Japanese Journal of Gastroenterology and Hepatology

Research Article ISSN: 2435-1210 | Volume 8

# In-Hospital Mortality of Patients with Liver Failure: A Propensity Score Matched Cohort Study

**Gong R[1], Wang M[1], Hu T[2] and Han X[3*]**

[1]Medical College of Qinghai University, Xining, Qinghai, 810016, P. R. China

[2]Department of Cardiology, The Second Affiliated Hospital, Chongqing Medical University, Chongqing, 400001, China

[3]Renal Department, Dezhou People's Hospital, Dezhou, Shandong Province, 25300, China

**[*]Corresponding author:**

Xiuxia Han,
Renal Department, Dezhou People's Hospital,
Dezhou, Shandong Province, 25300, China,
Tel: 0086 15615185106,
E-mail: 15615185106@163.com;
4728955@qq.com

Received: 11 Jan 2022
Accepted: 31 Jan 2022
Published: 04 Feb 2022
J Short Name: JJGH

**Copyright:**

©2022 Han X. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and build upon your work non-commercially.

**Citation:**

Han X, In-Hospital Mortality of Patients with Liver Failure: A Propensity Score Matched Cohort Study. Japanese J Gstro Hepato. 2022; V8(4): 1-9

**Keywords:**

MIMIC-IV; MELD; PSM; Liver failure; Prediction

**[&]Author contribution:**

Gong R, Wang M and these authors are contributed equally to this work

**Abbreviations**

MELD: Model for end-stage liver disease; LODS: Logical Organ Dysfunction System; SAPS II: Simplified Acute Physiology Score; OASIS: Oxford Acute Disease Severity Score; DCA: decision curve analysis; NIH: National Institutes of Health; ICU: Intensive Care Units; M±SD: mean ± standard deviation; IQR: Interquartile Range; ROC: Receiver Operating Characteristic; AUC: area under curves; PSM: propensity score matching.

## 1. Abstract

**1.1. Background:** Model for End-stage Liver Disease (MELD) has been considered as a poor prognostic indicator for patients with liver failure in most previous studies, however, some studies suggest that MELD is not ideal in clinical application. Therefore, new studies on MELD and poor outcomes in patients with liver failure are still necessary. We investigated whether MELD is a good predictor for in-hospital mortality of patients with liver failure based on the MIMIC-IV database.

**1.2. Method:** Four common clinical severe scores were compared. The subjects' operating characteristic curves of the four scores were drawn before and after the propensity score matching (PSM), and the areas under the curves were calculated and compared. In addition, the DCA curve was used to assess the difference in clinical benefit of each scoring system.

**1.3. Results:** Before and after PSM, MELD score was not the best among the four scores, and MELD score had the lowest sensitivity (58.64% and 54.91% before and after matching, respectively) and Youden's index (0.2227 and 0.2079 before and after matching, respectively).

**1.4. Conclusion:** Although the MELD score has made good progress in other clinical aspects, the prediction of poor prognosis in patients with liver failure still deserves improvement.

## 2. Background

Liver failure is a severe life-threatening disease, caused by numerous factors (including parasites, viral infection, drug abuse, bacteria, and bad lifestyle, etc.) that seriously damage the liver function, which leads to the disorder of liver physiology function[1-3]. Most cases are due to chronic hepatitis B virus (in developing countries) and drug abuse (in developed countries) [3-5]. Clinically, it is divided into explosive liver failure and chronic liver failure according to its onset rate [1, 4]. At present, liver failure has become a global public health problem [6]. The disease is known for its high fatality rate and high resource cost [6]. With advances in critical care medicine, the mortality rate has declined over the past decade, but it is still far from rosy, accounting for more than 50 percent of deaths [7]. The survival of a large proportion of patients could be improved by better understanding the early progression of the disease and timely intervention.

Model for End-stage Liver Disease (MELD) was originally created to predict survival in patients with liver disease with complications of portal hypertension and has subsequently been validated as an accurate predictor of survival in patients with different advanced liver diseases [8-11]. The score worked well within most patients. However, the survival rate of some patients with liver disease does not rest with the severity of the disease therefore the parameters of MELD are not disordered [12]. Consequently, it is unknown whether MELD score is appropriate for short-term survival in all patients with liver failure. In addition, Logical Organ Dysfunction System (LODS) [13], Simplified Acute Physiology Score (SAPS II) [14], and Oxford Acute Severity of Illness Score (OASIS) [15] are also widely used to assess critically ill outcomes in clinical practice. Here, we investigated the association of four scoring systems (MELD, LODS, SAPS II, and OASIS) with in-hospital mortality in patients with liver failure based on the Medical Information Mart for Intensive Care-IV (MIMIC-IV) database [16]. Furthermore, whether MELD score was a better predictor of in-hospital mortality in patients with liver failure among the four scores was determined. Considering that the factors of in-hospital death in patients with liver failure may be affected by other confounding factors, we conducted a paired comparison between patients who died and those who did not, to ensure the stability of the results. Particularly, we wanted to discuss the net benefit of the four scoring systems with respect to in-hospital mortality in patients with liver failure using decision curve analysis (DCA) [17], a suitable method for assessing alternative diagnostic and prognostic strategies.

## 3. Methods

### 3.1. Database

The MIMIC-IV database (https://mimic.mit.edu/) is based on the success of the MIMIVC-Ⅲ database [16]. The MIMIC-IV database contains factual hospitalization information for patients admitted to a tertiary Academic Medical Center in Boston, Massachusetts, USA, from 2008 to 2019. The researchers must be completed the "Protect Human Study Participants" exam and sign a data using agreement on the National Institutes of Health (NIH) website before granting access. The MIMIC-IV database contains detailed information about patients' clinical care. To protect patient privacy, all dates in the database have been changed and overall shifted, including patients' dates of birth, hospitalization, and discharge (e.g., December 12, 2008, expressed as December 12, 2108 after 100 years of overall shift). The patients in the database are anonymous, so there is no need for informed consent or ethical review. Author Tianyang Hu obtained the access and extraction rights of MIMIC-IV database (Record ID: 37474354).

### 3. 2. Population of Study and Data Extraction

All Intensive Care Units (ICU) patients diagnosed with liver failure were screened and identified by "long-title" in table "d-icd-diagnoses" in MIMIC-IV database (version 1.0). Since a patient may be repeatedly admitted to ICU, we only bring into each patient's first admission to the ICU. Navicat Premium 15.0 software was used to extract the following data from MIMIC-IV database for included patients, age, gender, race, start time of admission to ICU, the hospitalization time, Charison Comorbidity Index (a comprehensive measure of congestive heart failure, peripheral vascular disease, cerebrovascular disease, chronic lung disease, rheumatism, peptic ulcer disease, liver disease, diabetes, kidney disease, HIV/AIDS and other complications of index), LODS score, OASIS score, SAPS II score and MELD score.

### 3.3. Statistical Analysis

The Kolmogorov-Smirnov test was used to assess the normality of continuous variables. The normal distribution variables were expressed as mean ± standard deviation (M±SD), and independent sample T-test was used for comparison. If the normal distribution is not followed, continuity variables were expressed as median and interquartile range (IQR), and Wilcoxon rank-sum test was used for comparison. Dichotomous variables were expressed as numbers and percentages, and using chi-square test for comparison. Binomial logistic regression analysis was conducted to assess whether MELD was an independent risk factor for in-hospital death in patients with liver failure in intensive care after adjustment for potential confounders. The Receiver Operating Characteristic (ROC) curves of the four scoring system were plotted, and the area under curves (AUC) were calculated in this study to assess the predictive value of these scoring systems for nosocomial death in patients with liver failure ICU. DCA analysis was performed to evaluate the difference in clinical benefit of each scoring system. All analyses were performed using R software (version 4.0.3) and MedCalc statistical software (version 19.6.1). The Z-test was used to compare AUC differences between different scores, following the method of Delong et al.[18]. $P < 0.05$ was considered statistically significant.

## 4. Results

### 4.1. Baseline Characteristics of Patients with Liver Failure

76540 patients in MIMIC-IV (2008-2019) database were selected in our study. After strict inclusion and exclusion criteria were followed, 1074 patients were eventually included (Figure 1) that comprised males (649, 60%) and females (425, 40%). There are two groups (Table 1) based on whether they die or not, and matched them with a propensity score matching (PSM). Results before PSM revealed that age, CCI, LODS, OASIS, SAPS II, and MELD were significantly higher in the non-death group than in the non-death group, while hospitalization days were significantly higher in the non-death control group than that of the death group. Results after PSM indicated that the ratio is equal for males and females and there were no statistically significant differences in age or Charlson's index ($P< 0.05$). LODS, OASIS, and SAPS II in addition to MELD in death group were higher than those in non-death group. In contrast, the hospitalization time was greater in the non-death group than in the death group.

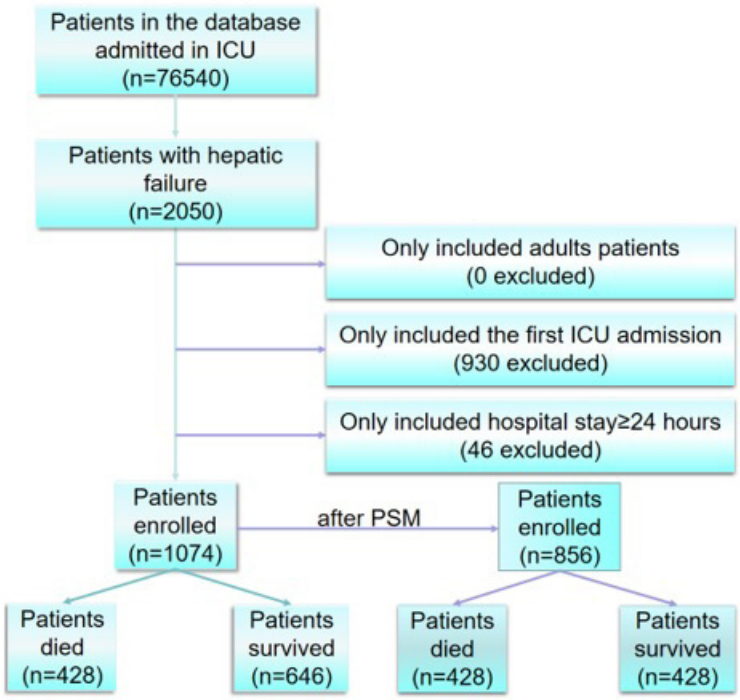


**Figure 1:** Flow charts of participants selected for this study.

**Table 1:** Demographic and clinical characteristics of the study population.

| | Before PSM | | | After PSM | | |
|---|---|---|---|---|---|---|
| **Characteristics** | **Death (n=428)** | **Survival (n=646)** | ***P-value*** | **Death (n=428)** | **Survival (n=428)** | ***P-value*** |
| **Age, year** | 61.4 ± 14.3 | 58.2 ± 15.2 | < 0.001 | 61.4 ± 14.3 | 60.3 ± 14.8 | 0.282 |
| **Gender** | | | 0.786 | | | 1 |
| **male** | 256 (59.8) | 393 (60.8) | | 256 (59.8) | 257 (60.0) | |
| **female** | 172 (40.2) | 253 (39.2) | | 172 (40.2) | 171 (40.0) | |
| **Ethnicity** | | | < 0.001 | | | < 0.001 |
| **white** | 215 (50.2) | 400 (61.9) | | 215 (50.2) | 269 (62.9) | |
| **black** | 28 (6.5) | 50 (7.7) | | 28 (6.5) | 33 (7.7) | |
| **others** | 185 (43.2) | 196 (30.3) | | 185 (43.2) | 126 (29.4) | |
| **LOS ICU, day** | 6.5 ± 6.9 | 6.3 ± 7.6 | 0.604 | 6.5 ± 6.9 | 6.5 ± 7.9 | 0.993 |
| **LOS HOS, day** | 12.6 ± 16.6 | 21.8 ± 20.8 | < 0.001 | 12.6 ± 16.6 | 21.7 ± 19.7 | < 0.001 |
| **CCI** | 6.7 ± 3.0 | 6.3 ± 3.1 | 0.02 | 6.7 ± 3.0 | 6.6 ± 3.1 | 0.477 |
| **LODS** | 9.9 ± 3.5 | 7.3 ± 3.7 | < 0.001 | 9.9 ± 3.5 | 7.4 ± 3.6 | < 0.001 |
| **OASIS** | 42.1 ± 9.4 | 36.0 ± 10.0 | < 0.001 | 42.1 ± 9.4 | 36.2 ± 9.9 | < 0.001 |
| **SAPS II** | 54.1 ± 16.0 | 42.6 ± 15.0 | < 0.001 | 54.1 ± 16.0 | 43.8 ± 14.9 | < 0.001 |
| **MELD** | 28.8 ± 9.0 | 24.2 ± 8.9 | < 0.001 | 28.8 ± 9.0 | 24.5 ± 8.9 | < 0.001 |

Values are expressed as Mean ± SD or n (%).

PSM=Propensity Score Matching, LOS ICU=Length of Intensive Care Unit Stay, LOS HOS=Length of Hospital Stay, CCI=Charlson Comorbidity Index, LODS=Logistic Organ Dysfunction System, OASIS=Oxford Acute Severity of Illness Score, SAPS II=Simplified Acute Physiology Score, MELD=Model for End-Stage Liver Disease.

### 4.2. Comparison of ROC Curves

ROC curves were selected to predict the in-hospital mortality of all cases in the cohort with four scoring systems. Before PSM (Figure 2), AUC of MELD, LODS, OASIS and SAPS II were 0.645, 0.703, 0.671 and 0.700, respectively. Then AUC of the four scoring systems were compared, LODS compared with MELD (Z = 2.844, P < 0.01), LODS compared with OASIS (Z = 2.528, P < 0.01), LODS to SAPS II (Z = 0.150, P = 0.8804), MELD to OASIS (Z = 1.192, P = 0.2333), MELD to SAPS II (Z = 2.939, P < 0.01) in addition to OASIS compared with SAPS II (Z = 1.900, P = 0.0575). After PSM (Figure 3), AUC of MELD, LODS, OASIS and SAPS II were 0.635, 0.698, 0.667 and 0.681, respectively. Likewise, the AUC of the four scoring systems were compared. LODs compared with MELD (z = 2.752, P < 0.01), LODs compared with OASIS (z = 2.205, P < 0.05), LODs to SAPS II (z = 1.034, P = 0.2969), MELD to OASIS (z = 1.274, P = 0.2027), MELD to SAPS II (z = 2.218, P < 0.05), OASIS to SAPS II (z = 0.805, P = 0.4206). The cutoff values corresponding to the Youden's index were used as the best diagnostic cutoff values for predicting in-hospital mortality in patients with liver failure. The sensitivity of MELD score and Youden's index before and after PSM had the lowest values were 58.64% and 54.91%, respectively (Table 2

and Table 3). The Youden's index before and after PSM was 0.2227 and 0.2079, respectively. LODS score had the highest specificity and Youden's index before and after PSM were 72.76% and 71.50%, respectively. OASIS scores had the highest sensitivity before and after PSM (71.73% and 65.65%, respectively). The remaining results are summarized in Table 2 (before match) and Table 3 (after match).

### 4.3. Comparison of Decision Curve

Generally, the net benefit ranges before and after PSM is LODS, SAPS II, OASIS and MELD from high to low according to DCA (Figure 4 and Figure 5). MELD was ranked last of the four rating systems. However, the interpretation of DCA should be combined with threshold, which will be detailed in the discussion section.

**Table 2:** Comparison of ROC Curves before PSM.

| Scoring system | AUC | 95%CI | Optimal cut-off | Sensitivity % | Specificity % | Youden's index |
|---|---|---|---|---|---|---|
| MELD | 0.645 | 0.615~0.673 | 27.475 | 58.64 | 63.62 | 0.2227 |
| LODS | 0.703 | 0.674~0.730 | 9 | 59.58 | 72.76 | 0.3233 |
| OASIS | 0.671 | 0.643~0.700 | 36 | 71.73 | 54.02 | 0.2575 |
| SAPS II | 0.700 | 0.672~0.728 | 47 | 64.49 | 66.72 | 0.3120 |

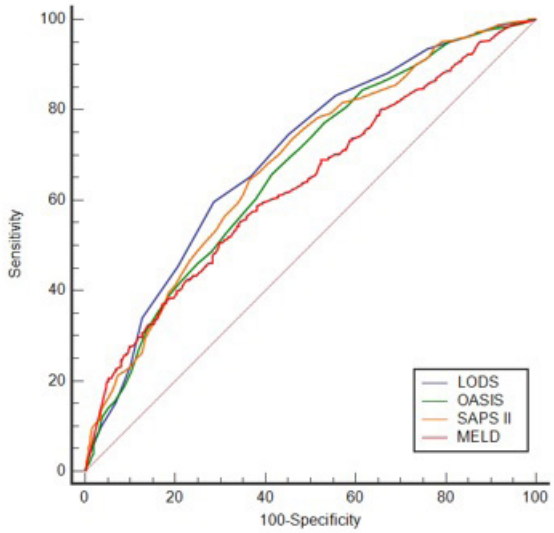


**Figure 2:** ROC curves of the four scoring systems for all cases in the cohort study before PSM.

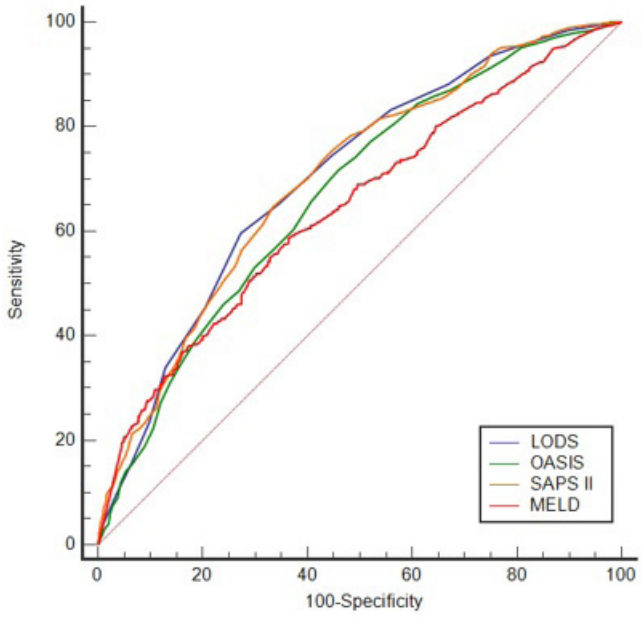


**Figure 3:** ROC curves of the four scoring systems for all cases in the cohort study after PSM.

**Table 3:** Comparison of ROC Curves after PSM.

| Scoring system | AUC | 95%CI | Optimal cut-off | Sensitivity % | Specificity % | Youden's index |
|---|---|---|---|---|---|---|
| MELD | 0.635 | 0.602~0.667 | 28.716 | 54.91 | 65.89 | 0.2079 |
| LODS | 0.698 | 0.666~0.728 | 9 | 59.58 | 71.50 | 0.3107 |
| OASIS | 0.667 | 0.634~0.699 | 38 | 65.65 | 58.64 | 0.2430 |
| SAPS II | 0.681 | 0.648~0.712 | 47 | 64.49 | 63.55 | 0.2804 |

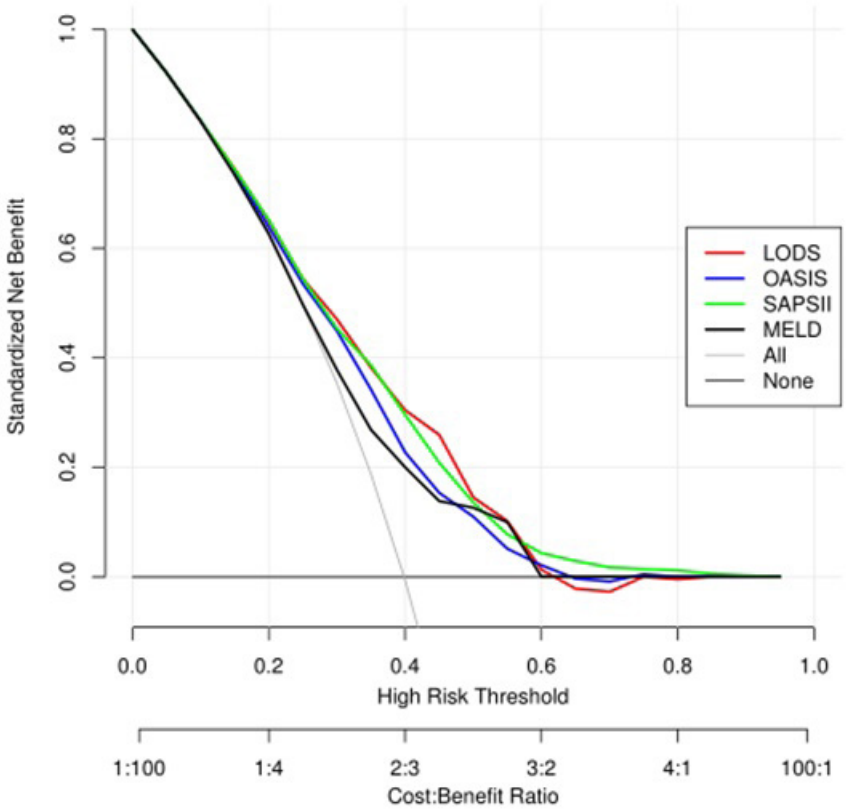


**Figure 4:** Decision curve analysis (DCA) of the four scoring systems before PSM.

**Table 4:** Binomial Logistic regression analysis (before PSM).

| | Univariable | | Multivariable | |
|---|---|---|---|---|
| **Variable** | **OR（95% CI）** | ***P-value*** | **OR（95% CI）** | ***P-value*** |
| **Age** | 1.015(1.006-1.023) | 0.001 | 1.003(0.990-1.016) | 0.691 |
| **Gender** | 1.044(0.813-1.340) | 0.737 | | |
| **LOS ICU** | 1.004(0.988-1.021) | 0.604 | | |
| **LOS HOS** | 0.960(0.950-0.970) | 0.000 | 0.954(0.943-0.965) | 0.000 |
| **CCI** | 1.048(1.007-1.091) | 0.021 | 1.035(0.975-1.098) | 0.262 |
| **LODS** | 1.223(1.178-1.269) | 0.000 | 1.151(1.081-1.226) | 0.000 |
| **OASIS** | 1.065(1.051-1.079) | 0.000 | 1.011(0.989-1.034) | 0.321 |
| **SAPS II** | 1.048(1.039-1.058) | 0.000 | 1.013(0.999-1.028) | 0.068 |
| **MELD** | 1.059(1.044-1.075) | 0.000 | 1.040(1.022-1.059) | 0.000 |

PSM=Propensity Score Matching, OR=Odds Ratio, CI=Confidence Interval, LOS ICU=Length of Intensive Care Unit Stay, LOS HOS=Length of Hospital Stay, CCI=Charlson Comorbidity Index, LODS=Logistic Organ Dysfunction System, OASIS=Oxford Acute Severity of Illness Score, SAPS II=Simplified Acute Physiology Score, MELD=Model for End-Stage Liver Disease.

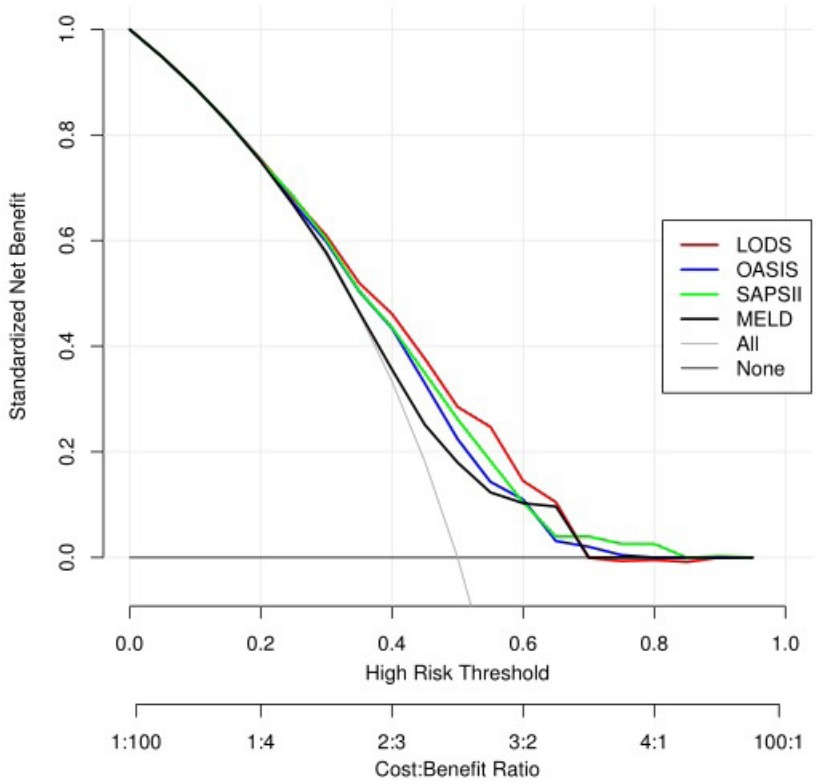


**Figure 5:** Decision curve analysis (DCA) of the four scoring systems after PSM.

### 4.4. Logistic Regression Analysis

The results before the PSM were described below. Before adjustment (Table 4), all four scoring systems were risk factors for in-hospital mortality in patients with liver failure (P < 0.001). Variables with P values less than 0.1 in univariate logistic regression analysis (Age, LOS HOS, CCI, LODS, OASIS, SAPS II, MELD) were included in multivariate logistic regression analysis and the results showed that LODS (OR: 1.151(1.081-1.226)) and MELD (OR: 1.040(1.022-1.059)) were independent risk factors for in-hospital mortality in patients with liver failure. The effect value could be interpreted as the risk of in-hospital mortality in patients with liver failure raised with each unit increase in LODS or MELD score. Age, CCI, OASIS and SAPS II were not associated with in-hospital mortality in patients with liver failure.

Here are the results of after PSM (Table 5). Before adjustment, all four scoring systems were risk factors for in-hospital mortality in patients with liver failure (P < 0.001). Variables with P value less than 0.1 in univariate Logistic regression analysis (Age, LOS HOS, CCI, LODS, OASIS, SAPS II, MELD) were included in multivariate logistic regression analysis, and the results showed that LODS (OR: 1.142(1.069-1.219)) and MELD (OR: 1.036(1.017-1.055)) were independent risk factors for in-hospital mortality in patients with liver failure. The effect value could be interpreted as an increased risk of in-hospital mortality in patients with liver failure for each unit addition in LODS or MELD. Age, CCI, OASIS and SAPS II were not associated with in-hospital mortality in patients with liver failure.

**Table 5:** Binomial Logistic regression analysis (after PSM).

| | Univariable | | Multivariable | |
|---|---|---|---|---|
| **Variable** | **OR（95% CI）** | ***P-value*** | **OR（95% CI）** | ***P-value*** |
| **Age** | 1.005(0.996-1.014) | 0.282 | | |
| **Gender** | 1.010(0.768-1.327) | 0.944 | | |
| **LOS ICU** | 1.000(0.982-1.018) | 0.993 | | |
| **LOS HOS** | 0.961(0.950-0.971) | 0.000 | 0.958(0.947-0.970) | 0.000 |
| **CCI** | 1.016(0.972-1.062) | 0.476 | | |
| **LODS** | 1.219(1.170-1.270) | 0.000 | 1.142(1.069-1.219) | 0.000 |
| **OASIS** | 1.065(1.049-1.080) | 0.000 | 1.016(0.992-1.041) | 0.187 |
| **SAPS II** | 1.044(1.034-1.054) | 0.000 | 1.011(0.997-1.025) | 0.136 |
| **MELD** | 1.054(1.038-1.071) | 0.000 | 1.036(1.017-1.055) | 0.000 |

PSM=Propensity Score Matching, OR=Odds Ratio, CI=Confidence Interval, LOS ICU=Length of Intensive Care Unit Stay, LOS HOS=Length of Hospital Stay, CCI=Charlson Comorbidity Index, LODS=Logistic Organ Dysfunction System, OASIS=Oxford Acute Severity of Illness Score, SAPS II=Simplified Acute Physiology Score, MELD=Model for End-Stage Liver Dis.

## 5. Discussion

The mortality rate of patients with liver failure was decreased with the development of clinical severe research, but it was still high in the ICU mortality rate [19, 20]. The MELD scores were widely proven to predict short-term mortality rates for patients with liver failure in previous studies. Nevertheless, some clinicians have recognized the high heterogeneity of clinical manifestations of large numbers of patients with hepatic failure under the common MELD score. As mentioned in a cohort study by Jennifer C. et al. [21], two patients had the equal value of MELD score, one patient had good ascites control and was working full time in job, while the other had high ascites with sarcopenia and was unable to stand alone. In theory, two patients with an equal MELD score would have the same predicted probability of dying, but the clinician would assess that the latter patient was clearly at higher risk of dying. Therefore, there is considerable controversy over whether MELD objectively reflects short-term mortality in patients with liver failure [22-24]. LODS, SAPS II and OASIS are also widely used in clinical severe cases [25-33], and have achieved meglio results in mass of clinical applications. In this study, we compared the association and ROC predictive effect of four scores (MELD, LODS, SAPS II, and OASIS) with in-hospital mortality in patients with liver failure, and conducted a one-to-one match based on patient death to control for confounding factors. Before PSM, multivariate logistic regression suggested that LODS and MELD were independent risk factors for in-hospital mortality in patients with liver failure. LODS score (OR: 1.151, (1.081-1.226)) was more strongly correlated with in-hospital mortality of liver failure than MELD score (OR: 1.040, (1.022-1.059). After PSM, we still observed a higher association between LODS score (1.142(1.069-1.219)) and in-hospital mortality in patients with liver failure than that of MELD score (1.036(1.017-1.055)). Both before and after PSM, MELD had the lowest AUC area, sensitivity and Youden's index, LODS score was higher in specificity and Youden's index than the other three scores, in addition to OASIS score exhibited higher in sensitivity than others. These results suggest that MELD score is not the optimal predictor of in-hospital mortality in patients with liver failure. The study of Lar E et al. also revealed that MELD scores represented no discriminative advantage over other scores in predicting liver failure and death in patients with acetaminophen-induced liver injury [34]. This confirmed our results that MELD was not the best choice for predicting the prognosis of patients with liver failure, and there is still plenty room for improvement.

MELD score is not excellent as a predictor of liver failure prognosis, but a large cohort study conducted by Gonwa et al. also demonstrated the superiority of MELD among liver transplantation candidates, and MELD continues to be used as a direct and effective strategy for identifying priority candidates in the US population [35]. This also explains our results to some extent, which suggest that MELD is a lower predictor of in-hospital mortality in patients with liver failure than the other three scores, but compared with SAPS II and OASIS, MELD is still an independent factor of in-hospital mortality in patients with liver failure after multiple logistic regression adjustment. Patients with higher MELD score were still at higher risk of death than patients with lower MELD score, suggesting that priority for liver transplantation in patients with higher MELD score may significantly reduce mortality in patients with liver failure, thereby improving clinical effectiveness. Additionally, the MELD score was also effective in predicting colorectal surgery [36], venous extracorporeal membrane oxygenation [37], high-risk acute pulmonary thromboembolism [38], and heart failure [39]. Taken together, MELD has an immense prospect in clinical application, but it still needs to be improved in predicting the prognosis of patients with liver failure.

The clinical benefit of a model is based on the identification of more true positives from positive patients, thus avoiding the waste of medical resources and reducing the harm caused by overtreatment of false positives. DCA can assist in determining the clinical benefits of applying various scoring systems. DCA curve takes probability threshold as abscissa and net income as ordinate [40, 41]. By placing several decision curves in the same coordinate, the clinical benefit of each prediction model can be intuitively judged by observing the ordinate of each decision curve under the same probability threshold. Before PSM (Figure 4) the horizontal solid black line represents the reference line without any treatment, with a net benefit of 0. The other is the reference line where all patients receive treatment (the gray slanted solid line in the graph), which the net benefit decreases as the probability threshold increases. We observed from the figure that the net benefit of the black curve corresponding to MELD is lower than the other three scores in most cases. Taking the probability threshold of 0.4 as an example, the net benefit corresponding to MELD is 0.2, indicating that when the predicted probability is 40%, MELD score can benefit about 20 out of 100 people, while the ordinate of the red curve corresponding to LODS score is about 0.3, at this point, LODS score can benefit about 30 out of 100 people. After PSM (Figure 5), we obtained almost similar results, confirming that the MELD scoring system was generally inferior to the other three scoring systems in terms of clinical benefits. There are still some deficiencies in our study. Limited by the database samples, we were unable to distinguish between explosive liver failure and chronic liver failure, which may not explain the specific differences of the four scores between the two types of liver failure. This problem can be resolved once the MIMIC database distinguishes the types of liver failure. Secondly, we only selected four representative scoring systems for comparison, owing to the diversity of sample scores, which does not mean that we believe LODS score is the best model for predicting the prognosis of liver failure. It is just that the advantage of LODS is more pronounced within four scores than in the others. Nevertheless, some work of this study is still worth mentioning. Patients with liver failure were matched one-to-one for the first time in order to reduce confounders' interference with the four scores. Meanwhile, we also used multiple interpolation to fill the missing

data, and strived to achieve the maximum statistical efficiency.

Overall, although MELD score has made great progress in other clinical aspects, it still deserves improvement in predicting adverse outcomes in liver failure. We will carry out large prospective cohort studies and establish relevant models by using XGboost and other machine learning technologies, attempting to establish a new and more accurate prediction model for predicting the poor prognosis of liver failure patients in future studies. Thus, early intervention in patients with potentially poor outcomes can reduce mortality.

## 6. Compliance with Ethical Standards

## 7. Funding

This work was Supported by Grants from National Natural Science Foundation of China (81860370, 32160233), General Project of Natural Science Foundation of Qinghai Province (2019-ZJ-970Q), CAS "Light of West China" Program (2019) and Qinghai Scientific Innovation Project of Traditional Chinese-Tibetan Medicine (J2020007).

## References

1. Squires JE, McKiernan P, Squires RH. Acute Liver Failure: An Update. Clin Liver Dis. 2018; 22: 773-805.
2. Grek A, Arasi L. Acute Liver Failure. AACN Adv Crit Care. 2016; 27: 420-429.
3. Olson JC. Acute-on-chronic and Decompensated Chronic Liver Failure: Definitions, Epidemiology, and Prognostication. Crit Care Clin. 2016; 32: 301-9.
4. Cai Q, Liu W, Zhu M, Sheng J, Microbial Infections as a Trigger for Acute-on-Chronic Liver Failure: A Review. Med Sci Monit. 2019; 25: 4773-83.
5. Wu D, Zhang S, Xie Z, Chen E, Rao Q, Liu X et al., Plasminogen as a prognostic biomarker for HBV-related acute-on-chronic liver failure. J Clin Invest. 2020; 130: 2069-80.
6. Bernal W, Auzinger G, Dhawan A Wendon J. Acute liver failure. Lancet. 2010; 376(9736): 190-201.
7. Kamath PS, Wiesner RH, Malinchoc M, Kremers W, Therneau TM, Kosberg CL et al., G. D'Amico, E.R. Dickson and W.R. Kim, A model to predict survival in patients with end-stage liver disease. Hepatology. 2001; 33: 464-70.
8. Kamath PS, Kim WR. The model for end-stage liver disease (MELD). Hepatology. 2007; 45(3): 797-805.
9. Sacleux SC, Samuel D. A Critical Review of MELD as a Reliable Tool for Transplant Prioritization. Semin Liver Dis. 2019; 39(4): 403-413.
10. Biggins SW, Bambha K. MELD-based liver allocation: who is underserved? Semin Liver Dis. 2006; 26(3): 211-20.
11. Cholankeril G, A.A. Li, B.B. Dennis, C. Gadiparthi, D. Kim, A.E. Toll, B.J. Maliakkal, S.K. Satapathy, S. Nair and A. Ahmed, Pre-Operative Delta-MELD is an Independent Predictor of Higher Mortality following Liver Transplantation. Sci Rep. 2019; 9(1): 8312.
12. Porte RJ, Lisman T, Tripodi A, S.H. Caldwell and J.F. Trotter, The International Normalized Ratio (INR) in the MELD score: problems and solutions. Am J Transplant. 2010; 10: 1349-53.
13. Heldwein MB, Badreldin AM, Doerr F, T. Lehmann, O. Bayer, T. Doenst and K. Hekmat, Logistic Organ Dysfunction Score (LODS): a reliable postoperative risk management score also in cardiac surgical patients? J Cardiothorac Surg. 2011; 6: 110.
14. Le Gall JR, Lemeshow S Saulnier F. A new Simplified Acute Physiology Score (SAPS II) based on a European/North American multicenter study. Jama. 1993; 270(24): 2957-63.
15. Zhu Y, Zhang J, Wang G, Yao R, Ren C, Chen G et al., Machine Learning Prediction Models for Mechanically Ventilated Patients: Analyses of the MIMIC-III Database. Front Med (Lausanne). 2021; 8: 662340.
16. Johnson AE, Pollard TJ, Shen L, Lehman LW, Feng M, Ghassemi M et al., Celi and R.G. Mark, MIMIC-III, a freely accessible critical care database. Sci Data. 2016; 3: 160035.
17. Van Calster B, Wynants L, Verbeek JFM, Verbakel JY, Christodoulou E, Vickers AJ et al., Reporting and Interpreting Decision Curve Analysis: A Guide for Investigators. Eur Urol, 2018; 74(6): 796-804.
18. DeLong 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.
19. Arroyo VR. Moreau PS. Kamath R, Jalan P, Ginès F et al., Acute-on-chronic liver failure in cirrhosis. Nat Rev Dis Primers. 2016; 2: 16041.
20. Moreau RR, Jalan P, Gines M, Pavesi P, Angeli J. Cordoba F et al., Acute-on-chronic liver failure is a distinct syndrome that develops in patients with acute decompensation of cirrhosis. Gastroenterology, 2013; 144(7): 1426-37, 1437.e1-9.
21. Lai JC, Dodge JL, Kappus MR, Dunn MA, Volk ML, Duarte-Rojo A et al., Changes in frailty are associated with waitlist mortality in patients with cirrhosis. J Hepatol. 2020; 73: 575-581.
22. Wiesner R, Edwards E, Freeman R, Harper A, Kim R, Kamath P, Kremers W et al., Model for end-stage liver disease (MELD) and allocation of donor livers. Gastroenterology, 2003; 124: p. 91-6.
23. Luo X, Leanza J, Massie AB, Garonzik-Wang JM, Haugen CE, Gentry SE. MELD as a metric for survival benefit of liver transplantation. Am J Transplant. 2018; 18: 1231-1237.
24. Verna EC, Connelly C, Dove LM, Adem P, Babic N, Corsetti J et al., Center-related Bias in MELD Scores Within a Liver Transplant UNOS Region: A Call for Standardization. Transplantation. 2020; 104(7): 1396-1402.
25. Hu C, Hu B, Li Z, Yang X, Song H, Li J et al., [Comparison of four scoring systems for predicting ICU mortality in patients with sepsis]. Nan Fang Yi Ke Da Xue Xue Bao. 2020; 40: 513-518.
26. Helbok R, Kendjo E, Issifou S, Lackner P, Newton CR, Kombila M et al., The Lambaréné Organ Dysfunction Score (LODS) is a simple clinical predictor of fatal malaria in African children. J Infect Dis. 2009. 200: 1834-41.
27. Kim TK, Yoon JR, Comparison of the predictive power of the LODS and APACHE II scoring systems in a neurological intensive care unit. J Int Med Res. 2012; 40: 777-86.
28. Godinjak A, Iglica A, Rama A, Tančica I, Jusufović S, Ajanović et al.,

Predictive value of SAPS II and APACHE II scoring systems for patient outcome in a medical intensive care unit. Acta Med Acad. 2016; 45(2): 97-103.

29. Choi JY, Jang JH, Lim YS, Jang JY, Lee G, Yang HJ et al., Performance on the APACHE II, SAPS II, SOFA and the OHCA score of post-cardiac arrest patients treated with therapeutic hypothermia. PLoS One, 2018; 13: p. e0196197.
30. Aegerter P, Boumendil A. Retbi, E. Minvielle, B. Dervaux and B. Guidet, SAPS II revisited. Intensive Care Med, 2005; 31(3): 416-23.
31. Granholm, A., C.F. Christiansen, S. Christensen, A. Perner and M.H. Møller, Performance of SAPS II according to ICU length of stay: Protocol for an observational study. Acta Anaesthesiol Scand. 2019; 63: 122-127.
32. Popuri, K., D. Ma, L. Wang and M.F. Beg, using machine learning to quantify structural MRI neurodegeneration patterns of Alzheimer's disease into dementia score: Independent validation on 8,834 images from ADNI, AIBL, OASIS, and MIRIAD databases. Hum Brain Mapp. 2020; 41: 4127-47.
33. El-Manzalawy, Y., M. Abbas, I. Hoaglund, A.U. Cerna, T.B. Morland, C.M. Haggerty, E.S. Hall and B.K. Fornwalt, OASIS+: leveraging machine learning to improve the prognostic accuracy of OASIS severity score for predicting in-hospital mortality. BMC Med Inform Decis Mak. 2021; 21: 156.
34. Schmidt 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.
35. Gonwa TA, McBride MA, Anderson K, Mai ML, Wadei H Ahsan N. Continued influence of preoperative renal function on outcome of orthotopic liver transplant (OLTX) in the US: where will MELD lead us? Am J Transplant. 2006; 6: 2651-9.
36. Coakley KM, Sarasani S, Prasad T, Steele SR, Paquette I, Heniford BT. MELD-Na Score as a Predictor of Anastomotic Leak in Elective Colorectal Surgery. J Surg Res. 2018; 232: 43-48.
37. Goussous N, Akbar H, LaMattina JC, Hanish SI, Barth RN, Bruno DA. Extracorporeal membrane oxygenation support following liver transplantation-A case series. Clin Transplant. 2019; 33(7): e13628.
38. İdin K, Dereli S, Kaya A, Yenerçağ M, Yılmaz AS, Tayfur K et al., Modified model for end-stage liver disease score predicts 30-day mortality in high-risk patients with acute pulmonary embolism admitted to intensive care units. Scand Cardiovasc J. 2021; 55(4): 237-244.
39. Acar, R.D., Ş. Acar, C. Doğan, Z. Bayram, A. Karaduman, S. Uysal, Y. Akbal Ö, A. Hakgör, C. Kaymaz and N. Özdemir, The TAPSE/PASP ratio and MELD score in patients with advanced heart failure. Herz, 2021; 46(Suppl 1): 75-81.
40. Zhang Z, Rousson V, Lee WC, Ferdynus C, Chen M, Qian X et al., Decision curve analysis: a technical note. Ann Transl Med, 2018; 6(15): 308.
41. Capogrosso P, Vickers AJ. A Systematic Review of the Literature Demonstrates Some Errors in the Use of Decision Curve Analysis but Generally Correct Interpretation of Findings. Med Decis Making, 2019; 39: 493-498.