Evaluating the Association Between Surrogate Endpoints and Liver-Related Outcomes in Primary Biliary Cholangitis: A Systematic Review and Meta-Analysis

CO193

Dilip Makhija[1], Marvin Rock[1], Chong H Kim[1], Mirko von Hein[2], Ryan Thaliffdeen[1], Ankita Sood[3], Ritesh Dubey[3], Barinder Singh[3]

[1]Gilead Sciences, Inc., Foster City, CA, USA, [2]Gilead Sciences, London, United Kingdom, [3]Pharmacoevidence, Mohali, India

Conclusions

- The current systematic literature review and meta-analysis demonstrates a robust and consistent association between alkaline phosphatase (ALP)-based surrogate markers, total bilirubin, and composite response definitions with long-term adverse liver outcomes such as liver transplantation (LT), cirrhosis, and mortality in primary biliary cholangitis (PBC)
- The study findings demonstrate the prognostic value of ALP reduction and normalization in guiding clinical decisions and monitoring treatment efficacy. However, evidence on ALP normalization remains limited, underscoring the need for further research, particularly as normalization is increasingly recognized as an important therapeutic goal given that even mildly elevated ALP levels above the normal range are associated with an increased risk of LT and mortality
- The results reinforce the use of ALP and total bilirubin as clinically meaningful biomarkers for guiding therapeutic decision-making, benchmarking treatment efficacy, and supporting health technology assessment and economic evaluations in PBC

Plain Language Summary

- Primary biliary cholangitis (PBC) is a rare, long-term autoimmune disease characterized by progressive immune mediated destruction of the intrahepatic bile ducts
- Outcomes such as cirrhosis, liver transplantation, and death often take many years to occur, making them difficult to measure directly in clinical studies. For this reason, well-established biochemical markers of liver function are widely used to assess disease progression and predict long-term outcomes
- This review included 49 studies evaluating the relationship between changes in key biochemical markers, such as alkaline phosphatase and bilirubin, and the risk of long-term liver-related complications
- Patients who did not show improvement or normalization of alkaline phosphatase had an approximately 2.5-fold higher risk of developing serious liver-related outcomes
- Patients who do not achieve a treatment response based on composite response criteria were more likely at a substantially higher risk of adverse outcomes
- Elevated total bilirubin was also strongly associated with adverse long-term outcomes across multiple studies, with patients failing to achieve improvement experiencing a two- to five-fold higher risk of events such as liver transplantation, hepatic decompensation, and mortality
- These results are consistent with our previous study, which demonstrated a strong association between lack of improvement in key surrogate markers and an increased risk of liver transplantation or death

References: 1. Kaplan MM, et al. *N Engl J Med;* 2005; 353(12):1261-73.
2. US FDA. Table of Surrogate Endpoints That Were the Basis of Drug Approval or Licensure. Accessed on 8 April 2026
3. Makhija D, at al. *Value Health*. 2025;28(S1):CO170
4. Makhija D, at al. *Value Health*. 2025;28(S1):CO156
5. Hilscher M, et al. *Ann. Hepatol*; 2016; 15(2):246-253.

Acknowledgment: This study was sponsored by Gilead Sciences, Inc. Medical writing assistance was provided by Pharmacoevidence and was funded by Gilead Sciences, Inc.

Disclosures: DM, MR, CK, MH, and **RT** are employees of Gilead Sciences, Inc. and may own stock in Gilead Sciences, Inc. **AS, RD** and **BS** are the employees of Pharmacoevidence.

Correspondence: Barinder Singh, Barinder.Singh@pharmacoevidence.com

Conflict of interest: DM, MR, CK, MH, RT, AS, RD, and BS, the authors, declare that they have no conflict of interest

Introduction

- Primary biliary cholangitis (PBC) is a rare, autoimmune cholestatic liver disease with slow progression, complicating direct measurement of long-term outcomes such as liver transplantation (LT) and mortality[1]
- The Food and Drug Administration has recognized a composite endpoint comprising of alkaline phosphatase (ALP) and total bilirubin for accelerated approval in PBC. These surrogate markers serve as valuable indicators of disease progression and treatment effectiveness[2]
- Previous findings have shown that patients who failed to achieve response based on ALP response, ALP normalization and composite criteria had a significantly higher risk of LT and mortality[3,4]
- Utilizing these surrogate endpoints allows for more timely and practical assessments of treatment efficacy, helping to guide clinical decision-making and improve the management of PBC by providing early insights into patient prognosis[5]
- This systematic literature review (SLR) and meta-analysis (MA) evaluates evidence linking ALP-based surrogate markers, total bilirubin, and composite response definitions to long-term adverse liver outcomes (LT, cirrhosis, hepatocellular carcinoma [HCC], hepatic decompensation, progression, mortality)

Methods

- Key biomedical databases were searched from inception to September 2024, to identify publications evaluating the association between surrogate endpoints and long-term liver outcomes in PBC. The prespecified eligibility criteria are presented in **Figure 1**
- The MA was performed using Stata 18.5 SE, with a DerSimonian-Laird random-effects model applied in the base case. Sensitivity analyses employed an inverse-variance fixed-effect model for hazard ratio (HR) estimates and the Mantel-Haenszel method for n/N data to calculate odds ratio (OR) estimates
- Statistical heterogeneity was assessed using I-squared (I^2) statistic

Figure 1. Eligibility Criteria of Systematic Literature Review

Results

- Overall, 49 studies were included, reporting an association of ALP response, ALP normalization, total bilirubin, and composite response (defined by key biomarkers) with long-term liver outcomes. The PRISMA flow for the SLR is provided in **Figure 2**
- Of these, 29 studies were peer-reviewed journal articles, while 20 were conference abstracts. Among the included studies, 33 were retrospective observational, 15 studies were prospective observational, while the remaining one study was both prospective and retrospective in design (**Figure 3**)
- In most of the studies, patients were treated with ursodeoxycholic acid ± off-label fibrates
- Overall, 28 studies reported an association between composite response, defined by key biomarkers and long-term liver outcomes. Commonly applied composite response criteria included PARIS I (ALP ≤3 × ULN, AST ≤2 × ULN and normal bilirubin; n=22), PARIS II (ALP and AST ≤1.5 × ULN and normal bilirubin; n=15), and Rotterdam (bilirubin ≤ 1 × ULN and/or albumin ≥ LLN; n=13) with some studies using Rochester (n=3) and POISE (n=2) criteria
- Further, 38 studies reported an association between ALP response, ALP normalization and clinical outcomes. The most commonly used ALP response criteria were Barcelona (ALP reduction ≥40% or ALP ≤1 × ULN; n=18) and Toronto I (ALP ≤1.67 × ULN; n=18), followed by ALP normalization (ALP ≤1 × ULN; n=7) and Toronto II (ALP ≤1.76 × ULN; n=3)
- Statistical heterogeneity across studies was not high, as indicated by low I^2 values, suggesting consistency of the observed associations

Figure 2. PRISMA Flowchart of Studies Identified Through the SLR

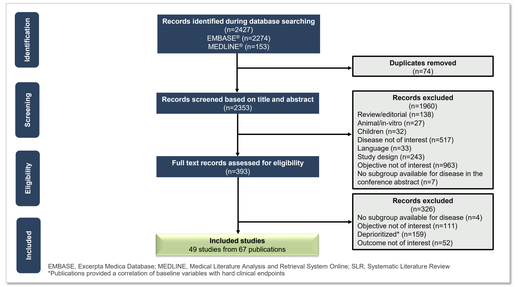


Figure 3. Study Characteristics of Included Studies

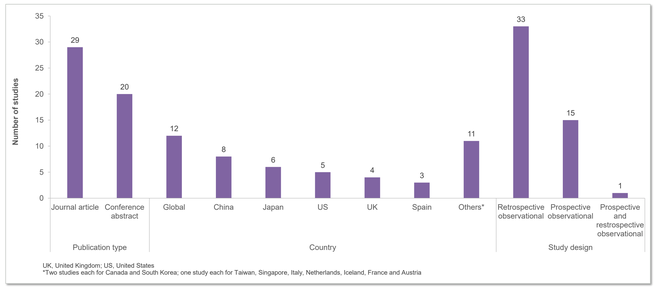


Figure 4. Base-Case Analysis for Association of ALP Response and ALP Normalization with Risk of Long-Term Liver Outcomes* (A) HR estimates (B) OR estimates

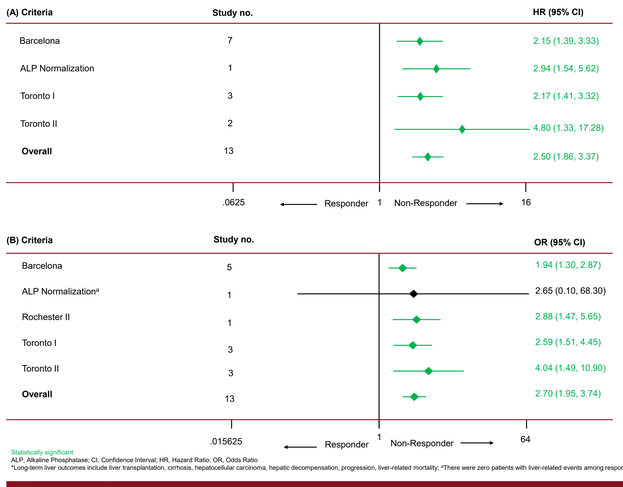
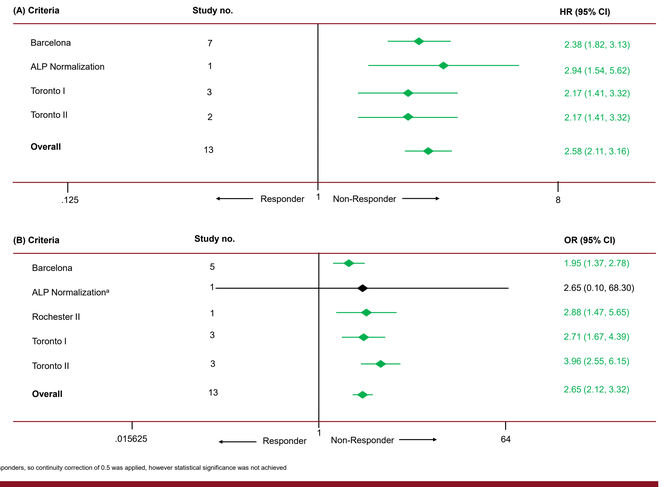


Figure 5. Sensitivity Analysis for Association of ALP Response and ALP Normalization with Risk of Long-Term Liver Outcomes* (A) HR estimates (B) OR estimates



Results (Contd.)

ALP normalization and risk of long-term adverse liver outcomes

- Patients who failed to achieve ALP normalization had a significantly higher risk of liver-related mortality, with effect estimate confirming this association (HR: 2.94, 95% CI: 1.54-5.62) (**Figure 4A**)
- Similar findings were reported for categorical data; however statistical significance was not achieved, possibly due to limited number of studies and absence of liver-related events among responders (HR: 2.65, 95% CI: 0.10-68.30) (**Figure 4B**)
- The results for sensitivity analyses were aligned with the base-case results (**Figure 5**)

ALP response and risk of long-term adverse liver outcomes

- Across studies assessing Barcelona criteria, non-responders had a significantly higher risk of adverse liver outcomes compared to responders, with the pooled effect estimate also reaching significance. Similar findings were reported for Toronto I/II criteria (**Figure 4A**)
- Overall, patients who did not meet the ALP response and ALP normalization criteria showed a significantly higher risk of liver outcomes compared to responders (HR: 2.50, 95% CI: 1.86-3.37) (**Figure 4A; Figure 4B**)

Composite response and risk of long-term adverse liver outcomes

- Across studies assessing PARIS I criteria (n=10), non-responders had a significantly higher risk of adverse liver outcomes compared to responders, with the pooled effect estimate also reaching significance (HR: 4.47, 95% CI: 3.29-6.07). Similar findings were reported for Paris II (n=6; HR: 5.05, 95% CI: 3.15-8.10) and Rotterdam criteria (n=8; HR: 6.16, 95% CI: 4.01-9.49)
- Two studies assessing composite response by POISE criteria (ALP <1.67×ULN, ≥15% ALP reduction, bilirubin ≤1×ULN) reported that non-responders had significantly higher risks of liver-related events than responders (HR: 4.45, 95% CI: 2.16-9.14)
- Consistent across all studies, patients who did not achieve composite response criteria were at a significantly higher risk of long-term liver outcomes, including LT, cirrhosis, HCC, hepatic decompensation, progression, mortality compared with responders (HR: 4.82, 95% CI: 3.85-6.05)
- Further, all studies with categorical data found significant associations between liver outcomes and PARIS I (OR: 6.80, 95% CI: 4.53-10.21), PARIS II (OR: 6.79, 95% CI: 3.62-12.74), Rotterdam (OR: 6.71, 95% CI: 2.65-17.00), and POISE criteria (OR: 10.42, 95% CI: 3.70-29.35), with pooled estimates confirming the association (OR: 6.94, 95% CI: 5.06-9.50)

Total bilirubin and risk of long-term adverse liver outcomes

- Overall, nine studies reported an association between total bilirubin and long-term liver outcomes
- Among the eight studies reporting quantitative data, seven studies reported a significantly higher risk of adverse long-term liver outcomes in non-responders compared with responders (**Table 1**)

Table 1. Effect Estimates Across Studies Reporting Association of Total Bilirubin With Long-term Liver Outcomes

Study name	Total Bilirubin cut-off	Effect estimates, HR/OR (95% CI) Non-Responders vs Responders	*P*-value
Harms 2019	1.0 x ULN	5.06 (4.34 , 5.89)	<0.001
Lammert 2014	>1.0 x ULN	4.90 (2.40 , 9.70)	<0.001
Yang 2023	≤1.0 x ULN	3.42 (1.95, 6.03)	<0.001
Kumagi 2010	13.5 μ mol/l	5.56* (1.25 , 25.0)	0.024
Pinyopornpanish 2022	ULN ratio	2.16 (1.85, 2.52)	<0.001
Papastergiou 2013	> ULN	5.64 (1.40, 22.66)	0.01
Corpechot 2022	≤ 0.6 x ULN	1.67 (0.92, 3.02)	NR
Murillo 2020	Bilirubin increase at 1 year	2.57 (1.53, 4.32)	<0.001

CI, Confidence Interval; HR, Hazard Ratio; NR, Not Reported; OR, Odds Ratio; ULN, Upper Limit of Normal
*Odds Ratio reported

Limitations

- The majority of included studies involved patients treated with UDCA alone or in combination with fibrates. The findings may not be generalizable to patients receiving approved second-line therapies, thereby underscoring the need for additional analyses
- The quality of data and reporting standards across the included studies can vary, potentially impacting the reliability of the meta-analysis results