CHARACTERIZATION AND PREDICTION OF ACUTE-ON-CHRONIC LIVER FAILURE IN PEDIATIC BILIARY ATRESIA PATIENTS LISTED FOR TRANSPLANT

Charles Ho
Rice University, Department of Statistics

STAT 450
Senior Undergraduate Capstone Project – Final Report
April 29, 2015

INSTRUCTOR
Dr. Rudy Guerra, PhD
Rice University, Department of Statistics

MENTORS
Dr. Philip Ernst, PhD
Rice University, Department of Statistics

Dr. Thaddaeus May, MD
Baylor College of Medicine / Texas Children's Hospital,
Section of Pediatric Gastroenterology, Hepatology and Nutrition

1. ABSTRACT

There is a need to accurately assess risk for acute-on-chronic liver failure (ACLF) for pediatric patients waiting for a liver transplant. The statistical analysis of this project aims to identify important clinical parameters indicating ACLF development as well as to develop a risk assessment model for the prediction of ACLF development based on a patient's data at time of listing. Data from a retrospective review of 114 biliary atresia patients at Texas Children's Hospital listed for transplant from 2001-2014 were analyzed. T-tests for significant differences suggest total bilirubin and INR were the most important clinical parameters that characterized ACLF patients. Using a Bayesian model averaging technique for variable selection, a logistic regression model was developed to predict ACLF outcome for patients in the dataset. The proposed model achieved an area under the ROC curve of 0.94 for the prediction of ACLF, which was significantly higher than those achieved by current scoring methods when applied to this dataset.

2. INTRODUCTION

Biliary atresia (BA), a liver disease associated with the blockage of the bile ducts, is the most common diagnosis for pediatric patients receiving a liver transplant. Some of these transplant patients may also experience acute-on-chronic liver failure (ACLF), which represents a rapid decompensation of the liver following a chronic liver condition. Previous research indicates that ACLF patients have poor survival outcomes, particularly in the short-term without serious interventions. However, there is tremendous improvement in survivorship among ACLF patients when they undergo liver transplantation.[1] The main concern at hand is that transplantation may not always be available for ACLF patients as there is a specific protocol for allocating transplant organs.

One of the primary criteria used by the United Network for Organ Sharing (UNOS), the organization which organizes the organ allocation process, is a scoring system named PELD, the Pediatric End-stage Liver Disease score.[2] This scoring model quantitatively measures the risk of death for patients. One critique of the use of PELD is that it does not always help allocate livers in a timely manner.[3] This previous study suggests that PELD "underestimates the near-term risk of death". Furthermore, for ACLF patients in particular, it is challenging to assess their need for transplantation since the acute nature of their disease is not observed until its rapid onset. Thus, they do not receive the priority on the transplant despite the excellent outcomes post-transplant.[1] Having a better understanding of ACLF risk may help address this dilemma.

2.1 Goals of Project

To better understand how to assess risk for ACLF for pediatric patients, the purpose of this statistical analysis project is to (1) identify key clinical indicators for ACLF development and (2) develop a risk assessment model to predict a patient's future development of ACLF based on their condition at time of listing on the transplant waitlist.

3. STUDY DESIGN & DATA COLLECTION

This data analysis project utilizes data collected from a retrospective medical record review of 114 biliary atresia patients at Texas Children's Hospital who were listed for liver transplantation from the years 2001 through 2014.

Data collected include blood laboratory values taken from the patient at the time of their listing on the transplant waitlist, which include patient's levels of total bilirubin, albumin, INR, etc. Anthropometric measures were also obtained for each patient, including their age-adjusted z-scores for weight, length (height), and BMI. These values were calculated according to the World Health Organization's child growth standards by inputting patient's raw measurements into the WHO Anthro software program (Version 3.2.2, 2011) Patient treatment information, such as

time on waitlist prior to transplant and age, are also included in the data. A complete listing of parameters included in this analysis is included in the appendix.

98 of the biliary atresia patients also had corresponding electrocardiogram (EKG) results near their time of listing on the transplant waitlist, which were also used in this analysis to identify potential association between electrical cardiac function with ACLF outcomes.

For all analysis, the primary outcome variable of interest is binary – whether or not the patient experienced ACLF while waiting for a transplant. A hepatologist identified twenty (20) ACLF patients through a specific criteria based upon review of patients' medical record.

4. METHODS FOR STATISTICAL ANALYSIS

All statistical analyses were conducted with R (Version 3.0.3). Packages used include BMA[4], for Bayesian Model Averaging and pROC[5] for ROC analysis.

4.1 Hypothesis testing for identification of significant clinical parameters indicating ACLF development

To identify important clinical parameters that can serve as indicators for ACLF development, two groups from the data were defined, one who were identified to have experienced ACLF ($n_1 = 20$) and the other not having experienced ACLF ($n_2 = 94$). Differences among blood lab values and anthropometric values of the ACLF and non-ACLF groups were assessed using two-sided student's t-test. Significance was assumed at $\alpha = 0.05$. An identical analysis was performed on the corresponding electrocardiogram (EKG) data, where there were 17 ACLF patients and 81 non-ACLF patients for the comparison of EKG parameters.

In order to analyze the relationship between certain binary predictors with the binary ACLF outcomes, a Pearson's chi-square test or, for small cell sizes, a Fisher's exact test was used to assess the significance of association. One such binary predictor is the corrected QT interval from the patients' EKG. A common rule-on-thumb used by clinicians for abnormal heart activity is if a patients' QT interval is above 445 milliseconds. A Fisher's exact test was used to assess this association with ACLF outcomes.

A component of the PELD score, which is the currently used quantitative model for evaluating priority on the liver transplant waitlist, utilizes a measure called "growth failure". This indicator is a binary variable defined as "yes" if the patient's age-adjusted height or weight is less than two standard deviations from the respective mean. Pearson's chi-square test was used to assess the relationship between growth failure with ACLF outcomes, which is suggested by the PELD model.

4.2 Development of risk assessment model for ACLF

The goal for this part of the project was to develop a statistical model that can predict the risk of a patient developing ACLF based on factors at the patient's time of listing on the transplant waiting list. The outcome is binary (whether or not patient develops ACLF). Logistic regression models were developed for this purpose.

4.2.1 Variable selection

A Bayesian model averaging method was used to select variables to include in the logistic regression models. The method is specified in a 2001 paper by Viallefont, Raftery, and Richardson.[6] In this project, all relevant single-term variables were initially considered. Based on the Bayesian model average method, the posterior estimates of the

parameters and standard errors were computed, as well as the posterior probability that the variable's regression parameter was not equal to zero. Rather than listing all twenty relevant variables studied, a subset of these results is shown in Table 1.

Coefficient	Parameter Estimate $E(\beta \mid data)$	Standard Error $SD(\beta \mid data)$	Posterior Probability $P(\beta \neq 0 \mid data)$
Total Bilirubin	0.08	0.06	1.00
INR	1.65	1.16	1.00
Listing Age	-0.01	0.01	0.60
Length-age z-score	-0.16	0.27	0.33
Growth failure factor	0.32	0.76	0.19
Ventricular Rate	0.01	0.02	0.16
GGT	0.00	0.00	0.12

Table 1. A sample of the results from an iteration of Bayesian model averaging.

The posterior probabilities were used as the criteria for variable selection. The "best" variable was those parameters whose posterior probability (i.e. $Pr(\beta \neq 0 \mid data)$) was the greatest. Here, the top five variables were selected in this first iteration of Bayesian model averaging. Then, in another iteration of Bayesian model averaging, two-way and three-way interaction terms were considered among these five selected variables. The best variables, as indicated by the highest posterior probabilities, were selected to include in a logistic regression model.

Because of the Bayesian approach utilized in this variable selection procedure, prior information can be used to influence variable selection. In this case, prior clinical knowledge was used. Based on suggestions from the hepatologist, the procedure was tuned to prioritize models which included total bilirubin and INR. This explains the fact that $Pr(\beta \neq 0 \mid data) = 1$ for those variables, as this was specified by the author throughout the variable selection process.

4.2.2 Logistic regression

Based on the analysis for variable selection, a logistic regression model was developed. Due to some missing data points for some of the variables, only 85 of the biliary atresia patients (17 of which were ACLF patients) could be used to develop the model. Using a logistic regression models assumes linearity among the predictor variables. This is assumed in this analysis because previous liver disease evaluation models, such as PELD, are linear as well.

4.2.3 Diagnostic measures of model fit

Additionally, to assess how well the model fits the data, several diagnostic measures were observed, such as residual plots, deviance residuals, and evaluation of observed Cook distances. An examination of a Cook distance plot identified influential points that could have significantly affected the regression model. Influential points were removed from the data and the regression model was developed without those patients in consideration.

4.2.4 Model evaluation via ROC analysis

The predictive performance of the developed logistic regression model was evaluated by ROC analysis. Area under of curve values for the model was used as a measure of predictive ability. Sensitivity and specificity at the optimal point (defined here as the operating point on the ROC curve closest to the top-left corner) were also determined.[7] These operating characteristics were then compared with that of current liver disease assessment models when applied to this dataset. These include the PELD score, which is currently used in practice, as well as the BALF model[8], a recently proposed model in liver transplantation literature to evaluate liver fibrosis. Significance testing for the comparison between ROC curves utilizes Delong's test for two correlated ROC curves.[9]

4.2.5 Model Validation

Because another source of biliary atresia patient data is not immediately available, external validation is not yet a feasible analysis to validate the model's ability to predict ACLF. Thus, a 4-fold cross validation method was used as one avenue for validating the developed logistic regression models. Five iterations of 4-fold cross validation were performed and the averaged results from each considered model were compared.

5. RESULTS

5.1 Assessment of significant clinical parameters for differentiating between ACLF and non-ACLF patients

The overall comparison between ACLF and non-ACLF via student's t-test indicate that age, total bilirubin, INR levels are significantly different between the two groups (Table 2).

Parameter	ACLF average $n_1 = 20$	Non-ACLF average $n_2 = 94$	p-value
Listing age	221.0 ± 80.7	709.8 ± 1122.0	<0.01
Total Bilirubin	16.6 ± 10.8	7.4 ± 6.9	<0.01
INR	1.8 ± 0.5	1.3 ± 0.4	<0.01
Platelets	149.1 ± 79.2	190.0 ± 136.7	0.09
GGT	326.3 ± 325.1	464.4 ± 465.0	0.12
BMI-age z-score	-0.4 ± 1.4	0.1 ± 1.3	0.16
Weight-age z-score	-1.2 ± 1.9	-0.8 ± 1.4	0.37
Sodium	136.3 ± 4.6	137.3 ± 3.5	0.40
Weight-length z-score	-0.2 ± 1.5	0.1 ± 1.2	0.48
Alkaline phosphate	623.8 ± 282.0	674.4 ± 368.5	0.50
Creatine	0.24 ± 0.1	0.2 ± 0.1	0.66
Length-age z-score	-1.6 ± 2.9	-1.4 ± 2.1	0.84
Albumin	3.1 ± 0.4	3.1 ± 0.6	0.93

Table 2. Blood lab values and anthropometric measures for ACLF and non-ACLF groups. Means are reported, with errors representing one standard deviation. P-values are from the student t-test.

The analysis of association between cardiac electrical parameters with liver failure is reflected in Table 3. None of the EKG parameters were deemed significantly different between the ACLF and non-ACLF groups, though the QRS complex interval was approaching significance.

Parameter	ACLF average $n_1 = 17$	Non-ACLF average $n_2 = 81$	p-value
QRS	60.8 ± 6.0	65.5 ± 19.3	0.08
R-axis	56.2 ± 50.8	72.0 ± 32.1	0.23
P-axis	46.4 ± 12.9	48.8 ± 17.1	0.53
T-axis	43.6 ± 18.9	46.7 ± 23.9	0.56
QT_C	424.2 ± 29.0	428.6 ± 28.0	0.57
Ventricular rate	125.8 ± 19.1	127.9 ± 22.6	0.70
PR interval	110.7 ± 13.6	112.2 ± 28.6	0.74
QT	295.5 ± 28.9	297.3 ± 33.3	0.82

Table 3. Comparison of electrocardiogram parameters for ACLF and non-ACLF groups. Means are reported, with errors representing one standard deviation. P-values are from the student t-test.

A Fisher's exact test ($p = 0.51$) indicates that the corrected QT results from EKG are not associated with ACLF outcomes. (Table 4) The estimated odds ratio (with small sample correction) is 1.84, with an associated 95% confidence interval of [0.44, 7.69].

	QT_C category	
Outcome	$QT_C < 445$	$QT_C \geq 445$
ACLF	15	2
Non-ACLF	62	18

Table 4. Two by two contingency table recording number of ACLF and non-ACLF patients with respect to their classification by the $QT_C = 445$ cutoff. The Fisher's exact test p-value was 0.51.

Pearson's chi-square test (p = 0.22) indicates that growth failure is not associated with ACLF outcomes (Table 5). The estimated odds ratio is 0.48, with an associated 95% confidence interval of [0.19, 1.29].

	Growth failure status	
Outcome	No growth failure	Growth failure
ACLF	10	10
Non-ACLF	64	31

Table 5. Two by two contingency table recording number of ACLF and non-ACLF patients with respect to their growth failure status. The Pearson chi-square test was 0.22

5.2 Development of risk assessment model for ACLF

Using the patients' data at their time of listing, logistic regression models were developed to characterize the relationship between listing parameters and subsequent development of ACLF. The outlook was for these models to act as a tool for clinicians to assess ACLF risk for patients initially placed on the liver transplant waitlist.

5.2.1 Variable selection

From the Bayesian model averaging variable selection described in the methods, the initially selected candidate variables were: total bilirubin, INR, listing age, length-age z-score, and the growth failure factor. Then, two-way and three-way interaction terms were considered in a second iteration of Bayesian model averaging. The variables which results in the best logistic regression model were: total bilirubin, INR, listing age and the 2-way interaction between total bilirubin and the growth failure factor. The best logistic regression model was selected based on the area under the curve results.

5.2.2 Proposed logistic regression model

The final logistic regression results, including the parameter estimate and associated odds ratio for each variable, are detailed in Table 6. Thus, total bilirubin, INR, listing age, and the interaction between total bilirubin and the growth failure factor at the patient's time of listing are utilized for the prediction of ACLF.

Coefficient	**Odds Ratio**	**Parameter Estimate**	**Standard Error**	**Test Statistic**	**p-value**
Intercept	0.11	-2.20	2.15	-1.02	0.31
Total Bilirubin	1.12	0.11	0.09	1.18	0.24
INR	14.67	2.69	2.02	1.33	0.18
Listing Age	0.97	-0.03	0.02	-1.83	0.07
Total Bili*Growth Failure	1.27	0.24	0.11	2.22	0.03

Table 6. Logistic regression results for the final proposed model to predict ACLF occurrence based on patient data at listing. P-values are from the Wald test.

5.2.3 ROC analysis results

ROC curves for the classification of ACLF patients at time of listing by the proposed model (Table 6), as well as by the existing PELD and BALF scoring models, were generated (Figure 1). Area under the curve values are used as a measure of overall predictive performance (Table 7). Sensitivity and specificity at the operating point closest to the top-left corner of the plot are also reported.

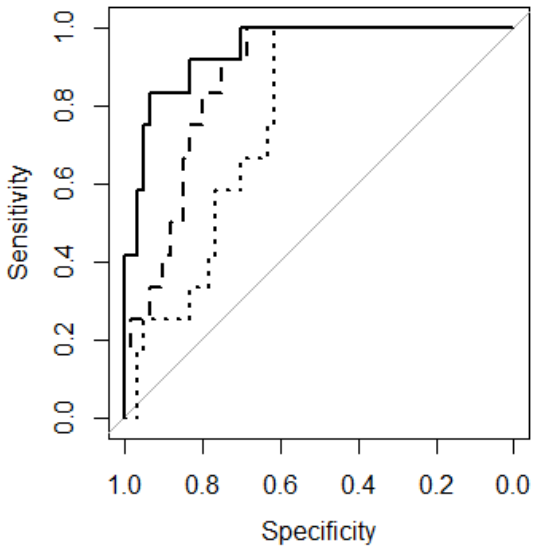

Figure 1. ROC curves representing sensitivities and specificities achieved by the proposed model (solid), PELD (dashed) and BALF (dotted) models.

Model	AUC	Sensitivity	Specificity
Proposed Model	0.941	93.3%	83.3%
PELD	0.872	80.0%	83.3%
BALF	0.768	61.7%	100.0%

Table 7. ROC results for comparing the proposed logistic regression model to the PELD and BALF scoring models

Significance testing using Delong's test indicates that the proposed model's area under the ROC curve is significantly higher than that of the PELD model ($p = 0.03$). The sensitivity for predicting ACLF from the proposed model is much greater than those yielded by the PELD and BALF models.

5.2.4 Cross-Validation

In Table 8, the average area under the ROC curve results from each of the five loops of four-fold cross validation are reported. Additionally, the average over all loops is computed and compared between the proposed model and the PELD model. The proposed model still achieves greater area under the curve values than the PELD score.

Model	Loop 1	Loop 2	Loop 3	Loop 4	Loop 5	**Average**
PELD	0.867	0.910	0.886	0.851	0.864	**0.876**
Proposed Model	0.923	0.942	0.893	0.901	0.923	**0.916**

Table 8. Area under the curve values for each of the five iterations under 4-fold cross validation, as well as the average area under the curve value achieved by the PELD and proposed model.

5.2.5 Final proposed model

Finally, a scoring formula is proposed based on the proposed logistic regression model. Clinicians prefer to use a simple additive equation to calculate a numeric value that represents patient's risk. In this case, the final numeric value computed from such derived equation represents a patient's risk for ACLF.

The PELD and BALF were also based upon regression models and the PELD[2] and BALF[8] equations (equations 1 and 2) were constructed by utilizing the regression coefficients as weights for the model's included models. The scoring formula associated with this project's proposed model is also constructed the same way, which is reported in equation 3. A constant term was included to re-scale the range of values, as clinicians prefer a numeric scale starting at zero.

PELD Score = 4.63*(age factor) – 6.87*ln(albumin) + 4.80*ln(TB) + 18.57*ln(INR) + 6.67*(growth failure) [Eq 1]

BALF Score = 7.196 + 1.438*ln(TB) + 0.434*ln(GGT) – 3.491*ln(albumin) – 0.670*(age) [Eq 2]

Proposed Model Score = 48.00 + 0.11*(TB) + 2.69*(INR) – 0.03*(age) + 0.24*(TB*growth failure) [Eq 3]

**Abbreviations: TB: total bilirubin, INR: international normalized ratio*

5.3 Summary of findings

- Listing age, total bilirubin, and INR levels were significantly different between the ACLF and non-ACLF cohorts.
- Electrocardiogram variables did not seem to be associated with ACLF outcomes.
- A logistic regression model for prediction of ACLF development was proposed and resulted in an area under curve value of 0.94, which was higher than that achieved by PELD.

6. DISCUSSION

6.1 Interpretation of hypothesis tests results

The result that total bilirubin and INR were significantly different between the ACLF and non-ACLF groups is consistent with current clinical knowledge (Table 2). It is interesting that GGT was approaching significance, when using a more relaxed significance level, as GGT was a factor included in the recently proposed BALF score, but not used the current PELD score. Platelets were also another somewhat significant factor, although the interpretation of this significance is complicated by the known non-linear relationship between platelets levels and liver disease.

The lack of significance among the EKG parameters (Table 3 and 4) with respect to ACLF outcome is not surprising, based on expert clinical knowledge. However, this exploration of the relationship liver failure and heart function is not yet complete, as other types of data representing heart function, such as those from an echocardiogram, may suggest other conclusions.

6.2 Possible confounding factors

The analysis of ACLF outcomes in this analysis may be complicated by several factors, as outlined below.

6.2.1 Time on the waitlist prior to transplant

It is important to mention that the patient's identification as having experienced ACLF was retrospective. This may mean that identification of ACLF patients out of the 114 biliary atresia patients may not reflect the true nature of the patients' liver disease. For instance, the time of transplantation for patients on the waitlist is not controlled and can

be highly variable and arbitrary. The availability of a donor organ for patients on the waitlist depends on patient's compatibility with the new organ as well as how highly the patient is advocated by their transplant team for the procedure.

The implication is that it may be possible for a patient in the non-ACLF group to have been transplanted early enough to avoid ACLF, even if they were on-track to develop ACLF. Such a phenomenon would seriously affect the comparison of ACLF and non-ACLF patients in this analysis. To preliminary assess this issue, a subset of the non-ACLF group was formed. This includes only non-ACLF patients who waited longer (n = 25) than the longest wait time experienced by an ACLF patient. Differences among blood lab values and anthropometric values of the ACLF and non-ACLF groups were assessed using two-sided student's t-test (Table 9).

Parameter	**ACLF average $n_1 = 20$**			**Subsetted Non-ACLF average $n_2 = 25$**			**p-value**
Total Bilirubin	16.6	±	10.8	4.9	±	4.8	<0.01
INR	1.8	±	0.5	1.2	±	0.2	<0.01
Platelets	149.1	±	79.2	266.8	±	182.0	0.01
GGT	326.3	±	325.1	636.2	±	485.3	0.02
Listing age	221.0	±	80.7	907.4	±	1537.9	0.04
Albumin	3.1	±	0.4	3.3	±	0.5	0.23
Sodium	136.3	±	4.6	137.9	±	3.4	0.23
Creatine	0.24	±	0.1	0.3	±	0.1	0.25
Length-age z-score	-1.6	±	2.9	-1.2	±	2.1	0.70
Weight-age z-score	-1.2	±	1.9	-1.0	±	1.4	0.73
Weight-length z-score	-0.2	±	1.5	-0.3	±	0.8	0.74
BMI-age z-score	-0.4	±	1.4	-0.3	±	1.8	0.90
Alkaline phosphate	623.8	±	282.0	624.0	±	351.6	0.99

Table 9. Blood lab values and anthropometric measures for ACLF and the subsetted non-ACLF group who waited longer on the waitlist. Means are reported, with errors representing one standard deviation. P-values are from the student t-test.

As shown in Table 2, the comparison between the subsetted non-ACLF group and the original ACLF group indicate that not only are total bilirubin and INR levels significantly different but also GGT and platelet levels. The differences between total bilirubin and INR are also more pronounced in this comparison. This may indicate that further examination of how patients are labeled in terms of their ACLF outcome is necessary, as it impacted the results when comparing between Tables 2 and 9.

6.2.2 Patient's age at time of listing

The very large variability among the ages among the non-ACLF patients at their time of listing is also another point of further analysis, as shown in Tables 2 and 9. The ACLF cohort consisted of mostly very young children (under a year old), while some non-ACLF patients were well into their teenage years. Controlling for age may provide another perspective in the comparison between ACLF and non-ACLF patients.

6.2.3 Previous medical history prior to listing for liver transplant

Significant and invasive procedures prior to listing on the transplant waitlist may seriously affect a patient's outcome with respect to ACLF. For biliary atresia patients, most patients undergo a Kasai HPE (hepatopotoenterostomy), which is considered to be a temporary life-extending treatment for biliary atresia before a more permanent intervention, such as a liver transplant, can be implemented. However, HPE alters the anatomy concerning the digestive and hepatic systems and, if performed at the wrong stages of a child's development, can be detrimental to a child's final health outcome. Thus, the relationship between whether or not a child received a Kasai and the subsequent development of ACLF should be investigated as well.

6.3 Interpretation of proposed logistic regression model

As shown in Table 6, there are four terms in the regression model. This rather small number of predictors is one strength of the model, as it offers simplicity and is one way to avoid significant overfitting. Other variable selection methods, such as a stepwise regression algorithm, yielded models with similar predictive performance, but included a much large number of predictors (usually more than six).

The incorporation of growth failure in the model is initially curious, since this factor is also included in the PELD score. Yet, it does not appear to be quite significant from hypothesis testing (Tables 2, 5, and 9). While unexplained in the PELD score, the proposed model (Table 6) includes growth failure as an interaction term with total bilirubin, which suggests that growth failure (and associated growth parameters, such as the length-age z-scores) may still be important in assessing ACLF risk in conjunction with information from total bilirubin levels. This interaction also can help explain the lack of significance for the parameter estimate associated with total bilirubin, as some the regression effect of total bilirubin was shifted over to the interaction term.

6.3.1 Consideration of other regression or classification models for ACLF prediction

Other more sophisticated regression and classification techniques were also considered to model the relationship between listing clinical parameters with ACLF risk. These included ridge regression, discriminant analysis, support vector machines, and random forests. However, these models yielded poorer classification performance.

6.3.2 Comments on model validation and over-fitting

The cross-validation results, in terms of the AUC values, are not as convincing as the results when testing the model on the trained data (Tables 7 and 8). The purpose of cross-validation was to examine the potential over-fitting by the proposed model and to simulate the performance of the prediction model if it were to be implemented on new datasets.

However, even if the cross-validation results indicate some degree of over-fitting, the originally reported AUC result still indicates that it is possible to generate a model that is more suited for ACLF prediction than the PELD scoring system. Furthermore, the relatively small number of ACLF cases included in this dataset makes it difficult to avoid the overfitting issue. A general guideline[10] is that ratio of subjects to the number of covariates in the logistic regression model should be less than 10. Since we have only 20 ACLF patients and 4 terms on the regression model, this guideline is not met. Performing this analysis on a larger dataset with a greater number of ACLF patients may provide more insight into this matter.

6.4 Directions for future work

This analysis can be expanded to also include echocardiogram data. There is a corresponding set of echocardiogram data for these same biliary atresia patients studied in this analysis. There is a clinical rationale for analyzing the relationship between geometric measure of heart function and liver failure.

Additionally, gaining access to a large database of biliary atresia patients listed for transplant, such as those coming from other major transplant institutions, can make this analysis and model building more significant.

6.5 Final remarks

While the results suggest that the proposed model in this paper provides better prediction performance than the currently implemented PELD model (Figure 1), this not necessarily meant to directly criticize or replace the PELD model. The PELD model was not designed to assess ACLF risk specifically, as it is a more general scheme for evaluating liver disease severity. A suggestion from this analysis is that, as a supplement to the PELD model, additional assessment methods, such as this paper's proposed model, can be used to evaluate liver transplant patients' priority for an organ in order to optimize survival outcomes. This may be particularly helpful for determining if special cases, such as future development of ACLF, require increased attention for transplant.

7. CONCLUSION

A logistic regression model, built through a Bayesian model averaging method of variable selection, is proposed as a risk assessment model for the prediction of acute-on-chronic liver failure in pediatric biliary atresia patients listed for transplant. This prediction model achieved an area under the ROC curve value of 0.94, when testing the model on the trained data. This AUC value was higher than those achieved by the PELD scoring model used clinically today. This should raise discussion for the use of supplemental models in addition to the PELD model for assessing transplant priority.

ACKNOWLEGEMENTS

I would like to thank Dr. Rudy Guerra for organizing the STAT 450 course and providing helpful feedback and discussion for this project. I thank Dr. Philip Ernst for useful feedback that helped lay the foundation of this work. I also thank the following Rice undergraduate students who were involved in collecting the data prior to the analysis of the data: Jacob Smith, Jesal Shah, Ross Pepper, Julia Zhang, Alex Tedesco and Sahar Noorani. I thank Dr. Thaddaeus May for proposing the project, providing access to the data, and giving necessary knowledge from the clinical perspective.

REFERENCES

[1] Finkenstedt, A.; *et.al*. "Acute-on-chronic liver failure: excellent outcomes after liver transplantation but high mortality on the wait list". *Liver Transplantation*, 2013. 19: 879-886

[2] United Network for Organ Sharing. "Question and answers for transplant candidates about MELD and PELD"

[3] Schneider, B.L.; *et.al*. "Critical analysis of the pediatric end-stage liver disease scoring system: a single center experience." *Liver Transplantation*, 2005. 11(7): 788-795

[4] Raftery, A.; *et.al*. "BMA: Bayesian Model Averaging". R Package Version 3.18.1.

[5] Xavier, R.; *et.al*."pROC: an open-source package for R and S+ to analyze and compare ROC curves". *BMC Bioinformatics*, 2011. 12:77

[6] Viallefont, V.; *et.al*. "Variable selection and Bayesian model averaging in case-control studies". *Statistics in Medicine*, 2001. 20:3215-3230

[7] Perkins, N.J.; Scheisterman E.F. "The inconsistency of optimal cutoff points using two ROC based criteria" *American Journal of Epidemiology*, 2006. 163(7): 670-675

[8] Tomita, H.; *et.al*. "Long-term native liver fibrosis in biliary atresia: development of a novel scoring system using histology and standard liver tests". *Journal of Hepatology*, 2014. 60: 1242-1248

[9] Delong, E.R.; *et.al*. "Comparing areas under two or more correlated receiver operating characteristic curves: a nonparametric approach". *Biometrics*, 1988. 44(3): 837-845

[10] Kalin, A.C.; *et.al*. "Recommendations for the assessment and reporting of multivariable logistic regression in transplantation literature". American Journal of Transplantation, 2010. 10(7): 1686-1694

APPENDIX

For the 114 biliary atresia patients studied, the following are a list of parameters collected and analyzed in this data analysis project, along with their units, when appropriate.

Analyzed parameters	Unit
Blood lab result parameters	
Albumin	g/dL
Alkaline phosphate	IU/L
Creatine	mg/dL
GGT (gamma-glutamyl transferase)	IU/L
INR (international normalized ratio)	
Platelets	10^{-3} µL
Sodium	mmol/L
Total bilirubin	mg/dL
Anthropometric measures	
Weight-length z-score[a]	
Weight-age z-score	
Length-age z-score	
BMI-age z-score	
Growth failure factor[b]	
Electrocardiogram parameters	
Ventricular rate	bpm
PR interval;	ms
QRS	ms
QT	ms
QT_c (corrected QT)	ms
P Axis	degree
R Axis	degree
T Axis	degree
Other information	
Age at listing	days
Time on waitlist prior to transplant/death	days

Table A. Complete listing of parameters analyzed in this data analysis project.

Notes

[a]Length refers to the patients' height

[b]Growth failure factor is coded as "1"if the weight-age z-score or length-age z-score were less than -2. Otherwise, it was coded as "0"