

# Comparison of Prognostic Models in Acute Liver Failure: Decision is to be Dynamic

Vandana Saluja<sup>1</sup>, Anamika Sharma<sup>2</sup>, Samba SR Pasupuleti<sup>3</sup>, Lalita G Mitra<sup>4</sup>, Guresh Kumar<sup>5</sup>, Prashant M Agarwal<sup>6</sup>

## ABSTRACT

**Background and aims:** Acute liver failure (ALF) is a rare disease entity with a high mortality. Management is dependent on accurate prognostication.

**Materials and methods:** One hundred consecutive patients presenting with ALF were prospectively evaluated. The King's college criteria (KCC), ALF early dynamic model (ALFED), sequential organ failure assessment (SOFA) score, and acute physiology and health evaluation II (APACHE II) scores were compared to predict mortality.

**Results:** There were significant differences in means of all the scores between survivors and nonsurvivors. The SOFA 48 hours had the highest area under receiver operating characteristic curve (AUC) (0.857) closely followed by the ALFED score (0.844). The optimal cutoff for the SOFA score at 48 hours to predict subsequent survival outcome is  $\geq 10$  and for the ALFED score is  $\geq 5$ . Sequential organ failure assessment 48 hours had a good sensitivity of 87%, and the ALFED score showed a good specificity of 84%. The decision curve analysis showed that between a threshold probability of 0.13 and 0.6, use of the SOFA score provided the maximum net benefit and at threshold probabilities of  $>0.6$ , the use of ALFED score provided the maximum clinical benefit.

**Conclusion:** Dynamic scoring results in better prognostication in ALF. The SOFA 48 hours and ALFED score have good prognostication value in nonacetaminophen-induced liver failure.

**Keywords:** Assessment, Liver injury, Net benefit, Scoring systems.

*Indian Journal of Critical Care Medicine* (2019): 10.5005/jp-journals-10071-23294

## INTRODUCTION

Determining prognosis is the deciding factor in the management of patients with ALF. Management challenges include offering a scarce organ with lifelong risk of immunosuppression vs continuing conservative medical management in the hope of native liver recovery.

Evaluation of prognostic models is influenced by the need for a strong positive or negative predictive accuracy.<sup>1</sup> Prognostication in ALF was introduced by the Kings College criteria in the year 1989.<sup>2</sup> Evaluation of these criteria in different studies has shown varying positive predictive values ranging from 70% to 100% and negative predictive values from 24% to 94%.<sup>2</sup> A meta-analysis has shown the criteria to have a sensitivity of 70% and specificity to be 80%.<sup>3</sup> Improvement in critical care management and better liver transplantation techniques have resulted in improved survival.<sup>4</sup> Therefore, subsequent studies have shown lower accuracy compared to the original study by O'Grady et al.<sup>2,5</sup>

The Clichy criteria were derived in 1986 for prognostication in patients with hepatitis B-related ALF. The degree of encephalopathy and decreased levels of Factor V, less than 20% (age  $<30$  years) and less than 30% (age  $>30$  years), were indicators of a poor outcome and hence signaled the need for a liver transplant.<sup>6,7</sup> Major disadvantages of this criteria are usefulness only in patients with hepatitis B and limited availability of Factor V measurement. Subsequent studies for search of prognostic indicators in ALF complicated by viral hepatitis concluded that raised intracranial pressure, prothrombin time more than 100 seconds on admission, age  $>50$  years, and jaundice encephalopathy interval adversely affected the outcome.<sup>8</sup>

The model for end-stage liver disease (MELD) score was proposed to estimate the survival of patients undergoing transjugular intrahepatic portosystemic shunt (TIPSS).<sup>9</sup> Later, it was adopted for

<sup>1,2,4,6</sup>Department of Anesthesia and Critical Care, Institute of Liver and Biliary Sciences, New Delhi, India

<sup>3</sup>Department of Statistics, Institute of Liver and Biliary Sciences, New Delhi, India

<sup>5</sup>Department of Research, Institute of Liver and Biliary Sciences, New Delhi, India

**Corresponding Author:** Lalita G Mitra, Department of Anesthesia Critical Care Medicine, Institute of Liver and Biliary Sciences, New Delhi, India, Phone: +91 9971792343, e-mail: lgmitra@hotmail.com

**How to cite this article:** Saluja V, Sharma A, Pasupuleti SSR, Mitra LG, Kumar G, Agarwal PM. Comparison of Prognostic Models in Acute Liver Failure: Decision is to be Dynamic. *Indian J Crit Care Med* 2019;23(12):574–581.

**Source of support:** Nil

**Conflict of interest:** None

organ allocation for liver transplant as a good predictor of 3-month mortality.<sup>10</sup> It was compared to the KCC and six prognostic indicators proposed by Dhiman et al.<sup>11</sup> They concluded that MELD score of  $\geq 33$  was the best discriminant between survivors and nonsurvivors. But it was inferior to the clinical prognostic indicators in patients with predominantly viral-induced fulminant hepatic failure. In this study, 52% patients had liver failure due to hepatitis B. Since 2007, hepatitis B vaccination has become a part of the universal immunization schedule in India and has resulted in reduction of infection.<sup>12,13</sup>

Most of the prognostic models available have been evaluated by measuring their discriminative ability, estimated by their c statistic. A c statistic can never have a value of 1, therefore there will be patients whose outcome cannot be predicted.<sup>3</sup> Measures of clinical utility have never been applied, to make them more meaningful.

Acute liver failure is a disease associated with fluctuations. Hence, there have been recent attempts of deriving dynamic models for better prognostication.<sup>14,15</sup> However, static and dynamic models have never been compared with respect to their net clinical benefit. In this study, we aimed at holistic comparison of scores, i.e., liver-derived scores vs intensive care unit (ICU) scores and static scoring vs dynamic scoring with respect to the clinical utility.

Institute of Liver and Biliary Sciences is a super specialty hospital dedicated to liver care in the country. As a result, it has been the referral unit for ALF and liver transplant in the country since 2009.

We follow a protocolized approach in the care of these patients. Most of the patients presenting are with nonparacetamol etiologies, mostly viral. We sought to evaluate already available prognostic scores. We selected four scores, out of which two, i.e., KCC<sup>2</sup> and the ALFED<sup>15</sup> model, have been derived and validated in these group of patients and the other two are ICU scores, i.e., SOFA score<sup>16</sup> and the APACHE II score,<sup>17</sup> and validated in patients with cirrhosis with acute deterioration.<sup>18</sup> The SOFA score was assessed at admission and at 48 hours.

## MATERIALS AND METHODS

Consecutive patients with ALF presenting to Institute of Liver and Biliary sciences from June 2014 to December 2017 were prospectively evaluated. Acute liver failure was defined as evidence of coagulation abnormality with international normalized ratio (INR) >1.5, any degree of mental alteration/encephalopathy in a patient without preexisting cirrhosis, and with an illness <26 weeks duration.

### INITIAL MANAGEMENT

Patients who developed or presented with encephalopathy at grade III or higher were intubated, sedated, and mechanically ventilated. Arterial blood gas parameters of pH lactate levels and partial pressure of oxygen (PO<sub>2</sub>) levels were used to support the decision for fluid resuscitation. Invasive hemodynamic monitoring with Flotrac assessment of fluid responsiveness was used to guide the fluid therapy. Coagulopathy was not corrected unless active bleeding was present. Norepinephrine was the primary vasopressor used and dobutamine was the primary inotropic agent with adjunctive use of intravenous low-dose hydrocortisone and vasopressin in patients not responsive to initial intravenous fluid therapy of 30 mL/kg. Indications for use of renal replacement therapy were acute kidney injury with anuria, relative oliguria, metabolic stabilization, control of acidosis, and hyperammonemia. Continuous renal replacement therapy was preferred. Sedation was achieved with fentanyl and propofol infusions, with use of atracurium for paralysis. Treatment for intracranial pressure crises was with bolus intravenous mannitol and hypertonic saline.

The severity of liver disease was evaluated with the King's College Hospital (KCH) criteria and the ALFED model. Acute physiology and health evaluation II was used for the classification of illness severity, and the SOFA score was used for grading organ dysfunction or failing organ systems. The KCH criteria were evaluated on admission after adequate resuscitation, SOFA score was calculated on admission and 48 hours subsequently (<https://www.mdcalc.com/sequential-organ-failure-assessment-sofa-score>), APACHE II severity score was calculated by using the calculators available (<https://reference.medscape.com/calculator/apache-ii-scoring-system?>) in the variables noted over 24 hours,

**Table 1:** Acute liver failure early dynamic scoring

Predictors of mortality based on variable dynamicity over 3 days	Score assigned
HE, persistent or progressed to $\geq 2$	2
INR, persistent or increased to $\geq 5$	1
Arterial ammonia, persistent or increased to level $\geq 123$ $\mu\text{mol/L}$	2
Serum bilirubin, persistent or increased to $\geq 15$ mg/dL	1

**Table 2:** Risk assessment based on the acute liver failure early dynamic score

ALFED score	Associated risk	Associated mortality (found by Kumar et al.) (%) <sup>15</sup>
0–1	Low	2.6
2–3	Moderate	33
4–6	High	88.5

and the ALFED score was assigned on the 3rd day, after noting the trend of encephalopathy, INR, ammonia, and bilirubin over the first 3 days (Table 1).

Based on the risk score and associated risk of mortality, patients could be stratified into three risk categories as follows (Table 2).

Finally, the mortality data were collected. The requirement of organ support in the form of mechanical ventilation, renal replacement, liver dialysis, and plasma exchange was further recorded. Survival was noted as survival to hospital discharge.

### Inclusion Criteria

Patients with ALF presenting for the first time in the medical ICU of our hospital.

### Exclusion Criteria

- Patients received with irreversible multiorgan dysfunction from outside hospitals.
- Patients leaving against medical advice.
- Patients receiving liver transplant.
- Patients with acute-on-chronic liver failure (ACLF).

### Statistical Analysis

For continuous variables like age, KCC, SOFA admission, SOFA 48 hours, APACHE II, and ALFED, descriptive statistics were presented in the form of mean  $\pm$  standard deviation. For categorical variables such as sex, descriptive statistics were present in the form of frequencies and percentages. All the statistical analyses were done using SAS University Edition. The Student's *t* test was used to test for a statistically significant difference between survivors and nonsurvivors in terms of their scores. Bivariate association between categorical variables was tested either by using the Chi-square test or by using the Fischer's exact test. Univariate logistic regression analyses were used to quantify the risk of death as a matter of various individual scores. For all the analyses, a *p* value of less than 0.05 was considered as statistically significant.

Performance of prognostic models was assessed by examining measures of discrimination, calibration, overall performance, and clinical utility.

Discrimination is the ability of the model to distinguish a patient with ALF with a higher risk of death from a patient with ALF with a lower risk of death or likely survival. This was examined by calculating

the area under receiver operating characteristic curve (AUC). A model with AUC between 0.7 and 0.8 is considered clinically useful, and a model with AUC between 0.8 and 0.9 has excellent accuracy.<sup>19</sup>

Calibration refers to the agreement between observed events and predictions (i.e., degree of correspondence between the predicted and observed mortality status). We used the Hosmer–Lemeshow test for this purpose. A high *p* value (close to 1) is considered a sign of good calibration.<sup>20</sup>

### Overall Performance

Nagelkerke *R*<sup>2</sup> was used to assess the goodness of fit or power of explanation of the model.

Clinical utility of model prediction was assessed using the decision curve analysis (DCA) proposed by Vickers and Elkin.<sup>21</sup> It integrates the impact of choices made by clinicians into analysis. A key concept of DCA is the threshold probability (Pt). Threshold probability is defined as “where the net benefit of treatment is the same as net benefit of avoiding treatment.”

Net benefit (NB) is the difference between the proportion of true positives and weighted proportion of false-positives for a given Pt. The decision curve is a plot of NB against a range of increasing threshold probabilities. The NB curves of all the five models (KCC, SOFA admission, SOFA 48 hours, APACHE II, and ALFED score) were compared with the NB curves of “treat all” (transplant all) and “treat none” (transplant none) strategies.

## RESULTS

A total of 205 patients with ALF were admitted to ICU of our institute, between January 2014 and December 2017.

Of these, 49 patients left the hospital against medical advice, 24 were received from other medical centers with irreversible organ failures, 26 underwent emergency liver transplantation, and six patients expired within 48 hours of admission.

The remaining 100 patients who met the selection criteria were considered for the present analysis. Out of 100 cases, 58 were of viral etiology and 20 were drug-induced, mostly due to antitubercular drugs. Five cases were autoimmune-mediated and in 17 patients we could not find the precipitating factor.

Out of 100 patients, 49 patients were male and 51 were female; 51 out of 100 patients survived. Demographic characteristics of survivors and nonsurvivors are shown in Table 3.

Out of 100 patients, 86 required mechanical ventilation. Of these 86 patients, 37 subsequently survived. There was a significant association between the requirement of mechanical ventilation and mortality.

Out of 100 patients, 20 underwent continuous renal replacement therapy and only one patient out of these survived. There was a significant association between receipt of renal replacement therapy and mortality.

Out of 100 patients, 21 underwent plasma exchange and 8 of them survived. There was no significant association between the receipt of plasmapheresis and mortality.

Out of 100 patients, 43 required vasopressor support in the form of noradrenaline, vasopressin, or both. Seven out of these 43 survived. There was a significant association between the need for vasopressor support and mortality.

### Distribution of Prognostic Scores

The distribution of the various scores by subsequent survival status of the patient has been shown in Figure 1. It shows that the scores were higher for patients who subsequently died than for those who have survived. Compared to other scores, the SOFA score at admission varies little by the subsequent survival status of patient. Table 4 shows that there are significant differences in means of all the scores between survivors and nonsurvivors.

Table 5 shows univariate logistic regression results. It shows that all the considered scores are statistically significant predictors of subsequent survival status but with differential overall performance and calibration. Odds of subsequent death are found to increase by 157% with one unit increase in KCC. The same for SOFA at admission, SOFA at 48 hours, APACHE II, and ALFED are 32, 74, 15, and 253%, respectively.

### Discrimination Measures

Results of the receiving operating characteristics (ROC) curve analysis are shown in Table 6. It shows that the SOFA score at 48 hours has the highest discriminative score, with AUC being 0.857 (95% CI 0.782–0.931).

We followed the Youden method (the value of test score where (sensitivity + specificity – 1) is the maximum) to choose the optimal cutoff point. Using this criterion, we found the optimal cutoff point for the SOFA score at 48 hours to predict subsequent survival outcome is  $\geq 10$ , i.e., the ALF patient with a SOFA score at 48 hours of greater than or equal to 10 is predicted to die and with score less than 10 is predicted to survive. Similarly, we found the optimal cutoff point for the ALFED score is  $\geq 5$ , for the KCC score it is  $\geq 3$ , for APACHE II it is  $\geq 15$ , and for SOFA at admission it is  $\geq 10$ . Various predictive evaluation measures associated with these cutoff points are shown in Table 6.

Binary prediction criteria based on SOFA at 48 hours have the highest accuracy with 81% of the patients being correctly predicted for their actual subsequent survival status. It correctly predicted 43 out of 49 patients who have died but only correctly predicted 38 cases out of 51 survived patients. Binary decision criteria based on the ALFED score, on the other hand, had an overall accuracy score of 79% but had better performance while predicting true survivals (43 out of 51 patients survived were correctly predicted) than those who died (36 out of 49). Therefore, SOFA at 48 hours has the highest sensitivity (0.88), highest negative predictive value (NPV, 0.864), and least negative likelihood ratio (LR, 0.164). Specificity is highest for criteria based on the ALFED score (0.843).

Therefore, SOFA at 48 hours and the ALFED score were the best scores to discriminate the actual survival status of ALF patients. Sequential organ failure assessment at 48 hours has highest accuracy, highest sensitivity, and the ALFED model had highest specificity.

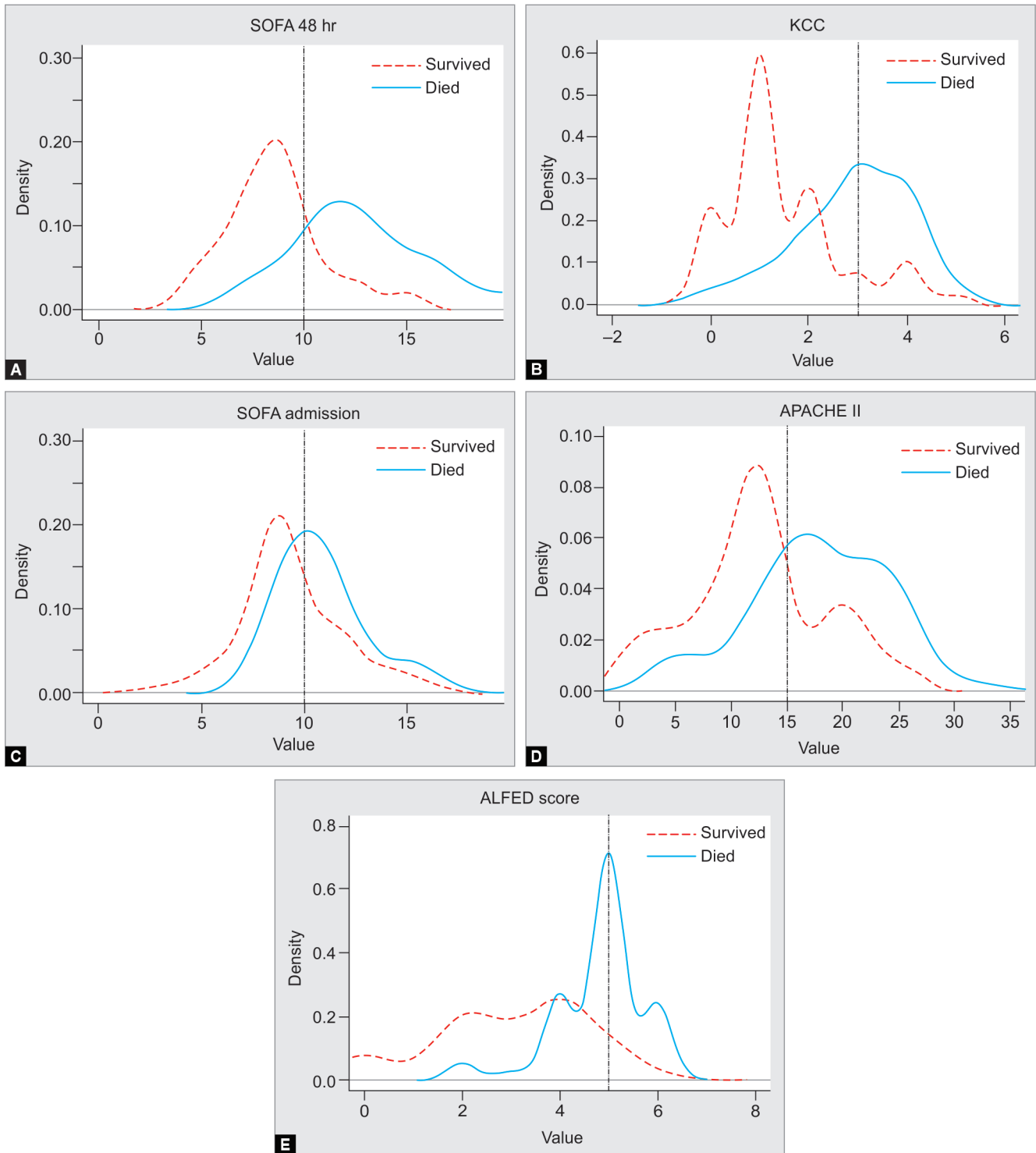
### Calibration Measures

Hosmer and Lemeshow test results from Table 5 show that the ALFED has the highest calibration (*p* value = 0.690) and is

**Table 3:** Demographic characteristics of present study respondents by their survival status

Characteristic variable	Dead (49)	Survived (51)	<i>p</i> (<0.05 = S/>0.05 = NS)
Age (years) ± SD	28.94 ± 14.77	22.03 ± 13.78	0.017
Sex			
Male (49)	21 (42.9%)	28 (54.9%)	0.23
Female (51)	28 (57.1%)	23 (45.1%)	





**Figs 1A to E:** Distribution of the scores by the subsequent survival status of the patients. *Note:* The vertical dotted line in each graph indicates the best cut off point (where the sum of sensitivity and specificity is maximum)

followed by SOFA at 48 hours ( $p$  value = 0.407), SOFA at admission ( $p$  value = 0.233), KCC ( $p$  value = 0.1216), and APACHE II (0.010).

**Overall Performance**

Overall performance as assessed by Nagelkerke  $R$ -square (Table 5) is highest for SOFA at 48 hours (0.478) and is closely followed by the ALFED score (0.463). These results suggest SOFA at 48 hours has the best overall performance and is followed by the ALFED model.

**Clinical Utility**

Prognostication in a patient with ALF involves predicting a poor outcome, i.e., death, hence offering the patient the only definitive treatment, i.e., transplant vs opting for a conservative management with the hope of native liver recovery.

Both the decisions are based on prognostic models and carry a trade-off between benefit and harm.

The benefit of going for a transplant would be offering a chance of survival to high-risk patients, but at the same time few patients would be put at risk of unnecessary surgery and lifelong immune suppression, those who were overtreated.

At the same time, conservative management would deny high-risk patients' early intervention (in the form of lifesaving transplant), but low-risk patients would be saved from unnecessary surgery.

The probability threshold represents this benefit to harm trade-off. Probability threshold is the risk threshold of a clinician on which treatment is decided.

In simpler words, the clinician has to decide the number of unnecessary episodes of early management (i.e., liver transplant) that she or he would be willing to recommend for a chance at preventing one death.

Figure 2 shows the NB of banking on the considered prognostic scores. The NB is the highest if the decision to treat aggressively (transplant) is taken based on SOFA at 48 hours, in the threshold probability range 0.12–0.58. After this threshold probability range, taking decision based on the ALFED score has the highest NB.

Table 7 shows the benefit of using various prognostic scores, in terms of reduction in number of aggressive interventions (unnecessary liver transplantations) per 100 ALF patients, at various threshold probabilities. At the very low threshold probabilities (i.e., in a situation where a doctor recommends or considers to aggressively treat more than eight patients to prevent one death), none of the considered scores have any benefit over the "treating all" option. But if the threshold probability is between 0.13 and 0.6 (i.e., to prevent one death among ALF patients if the doctor prefers to aggressively treat 1.5 to 8 ALF patients), then taking decisions based on the SOFA at 48 hours criteria would result in reduced number of unnecessary aggressive interventions. For instance, at a threshold probability of 0.4, taking decisions based on the SOFA at 48 hours criteria would result in reduction of 29 unnecessary aggressive interventions per 100 patients, as compared to the "treat all" patients strategy, while the same based on ALFED, KCC, APACHE II, and SOFA at admission criteria are 24, 21, 20, and 12, respectively.

At the higher threshold probabilities (i.e., to prevent one death among ALF patients if the doctor recommends to aggressively treat less than 1.5 but more than one patient), banking on the ALFED model-based criteria would result in averting more number of unnecessary interventions as compared to others.

**Table 4:** Distribution of various prognostic scores by survival status of patients considered in this study

	Mortality (n = 49)		Survival (n = 51)	
	Mean ± SD	Mean ± SD	Mean ± SD	p value
KCC	2.94 ± 1.34	1.47 ± 1.20		<0.001
SOFA admission	10.88 ± 2.20	9.39 ± 2.49		0.002
SOFA 48 hours	12.74 ± 3.24	8.61 ± 2.34		<0.001
APACHE II	17.85 ± 6.18	12.58 ± 6.01		<0.001
ALFED score	4.82 ± 0.90	3.02 ± 1.55		<0.001

**Table 5:** Logistic regression results

Criteria	Odds ratio	95% CI	p value	Nagelkerke R square	Hosmer and Lemeshow goodness-of-fit test	
					Chi-square	p
KCC	2.565	1.732–3.797	<0.0001	0.3497	5.8030	0.1216
SOFA admission	1.321	1.092–1.598	0.0042	0.1236	6.8337	0.2333
SOFA 48 hours	1.739	1.392–2.171	<0.0001	0.4777	8.2788	0.4067
APACHE II	1.151	1.069–1.239	0.0002	0.2116	16.8324	0.0099
ALFED score	3.532	2.080–5.996	<0.0001	0.4632	2.2503	0.6898

KCC, Kings college criteria; SOFA adm, sequential organ failure assessment score at admission; SOFA 48 hours, sequential organ failure assessment score at 48 hours; APACHE II, acute physiology and chronic evaluation; ALFED, acute liver failure early dynamic model

**Table 6:** Best possible cutoff for various considered scores, and their various diagnostic characteristics

Test	KCC	SOFA admission	SOFA 48 hours	APACHE II	ALFED score
AUC	0.803	0.687	0.857	0.741	0.844
95% CI for AUC	(0.714, 0.892)	(0.583, 0.791)	(0.782, 0.931)	(0.643, 0.840)	(0.770, 0.918)
Optimal cut point for predicting death	≥3	≥10	≥10	≥15	≥5
True-positives	34	35	43	37	36
True-negatives	43	33	38	38	43
False-positives	8	18	13	13	8
False-negatives	15	14	6	12	13
Accuracy	0.770	0.680	0.810	0.750	0.790
Sensitivity	0.694	0.714	0.878	0.755	0.735
Specificity	0.843	0.647	0.745	0.745	0.843
Positive predictive value	0.810	0.660	0.768	0.740	0.818
Negative predictive value	0.741	0.702	0.864	0.760	0.768
LR positive	4.423	2.024	3.443	2.962	4.684
LR negative	0.363	0.442	0.164	0.329	0.315

LR, likelihood ratio



**DISCUSSION**

The main cause of death in fulminant liver failure is brain herniation due to cerebral edema.<sup>22</sup>

Spontaneous survival has increased from 17% to 48%. Survival after emergency liver transplant has improved from 56% to 86%. Proportion of patients with intracranial hypertension (ICH) have reduced from 76% (1984–88) to 20% (2004–2008). Also there has been a reduction in mortality due to ICH from 95% to 55%.<sup>4</sup>

But all patients may not have access to the superspeciality care required for this condition in a developing country like India.<sup>23</sup> Therefore, appropriate prognostication may help in early referral or successful conservative management with aid of telemedicine

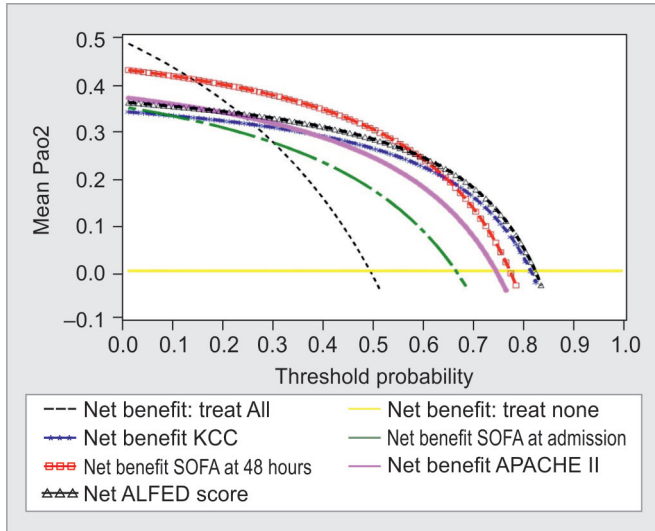
wherever possible if robust methods for evaluating such patients are available.<sup>24,25</sup>

Though posttransplantation survival has improved,<sup>4</sup> a word of caution is advisable in the setting of living donor liver transplantation. Problems of possible donor coercion and expedited evaluation leading to increased complication rates of 34% even in centers with great expertise are other factors to be taken into account, making accurate prognostication even more mandatory.<sup>26</sup>

Prognostication in ALF has undergone recent changes with development of organ failure scores and dynamic scoring systems for this group of patients.<sup>27,28</sup> Since ALF is a critical illness with multiorgan involvement, we sought to compare the already existing ICU scores with the liver-derived scores. The ALFED model is a dynamic score, derived from patients with ALF with predominant viral etiology.<sup>15</sup> Shalimar et al. compared the ALFED score with other liver-derived scores like MELD, MELD Na (MELD-Sodium), KCH, and CLIF-ACLF scores.<sup>29</sup> They recorded the scores on day 1 and day 3 and concluded that the ALFED model performed the best in predicting outcomes in patients with viral etiology. They used the CLIF-ACLF score that has been derived from the SOFA score. Our aim was to compare the static scores with the dynamic scores, out of which we selected two dynamic scores: SOFA that is an ICU-related score and ALFED that is a liver-derived score. We have assessed the APACHE II and KCC scores on day 1 after adequate resuscitation and stabilization of the patient.

Our results showed that the younger age group had significant survival. This finding is similar to Schiødt et al. where they showed that spontaneous survival in the nonacetaminophen category was higher in the age group of 20–24 years than the age group of 25–29 years.<sup>30</sup>

Cholongitas et al. compared the KCC, MELD, SOFA, and APACHE II scores in ALF due to paracetamol-induced injury.<sup>31</sup> They evaluated



**Fig. 2:** Decision curves showing the net clinical benefit of the prognostic scores (treat all means transplant all; treat none means transplant none)

**Table 7:** Benefit of using various criteria, in terms of reduction in number of unnecessary aggressive interventions (like liver transplantations) per 100 ALF patients, instead of treating all patients, at various threshold probabilities

Threshold probability	Its meaning—number of liver transplantations doctor would recommend to prevent one death	Net benefit of treating all	Reduction in unnecessary aggressive interventions per 100 ALF patients, if we make decisions based on binary decision criteria on				
			KCC	SOFA at admission	SOFA at 48 hours	APACHE II	ALFED score
0.1	10.00	0.433	-92	-93	-16	-70	-74
0.15	6.67	0.400	-42	-46	4	-30	-31
0.2	5.00	0.363	-17	-23	14	-10	-9
0.25	4.00	0.320	-2	-9	20	2	4
0.3	3.33	0.271	8	0	24	10	13
0.35	2.86	0.215	15	7	27	16	19
0.4	2.50	0.150	21	12	29	20	24
0.45	2.22	0.073	25	16	31	23	27
0.5	2.00	-0.020	28	19	32	26	30
0.55	1.82	-0.133	31	22	33	28	32
0.6	1.67	-0.275	33	24	34	30	34
0.65	1.54	-0.457	35	25	35	32	36
0.7	1.43	-0.700	37	27	35	33	37
0.75	1.33	-1.040	38	28	36	34	39
0.8	1.25	-1.550	39	30	37	35	40
0.85	1.18	-2.400	40	31	37	36	41
0.9	1.11	-4.100	41	31	37	37	42
0.95	1.05	-9.200	42	32	38	37	42

125 patients and included those patients who were transplanted as well. They considered death and transplant as an equivalent outcome. Our study represents an uninterrupted natural history of ALF in the liver transplantation era. We believe the need for transplant and death cannot be considered equivalent. The transplant cohort is composed of patients who might have survived or died if liver transplant had not interrupted their natural course.

Studies to compare prognostic models or to find new markers have mainly been done in acetaminophen-induced liver failure. We have evaluated existing models in patients with nonacetaminophen etiology of ALF. Cholongitas et al.<sup>31</sup> showed that the SOFA score provided the best discriminative ability in acetaminophen-induced liver failure.

Mitchell et al.<sup>32</sup> compared the KCC criteria with the APACHE II score in 102 patients with acetaminophen-induced liver failure. They concluded that the APACHE II score of >15 helped in early identification of high-risk patients with acetaminophen-induced liver failure when compared to the KCC. We found the ALFED and SOFA 48 hours to be the best models. In our study, the APACHE II score of 15 and above was found to be useful. However, it was inferior to the KCC score. Our patient group mainly comprised of nonacetaminophen etiology, which could explain the different result.

We found a significant association between the need of organ supports like ventilator support, renal replacement therapy, and vasopressors with mortality. Need for ventilator support in ALF is mainly dictated by the presence of cerebral edema that correlates well with the ammonia levels.<sup>33</sup> Also presence of acute kidney injury and failure of ammonia clearance predict a high mortality.<sup>34,35</sup> All these factors indicate the severity and probable irreversibility of the disease itself, which contributed to the mortality in our study.

Plasma exchange has shown to reduce the innate immune activation and subsequent improvement in multiorgan dysfunction in ALF. This results in improved outcomes with increased transplant-free survival.<sup>36</sup> However, we did not find any such benefit in our study, as our study was not designed for this purpose.

Our study is the first to analyze the net clinical benefit of different prognostic scores in the setting of ALF. Traditional methods of performance like sensitivity, specificity, and area under curve measure the diagnostic accuracy of one prediction model against another. However, accuracy does not guarantee improved decision making. Statistical measures of predictive accuracy such as discrimination, calibration, and model fit are difficult for clinicians to interpret. The decision curve analysis estimates the NB of basing a clinical decision on a patient's prognostic score and compares this with other decision-making strategies.<sup>37</sup>

Our results show that SOFA 48 hours and the ALFED model have performed the best with almost similar AUC and overall performance measures. While SOFA 48 hours has shown good sensitivity, the ALFED model has shown good specificity. These results still do not clearly indicate to a clinician, as to when one score should be preferred over the other.

The threshold probability takes into account the treating preferences of individual clinicians. A clinician with low probability threshold (<50%) weighs the consequence of undertreatment more heavily than consequences of unnecessary treatment.

A clinician with a high probability threshold (>50%) weighs the consequence of unnecessary treatment more heavily than the consequences of undertreatment.<sup>37</sup>

The decision curve analysis has built the bridge between the mathematical models and clinical utility where it has shown that at

low probability thresholds the use of SOFA 48 hours would provide the maximum clinical benefit with avoidance of maximum number of unnecessary transplants, whereas at high probability thresholds the use of the ALFED score will provide the maximum benefit.

Drawbacks of the study are its observational nature coming from a single center. We suggest further validation of the results obtained.

## CONCLUSION

Given the heavy costs associated with intervention of ALF patients and the limited number of medical experts available for ALF intervention in India, decision for aggressive intervention should be taken based on either SOFA 48 hours or the ALFED score depending on the individual clinician treating thresholds.

Evaluation of prognostic models in such heavy stake settings should be done with robust methods like the decision curve analysis to provide a clear picture.

## ACKNOWLEDGMENTS

Authors are thankful to Dr Andrew J Vikers and Dr Daniel Sjoberg, from Memorial Sloan Kettering Cancer Center, New York, USA, for sharing their Decision Curve Analysis SAS macro with us.

## REFERENCES

- O'Grady JG. Prognostication in acute liver failure: a tool or an anchor? *Liver Transpl* 2007;13(6):786–787. DOI: 10.1002/lt.21159.
- O'Grady JG, Alexander GJ, Hallyar KM, Williams R. Early indicators of prognosis in fulminant hepatic failure. *Gastroenterology* 1989;97(2):439–445. DOI: 10.1016/0016-5085(89)90081-4.
- Mishra A, Rustgi V. Prognostic models in acute liver failure. *Clin Liver Dis* 2018;22(2):375–388. DOI: 10.1016/j.cld.2018.01.010.
- Bernal W, Hyyrylainen A, Gera A, Audimoolam VK, McPhail MJ, Auzinger G, et al. Lessons from look-back in acute liver failure? A single centre experience of 3300 patients. *J Hepatol* 2013;59(1):74–80. DOI: 10.1016/j.jhep.2013.02.010.
- 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(1):62–68. DOI: 10.1016/S0168-8278(97)80010-4.
- Bernuau J, Goudeau A, Poynard T, Dubois F, Lesage G, Yvonne B, et al. Multivariate analysis of prognostic factors in fulminant hepatitis B. *Hepatology* 1986;6(4):648–651. DOI: 10.1002/hep.1840060417.
- Bernuau J, Benhamou JP. Fulminant and subfulminant liver failure. In: McIntyre N, Benhamou JP, Bircher J, Rizzeto M, Rodes J, ed. *Oxford Textbook of Clinical Hepatology*. Oxford: Oxford Medical Publications; 1991. pp. 923–942.
- Dhiman RK, Seth AK, Jain S, Chawla YK, Dilawari JB. Prognostic evaluation of early indicators in fulminant hepatic failure by multivariate analysis. *Dig Dis Sci* 1998;43(6):1311–1316. DOI: 10.1023/A:1018876328561.
- Malinchoc M, Kamath PS, Gordon FD, Peine CJ, Rank J, ter Borg PC. A model to predict poor survival in patients undergoing transjugular intrahepatic portosystemic shunts. *Hepatology* 2000;31(4):864–871. DOI: 10.1053/he.2000.5852.
- Wiesner R, Edwards E, Freeman R, Harper A, Kim R, Kamath P, et al. Model for end-stage liver disease (MELD) and allocation of donor livers. *Gastroenterology* 2003;124(1):91–96. DOI: 10.1053/gast.2003.50016.
- Dhiman RK, Jain S, Maheshwari U, Bhalla A, Sharma N, Ahluwalia J, et al. Early indicators of prognosis in fulminant hepatic failure: an assessment of the model for end-stage liver disease (MELD) and king's college hospital criteria. *Liver Transpl* 2007;13(6):814–821. DOI: 10.1002/lt.21050.

12. Lahariya C, Subramanya BP, Sosler S. An assessment of hepatitis B vaccine introduction in India: lessons for roll out and scale up of new vaccines in immunization programs. *Indian J Public Health* 2013;57(1):8–14. DOI: 10.4103/0019-557X.111357.
13. Aggarwal R, Babu JJ, Hemalatha R, Reddy AV, Sharma D, Kumar T. Effect of inclusion of hepatitis B vaccine in childhood immunization program in India: a retrospective cohort study. *Indian Pediatr* 2014;51(11):875–889. DOI: 10.1007/s13312-014-0520-y.
14. Bernal W, Williams R. Beyond KCH selection and options in acute liver failure. *Hepatol Int* 2018;12(3):204–213. DOI: 10.1007/s12072-018-9869-7.
15. Kumar R, Shalimar, Sharma H, Goyal R, Kumar A, Khanal S, et al. Prospective derivation and validation of early dynamic model for predicting outcome in patients with acute liver failure. *Gut* 2012;;61(7):1068–1075. DOI: 10.1136/gutjnl-2011-301762.
16. Vincent JL, Moreno R, Takala J, Willatts S, De Mendonca A, Bruining H, et al. For working group on sepsis-related problems of the European society of intensive care medicine. The SOFA (sepsis-related organ failure assessment) score to describe organ dysfunction/failure. *Intensive Care Med* 1996;22(7):707–710. DOI: 10.1007/BF01709751.
17. Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. *Crit Care Med* 1985;13(10):818–829. DOI: 10.1097/00003246-198510000-00009.
18. Cholongitas E, Senzolo M, Patch D, Kwong K, Nikolopoulou V, Leandro G, et al. Risk factors, sequential organ failure assessment and model for end-stage liver disease scores for predicting short term mortality in cirrhotic patients admitted to intensive care unit. *Aliment Pharmacol Ther* 2006;23(7):883–893. DOI: 10.1111/j.1365-2036.2006.02842.x.
19. Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology* 1983;148(3):839–843. DOI: 10.1148/radiology.148.3.6878708.
20. Lemeshow S, Hosmer DW Jr. A review of goodness of fit statistics for use in the development of logistic regression models. *Am J Epidemiol* 1982;115(1):92–106. DOI: 10.1093/oxfordjournals.aje.a113284.
21. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making* 2006;26(6):565–574. DOI: 10.1177/0272989X06295361.
22. Butterworth RF. Pathogenesis of hepatic encephalopathy and brain edema in acute liver failure. *J Clin Exp Hepatol* 2015;5(Suppl 1):S96–S103. DOI: 10.1016/j.jceh.2014.02.004.
23. Panagariya A. The challenges and innovative solutions to rural health dilemma. *Ann Neurosci* 2014;21(4):125–127. DOI: 10.5214/ans.0972.7531.210401.
24. Bassi A, John O, Praveen D, Maulik PK, Panda R, Jha V. Current status and future directions of mHealth interventions for health system strengthening in India: systematic review. *JMIR Mhealth Uhealth* 2018;6(10):e11440. DOI: 10.2196/11440.
25. Sahu M, Grover A, Joshi A. Role of mobile phone technology in health education in Asian and African countries: a systematic review. *Int J Electron Healthc* 2014;7(4):269–286. DOI: 10.1504/IJEH.2014.064327.
26. Mendizabal M, Silva MO. Liver transplantation in acute liver failure: a challenging scenario. *World J Gastroenterol* 2016;22(4):1523–1531. DOI: 10.3748/wjg.v22.i4.1523.
27. Bernal W, Wang Y, Maggs J, Willars C, Sizer E, Auzinger G, et al. Development and validation of a dynamic outcome prediction model for paracetamol-induced acute liver failure: a cohort study. *Lancet Gastroenterol Hepatol* 2016;1(3):217–225. DOI: 10.1016/S2468-1253(16)30007-3.
28. Figorilli F, Putignano A, Roux O, Housel-Debry P, Francoz C, et al. Development of an organ failure score in acute liver failure for transplant selection and identification of patients at high risk of fatality. *PLoS ONE* 2017;5(12):12. DOI: 10.1371/journal.pone.0188151.
29. Shalimar, Sonika U, Kedia S, Mahapatra SJ, Nayak B, Yadav DP, et al. Comparison of dynamic changes among various prognostic scores in viral hepatitis-related acute liver failure. *Ann Hepatol* 2018;17(3):403–412. DOI: 10.5604/01.3001.0011.7384.
30. Schiødt FV, Chung RT, Schilsky ML, Hay JE, Christensen E, et al. Acute liver failure study group. Outcome of acute liver failure in the elderly. *Liver Transpl* 2009;15(11):1481–1487. DOI: 10.1002/lt.21865.
31. Cholongitas E, Theocharidou E, Vasiopoulou P, Betrosian A, Shaw S, Patch D, et al. 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(4):405–412. DOI: 10.1002/lt.23370.
32. Mitchell I, Bihari D, Chang R, Wendon J, Williams R. Earlier identification of patients at risk from acetaminophen-induced acute liver failure. *Crit Care Med* 1998;26(2):279–284. DOI: 10.1097/00003246-199802000-00026.
33. Scott TR, Kronsten VT, Hughes RD, Shawcross DL. Pathophysiology of cerebral oedema in acute liver failure. *World J Gastroenterol* 2013;19(48):9240–9255. DOI: 10.3748/wjg.v19.i48.9240.
34. Coelho S, Fonseca JN, Gameiro J, Jorge S, Velosa J, Lopes JA. Transient and persistent acute kidney injury in acute liver failure. *J Nephrol* 2018;32(2):289–296. DOI: 10.1007/s40620-018-00568-w.
35. Deep A, Stewart CE, Dhawan A, Douiri A. Effect of continuous renal replacement therapy on outcome in pediatric acute liver failure. *Crit Care Med* 2016;44(10):1910–1919. DOI: 10.1097/CCM.0000000000001826.
36. Larsen FS, Schmidt LE, Bernsmeier C, Rasmussen A, Isoniemi H, Patel VC, et al. High-volume plasma exchange in patients with acute liver failure: an open randomised controlled trial. *J Hepatol* 2016;64(1):69–78. DOI: 10.1016/j.jhep.2015.08.018.
37. Traeger AC, Hübscher M, McAuley JH. Understanding the usefulness of prognostic models in clinical decision-making. *J Physiother* 2017;63(2):121–125. DOI: 10.1016/j.jphys.2017.01.003.