Investigating Composite Response Definitions Based on Key Surrogate Markers for Predicting Clinical Outcomes in Primary Biliary Cholangitis: A Systematic Review and Meta-Analysis

CO156

Dilip Makhija, MS[1], Marvin Rock, MPH, DrPH[1], Chong H Kim, MPH, MS, PhD[1], Mirko von Hein, M.Sc.[2], Ryan Thaliffdeen, BS, PharmD, MS[1], Oskar Eklund, MSc[3], Pankaj Rai, MS[4], Ritesh Dubey, PharmD[4], Howard Thom, MSc, PhD[5], Gianluca Baio, PhD[6], Barinder Singh, RPh[4]

[1]Gilead Sciences, Inc., Foster City, CA, USA, [2]Gilead Sciences, London, United Kingdom, [3]Gilead Sciences AB, Solna, Sweden, [4]Pharmacoevidence, Mohali, India, [5]University of Bristol, Bristol, United Kingdom, [6]University College London, London, United Kingdom

CONCLUSIONS

- Key disease markers, including ALP and total bilirubin, show a strong association with long-term clinical outcomes in PBC when incorporated into composite response definitions
- The findings align with our previous evidence that links ALP decline to improved clinical outcomes, such as LT and death
- These findings support the use of surrogate markers in guiding treatment decisions, evaluating therapeutic efficacy, and reinforcing the value of biomarker-driven strategies in PBC management and reducing patient burden

PLAIN LANGUAGE SUMMARY

- Researchers defined treatment response using different 'composite response criteria,' which relied on blood tests like ALP and bilirubin. Paris I, Paris II, and Rotterdam were the most common, with a few studies using Rochester or POISE
- Patients who do not achieve a treatment response based on composite response criteria were more likely at a substantially higher risk of liver transplantation or death
- Regardless of the composite response definition used, the consistent associations observed across diverse studies validate their utility as reliable indicators for adverse outcomes, including liver transplantation or death
- The most recent studies using the POISE criteria confirmed that non-responders had significantly worse clinical outcomes

INTRODUCTION

- Primary biliary cholangitis (PBC) is a rare, chronic autoimmune liver disease with slow progression, making it challenging to assess long-term clinical outcomes like liver transplantation (LT) and death[1]
- Surrogate markers, such as ALP and bilirubin, whether alone or in composite response criteria, are strong indicators of disease progression and treatment response. Lower ALP and bilirubin levels correlate strongly with improved transplant-free survival and reduced risk of adverse outcomes[1,2]
- Utilizing these surrogate endpoints allows for more timely and practical assessments of treatment efficacy, helping to guide treatment decisions while long-term outcomes remain difficult to predict accurately
- Our previous systematic literature review (SLR) and meta-analysis (MA) demonstrated that lower alkaline phosphatase (ALP) levels and ALP normalization were linked to reduced risk of adverse outcomes, such as liver transplantation (LT) or death, among patients with PBC[3]

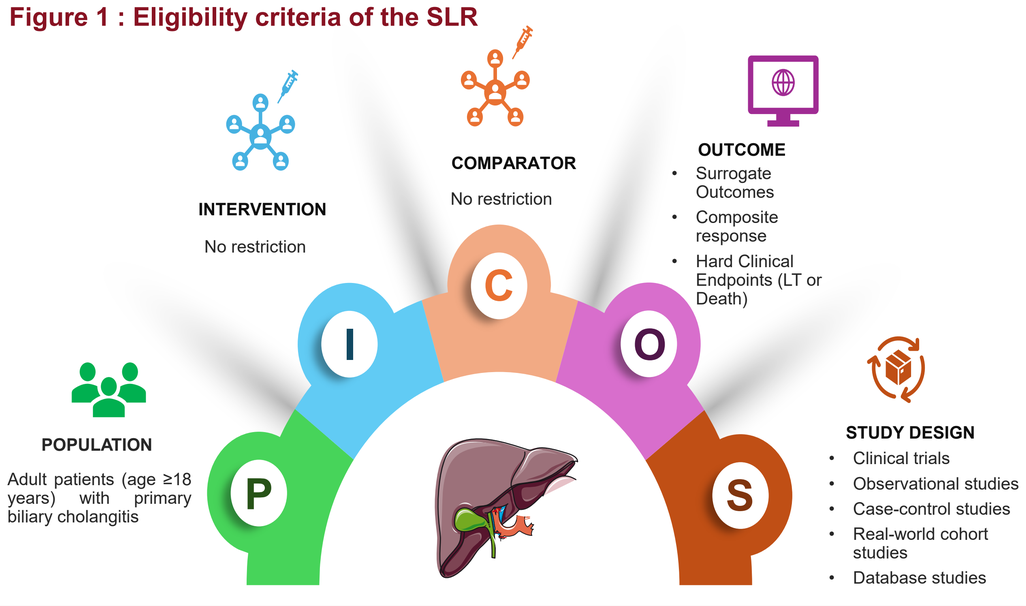
OBJECTIVE

- To better understand the relationship between surrogate endpoints and clinical outcomes, this study systematically reviewed the evidence to assess whether composite response definitions based on surrogate markers (e.g., ALP, bilirubin) can reliably predict long-term outcomes in PBC

METHODS

- This study adhered to National Institute for Health and Care Excellence (NICE) and Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines for SLRs, following standard methodology with a transparent, reproducible, and unbiased approach
- EMBASE® and PubMed® were searched for English-language articles from database inception to September 2024, to identify publications evaluating the association between composite response-based surrogate endpoints and long-term clinical outcomes in PBC. The prespecified eligibility criteria are presented in **Figure 1**
- Two independent reviewers performed the data collection and data extraction activities, with conflicts resolved by a third independent reviewer
- The MA was performed using Stata 17, employing the DerSimonian-Laird (DL) + Hartung-Knapp-Sidik-Jonkman (HKSJ)[4] method for hazard ratio (HR) estimates, which combines the DL random-effects model with the HKSJ adjustment to improve confidence interval (CI) accuracy; the Mantel-Haenszel (MH) method was employed for categorical data (i.e., n/N data)
- Statistical heterogeneity was assessed using the parameters such as Cochrane's Q, I-squared (I^2), and Tau-squared (τ^2)

Figure 1 : Eligibility criteria of the SLR



RESULTS

- Overall, 28 studies were included, reporting an association between a composite response defined by key biomarkers and clinical outcomes. The PRISMA flow for the SLR and MA is provided in **Figure 2**
- Of these, 19 studies were peer-reviewed journal articles, while 9 were conference abstracts. Among the included studies, 19 were retrospective observational, while nine were prospective observational
- Six studies were conducted in China, followed by five studies each conducted in Japan and globally. Additionally, three were conducted in the UK, two each in the Netherlands and South Korea, and one each in Spain, France, Italy, Iceland, and the USA (**Figure 3**)
- Across these studies, death or LT were the most frequently reported hard clinical endpoints (n=15), with most patients treated with UDCA alone or with fibrates
- 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
- All six studies evaluating the association between Paris I criteria and LT or death reported statistically significant results, with the pooled effect estimate also reaching significance (HR: 5.60, 95% CI: 3.25-9.66) (**Figure 4**)

Figure 2: Flow of studies in the SLR

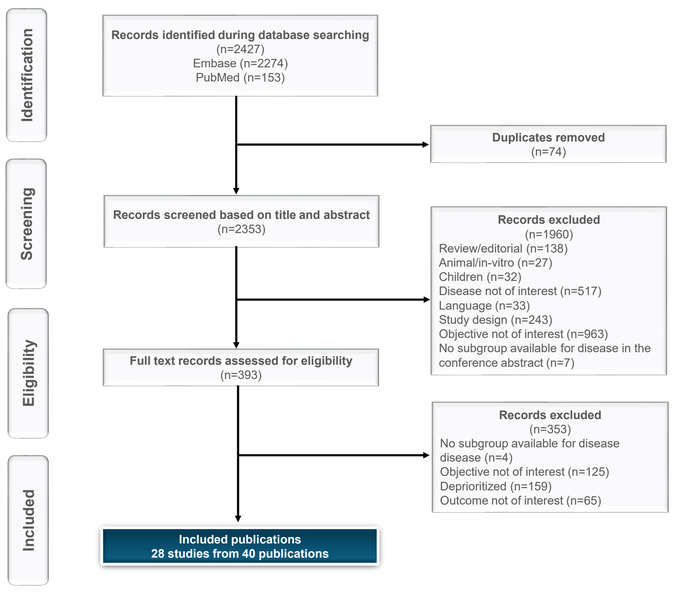


Figure 3: Study characteristics of included studies

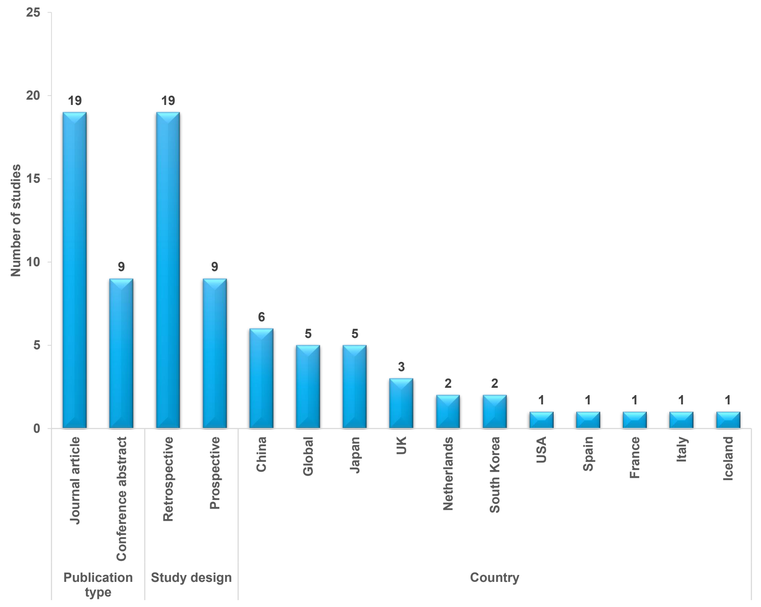


Figure 4: Association of composite response (Paris I, Paris II, Rotterdam, and Rochester) with LT or death (HR estimates)

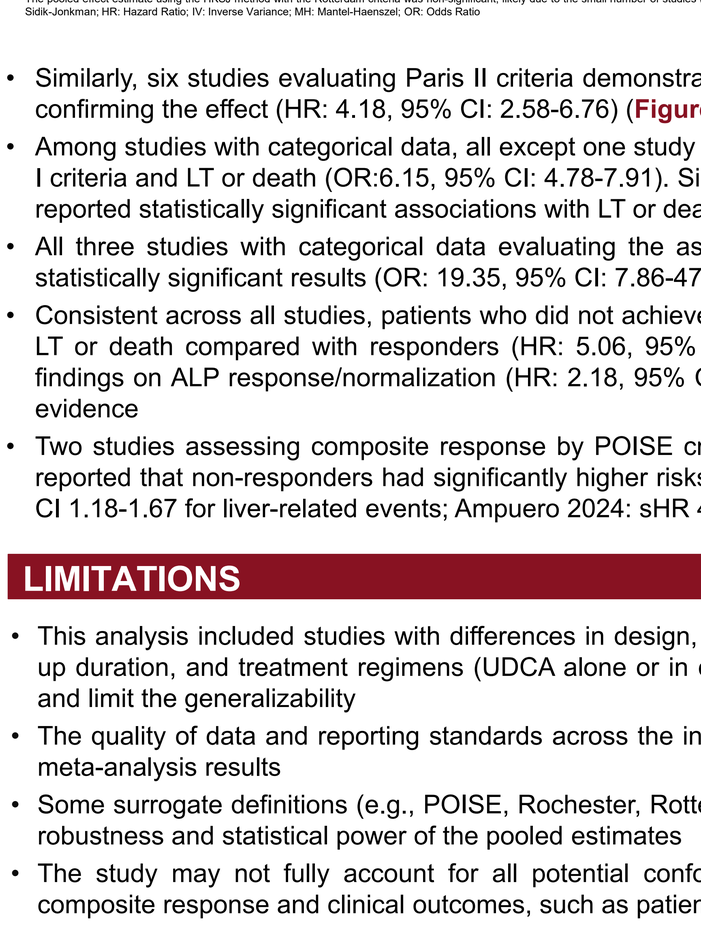
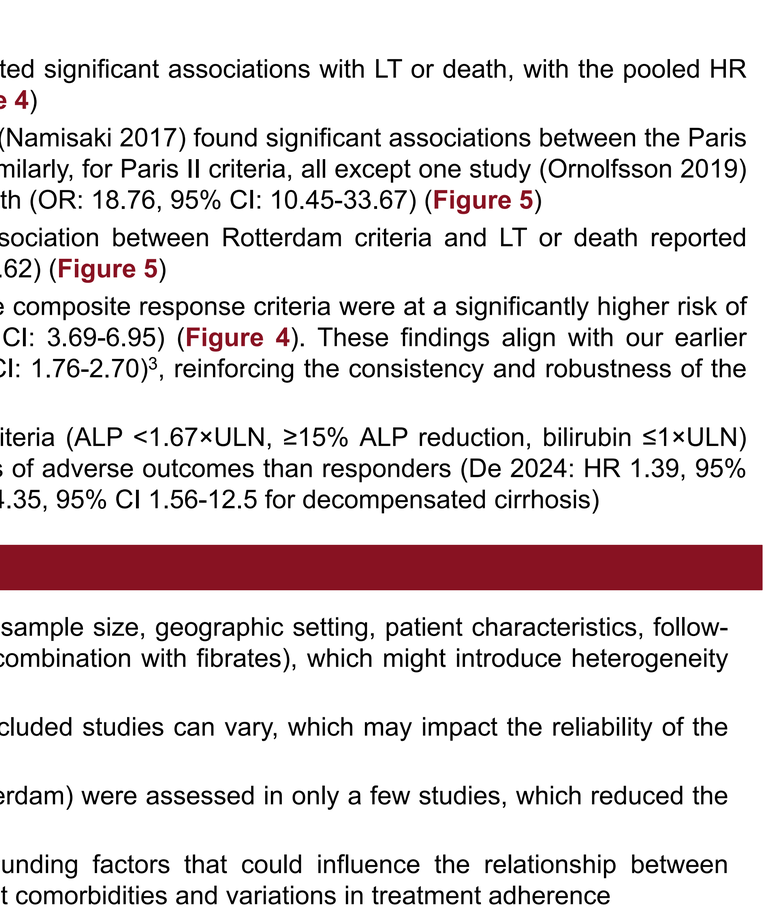


Figure 5: Association of composite response (Paris I, Paris II, and Rotterdam) with LT or death (OR estimates derived from n/N data)



The pooled effect estimate using the HKSJ method with the Rotterdam criteria was non-significant, likely due to the small number of studies (n=2). In contrast, alternative methods produced significant results (DL: 1.62 [1.33-1.91]; IV: 1.65 [1.48–1.81]); DL: DerSimonian-Laird; HKSJ: Hartung-Knapp-Sidik-Jonkman; HR: Hazard Ratio; IV: Inverse Variance; MH: Mantel-Haenszel; OR: Odds Ratio

- Similarly, six studies evaluating Paris II criteria demonstrated significant associations with LT or death, with the pooled HR confirming the effect (HR: 4.18, 95% CI: 2.58-6.76) (**Figure 4**)
- Among studies with categorical data, all except one study (Namisaki 2017) found significant associations between the Paris I criteria and LT or death (OR:6.15, 95% CI: 4.78-7.91). Similarly, for Paris II criteria, all except one study (Ornolfsson 2019) reported statistically significant associations with LT or death (OR: 18.76, 95% CI: 10.45-33.67) (**Figure 5**)
- All three studies with categorical data evaluating the association between Rotterdam criteria and LT or death reported statistically significant results (OR: 19.35, 95% CI: 7.86-47.62) (**Figure 5**)
- Consistent across all studies, patients who did not achieve composite response criteria were at a significantly higher risk of LT or death compared with responders (HR: 5.06, 95% CI: 3.69-6.95) (**Figure 4**). These findings align with our earlier findings on ALP response/normalization (HR: 2.18, 95% CI: 1.76-2.70)[3], reinforcing the consistency and robustness of the evidence
- 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 adverse outcomes than responders (De 2024: HR 1.39, 95% CI 1.18-1.67 for liver-related events; Ampuero 2024: sHR 4.35, 95% CI 1.56-12.5 for decompensated cirrhosis)

LIMITATIONS

- This analysis included studies with differences in design, sample size, geographic setting, patient characteristics, follow-up duration, and treatment regimens (UDCA alone or in combination with fibrates), which might introduce heterogeneity and limit the generalizability
- The quality of data and reporting standards across the included studies can vary, which may impact the reliability of the meta-analysis results
- Some surrogate definitions (e.g., POISE, Rochester, Rotterdam) were assessed in only a few studies, which reduced the robustness and statistical power of the pooled estimates
- The study may not fully account for all potential confounding factors that could influence the relationship between composite response and clinical outcomes, such as patient comorbidities and variations in treatment adherence

REFERENCES

1. Kaplan MM, et al.. N Engl J Med. 2005 Sep 22;353(12):1261-73
2. Lammers WJ, et al. Gastroenterology. 2014 Dec;147(6):1338-49.
3. Makhija et al. Value Health. 2025;28(S1):CO170.
4. IntHout J, et al. BMC Med Res Methodol. 2014 Feb 18;14:25

ACKNOWLEDGEMENTS

We extend our thanks to the patients, their families, and all participating investigators. This study was funded by Gilead Sciences, Inc., while Pharmacoevidence provided medical writing and editorial support.

DISCLOSURES

DM, MR, CK, MH, RT, and OE are employees of Gilead Sciences, Inc. PR, RD, and BS are employees of Pharmacoevidence. HT is associated with the University of Bristol, and GB is associated with University College London