# Risk Stratification in Autoimmune Cholestatic Liver Diseases: Opportunities for Clinicians and Trialists

Palak J. Trivedi, [1] Christophe Corpechot, [2] Albert Pares [3] and Gideon M. Hirschfield [1*]

1. National Institute for Health Research (NIHR) Birmingham Liver Biomedical Research Unit (BRU) and Centre for Liver Research, University of Birmingham, Birmingham, UK.
2. National Reference Center for Inflammatory Diseases of the Biliary Tract (MIVB), Rare Liver Diseases Health Network (FILFOIE), Saint-Antoine Hospital, Assistance Publique – Hôpitaux de Paris (APHP), Paris, France
3. Liver Unit, Hospital Clínic, CIBERehd, IDIBAPS, University of Barcelona, Barcelona, Spain

Palak J. Trivedi: p.j.trivedi@bham.ac.uk

Christophe Corpechot: christophe.corpechot@aphp.fr

Albert Pares pares@ub.edu

Gideon M. Hirschfield g.hirschfield@bham.ac.uk

* Correspondence:

Dr. Gideon M Hirschfield, Centre for Liver Research, Institute of Biomedical Research, University of Birmingham, B15 2TT, UK. Email: g.hirschfield@bham.ac.uk

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/hep.28128

**Acknowledgements**

GMH is a co-investigator for UK-PBC (www.uk-pbc.com) supported by a Stratified Medicine Award from the UK Medical Research Council, and Principal Investigator for UK-PSC, a NIHR Rare Disease Translational Collaboration.

The views expressed are those of the authors(s) and not necessarily those of the National Health System, the National Institute for Health Research, or the Department of Health.

**List of Abbreviations**

| | |
|---|---|
| **ALP** | Alkaline phosphatase |
| **AMA** | Anti-mitochondrial antibody |
| **ANA** | Anti-nuclear antibody |
| **AST** | Aspartate aminotransferase |
| **APRI** | AST/platelet ratio |
| **AUROC** | Area under the receiver operator curve |
| **CCA** | Cholangiocarcinoma |
| **DS** | Dominant strictures |
| **ELF®** | Enhanced liver fibrosis score |
| **ERC** | Endoscopic retrograde cholangiography |
| **GRADE** | Grading of Recommendations Assessment, Development and Evaluation criteria |
| **HPB** | Hepatopancreatobiliary |
| **HR** | Hazard ratio |
| **IBD** | Inflammatory bowel disease |
| **IgG4** | Immunoglobulin G subclass 4 |
| **LSM** | Liver stiffness measurement |
| **MRC** | Magnetic resonance cholangiography |
| **MRI** | Magnetic resonance imaging |
| **PBC** | Primary biliary cirrhosis |
| **PHT** | Portal hypertension |
| **PSC** | Primary sclerosing cholangitis |
| **sdPSC** | Small duct primary sclerosing cholangitis |
| **SMR** | Standardized mortality ratio |
| **UDCA** | Ursodeoxycholic acid |
| **ULN** | Upper limit of normal |
| **VCTE** | Vibration controlled transient elastography |

**Abstract**

Primary biliary cirrhosis (PBC) and primary sclerosing cholangitis (PSC) are infrequent autoimmune cholestatic liver diseases, that disproportionate to their incidence and prevalence, remain very important causes of morbidity and mortality for patients with liver disease. Mechanistic insights spanning genetic risks and biologic pathways to liver injury and fibrosis have lead to a renewed interest in developing therapies beyond ursodeoxycholic acid, that are aimed at both slowing disease course and improving quality of life. International cohort studies have facilitated a much greater understanding of disease heterogeneity and in so doing highlight the opportunity to provide patients with a more individualized assessment of their risk of progressive liver disease, based on clinical, laboratory or imaging findings. This has lead to a new approach to patient care that focuses on risk stratification (both high and low risk) and furthermore allows such stratification tools to help identify patient subgroups at greatest potential benefit from inclusion in clinical trials. We review the applicability and validity of risk stratification in autoimmune cholestatic liver disease, highlighting strengths and weaknesses of current and emergent approaches.

## Introduction

Primary biliary cirrhosis (PBC) and primary sclerosing cholangitis (PSC) are chronic autoimmune cholestatic liver diseases, for which clinical outcome is largely dictated by development of cirrhosis, portal hypertension (PHT) and variable predisposition to malignancy (1–4). Rates of clinical progression vary; and accurately identifying disease course is of importance to patients, clinicians, as well as those committed to developing new effective and affordable treatments (5). Patients seek reassurance and guidance as to their own prognosis, and clinicians wish to confidently recognize those at highest risk of poor outcomes as equally as they strive to reassure individuals with good prognosis. Partnerships with industry are essential to drug development; and collectively all those involved in clinical trial design, recruitment and analysis wish to understand unmet need and conduct studies of new therapies as carefully constructed interventions that deliver **S**pecific, **M**easurable, **A**chievable, **R**elevant and **T**ime-cost limited outputs. Such ventures seek to 'de-risk' drug development pathways where possible, but maximize opportunity to advance therapy for patient benefit in a timely way.

Herein, we present an appraisal of existing parameters that stratify individuals with PBC and PSC, prior to examining effectiveness and applicability of more incipient classification systems (**Figure 1**). The strengths and weaknesses of various approaches are highlighted specifically throughout, as well more generally with regard to study design (**Table 1)**.

**Clinical history and phenotypes**

The full appreciation of the breadth of PBC as a disease has evolved as awareness has risen, particularly given widespread access to anti-mitochondrial antibody (AMA) testing; reactivity of which in the presence of cholestasis, facilitates robust and timely patient identification without need for histological confirmation (6,7). PBC is increasingly identified at an earlier pre-cirrhotic stage (8), and well-conducted multi-center cohort studies have aided in the recognition of variant presentations (**Table 2**); including male patients and women aged <50 years (9). Ursodeoxycholic acid (UDCA) is the only approved therapy, with diminished disease progression in treated patients and significantly improved 10-year transplant-free survival (78% vs. 66%; $p<0.001$) (3,8,10–12). Pooled survival indices nevertheless remain lower than age- and sex-matched control populations (10–12).

Modeling the clinical course of PSC, in contrast to PBC, is more testing; perhaps inevitably so given a lower incidence and absence of a defined serologic marker. This is paralleled by a clinical phenotype driven by variable, unpredictable consequences related to chronic inflammation, fibrosis and neoplasia of medium–large sized bile ducts. In the largest population-based study to date ($n$=590), disease was validated as being male predominant (~60%) with a median age at diagnosis of ~40 years (13). However, PSC can develop at any age, with younger patients frequently manifesting a hepatitic presentation (14). Associations with inflammatory bowel disease (IBD) are well-recognized and ~70% of PSC patients have a history of colitis, which confers a fivefold greater risk of colonic cancer relative to IBD alone, as well as increased susceptibility to cholangiocarcinoma (CCA) independent of liver disease stage. PSC portends a standardized mortality ratio (SMR) more than fourfold that of a matched control population, although there is discrepancy between event-free survival

times across transplant centers versus true population-based cohorts (median 13.2 vs. 21.3 years; $p<0.001$ (13)). Population-level data thus highlight significant challenges to prognostic modeling, and unmask the inadequate phenotypic representation of early-stage disease and inherent selection bias with tertiary center-restricted reporting.

***Symptom complex***

Pruritus and fatigue are frequent symptoms associated with cholestasis (15), and approximately 60% of patients with PBC are asymptomatic at diagnosis with as few as 5% remaining symptom-free over time (16). The prognostic importance of fatigue in PBC is contentious, but concern is best highlighted in the prospective cohort study from Jones et al. ($n$=136) (12), wherein transplant-free survival was significantly shorter among fatigued patients relative to non-fatigued, disease-matched controls (56% vs. 74%; $p<0.0001$), independent of UDCA provision. Although a consensus biological explanation for fatigue is lacking, presenting age and sex heavily influence the clinical phenotype, with young women (a group failing UDCA therapy more commonly), having the greatest symptom burden (9,17). There is no evidence however, that symptomatic presentations impart additional discriminatory value to existing risk-prediction models.

Symptomatic presentations in PSC similarly vary (36–56%) with over 20% developing symptoms *de novo* during follow-up (18–20). Relapsing-remitting episodes of acute cholangitis are a frequent concern; and data from several cohorts suggest symptomatic presentations carry poorer transplant-free and malignancy-free survival (18,20). One-third of CCA are diagnosed within the first year of PSC presentation (annual incidence thereafter 0.5–1.5%; lifetime risk 7–15%) (13,18), and patients often report abdominal pain prior to diagnosis, particularly those with a prolonged history of IBD (>1 year) (18,21).

**Biochemical response criteria in PBC**

Serum bilirubin is well-established as a predictor of outcome and incorporated into several prognostic scoring systems (22,23). However, 'time-constrained' models such as the Mayo-score, which include bilirubin together with other markers of cirrhosis, are limited to prediction of short-term survival (<2 years) in relatively late-stage disease. A potentially more applicable surrogate is serum alkaline phosphatase (ALP); and in the largest ever meta-analysis of individual patient data ($n$=4845) a near log-linear relationship was illustrated between ALP and subsequent risk of transplantation/death across several time points (8). This study demonstrated that ALP bestows prognostic information early in the clinical course, incremental to the predictive power of bilirubin and independent of follow-up time, presenting age, sex, disease stage, and treatment status.

To this effect, several studies illustrate strong associations between percentage reduction or absolute decreases/normalization in serum ALP (in isolation or combination with other biochemical covariates) and significantly improved clinical outcome (10,11,24,25). Indeed, the majority who successfully attain pre-defined biochemical thresholds 1–2 years after UDCA-treatment (13–15 mg/kg/day) experience survival patterns akin to that of an age- and sex-matched control population (**Table 3A**). All response criteria have been independently and externally validated, with Paris-I capturing the greatest breadth of biochemical changes. Furthermore, there is clear, negative prognostic impact of biochemical non-response on future hepatocellular carcinoma risk in PBC patients, independently and additive to the effects posed by male sex and advanced baseline disease stage (2).

Although a small proportion with early-stage disease meet criteria free of therapy (26), this represents an understudied population and presently it is not possible to identify patients with a good prognosis regardless of intervention. Inversely, paradigms reliant on waiting one-year for therapeutic evaluation may leave high-risk patients (future non-responders) on therapy lacking benefit and reduce impact of second line therapy because of delayed initiation. To this effect, a prospective study from China suggested that attainment rates as well as predictive value is identical when biochemical response is assessed at 6 vs. 12 months (**Table 3B**) (27), but this needs validation.

***Demographic variations***

Population-level and international multi-center studies have substantiated the predictive performance of biochemical response criteria, independently of disease stage and UDCA exposure (9,28). Perhaps most notable, is the UK-PBC study ($n$=2353), which not only recognized an increasing prevalence of younger presenting women (25% aged <50) but also an inverse correlation of patient age and likelihood of meeting biochemical response (9). Attainment rates were reportedly ≤50% in women aged below 40, and echo results of an earlier, single-center study wherein age <55 years conferred poorer relative survival relative to matched controls (SMR: 7.4) (29). Younger women often present with more pronounced elevations in serum ALP (17), but frequently fail therapy due to transaminase elevations (9); possibly reflecting a more hepatitic phenotype – noteworthy given that the degree of interface activity is recognized to influence disease progression (10,14,30,31). The influence of presenting age was less apparent in men (9), who despite being older at diagnosis exhibited greater frequency of non-response overall; possibly reflecting more advanced baseline fibrosis at presentation (32).

The strong influence of presenting age may allow more timely stratification of at-risk groups (prior to assessment of 12-month biochemical response), who, because of a relatively poor predicted survival would be potentially eligible for early clinical trial entry. Application of response criteria at earlier time-points would conceivably allow for more opportune recognition of at-risk individuals and must also ensure low-risk patients are not over treated (27).

***Optimization of criteria***

Existing biochemical response criteria remain to be refined, with a subgroup of responders still at risk of developing adverse events. There is evidence that reduction in hepatic veno-portal gradient whilst on UDCA-treatment associates with improved transplant-free survival in PBC, stratifying through a 20% gradient-decline over 2 years (3). Conversely, the presence of gastroesophageal varices (GEV) is a poor prognostic factor (4); and as portal hypertension (PHT) can develop in the absence of cirrhosis secondary to pre-sinusoidal resistance, several algorithms for prediction of GEV are proposed. Although advocated for guiding variceal surveillance, such models carry pre-selection bias, as study populations from which they derive were included following endoscopy referral. Moreover, no current strategy allows non-invasive discrimination of clinically significant PHT.

With regards to patient survival, performance characteristics of the AST/platelet ratio index (APRI) have been ascertained given ability to infer not only PHT but also fibrosis (3,28). When applied at baseline or at 1 year, APRI was identified as an independent predictor of transplant-free survival across a tertiary center population ($n$=386), with a discriminatory cut-point of 0.54 externally validated in three international cohorts (28,33). Moreover, 1-year APRI identified the sub-group at risk of disease progression and earlier mortality

despite successful attainment of biochemical response (**Table 3C**); indicating independent and additive prognostic information to existing criteria (28,34,35).

Newer, highly complex and robust computational algorithms incorporating facets of APRI in addition to conventional biochemical response parameters are soon to be published. These scoring systems derive from large multi-center cohorts as part of UK-PBC as well as the Global PBC Study Group (34,35) and convey probability of transplant-free survival on a continuous as oppose to dichotomous scale (AUROC >0.9). In addition to being internally validated, the latter in particular has been compared against a healthy age/sex matched control population, demonstrating comparable prognostic performance to Paris-I + APRI (35). However it is uncertain above what point patients will be deemed high-risk enough for clinical trial stratification, how the modifier effects of UDCA on risk-score will influence outcome (delta-change), and which additional stratifiers will continue to retain independent clinical impact.

**Can biochemical surrogates be extrapolated to PSC?**

Serum bilirubin is inherent to many historic PSC prognostic models including the disease-specific Mayo-score (36). Despite widespread application, the series from which the latter derives antedates modern management of variceal bleeding and receives further criticism given inability to foreshadow adverse events in prior clinical trials (37). Although a persistently elevated bilirubin for >3 months incites concern for hepatobiliary malignancy (18), levels have a propensity to fluctuate with flares of cholangitis and potentially influenced by biliary interventions.

There is no proven survival advantage or reduction in hepatobiliary/colorectal malignancy risk for PSC patients receiving UDCA, and an increased predisposition toward adverse events well documented with high dosages (28–30mg/kg/day) (1,5). Several groups have nevertheless attempted construction of 'ALP-based' biochemical response criteria (**Table 4**) (38–43), but ultimately each has failed cross-validation at the originally conceived time-points. For instance, the 1.5xULN cut-point proved discriminatory at 2 years in the Oxford cohort (irrespective of UDCA receipt (40)) but was only predictive when applied at 6 and 12 months in the Heidelberg and national UK series, respectively. Moreover, in only one published study has the predictive value of ALP as a continuous variable been confirmed prior to establishing utility through dichotomization (43); however, full statistical methodology was not presented and clinical endpoints incorrectly assessed as time-constrained events.

Systematic efforts to validate the prognostic utility of serum ALP in PSC remain in their infancy and none of the studies thus far incorporate a comparator group. It is therefore difficult to infer what an improved serum ALP truly means, as 'PSC biochemical responders' may still benefit from trials of new therapy if survival significantly deviates from the healthy population. Spontaneous normalization has been reported in up to ~40% of patients (38); and whilst this may indicate a slowly progressive form of disease, based on available evidence ALP cannot be recommended as a 'standalone' stratifier of risk in PSC.

**Immunoserological indices and coexisting autoimmunity**

***PBC-specific anti-nuclear antibodies (ANA)***

Unlike AMA, which holds no prognostic value (8,9,28), there exist several ANA subtypes that may associate with adverse clinical outcome in PBC. Baseline anti-gp210 reactivity imparts over a six-fold risk of progression to liver failure/transplantation (44); and although neither independent nor additive to biochemical response (12,30) may assist in the earlier, prospective identification of high-risk patients (27,44). Anti-centromere antibodies similarly associate with a PHT (44), although more often present in autoimmune connective tissue disease. Extra-hepatic autoimmunity develops in ~60% of PBC patients however impact on liver-related outcomes is not readily apparent (45).

***Serum IgG4 in PSC***

Between 9–15% of PSC patients have raised serum immunoglobulin subclass 4 values (IgG4) (46–49), and at least three separate studies support clinical distinctions based on elevations; those having higher than normal values (>1.4g/L) exhibiting greater derangements in liver biochemistry (46–48). One group identified shorter median time to transplantation in patients harboring elevated serum IgG4 (48), although this observation has failed replication in several international centers (49). The stratifying properties of serum IgG4 therefore remain unsubstantiated and require further evaluation.

***Impact of colitis in PSC***

Several historic studies suggest that the presence of colitis influences liver disease progression. However, many were flawed given their assessment of IBD as a time-fixed covariate; and the chronological displacement of disease presence and activity between gut and liver manifestations impart significant difficulties in examining colitis as a risk-

stratifier. Nevertheless in a prospective follow-up of nearly 200 PSC patients, all hepatobiliary malignancies were seen to develop on a background of concurrent colitis, with no cancers in the absence of IBD (50). Moreover, transplant-free survival independent of CCA was also significantly different between groups (23% vs. 80%; $p$=0.045). The negative prognostic impact of colitis on liver-related outcomes has since been confirmed in a large Dutch PSC cohort ($n$=161) as well as two population-based series (13,51–53).

**Cholangiographic stratification in PSC**

Several cholangiographic prognostic models derived from endoscopic retrograde cholangiographic (ERC) appearances have been proposed (54); however, diagnostic paradigms have evolved and no correlation between severity of ductal involvement and survival through 2-dimensional magnetic-resonance cholangiography (MRC) demonstrated. Nevertheless, a promising study utilizing annual 3D MRC to score liver parenchymal appearances, PHT and bile duct lesions, predicted radiological progression from baseline with high accuracy (AUROC>0.8) (55). 60% of patients developed evolving changes over ~4 years and preliminary data indicate baseline radiological score a highly sensitive prognosticator of clinical outcome with the most predictive components relating to parenchymal, opposed to ductal changes (56).

***Dominant strictures***

Dominant strictures (DS) were originally defined based on historical ERC findings, and consensus opinion as to how such lesions are to be classified non-invasively is yet to be delivered. Observational studies report a presenting frequency of 12–60% (57,58), with no population-level indications of true incidence. Natural history data are similarly restricted, with reduced survival largely reflecting difficulties in CCA recognition (18,50,58,59).

However, more recent reports suggest actuarial transplant-free survival as significantly poorer irrespective of cancer development and heavily influenced by presence of colitis (50,60). Several investigators report biochemical and clinical improvements following endoscopic therapy (61), but the prognostic impact of intervention needs assessment.

***Small duct PSC***

Small duct PSC (sdPSC) represents 10–15% of the disease spectrum, with affected individuals less often symptomatic (62). There is now well-validated evidence that disease progression is relatively infrequent, occurring over a longer time period than the classical form (13,63,64). Although colitis manifests to a similar degree there is little to suggest an impact on liver-related outcomes; and as survival patterns mirror those of an age- and sex-matched population, the need for investigative therapy is perhaps less perceptible.

**Histological stage and non-invasive evaluation**

Disease identification in PBC and PSC is largely reliant on serology and cholangiography, respectively, in the appropriate clinical and biochemical context. Nevertheless, liver biopsy is invaluable in cases of diagnostic doubt, and provides key information with regards to disease activity and severity that may improve predictive power of existing algorithms (31).

Several contemporary histological systems have emerged for PBC (65,66), with the aim of accurately representing interface activity, ductopenia, chronic cholestasis and fibrotic indices – variables well known to forecast biochemical non-response and clinical outcome **(Table 5)** (9,10,14,25,30,31,67,68). Common histological changes in PSC include interface activity, ductopenia and concentric periductal fibrosis, although individual prognostic weightings are unclear, and no disease-specific classification exists. Nevertheless, data

extrapolated from the Dutch population-based registry (*n*=64) indicates that scoring through PBC-based classification systems, as well as lobular fibrosis stage (Ishak), significantly associates with time to transplantation in PSC patients (69).

Histology remains the 'gold-standard' for assessing fibrosis progression – a clear determinant of clinical outcome. However, the intrusiveness coupled with well-known sampling variability and disconcordant reporting in cholestatic disease has fostered development of several non-invasive surrogates (**Table 6**). In the current clinical climate, histological stratification holds limited routine applicability, although staging systems and evaluation of prognosis-related histological lesions may have a place as surrogate endpoints in clinical trials – a topic beyond the scope of this review.

***Vibration controlled transient elastography (VCTE)***

The accuracy of VCTE in fibrosis staging has been demonstrated in at least two large PBC cohorts (70,71); with prognostic capabilities independent of biochemical response evident in a recent single-center retrospective study of 150 patients (70). Whilst, VCTE outperforms APRI as well as several non-invasive surrogates of fibrosis, it remains unclear if the former confers additive discrimination to biochemical response. The prognostic impact of LSM in PSC has also recently been described (72); and as with prior descriptors correlated well with degree of liver fibrosis, performing best at extremes of histological stage ($\leq$F1 and $\geq$F3). More striking was the observation that increased baseline measurements and rate-of-change in LSM were strongly and independently linked with PSC-specific clinical events (72).

LSM in addition to reflecting severity of fibrosis can also be influenced by extrahepatic cholestasis, and may not necessarily capture disease facets such as hepatic

necroinflammatory activity, ductopenia and portal hypertension. Nevertheless, encouraging data from existing series strongly support VCTE-derived LSM – absolute values as well as fluctuations over time – as major predictors of adverse events. Given correlations with mortality and liver transplantation in PBC and PSC, VCTE may represent a generic surrogate in chronic cholestatic liver disease, and prospective validations as part of multi-center collaborative efforts continue.

***Enhanced liver fibrosis (ELF®) score***

The ELF® score bears similar prognostic utility to histological fibrosis staging in PBC (73); although akin to VCTE, additive predictive value to biochemical response has not been demonstrated. More recent focus on the stratifying properties in PSC led to a notable publication by the Norwegian Study Group. Therein, patients exhibited significantly divergent transplant-free survival curves according to tertile distribution, or through a dichotomous Youden-index-derived cut-point (74). Moreover, ELF® score correlated well with elastography and provided incremental prognostic utility to Mayo risk. One caveat however, is the relatively short disease duration experienced by transplant-free survivors (median 0.2 years), and of further uncertainty is how dynamic fluctuations impact outcome longitudinally. Nevertheless, this study represents the first non-invasive, externally validated serum biomarker panel in PSC.

**Clinical integration and prospective outlook**

Biochemical non-responders represent the most readily identifiable 'at-risk' group in PBC, and incorporating a stepwise algorithm with response criteria as the central feature is likely to capture the greatest breadth of individuals who will benefit from clinical trials (**Figure**

**2a**). Validation at interim time points for groups who commonly experience treatment failure is urgently decreed, and may assist in the earlier identification of high-risk patients. Along similar lines, prospective banking of biological materials with paired long-term clinical follow-up data could yield predictive markers from the point of diagnosis through interrogation of key pathways underlying non-response. The few PBC patients who endure adverse events despite attainment of response remain poorly-defined but increasingly recognized (28,35); and the additional impact of 'biochemical escape' – wherein prior responders develop subsequent elevations in laboratory parameters – yet to be explored. The additive predictive value of histology and its non-invasive surrogates to existing criteria also requires further validation in a manner similar to that presented for APRI; and formal publication of newer biochemical response criteria with dynamic predictive capabilities eagerly awaited (31,34,35).

By contrast, safe discrimination of risk-phenotypes in PSC is not possible through early application of a single modality, and timely assessment requires harnessing multiple predictive techniques collectively (**Figure 2b**). Despite invasiveness of histological stratification, the advent of VCTE and related biomarkers hold promise although predictive performance is best at stages of advanced fibrosis implying surrogacy toward disease stage rather than severity; and prospective validation currently remaining. Present biochemical surrogates are far from robust, and it is crucial for future endeavors to secure appropriate control groups prior to stratifying PSC patients as low-risk based on serum ALP alone; particularly as 20% of UDCA-treated patients with normal values still develop progressive disease (38). Further efforts are also needed to appraise the relative independence of existing parameters that stratify risk, both consequentially and concurrently.

## Conclusion

Patients with PBC and PSC remain a heterogeneous cohort with concerns as regard to forecasting outcome reliably. Stratification paradigms are shifting with increased efforts toward recognition of at-risk phenotypes. The increased utilization of such tools both clinically and in trial-settings is hoped to allow for more personalized care. In so doing low-risk patients can be reassured and managed accordingly, whilst higher-risk individuals offered tailored care, and access to carefully designed trials relevant to their disease course.

## References


1. Hirschfield GM, Karlsen TH, Lindor KD, Adams DH. Primary sclerosing cholangitis. Lancet. 2013;382:1587–1599.

2. **Trivedi PJ, Lammers WJ,** van Buuren HR, Parés A, Floreani A, Janssen HLA, et al. Stratification of hepatocellular carcinoma risk in primary biliary cirrhosis: a multicentre international study. Gut. 2015

3. Huet P-M, Vincent C, Deslaurier J, Coté J, Matsutami S, Boileau R, et al. Portal hypertension and primary biliary cirrhosis: effect of long-term ursodeoxycholic acid treatment. Gastroenterology. 2008;135:1552–1560.

4. Patanwala I, McMeekin P, Walters R, Mells G, Alexander G, Newton J, et al. A validated clinical tool for the prediction of varices in PBC: the Newcastle Varices in PBC Score. J. Hepatol. 2013;59:327–335.

5. Trivedi PJ, Hirschfield GM. Treatment of autoimmune liver disease: current and future therapeutic options. Ther. Adv. Chronic Dis. 2013;4:119–141.

6. EASL Clinical Practice Guidelines: management of cholestatic liver diseases. J. Hepatol. 2009;51:237–267.

7. Lindor KD, Gershwin ME, Poupon R, Kaplan M, Bergasa NV, Heathcote EJ. Primary biliary cirrhosis. Hepatology. 2009;50:291–308.

8. Lammers WJ, van Buuren HR, Hirschfield GM, Janssen HLA, 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;147:1338–1349.

9. **Carbone M, Mells GF, Pells G,** Dawwas MF, Newton JL, Heneghan MA, 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.

10. 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;48:871–877.

11. Parés A, Caballería L, Rodés J. Excellent long-term survival in patients with primary biliary cirrhosis and biochemical response to ursodeoxycholic Acid. Gastroenterology. 2006;130:715–720.

12. Jones DE, Rifai A Al-, Frith J, Patanwala I, Newton JL. The independent effects of fatigue and UDCA therapy on mortality in primary biliary cirrhosis: results of a 9 year follow-up. J. Hepatol. 2010;53:911–917.

13. Boonstra K, Weersma RK, van Erpecum KJ, Rauws EA, Spanier BWM, Poen AC, et al. Population-based epidemiology, malignancy risk, and outcome of primary sclerosing cholangitis. Hepatology. 2013;58:2045–2055.

14. Trivedi PJ, Hirschfield GM. Review article: overlap syndromes and autoimmune liver disease. Aliment. Pharmacol. Ther. 2012;36:517–533.

15. Mells GF, Pells G, Newton JL, Bathgate AJ, Burroughs AK, Heneghan MA, et al. Impact of primary biliary cirrhosis on perceived quality of life: The UK-PBC national study. Hepatology. 2013;58:273–283.

16. Prince M, Chetwynd A, Newman W, Metcalf JV, James OFW. Survival and symptom progression in a geographically based cohort of patients with primary biliary cirrhosis: follow-up for up to 28 years. Gastroenterology. 2002;123:1044–1051.

17. Quarneti C, Muratori P, Lalanne C, Fabbri A, Menichella R, Granito A, et al. Fatigue and pruritus at onset identify a more aggressive subset of primary biliary cirrhosis. Liver Int. 2015;35:636–641.

18. Tischendorf JJW, Hecker H, Krüger M, Manns MP, Meier PN. Characterization, Outcome, and Prognosis in 273 Patients with Primary Sclerosing Cholangitis: A Single Center Study. Am. J. Gastroenterol. 2007;102:107–114.

19. Broomé U, Olsson R, Lööf L, Bodemar G, Hultcrantz R, Danielsson A, et al. Natural history and prognostic factors in 305 Swedish patients with primary sclerosing cholangitis. Gut. 1996;38:610–615.

20. Claessen MMH, Vleggaar FP, Tytgat KMAJ, Siersema PD, van Buuren HR. High lifetime risk of cancer in primary sclerosing cholangitis. J. Hepatol. 2009;50:158–164.

21. Boberg KM, Bergquist A, Mitchell S, Pares A, Rosina F, Broomé U, et al. Cholangiocarcinoma in primary sclerosing cholangitis: risk factors and clinical presentation. Scand. J. Gastroenterol. 2002;37:1205–1211.

22. Murtaugh PA, Dickson ER, Van Dam GM, Malinchoc M, Grambsch PM, Langworthy AL, et al. Primary biliary cirrhosis: prediction of short-term survival based on repeated patient visits. Hepatology. 1994;20:126–134.

23. Borg PCJ ter, Schalm SW, Hansen BE, van Buuren HR, Dutch PBC Study Group. 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.

24. Corpechot C, Chazouillères O, Poupon R. Early primary biliary cirrhosis: biochemical response to treatment and prediction of long-term outcome. J. Hepatol. 2011;55:1361–1367.

25. 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;105:2186–2194.

26. Papastergiou V, Tsochatzis EA, Rodriquez-Peralvarez M, Thalassinos E, Pieri G, Manousou P, et al. Biochemical criteria at 1 year are not robust indicators of response to ursodeoxycholic acid in early primary biliary cirrhosis: results from a 29-year cohort study. Aliment. Pharmacol. Ther. 2013;38:1354–1364.

27. Zhang L-N, Shi T-Y, Shi X-H, Wang L, Yang Y-J, 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;58:264–272.

28. **Trivedi PJ, Bruns T,** Cheung A, Li K-K, Kittler C, Kumagi T, 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.

29. Kubota J, Ikeda F, Terada R, Kobashi H, Fujioka S, Okamoto R, et al. Mortality rate of patients with asymptomatic primary biliary cirrhosis diagnosed at age 55 years or older is similar to that of the general population. J. Gastroenterol. 2009;44:1000–1006.

30. Corpechot C, Carrat F, Poupon R, Poupon R-E. Primary biliary cirrhosis: incidence and predictive factors of cirrhosis development in ursodiol-treated patients. Gastroenterology. 2002;122:652–658.

31. Carbone M, Sharp, SJ, Heneghan MA, Neuberger JM, Hirschfield GM, Burroughs AK, et al. Histological Stage is Relevant for Risk-Stratification in Primary Biliary Cirrhosis. J. Hepatol. 2015;62:S805.

32. Cheung A, Lammers WJ, Hirschfield GM, Invernizzi P, Mason AL, Ponsioen CY, et al. Age, Bilirubin and Albumin, Regardless of Sex, are the Strongest Independent Predictors of Biochemical Response and Transplantation-Free Survival in Patients with Primary Biliary Cirrhosis. J. Hepatol. 2015;62:S798.

33. Joshita S, Umemura T, Ota M, Tanaka E. AST/platelet ratio index associates with progression to hepatic failure and correlates with histological fibrosis stage in Japanese patients with primary biliary cirrhosis. J. Hepatol. 2014;61:1443–1445.

34. Carbone M, Sharp, SJ, Heneghan MA, Neuberger JM, Hirschfield GM, Burroughs AK, et al. The UK-PBC Risk Score: Derivation and Validation of a Risk Score to Predict Liver Events in the UK-PBC Research Cohort. J. Hepatol. 2015;62:S201.

35. Lammers WJ, Pares A, Corpechot C, Nevens F, Janssen HLA, Ponsioen CY, et al. Identification of PBC Patients in Need of Additional Therapy During the Course of UDCA Treatment - An International Multicenter Study. J. Hepatol. 2015;62:s786.

36. Kim WR, Therneau TM, Wiesner RH, Poterucha JJ, Benson JT, Malinchoc M, et al. A revised natural history model for primary sclerosing cholangitis. Mayo Clin. Proc. 2000;75:688–694.

37. Lindor KD, Kowdley KV, Luketic VAC, Harrison ME, McCashland T, Befeler AS, et al. High-dose ursodeoxycholic acid for the treatment of primary sclerosing cholangitis. Hepatology. 2009;50:808–814.

38. Stanich PP, Björnsson E, Gossard AA, Enders F, Jorgensen R, Lindor KD. Alkaline phosphatase normalization is associated with better prognosis in primary sclerosing cholangitis. Dig. Liver Dis. 2011;43:309–313.

39. Rupp C, Rössler A, Halibasic E, Sauer P, Weiss K-H, Friedrich K, et al. Reduction in alkaline phosphatase is associated with longer survival in primary sclerosing

cholangitis, independent of dominant stenosis. Aliment. Pharmacol. Ther. 2014;40:1292–1301.

40. Mamari S Al, Djordjevic J, Halliday JS, Chapman RW. Improvement of serum alkaline phosphatase to <1.5 upper limit of normal predicts better outcome and reduced risk of cholangiocarcinoma in primary sclerosing cholangitis. J. Hepatol. 2013;58:329–334.

41. Goode E, Srivastava, B, Clark AB, Spiess K, Hirschfield GM, Gelson WTH, et al. Early Clinical Features Associated with Long-Term Risk of Transplantation in Primary Sclerosing Cholangitis: Results from the UK-PSC Consortium. J. Hepatol. 2015;62:S230.

42. Lindström L, Hultcrantz R, Boberg KM, Friis–Liby I, Bergquist A. Association Between Reduced Levels of Alkaline Phosphatase and Survival Times of Patients With Primary Sclerosing Cholangitis. Clin. Gastroenterol. Hepatol. 2013;11:841–846.

43. Watanabe T, Hirano K, Tada M, Isayama H, Mizuno S, Arizumi T, et al. Short-term prognostic factors for primary sclerosing cholangitis: Prognostic Factors for PSC. J. Hepato-Biliary-Pancreat. Sci. 2015;22:486–490.

44. Nakamura M, Kondo H, Mori T, Komori A, Matsuyama M, Ito M, et al. Anti-gp210 and anti-centromere antibodies are different risk factors for the progression of primary biliary cirrhosis. Hepatology. 2007;45:118–127.

45. Floreani A, Franceschet I, Cazzagon N, Spinazzè A, Buja A, Furlan P, et al. Extrahepatic autoimmune conditions associated with primary biliary cirrhosis. Clin. Rev. Allergy Immunol. 2015;48:192–197.

46. Alswat K, Harthy N Al-, Mazrani W, Alshumrani G, Jhaveri K, Hirschfield GM. The spectrum of sclerosing cholangitis and the relevance of IgG4 elevations in routine practice. Am. J. Gastroenterol. 2012;107:56–63.

47. **Boonstra K, Culver EL,** de Buy Wenniger LM, van Heerde MJ, van Erpecum KJ, Poen AC, et al. Serum immunoglobulin G4 and immunoglobulin G1 for distinguishing immunoglobulin G4-associated cholangitis from primary sclerosing cholangitis. Hepatology. 2014;59:1954–1963.

48. Mendes FD, Jorgensen R, Keach J, Katzmann JA, Smyrk T, Donlinger J, et al. Elevated serum IgG4 concentration in patients with primary sclerosing cholangitis. Am. J. Gastroenterol. 2006;101:2070–2075.

49. Benito de Valle M, Müller T, Björnsson E, Otten M, Volkmann M, Guckelberger O, et al. The impact of elevated serum IgG4 levels in patients with primary sclerosing cholangitis. Dig. Liver Dis. 2014;46:903–908.

50. Rudolph G, Gotthardt D, Kloeters-Plachky P, Rost D, Kulaksiz H, Stiehl A. In PSC with dominant bile duct stenosis, IBD is associated with an increase of carcinomas and reduced survival. J. Hepatol. 2010;53:313–317.

51. Burak K, Angulo P, Pasha TM, Egan K, Petz J, Lindor KD. Incidence and Risk Factors for Cholangiocarcinoma in Primary Sclerosing Cholangitis. Am. J. Gastroenterol. 2004;99:523–526.

52. Ngu JH, Gearry RB, Wright AJ, Stedman CAM. Inflammatory bowel disease is associated with poor outcomes of patients with primary sclerosing cholangitis. Clin. Gastroenterol. Hepatol. 2011;9:1092–1097.

53. Janse M, Lamberts LE, Verdonk RC, Weersma RK. IBD is associated with an increase in carcinoma in PSC irrespective of the presence of dominant bile duct stenosis. J. Hepatol. 2012;57:473–474.

54. Ponsioen CY, Reitsma JB, Boberg KM, Aabakken L, Rauws EA, Schrumpf E. Validation of a cholangiographic prognostic model in primary sclerosing cholangitis. Endoscopy. 2010;42:742–747.

55. Ruiz A, Lemoinne S, Carrat F, Corpechot C, Chazouillères O, Arrivé L. Radiologic course of primary sclerosing cholangitis: assessment by three-dimensional magnetic resonance cholangiography and predictive features of progression. Hepatology. 2014;59:242–250.

56. Lemoinne S, Arrivé L, Fankem ADK, Housset C, Corpechot C, Chazouillères O. Ability of a Simple Radiologic Score, Assessed by Three Dimensional Magnetic Resonance Cholangiography (MRC), to Predict Clinical Outcome of Patients with Primary Sclerosing Cholangitis (PSC). J. Hepatol. 2015;62:S794.

57. Weismüller TJ, Lankisch TO. Medical and endoscopic therapy of primary sclerosing cholangitis. Best Pract. Res. Clin. Gastroenterol. 2011;25:741–752.

58. Chapman MH, Webster GJ, Bannoo S, Johnson G, Wittmann J, Pereira SP. Cholangiocarcinoma and dominant strictures in patients with primary sclerosing cholangitis; a 25 year single centre experience. Eur. J. Gastroenterol. Hepatol. 2012;24:1051–1058.

59. Björnsson E, Lindqvist-Ottosson J, Asztely M, Olsson R. Dominant Strictures in Patients with Primary Sclerosing Cholangitis. Am. J. Gastroenterol. 2004;99:502–508.

60. Rudolph G, Gotthardt D, Klöters-Plachky P, Kulaksiz H, Rost D, Stiehl A. Influence of dominant bile duct stenoses and biliary infections on outcome in primary sclerosing cholangitis. J. Hepatol. 2009;51:149–155.

61. Gotthardt DN, Rudolph G, Klöters-Plachky P, Kulaksiz H, Stiehl A. Endoscopic dilation of dominant stenoses in primary sclerosing cholangitis: outcome after long-term treatment. Gastrointest. Endosc. 2010;71:527–534.

62. Björnsson E, Boberg KM, Cullen S, Fleming K, Clausen OP, Fausa O, et al. Patients with small duct primary sclerosing cholangitis have a favourable long term prognosis. Gut. 2002;51:731–735.

63. Angulo P, Maor-Kendler Y, Lindor KD. Small-duct primary sclerosing cholangitis: A long-term follow-up study. Hepatology. 2002;35:1494–1500.

64. Björnsson E, Olsson R, Bergquist A, Lindgren S, Braden B, Chapman RW, et al. The Natural History of Small-Duct Primary Sclerosing Cholangitis. Gastroenterology. 2008;134:975–980.

65. Kakuda Y, Harada K, Sawada-Kitamura S, Ikeda H, Sato Y, Sasaki M, et al. Evaluation of a new histologic staging and grading system for primary biliary cirrhosis in comparison with classical systems. Hum. Pathol. 2013;44:1107–1117.

66. Wendum D, Boëlle P-Y, Bedossa P, Zafrani E-S, Charlotte F, Saint-Paul M-C, et al. Primary biliary cirrhosis: proposal for a new simple histological scoring system. Liver Int. 2015;35:652–659.

67. Vleggaar FP, van Buuren HR, Zondervan PE, Kate FJ ten, Hop WC. Jaundice in non-cirrhotic primary biliary cirrhosis: the premature ductopenic variant. Gut. 2001;49:276–281.

68. Ozaslan E, Efe C, Heurgué-Berlot A, Kav T, Masi C, Purnak T, et al. Factors associated with response to therapy and outcome of patients with primary biliary cirrhosis with features of autoimmune hepatitis. Clin. Gastroenterol. Hepatol. 2014;12:863–869.

69. de Vries EMG, Verheij J, Hubscher SG, Leeflang MMR, Boonstra K, Beuers U, et al. Applicability and Prognostic Value of Histologic Scoring Systems in Primary Sclerosing Cholangitis. J. Hepatol. 2015

70. Corpechot C, Carrat F, Poujol-Robert A, Gaouar F, Wendum D, Chazouillères O, et al. Noninvasive elastography-based assessment of liver fibrosis progression and prognosis in primary biliary cirrhosis. Hepatology. 2012;56:198–208.

71. Floreani A, Cazzagon N, Martines D, Cavalletto L, Baldo V, Chemello L. Performance and utility of transient elastography and noninvasive markers of liver fibrosis in primary biliary cirrhosis. Dig. Liver Dis. 2011;43:887–892.

72. Corpechot C, Gaouar F, Naggar A El, Kemgang A, Wendum D, Poupon R, et al. Baseline values and changes in liver stiffness measured by transient elastography are associated with severity of fibrosis and outcomes of patients with primary sclerosing cholangitis. Gastroenterology. 2014;146:970–979.

73. Mayo M, Parkes J, Adams-Huet B, Combes B, Mills A, Markin R, et al. Prediction of Clinical Outcomes in Primary Biliary Cirrhosis by Serum Enhanced Liver Fibrosis (ELF) Assay. Hepatology. 2008;48:1549–1557.

74. **Vesterhus M, Hov JR,** Holm A, Schrumpf E, Nygård S, Godang K, et al. Enhanced liver fibrosis score predicts transplant-free survival in primary sclerosing cholangitis. Hepatology. 2015;62:188–97.

75. Halliday JS, Djordjevic J, Lust M, Culver EL, Braden B, Travis SPL, et al. A unique clinical phenotype of primary sclerosing cholangitis associated with Crohn's disease. J. Crohns Colitis. 2012;6:174–181.

76. Kuiper EMM, Hansen BE, de Vries RA, Ouden–Muller JW den, van Ditzhuijsen TJM, Haagsma EB, et al. Improved Prognosis of Patients With Primary Biliary Cirrhosis That Have a Biochemical Response to Ursodeoxycholic Acid. Gastroenterology. 2009;136:1281–1287.

77. Parés A. Old and novel therapies for primary biliary cirrhosis. Semin. Liver Dis. 2014;34:341–351.

78. Iorio A, Spencer FA, Falavigna M, Alba C, Lang E, Burnand B, et al. Use of GRADE for assessment of evidence about prognosis: rating confidence in estimates of event rates in broad categories of patients. The BMJ. 2015;350:h870.

**Table 1: Common Pitfalls in Risk Stratification Studies**

| Precedents | Apprehensions |
|---|---|
| **Well-recognized; frequently addressed** | |
| Inadequate follow-up time | Restricts assessment to late-stage events.<br>Poorly predictive of at-risk groups from disease outset. |
| Single-center studies | Lack of independent, external validation<br>May not be generalizable. |
| Small sample size | Inadequate statistical power.<br>Increased type-I error. |
| **Well-recognized; infrequently addressed** | |
| Representation limited to retrospective data collection | Incomplete data.<br>Selection bias and confounding factors.<br>Reduces statistical power |
| Extrapolation of risk-stratifiers beyond original intended end-point / time-point | Increased type-II error. |
| Tertiary-center restricted studies | Recruitment /referral bias; not necessarily representative of the disease globally. |
| Logistic regression (odds' ratios; OR) applied for clinical endpoints that are not time-constrained | Time-variable events (e.g. death, transplantation) necessitate Cox-regression (hazard ratios; HR), or equivalent. |
| 'Time-dependent' covariates modeled as 'time-constant' | Inadequate representation of chronal displacement between potential risk factors and the measured endpoint. |
| **Poorly-recognized; infrequently addressed** | |
| Predictive utility of a continuous variable not proven prior to dichotomization | Subjugation of variability<br>Increased type-I error<br>Conceals non-linearity between covariates and clinical outcomes |
| Independently predictive covariates deemed additive | Incorrect classification of risk through weaker stratifiers |
| Incorrect estimation of median event-free survival times | *Incorrect approach:*<br>Calculated median: mid-point of all pts.' follow-up times.<br><br>*Correct approach:*<br>Actuarial median: time at which 50% of the cohort meets the clinical endpoint according to Kaplan-Meier estimates. |
| Contrary inferences | Favorable clinical outcomes suggested for 'low-risk' groups stratified according to dichotomous variables, despite event-free survival dwarfing that of a matched control population. |

**Table 2: Variant Presentations**

| **A) Primary biliary cirrhosis** | |
|---|---|
| **Phenotypic variant** (% of pt. population) | **Prognostic implication** |
| Male sex (2,9,32)<br>(~5-10%) | • Older age at diagnosis relative to women (60 vs. 55 yrs. $p<0.001$)<br>• Greater frequency of non-response (63% vs. 76%; $p<0.001$)<br>– Likely attributable to more advanced baseline disease<br>• Increased HCC risk in biochemical non-responders, as well as cirrhotic pts. |
| Young presenting age (9,32)<br>(~25%) | • Biochemical response rate in women <40 yrs. old less than 50% |
| AMA negative (8,9,28)<br>(~5 – 10%) | • Clinical course identical to AMA-positive PBC |
| Intractable pruritus (12,17)<br>(dynamic frequency reported) | • Consider referral for clinical trials specifically targeting pruritus<br>• Rarely can be severe enough to merit transplantation as the solitary indication |
| Overlap* with autoimmune hepatitis (AIH) (14)<br>(~3 – 20%; diagnostic criteria inconsistent) | • Severity of interface hepatitis predictive of biochemical non-response<br>– Unclear if biochemical response criteria predict prognosis in overlap<br>• Inconclusive data regarding clinical outcome relative to PBC alone. |
| **B) Primary sclerosing cholangitis** | |
| **Phenotypic variant** (% of pt. population) | **Prognostic implication** |
| Young presenting age (14)<br>(frequency unclear – lack of consensus on disease nomenclature between children and adult pts.) | • Under 16s more commonly present with an inflammatory phenotype<br>• Speculated to endure a more rapidly progressive course (unconfirmed) |
| Small duct disease (13,62–64)<br>(10 – 15%) | • Less often symptomatic (30% vs. 53%; $p<0.01$)<br>• Disease progression infrequent relative to classical PSC<br>– Transplantation rate ~10%<br>– LTx-free survival relative to classical form: 29 yrs. vs. 17 yrs.; $p=0.04$.<br>• Increased risk of CCA not evident in small duct disease<br>• Up to 25% may progress to large duct disease over ~7 yrs. |
| Dominant strictures (50,60)<br>(12 – 60%) | • Strictly defined on ERC criteria<br>• Increased mortality largely attributed to inability in differentiating benign from malignant lesions<br>• Reduced transplant-free survival (particularly if coexists with colitis)<br>– 25% vs. 73% over 20 years; $p=0.011$ |
| Coexisting Crohn's disease (75)<br>(13% - one study) | • 50% female<br>• 22% have small duct PSC – consequently exhibit improved LTx-free survival<br>• Impact of small intestinal versus colonic Crohn's not yet discerned |
| Elevated serum IgG4 levels (46–49)<br>(9 – 15%) | • More significant elevations in serum liver biochemistry/Mayo score<br>• Impact on clinical outcome unclear |
| Overlap* with autoimmune hepatitis (AIH) (14)<br>(~1 – 53%; diagnostic criteria inconsistent) | • 50% of children diagnosed with AIH aged <16 yrs. 'evolve' into PSC<br>• Inconclusive data regarding clinical outcome relative to PSC alone.<br>• Frequency of coexisting IBD similar (PSC likely dominant disease process) |

* The prevalence of overlap is difficult to ascertain because of publication bias, variable definitions, and considerable heterogeneity between syndrome designations. Moreover, the limitations of applying surrogates of outcome to settings distinct from which they

were originally intended must be recognised (covered elsewhere (14)). Given small numbers of patients comprising few non-randomised, non-blinded studies, evidenced-based risk-stratification centred on the relative presence/absence of overlap features is not possible currently, and worthy of prospective multi-centre collaborative investigation.

**Table 3: Biochemical Response in Primary Biliary Cirrhosis.**

| **A) Established models derived from biochemical/laboratory parameters (2006 – 2010)** | | | | |
|---|---|---|---|---|
| **Proposed response criteria** *(Attainment rate)* | **Derivation cohort;** No. of pts. | **Clinical event rate;** R. vs. N.R.; log-rank *(Interval)* | **Validation series;** No. of pts. | **Comments** |
| **Barcelona (11)**<br>> 40% decrease in ALP (or normalization) from baseline<br><br>*(Achieved by 61% of pts.)* | *n*=192; single-center | 3% vs. 17%; $p<0.01$<br>*(Median F/Up 7.5 yrs.)* | *n*=2353; multi-center (N) (9)<br>*n*=1015; multi-center (I) (28) | Delta change in ALP less predictive than absolute reduction in subsequent series (8) |
| **Paris – I (10)**<br>ALP < 3 x ULN, and<br>AST < 2x ULN, and<br>bilirubin ≤ 1mg/dL<br><br>*(Achieved by 61% of pts.)* | *n*=292; single-center | 10% vs. 49%; $p<0.001$<br>*(At 10 yrs.)* | *n*=2353; multi-center (N) (9)<br>*n*=1015; multi-center (I) (28) | Non-response also predictive of HCC risk (2)<br><br>Most robustly validated of all criteria |
| **Rotterdam (76)**<br>Albumin and bilirubin normalization<br><br>*(Achieved by 76% of pts.)* | *n*=311; multi-center (N) | 19% vs. 44%; $p<0.001$<br>*(At 10 yrs.)* | *n*=1015; multi-center (I) (28) | Non-response also predictive of HCC risk (2)<br><br>Limited applicability in early-stage disease |
| **Toronto (25)**<br><br>ALP < 1.67 x ULN | *n*=69; single-center | *N/A* | *n*=2353; multi-center (N) (9)<br>*n*=1015; multi-center (I) (28) | Non-response also predictive of HCC risk (2)<br><br>Endpoint in the derivation study was histological progression, whereas association with clinical outcome verified in subsequent validation series. |

**B) Modification of existing response criteria (2011 – 2013)**

| Proposed response criteria | Derivation cohort; No. of pts. | Clinical event rate; R. vs. N.R.; log-rank *(Interval)* | Validation series; No. of pts. | Comments |
|---|---|---|---|---|
| **Paris – II (24)** – *early stage PBC only*<br>ALP ≤ 1.5 x ULN,<br>AST ≤ 1.5 x ULN, and<br>bilirubin ≤ 1 mg/dL<br><br>*(Achieved by 48% of pts.)* | *n*=165; single-center | 0% vs. 13%; *p*<0.001<br>*(Mean F/Up 7 yrs.)* | *n*=2353; multi-center (N) (9)<br>*n*=1015; multi-center (I) (28) | Externally validated;<br>Non-response also predictive of HCC risk (2)<br><br>Concern over modest sensitivity (50%) and low NPV (13%) may incorrectly stratify high-risk pts. |
| **Early biochemical response (27)**<br>Paris; Barcelona; Toronto met at 6 mo. | *n=187;* single-center | See comments | *Not yet validated* | Biochemical response rates<br>(Paris–I; 71% vs. 70%) and respective<br>PPV (0.90 vs. 0.91) / NPV (0.45 vs. 0.47) of future clinical events identical at 6 vs. 12 mo. |

**C) Optimization of prior response criteria (2013 – 2015)**

| Proposed response criteria | Derivation cohort; No. of pts. | F/Up interval | Validation series; No. of pts. | Net reclassification index (N.R.I) |
|---|---|---|---|---|
| **Biochemical response + APRI-r1 ≤0.54* (28)** | *n*=352; single-center | Median: 7 yrs. | *n*=446; multi-center (I) (28) | 20% |
| **UK-PBC risk score (34) ****<br>Prognostic index comprising bilirubin, ALP ratio, AST, ALT albumin and platelet count. | *n*=1916; multi-center (N) | Median: 7 yrs. | *n*=1249; multi-center (N) (34) | Not specified |
| **GLOBE score; www.globalpbc.com (35) ****<br>Prognostic index comprising pt. age, bilirubin, ALP ratio, albumin and platelet count. | *n*=2488; multi-center (I) | Median: 8 yrs. | *n*=1631; multi-center (I) (35) | 25 – 35% |

Several response criteria are proposed in primary biliary cirrhosis (PBC), wherein liver transplant (LTx)-free survival akin to that of a matched population is predicted following attainment of well-defined parameters; most-often applied at 1 – 2 years following ursodeoxycholic acid (UDCA) treatment/PBC diagnosis. Although the optimum cut-point for serum ALP is difficult to define, it is apparent that absolute levels during follow-up predicts outcome with higher accuracy relative to percentage decrease.

Event rates in responders (R) vs. non-responders (N.R.) are provided for principle studies in (**A**). Modifications to existing criteria; specifically, targeting patients with early–stage disease, as well as 6-month vs. 12 month biochemical response have been attempted, although the latter approach is awaiting validation (**B**). Improvements of prior response criteria are being attempted, with examples provided for current studies **(C)**. Single-center and multi-center national (N) and international (I) studies are denoted accordingly.

* Application of APRI-score at 1-year (APRI-r1) to all pre-existing biochemical criteria has been shown to improve predictive performance.

** Full results yet to be published.

**Table 4: Proposed ALP Thresholds in Primary Sclerosing Cholangitis**

| **Proposed criteria** *(Attainment rate)* | **Derivation cohort;** No. of pts. | **Clinical event rate;** R. vs. N.R. *(Interval)* | **Endpoints tested** | **Apprehensions** |
|---|---|---|---|---|
| **Rochester (38) *** ALP normalization at any point (median attainment time: 1 yr.) *(Achieved by 40% of pts.)* | *n*=87; single-center | 14% vs. 33%; *p*=0.02 *(Median follow-up 7.3 yrs.)* | Death, LTx., CCA | Small number of pts. 20% of UDCA-treated individuals reached clinical endpoint despite normal serum ALP |
| **Oxford (40)** ALP < 1.5 x ULN at 2 yrs. (*Achieved by 40% of pts.)* | *n*=139; single-center | 6% vs. 38%; *p*<0.0001 *(Median follow-up 10 yrs.)* | Decompensation, Death, LTx., CCA | Tested in Heidelberg study at 6 mo. (39) and UK-PSC multi-center study at 1-year time-point (41). ALP threshold not successfully validated at original 2-yr. time-point. |
| **Scandinavian multi-center (42)** ALP > 40% decline from baseline or normal at 1 yr. (*Achieved by 40% of pts.)* | *n*=195; multi-center (I) | Rate not specified*;* Diff. between groups: *p*<0.001 | Death, LTx., CCA | Not successfully validated (39) |
| **Heidelberg (39)** ALP <1.5 x ULN., or ALP ≥ 50% decline, or ALP normal 6 mo. from baseline *(Achieved by 51% of pts. – any above)* | *n*=185; single-center | 13% vs. 49%; *p*<0.05 *(Median follow-up 10 yrs.)* | Death, LTx, CCA. | Clinical event rate not significantly different between groups when ALP < 1.5xULN threshold applied (in isolation) at 1 yr. |
| **UK-PSC (41) **** Criteria 1) ALP < 1.5 x ULN at 1 yr. Criteria 2) ALP < 2.0 X ULN at 2 yrs. (*Attainment rates not yet available)* | *n*=1200; multi-center (N) | Rate not specified*;* Significant difference between gps.: 1) *p*<0.001, and 2) *p*=0.015 | LTx. only | Threshold of ALP < 1.5xULN did not prove discriminatory when applied at 2 yrs. |

Emerging biochemical response criteria in PSC patients based on varying thresholds of serum ALP applied 6 – 24 months following diagnosis. Attainment of these criteria is purported to infer significantly improved clinical outcome in the individual cohorts studied, although comparisons to matched control population are yet to be drawn, and none of the inclusive studies have assessed serum ALP as a continuous variable prior to application of presented cut-points.

* Pre-defined time-point not specified.

** Full results yet to be published.

**Table 5: Prognosis-Related Histological Themes in Primary Biliary Cirrhosis**

| Feature | Comments |
|---|---|
| **Fibrosis** | Advanced septal fibrosis predictive of UDCA-failure and clinical outcome (10,25,68). |
| **Interface hepatitis** | Positive correlation with AST / ALT (Spearman's $\rho$: 0.469 / 0.395; $p$<0.05) (66).<br><br>Moderate–severe activity* independently predictive of biochemical non-response, histological stage progression, progression to transplantation and death (relative risk 1.9; $p$=0.002) (10,30,31,72).<br><br>Improvement in interface activity (in the absence of significant fibrosis) with corticosteroids reported in at least two randomized trials and one multi-center observational study (77) |
| **Ductopenia** | Negative correlation of bile duct ratio** with serum ALP (Spearman's $\rho$: –0.362; $p$<0.05) (66)<br><br>Duct loss in >50% of portal tracts predicts histological disease progression and failure to meet biochemical response (25).<br><br>Premature ductopenic variant affects 5–10% of pts: characterized by: rapid onset bile duct loss without significant baseline fibrosis, severe icteric cholestasis and rapid progression toward cirrhosis (<5 years (67)). |
| **Chronic cholestasis** | Deposition of orcein-positive granules in periportal hepatocytes predictive of development of cirrhosis-related conditions (65). |

* Moderate: segmental necrosis at periphery of >50% of portal tracts or circumferential necrosis in <50% of portal tracts
Severe: Circumferential necrosis in >50% of portal tracts.
** Ratio of the number of portal tracts with ducts to total number of portal tracts

**Table 6: Non-Invasive Evaluation of Liver Fibrosis**

| Modality | Primary Biliary Cirrhosis (PBC) | Primary Sclerosing Cholangitis (PSC) |
|---|---|---|
| **Vibration controlled transient elastography (VCTE)** | *Precedents:*<br>Increased risk of clinical events (decompensation, LTx and liver-related mortality) independent of biochemical response in patients with LSM >9.6kPa, or ΔLSM>2.1kPa/yr. (70).<br><br>*Studied cohorts*:<br>*n*=150; single-center and UDCA treated<br><br>*Comment:*<br>Proven surrogate of fibrosis in PBC. However, validation as an outcome predictor pending.<br><br>Unclear whether adds predictive value to biochemical response status. | *Precedents:*<br>Higher baseline LSM (>9.9kPa most sensitive), or ΔLSM>1.3kPa/yr. predictive of adverse clinical events (72). Low-risk group best discriminated by LSM ≤6.5kPa<br><br>*Studied cohorts:*<br>*n*=167; single-center<br><br>*Comment:*<br>Impact of severe cholestasis/ cholangitis/ IBD activity uncertain.<br><br>Validation as an outcome predictor pending |
| **Enhanced Liver Fibrosis score (ELF®)** | *Precedents:*<br>Significant differences in clinical event rate between score tertiles (73).<br><br>Δ1-point increase imparts 3-fold greater risk of liver-related events<br><br>*Studied cohorts*:<br>*n*=161; multicenter national data extrapolated from a clinical trial of methotrexate and UDCA.<br><br>*Comment:*<br>Unclear whether adds predictive value to biochemical response status.<br><br>Impact of longitudinal stability versus fluctuations over time yet to be determined.<br><br>Not yet validated.<br><br>Unclear whether stratifier of disease severity vs. stage. | *Precedents:*<br>LTx.-free survival significantly greater in pts. harboring low (9.7 yrs.) vs. high (1.3 yrs.) ELF® scores (threshold: 10.6) (74).<br><br>*Studied cohorts:*<br>*n*=167 (derivation) + 138 (validation)<br><br>*Comment:*<br>Internal multi-center validation. However, short disease duration in LTx-free survivors in the original report (<5 yrs.).<br><br>Impact of longitudinal stability versus fluctuations over time yet to be determined.<br><br>Unclear whether stratifier of disease severity vs. stage. |

**Figure 1. Approaches to Risk Stratification in Autoimmune Cholestatic Liver Disease**

The presented infographic illustrates the authors' ranking with regard to currently available prognostic models and scoring systems, ordered dependent on predictive performance, validation status and routine clinical applicability. For instance, biochemical response criteria represent the most robust discrimination method of at-risk populations in PBC, and can be assessed non-invasively by clinically acceptable means. However, liver histology is perhaps the most biologically representative index of disease progression (PBC and PSC); yet routine, ongoing assessment through serial liver biopsies clearly unacceptable in routine clinical practice. Nevertheless, application of robust non-invasive surrogates holds promise, (particularly for transient elastography, which as a fibrosis indicator is very well-substantiated in PBC) and may be extrapolated to forecasting clinical outcomes. Further validation of these modalities as independent and more so additive predictors is eagerly awaited, particularly to discriminate severity versus stage of disease. Conversely, the emergence of serum IgG4 and serum ALP as putative risk stratifiers in PSC are not supported by well-controlled or high-quality validation, and studies incorporating assessment as continuous variables with inclusive control populations urgently commanded.

Levels of evidence for each stratagem are indicated in superscript according to the recently revised GRADE criteria for assessment of prognosis (high=A, intermediate=B, low=C, very low=D) (**Supplementary Table 1** (78)).

**Figure 2a. Proposed Pathway to Clinical Trial Recruitment: Primary Biliary Cirrhosis**

Biochemical response criteria are the most robust of all predictive modalities, with greatest chance of attainment following UDCA provision. Current strategies require assessment at 12 months (a), although increasing identification of presenting phenotypes in which therapeutic failure is more common may call for earlier application of response criteria (e.g. at 6 months) if validation holds true (b). This group is speculated to include young women; and where available those who test positive for anti-gp210 reactivity or who exhibit an elevated baseline liver stiffness measurement (LSM) as measured by transient elastography. Up till now, Paris–I has been externally validated as the most accurate discriminator (optimal response models may differ according to study population), and certain individuals fail therapy predominantly on transaminase indices (c). Whilst not necessarily classifying an 'overlap syndrome,' significant interface hepatitis may be conducive to adjuvant corticosteroids, and an argument for stratification through liver histology is presented at this stage (14,68). Biochemical non-response imparts additional hepatocellular carcinoma (HCC) risk (d), with highest incidence in cirrhotic patients; and men irrespective of disease stage. Additionally, some patients experience progressive liver disease despite fulfilling response criteria (e), and sequential application of APRI is currently the only validated method that assists in their early recognition – until such time as the additional discriminatory value of annual change in LSM is substantiated. All patients with evidence suggestive of portal hypertension (PHT), irrespective of liver disease stage are also recommended to undergo endoscopic variceal surveillance according to current guidelines and local expertise, given the negative clinical impact of varices on disease outcome (4). Incorporating such a stepwise algorithm to all newly presenting, well-compensated patients (outside of transplantation criteria) will likely capture the greatest breadth of at-risk individuals, wherein therapeutic shortfall is most evident.

**Figure 2b. Proposed Pathway to Clinical Trial Recruitment: Primary Sclerosing Cholangitis**

The unpredictable clinical nature and dearth of effective medical therapy in PSC means that the vast majority of patients outside of transplant criteria currently harbor >1 high-risk classifier at time of presentation (a), including the presence of colitis, persistently elevated liver biochemistry, or features predictive of advancing fibrosis or future cholangiographic progression. Symptomatic presentations in addition to indicators of advancing fibrosis also predict adverse clinical outcome although the relative and independent predictive value between modalities are yet to be established in PSC, with ELF® score being somewhat restricted and of limited routine availability. Moreover, as a continuous variable the optimum stratification threshold utilizing elastography is not yet defined, with liver stiffness measurement (LSM) >9.9kPa the best discriminator for identifying high-risk individuals, yet ≤6.5kPa most indicative of early disease. Nevertheless, the dynamic impact of chronal increments is well-demonstrated for elastography (b) and possibly for progressive MRC scores (not illustrated; formal publication pending); signifying further groups in whom clinical trials should be encouraged. Conversely, asymptomatic patients with small duct disease, as well as those with classical PSC achieving persistently low/normal liver biochemistry who maintain stable fibrotic indices in the absence of cholangiographic progression, likely herald a more consistent low-risk profile (c); albeit with need for longitudinal appraisal (d) given that early predictive models of disease progression are not yet available. Indeed, regular risk-assessment of malignant complications is critical to ensure long-term patient safety, for no early or robust predictors of future CCA currently exist. To this effect, a position for even those in the lower risk category (with large duct disease) to be considered for clinical trials specifically targeted at reducing CCA incidence can also be argued (open arrows), whilst accepting the strong probability that other PSC-related clinical events develop at a low incidence. The optimum frequency of routine radiological

surveillance is often debated (e), with no evidence-based guidance in this regard. A suggested policy of 12-monthly (detection of gallbladder polyps), or 6-monthly in cirrhotic patients (hepatocellular carcinoma surveillance) is proposed in keeping with current guidelines.

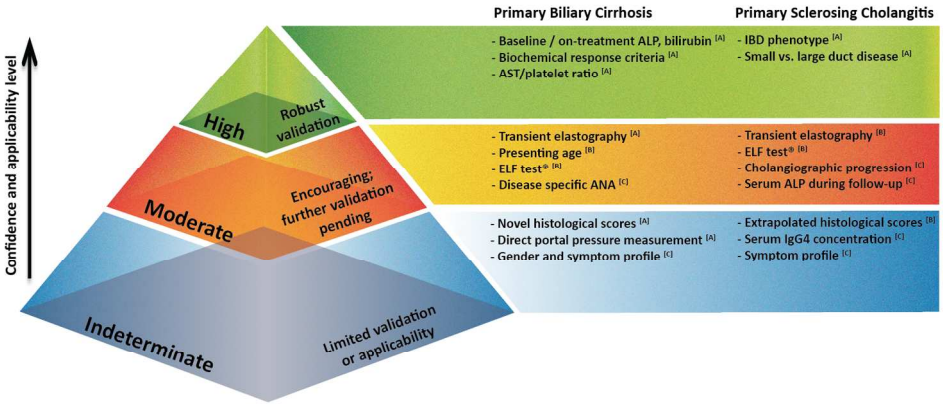
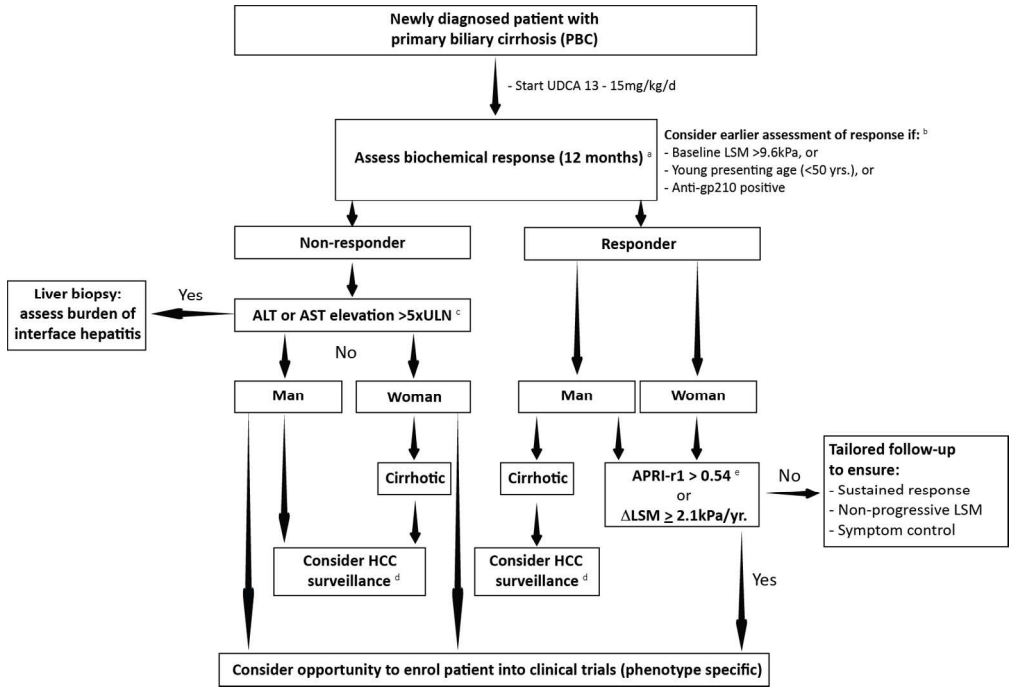
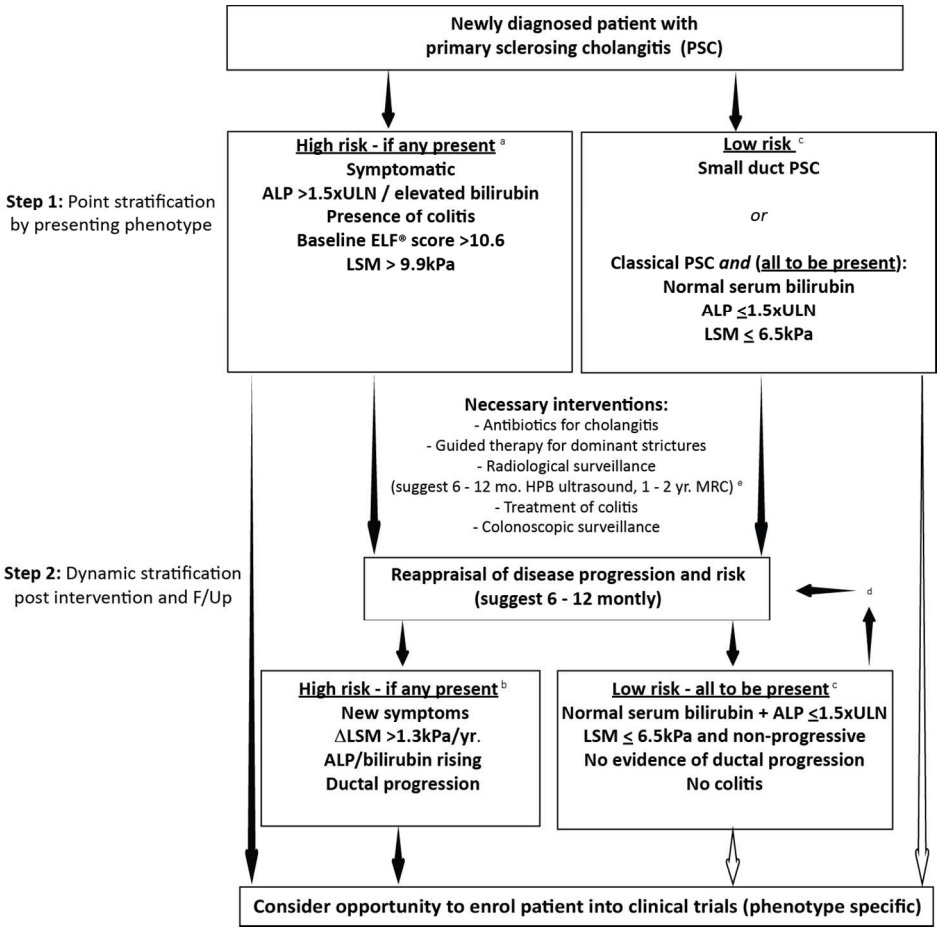



Figure 1. Approaches to Risk Stratification in Autoimmune Cholestatic Liver Disease
The presented infographic illustrates the authors' ranking with regard to currently available prognostic models and scoring systems, ordered dependent on predictive performance, validation status and routine clinical applicability. For instance, biochemical response criteria represent the most robust discrimination method of at-risk populations in PBC, and can be assessed non-invasively by clinically acceptable means. However, liver histology is perhaps the most biologically representative index of disease progression (PBC and PSC); yet routine, ongoing assessment through serial liver biopsies clearly unacceptable in routine clinical practice. Nevertheless, application of robust non-invasive surrogates holds promise, (particularly for transient elastography, which as a fibrosis indicator is very well-substantiated in PBC) and may be extrapolated to forecasting clinical outcomes. Further validation of these modalities as independent and more so additive predictors is eagerly awaited, particularly to discriminate severity versus stage of disease. Conversely, the emergence of serum IgG4 and serum ALP as putative risk stratifiers in PSC are not supported by well-controlled or high-quality validation, and studies incorporating assessment as continuous variables with inclusive control populations urgently commanded.

Levels of evidence for each stratagem indicated in superscript according to the recently revised GRADE criteria for assessment of prognosis (high=A, intermediate=B, low=C, very low=D) (Supplementary Table 1 (78)).
155x71mm (300 x 300 DPI)




Biochemical response criteria are the most robust of all predictive modalities, with greatest chance of attainment following UDCA provision. Current strategies require assessment at 12 months (a), although increasing identification of presenting phenotypes in which therapeutic failure is more common may call for earlier application of response criteria (e.g. at 6 months) if validation holds true (b). This group is speculated to include young women; and where available those who test positive for anti-gp210 reactivity or who exhibit an elevated baseline liver stiffness measurement (LSM) as measured by transient elastography. Up till now, Paris–I has been externally validated as the most accurate discriminator (optimal response models may differ according to study population), and certain individuals fail therapy predominantly on transaminase indices (c). Whilst not necessarily classifying an 'overlap syndrome,' significant interface hepatitis may be conducive to adjuvant corticosteroids, and an argument for stratification through liver histology is presented at this stage (14,68). Biochemical non-response imparts additional hepatocellular carcinoma (HCC) risk (d), with highest incidence in cirrhotic patients; and men irrespective of disease stage. Additionally, some patients experience progressive liver disease despite fulfilling response criteria (e), and sequential application of APRI is currently the only validated method that assists in their early recognition – until such time as the additional discriminatory value of annual change in LSM is substantiated. All patients with evidence suggestive of portal hypertension (PHT), irrespective of liver disease stage are also recommended to undergo endoscopic variceal surveillance according to current guidelines and local expertise, given the negative clinical impact of varices on disease outcome (4). Incorporating such a stepwise algorithm to all newly presenting, well-compensated patients (outside of transplantation criteria) will likely capture the greatest breadth of at-risk individuals, wherein therapeutic shortfall is most evident.
146x100mm (300 x 300 DPI)




The unpredictable clinical nature and dearth of effective medical therapy in PSC means that the vast majority of patients outside of transplant criteria currently harbor >1 high-risk classifier at time of presentation (a), including the presence of colitis, persistently elevated liver biochemistry, or features predictive of advancing fibrosis or future cholangiographic progression. Symptomatic presentations in addition to indicators of advancing fibrosis also predict adverse clinical outcome although the relative and independent predictive value between modalities are yet to be established in PSC, with ELF® score being somewhat restricted and of limited routine availability. Moreover, as a continuous variable the optimum stratification threshold utilizing elastography is not yet defined, with liver stiffness measurement (LSM) >9.9kPa the best discriminator for identifying high-risk individuals, yet <6.5kPa most indicative of early disease. Nevertheless, the dynamic impact of chronal increments is well-demonstrated for elastography (b) and possibly for progressive MRC scores (not illustrated; formal publication pending); signifying further groups in whom clinical trials should be encouraged. Conversely, asymptomatic patients with small duct disease, as well as those with classical PSC achieving persistently low/normal liver biochemistry who maintain stable fibrotic indices in the absence of cholangiographic¬¬ progression, likely herald a more consistent low-risk profile (c); albeit with need for longitudinal appraisal (d) given that early predictive models of disease progression are not yet available. Indeed, regular risk-assessment of malignant complications is critical to ensure long-term patient safety, for no early or robust predictors of future CCA currently exist. To this effect, a position for even those in the lower risk category (with large duct disease) to be considered for clinical trials specifically targeted at reducing CCA incidence can also be argued (open arrows), whilst accepting the strong probability that other PSC-related clinical events develop at a low incidence. The optimum frequency of routine radiological surveillance is often debated (e), with no evidence-based guidance in this regard. A suggested policy of 12-monthly (detection of gallbladder polyps), or 6-monthly in cirrhotic patients (hepatocellular carcinoma surveillance) is proposed in keeping with current

guidelines.
124x114mm (300 x 300 DPI)

**Supplementary Table 1: GRADE criteria for assessment of prognosis (78)**

| Quality level | Description | GRADE Score |
|---|---|---|
| High | We are very confident that the true prognosis (probability of future events) lies close to that of the estimate | A |
| Intermediate | We are moderately confident that the true prognosis (probability of future events) is likely to be close to the estimate, but there is a possibility that it is substantially different | B |
| Low | Our confidence in the estimate is limited: the true prognosis (probability of future events) may be substantially different from the estimate | C |
| Very low | We have very little confidence in the estimate: the true prognosis (probability of future events) is likely to be substantially different from the estimate | D |