



Original Article



A Dual Time Window-driven Strategy to Optimize Primary Biliary Cholangitis Treatment via Alkaline Phosphatase Normalization

Han Zhao^{1#}, Yansheng Liu^{1#}, Yingmei Tang^{2#}, Ningning Wang³, Yanmin Liu⁴, Yiling Li³, Chunyang Huang⁴, Jieting Duan², Yan Feng³, Linhua Zheng¹, Ruiqing Sun¹, Xiufang Wang¹, Juan Deng¹, Gui Jia¹, Patrick S.C. Leung⁵, M. Eric Gershwin^{5*}, Yulong Shang^{1*}  and Ying Han^{1*} 

¹National Clinical Research Center for Digestive Diseases and Xijing Hospital of Digestive Diseases, Xijing Hospital, Air Force Medical University, Xi'an, Shaanxi, China; ²Department of Gastroenterology, The Second Affiliated Hospital of Kunming Medical University, Kunming, Yunnan, China; ³Department of Gastroenterology, The First Hospital of China Medical University, Shenyang, Liaoning, China; ⁴Second Department of Liver Disease Center, Beijing You'an Hospital, Capital Medical University, Beijing, China; ⁵Division of Rheumatology, Allergy and Clinical Immunology, University of California, Davis, CA, USA

Received: March 17, 2026 | Revised: April 17, 2026 | Accepted: April 27, 2026 | Published online: May 15, 2026

Abstract

Background and Aims: The current criterion of biochemical response to ursodeoxycholic acid in primary biliary cholangitis is an alkaline phosphatase (ALP) level of $\leq 1.67 \times$ the upper limit of normal (ULN) after 12 months of treatment. However, a proportion of patients who meet this parameter may still progress to liver decompensation. This study aimed to optimize the clinical management of primary biliary cholangitis by (1) establishing ALP normalization as a core treatment target, (2) identifying early intervention windows, and (3) developing risk stratification criteria. **Methods:** This multicenter retrospective study included an internal cohort and an external validation cohort. We assessed the prognostic impact of ALP normalization with Kaplan-Meier and Cox regression. Sankey diagrams and segmented Poisson regression analysis mapped dynamic risk transitions to identify critical intervention windows. Predictive performance (sensitivity/specificity/positive predictive value/negative predictive value (NPV)) of Mayo, Paris II, and Toronto criteria for 12-month ALP normalization was compared. **Results:** Patients achieving ALP normalization showed significantly higher complication-free survival versus those with ALP $1.0\text{--}1.67 \times$ ULN (89.8% vs. 79.8%; $P = 0.016$). Segmented Poisson regression identified significant change points at 3.73 and 5.5 months for high-to-medium and medium-to-low risk transitions, respectively. Failure to meet the Toronto criteria at month 3

predicted non-normalization with 95% NPV, whereas Paris II criteria at month 6 provided optimal specificity (73%) for identifying patients who failed to achieve ALP normalization.

Conclusions: ALP normalization significantly improves clinical outcomes. Two subgroups demonstrate low normalization probability and warrant early intervention: (1) patients with ALP $\geq 1.67 \times$ ULN after 3 months and (2) those not meeting Paris II criteria by month 6.

Citation of this article: Zhao H, Liu Y, Tang Y, Wang N, Liu Y, Li Y, *et al.* A Dual Time Window-driven Strategy to Optimize Primary Biliary Cholangitis Treatment via Alkaline Phosphatase Normalization. J Clin Transl Hepatol 2026. doi: 10.14218/JCTH.2026.00082.

Introduction

Primary biliary cholangitis (PBC) is a chronic cholestatic liver disease that, if inadequately treated, can progress to end-stage liver disease through progressive destruction of the intrahepatic bile ducts.¹ Ursodeoxycholic acid (UDCA) is the established first-line therapy, known to improve biochemical markers, slow fibrosis progression, and extend transplant-free survival.² However, approximately 30%–40% of patients exhibit an inadequate biochemical response to UDCA, maintaining a heightened risk of liver-related complications and mortality.^{3,4}

For patients with an inadequate response to UDCA, second-line agents such as fibrates (e.g., fenofibrate) are recommended to enhance biochemical response.⁵ Several binary response criteria have been validated to identify PBC patients who remain at increased risk of death or liver transplantation (LT) despite UDCA therapy. These include the Rochester, Barcelona, Paris I, Paris II, Rotterdam, Toronto, and Ehime criteria.^{3,4,6–10} Although they vary in treatment duration, cutoff values, and included parameters, alkaline

Keywords: Primary biliary cholangitis; Alkaline phosphatase; Ursodeoxycholic acid; Stratified therapy; Early prediction; Normalization.

#Contributed equally to this work.

*Correspondence to: Ying Han and Yulong Shang, Xijing Hospital of Digestive Diseases, The Fourth Military Medical University, Xi'an, Shaanxi 710032, China. ORCID: <https://orcid.org/0000-0003-3046-9507> (YH) and <https://orcid.org/0000-0002-8576-3175> (YS). Tel: +86-29-84771509, Fax: +86-29-82539041, E-mail: hanying1@fmmu.edu.cn (YH) and shangyul870222@163.com (YS); M. Eric Gershwin, Division of Rheumatology, Allergy and Clinical Immunology, University of California, Davis, CA 95616, USA. ORCID: <https://orcid.org/0000-0001-5245-2680>. E-mail: megershwin@ucdavis.edu.

phosphatase (ALP) is a central component in all but one of these models. In the context of drug development, the PBC Obeticholic Acid International Study of Efficacy (POISE) criteria—loosely based on the Toronto model—have been widely adopted. These criteria require a serum ALP $< 1.67 \times$ the upper limit of normal (ULN), a reduction of at least 15% from baseline, and a total bilirubin (TBIL) level within the normal range, and have served as surrogate clinical endpoints in therapeutic trials.^{11,12}

However, emerging evidence challenges the adequacy of static biochemical thresholds for risk stratification in PBC. Subgroup analyses have shown that higher TBIL values within the normal range ($\leq 1.0 \times$ ULN) are associated with a worse prognosis.¹³ Specifically, patients with TBIL levels between 0.6 and $1.0 \times$ ULN had significantly lower 10-year survival rates compared with those with $TBIL \leq 0.6 \times$ ULN. More importantly, complete normalization of ALP ($\leq 1.0 \times$ ULN) has emerged as an independent protective factor. Among patients with mildly increased ALP ($< 1.67 \times$ ULN), those who failed to achieve normalization exhibited lower 10-year survival rate. Corpechot *et al.* further reinforced the prognostic value of ALP normalization, highlighting its role as a key therapeutic target.¹⁴ Among patients with ALP levels $< 1.67 \times$ ULN, achieving complete normalization ($\leq 1 \times$ ULN) was identified as an independent protective factor for complication-free survival. In contrast, reducing TBIL to $< 0.6 \times$ ULN did not confer a similar survival benefit. Collectively, these findings highlight the strong association between ALP normalization and improved treatment outcomes in PBC and support a clear dose–response relationship between ALP levels and clinical prognosis.

Furthermore, recent therapeutic advances have reinforced the feasibility of achieving ALP normalization in PBC. Phase III clinical trials of novel agents such as elafibranor and seladelpar have demonstrated significantly higher rates of ALP normalization compared with placebo.^{15,16} In addition, a randomized controlled trial by Liu *et al.* evaluating initial combination therapy with fenofibrate and UDCA showed superior biochemical response and ALP normalization rates at 12 months compared with UDCA monotherapy.¹⁷

Building on increasing evidence on the association of ALP with improved clinical outcomes, this study aimed to (1) validate ALP normalization as a definitive therapeutic target that surpasses conventional response thresholds and (2) identify predictive risk stratification criteria for early intervention.

Methods

Study population and design

We conducted a multicenter retrospective cohort study using a three-stage progressive validation design, encompassing phases to (1) validate the prognostic value of ALP normalization, (2) identify critical intervention windows, and (3) compare the predictive performance of various response criteria. The study population was divided into an internal development cohort and an external validation cohort. The internal cohort, comprising patients with PBC treated at Xijing Hospital between October 2004 and June 2024, was used for all three stages of analysis. The external validation cohort included independent patients from three additional centers, enrolled between June 2013 and December 2024, and was used specifically to validate the results of the criteria comparison in the third phase. Detailed information on patient sources from each external center is provided in Supplementary Table 1.

The inclusion criteria for patient enrollment were: (1) a

confirmed diagnosis of PBC; (2) initiation of daily UDCA therapy at a dosage of 13–15 mg/kg from the date of diagnosis; (3) a follow-up period of more than one year; and (4) absence of liver failure events during the initial 12-month treatment period. Exclusion criteria included: (1) the presence of concurrent liver diseases, such as hepatitis B or C, alcoholic liver disease, primary sclerosing cholangitis, or autoimmune hepatitis; and (2) administration of second-line therapies—including obeticholic acid or fibrate drugs—at any time during the study.

The diagnosis and treatment of PBC followed international guidelines.¹⁸ Briefly, a diagnosis of PBC was established when at least two of the following three criteria were met: (1) biochemical evidence of cholestasis indicated by increased ALP levels; (2) positivity for anti-mitochondrial antibodies; and (3) histological findings consistent with PBC on liver biopsy.

In the first phase, which aimed to validate the prognostic significance of ALP normalization, survival analysis endpoints included death, LT, or severe liver-related complications (such as esophageal variceal bleeding, ascites, hepatic encephalopathy, hepatorenal syndrome, or hepatocellular carcinoma). The assessment followed a hierarchical priority: death or LT was considered the endpoint if either occurred; otherwise, severe liver-related complications were used as the endpoint. The baseline for survival analysis was set at 12 months after initiation of UDCA monotherapy, corresponding to the timing of key biochemical assessments and serving as a critical point for evaluating long-term treatment efficacy. Patients who did not experience an endpoint event were censored at their last recorded follow-up.

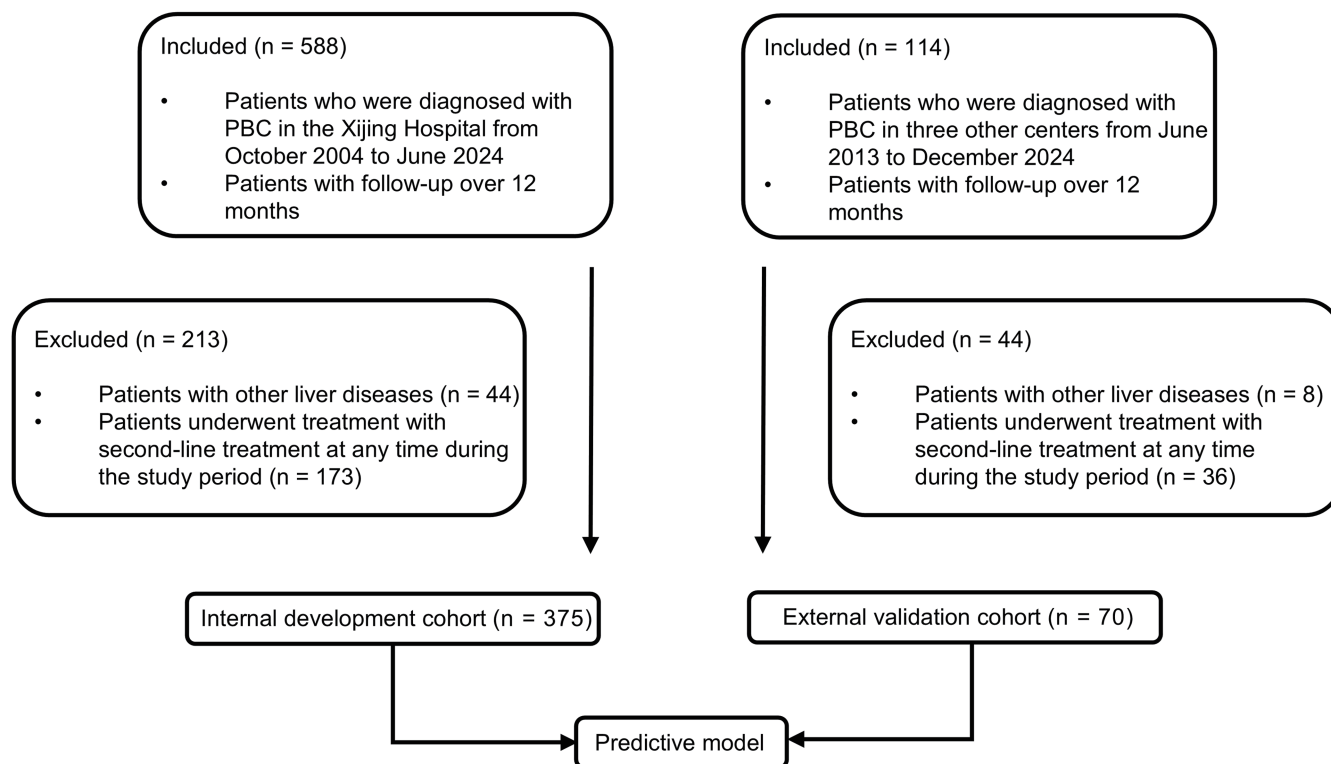
In the second phase of the study, a Sankey diagram was used to visualize changes in patient risk levels over time. Risk categories were defined based on ALP levels at each time point: high risk as $ALP > 1.67 \times$ ULN, medium risk as ALP between $1.0 \times$ ULN and $1.67 \times$ ULN, and low risk as $ALP \leq 1.0 \times$ ULN. These visualizations helped track patient transitions across risk strata during treatment.

The third phase of the study evaluated the predictive efficacy of established response criteria, using ALP normalization at 12 months following UDCA initiation as the primary endpoint. Predictive performance for ALP normalization at 12 months was assessed at 3 and 6 months using the following criteria: (1) the Mayo criteria, defined as $ALP < 2.0 \times$ ULN; (2) the Paris II criteria, defined as $ALP < 1.5 \times$ ULN, $AST < 1.5 \times$ ULN, and $TBIL < 1$ mg/dL; and (3) the Toronto criteria, defined as $ALP < 1.67 \times$ ULN.

Clinical data, including demographic characteristics, objective symptoms, and laboratory findings, were extracted from the electronic medical records of the included patients. Cirrhosis was diagnosed based on histological evidence or imaging findings obtained via ultrasound, computed tomography, or magnetic resonance imaging. Liver histology was staged as early (stage I/II) or late (stage III/IV) according to the Ludwig classification.¹⁹ Data for the training cohort were collected from Xijing Hospital, Fourth Military Medical University (Xi'an, China), while the validation cohort data were sourced from the Second Affiliated Hospital of Kunming Medical University (Kunming, China), Beijing You'an Hospital affiliated with Capital Medical University (Beijing, China), and the First Affiliated Hospital of China Medical University (Shenyang, China).

Follow-up assessments were conducted at 1, 3, 6, and 12 months after UDCA initiation, and annually thereafter. Liver-related clinical events (including death, LT, and hepatic complications) were ascertained through telephone follow-up.

The study was conducted in accordance with the principles of the Declaration of Helsinki (as revised in 2024) and was



Note: participants may meet more than one exclusion criterion; therefore the sum of the numbers for individual reasons may exceed the total number excluded

Fig. 1. Patient selection flow for internal and external cohorts.

approved by the Ethics Committee of Xijing Hospital (approval number: KY20253468-1). Informed consent was obtained from all participants. This study was reported according to the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines for cohort studies. A completed STROBE checklist is provided as Supplementary File 1.

Statistical analysis

Statistical analyses were performed using R (version 4.5.0) and Python (version 3.13.3). Categorical variables are presented as frequencies and percentages. Continuous variables are expressed as mean \pm standard deviation for normally distributed data, and as median with interquartile range for non-normally distributed data. Comparisons of continuous variables were conducted using the Student's t-test for normally distributed data and the Mann-Whitney U test for non-normally distributed data. Differences in categorical variables were assessed using the chi-squared test or Fisher's exact test, as appropriate.

Survival analyses were conducted using the Kaplan-Meier method, with group comparisons assessed via the log-rank test. Multivariate Cox proportional hazards models were applied to evaluate the association between covariates and the primary composite endpoint, with results reported as hazard ratios (HRs) and 95% confidence intervals (CIs). Patient risk stratification over time was visualized using a Sankey diagram, while temporal trends were assessed through segmented Poisson regression. The predictive performance of each response criterion for ALP normalization at 12 months post-UDCA initiation was evaluated by calculating sensitivity, specificity, positive predictive value (PPV), and nega-

tive predictive value (NPV). Regarding missing data, minimal missingness among covariates in the multivariate Cox model (Supplementary Fig. 1) rendered imputation unnecessary during the development phase; affected cases were excluded from the analysis. In contrast, for the longitudinal follow-up data, missing values were addressed using Markov chain Monte Carlo (MCMC) multiple imputation. This method, which relies on observed state transition probabilities, generated five complete datasets, and final estimates were obtained by pooling the results of these analyses. The robustness of the imputation approach was evaluated through sensitivity analyses comparing imputed and non-imputed outcomes (Supplementary Tables 2-3 and Supplementary Figs. 2-3). All statistical tests were two-sided, with a significance threshold of $P < 0.05$.

Results

Study population characteristics

Figure 1 illustrates the study flowchart. Of the 588 patients initially screened in the internal development cohort, 375 met the eligibility criteria. The remaining 213 were excluded based on the predefined exclusion criteria. The external validation cohort included 70 eligible patients from three regional centers. The demographic and clinical characteristics of both cohorts are summarized in Table 1. The mean age was 52.63 ± 9.69 years in the internal development cohort and 48.92 ± 9.95 years in the external validation cohort ($P = 0.038$). In terms of gender distribution, 88.5% of patients in the development cohort and 81.4% in the validation cohort were

Table 1. Baseline characteristics of the internal development and external validation cohorts

Characteristics	Internal development cohort (N = 375)	External validation cohort (N = 70)	P-value
Age, years	52.63 ± 9.69	48.92 ± 9.95	0.038
Female (%)	332 (88.5)	57 (81.4)	0.147
RBC, 10 ¹² /L	4.00 (3.65–4.29)	4.28 (3.75–4.57)	0.005
HGB, g/L	120.00 (109.50–132.00)	123.00 (113.25–135.75)	0.100
PLT, 10 ⁹ /L	141.0 (87.00–195.50)	203.50 (138.00–246.50)	0.002
ALT, IU/L	50.00 (29.00–85.00)	65.90 (45.10–97.00)	0.006
AST, IU/L	55.00 (36.00–83.00)	67.00 (42.00–106.00)	0.038
TBIL, μmol/L	16.85 (11.85–26.65)	18.90 (13.70–24.40)	0.158
ALP, IU/L	205.00 (127.00–330.25)	250.00 (152.00–392.00)	0.158
GGT, IU/L	209.00 (101.00–384.50)	312.00 (159.60–548.00)	0.010
IgG, g/L	15.50 (13.05–19.55)	16.00 (13.03–19.91)	0.399
IgM, g/L	3.01 (2.11–5.04)	3.17 (2.19–5.16)	0.537
Autoantibodies (positive, %)			
AMA	240/282 (85.1)	21/24 (87.5)	0.876
ANA	270/291 (92.8)	32/39 (82.1)	0.053
AMA-M2	176/288 (61.1)	26/37 (70.3)	0.234
gp210	88/276 (31.9)	10/21 (47.6)	0.123
sp100	35/276 (12.7)	7/21 (33.3)	0.012
Histological stage (%) ^a			
Early-stage (I–II)	226/257 (87.9)	13/17 (76.5)	0.145
Late-stage (III–IV)	31/257 (12.0)	4/17 (23.5)	0.145

Continuous variables are presented as mean ± standard deviation or median and interquartile range, while categorical data are shown as frequencies and percentages. ^aAvailable for 257 (68.5%) PBC patients in the internal development cohort and 17 (24.3%) PBC patients in the external validation cohort, respectively. ALP, alkaline phosphatase; ALT, alanine transaminase; AST, aspartate aminotransferase; AMA, anti-mitochondrial antibody; AMA-M2, subtype of anti-mitochondrial antibody; ANA, antinuclear antibody; GGT, gamma-glutamyl transferase; HGB, hemoglobin; IgG, immunoglobulin G; IgM, immunoglobulin M; PBC, primary biliary cholangitis; PLT, platelet; RBC, red blood cell.

female; this difference was not statistically significant ($P = 0.147$). For histological staging, 87.9% (226/257) of patients in the development cohort were classified as stage I–II, while 12.0% (31/257) were classified as stage III–IV according to the Ludwig system. In the validation cohort, 76.5% (13/17) of patients were classified as stage I–II, and 23.5% (4/17) as stage III–IV. The difference in histological staging between the development and validation cohorts was not statistically significant ($P = 0.145$). Additionally, the development cohort exhibited lower platelet and red blood cell counts, a lower positive rate of anti-sp100 antibodies, and reduced levels of ALT, AST, and GGT compared with the validation cohort ($P = 0.002$, $P = 0.005$, $P = 0.012$, $P = 0.006$, $P = 0.038$, and $P = 0.010$, respectively).

Notably, the internal development cohort was followed for a median duration of 76.9 months. During this period, 69 serious clinical events were recorded, including 22 deaths and two liver transplants. This extensive follow-up provided robust survival data for subsequent analyses.

Prognostic significance of ALP normalization after 12 months of treatment

Kaplan–Meier analyses (Fig. 2) were conducted to assess complication-free survival rates stratified by ALP levels at one year ($>1.67 \times \text{ULN}$, $1.0\text{--}1.67 \times \text{ULN}$, and $\leq 1.0 \times \text{ULN}$). Complication-free survival was inversely correlated with ALP

levels, with rates of 62.8% in the $>1.67 \times \text{ULN}$ group, 79.8% in the $1.0\text{--}1.67 \times \text{ULN}$ group, and 89.8% in the $\leq 1.0 \times \text{ULN}$ group. Patients achieving ALP normalization ($\leq 1.0 \times \text{ULN}$) demonstrated the lowest risk of LT, liver-related mortality, or hepatic complications. Statistically significant differences were found between the normalized group and those with ALP levels between 1.0 and $1.67 \times \text{ULN}$ ($P = 0.016$), as well as between the medium-risk group and those with ALP levels exceeding $1.67 \times \text{ULN}$ ($P = 0.014$).

Following the Kaplan–Meier analysis, a multivariable Cox regression was conducted focusing on patients with ALP levels below $1.67 \times \text{ULN}$ after 12 months of treatment (Table 2). Several binary variables were examined, including ALP $> 1.0 \times \text{ULN}$, age > 55 years, TBIL $> 0.6 \times \text{ULN}$, and ALT $> 1.0 \times \text{ULN}$. ALP levels between 1.0 and $1.67 \times \text{ULN}$ emerged as a significant predictor of adverse outcomes (HR = 2.27; 95% CI: 1.21–4.26; $P = 0.011$), followed by age over 50 years (HR = 4.27; 95% CI: 1.66–10.96; $P = 0.003$). In contrast, neither ALT $> 1.0 \times \text{ULN}$ (HR = 0.45; 95% CI: 0.16–1.27; $P = 0.133$) nor TBIL $> 0.6 \times \text{ULN}$ (HR = 2.37; 95% CI: 0.73–7.73; $P = 0.152$) was significantly associated with increased risk.

Analysis of the critical time window for patient risk transitions

A Sankey diagram was employed to visualize the dynamic shifts in risk levels among 375 PBC patients over a 12-month

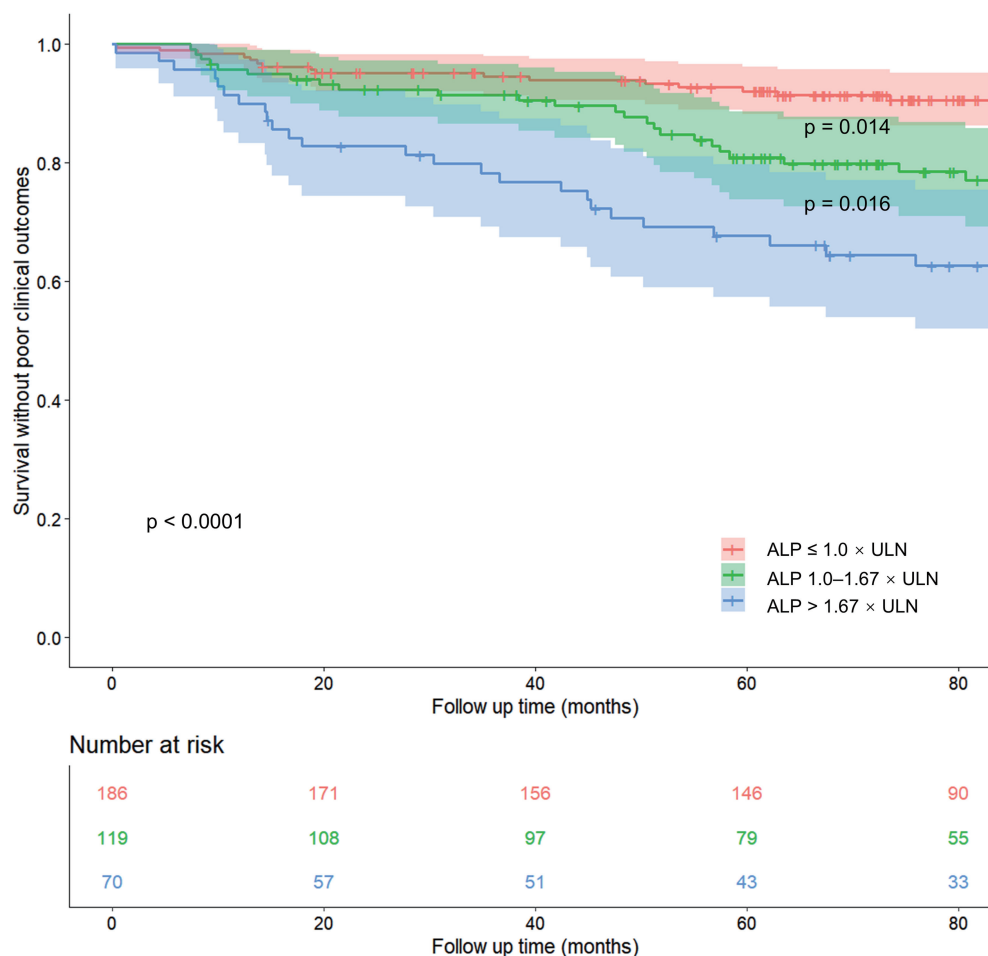


Fig. 2. Survival without poor clinical outcomes stratified by ALP levels at entry. Kaplan-Meier curves illustrate complication-free survival in patients grouped by ALP levels at entry: ALP $\leq 1.0 \times \text{ULN}$ (blue-gray line), $1.0\text{--}1.67 \times \text{ULN}$ (green-gray line), and $> 1.67 \times \text{ULN}$ (red-gray line). Shaded areas represent the 95% confidence intervals for each group. The p-value (< 0.0001) from the log-rank test indicates a statistically significant difference in survival among the groups. Pairwise comparisons yielded $P = 0.016$ for ALP $\leq 1.0 \times \text{ULN}$ versus $1.0\text{--}1.67 \times \text{ULN}$ and $P = 0.014$ for $1.0\text{--}1.67 \times \text{ULN}$ versus $> 1.67 \times \text{ULN}$. ALP, alkaline phosphatase; ULN, upper limit of normal.

follow-up period (Fig. 3). Risk stratification was based on ALP levels, categorized as high risk ($>1.67 \times \text{ULN}$), medium risk ($1.0\text{--}1.67 \times \text{ULN}$), and low risk ($\leq 1.0 \times \text{ULN}$). The width of the arrows corresponds to the number of patients transitioning between risk groups. Missing data were addressed using MCMC multiple imputation.

The high-to-medium risk transition rate showed a biphasic pattern over the 12-month follow-up period, with an initial rapid decline followed by a plateau (Fig. 4A). Segmented Poisson regression analysis identified a significant joinpoint at 3.73 months ($P = 0.003$; AIC improvement = 7.8, Supple-

mentary Table 4). The trend was divided into two distinct periods: an initial rapid decline phase from baseline to approximately 3.7 months, followed by a plateau phase in which the transition rate remained stable (monthly percent change (MPC) = -0.09% , 95% CI: -3.52 to 0.45% , Supplementary Table 4), highlighting that the first 3 months following intervention constitute a critical window for transitioning high-risk patients.

The downward trend observed in the moderate-risk patient group followed a similar pattern, with segmented Poisson regression identifying a changepoint at approximately

Table 2. Time-dependent, multivariable-adjusted Cox regression analysis of poor clinical outcomes in patients with ALP $< 1.67 \times \text{ULN}$

Parameter	Events/N	HR	SE	z	$P > z $	95% CI
Age > 50 y	38/190	4.2686	0.4813	3.015	0.003	1.66–10.96
ALP $> 1.0 \times \text{ULN}$	24/119	2.2686	0.3210	2.552	0.011	1.21–4.26
TBIL $> 0.6 \times \text{ULN}$	39/242	2.3703	0.6028	1.432	0.152	0.73–7.73
ALT $> 1.0 \times \text{ULN}$	4/54	0.4503	0.5310	-1.503	0.133	0.16–1.27

Results are ranked in order of decreasing statistical significance. A total of 305 patients with ALP $< 1.67 \times \text{ULN}$ were included, of whom 43 experienced poor clinical outcomes. ALP, alkaline phosphatase; ALT, alanine transaminase; TBIL, total bilirubin; ULN, upper limit of the normal range.

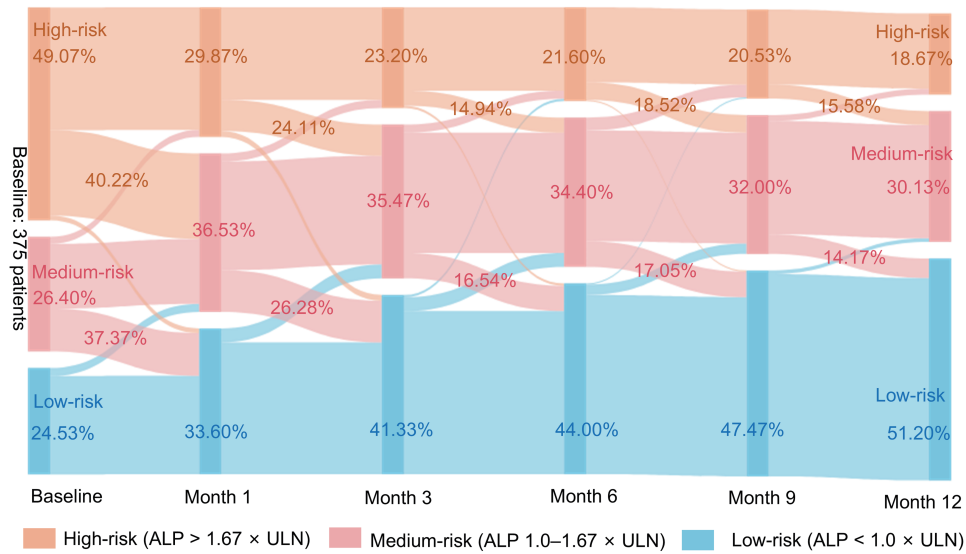


Fig. 3. Analysis of the dynamic migration of risk levels within one year. The Sankey diagram illustrates transitions between risk categories over 12 months for 375 PBC patients who underwent six follow-up assessments (baseline and at 1, 3, 6, 9, and 12 months). Categories are represented by the band width that correspond to the number of patients, and labels show absolute counts (N) and percentages (%), relative to the total population of the preceding node. Risk categories are defined as follows: high risk (orange, ALP > 1.67 × ULN), medium risk (pink, ALP 1.0–1.67 × ULN), and low risk (blue, ALP ≤ 1.0 × ULN). Missing follow-up data were imputed using Markov chain Monte Carlo multiple imputation. Details of the original dataset and sensitivity analyses are provided in Supplementary Tables 2–3. Minor transitions (<5% of total flow) were aggregated as “other pathways” and omitted from the labels. ALP, alkaline phosphatase; ULN, upper limit of normal; PBC, primary biliary cholangitis.

5.5 months ($P = 0.043$, Fig. 4B). The trend evolved through two periods: a declining phase from baseline to about 5.5 months, after which the rate of decline substantially weakened, forming a plateau (MPC = -0.30% , 95% CI: -10.92 to 0.36% ; Supplementary Table 5). These findings suggest that intervention strategies should be adjusted around month 6 to address the observed plateau in treatment efficacy. The conclusions based on uninterpolated data are consistent (Sup-

plementary Tables 2–3, Supplementary Figs. 2–3), confirming the robustness and reliability of the results.

Predictive performance of biochemical criteria for ALP normalization

Based on the critical time windows identified at 3 and 6 months for changes in patient risk, we calculated sensitivity, specificity, PPV, and NPV to assess the effectiveness of the

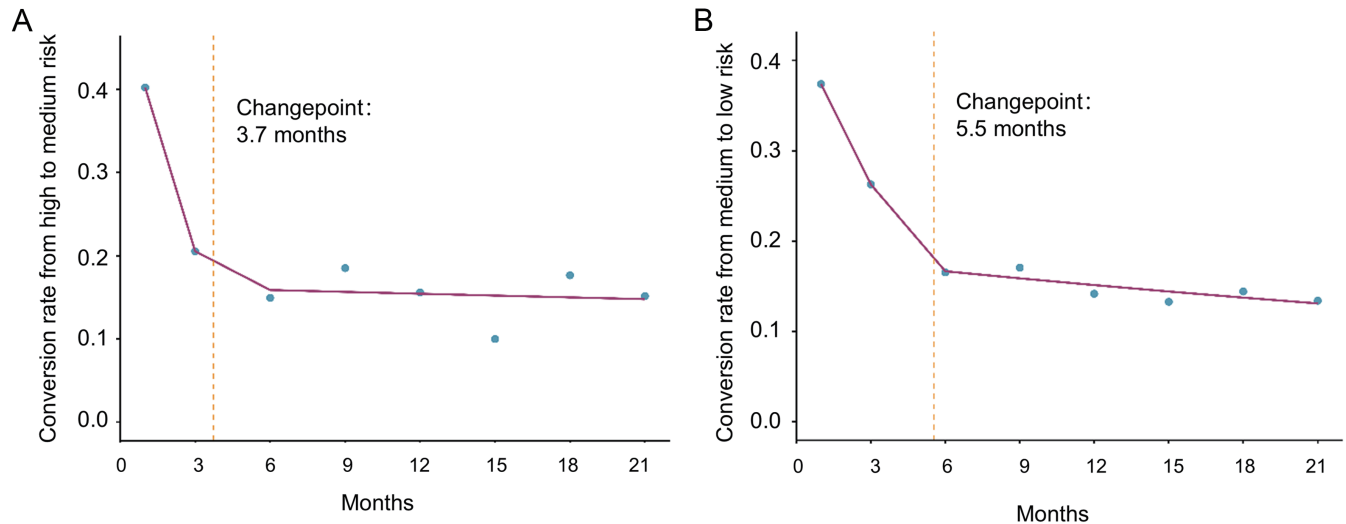


Fig. 4. Segmented Poisson regression analysis of recovery rates over 24-month follow-up. (A) Transition from high-risk to medium-risk status. (B) Transition from medium-risk to low-risk status. The segmented Poisson regression plots illustrate the temporal trends in transition rates over the 24-month period. Solid lines represent the fitted models; circles indicate observed recovery rates. Vertical dashed lines mark the automatically estimated join points. In panel A, the transition rate from high- to medium-risk showed a decline phase from baseline to the join point at 3.73 months ($P = 0.003$), followed by a plateau phase thereafter. In panel B, the transition rate from medium- to low-risk also displayed two distinct phases: a decline from baseline to the join point at 5.55 months ($P = 0.043$), after which the trend stabilized. Follow-up time points (interval starting points) included baseline, 1, 3, 6, 9, 12, 15, 18, and 21 months. Data were derived from post-imputation datasets; raw data and sensitivity analyses are detailed in Supplementary Figure 3.

Table 3. Predictive performance of biochemical response criteria at months 3 and 6 for 12-month ALP normalization in the PBC internal development cohort

	Sensi- tivity	Speci- ficity	PPV	NPV
Month 3				
Mayo criteria	0.98	0.33	0.60	0.96
Paris II criteria	0.67	0.75	0.73	0.69
Toronto criteria	0.98	0.44	0.64	0.95
Month 6				
Mayo criteria	1.00	0.31	0.60	1.00
Paris II criteria	0.71	0.73	0.73	0.71
Toronto criteria	1.00	0.44	0.65	1.00

PPV, positive predictive value; NPV, negative predictive value.

three criteria.

The calculation method described by Corpechot *et al.* was employed, defining biochemical response as a positive test and ALP normalization at 12 months as the event of interest.⁷

At month 3, both the Mayo and Toronto criteria demonstrated excellent NPVs (≥ 0.95) and sensitivity (0.98), indicating that patients failing these criteria were highly unlikely to achieve normalization. Notably, the Toronto criteria exhibited superior specificity compared to the Mayo criteria (44% vs. 33%), making it more advantageous for early screening to balance timely intervention with the avoidance of over-treatment. By month 6, while the Mayo and Toronto criteria maintained 100% NPV, their specificity remained limited. In contrast, the Paris II criteria achieved the highest specificity (73%), providing superior accuracy in identifying high-risk non-responders who require intensified treatment, thereby minimizing the risk of overlooking patients in need of escalation (Table 3).

External validation results further supported the robustness of our findings: the Toronto criteria achieved an NPV of 0.94 at month 3, and the Paris II criteria demonstrated a specificity of 63% at 6 months, comparable to the 73% specificity observed in the development cohort. These results confirm the consistency of the core findings across different populations (Table 4).

Discussion

Normalization of serum ALP levels is associated with better long-term outcomes in PBC patients treated with UDCA¹⁴ and serves as an important prognostic indicator.²⁰ This study confirms that UDCA-treated PBC patients with ALP levels between 1.0 and $1.67 \times \text{ULN}$ remain at risk of poor outcomes. In addition, age over 50 years was identified as an independent risk factor for poor prognosis among patients with $\text{ALP} < 1.67 \times \text{ULN}$. Importantly, we identified two critical dynamic risk transition windows: (1) after 3 months of treatment, patients remaining at high risk ($\text{ALP} > 1.67 \times \text{ULN}$) are less likely to transition to medium risk ($\text{ALP} 1.0\text{--}1.67 \times \text{ULN}$); and (2) after 6 months, patients remaining at medium risk exhibit a significant decline in normalization rates. Based on these time windows, we further assessed the potential for ALP normalization in UDCA-treated PBC patients: (1) at 3 months, patients with $\text{ALP} > 1.67 \times \text{ULN}$ have a markedly reduced likelihood of ALP normalization; and (2) at 6 months, non-responders according to the Paris II criteria ($\text{ALP} < 1.5 \times \text{ULN}$, $\text{AST} < 1.5 \times \text{ULN}$,

Table 4. Predictive performance of biochemical response criteria at months 3 and 6 for 12-month ALP normalization in the PBC external validation cohort

	Sensi- tivity	Speci- ficity	PPV	NPV
Month 3				
Mayo criteria	0.91	0.52	0.50	0.92
Paris II criteria	0.58	0.87	0.70	0.80
Toronto criteria	0.92	0.70	0.61	0.94
Month 6				
Mayo criteria	0.90	0.32	0.41	0.86
Paris II criteria	0.50	0.63	0.42	0.71
Toronto criteria	0.80	0.42	0.42	0.80

PPV, positive predictive value; NPV, negative predictive value.

and $\text{TBIL} < 1 \text{ mg/dL}$) similarly have significantly limited potential for normalization.

Data from our study cohort confirmed that patients with normalized ALP levels had better outcomes than those with ALP between 1.0 and $1.67 \times \text{ULN}$.^{13,14} Murillo-Pérez *et al.* initially reported that both $\text{TBIL} \leq 0.6 \times \text{ULN}$ and $\text{ALP} \leq 1.0 \times \text{ULN}$ were associated with improved survival in the overall PBC population, and that $\text{TBIL} \leq 0.6 \times \text{ULN}$ conferred a protective effect even among patients with $\text{ALP} < 1.67 \times \text{ULN}$.¹³ However, proteomic research by Jones *et al.* found that, even under the most stringent current response criteria (the Paris II criteria), a subset of patients labeled as “UDCA responders” exhibits persistently elevated disease biomarkers compared to those achieving ALP and bilirubin normalization, with no significant difference observed between patients with $\text{TBIL} \leq 0.6 \times \text{ULN}$ and $0.6\text{--}1 \times \text{ULN}$ when ALP levels are in the normal range.²¹ Our multivariate analysis revealed that, after adjusting for confounding factors such as age, ALP normalization emerged as an independent protective factor within the subgroup with $\text{ALP} < 1.67 \times \text{ULN}$, whereas $\text{TBIL} \leq 0.6 \times \text{ULN}$ was not a consistently stable protective indicator among responders. Therefore, complete ALP normalization should be prioritized as the primary treatment goal, with precise TBIL management as a secondary objective, especially in cases where ALP targets are unmet. This dependency likely arises from the distinct pathological roles of the two markers²²⁻²⁴: ALP elevation directly reflects cholestatic activity through damaged cholangiocyte membrane shedding and progressive bile duct damage. Conversely, elevated TBIL indicates impaired bilirubin absorption or conjugation in hepatocytes, or severe excretory failure, reflecting advanced hepatocellular dysfunction. Thus, ALP normalization captures early, modifiable ductal pathology, whereas TBIL reduction may signal irreversible hepatic dysfunction.

Notably, our data indicated that within the subgroup of responders with $\text{ALP} < 1.67 \times \text{ULN}$, multivariate Cox regression analysis identified age over 50 years as an independent risk factor for poor prognosis (HR = 4.27; 95% CI: 1.66–10.96; $P = 0.003$), highlighting the need for increased clinical vigilance in older patients.

The duration and magnitude of biochemical abnormalities significantly impact prognosis in PBC patients.²⁵ Therefore, patients who fail to achieve biochemical normalization should be identified and managed promptly. Our longitudinal risk dynamics analysis further identified two critical time windows: 3 months and 6 months after initial UDCA treatment. The

minimal direct high-to-low risk transitions (<3%) reinforce the concept that risk reduction in PBC is typically a gradual process rather than an abrupt shift. Thus, ongoing monitoring within the first 6 months after initial UDCA treatment is highly recommended. Our data are also consistent with previous reports of an early rapid decline in biochemical markers following UDCA initiation²⁶; however, our segmented Poisson regression provides a more granular dynamic perspective, indicating that this initial decline phase extends to approximately 3 months before reaching a plateau. Herein, we emphasize the importance of this early period for assessing drug sensitivity in baseline high-risk patients and establish a high NPV threshold (ALP $\geq 1.67 \times$ ULN) to accurately identify those unlikely to achieve biochemical normalization. Zhang *et al.*²⁷ suggested that predictive efficacy is reached at 6 months, marking the stabilization of patient indicators post-treatment. Through trend analysis, our study highlights the critical significance of this six-month time point for biochemical normalization. We also note that the Paris II criteria allow for effective screening of patients requiring intervention at this stage.

Of note, Corpechot *et al.* reported that the survival benefit associated with ALP normalization was restricted to patients with liver stiffness measurement > 10 kPa and age < 62 years, suggesting that this benefit may be most pronounced in those with more advanced fibrosis. In contrast, our cohort predominantly comprised patients with early-stage fibrosis (Ludwig stage I-II, 87.9%), yet ALP normalization was still independently associated with improved complication-free survival. This discrepancy may reflect differences in patient selection, disease stage distribution, or follow-up duration between the two studies. Future studies should prospectively examine whether the prognostic value of ALP normalization is modified by fibrosis severity and patient age.

In summary, this study proposes a time-window refinement model grounded in risk evolution, offering an evidence-based framework to guide individualized intensive therapy. However, several limitations remain: (a) the efficacy of this time window-guided approach in improving hard clinical endpoints requires validation in larger prospective cohorts; (b) the limited sample size of the liver pathology repository restricts deeper investigation into the mechanisms underlying the conversion from histological to biochemical response; (c) several baseline differences were observed between the internal and external cohorts, including lower platelet counts and liver enzyme levels in the development cohort, which may partly reflect differences in disease severity or sample collection protocols across centers. Additionally, the relatively small size of the external validation cohort ($n = 70$) may limit the generalizability of the validation findings. Nevertheless, such an approach will allow the development of a stratified intervention strategy by identifying critical time windows—the third and sixth months of treatment—that balance precision therapy with the avoidance of overtreatment.

Conclusions

This study establishes ALP normalization ($\leq 1.0 \times$ ULN) as a superior therapeutic target in PBC, demonstrating a 10.0 percentage-point improvement in complication-free survival (89.8% vs. 79.8% at median follow-up) compared with patients maintaining ALP levels between 1.0 and $1.67 \times$ ULN. These findings indicate that current biochemical response criteria may inadequately stratify risk, as a substantial proportion of patients classified as “responders” remain at elevated risk for adverse outcomes.

Our longitudinal analysis identified two critical intervention

windows for treatment intensification. At 3 months, patients with ALP $\geq 1.67 \times$ ULN show markedly reduced normalization probability (NPV 95% by Toronto criteria), warranting early consideration of second-line therapies. At 6 months, failure to meet Paris II criteria reliably identifies patients unlikely to achieve normalization (specificity 73%). This dual time window-driven strategy enables timely intervention while avoiding premature treatment escalation.

The clinical implications are threefold: treatment goals should prioritize complete ALP normalization; biochemical monitoring at 3 and 6 months provides actionable decision points for therapy intensification; and older patients (>50 years) warrant enhanced surveillance despite meeting conventional response criteria. While prospective validation is needed, this time window-based framework offers a practical approach to implementing precision medicine in PBC management, potentially reducing liver-related complications and improving long-term outcomes.

Funding

This study was supported by Prevention and Control of Emerging and Major Infectious Diseases-National Science and Technology Major Project (2025ZD01906300 & 2025ZD01906304), the National Natural Science Foundation of China (No. 822 70551 to Ying Han, 82200577 to Yansheng Liu), the Innovation Capacity Support Program of Shaanxi Province (No. 2024RS-CXTD-79 to Yulong Shang, 2025ZC-KJXX-109 to Yansheng Liu), and the Key Research and Development Program of Shaanxi (2023-ZDLSF-33 to Yulong Shang).

Conflict of interest

YH has been an Editorial Board Member of *Journal of Clinical and Translational Hepatology* since 2013. The other authors have no conflict of interests related to this publication.

Author contributions

Conceptualization (HZ, YsL, YT, NW, YmL, MEG, YS, YH), methodology (HZ, YsL, LZ, RS, XW, JD, GJ, PSCL), formal analysis (HZ, YsL), investigation (HZ, YsL, YT, NW, YmL, YIL, CH, JD, YF), writing - original draft (HZ, YsL, YT, NW, YmL), writing - review and editing (HZ, YsL, MEG, YS, YH), project administration (HZ, YsL, LZ, RS, XW, JD, GJ, PSCL), data curation (HZ, YsL, YT, NW, YmL, YIL, CH, JD, YF, LZ, RS, XW, JD, GJ, PSCL), visualization (HZ, YsL), validation (YT, NW, YmL, YIL, CH, JD, YF), supervision (MEG, YS, YH), resources (YS, YH), and funding acquisition (YS, YH). All authors approved the final manuscript.

Ethical statement

This study was approved by the Ethics Committee of Xijing Hospital (approval number: KY20253468-1). All procedures were conducted in accordance with the ethical standards of the Declaration of Helsinki (as revised in 2024) and its later amendments. Informed consent was obtained from all participants.

Data sharing statement

The data that support the findings of this study are not publicly available due to privacy and ethical restrictions. The datasets contain sensitive patient information that cannot be shared, even in de-identified form, according to institutional data protection policies.

References

[1] Hirschfield GM, Dyson JK, Alexander GJM, Chapman MH, Collier J, Hübscher S, *et al*. The British Society of Gastroenterology/UK-PBC primary biliary cholangitis treatment and management guidelines. *Gut* 2018;67(9):1568–1594. doi:10.1136/GUTJNL-2017-315259, PMID:29593060.

[2] Lindor KD, Bowlus CL, Boyer J, Levy C, Mayo M. Primary Biliary Cholangitis: 2018 Practice Guidance from the American Association for the Study of Liver Diseases. *Hepatology* 2019;69(1):394–419. doi:10.1002/hep.30145, PMID:30070375.

[3] Corpechot C, Abenavoli L, Rabahi N, Chrétien Y, Andréani T, Johanet C, *et al*. Biochemical response to ursodeoxycholic acid and long-term prognosis in primary biliary cirrhosis. *Hepatology* 2008;48(3):871–877. doi:10.1002/hep.22428, PMID:18752324.

[4] Parés A, Caballería L, Rodés J. Excellent long-term survival in patients with primary biliary cirrhosis and biochemical response to ursodeoxycholic acid. *Gastroenterology* 2006;130(3):715–720. doi:10.1053/j.gastro.2005.12.029, PMID:16530513.

[5] Tanaka A, Hirohara J, Nakano T, Matsumoto K, Chazouillères O, Takikawa H, *et al*. Association of bezafibrate with transplant-free survival in patients with primary biliary cholangitis. *J Hepatol* 2021;75(3):565–571. doi:10.1016/j.jhep.2021.04.010, PMID:33882268.

[6] Angulo P, Lindor KD, Thorneau TM, Jorgensen RA, Malinchoc M, Kamath PS, *et al*. Utilization of the Mayo risk score in patients with primary biliary cirrhosis receiving ursodeoxycholic acid. *Liver* 1999;19(2):115–121. doi:10.1111/j.1478-3231.1999.tb00020.x, PMID:10220741.

[7] Corpechot C, Chazouillères O, Poupon R. Early primary biliary cirrhosis: biochemical response to treatment and prediction of long-term outcome. *J Hepatol* 2011;55(6):1361–1367. doi:10.1016/j.jhep.2011.02.031, PMID:21703194.

[8] Kuiper EM, Hansen BE, de Vries RA, den Ouden-Muller JW, van Ditzhuijsen TJ, Haagsma EB, *et al*. Improved prognosis of patients with primary biliary cirrhosis that have a biochemical response to ursodeoxycholic acid. *Gastroenterology* 2009;136(4):1281–1287. doi:10.1053/j.gastro.2009.01.003, PMID:19208346.

[9] Kumagi T, Guindi M, Fischer SE, Arenovich T, Abdalian R, Coltescu C, *et al*. Baseline ductopenia and treatment response predict long-term histological progression in primary biliary cirrhosis. *Am J Gastroenterol* 2010;105(10):2186–2194. doi:10.1038/ajg.2010.216, PMID:20502446.

[10] Azemoto N, Abe M, Murata Y, Hiasa Y, Hamada M, Matsuura B, *et al*. Early biochemical response to ursodeoxycholic acid predicts symptom development in patients with asymptomatic primary biliary cirrhosis. *J Gastroenterol* 2009;44(6):630–634. doi:10.1007/s00535-009-0051-9, PMID:19370305.

[11] Nevens F, Andreone P, Mazzella G, Strasser SI, Bowlus C, Invernizzi P, *et al*. A Placebo-Controlled Trial of Obeticholic Acid in Primary Biliary Cholangitis. *N Engl J Med* 2016;375(7):631–643. doi:10.1056/NEJMoa1509840, PMID:27532829.

[12] Mells G, Jones D, Diggall K. P23 predicted risk of end-stage liver disease utilizing the UK-PBC risk score with continued standard of care and subsequent addition of OCA for 60 months in patients with PBC. *Gut* 2020;69:A18–A19. doi:10.1136/gutjnl-2020-BASL.34.

[13] Murillo Perez CF, Harms MH, Lindor KD, van Buuren HR, Hirschfield GM, Corpechot C, *et al*. Goals of Treatment for Improved Survival in Primary Biliary Cholangitis: Treatment Target Should Be Bilirubin Within the Normal Range and Normalization of Alkaline Phosphatase. *Am J Gastroenterol* 2020;115(7):1066–1074. doi:10.14309/ajg.0000000000000557, PMID:32618657.

[14] Corpechot C, Lemoine S, Soret PA, Hansen B, Hirschfield G, Gulamhusein A, *et al*. Adequate versus deep response to ursodeoxycholic acid in primary biliary cholangitis: To what extent and under what conditions is normal alkaline phosphatase level associated with complication-free survival gain? *Hepatology* 2024;79(1):39–48. doi:10.1097/HEP.0000000000000529, PMID:37399238.

[15] Kowdley KV, Bowlus CL, Levy C, Akarca US, Alvares-da-Silva MR, Andreone P, *et al*. Efficacy and Safety of Elafibranor in Primary Biliary Cholangitis. *N Engl J Med* 2024;390:795–805. doi:10.1056/NEJMoa2306185, PMID:37962077.

[16] Hirschfield GM, Bowlus CL, Mayo MJ, Kremer AE, Vierling JM, Kowdley KV, *et al*. A Phase 3 Trial of Seladelpar in Primary Biliary Cholangitis. *N Engl J Med* 2024;390(9):783–794. doi:10.1056/NEJMoa2312100, PMID:38381664.

[17] Liu Y, Guo G, Zheng L, Sun R, Wang X, Deng J, *et al*. Effectiveness of Fenofibrate in Treatment-Naive Patients With Primary Biliary Cholangitis: A Randomized Clinical Trial. *Am J Gastroenterol* 2023;118(11):1973–1979. doi:10.14309/ajg.0000000000002238, PMID:36892506.

[18] European Association for the Study of the Liver. EASL Clinical Practice Guidelines: The diagnosis and management of patients with primary biliary cholangitis. *J Hepatol* 2017;67(1):145–172. doi:10.1016/j.jhep.2017.03.022, PMID:28427765.

[19] Ludwig J, Dickson ER, McDonald GS. Staging of chronic nonsuppurative destructive cholangitis (syndrome of primary biliary cirrhosis). *Virchows Arch A Pathol Anat Histol* 1978;379(2):103–112. doi:10.1007/BF00432479, PMID:150690.

[20] Lammers WJ, van Buuren HR, Hirschfield GM, Janssen HL, Invernizzi P, Mason AL, *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(6):1338–1349. e5, quiz e15. doi:10.1053/j.gastro.2014.08.029, PMID:25160979.

[21] Jones DEJ, Wetten A, Barron-Millar B, Ogle L, Mells G, Flack S, *et al*. The relationship between disease activity and UDCA response criteria in primary biliary cholangitis: A cohort study. *EBioMedicine* 2022;80:104068. doi:10.1016/j.ebiom.2022.104068, PMID:35609437.

[22] Alvaro D, Benedetti A, Marucci L, Delle Monache M, Monterubbianesi R, Di Cosimo E, *et al*. The function of alkaline phosphatase in the liver: regulation of intrahepatic biliary epithelium secretory activities in the rat. *Hepatology* 2000;32(2):174–184. doi:10.1053/jhep.2000.9078, PMID:10915721.

[23] Ramírez-Mejía MM, Castillo-Castañeda SM, Pal SC, Qi X, Méndez-Sánchez N. The Multifaceted Role of Bilirubin in Liver Disease: A Literature Review. *J Clin Transl Hepatol* 2024;12(11):939–948. doi:10.14218/JCTH.2024.00156, PMID:39544246.

[24] Xue R, Meng Q, Dong J, Li J, Yao Q, Zhu Y, *et al*. Clinical performance of stem cell therapy in patients with acute-on-chronic liver failure: a systematic review and meta-analysis. *J Transl Med* 2018;16(1):126. doi:10.1186/s12967-018-1464-0, PMID:29747694.

[25] Kowdley KV, Victor DW 3rd, MacEwan JP, Nair R, Levine A, Hernandez J, *et al*. Longitudinal Relationship Between Elevated Liver Biochemical Tests and Negative Clinical Outcomes in Primary Biliary Cholangitis: A Population-Based Study. *Aliment Pharmacol Ther* 2025;61(11):1775–1784. doi:10.1111/apt.70120, PMID:40176393.

[26] Yang C, Guo G, Li B, Zheng L, Sun R, Wang X, *et al*. Prediction and evaluation of high-risk patients with primary biliary cholangitis receiving ursodeoxycholic acid therapy: an early criterion. *Hepatol Int* 2023;17(1):237–248. doi:10.1007/s12072-022-10431-7, PMID:36309918.

[27] Zhang LN, Shi TY, Shi XH, Wang L, Yang YJ, Liu B, *et al*. Early biochemical response to ursodeoxycholic acid and long-term prognosis of primary biliary cirrhosis: results of a 14-year cohort study. *Hepatology* 2013;58(1):264–272. doi:10.1002/hep.26322, PMID:23408380.