Published in final edited form as:

Clin Gastroenterol Hepatol. 2016 August ; 14(8): 1199–1206.e2. doi:10.1016/j.cgh.2016.03.046.

Development of a Model to Predict Transplant-free Survival of Patients with Acute Liver Failure

David G Koch, MD, MSCR¹, Holly Tillman, MS,² Valerie Durkalski, PhD,² William M Lee, MD³, and Adrian Reuben, BSc, MBBS, FRCP, FACP¹

¹Division of Gastroenterology and Hepatology, Medical University of South Carolina, Charleston

²Department of Public Health Sciences, Medical University of South Carolina, Charleston

³Division of Digestive and Liver Diseases, Department of Internal Medicine, University of Texas Southwestern Medical Center at Dallas

Abstract

Background & Aims: Patients with acute liver failure (ALF) have a high risk of death that can be substantially reduced with liver transplantation. It is a challenge to predict which patients with ALF will survive without liver transplant because available prognostic scoring systems are inadequate. We devised a mathematical model, using a large dataset collected by the Acute Liver Failure Study Group (ALFSG), that can predict transplant-free survival in patients with ALF.

Methods: We performed a retrospective analysis of data from 1974 subjects who met criteria for ALF (coagulopathy and hepatic encephalopathy within 26 weeks of the first symptoms, without pre-existing liver disease) enrolled in the ALFSG database from January 1, 1998 through June 11, 2013. We randomly assigned the subjects to development and validation cohorts. Data from the development cohort were analyzed to identify factors associated with transplant-free survival (alive without transplantation by 21 days after admission to the study). Statistically significant variables were used to create a multivariable logistic regression model.

Results: Most subjects were women (70%) and Caucasian (78%); acetaminophen overdose was the most common etiology (48% of subjects). The rate of transplant-free survival was 50%. Admission values of coma grade, ALF etiology, vasopressor use, and log transformations of bilirubin and International Normalized Ratio were significantly associated with transplant-free

Correspondence: David Koch, MD, MSCR, MUSC Division of Gastroenterology and Hepatology, 114 Doughty Street, Charleston, SC 29425, kochd@musc.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Author Contributions:

David Koch- Implemented the study concept and design, assisted in data analysis and interpretation, and drafted the manuscript.

Holly Tillman- Acquired the data, performed the primary data analyses, and assisted in critical revisions of the manuscript.

Valerie Durkalski- Assisted in data analysis and interpretation, study supervision, and critical revisions of the manuscript.

William M. Lee- Primary Investigator for the Acute Liver Failure Study Group who was involved in study supervision as well as critical revisions to the manuscript.

Adrian Reuben- Implemented the study concept and design, and assisted in study supervision as well as critical revisions to the manuscript.

Disclosures: None

survival, based on logistic regression analysis. In the validation cohort, the resulting model predicted transplant-free survival with a c-statistic value of 0.84, 66.3% accuracy (95% confidence interval [CI], 63.1%–69.4%), 37.1% sensitivity (95% CI, 32.5%–41.8%), and 95.3% specificity (95% CI, 92.9%–97.1%).

Conclusion: Using data from the ALFSG, we developed a model that predicts transplant-free survival of patients with ALF based on easily identifiable hospital admission data. External validation studies are required.

Keywords

acute liver failure; prognosis; predictive model; mortality

Introduction:

Acute liver failure (ALF) is an uncommon disease wherein a catastrophic injury to the liver results in coagulopathy and hepatic encephalopathy (HE) in patients without preexisting liver disease. Spontaneous survival (SS) rates in ALF are low (approximately 45%)^{1, 2} and vary depending on the etiology, but liver transplantation (LT) significantly improves survival². The rarity of ALF and variability in its clinical course, however, complicate the decision process. The King's College Hospital (KCH) Criteria, developed in 1989 using data largely collected prior to use of LT, is the most utilized prognostic score³ but suffers from poor sensitivity and negative predictive value^{4, 5}, as many patients who do not meet criteria die nonetheless. Moreover, use of KCH criteria appears to have become less reliable in recent years^{4, 6}. Other scores have drawbacks too, such as the ALFSG index that relies on non-routine serologic markers of apoptosis and necrosis that have been shown to be predictive of outcomes in ALF,⁷ and the Sequential Organ Failure Assessment (SOFA) and Model of End Stage Liver Disease (MELD) scores^{8–13} that predict outcomes with varying degrees of reliability. A recent systematic review of currently available prediction models in ALF identified many limitations and concluded that the SOFA and MELD models need improvement.¹⁴ To overcome these limitations and to develop a prognostic model to be used at hospital admission that is generalizable to all causes of ALF, clinically applicable, and reliable in the prediction of SS without transplant, we utilized clinical data from subjects enrolled in the Acute Liver Failure Study Group (ALFSG) database, a large, multicenter prospective registry of patients with ALF.

Methods:

Study Population.

We studied 1974 subjects enrolled from 28 academic centers in the United States into the NIH-funded ALFSG database from January 1, 1998 through June 11, 2013 who met criteria for ALF, namely coagulopathy (International Normalized Ratio [INR] ≥ 1.5) and any degree of HE within 26 weeks of the first symptoms without preexisting liver disease (ClinicalTrials.gov:NCT00518440). Written informed consent was obtained from the legal next-of-kin since subjects were encephalopathic. All centers complied with their local Institutional Review Boards' requirements.

Data Management and Integrity

Patient demographics, medical history, clinical features and laboratory values were collected at study enrollment and clinical status and laboratory results were also recorded serially for up to 7 days, or until discharge, death, or transplant if prior to 7 days. Etiologic diagnoses were made by each site's primary investigator. If an exhaustive search for the cause of ALF by experienced hepatologists was inconclusive, the etiology was recorded as Indeterminate. All data are managed and housed on a central server by the ALFSG Statistical and Data Management Center at the Medical University of South Carolina, and annual site visits are conducted by the ALFSG clinical leadership.

Primary Outcome

The primary outcome, SS, was defined as being alive without transplantation at 21 days following admission to the study. This has been the primary clinical endpoint defined by the with the assumption being that patients undergoing LT would have otherwise died, and are therefore censored from the study when LT is performed. Subjects with missing data for given variables were excluded from the respective analyses. The final model was tested to determine its ability to predict 21-day SS.

Covariate Selection and Univariate Analyses

Potential candidate variables for univariate analyses of the association with 21-day SS were chosen based on clinical input. The inclusion into multivariable modeling was based on clinical and statistical significance. HE grades 1–4¹⁵ (HE grade 3, and HE grade 4 or coma), and etiology was dichotomized as either (i) Favorable (acetaminophen overdose, pregnancy, ischemia, or hepatitis A) or (ii) Unfavorable (all other causes) for outcomes that vary according to etiology groups^{2, 16}. The predictive value of dynamic changes in clinical and laboratory variables obtained on a daily basis was also examined.

Univariate analyses assessed the association between the selected covariates and SS. Continuous variables were expressed as medians (interquartile ranges, IQR) and compared using the Wilcoxon rank sum test. Categorical variables were assessed using a chi-square test. Associations with a p-value < 0.05 were entered into the logistic regression model.

Modeling SS

A split-sample approach was used to develop and internally validate the model¹⁷. Split sampling was chosen based on the large cohort size and validated using a bootstrap approach¹⁸. For the creation of the split-sample, each subject in the database was assigned a random number and ranked in ascending order. Half were randomly assigned to a development cohort to construct a logistic regression model that predicted SS. The remaining subjects were entered into a validation cohort. Continuous variables were assessed for linearity in the log-odds with the Loess procedure and appropriate transformations were done if necessary. Variables were chosen by manual selection based on the Hosmer-Lemeshow goodness-of-fit test and the area under the receiver operating curve (AUROC) or concordance (c)-statistic – a measure of discriminatory ability¹⁹. The final model was then applied to the validation cohort, and the AUROC in the validation cohort was compared to the result for the development cohort. Further validation was done

independently with a bootstrap procedure to confirm the choice of variables as well as final model performance²⁰. For the bootstrap approach, 1000 random samples of 500 observations each were drawn from the complete original data set. Variable selection was conducted as defined above and resulted in a similar choice of variables. A similar approach was implemented when calculating the c-statistic using the bootstrap method. Finally, the KCH Criteria and MELD scores were applied to subjects in the database and their performances in predicting SS, as assessed by the AUROC, were compared to our model. Statistical significance was defined as a two-sided p-value less than 0.05. Data analyses were performed using SAS (v 9.1.03 Cary, NC).

Results

Demographics of Patients Enrolled in the ALFSG Database

Of the 1974 patients enrolled into the registry during the study period, the majority were women (70%) and Caucasian (78%), Table 1. The most common ALF etiology was acetaminophen overdose (48%) followed by Indeterminate (12%) and Drug-Induced Liver Injury (DILI, 11%). On admission to the study, patients had severe liver dysfunction with a median bilirubin 6.9 mg/dL (3.7 mg/dL – 18.9 mg/dL) and INR 2.7 (2.0 – 4.1). Arterial and venous ammonia levels were similarly elevated, and HE grades were evenly dispersed as mild (52%) or deep (48%). By 21 days after enrollment, 23% were transplanted and 30% had died (including 35 post-transplant deaths and 527 deaths without LT). Of the subjects who died without LT, 128 were listed for LT, 397 were not listed, and the listing status was unknown for 2 subjects. The overall SS rate was 50%.

Clinical Predictors of SS

Half of the subjects (987) were randomly allocated to a development cohort and the other half to a validation cohort. Demographic and clinical variables did not differ significantly between the two data subsets, except that the subjects in the validation cohort were slightly older 41.0 (29.0 – 52.0) years compared to 38.0 (28.0 – 50.0) years ($p=0.02$) in the development cohort (Supplemental Table 1). The SS rate was 50% for both groups.

In the development cohort, subjects with SS were more likely to be Caucasian (83%) and younger (35 vs. 40 years; $p=0.01$, Table 2) than non-SS subjects; there was no difference in SS between the genders (48% for men vs. 51% for women; $p=0.54$). As expected, etiology did significantly impact survival, with Favorable etiologies having an average SS rate of 68%, compared to 27% for Unfavorable etiologies. HE grades correlated inversely with SS and, when combined with etiology had a synergistic effect on the outcome. However, the combination of etiology and depth of HE alone did not adequately predict outcome, leading us to assess the impact of additional variables. Hemodynamic instability and organ dysfunction (such as need for vasopressive agents for blood pressure support, hemodialysis, or mechanical ventilation) portended a poor prognosis, as did the severity of liver synthetic dysfunction (bilirubin and INR levels) and higher blood levels of creatinine, phosphate, ammonia (both venous and arterial) and lactate. A rise in INR and creatinine during the study period did predict lower SS, but this was not the case for bilirubin. Finally, year of

enrollment into the study was a predictor of SS since subjects from more recent enrollment years had better survival than earlier recruits.

Creation and Validation of the Logistic Regression Model for SS

Statistically significant variables from 963 subjects were entered into a logistic regression model. Neither bilirubin nor INR met the assumption of linearity in the log odds of SS, so the natural logarithmic transformation (ln) of each was used for the multivariable analysis. Admission variables that remained statistically significant in the regression analysis included: HE grade, ALF etiology severity, vasopressor use, bilirubin and INR (Supplemental Table 2). The resulting ALFSG prediction model that computes the log odds (logit) for 21-day SS created from 878 of the 963 subjects with complete data for all included variables is:

$$\text{Logit SS} = 2.67 - 0.95(\text{HE}^*) + 1.56(\text{Etiology}^*) - 1.25(\text{Vasopressor Use}^*) - 0.70(\ln \text{bilirubin}) - 1.35(\ln \text{INR}).$$

*For Light HE insert 0, for Deep HE insert 1; for Unfavorable Etiology insert 0, for Favorable Etiology insert 1 – as defined in Methods; for absence of vasopressor use insert 0, for vasopressor use insert 1.

The logit SS can be transformed into the predicted probability of SS with the following formula: Predicted SS = $1/(1 + e^{(-1 * \text{Logit SS})})$. The model's predictability was high with a c-statistic of 0.84, 95% CI 0.8170 – 0.8681, p-value: <0.0001 (Figure 1).

The final model was then applied to the validation cohort (n= 885 subjects), and it performed similarly (c-statistic 0.84, 95% CI 0.8173 – 0.8687, p-value: <0.0001, Figure 1 and Supplemental Table 3). The model fit was good by the Hosmer and Lemeshow Goodness-of-Fit test (p-value 0.7567). The model performance was further independently validated with the bootstrap procedure. The bootstrap approach resulted in an average c-statistic of 0.868 (0.867–0.869).

The minimum predicted survival threshold that a clinician will accept and consider useful in aiding decisions in ALF will vary with experience and the philosophy of each program. The current model provides a continuum of probability for SS in ALF. The observed survival rates for cohorts of subjects from the ALFSG database that are stratified by their predicted SS according to the logistic regression model are shown in Table 3. For example, for subjects with predicted probabilities of SS between 50% - 60% and 90% – 100%, respectively, the observed SS rates were 56% and 96%. An optimal classification threshold of 80% was derived from the AUROC, which is similar to the 82% survival predicted by KCH Criteria for non-acetaminophen cases¹⁸. In the validation cohort, at the 80% threshold, the model had a sensitivity of 37.1% (32.5% - 41.8%), specificity of 95.3% (92.9% - 97.1%), positive predictive value (PPV) 88.6% (83.1% - 92.8%), negative predictive value (NPV) 60.5% (56.8% - 64.1%), and accuracy (overall correct classifications) of 66.3% (63.1% - 69.4%)%. These values are provided in Supplemental Table 3 for the development cohort. The performance of the model at other thresholds is provided in Supplemental Table 4, which illustrates a decrease in accuracy as the SS threshold increases. The decrease in the

incorrect predictions of SS - the goal of this study - is accompanied necessarily by an increase in incorrect predictions of death/transplant.

In addition to the variables chosen for the model, several others were also statistically significant in the univariate analysis (Table 2), including phosphate, year of enrollment into the study, prior illegal drug use, and transfusion of blood products (red blood cells, platelets, or fresh frozen plasma). Despite the fact that these variables make sense clinically, they did not increase the model's performance and thus there was no benefit to including them in the final model. Intuitively, changes over time in ammonia levels and other statistically significant variables in the univariate analysis (Table 2) would be expected to have greater power in predicting SS than variables obtained on admission. However, except for INR, dynamic changes in such variables over time were not statistically significant in the multivariable analysis. Paradoxically and contrary to expectation, even the change in INR during the study period offered no statistical advantage over the admission INR value with respect to predicting SS (data not shown). Thus, we have settled upon an economic model that fulfills the goal of this project, namely the most reliable prediction of SS, yet one that is parsimonious, comprehensible to clinicians, and being relatively easy to compute would be suitable for a smartphone application.

There were statistically significant interactions between the variables included in the final model. Each interaction was added to the model and compared to the final model with no interactions, and none improved performance (i.e., the interaction terms did not affect the AUROC). Finally, we tested the model's accuracy using data entered on each of the 7 days during the study period. The model performed similarly, irrespective of the post-enrollment day of the data, and, of importance as the number of subjects available for analysis decreased significantly later in the 7 days of enrollment in the study due to discharge, death or LT.

Model Performance According to Etiology and Coma Grade

Since etiology and HE grade are both highly significant predictors of SS, we stratified the 885 subjects from the validation cohort with complete data into 4 groups, to examine the model's accuracy for each combination of HE (i.e., mild or deep) and favorability of etiology as follows: (i) Favorable Etiology/ Mild HE (N = 209), (ii) Favorable Etiology/ Deep HE (N = 278), (iii) Unfavorable Etiology/Mild HE (N = 228), and (iv) Unfavorable Etiology/Deep HE (N = 170). The four groups were of approximately similar sizes.

Overall, the model accurately predicted SS in 66.3% of subjects (Table 4). It performed best in the most at-risk ALF subjects (Unfavorable Etiology/Deep HE or coma, - accuracy 79.4%), but was least accurate in the Favorable Etiology patients with Deep HE or coma (accuracy 56.5%). Using the most accurate (as defined above) but conservative threshold of 80% SS derived from the AUROC (as described above), the model falsely predicted the outcome in 298 subjects, since 277 of them who were not expected to survive - did, and the 21 who were predicted to survive - did not. This shows that the model errs on being conservative, as it incorrectly predicts SS in only 2.4% of subjects (21/885) who died or were transplanted, while for the remaining 277 cases (31.3%) death/LT was predicted when they actually spontaneously survived.

Comparison of the ALFSG Predictive Model for SS, to the Kings College Criteria and MELD Score

The KCH Criteria and MELD scores were applied to subjects in the dataset, and those models' abilities to predict SS were determined. The AUROC curves and corresponding c-statistic results were compared to our model using the DeLong method²¹. All comparisons were statistically significantly ($p < 0.0001$), confirming superiority of our model to predict ALF SS over the KCH criteria and MELD score (Figure 2).

Discussion

Using clinical data on subjects enrolled in the ALFSG database, we derived a model that predicts 21-day SS in ALF subjects, who would therefore not need liver transplantation. At a survival threshold of 80% which gives the best balance of sensitivity and specificity by ROC analysis, the model is conservative, as it very rarely (2%) predicted survival in patients who died or were transplanted. By contrast, the KCH Criteria failed to identify 28% of patients in the ALFSG dataset who eventually died or were transplanted, and therefore would not have been offered LT. Of greater importance and practical value, our model also accurately predicts SS across varying actual rates of survival, proving its utility as a continuous predictor of SS rather than just a threshold. The laboratory variables, bilirubin and INR, are readily available in any clinical setting, an important advantage over the ALFSG Index⁷ that had similar accuracy (c-statistic 0.82) but relies on research biomarkers of apoptosis and necrosis that are not readily available. Also, despite the fact that ALF is a very dynamic disease process in which the changes in clinical status that occur continually would be expected to predict outcome, our model is actually able to accurately foretell survival utilizing clinical and laboratory data from first day of enrollment in the study. This was supported by the fact that inputting data from a later time in the admission did not improve the model's performance. This does not imply that other variables, like ammonia, lactate, phosphate, respiratory failure and renal impairment - either static or dynamic - are not important in the pathogenesis or progress of ALF, but simply that mathematically their inclusion does not improve the model's performance.

Prior prognostic scores, including the KCH Criteria, have utilized etiology and initial HE grade as essential determinants of outcome. Since the causes of ALF can be divided into those with poor SS, usually $< 30\%$, and those with SS likelihood $> 50\%$, we developed the model based on this ALF severity stratification, rather than comparing acetaminophen hepatotoxicity to all other etiologies together.^{2,21}, and in this regard provide proof of concept for this etiologic classification. We confirmed the importance to outcome of the hierarchy of etiology and HE combinations. Other variables significantly associated with SS in the multivariable model were the need for vasopressive agents, bilirubin, and INR. Regarding kidney injury, neither serum creatinine nor need for renal replacement therapy predicted SS in the multivariable analysis, despite the fact that the MELD score (which includes creatinine) has been considered predictive for outcome in ALF^{10, 13, 22, 23}. We anticipated that serum ammonia and lactate would retain their predictive significance in the final model, based on prior experience^{1,24, 25}. However, this was not the case.

Both the goodness-of-fit (according to the Hosmer-Lemeshow statistic) and predictability were good in our model (c-statistic = 0.84). However, reporting the c-statistic and goodness-of-fit alone is insufficient to determine a model's accuracy, given the heterogeneous nature of the causes of ALF that have different outcomes^{2, 16}. Our model did perform quite differently, depending on the ALF etiology and degree of coma. Therefore, further modifications to this or any other model should be applied to each of these groups to better define the model's accuracy.

Aside from the variability in accuracy of the model according to etiology and coma grade, another limitation of this model and any one derived from patients in the era of LT is that the true SS in ALF is never actually known, because, in practice, LT is performed in patients who otherwise would have survived. Although standard diagnostic and management protocols are used across the ALFSG sites, criteria for deciding on the need for LT in a given patient may vary between transplant centers, and equally important, organ availability plays a role too. For these reasons, it is difficult to know the impact our model would have on transplant practices for ALF, including the use of scarce organs in treating these patients.

Conclusion

Using data derived from a large, multicenter, prospectively collected registry database, we created a multivariable logistic regression model that uses simple, readily available clinical and laboratory variables to predict SS in patients with ALF. The current ALFSG model provides a rigorously validated clinical decision tool that is conservative for predicting SS. Future studies are needed to confirm the applicability of our model in the clinical setting (including hospitals outside of tertiary referral medical centers and, perhaps, in non-US settings) and to determine the effect our model may have on the utilization of LT in the management of ALF.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements:

Members and institutions participating in the Acute Liver Failure Study Group 1998–2012 are as follows: W.M. Lee, M.D. (Principal Investigator); Anne M. Larson, M.D., Iris Liou, M.D., University of Washington, Seattle, WA; Timothy Davern, M.D., University of California, San Francisco, CA (current address: California Pacific Medical Center, San Francisco, CA); Oren Fix, M.D., University of California, San Francisco; Michael Schilsky, M.D., Mount Sinai School of Medicine, New York, NY (current address: Yale University, New Haven, CT); Timothy McCashland, M.D., University of Nebraska, Omaha, NE; J. Eileen Hay, M.B.B.S., Mayo Clinic, Rochester, MN; Natalie Murray, M.D., Baylor University Medical Center, Dallas, TX; A. Obaid S. Shaikh, M.D., University of Pittsburgh, Pittsburgh, PA; Andres Blei, M.D., Northwestern University, Chicago, IL (deceased), Daniel Ganger, M.D., Northwestern University, Chicago, IL; Atif Zaman, M.D., University of Oregon, Portland, OR; Steven H.B. Han, M.D., University of California, Los Angeles, CA; Robert Fontana, M.D., University of Michigan, Ann Arbor, MI; Brendan McGuire, M.D., University of Alabama, Birmingham, AL; Raymond T. Chung, M.D., Massachusetts General Hospital, Boston, MA; Alastair Smith, M.B., Ch.B., Duke University Medical Center, Durham, NC; Robert Brown, M.D., Cornell/Columbia University, New York, NY; Jeffrey Crippin, M.D., Washington University, St Louis, MO; Edwyn Harrison, Mayo Clinic, Scottsdale, AZ; Adrian Reuben, M.B.B.S., David Koch, M.D., M.S.C.R., Medical University of South Carolina, Charleston, SC; Santiago Munoz, M.D., Albert Einstein Medical Center, Philadelphia, PA; Rajender Reddy, M.D., University of Pennsylvania, Philadelphia, PA; R. Todd Stravitz, M.D., Virginia Commonwealth University, Richmond, VA; Lorenzo Rossaro, M.D., University of California Davis, Sacramento, CA; Raj Satyanarayana, M.D., Mayo Clinic, Jacksonville, FL; and Tarek Hassanein, M.D., University

of California, San Diego, CA. The University of Texas Southwestern Administrative Group included Grace Samuel, Ezmina Lalani, Carla Pezzia, Corron Sanders, Ph.D., Nahid Attar, Linda S. Hynan, Ph.D. and Angela Bowling and the Medical University of South Carolina Data Coordination Unit included Valerie Durkalski, Ph.D., Wenle Zhao, Ph.D., Holly Tillman MS, and Sarah Williams

Grant Support:

This study was funded by a National Institutes of Health grant (DK U-01 58369) for the Acute Liver Failure Study Group provided by the National Institute of Diabetes and Digestive and Kidney Diseases. Dr. Koch is also funded by the American College of Gastroenterology Junior Faculty Development Award.

Abbreviations:

(ALF)	Acute Liver Failure
(ALFSG)	Acute Liver Failure Study Group
(AUROC)	area under the receiver operating curve
(APACHE II)	Acute Physiology and Chronic Health Evaluation II score
(c)-statistic	concordance
(INR)	International Normalized Ratio
(KCH)	King's College Hospital
(LT)	liver transplantation
(MELD)	Model of End Stage Liver Disease
(SOFA)	Sequential Organ Failure Assessment score
(SS)	Spontaneous Survival

References

- Bernal W, Donaldson N, Wyncoll D. Blood lactate as an early predictor of outcome in paracetamol-induced acute liver failure: a cohort study *Lancet*. 2002; 359:558–63. [PubMed: 11867109]
- Ostapowicz G, Fontana RJ, Schiodt FV. Results of a prospective study of acute liver failure at 17 tertiary care centers in the United States *Ann Intern Med*. 2002; 137:947–54. [PubMed: 12484709]
- O'Grady JG, Alexander GJ, Hayllar KM. Early indicators of prognosis in fulminant hepatic failure *Gastroenterology*. 1989; 97:439–45. [PubMed: 2490426]
- McPhail MJ, Wendon JA, Bernal W. Meta-analysis of performance of King's College Hospital Criteria in prediction of outcome in non-paracetamol-induced acute liver failure *J Hepatol*. 2010; 53:492–9. [PubMed: 20580460]
- Shakil AO, Kramer D, Mazariegos GV. Acute liver failure: clinical features, outcome analysis, and applicability of prognostic criteria *Liver Transpl*. 2000; 6:163–9. [PubMed: 10719014]
- Anand AC, Nightingale P, Neuberger JM. Early indicators of prognosis in fulminant hepatic failure: an assessment of the King's criteria *J Hepatol*. 1997; 26:62–8. [PubMed: 9148024]
- Rutherford A, King LY, Hynan LS. Development of an accurate index for predicting outcomes of patients with acute liver failure *Gastroenterology*. 2012; 143:1237–43. [PubMed: 22885329]
- Cholongitas E, Theodoridou E, Vasianopoulou P. Comparison of the sequential organ failure assessment score with the King's College Hospital criteria and the model for end-stage liver disease score for the prognosis of acetaminophen-induced acute liver failure *Liver Transpl*. 2012; 18:405–12. [PubMed: 22213443]

- 9Craig DG, Reid TW, Wright EC. The sequential organ failure assessment (SOFA) score is prognostically superior to the model for end-stage liver disease (MELD) and MELD variants following paracetamol (acetaminophen) overdose *Aliment Pharmacol Ther.* 2012; 35:705–13. [PubMed: 22260637]
- 10Parkash O, Mumtaz K, Hamid S. MELD score: utility and comparison with King's College criteria in non-acetaminophen acute liver failure *J Coll Physicians Surg Pak.* 2012; 22:492–6. [PubMed: 22868013]
- 11Pelaez-Luna M, Martinez-Salgado J, Olivera-Martinez MA. Utility of the MAYO End-Stage Liver Disease score, King's College Criteria, and a new in-hospital mortality score in the prognosis of in-hospital mortality in acute liver failure *Transplant Proc.* 2006; 38:927–9. [PubMed: 16647512]
- 12Schmidt LE, Larsen FS. MELD score as a predictor of liver failure and death in patients with acetaminophen-induced liver injury *Hepatology.* 2007; 45:789–96. [PubMed: 17326205]
- 13Yantorno SE, Kremers WK, Ruf AE. MELD is superior to King's college and Clichy's criteria to assess prognosis in fulminant hepatic failure *Liver Transpl.* 2007; 13:822–8. [PubMed: 17539002]
- 14Wlodzimirow KA, Eslami S, Chamuleau RA. Prediction of poor outcome in patients with acute liver failure-systematic review of prediction models *PLoS One.* 2012; 7:e50952. [PubMed: 23272081]
- 15Atterbury CE, Maddrey WC, Conn HO. Neomycin-sorbitol and lactulose in the treatment of acute portal-systemic encephalopathy. A controlled, double-blind clinical trial *Am J Dig Dis.* 1978; 23:398–406. [PubMed: 354373]
- 16Ichai P, Samuel D. Etiology and prognosis of fulminant hepatitis in adults *Liver Transpl.* 2008; 14(Suppl 2):S67–79. [PubMed: 18825677]
- 17Steyerberg EW, Harrell FE Jr., Borsboom GJ. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis *J Clin Epidemiol.* 2001; 54:774–81. [PubMed: 11470385]
- 18Steyerberg EW, Bleeker SE, Moll HA. Internal and external validation of predictive models: a simulation study of bias and precision in small samples *J Clin Epidemiol.* 2003; 56:441–7. [PubMed: 12812818]
- 19Steyerberg EW, Vickers AJ, Cook NR. Assessing the performance of prediction models: a framework for traditional and novel measures *Epidemiology.* 2010; 21:128–38. [PubMed: 20010215]
- 20Harrell FE Jr., Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors *Stat Med.* 1996; 15:361–87. [PubMed: 8668867]
- 21DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach *Biometrics.* 1988; 44:837–45. [PubMed: 3203132]
- 22Katoonizadeh A, Decaestecker J, Wilmer A. MELD score to predict outcome in adult patients with non-acetaminophen-induced acute liver failure *Liver Int.* 2007; 27:329–34. [PubMed: 17355453]
- 23Zaman MB, Hoti E, Qasim A. MELD score as a prognostic model for listing acute liver failure patients for liver transplantation *Transplant Proc.* 2006; 38:2097–8. [PubMed: 16980011]
- 24Hadem J, Stiefel P, Bahr MJ. Prognostic implications of lactate, bilirubin, and etiology in German patients with acute liver failure *Clin Gastroenterol Hepatol.* 2008; 6:339–45. [PubMed: 18328438]
- 25Baquerizo A, Anselmo D, Shackleton C. Phosphorus as an early predictive factor in patients with acute liver failure *Transplantation.* 2003; 75:2007–14. [PubMed: 12829902]

Figure 1. ROC Curves for Development and Validation Cohorts

*Depicts the threshold for 80% SS in each cohort

Figure 2. Comparison of the ALFSG Model to the King's College Criteria and MELD Score, in Predicting 80% survival

Model c-statistics: ALFSG Model = 0.843; MELD = 0.717; King's College Criteria APAP (0.560), King's College Criteria Non-APAP (0.655)

Table 1.

Baseline Demographics and 21-day Outcomes of the Study Population (N=1974)

Variables [*]	N ^{**}	N (%) or Median (IQR)
Age (years)	1974	39 (29 – 51)
Gender	1974	
Male		600 (30.4)
Female		1374 (69.6)
Race	1974	
Caucasian		1531 (77.6)
African American		274 (13.9)
Other		169 (8.6)
ALF Etiology	1961	
Acetaminophen Overdose		933 (47.6)
Indeterminate		236 (12.0)
Drug Induced Liver Injury		215 (11.0)
Autoimmune Hepatitis		114 (5.8)
Ischemia		111 (5.7)
Hepatitis A		34 (1.7)
Other ^{***}		318 (16.2)
BMI	1563	26.8 (23.2 – 31.4)
Systolic Blood Pressure (mm Hg)	1962	124 (110 – 140)
Diastolic Blood Pressure (mm Hg)	1962	68 (58 – 79)
Bilirubin (mg/dL)	1941	6.9 (3.7 – 18.9)
INR	1875	2.7 (2.0 – 4.1)
Serum creatinine (mg/dL)	1962	1.6 (0.9 – 3.1)
Lactate (mmol/L)	993	4.4 (2.5 – 9.1)
Arterial Ammonia	428	93.5 (58.0 – 161.0)
Venous Ammonia	646	92.0 (61.0 – 140.0)
Phosphate	1708	3.0 (2.0 – 4.4)
Coma Grade	1922	
I or II		992 (52%)
III or IV		930 (48%)
Renal replacement therapy	1968	403 (21%)
Vasopressor Use	1968	366 (19%)
Mechanical Ventilation	1968	948 (48%)
Outcomes	1905	
Alive		1340 (70%)
Dead		565 (30%)
Liver Transplant	1917	431 (23%)
Spontaneous Survival [†]	1920	959 (50%)

Variables *	N **	N (%) or Median (IQR)
Listed for Liver Transplant	959	151 (16%)
Not Listed for Liver Transplant	959	799 (83%)
Non-spontaneous Survival ††	1920	961 (49%)
Listed for Liver Transplant	961	560 (58%)
Not Listed for Liver Transplant	961	398 (41%)

* Clinical variables represent values at the time of entry (Day 1) in the database.

** N provided for each variable to account for missing data.

*** Any ALF in pregnancy, Budd-Chiari syndrome, Hepatitis B, Hepatitis C, Hepatitis E, Mushroom intoxication, Shock/ischemia, Wilson's disease, Other Viruses, and Other.

† Listing status unknown for 9 subjects

†† Listing status unknown for 3 subjects

Table 2.

Univariate Analyses of Clinical Predictors of Spontaneous Survival in the Development Cohort

	Spontaneous Survival				p-value
	Yes(N=481)		No (N=482)		
Admission Variables	N [*]	N (%) or Median (IQR)	N [*]	N (%) or Median (IQR)	
Age (yrs.)	481	35.0 (28.0 – 48.0)	482	40.0 (28.0 – 53.0)	0.01
Gender	481		482		0.54
Male		144 (29.9)		153 (31.7)	
Female		337 (70.1)		329 (68.3)	
Race	481		482		0.001
Caucasian		398 (82.7)		357 (74.1)	
African American		62 (12.9)		80 (16.6)	
Other		21 (4.4)		45 (9.3)	
ALF Etiology	474		480		<0.0001
APAP		307 (64.8)		138 (28.8)	
Ischemia		39 (8.2)		20 (4.2)	
Hepatitis A		7 (1.5)		9 (1.9)	
Indeterminate		27 (5.7)		90 (18.8)	
DILI		33 (7.0)		75 (15.6)	
Autoimmune Hepatitis		12 (2.5)		39 (8.1)	
Other ^{**}		49 (10.3)		109 (22.7)	
Etiology Severity	474		480		<0.0001
Favorable		359 (75.7)		170 (35.4)	
Unfavorable		115 (24.3)		310 (64.6)	
Coma Grade	470		472		<0.0001
I or II		291 (61.9)		214 (45.3)	
III or IV		179 (38.1)		258 (54.7)	
Vasopressor Use	479		479		<0.0001
Yes		59 (12.3)		131 (27.4)	
No		420 (87.7)		348 (72.7)	
Mechanical Ventilation	479		479		<0.0001
Yes		184 (38.4)		266 (55.5)	
No		295 (61.6)		213 (44.5)	
Renal replacement therapy	479		479		0.003
Yes	82 (17.1)		119 (24.8)		
No	397 (82.9)		360 (75.2)		

	Spontaneous Survival				p-value
	Yes(N=481)		No (N=482)		
Admission Variables	N [*]	N (%) or Median (IQR)	N [*]	N (%) or Median (IQR)	
Bilirubin (mg/dL)	474	4.6 (2.3 – 9.2)	477	14.7 (5.5 – 24.6)	<0.0001
INR	453	2.3 (1.8 – 3.3)	457	3.0 (2.2 – 4.6)	<0.0001
Creatinine (mg/dL)	477	1.5 (0.8 – 3.1)	480	1.9 (1.0 – 3.2)	0.014
Lactate (mmol/L)	256	3.2 (1.9 – 5.5)	234	6.6 (3.9 – 13.6)	<0.0001
Phosphate (mmol/L)	422	2.7 (1.8 – 3.8)	413	3.6 (2.5 – 5.3)	<0.0001
Arterial Ammonia (pg/dL)	97	64.0 (39.0 – 109.0)	115	109.0 (68.0 – 196.0)	<0.0001
Venous Ammonia (pg/dL)	170	80.5 (49.0 – 117.0)	157	123.0 (79.0 – 177.0)	<0.0001
BMI	392	26.1 (22.9 – 30.5)	385	29.0 (23.4 – 32.2)	0.02
Systolic BP (mm Hg)	477	125.0 (110.0 – 140.0)	480	120.0 (108.0 – 137.0)	0.005
Diastolic BP (mmHg)	477	71.0 (60.0 – 80.0)	480	65.0 (55.0 – 77.0)	<0.0001

* N provided for each variable to account for missing data.

*** Any ALF in pregnancy, Budd-Chiari syndrome, Hepatitis B, Hepatitis C, Hepatitis E, Mushroom intoxication, Shock/ischemia, Wilson's disease, Other Viruses, and Other

Table 3.

Predicting 21-day Spontaneous Survival in Acute Liver Failure

	Predicted Probability of Spontaneous Survival at 21 Days [*]					
	0%–50%	50%–60%	60%–70%	70%–80%	80–90%	90–100%
Number of patients	931	138	132	179	217	166
Number of Spontaneous Survival	229	77	89	135	186	157
Observed Spontaneous Survival	24.6%	55.8%	67.4%	75.4%	85.7%	94.6%

^{*} Predicted SS is based on results from the ALFSG Model.

Table 4.

ALFSG Model Performance According to Hepatic Encephalopathy (HE) Grade and ALF Etiology (Validation Cohort)

HE Grade	Etiology	SS %	Total Subjects	Actual Equals Predicted Outcome	% Correctly Predicted (SS or non-SS)	Model Incorrectly Predicts Spontaneous Survival [†]	Model Incorrectly Predicts Negative Outcome (Death or LT) ^{††}
Mild	Favorable	85	209	135	64.6%	11 (5.3%)	63 (30.1%)
Deep	Favorable	55	278	157	56.5%	10 (3.6%)	111 (39.9%)
Mild	Unfavorable	33	228	160	70.2%	0 (0.0%)	68 (29.8%)
Deep	Unfavorable	21	170	135	79.4%	0 (0%)	35 (20.6%)
Total			885 [*]	587	66.3%	21 (2.4%)	277 (31.3%)

^{*} Predictor variables missing from 102/987 subjects leaving 885 for analysis.

[†] Subjects in which the model incorrectly predicts spontaneous survival when death or LT occurs.

^{††} Subjects in which the model incorrectly predicts the negative outcome (death or LT) when spontaneous survival occurs.