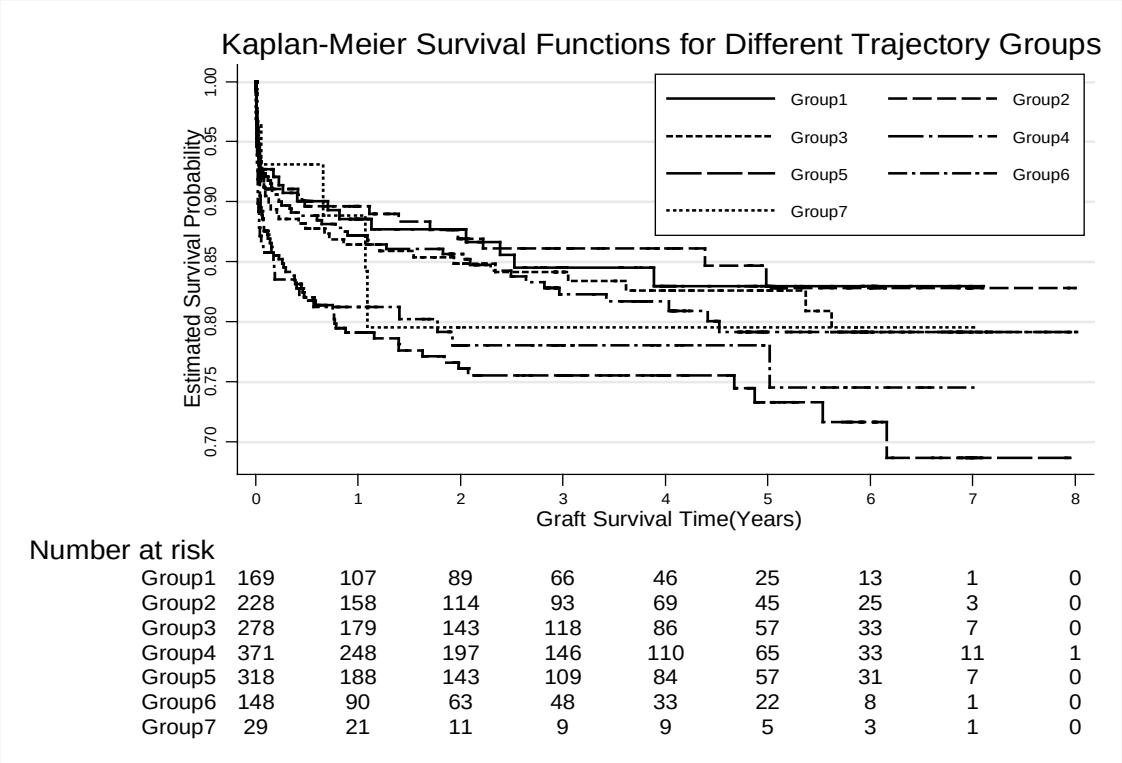


Figure 6 Graft survival functions for different PELD trajectory groups excluding liver cancer patients

After taking cancer patients out of the data set, the graft failure rates among different diagnosis groups were not significantly different from one another (Table 11, Table 7, and Figure 7). If we compare the test results in Table 6 (with liver cancer patients) and Table 9 (without

liver cancer patients), the difference was a good reflection of the specificity of the survival outcomes for the cancer patients. The prediction power of PELD was sufficient for patients diagnosed with biliary atresia, metabolic disorder, acute liver failure, and autoimmune disorder, but may not be sufficient for cancer patients.

Table 11 Tests of equality of graft survival functions for different diagnosis groups excluding liver cancer patients (n =1,541)

Test	Test of Equality	
	Chi2(4)	p-value
Breslow	1.61	0.8065
Peto-Peto	1.05	0.9024
Tarone-Ware	1.16	0.8840

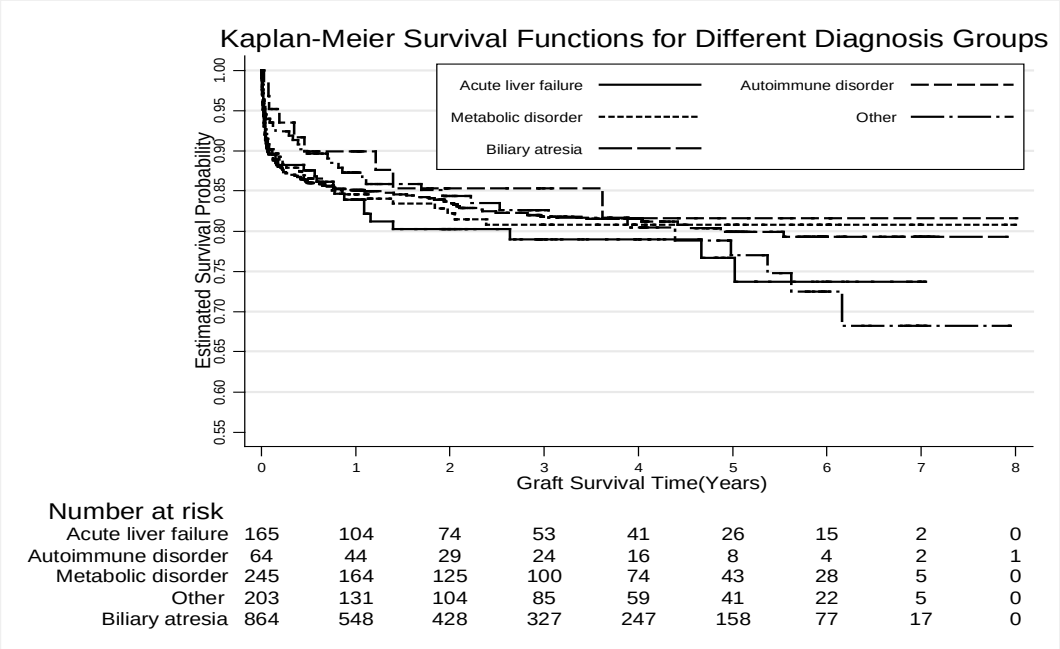


Figure 7 Graft survival functions for different diagnosis groups excluding liver cancer patients

Listing PELD score and transplant PELD, the first and last possible PELD measures for every patient, have been used by other researchers as the primary predictor for post-transplant survival. Although this project seeks to explore the longitudinal profile of PELD score as predictor of post-transplant graft survival, it was also of interest to find out about the effects of listing and transplant PELD score and compare to that of trajectory pattern.

Our analyses showed that both listing and transplant PELD scores were statistically significant predictors of graft survival: the higher the listing/transplant PELD scores, the higher the graft failure rate. To compare the Cox models based on listing PELD, transplant PELD, and trajectory PELD, Akaike information criterion (AIC) was used because none of the 3 models were nested within another. The results showed that the model with transplant PELD as a single predictor had slightly lower AIC than the other two models (Table 12).

Table 12 Comparison of 3 Cox models of graft survival (n =1,541)

Predictor	p-value	Log likelihood	AIC	BIC
Trajectory	0.0263	1878.052	3768.10	3800.15
Listing PELD	0.0043 *HR=1.0122	1881.130	3764.26	3769.60
Transplant PELD	0.0291 *HR=1.0095	1882.827	3767.65	3773.00

3.7 Additional Analysis I: Change in PELD Score from Listing to Transplant

In the current data set with 1,541 patients, PELD trajectories, listing PELD, and transplant PELD all had been identified as significant predictors of graft survival. To create a simpler or more

effective model that can accommodate two or more of these predictors, we created a change measure in PELD score from listing to transplant as done so by previous researchers. [11]

The change score for PELD was calculated by subtracting the listing PELD score from its counterpart at the time of transplant so this measure would be positive for most patients. The change scores had a mean of 2.62 and a median of 2 and were approximately normally distributed. Using a paired-sample t-test, the change score was significantly larger than 0 ($p<0.0001$), meaning that transplant PELD was significantly higher than the listing PELD.

3.8 Additional Analysis II: Potential Cut-off Point of the PELD Scores

Although the model with transplant PELD was identified as the statistically best model, the PELD trajectories still have their special merits given that they were derived using all the PELD scores within a 90-day period at the time of transplant. Groups 5 and 6 had the highest graft failure rates and their corresponding PELD scores were above 25 in general. The group with the next highest PELD scores (group 4) has a mean PELD score of 20. Based on these results, we tried to use 25 as a potential cut-off point for the listing and transplant PELD scores.

The results showed that about 20% of the patients had a listing PELD score greater than 25 while for transplant PELD, this proportion increased to 28.2%, which reflects the same trend of increase in PELD scores overtime as trajectory pattern. We also found that among those who died or underwent re-transplantation, the proportion of listing/transplant PELD scores greater than 25 was significantly higher than those who ended up censored. The proportion of graft failure among patients with a listing PELD score >25 and <25 was 21% and 16.45%, respectively ($p=0.054$). When using 25 as a cut-off for transplant PELD, the difference in graft

failure rates between the 2 groups was highly significant: 21.4% vs. 15.35% (p=0.008) (Table 13).

Table 13 Using 25 as the cut-off in listing and transplant PELD scores (n =1,541)

	Listing PELD		Transplant PELD	
	< 25	≥ 25	< 25	≥ 25
# of Censoring	1,030	244	932	342
# of Events	202	65	174	93
Total	1,232	309	1,106	435
Chi2(1)	3.7122		6.9501	
p-value	0.054		0.008	

4. 0 DISCUSSION

4.1 Longitudinal Profile of PELD Scores: Statistical Significance vs. Clinical Significance

As a criterion in the selection of trajectory patterns, all parameter estimates and group membership should be significant at a 0.05 level. The 7-group trajectory pattern we chose sufficed the criterion above; its coefficients for linear and quadratic terms were all very close to 0, which mean that the magnitude of the change in PELD scores, whatever the direction, was not large in an absolute sense. To take one step further, most of the lower PELD- score trajectory groups (groups 1-4) tended to have more flat curves than the higher PELD-score groups (groups 5-7). This suggests that patients with lower PELD scores, deemed to be less severe in their diseases, will also appear to be more stable overtime. Groups 6 and 7 have more obvious fluctuations in the PELD scores. This may be caused by smaller sample sizes, which allow for larger variability.

Results from the trajectory pattern analyses were echoed by the change score in PELD from listing to transplant, which had a mean value of 2.62 and a median of 2, while the range of PELD score in this data set was as large as 88. Although statistically significant, this difference may not be important in actual clinical practice. Statistically, given a sample size over 1, 500, substantial power of all the analyses is guaranteed. This is because that the variances of the measures will become very small such that even a slight change will end up a statistically significant. However, in clinical practice, which focuses on treating individual cases, the magnitude of the absolute difference often exceeds statistical significance. For a measure like the PELD score, which is measured under the same system, could still conceal the possible bias

and errors in its derivation and recording. When we take such potential problems into account, a mean difference between 2 to 3 may be rendered more non-significant for clinical practitioners than to biostatisticians.

On the other hand, the primary purpose of this project was to explore the underlying longitudinal patterns of the repeated PELD measures. Such trajectory patterns make use of all the PELD measures within 90 days prior to transplant and could provide some insight into the change in a patient's disease status which cannot be provided by a single measure of the PELD score.

There was a relatively small group in our 7-group trajectory patterns but it did not have the worst graft survival as indicated by its highest PELD scores. After examining data carefully, this group can be set aside as the miscellaneous group, as well as the fraction of patients with rarely high PELD scores. Combined with etiology, it is possible for patient from any diagnosis group to demonstrate a very high PELD score, yet such high values in PELD scores are not quite prevalent in clinical practice. When such cases arise in practice, it is feasible for clinicians to treat such patients with individualized plans. It is reasonable to pay more serious attention to patients in this group from the public health or population perspectives.

When a potential recipient enters the organ allocation system with a PELD score over 25, clinicians can have some confidence to say that in addition to what will happen during the waiting period, he or she will also have a non-optimal post-transplant survival experience. Clinicians will be more certain about their prediction if a transplant PELD score >25 is observed. This may not always be true for every patient but it will help the decision in organ allocation and patient prioritization.

4.2 "Best" Predictor of Post-transplant Survival: Trajectory Pattern vs. Listing/Transplant PELD

In previous studies, PELD was used as a predictor of mortality on the waiting list as it had been designed to. There are also some researchers who used single measure of PELD along with other covariates like donor factors to form a model to predict patients' post-transplant survival. However, previous studies were usually carried out using data from one medical center with a sample size around 100. We took the advantage of our resource of national data with over 1,500 patients to rerun the analyses with single PELD measure as predictor along wiht listing and transplant PELD scores for comparison.

The PELD trajectory patterns, the listing PELD, and the transplant PELD have all been shown to be statistically significant in predicting graft survival. It is more convenient to use a single measured PELD as a predictor given it would provide as strong statistical power as that provided by the trajectory patterns, which may be difficult and time consuming to derive. In our case, transplant PELD was the most statistically favorable by comparing AIC's of all the models, but its superiority was not evident. Still, trajectory PELD will provide additional information regarding patients' status overtime with parameter estimates for the actual pattern. Additionally, the application of trajectory pattern will facilitate the identification of rare or miscellaneous sub-group in the whole population as the way group 7 was identified in this project.

To obtain a balance between the concern of having to pick single PELD measure among numerous records available and the complexity of trajectory pattern, a change score in PELD from initial listing to final transplant is worth looking into.

4.3 "Best" Predictor of Post-transplant Survival: Trajectory Pattern vs. Etiology

The PELD scoring system was designed to act as an estimation of mortality on the waiting list and optimize organ allocation while regarding post-transplant survival estimation, measures like PELD score may not be the first choice for the majority. It will definitely raise the significance of PELD score in practice if it can be proven to be informative throughout patients' whole process of liver transplantation, yet the importance of etiology should never be overlooked.

After excluding cancer patients, the graft survival among the rest of the diagnosis groups were not significantly different from each other ($p=0.8929$). The Kaplan-Meier plot, complimentary to statistical tests, showed some interesting distinction among etiologies. The most obvious drops in survival probabilities could be observed in the first two years after transplantation and stayed stable from years 3 to 8. The survival curves crossed with each other during the whole study period, which indicated that relative hazards among etiologies had not been proportional over time ($p=0.004$). This phenomenon is foreseeable since diseases differ in their timelines and severity and the respective treatments for different diseases will trigger divergent curative and side effects among patients, which will also have impacts on graft failure.

Liver cancer patients, initially included in the data set and then taken out, clearly stand out by themselves regarding the outcome of interest. It is well known that the primary regimens for most liver cancer cases are hepatectomy and chemotherapy, liver transplant is a potential treatment but more appropriate for liver cancer patients with complications such as cirrhosis, or metastatic cancer in the liver or for whom the tumor size is relatively small. The treatments specific to cancer as compared to other liver diseases will strongly influence quality of life and outcomes of a patient.

The primary reason for excluding cancer patients lies in the fact that they tend to have a distinct survival experience. Liver cancer is a leading cause of death for Americans with a 5-year survival rate of only 28% [16], while for patients in the UNOS data set; the survival rate at the 8th year is as high as 78%. Second, liver transplant is regarded as the primary treatment for a special sub-group of all liver cancer patients as mentioned above so cancer patients in the current data may not be representative of its underlying population. Third, cancer patients turn out to have unusually low PELD scores as the majority of them belong to group 1 in PELD trajectory pattern and such uneven distribution can be problematic for secondary analysis. Also, the proportion of cancer patients is not high to start with (less than 10%) so its exclusion will not affect the statistical power of our analysis.

It is true that etiology is important for prognosis, yet its variation and possible complication will make it hard to make inferences to trace the real reasons for a certain kind of result. For patients in need of liver transplant, etiology can be taken into consideration and must be handled carefully.

4.4 Strengths, Weaknesses, and Future Plans

This project is one of the very first to explore the longitudinal profiles of PELD measures and associate them with post-transplant survival. We have also conducted these analyses using recently updated national registry data to ensure the generalizability and power.

As an exploratory research project, new questions always arise as the analysis results accumulate. The survival analysis was carried out univariately without adjusting for potential confounding variables, such as the donor factors and the presence of certain complications in

patients. Also, the data we have may not suffice certain assumptions of the Cox proportional hazards model thus alternative methods should be recognized.

The use of traditional survival analysis is an intuitive and convenient way and the trajectory pattern of repeated measurements is not necessarily the best predictor available. In future studies, more advanced methodology could be applied to answer the questions proposed. Other researchers have proposed to combine trajectory analysis with propensity scores or to model trajectory pattern and survival functions jointly in one model. In summary, liver transplant/graft survival is an important medical issue that needs to be studied and reported continuously.

5.0 CONCLUSION

The PELD score has been proven to be a statistically important predictor for both pre and post liver transplant survival experiences, and its usage can be realized in different forms: researchers can choose from PELD score trajectory pattern, listing/transplant PELD score, change score in PELD from listing to transplant, as well as a cut-off of 25 based on their specific research questions and data available; medical practitioners can make use of such evidence to support their clinical decision. As our primary interest, PELD score trajectory has its own advantages in that it is a detailed description of the changes in patient's disease status overtime and it can help identify underlying group(s) and particular subgroup.

Etiology is a key component for prognosis and should still be given attention in future studies. Because PELD score is designed to disregard etiology and waiting list time when allocating organs, all end-stage liver diseases (excluding liver cancer) do not actually differ that much regarding post-transplant graft survival. Liver cancer patients, due to their specific treatments and survival experiences, need to be studied separately.

More advanced techniques such as propensity score, joint modeling can be applied and more covariates can be incorporated to gain deeper knowledge in this field, and such research will carry on as the UNOS data registry keeps updating.

BIBLIOGRAPHY

[1] McDiarmid S, Merion RM, Dykstra DM, Harper A. Selection of pediatric candidates under the PELD system. Liver Transplantation 2004; 10(Suppl 2): S23-S30.

[2] McDiarmid S. New liver allocation policies and their potential effect on pediatric patients awaiting liver transplantation. Pediatric Transplant 2002; 6:180-186.

[3] Components of the CTP score employed in the previous liver allocation policy. Available at United Network for Organ Sharing website: www.unos.org.

[4]MELD/PELD calculation documentation. Available at United Network for Organ Sharing website: http://www.unos.org/docs/MELD_PELD_Calculator_Documentation.pdf

[5] Cowles RA, et al. Timing of liver transplantation in biliary atresia-results in 71 children managed by a multidisciplinary team. Journal of Pediatric Surgery 2008; 43(9): 1605-1609.

[6] Farmer DJ, Venick RS, McDiarmid S, et al. Fulminant hepatic failure in children: superior and durable outcomes with liver transplantation over 25 years at a single center. Annals of Surgery 2009; 250(3): 484-493.

[7] Oh SH, Kim KM, Kim DY, Lee YJ, et al. Long-term outcomes of pediatric living donor liver transplantation at a single institution. Pediatric Transplant 2010; 14(7): 840-847.

[8] Carroll CL, Goodman DM, Superina RA, Whitington PF, Alonso EM. Timed Pediatric Risk of Mortality Scores predict outcomes in pediatric liver transplant recipients. Pediatric Transplant 2003; 7(4): 289-295.

[9] Freeman AB, Harper A, Edwards EB. Excellent liver transplant survival rates under the MELD/PELD system. Transplantation proceedings 2005; 37(2) 585-588.

[10] Bourdeaux C, Tri TT, et al. PELD score and posttransplant outcome in pediatric liver transplantation: a retrospective study of 100 recipients. Transplantation 2005; 79(9): 1273-1276.

[11] Nagin DS. Analyzing developmental trajectories: a semiparametric, grou-based approach. Psychological Methods 1999; 4(2): 139-157.

[12] Nagin DS, Odgers CL. Group-based trajectory modeling (nearly) two decades later. Journal of Quantitative Criminology 2010; 26: 445-453.

[13] Jones BL, Nagin DS. Advances in group-based trajectory modeling and a SAS procedure for estimating them. Sociological Methods Research 2007; 35: 542-571.

[14] Jones BL, Nagin DS. Roeder K. A SAS procedure based on mixture models for estimating developmental trajectories. Sociological Methods and Research 2001; 29(3): 374-393.

[15] Arrandale V, Koehoorn M, MacNab Y, Kennedy SM. How to use SAS® Proc Traj and SAS® Proc Glimmix in Respiratory Epidemiology. Part of Arrandale MS thesis https://circle.ubc.ca/bitstream/handle/2429/834/Arrandale_guidance_doc.pdf?sequence=1

[16] Survival rates for liver cancer, available at American Cancer Society website: http://www.cancer.org/Cancer/LiverCancer/DetailedGuide/liver-cancer-survival-rates.