

84658_Auto_Edited.docx

Name of Journal: *World Journal of Gastrointestinal Surgery*

Manuscript NO: 84658

Manuscript Type: ORIGINAL ARTICLE

Observational Study

Prognostic scores in primary biliary cholangitis patients with advanced disease

Feng J *et al.* Prognostic scores in advanced PBC

Juan Feng, Jia-Min Xu, Hai-Yan Fu, Nan Xie, Wei-Min Bao, Ying-Mei Tang

Abstract

BACKGROUND

Due to the chronic progressive disease characteristics of primary biliary cholangitis (PBC), patients with advanced PBC should not be ignored; most prognostic score studies focused on early-stage PBC.

AIM

To compare the prognostic value of various risk scores in advanced PBC to help PBC patients obtain more monitoring and assessment.

METHODS

This study considered patients diagnosed with PBC during hospitalization between 2015 and 2021. The clinical stage was primarily middle and late, and patients usually took ursodeoxycholic acid (UDCA) after diagnosis. The discriminatory performance of the scores was assessed with concordance statistics (C-statistics) at baseline and one year of UDCA. Telephone follow-up was conducted to analyze the course and disease-associated outcomes. The follow-up deadline was December 31, 2021. We compared the risk score indexes between those patients who reached a composite end point of death or liver transplantation (LT) and those who remained alive at the deadline. The combined performance of prognostic scores in estimating the risk of death or LT at one year of UDCA was assessed using Cox regression analyses. Predictive accuracy was evaluated by comparing predicted and actual survival through Kaplan-Meier analyses.

RESULTS

We included 397 patients who were first diagnosed with PBC during hospitalization and received UDCA treatment; most disease stages were advanced. After an average of 6.4 ± 1.4 years of follow-up, 82 patients had died, and four had undergone LT. After receiving UDCA treatment for one year, the score with the best discrimination performance was the Mayo, with a C-statistic of 0.740 (95% confidence interval: 0.690-

0.791). The albumin-bilirubin (ALBI), GLOBE, and Mayo scores tended to overestimate transplant-free survival. Comparing seven years of calibration results showed that the Mayo score was the best model.

CONCLUSION

Mayo, GLOBE, UK-PBC, and ALBI scores demonstrated comparable discriminating performance for advanced-stage PBC. The Mayo score showed optimal discriminatory performance and excellent predictive accuracy.

Key Words: Primary biliary cholangitis; Prognostic value; Liver transplantation; Cholangitis

Feng J, Xu JM, Fu HY, Xie N, Bao WM, Tang YM. Prognostic scores in primary biliary cholangitis patients with advanced disease. *World J Gastrointest Surg* 2023; In press

Core Tip: Primary biliary cholangitis (PBC) is a chronic progressive liver disease that destroys the intrahepatic small bile ducts. PBC in the middle and late stages cannot be ignored. The present study enrolled patients first diagnosed with PBC during hospitalization whose disease stages were primarily in the middle and late stages. We compared the prognostic value of various risk scores in PBC patients with advanced disease stages so that a significant proportion would undergo monitoring, disease evaluation, and timely treatment.

INTRODUCTION

5

Primary biliary cholangitis (PBC) is a chronic progressive liver disease that causes the gradual destruction of the intrahepatic small bile ducts^[1]. Preclinical PBC may present with specific diagnostic antibodies (anti-mitochondrial antibody, AMA) but remain asymptomatic and normal liver function for over a decade. Approximately 50%-60% are asymptomatic at diagnosis^[2]. Ursodeoxycholic acid (UDCA) is the first-line treatment. It

increases long-term survival; however, approximately 40% of patients with PBC have incomplete responses, and these patients progress rapidly to the middle and late stages after early diagnosis and treatment^[3]. Because of the chronic progressive disease characteristics, PBC patients in the middle and late stages should not be ignored.

Over the past 20 years, several risk-scoring models for PBC have been proposed as tools to estimate the risk of adverse outcomes and to guide management^[4]. The most influential scores are GLOBE and UK-PBC, developed for early PBC patients. Recent studies reported that these scores accurately predict outcomes in patients treated with UDCA treatment at various disease stages^[5-7] however, their application to middle and late PBC patients remains to be studied. The Mayo score was developed to determine the timing of liver transplantation (LT) in PBC and is now a model for predicting PBC survival^[8-10]. ² The aspartate aminotransferase-to-platelet ratio index (APRI) and fibrosis-4 index (FIB-4) are non-invasive fibrosis scores based on biochemical indicators^[11]. All parameters, including aminotransferase, platelets, and age, are associated with PBC outcomes^[12,13]. The albumin-bilirubin (ALBI) score was initially developed to assess liver function in hepatocellular carcinoma (HCC) patients^[14]. The total bilirubin (TBil) and albumin (ALB) in the score are associated with PBC progression, and some studies have used them to predict PBC outcomes^[15,16]. There are few studies on the efficacy and differences of the various prognostic scoring systems in PBC patients, especially in patients in advanced stages^[17-19].

Some patients with decompensated cirrhosis return to a clinical state consistent with compensated cirrhosis when they undergo appropriate etiological and symptomatic supportive treatment, the “recompensation phenomenon”^[20]. Portal hypertension and systemic inflammation can lead to the progression of decompensated cirrhosis. Recently, studies have been performed on the mechanism and clinical feasibility of reversing decompensation and recompensation in cirrhosis^[21-23]. These findings led to updating the stage evaluation concept and an outcomes estimate system for decompensated cirrhosis.

The present study enrolled patients diagnosed with PBC during hospitalization whose disease stages were in the middle and late stages. We compared the effectiveness and differences of various prognostic scoring systems to optimize monitoring, disease evaluation, and timely treatment for advanced-stage PBC.

MATERIALS AND METHODS

Population and study design

Patient data were derived from nine hospitals in Yunnan Province, China. Patients whose disease was on the first page of the medical record were diagnosed with PBC (ICD-10 code K74.3) and were treated with UDCA after diagnosis. The diagnostic criteria were as follows: elevated serum ALP; AMA-positive or AMA-negative when there were PBC-specific autoantibodies such as spl00 and gp210; histological evidence suggesting non-suppurative destructive cholangitis and interlobular bile duct injury. PBC can be diagnosed when two criteria are met, and the diagnostic criteria meet the 2018 AASL guidelines^[24].

Patients were excluded if they underwent follow-up for less than six months or if the dates of treatment initiation or major clinical events were unknown.

Data collection

Clinical data were obtained from 397 PBC patients diagnosed during hospitalization from May 1, 2015 to December 31, 2021. Clinical data collected from these patients included age, gender, ethnicity, date of PBC diagnosis, past medical and personal histories, clinical manifestations, liver disease complications, liver biopsy results, imaging results, gastroscopy results, laboratory values (immunological tests, serum biochemistries, complete blood counts, and coagulation times). UDCA (13-15 mg/kg/day) was prescribed after diagnosis, and laboratory results were collected at the one-year follow-up. Current guidelines and the reports from centers worldwide state that biochemical improvement after one year of UDCA treatment accurately

predicts long-term outcomes and survival^[24-26]; therefore, we collected laboratory results at a one-year follow-up for prognostic assessment.

All patients were followed up by telephone with a deadline of December 31, 2021. Endpoint events were liver-related death or LT. No endpoint event is non-transplantation survival. Classification of the disease stage according to patients' clinical characteristics and examination data. A cirrhosis diagnosis was based on a liver imaging examination (B-ultrasound, computed tomography), liver biopsy, or liver transient elastic imaging in the medical records, which indicated cirrhosis. The diagnosis standard was derived from the 2020 guidelines^[27]. We divided the patients into groups without cirrhosis, compensated cirrhosis, and decompensated cirrhosis.

Ethical considerations

This study was performed as per the Declaration of Helsinki. The Ethics Committee of the second affiliated hospital of Kunming medical university approved the study (approval No. YJ-2022-14). Each participating center approved the protocol. We analyzed all data anonymously.

Statistical analyses

The baseline time was the start of UDCA treatment, and the primary endpoint was a composite of death or LT. Patients not meeting this endpoint during follow-up were censored at their final follow-up visit. The formulas of prognostic scores can be found in the Supplementary material.

These scores were computed at baseline and after one-year UDCA treatment. These risk scores were descriptive statistics to compare patients that did or did not meet the composite endpoint.

Predictive validity was based on model discrimination and calibration. Cox proportional hazard regression analyses were performed to assess the discriminative performance of the risk-scoring models at baseline and after UDCA treatment for one year. The overall discriminative performance of these models was calculated using the

concordance (C)-statistic. Combining these predictive models when assessing the risk of death or LT based on data collected following UDCA treatment for one year was further evaluated using Cox regression analyses. C-statistic values were also assessed for various combinations of risk prediction models.

A graphical approach was used to assess model calibration by comparing Kaplan-Meier transplant-free survival estimates produced by these risk prediction models after one year of UDCA treatment.

All analyses were performed using R v 4.2.1. To account for missing values, the predictive mean matching of the mice package was applied to interpolate the missing data of laboratory results using multiple interpolation methods. Continuous data were expressed as the median and interquartile range. $P < 0.05$ was the threshold of significance.

RESULTS

Study population characteristics

We enrolled 397 PBC patients initially diagnosed while hospitalized and underwent UDCA treatment. The mean age was 56.84 (standard deviation 11.2) years and included 343 (86.4%) women. The specific staging, ¹clinical and biochemical characteristics at the beginning of UDCA treatment are displayed in Table 1.

The patients were followed for 6.4 ± 1.4 years, with three lost to follow-up at the final follow-up. During follow-up, 86 experienced a clinical endpoint, four underwent LT, and ⁶82 patients died. Liver disease was related to the cause of death in 79/82 (96.3%). The ⁶3-, 5-, and 7-year transplant-free survival rates were 94.0%, 86.9%, and 78.3%, respectively (Figure 1). Advanced stages correlated with lower survival ($P < 0.001$).

At the start of UDCA therapy, 80 (20.2%) patients had no cirrhosis, 43 (10.9%) had compensated cirrhosis, and 274 (69.0%) had decompensated cirrhosis.

Discriminative performance of different prognostic risk-scoring models

The overall discriminative performance of the Mayo, APRI, FIB-4, and ALBI models was assessed at baseline based on C-statistic values when used to predict death or LT. GLOBE and UK-PBC scores were based on values measured at baseline and after UDCA treatment for one year; therefore, the C-statistic values were assessed after one year. The baseline C-statistic values for the Mayo and ALBI scores were 0.702 [95% confidence interval (CI): 0.653-0.751] and 0.705 (95%CI: 0.656-0.755), respectively, while the FIB-4 and APRI scores showed poorer performance (Table 2).

Following UDCA treatment for one year, the C-statistic values for Mayo, GLOBE, UK-PBC, and ALBI scores were 0.740 (95%CI: 0.678-0.776), 0.731 (95%CI: 0.681-0.782), 0.727 (95%CI: 0.678-0.776), and 0.725 (95%CI: 0.672-0.778), respectively. In contrast, the FIB-4 showed poorer discriminatory power, and the APRI scores showed virtually no discriminatory performance (Table 2; Supplementary Figure 1).

There were no significant differences between the GLOBE, UK-PBC, Mayo, and ALBI scores concerning predictive performance at the start of UDCA treatment or one year after (Supplementary Table 1).

Analysis of the combined performance of different risk prediction scores

Cox regression analyses were used to evaluate the availability of combining predictive models when assessing the odds of death or LT based on data collected following UDCA treatment for one year. In univariate Cox regression analyses, the UK-PBC, ALBI, GLOBE, and Mayo scores were all significantly associated with death or LT ($P < 0.001$) (Table 3). The hazard ratio (HR) of UK-PBC was the largest (HR 6.046, 95%CI: 3.479-10.510). In multivariate analysis, only the GLOBE scores remained significantly associated with death or LT (Table 3).

Adding the UK-PBC, APRI, FIB-4, Mayo, and ALBI to the GLOBE score did not significantly improve the discriminative performance, with a C-statistic value that remained at 0.73 (Supplementary Table 2). The C-statistics of all scores before adding are displayed in Table 1.

Combining the UK-PBC score with APRI, FIB-4, and ALBI did not cause a significant increase in discrimination performance. The C-statistic remained at 0.72 (Supplementary Table 2); only with the addition of the Mayo score did the C-statistic increase (+0.02).

The most significant increase in C-statistic values was observed when the Mayo score was combined with the others. The APRI score increased to 0.740 (95% CI: 0.689-0.791), and the FIB-4 score increased to 0.741 (95% CI: 0.69-0.791) (Supplementary Table 2)

Calibration analyses of different predictive risk scores

The ALBI, GLOBE, and Mayo scores with superior discriminatory performance were selected to evaluate the predicted and observed survival (Figure 2). The UK-PBC score was omitted from the analyses because it primarily predicts liver-related death rather than transplant-free survival^[28]. The three risk prediction models tended to overestimate transplant-free survival. They showed good calibration for short-term survival; the deviation from observed survival at 1 to 3 years for ALBI, GLOBE, and Mayo (< 0.2%). After three years, the deviation tends to be greater yearly. The most significant deviation was for the GLOBE score (2%-4.3%), and the most minor was for the Mayo score (1%-2.4%). When these scores were evaluated at yearly intervals for up to seven years, the deviation of the GLOBE score was the greatest, and the Mayo score was the most minor. By comparison, the Mayo score demonstrated the best calibration.

DISCUSSION

We assessed the PBC-specific scores GLOBE, UK-PBC, and Mayo and compared the ALBI, APRI, and FIB-4 scores. These analyses revealed that the ALBI and Mayo scores showed adequate discriminatory performance and good predictive accuracy at baseline. The Mayo score demonstrated superior discriminatory performance and calibration singly and combined with other risk models, suggesting that this score is the best risk prediction model for predicting liver-related death or LT in PBC patients in the

advanced stage. These findings also suggest that the performance of the PBC-specific risk scores is superior to other prognostic scores for advanced PBC.

¹ Models with a C-statistic value greater than 0.7 are considered good prognostic models. The Mayo score was the only model consistently reaching this threshold at baseline and one year of UDCA treatment. The C-statistic of the Mayo score was greater after patients received UDCA for one year, suggesting an increase in discriminatory performance with prolonged UDCA treatment. The next most effective predictive models were the GLOBE, UK-PBC, and ALBI scores, with no significant differences in predicting liver-related death or LT following UDCA treatment for one year.

The Mayo score exhibited consistently better discriminative performance than other scores in this PBC patient cohort. The Mayo score is a traditional risk prediction model developed for PBC patients, although it was primarily developed to evaluate untreated PBC patients while this study enrolled patients that had undergone UDCA treatment and were in an advanced stage. The Mayo score has previously been linked to transplant-free survival among patients that underwent UDCA treatment, enabling their stratification into low- and high-risk groups based on the original thresholds^[29,30]. However, the reliance of this scoring model on ascites, which can be subjective, may limit its clinical applicability. In this study, the parameter was derived from the results of imaging examinations during hospitalization, and this examination is a routine item of these hospitalized patients, so the judgment of the parameter of ascites was relatively objective. In theory, the superior discriminatory performance of the Mayo score may be promoted by ascites and prothrombin time, which are the most relevant parameters in late-stage PBC; other parameters are TBil and ALB, which also are indicators of significant changes in patients with more advanced-stage. Based on these characteristics, the Mayo score may be more applicable for prognosis assessment in advanced-stage PBC patients. Our study verified this point, but the actual evidence remains to be further verified in a large population or more studies.

The discriminatory performance of GLOBE, UK-PBC, and ALBI scores is secondary to the Mayo score. Both the UK-PBC and GLOBE scores were developed as PBC-specific

scoring systems and have previously been applied to evaluate the prognosis of early PBC patients, our cohort was mainly late-stage patients, and the results were inferior to the Mayo score. ALBI score is calculated using two indicators (TBil, ALB), validated biomarkers associated with PBC disease progression^[31-33]. APRI and FIB-4 scores had inferior discriminatory performance in this study, while the two were liver fibrosis scores based on biochemical indicators. However, this study's poor performance may be because most patients had cirrhosis without significant differences in the progression of liver fibrosis, which is not applicable to predicting advanced PBC patients.

The different combinations of prognostic models were evaluated for their ability to predict death or LT. The study results showed that GLOBE and UK-PBC were relatively stable, with little change in C-statistic when other scores were added. Moreover, all predictive models' univariate and multivariate Cox regression analyses also support this point. Adding the APRI, FIB-4, and other scores, the highest C-statistic value increases were observed when the Mayo scores were combined. The results demonstrated that GLOBE and UK-PBC score models have good stability and are applicable for prognosis assessment exclusively. While the APRI and FIB-4 scores were applied to combine with other scores, and the best discriminatory performance was combined with the Mayo score.

We chose the ALBI, GLOBE, and Mayo scores for model calibration, which had superior discriminatory performance after UDCA therapy for one year, while the UK-PBC model was omitted because it predicts liver-related death and not transplant-free survival^[28]. These scores all tended to overestimate the transplant-free survival rate, with better calibration at 1-3 years. The deviation tended to increase yearly after three years. In the 1-7-year interval, the deviation of the GLOBE score was the greatest, and the Mayo score was the most minor. In contrast, the best model calibration was the Mayo score. These findings suggest the Mayo score has the best prediction performance and accuracy for advanced PBC patients.

This study has several limitations. First, we did not have a large study cohort, and the comparison of prognostic scores was calculated at baseline and one year later; this

limitation indicates the need for verification using large sample sizes and prospective studies. Second, this was a retrospective analysis; some of the included data were missing. We applied predictive mean matching to interpolate the missing values. Third, while the UK-PBC risk score was developed to predict liver-related death and not transplant-free survival (unlike the other score models), the same analyses used the endpoints and indicated similar discriminatory performance. Despite the limitations, the study is significant because of the lack of the comparison of prognostic scores in advanced PBC patients.

CONCLUSION

The Mayo, GLOBE, UK-PBC, and ALBI scores had excellent prediction performance for death and LT. Mayo scores had the best prediction efficacy in discriminating performance and predicting outcomes. The significance of this study is that it enables advanced PBC patients to be monitored and assessed closely in clinical practice to delay PBC progression.

3%

SIMILARITY INDEX

PRIMARY SOURCES

- 1 [Jorn C. Goet, Carla F. Murillo Perez, Maren H. Harms, Annarosa Floreani et al. "A Comparison of Prognostic Scores \(Mayo, UK-PBC, and GLOBE\) in Primary Biliary Cholangitis", American Journal of Gastroenterology, 2021](#) 42 words — 1%
Crossref
- 2 [aasldpubs.onlinelibrary.wiley.com](#) 18 words — 1%
Internet
- 3 [www.science.gov](#) 16 words — < 1%
Internet
- 4 [Guangqin Xiao, Jiayin Yang, Lunan Yan. "Comparison of diagnostic accuracy of aspartate aminotransferase to platelet ratio index and fibrosis-4 index for detecting liver fibrosis in adult patients with chronic hepatitis B virus infection: A systemic review and meta-analysis", Hepatology, 2015](#) 14 words — < 1%
Crossref
- 5 [www.researchgate.net](#) 11 words — < 1%
Internet
- 6 [core.ac.uk](#) 10 words — < 1%
Internet

EXCLUDE QUOTES ON
EXCLUDE BIBLIOGRAPHY ON

EXCLUDE SOURCES < 10 WORDS
EXCLUDE MATCHES < 10 WORDS