Citation: Clinical and Translational Gastroenterology (2017) 8, e100; doi:10.1038/ctg.2017.23
Official journal of the American College of Gastroenterology

www.nature.com/ctg

Prognostic Factors for Transplant-Free Survival and Validation of Prognostic Models in Chinese Patients with Primary Biliary Cholangitis Receiving Ursodeoxycholic Acid

Ka-Shing Cheung, MBBS, MPH[1], Wai-Kay Seto, MD[1,2], James Fung, MD[1,2], Ching-Lung Lai, MD[1,2] and Man-Fung Yuen, MD, PhD[1,2]

OBJECTIVES: We aimed to validate the prognostic models for primary biliary cholangitis (PBC) in Chinese patients receiving ursodeoxycholic acid (UCDA), and to compare their performances in predicting the long-term survival.
METHODS: Chinese patients with PBC from a tertiary center were identified via electronic search of hospital medical registry. Risk factors associated with adverse events (liver transplantation or death from liver-related causes including hepatocellular carcinoma (HCC) and liver decompensation) were determined. Transplant-free survival was defined as survival free of liver-related death or transplantation.
RESULTS: Of the 144 patients, 41 (28.5%) had baseline cirrhosis. The median age at diagnosis was 57.8 years. During a median follow-up of 7.0 years, 40 patients died (21 liver-related; 19 non-liver-related), 12 developed HCC, and 10 underwent transplantations. The 5-, 10-, and 15-year transplant-free survival probabilities were 91.0%, 78.1%, and 58.9%, respectively. Independent risk factors for adverse events were increasing age (hazard ratio (HR) 1.05), cirrhosis (HR 8.53), and suboptimal treatment response (HR 3.06). Aspartate aminotransferase/platelet ratio index at 1 year (APRI-r1) in combination with treatment response optimized the risk stratification. The performances of the GLOBE, UK-PBC scores, Rotterdam criteria, and APRI-r1 were comparable in predicting adverse events. The area under receiver operating curves within 5, 10, and 15 years were as follows—GLOBE score: 0.83, 0.85, and 0.85, respectively; UK-PBC score: 0.89, 0.83, and 0.79, respectively; Rotterdam criteria: 0.82, 0.76, and 0.80, respectively; APRI-r1: 0.80, 0.83, and 0.77, respectively.
CONCLUSIONS: The UK-PBC, GLOBE scores, Rotterdam criteria, and APRI-r1 had good and comparable prognostic prediction values for Chinese PBC patients receiving UCDA.
Clinical and Translational Gastroenterology (2017) **8,** e100; doi:10.1038/ctg.2017.23; published online 22 June 2017
Subject Category: Liver

INTRODUCTION

Primary biliary cholangitis (PBC) is a chronic cholestatic liver disease due to an immune-mediated damage on the small-sized biliary ducts leading to progressive ductal destruction and loss.[1] Cirrhosis is one of the major risk factors for poor prognosis, posing a higher risk of hepatocellular carcinoma (HCC) development and death.[2]

Without treatment, patients with PBC are estimated to have 5-, 10-, and 15-year transplant-free survival rates of 79%, 59%, and 32%, respectively.[3] Currently, ursodeoxycholic acid (UDCA) is the global standard of treatment that has consistently shown to be beneficial in delaying histologic progression, reducing cirrhotic complications, and liver transplantations, as well as improving long-term survival.[4–7] In fact, among those who have an adequate response to UDCA, the survival is comparable to that of age- and sex-matched healthy subjects, while the disease progression is faster for suboptimal responders.[5]

Around 35% of PBC patients do not have an optimal response to UDCA.[8] Therefore, a number of prognostic models exist to define biochemical response to UCDA, including the Rotterdam criteria,[9] Barcelona criteria,[5] Paris I criteria,[8] Paris II criteria,[10] and Toronto criteria.[11] They have been shown to be able to stratify the risks of developing adverse outcomes including cirrhosis, HCC, and mortality. However, these prognostic models are based on treatment response only. Newer models, the UK-PBC score and GLOBE score, have been developed by including other prognostic factors such as age, platelet count,[12] aspartate aminotransferase (AST). The AST/platelet ratio index (APRI) at baseline and at 1 year (APRI-r1) were also found to be able to optimize risk stratification.[12–14] The UK-PBC score and GLOBE score were found to outperform the other models based on treatment response only. Application of APRI-r1 to all biochemical criteria has also been shown to improve the predictive

[1]Department of Medicine, The University of Hong Kong, Queen Mary Hospital, Hong Kong, Hong Kong and [2]State Key Laboratory for Liver Research, The University of Hong Kong, Hong Kong, Hong Kong
Correspondence: Man-Fung Yuen, MD, PhD, Department of Medicine, The University of Hong Kong Queen Mary Hospital, Pokfulam Road, Hong Kong, Hong Kong. E-mail: mfyuen@hkucc.hku.hk
Received 5 January 2017; accepted 28 April 2017

performances.[15] As such, the net reclassification indexes were similar for APRI in combination with other biochemical criteria and GLOBE score (20–35%).

There is a recent study conducted by Yang *et al.*[16] validating various prognostic models in the Chinese population. However, this study has a short duration of follow-up (a mean of 36 ± 16 months), and therefore the performances of these models to predict adverse outcomes can only be determined at the time point of 5 years. In addition, since only seven patients had liver transplantations or liver-related deaths, the study end point was defined as development of cirrhotic complications, while the UK-PBC and PBC scores are developed to predict adverse outcomes of liver transplantation and death.

The aim of our study was to validate various prognostic models and to compare their performances in predicting the long-term survival in Chinese patients with PBC.

METHODS

Data sources and study population. All patients with PBC who were followed up in the Hepatology Clinic of Queen Mary Hospital between January 2000 and October 2015 were recruited. Queen Mary Hospital is one of the major public hospitals in Hong Kong with a population of ~7.3 million, and is also a tertiary referral center.

Our departmental practice was to prescribe UCDA to all PBC patients. Patients prescribed with UCDA in Queen Mary Hospital between 2000 and 2015 were first identified by electronic search of the hospital medical registry. Non-PBC cases were subsequently excluded by reviewing both case notes and electronic medical records of these patients. Patients who were non-Chinese or who received UCDA for <1 year were also excluded. Other exclusion criteria included the overlap syndrome (as defined by the Paris Study Group Criteria)[17] and other concomitant liver diseases such as chronic hepatitis B virus infection, chronic hepatitis C virus infection, alcoholic liver disease, steatohepatitis, and Wilson's disease. The search and identification of PBC patients in our study are illustrated in Figure 1.

The study protocol was approved by the Institutional Review Board, the University of Hong Kong and West Cluster of Hospital Authority, Hong Kong.

Outcomes and covariates

Diagnosis of PBC. PBC was diagnosed when two out of three criteria were satisfied: (1) cholestatic liver biochemistry with raised alkaline phosphatase (ALP) at least 1.5 times the upper limit of normal (ULN); (2) anti-mitochondrial antibody (AMA) positivity; (3) compatible histological features "non-suppurative destructive cholangitis with destruction of interlobular biliary ducts".[2]

For cases with liver biopsies, histologic staging was classified according to the staging system proposed by Ludwig *et al.*[18] Stage 1 is characterized by inflammation and fibrosis limited to the portal triads. Stage 2 involves the extension of inflammation and fibrosis to the periportal areas. Stage 3 is defined by bridging fibrosis, whereas stage 4 is defined by the presence of cirrhosis.

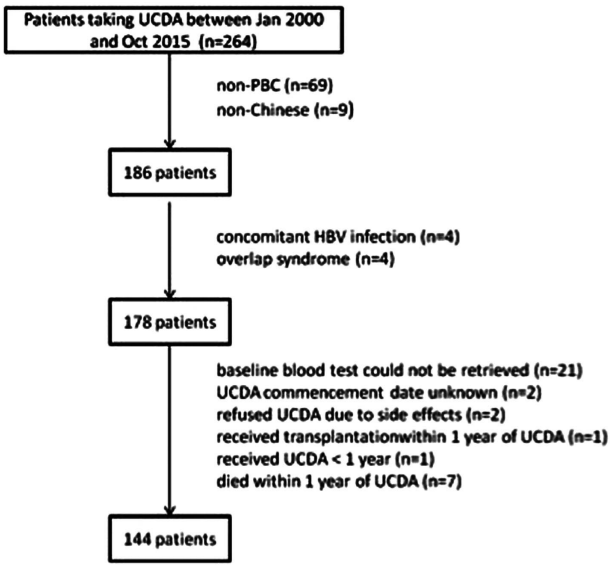

Figure 1 Flowchart illustrating case search and identification. HBV, hepatitis B virus; PBC, primary biliary cholangitis; UCDA, ursodeoxycholic acid.

Diagnosis of early PBC. The Paris II criteria were specifically developed to predict adverse outcomes in patients with early-stage disease in the study by Corpechot *et al.*[10] Both histological (Ludwig's stages 1 to 2) and biochemical criteria (normal bilirubin and albumin levels at baseline) are shown to be valid in defining early PBC.

Surveillance and diagnosis of adverse outcomes. Patients were followed up every 3 to 6 months with regular monitoring of platelet count, transaminases, ALP, γ-glutamyl transferase (GGT), bilirubin, albumin, prothrombin time (PT), international normalized ratio, α-fetoprotein, and creatinine. Patients requiring HCC surveillance as suggested by standard guidelines[19] were advised to have ultrasonography of the liver at 6-monthly intervals.

HCC was diagnosed by histology or typical radiological features (arterial enhancement and venous washout by triphasic computed tomography scan or contrast magnetic resonance imaging). Cirrhosis was diagnosed by imaging (ultrasonography/computed tomography/magnetic resonance imaging showing small, nodular liver, or features of portal hypertension namely splenomegaly, varices, and ascites), transient elastography with a fibrosis score >16.9 kPa[20] or clinical features including thrombocytopenia, coagulopathy, gastroesophageal varices, ascites, and hepatic encephalopathy.

Adverse events included liver transplantation and/or death from liver-related causes, including HCC and liver decompensation as defined by variceal bleeding, spontaneous bacterial peritonitis, hepatorenal syndrome, and/or hepatic encephalopathy. Transplant-free survival was defined as survival free of liver-related death or liver transplantation.[13] For the GLOBE score, adverse events were defined as liver transplantation and/or all-cause mortality.[14] Patients were censored at the date of latest follow-up or date of non-liver-related death. The duration of follow-up referred to the period of observation from the date of diagnosis to the censored date.

Suboptimal response to treatment. In the initial analysis of the risk factors for adverse events, suboptimal response to UCDA was identified by using the Paris I criteria, defined as ALP more than three times the ULN, AST more than two times the ULN, or bilirubin more than 1 mg/dl. The Paris I criteria were used because they were shown in previous studies to outperform other treatment response criteria.[9,12,21,22] The descriptions of other prognostic risk models are shown in Supplementary Table 1. As proposed by Trivedi *et al.*,[12] APRI-r1 in combination with treatment response criteria enables the classification of PBC patients into low-risk (biochemical response with APRI-r1 ≤ 0.54), intermediate risk (suboptimal biochemical response with APRI-r1 ≤ 0.54, or biochemical response with APRI-r1 > 0.54), and high risk (suboptimal biochemical response with APRI-r1 > 0.54).

Statistical analyses. All statistical analyses were performed using R version 3.2.3 (A language and environment for statistical computing, Vienna, Austria, ISBN 3-900051-07-0, URL http://www.R-project.org) statistical software. Continuous variables were expressed as median and interquartile range. The Mann–Whitney *U*-test was used to compare continuous variables of two groups. The χ^2 test or Fisher's exact test when appropriate was applied for comparing categorical variables. The Cox proportional hazards model was used to identify variables that were associated with adverse events. To deal with missing data in the Cox model, multiple imputation was used to construct 50 complete data sets by imputing the missing variables.[23] The Kaplan–Meier method was used to analyze the adverse outcomes, and statistical significance was determined by log-rank test. The receiver operating curve was generated by plotting 'sensitivity' against '1 - specificity' at different values. The performances of various prognostic models were expressed in terms of area under receiver operating curve (AUROC), with the 95% confidence interval (95% CI) being derived from bootstrapping by sampling with replacement from the original sample and repeating the process by 1,000 times. A two-sided *P* value of < 0.05 was used to define statistical significance.

Results

Patient characteristics. A total of 144 patients with PBC were identified. The baseline demographics, laboratory results, and histology staging are shown in Table 1a. One hundred and twenty-seven patients (88.2%) were female, and the median age at diagnosis was 57.8 years (interquartile range: 48.7–71.5 years). Forty-one patients (28.5%) had cirrhosis before treatment commencement. For patients with liver biopsy reports available for review ($n = 52$), 44.2% had stage 3 or 4 disease. The median dose of UCDA that patients received was 750 mg. The proportions of patients with suboptimal response to treatment are listed in Table 1b, ranging from 33.3 to 42.4% according to different criteria. Table 1c shows the proportions of patients with suboptimal

Table 1a Baseline characteristics of the study cohort (n = 144)

Variable	Whole cohort ($n = 144$)	Patients with adverse events ($n = 30$)	Patients without adverse events ($n = 114$)	*P* value
Age (years)	57.8 (48.7–71.5)	59.7 (45.3–71.0)	57.6 (49.0–72.1)	0.678
Female sex, *n* (%)	127 (88.2%)	24 (80.0%)	103 (90.4%)	0.118
Duration of follow-up (years)	7.0 (3.6–10.6)	7.4 (4.5–11.8)	7.0 (3.3–10.4)	0.376
Ursodeoxycholic acid (mg)	750 (750–750)	750 (750–750)	750 (750–750)	0.491
Suboptimal treatment response (Paris I criteria)	52 (36.1%)	21 (70.0%)	31 (27.2%)	<0.001
Diabetes, *n* (%)	29 (20.1%)	6 (20.0%)	23 (20.2%)	0.983
Smoking, *n* (%)[a]	13 (9.5%)	4 (14.3%)	9 (8.3%)	0.303
Alcohol, *n* (%)[a]	17 (13.7%)	2 (6.9%)	15 (13.4%)	0.525
Cirrhosis, *n* (%)	41 (28.5%)	21 (70.0%)	20 (17.5%)	<0.001
Histological stages 3 and 4, *n* (%)[b]	23 (44.2%)	12 (75.0%)	11 (30.6%)	0.006
Platelet ($x10^9$/l)[a]	216 (152–262)	135 (90–202)	234 (182–269)	<0.001
Creatinine (μmol/l)[a]	69 (60–82)	75 (62–87)	68 (60–81)	0.210
Albumin (g/l)[a]	40 (36–42)	35 (31–40)	40 (38–43)	<0.001
Bilirubin (μmol/l)[a]	14 (10–26)	30 (19–55)	13 (9–22)	<0.001
ALP (U/l)	284 (196–484)	332 (224–486)	267 (194–483)	0.376
ALT (U/l)[a]	74 (54–130)	88 (54–139)	73 (54–126)	0.540
AST (U/l)[a]	68 (51–115)	89 (62–125)	65 (40–108)	0.079
GGT (U/l)[a]	517 (256–771)	619 (457–776)	436 (219–765)	0.134
PT (s)	11.3 (10.5–11.7)	11.7 (11.4–12.8)	11.1 (10.5–11.5)	<0.001
AMA positivity	119 (82.6%)	22 (73.3%)	97 (85.1%)	0.130
Globulin (mg/dl)[a]	41 (37–46)	44 (38–50)	41 (37–45)	0.161
IgM (mg/dl)[a]	363 (250–502)	462 (289–593)	356 (240–444)	0.044
Mayo risk score	4.7 (3.8–5.5)	5.5 (4.8–6.6)	4.4 (3.7–5.2)	<0.001
MELD score	6 (6–8)	7 (6–9)	6 (6–6)	0.005
CPS	5 (5–6)	6 (5–8)	5 (5–6)	<0.001
CP class B/C	29 (20.1%)	12 (40.0%)	17 (14.9%)	<0.001

ALP, alkaline phosphatase; ALT, alanine aminotransferase; AMA, anti-mitochondrial antibody; AST, aspartate aminotransferase; CPS, Child–Pugh score; GGT, γ-glutamyl transferase; IgM, immunoglobulin M; INR, international normalized ratio; MELD, model for end-stage liver disease; PT, prothrombin time.
Adverse events were defined as liver transplantation or liver-related death.
All continuous variables are expressed in median (interquartile range).
[a]Missing data: smoking (7), alcohol (3), platelet (10), creatinine (1), albumin (2), bilirubin (2), ALT (2), AST (2), GGT (2), globulin (6), and IgM (20).
[b]Sixty-two patients had liver biopsies performed with reports available for review in 52.

treatment response according to different age strata. The rates of treatment response did not differ significantly across different age strata by using the Rotterdam, Barcelona, or Paris I criteria. However, when the Toronto criteria were used, younger patients had significantly lower rates of treatment response. None of the patients in the present cohort received fibric acid derivatives.

The median follow-up was 7.0 years (interquartile range: 3.6–10.6 years), and the total follow-up was 1,151 patient-years. Forty patients died (liver-related deaths: 21; non-liver-related deaths: 19), 12 patients developed HCC, and 10 patients underwent liver transplantations.

Sixty-two patients had liver biopsies performed with reports available for review in 52. Using the Paris II criteria, only 29 patients had early PBC according to the histological criteria, whereas 52 patients had early-stage disease according to the biochemical criteria. As a result of the relatively small sample size and number of events ($n=4$) in this subgroup of patients, Cox regression analysis and determination of AUROCs were not performed for the Paris II criteria.

Risk factors for adverse events. Table 2a shows the association between various factors and adverse events. On univariate analysis, male sex, presence of cirrhosis, lower platelet count, lower serum albumin, higher serum bilirubin, longer PT, higher serum globulin and immunoglobulin M (IgM) on presentation, and suboptimal response to UCDA (as defined by the Paris I criteria) were significant risk factors for liver transplantation and/or liver-related death. Baseline serum transaminases, ALP, and GGT were not significant risk factors. All significant risk factors on univariate analysis were further assessed by multivariate analysis. Risk factors that remained to be significant were older age (hazard ratio (HR) 1.05; 95% CI: 1.02–1.08), presence of cirrhosis (HR 8.53; 95% CI: 2.80–25.96) and suboptimal treatment response using the Paris I criteria (HR 3.21; 95% CI: 1.24–8.33). Table 2b shows the HR of suboptimal response as defined by different prognostic models. Multivariate analysis (in which statistically significant factors in Table 2a were included) was performed separately for each prognostic model. Suboptimal treatment response was a significant risk factor when the Rotterdam criteria (HR 4.05; 95% CI: 1.39–11.84), Barcelona criteria (HR 3.25; 95% CI: 1.27–8.34), or Paris I criteria (HR 3.21; 95% CI: 1.24–8.33) were adopted, but not for the Toronto criteria.

Patients were classified into low risk, intermediate risk, and high risk according to the biochemical response and APRI-r1 as mentioned in the previous section.[12] High-risk patients (suboptimal biochemical response with APRI-r1 >0.54) had the poorest prognosis compared with low-risk and intermediate-risk patients, with consistent findings using the Paris I, Rotterdam, and Toronto criteria (Table 2c).

Survival analysis. The overall 5-, 10-, and 15-year transplant-free survival rates were 91.0% (95% CI: 85.8–96.5), 78.1% (95% CI: 69.6–87.5), and 58.9% (95% CI: 46.1–75.4), respectively (Figure 2a).

Transplant-free survival was significantly better for patients without baseline cirrhosis than for those with cirrhosis (log rank $P<0.001$; Figure 2b). Among patients without baseline cirrhosis, the 5-, 10-, and 15-year transplant-free survival rates were 97.1% (95% CI: 93.3–100), 93.5% (95% CI: 87.4–99.3), and 86.4% (95% CI: 75.8–98.4). Among patients with baseline cirrhosis, the 5-, 10-, and 15-year transplant-free survival rates were 75.3% (95% CI: 61.7–91.9), 44.9% (95% CI: 29.4–68.6), and 15.7% (95% CI: 4.8–51.3).

Survival was also significantly better for patients who showed response to UCDA according to different treatment response criteria (all log rank $P<0.05$) (Figure 3a–c), except for the Toronto criteria (log rank $P=0.159$). The classification of patients into low risk, intermediate risk, and high risk according to the biochemical response and APRI-r1 further stratified the risks of transplant-free survival, regardless of which treatment response criteria was being used (all log rank $P<0.05$) (Figure 4a–d).

Validation of various prognostic models. The performances of the GLOBE and UK-PBC scores were comparable in predicting adverse events in our cohort, with an overlap of the 95% CIs. For the GLOBE score, the AUROCs for adverse events within 5, 10, and 15 years were 0.83 (95% CI: 0.74–0.90), 0.85 (95% CI: 0.75–0.92) and 0.85 (95% CI:

Table 1b Number of patients with suboptimal treatment response according to different criteria

Criteria	Number of patients with suboptimal treatment response (%)
Rotterdam criteria	61 (42.4%)
Barcelona criteria	48 (33.3%)
Paris I criteria	52 (36.1%)
Toronto criteria[a]	50 (38.8%)

[a]Fifteen patients had missing data on the ALP level at 2 years after treatment.

Table 1c Number and proportions of patients with suboptimal treatment response according to different criteria among various age strata

	Age <40 years ($n=7$)	Age 40–49.9 years ($n=35$)	Age 50–59.9 years ($n=36$)	Age 60–69.9 years ($n=25$)	Age ≥ 70 years ($n=41$)	*P* value
Rotterdam criteria	4 (57.1%)	11 (31.4%)	11 (30.6%)	11 (44.0%)	24 (58.5%)	0.063
Barcelona criteria	2 (28.6%)	7 (20%)	11 (30.6%)	14 (56.0%)	14 (34.1%)	0.070
Paris I criteria	4 (57.1%)	14 (40%)	13 (36.1%)	8 (32.0%)	13 (31.7%)	0.719
Toronto criteria	6 (85.7%)	19 (57.6%)[a]	11 (35.5%)[b]	6 (27.3%)[c]	8 (22.2%)[d]	0.002

[a]Two patients had missing data on the ALP level at 2 years after treatment.
[b]Five patients had missing data on the ALP level at 2 years after treatment.
[c]Three patients had missing data on the ALP level at 2 years after treatment.
[d]Five patients had missing data on the ALP level at 2 years after treatment.

Table 2a HRs and 95% CIs for the association between variables and adverse events (liver-related death or liver transplantation)

Variables	Univariate analysis			Multivariate analysis		
	HR	95% CI	*P* value	HR	95% CI	*P* value
Age (years)	1.02	1.00–1.05	0.079	1.05	1.02–1.08	**<0.001**
Male	3.26	1.28–8.33	**0.014**	0.81	0.25–2.61	0.721
Smoking	2.23	0.76–6.59	0.146			
Alcohol	0.73	0.17–3.12	0.674			
Diabetes mellitus	0.78	0.31–1.92	0.583			
Cirrhosis	9.64	4.23–22.00	**<0.001**	8.53	2.80–25.96	**<0.001**
Platelet (x10^9/l)	0.99	0.986–0.997	**0.001**	1.00	0.99–1.01	0.755
Creatinine (μmol/l)	1.01	0.98–1.03	0.600			
Albumin (g/l)	0.83	0.77–0.89	**<0.001**	0.95	0.85–1.08	0.452
Bilirubin (μmol/l)[a]	1.02	1.01–1.03	**0.001**			
ALP (U/l)	1.00	0.997–1.001	0.291			
ALT (U/l)	1.00	0.992–1.002	0.245			
AST (U/l)	1.00	0.995–1.005	0.863			
GGT (U/l)	1.00	0.999–1.001	0.697			
PT (s)	1.31	1.05–1.63	**0.018**	1.04	0.74–1.45	0.829
AMA positivity	0.70	0.31–1.60	0.400			
Globulin (mg/dl)	1.06	1.01–1.11	**0.023**	1.04	0.97–1.12	0.298
IgM (mg/dl)	1.001	1.000–1.002	**0.035**	1.00	0.999–1.001	0.281
Suboptimal treatment response (Paris I criteria)[a]	4.55	2.01–10.28	**<0.001**	3.21	1.24–8.33	**0.017**

ALP, alkaline phosphatase; ALT, alanine aminotransferase; AMA, anti-mitochondrial antibody; AST, aspartate aminotransferase; 95% CI, 95% confidence interval; GGT, γ-glutamyl transferase; HR, hazard ratio; IgM, immunoglobulin M; INR, international normalized ratio; PT, prothrombin time.
P values <0.05 are highlighted in bold.
[a]In the multivariate analyses, bilirubin was not included as this variable was already included in the Paris I criteria.

Table 2b HRs and 95% CIs for the association between suboptimal treatment response (as defined by various prognostic models) and adverse events (liver-related death or liver transplantation)

Criteria	Univariate analysis			[a]Multivariate analysis		
	HR	95% CI	*P* value	HR	95% CI	*P* value
Rotterdam	9.07	3.44–23.88	**<0.001**	4.05	1.39–11.84	**0.011**
Barcelona	3.75	1.72–8.20	**<0.001**	3.25	1.27–8.34	**0.014**
Paris I	4.55	2.01–10.28	**<0.001**	3.21	1.24–8.33	**0.017**
Toronto	1.54	0.72–3.32	0.265	1.62	0.57–4.57	0.362

95% CI, 95% confidence interval; HR, hazard ratio.
P values <0.05 are highlighted in bold.
[a]The adjusted HR for suboptimal response was derived by multivariate analysis with other significant variables in Table 2a (age, male sex, cirrhosis, platelet, albumin, bilirubin, prothrombin time, globulin and immunoglobulin M) included. Separate multivariate analysis was performed for each criteria in defining suboptimal response. In the multivariate analyses, bilirubin was not included for the Paris I criteria, whereas both bilirubin and albumin were not included for the Rotterdam criteria, as these variables were already included in the criteria.

0.73–0.95), respectively. For the UK-PBC score, the AUROCs for adverse events within 5, 10, and 15 years were 0.89 (95% CI: 0.77–0.98), 0.83 (95% CI: 0.73–0.91), and 0.79 (95% CI: 0.65–0.91), respectively. In comparison, the AUROCs by using other prognostic models within 5, 10, and 15 years were as follows: (Rotterdam = 0.82, 0.76, 0.80; Barcelona = 0.69, 0.60, 0.68; Paris I = 0.72, 0.70, 0.67; Toronto = 0.55, 0.52, 0.51; APRI = 0.61, 0.69, 0.40; APRI-r1 = 0.80, 0.83, 0.77). With regard to non-PBC-specific scores, the AUROCs within 5, 10, and 15 years were 0.67, 0.69, 0.71 for model for end-stage liver disease score and 0.84, 0.69, 0.70 for Child–Pugh score (CPS).

Table 3 summarizes the predictive performances of various models. Among the criteria based on treatment response only, the sensitivity and negative predictive value (NPV) were highest using the Rotterdam criteria. Values of sensitivity, specificity, positive predictive value (PPV), and NPV were not available for the UK-PBC, GLOBE, APRI, APRI-r1, MELD, and CPS as these scores were on a continuous scale.

DISCUSSION

The present study recruited 144 Chinese PBC patients who had received UDCA for at least 1 year. The median age at diagnosis was 57.8 years in our cohort, which was similar to that described in studies from other Asian countries (including Taiwan, Singapore, and South Korea), ranging from 55.6 to 57.4 years.[24–26] With the mean age of diagnosis of 54.5 years in the western population,[3] Asian patients appear to be diagnosed with PBC at a slightly older age. Females accounted for the majority of our cohort (88%), and 43% of our patients had stage 3 or 4 disease at the time of diagnosis. The female preponderance and proportion of patients with advanced histologic disease are similar to those reported in the West.[2,8]

Table 2c Prediction of adverse events (liver-related death or liver transplantation) by suboptimal treatment response in combination with APRI-r1

Criteria	Univariate analysis			[a]Multivariate analysis		
	HR	95% CI	*P* value	HR	95% CI	*P* value
Rotterdam[b]						
Low risk	Reference	—	—	Reference	—	—
Intermediate risk	1.59	0.27–9.51	0.613	1.05	0.16–7.07	0.956
High risk	11.80	2.76–50.36	**<0.001**	5.30	1.03–27.34	**0.046**
Barcelona[b]						
Low risk	Reference	—	—	Reference	—	—
Intermediate risk	5.94	0.75–47.01	0.091	2.29	0.24–21.82	0.473
High risk	22.33	2.91–171.26	**0.003**	8.38	0.96–73.17	0.055
Paris I[b]						
Low risk	Reference	—	—	Reference	—	—
Intermediate risk	3.96	0.77–20.46	0.100	1.94	0.32–11.55	0.469
High risk	10.02	2.33–43.10	**0.002**	5.48	1.09–27.60	**0.039**
Toronto						
Low risk	Reference	—	—	Reference	—	—
Intermediate risk	9.58	1.20–76.88	**0.033**	9.73	0.84–112.80	0.069
High risk	12.60	1.66–95.68	**0.014**	18.67	1.49–234.17	**0.023**

APRI-r1, AST/platelet ratio index at 1 year; 95% CI, 95% confidence interval; HR, hazard ratio.
P values <0.05 are highlighted in bold.
[a]The adjusted HR for suboptimal response was derived by multivariate analysis with other significant variables in Table 2a (age, male sex, cirrhosis, platelet, albumin, bilirubin, prothrombin time, globulin and immunoglobulin M) included. Separate multivariate analysis was performed for each criteria in defining suboptimal response. In the multivariate analyses, bilirubin was not included for the Paris I criteria, whereas both bilirubin and albumin were not included for the Rotterdam criteria, as these variables were already included in the criteria.
[b]Missing data: 6 (Paris I, Rotterdam and Barcelona criteria in combination with APRI-r1), 20 (Toronto criteria in combination with APRI-r1).

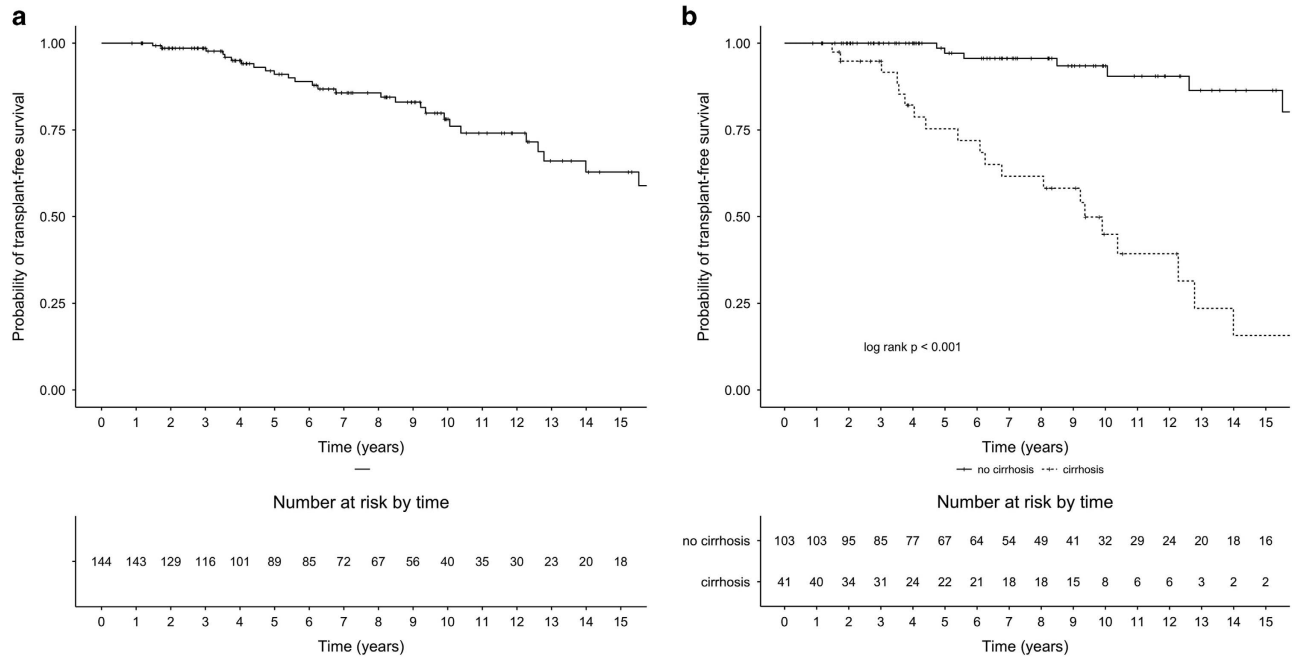

Figure 2 Baseline cirrhosis was associated with poorer transplant-free survival. (**a**) Kaplan–Meier survival plot for overall survival of the whole cohort. (**b**) Kaplan–Meier survival plot stratified by baseline cirrhosis.

The treatment response rate was also comparable, ~33–42% according to different criteria. In addition, the survival rates of our cohort were similar to those reported in a meta-analysis,[3] with the 5-, 10-, and 15-year transplant-free survival rates being 91%, 78%, and 59% vs. 88%, 77%, and 63%, respectively. Overall, the characteristics of our cohort of Chinese patients were similar to those of the Western population.

Our study validated different prognostic models and compared their performances in Chinese PBC patients receiving UDCA. In our study, both the recently developed

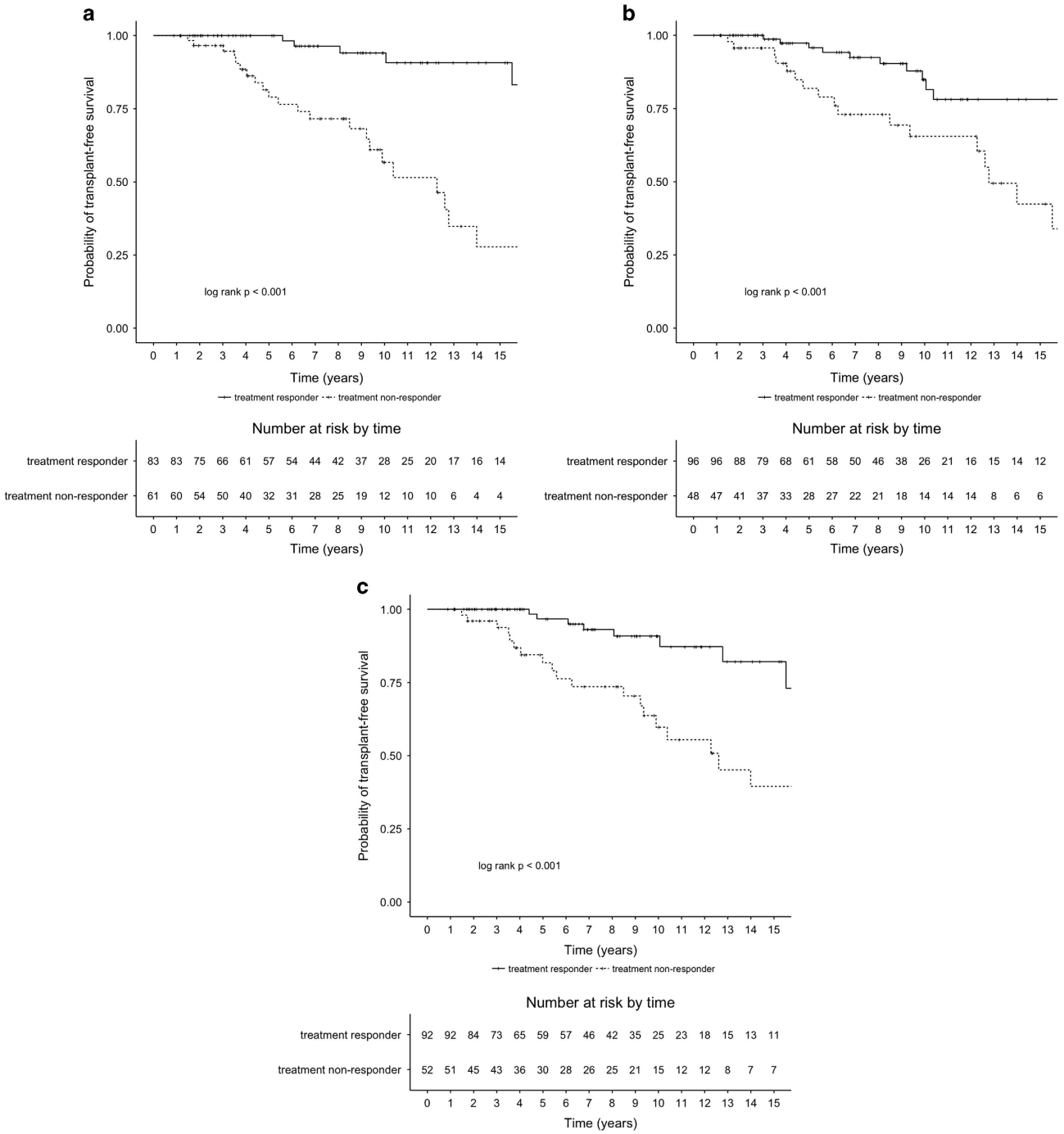

Figure 3 Suboptimal treatment response was associated with poorer transplant-free survival. (**a**) Kaplan–Meier survival plot stratified by treatment response (Rotterdam criteria). (**b**) Kaplan–Meier survival plot stratified by treatment response (Barcelona criteria). (**c**) Kaplan–Meier survival plot stratified by treatment response (Paris I criteria).

models (the GLOBE and UK-PBC scores) had better predictability (AUROCs close to or >0.80 for predicting adverse events within 5, 10, and 15 years) than other criteria based only on treatment response (except for the Rotterdam criteria), the results of which were consistent with studies of the western population.[13,14] There are several possible reasons to explain this. First, a number of important independent factors are included in the GLOBE and UK-PBC scores, e.g., age,[21] bilirubin,[3,27] ALP,[5,8,10] albumin,[28] and platelet count,[12] while the other criteria are based only on treatment response, and therefore cirrhosis may not be taken into consideration. Second, dichotomization of continuous variables in previous criteria will affect the robustness of the model.[29]

The overall performance (in terms of AUROC) of the GLOBE score in the validation cohort from the Global PBC Study Group was 0.82,[14] while that of the UK-PBC score in the

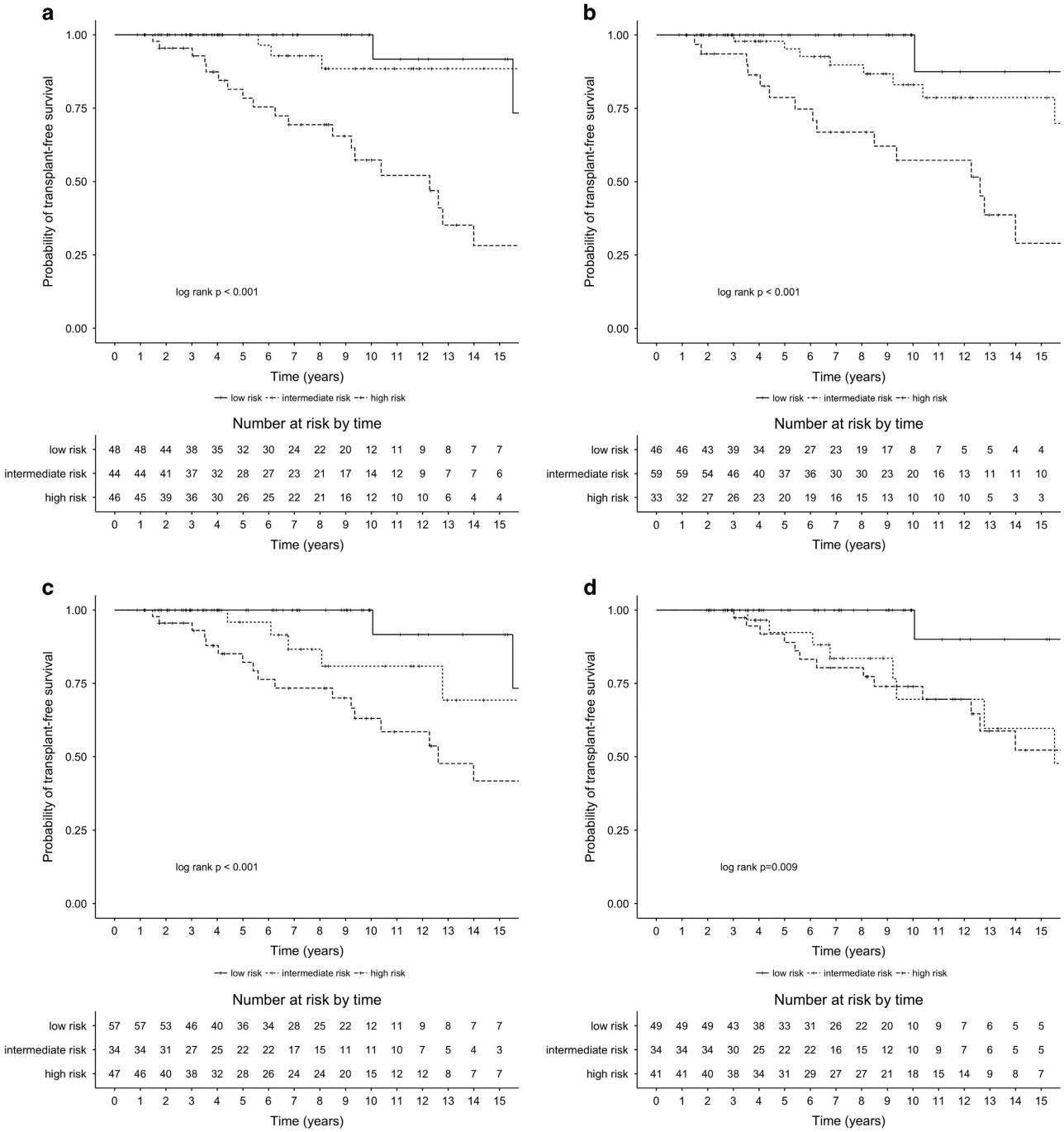

Figure 4 Treatment response criteria in combination with APRI-r1 further stratified risks of transplant-free survival. (**a**) Kaplan–Meier survival plot stratified by treatment response (Rotterdam criteria) and APRI-r1. (**b**) Kaplan–Meier survival plot stratified by treatment response (Barcelona criteria) and APRI-r1. (**c**) Kaplan–Meier survival plot stratified by treatment response (Paris I criteria) and APRI-r1. (**d**) Kaplan–Meier survival plot stratified by treatment response (Toronto criteria) and APRI-r1. APRI-r1, AST/platelet ratio index at 1-year; low risk (biochemical response with APRI-r1 ≤ 0.54), intermediate risk (suboptimal biochemical response with APRI-r1 ≤ 0.54; or biochemical response with APRI-r1 > 0.54) and high risk (suboptimal biochemical response with APRI-r1 > 0.54).

validation cohort from the UK-PBC Research Cohort was as follows: AUROCs of 0.96, 0.95, and 0.94 within 5, 10, and 15 years, respectively.[13] APRI-r1 was another prognostic model with a good performance (AUROUC of ~ 0.80)[12]

Our study showed that the performances of these two newly developed models were comparable in terms of predicting adverse events in the Chinese population, as evidenced by the overlapping of the 95% CIs across different periods. The performances of the GLOBE score were similar in the western and Chinese patients, while the UK-PBC score performed better in the western population. However, these results should be interpreted with caution as the number of patients recruited in our study was less than the Global PBC Study Group and UK-PBC Research Cohort.

Table 3 Predictive performances of prognostic models for adverse events

	Rotterdam	Barcelona	Paris I	Toronto	UK-PBC	GLOBE	APRI	APRI-r1	MELD	CPS
Event within 5 years										
AUROC (95% CI)	0.82 (0.77–0.86)	0.69 (0.54–0.85)	0.72 (0.57–0.84)	0.55 (0.34–0.74)	0.89 (0.77–0.98)	0.83 (0.74–0.90)	0.61 (0.33–0.86)	0.80 (0.62–0.93)	0.67 (0.52–0.81)	0.84 (0.69–0.95)
Sensitivity	100%	70.0%	80.0%	50.0%	—	—	—	—	—	—
Specificity	63.5%	68.8%	35.4%	58.9%	—	—	—	—	—	—
PPV	22.2%	18.9%	19.0%	9.3%	—	—	—	—	—	—
NPV	100%	95.7%	96.9%	93.5%	—	—	—	—	—	—
Event within 10 years										
AUROC (95% CI)	0.76 (0.65–0.85)	0.60 (0.48–0.72)	0.70 (0.58–0.80)	0.52 (0.40–0.65)	0.83 (0.73–0.91)	0.85 (0.75–0.92)	0.69 (0.54–0.83)	0.83 (0.72–0.92)	0.69 (0.56–0.83)	0.69 (0.56–0.82)
Sensitivity	81.8%	54.5%	72.7%	52.6%	—	—	—	—	—	—
Specificity	70.4%	66.7%	66.7%	51.9%	—	—	—	—	—	—
PPV	52.9%	40.0%	47.1%	27.8%	—	—	—	—	—	—
NPV	90.5%	78.3%	85.7%	75.7%	—	—	—	—	—	—
Event within 15 years										
AUROC (95% CI)	0.80 (0.67–0.92)	0.68 (0.54–0.82)	0.67 (0.52–0.80)	0.51 (0.36–0.66)	0.79 (0.65–0.91)	0.85 (0.73–0.95)	0.40 (0.19–0.62)	0.77 (0.61–0.92)	0.71 (0.55–0.84)	0.70 (0.54–0.85)
Sensitivity	81.5%	63.0%	70.4%	58.3%	—	—	—	—	—	—
Specificity	78.9%	73.7%	63.2%	42.1%	—	—	—	—	—	—
PPV	84.6%	77.3%	73.1%	56.0%	—	—	—	—	—	—
NPV	75.0%	58.3%	60.0%	44.4%	—	—	—	—	—	—

APRI, AST to platelet ratio index at baseline; APRI-r1, APRI at 1 year; AUROC, area under receiver operating curve; CPS, Child–Pugh score; MELD, model for end-stage liver disease; NPV, negative predictive value; PPV, positive predictive value.
Adverse events were defined as liver transplantation or all-cause mortality (GLOBE score) and liver transplantation or liver-related death (other prognostic scores).

As for the other prognostic models, the Rotterdam criteria and the APRI-r1 had overall better performances for prediction of adverse events, followed by the Paris I criteria and the Barcelona criteria, whereas the Toronto criteria did not have satisfactory performance. A previous study also showed that the Rotterdam and the Paris I criteria had better performance among criteria based on treatment response.[14]

All of these criteria had high NPV (>90% at 5 years, >75% at 10 years) but low PPV (<23% at 5 years, <53% at 10 years) in predicting adverse events. Overall, with increasing years, the sensitivity and NPV decreased, while the specificity and PPV increased. These results suggest that treatment response criteria, particularly the Rotterdam and the Paris I criteria, were accurate at recognizing patients with good prognosis.

One of the strengths of our study is the case recruitment process, in which cases were identified via electronic search of the hospital medical registry, with subsequent verification by reviewing case notes and electronic medical records. This ensures the completeness of case recruitment. Moreover, the relatively long duration of follow-up (a median duration of 7 years with the longest duration up to 26 years) allowed for more meaningful comparison of the hazards of adverse events with regard to different predictive variables, despite the long lag time for the development of adverse events from diagnosis that was typical of PBC.

There are some limitations of our study. First, the sample size was relatively small. However, the results obtained from this study appear to be valid, as evidenced by the shared patient characteristics between our cohort and other studies, as well as the consistency in terms of the performance of various prognostic models. Second, as our hospital is a tertiary center in the territory, the study cohort may represent a selected group of patients (which explains the relatively large number of events including HCC, transplantation and deaths in this study). Therefore, the findings of this study may not be applicable to all Asian PBC patients. Third, we could not ascertain whether UCDA was prescribed by a weight-based approach at a dose of 13–15 mg/kg per day,[30] because data of the baseline body weight were missing in some patients in this retrospective study. However, the survival rates of our cohort were similar to those reported in a meta-analysis,[3] indicating that the bias related to the UCDA dose was minimal. In fact, the patients in the UK-PBC Research Cohort also received a lower median dose of 12 mg/kg per day (interquartile range: 9–14 mg/kg per day).[13]

In conclusion, our study validated the usefulness of the GLOBE, UK-PBC scores, Rotterdam criteria and APRI-r1 in predicting the prognosis of long-term survival in Chinese PBC patients treated with UDCA. These prognostic models can potentially stratify the risks of individual patients, hence identifying those who require closer follow-up and at need of alternative therapies, such as obeticholic acid.[31] Further studies with larger sample sizes are warranted to confirm our findings.

CONFLICT OF INTEREST

Guarantors of the article: Ka-Shing Cheung, MBBS, MPH and Man-Fung Yuen, MD, PhD.

Specific author contributions: Ka-Shing Cheung and Wai-Kay Seto were involved with study concept and design; acquisition of data; analysis and interpretation of data; drafting of manuscript. James Fung was involved with analysis and interpretation of data; critical revision of the manuscript for important intellectual content. Ching-Lung Lai and Man-Fung Yuen were involved with the study concept and design; analysis and interpretation of data; drafting of manuscript; critical revision of the manuscript for important intellectual content; and study supervision. The corresponding author had full access to all data, and was fully responsible for the data integrity and statistical analysis. All authors revised the manuscript and approved the final version of this article.

Financial support: None.

Potential competing interests: WK Seto is an advisory board member of Bristol-Myers Squibb and Gilead Sciences, and received speaker fees from Bristol-Myers Squibb, Gilead Sciences, and Novartis. J Fung received research funding from Novartis. CL Lai received speaker fees and is an advisory board member of Bristol-Myers and Gilead Sciences. MF Yuen received speaker fees and research funding and is an advisory board member of Bristol-Myers Squibb, Novartis, Gilead Sciences, and Roche Diagnostics. The remaining authors declare no conflict of interest.

Study Highlights

WHAT IS CURRENT KNOWLEDGE

- ✓ Primary biliary cholangitis (PBC) can lead to hepatocellular carcinoma (HCC) and death.
- ✓ Studies on prognostic models predicting the long-term survival in PBC patients are lacking in Chinese population.

WHAT IS NEW HERE

- ✓ UK-PBC and GLOBE scores had good prognostic prediction values among Chinese PBC patients receiving ursodeoxycholic acid (UCDA).
- ✓ Rotterdam criteria and APRI-r1 also had satisfactory prognostic prediction values.
- ✓ The 5-, 10- and 15-year transplant-free survival rates were similar between Chinese and western patients.

1. Carey EJ, Ali AH, Lindor KD. Primary biliary cirrhosis. *Lancet* 2015; **386**: 1565–1575.
2. Lindor KD, Gershwin ME, Poupon R *et al.* Primary biliary cirrhosis. *Hepatology* 2009; **50**: 291–308.
3. Lammers WJ, van Buuren HR, Hirschfield GM *et al.* Levels of alkaline phosphatase and bilirubin are surrogate end points of outcomes of patients with primary biliary cirrhosis: an international follow-up study. *Gastroenterology* 2014; **147**: 1338–49.e5; quiz e15.
4. Heathcote EJ, Cauch-Dudek K, Walker V *et al.* The Canadian Multicenter Double-blind Randomized Controlled Trial of ursodeoxycholic acid in primary biliary cirrhosis. *Hepatology* 1994; **19**: 1149–1156.
5. Pares A, Caballeria L, Rodes J. Excellent long-term survival in patients with primary biliary cirrhosis and biochemical response to ursodeoxycholic acid. *Gastroenterology* 2006; **130**: 715–720.
6. Poupon RE, Lindor KD, Cauch-Dudek K *et al.* Combined analysis of randomized controlled trials of ursodeoxycholic acid in primary biliary cirrhosis. *Gastroenterology* 1997; **113**: 884–890.
7. Lee J, Belanger A, Doucette JT *et al.* Transplantation trends in primary biliary cirrhosis. *Clin Gastroenterol Hepatol* 2007; **5**: 1313–1315.
8. Corpechot C, Abenavoli L, Rabahi N *et al.* Biochemical response to ursodeoxycholic acid and long-term prognosis in primary biliary cirrhosis. *Hepatology* 2008; **48**: 871–877.
9. Kuiper EM, Hansen BE, de Vries RA *et al.* Improved prognosis of patients with primary biliary cirrhosis that have a biochemical response to ursodeoxycholic acid. *Gastroenterology* 2009; **136**: 1281–1287.
10. Corpechot C, Chazouilleres O, Poupon R. Early primary biliary cirrhosis: biochemical response to treatment and prediction of long-term outcome. *J Hepatol* 2011; **55**: 1361–1367.
11. Kumagi T, Guindi M, Fischer SE *et al.* Baseline ductopenia and treatment response predict long-term histological progression in primary biliary cirrhosis. *Am J Gastroenterol* 2010; **105**: 2186–2194.
12. Trivedi PJ, Bruns T, Cheung A *et al.* Optimising risk stratification in primary biliary cirrhosis: AST/platelet ratio index predicts outcome independent of ursodeoxycholic acid response. *J Hepatol* 2014; **60**: 1249–1258.
13. Carbone M, Sharp SJ, Flack S *et al.* The UK-PBC risk scores: derivation and validation of a scoring system for long-term prediction of end-stage liver disease in primary biliary cholangitis. *Hepatology* 2016; **63**: 930–950.
14. Lammers WJ, Hirschfield GM, Corpechot C *et al.* Development and validation of a scoring system to predict outcomes of patients with primary biliary cirrhosis receiving ursodeoxycholic acid therapy. *Gastroenterology* 2015; **149**: 1804–1812.e4.
15. Trivedi PJ, Corpechot C, Pares A *et al.* Risk stratification in autoimmune cholestatic liver diseases: Opportunities for clinicians and trialists. *Hepatology* 2016; **63**: 644–659.
16. Yang F, Yang Y, Wang Q *et al.* The risk predictive values of UK-PBC and GLOBE scoring system in Chinese patients with primary biliary cholangitis: the additional effect of anti-gp210. *Aliment Pharmacol Ther* 2017; **45**: 733–743.
17. Chazouilleres O, Wendum D, Serfaty L *et al.* Primary biliary cirrhosis-autoimmune hepatitis overlap syndrome: clinical features and response to therapy. *Hepatology* 1998; **28**: 296–301.
18. Ludwig J, Dickson ER, McDonald GS. Staging of chronic nonsuppurative destructive cholangitis (syndrome of primary biliary cirrhosis). *Virchows Arch A* 1978; **379**: 103–112.
19. Bruix J, Sherman M. Management of hepatocellular carcinoma: an update. *Hepatology* 2011; **53**: 1020–1022.
20. Corpechot C, Carrat F, Poujol-Robert A *et al.* Noninvasive elastography-based assessment of liver fibrosis progression and prognosis in primary biliary cirrhosis. *Hepatology* 2012; **56**: 198–208.
21. Carbone M, Mells GF, Pells G *et al.* Sex and age are determinants of the clinical phenotype of primary biliary cirrhosis and response to ursodeoxycholic acid. *Gastroenterology* 2013; **144**: 560–569.e7; quiz e13-4.
22. Trivedi PJ, Lammers WJ, van Buuren HR *et al.* Stratification of hepatocellular carcinoma risk in primary biliary cirrhosis: a multicentre international study. *Gut* 2016; **65**: 321–329.
23. White IR, Royston P. Imputing missing covariate values for the Cox model. *Stat Med* 2009; **28**: 1982–1998.
24. Kim KA, Ki M, Choi HY *et al.* Population-based epidemiology of primary biliary cirrhosis in South Korea. *Aliment Pharmacol Ther* 2016; **43**: 154–162.
25. Wong RK, Lim SG, Wee A *et al.* Primary biliary cirrhosis in Singapore: evaluation of demography, prognostic factors and natural course in a multi-ethnic population. *J Gastroenterol Hepatol* 2008; **23**: 599–605.
26. Su CW, Hung HH, Huo TI *et al.* Natural history and prognostic factors of primary biliary cirrhosis in Taiwan: a follow-up study up to 18 years. *Liver Int* 2008; **28**: 1305–1313.
27. Bonnand AM, Heathcote EJ, Lindor KD *et al.* Clinical significance of serum bilirubin levels under ursodeoxycholic acid therapy in patients with primary biliary cirrhosis. *Hepatology* 1999; **29**: 39–43.
28. ter Borg PC, Schalm SW, Hansen BE *et al.* Prognosis of ursodeoxycholic acid-treated patients with primary biliary cirrhosis. Results of a 10-yr cohort study involving 297 patients. *Am J Gastroenterol* 2006; **101**: 2044–2050.
29. Royston P, Altman DG, Sauerbrei W. Dichotomizing continuous predictors in multiple regression: a bad idea. *Stat Med* 2006; **25**: 127–141.
30. European Association for the Study of the Liver. EASL Clinical Practice Guidelines: management of cholestatic liver diseases. *J Hepatol* 2009; **51**: 237–267.
31. Nevens F, Andreone P, Mazzella G *et al.* A placebo-controlled trial of obeticholic acid in primary biliary cholangitis. *N Engl J Med* 2016; **375**: 631–643.

Supplementary Information accompanies this paper on the Clinical and Translational Gastroenterology website (http://www.nature.com/ctg)