LIVER

Goals of Treatment for Improved Survival in Primary Biliary Cholangitis: Treatment Target Should Be Bilirubin Within the Normal Range and Normalization of Alkaline Phosphatase

Carla F. Murillo Perez, MSc[1,2], Maren H. Harms, MD[2], Keith D. Lindor, MD[3], Henk R. van Buuren, MD[2], Gideon M. Hirschfield, MD[1], Christophe Corpechot, MD[4], Adriaan J. van der Meer, MD[2], Jordan J. Feld, MD[1], Aliya Gulamhusein, MD[1], Willem J. Lammers, MD[2], Cyriel Y. Ponsioen, MD[5], Marco Carbone, MD[6], Andrew L. Mason, MD[7], Marlyn J. Mayo, MD[8], Pietro Invernizzi, MD[6], Pier Maria Battezzati, MD[9], Annarosa Floreani, MD[10], Ana Lleo, MD[11], Frederik Nevens, MD[12], Kris V. Kowdley, MD[13], Tony Bruns, MD[14,15], George N. Dalekos, MD[16], Nikolaos K. Gatselis, MD[16], Douglas Thorburn, MD[17], Palak J. Trivedi, MD[18], Xavier Verhelst, MD[19], Albert Parés, MD[20], Harry L.A. Janssen, MD[1] and Bettina E. Hansen, PhD[1,21] on behalf of the GLOBAL PBC Study Group

INTRODUCTION: **In primary biliary cholangitis (PBC), bilirubin and alkaline phosphatase (ALP) are widely established as independent predictors of prognosis. Current treatment goals do not aim for normalization of surrogate markers because their association with survival has not been defined.**

METHODS: **The patient cohort from the GLOBAL PBC Study Group was used, comprising of long-term follow-up data from European and North American centers. Ursodeoxycholic acid-treated and untreated patients with bilirubin levels ≤1 × upper limit of normal (ULN) at baseline or 1 year were included. The association of normal ALP with transplant-free survival was assessed in a subgroup with ALP ≤1.67 × ULN at 1 year. Optimal thresholds of bilirubin and ALP to predict liver transplantation (LT) or death were evaluated.**

RESULTS: **There were 2,281 patients included in the time zero cohort and 2,555 patients in the 1-year cohort. The bilirubin threshold with the highest ability to predict LT or death at 1 year was 0.6 × ULN (hazard ratio 2.12, 95% CI 1.69–2.66, $P < 0.001$). The 10-year survival rates of patients with bilirubin ≤0.6 × ULN and >0.6 × ULN were 91.3% and 79.2%, respectively ($P < 0.001$). The risk for LT or death was stable below the bilirubin levels of 0.6 × ULN, yet increased beyond this threshold. Ursodeoxycholic acid-induced reduction in bilirubin below this threshold was associated with an 11% improvement in 10-year survival. Furthermore, ALP normalization was optimal, with 10-year survival rates of 93.2% in patients with ALP ≤ 1 × ULN and 86.1% in those with ALP 1.0–1.67 × ULN.**

DISCUSSION: **Attaining bilirubin levels ≤0.6 × ULN or normal ALP are associated with the lowest risk for LT or death in patients with PBC. This has important implications for treatment targets.**

SUPPLEMENTARY MATERIAL accompanies this paper at http://links.lww.com/AJG/B408

Am J Gastroenterol 2020;115:1066–1074. https://doi.org/10.14309/ajg.0000000000000557

[1]Toronto Centre for Liver Disease, Toronto General Hospital, University Health Network, Toronto, Ontario, Canada; [2]Gastroenterology and Hepatology, Erasmus University Medical Center, Rotterdam, Netherlands; [3]College of Health Solutions, Arizona State University, Phoenix, Arizona; [4]Centre de Référence des Maladies Inflammatoires des VoiesBiliaires, Hôpital Saint-Antoine, Paris, France; [5]Department of Gastroenterology and Hepatology, Academic Medical Center, Amsterdam, Netherlands; [6]Division of Gastroenterology and Center for Autoimmune Liver Diseases, Department of Medicine and Surgery, University of Milan-Bicocca, Milan, Italy; [7]Divison of Gastroenterology and Hepatology, University of Alberta, Edmonton, Alberta, Canada; [8]Digestive and Liver Diseases, UT Southwestern Medical Center, Dallas, Texas; [9]Department of Health Sciences, Università degli Studi di Milano, Milan, Italy; [10]Department of Surgery, Oncology and Gastroenterology, University of Padua, Padua, Italy; [11]Division of Internal Medicine and Hepatology, Humanitas Clinical Research Center IRCSS, Humanitas University, Rozzano (Milan), Italy; [12]Department of Hepatology, University Hospitals Leuven, KU Leuven, Leuven, Belgium; [13]Liver Care Network, Swedish Medical Center, Seattle, Washington; [14]Department of Internal Medicine IV, Jena University Hospital, Friedrich Schiller University, Jena, Germany; [15]Department of Internal Medicine III, University Hospital RWTH Aachen, Aachen, Germany; [16]Department of Medicine and Research Laboratory of Internal Medicine, University of Thessaly, Larissa, Greece; [17]The Sheila Sherlock Liver Centre, The Royal Free Hospital, London, United Kingdom; [18]National Institute for Health Research Birmingham Biomedical Research Centre and Centre for Liver Research, University of Birmingham, Birmingham, United Kingdom; [19]Department of Gastroenterology and Hepatology, Ghent University Hospital, Ghent, Belgium; [20]Liver Unit, Hospital Clínic, CIBERehd, IDIBAPS, University of Barcelona, Barcelona, Spain; [21]Institute of Health Policy, Management and Evaluation, University of Toronto, Toronto, Ontario, Canada. **Correspondence:** Bettina E. Hansen, MSc, PhD. E-mail: Bettina.hansen@utoronto.ca.
Received September 10, 2019; accepted January 7, 2020; published online February 20, 2020

INTRODUCTION

Primary biliary cholangitis (PBC) is an autoimmune cholestatic liver disease that is characterized by chronic nonsuppurative inflammation of the small intrahepatic bile ducts (1). The disease usually has a slow progressive course, which may eventually lead to cirrhosis and ultimately liver failure or premature death in the absence of liver transplantation (LT). However, the prolonged number of years it may take for patients to develop such clinical outcomes poses a significant obstacle in randomized controlled trials that aim to evaluate the clinical benefit of therapeutic interventions. Owing to these feasibility concerns, various surrogate markers have been evaluated for their prognostic value on clinical outcomes (2). Such surrogate markers can allow risk stratification of patients without the need for an extended follow-up and can be implemented by healthcare providers or in clinical trials to promptly assess the need and benefit of a therapeutic agent.

It is established that bilirubin is an independent predictor of prognosis in both ursodeoxycholic acid (UDCA)-treated and untreated patients with PBC (2–4). The normalization of bilirubin prompted by UDCA has been associated with improved transplant-free survival (4). Furthermore, bilirubin and alkaline phosphatase (ALP) have been established as surrogate endpoints that are "reasonably likely to predict clinical benefit," and the widely accepted thresholds are 1× the upper limit of normal (ULN) and 1.67 × ULN, respectively. Normal bilirubin is also a component of multiple response criteria, such as the Rotterdam, Paris-I, and Paris-II criteria (5–7). Abnormal bilirubin levels are observed during later stages of PBC and are indicative of impaired liver function (8). Over the past decades, however, there has been an increase in the proportion of patients who present with normal bilirubin levels, and this group now represents most patients with PBC (9). Because bilirubin is usually not elevated above the ULN until later stages of the disease, it is considered to be an inadequate marker for risk stratification in early stage PBC. The prognostic value of bilirubin and ALP below the ULN has not been previously assessed. We sought to evaluate whether bilirubin or ALP levels within the normal range (≤1 × ULN) are associated with survival in patients with PBC to optimize treatment goals and the number of patients who may benefit from second-line therapy.

PATIENTS AND METHODS

Population and study design

We used the GLOBAL PBC Study Group database to study the predictive value of normal bilirubin for survival in PBC. The Global PBC Study Group is an international collaboration of 17 centers across Europe and North America that includes long-term follow-up data of patients with PBC. During the time frame analyzed, no patients were on obeticholic acid. We included UDCA-treated and untreated patients diagnosed with PBC according to the internationally accepted guidelines and whose bilirubin levels were normal (≤1 × ULN as defined by each local center) at time zero or 1 year (8,10,11). Those with short follow-up (<6 months for time zero cohort; <1 year for 1-year cohort), UDCA discontinuation, unknown clinical event dates, autoimmune hepatitis overlap, or other concomitant liver diseases were excluded from the study. Patients were allocated to 2 independent cohorts based on the time point(s) at which their bilirubin levels were normal (time zero and 1 year) in which inclusion is not mutually exclusive. This study was conducted in accordance with the 1975 Declaration of Helsinki. The protocol was approved by the institutional research board at all participating centers as per local regulations.

Data collection

Time zero (study entry) is defined as the date UDCA was initiated in treated patients and the date of the first visit in untreated patients. At study entry, the following data were available: sex, age at diagnosis, antimitochondrial antibody serological status, liver histology, biochemical disease stage (Rotterdam criteria (5)), and UDCA therapy. Three biochemical disease stages, mild, moderately advanced, and advanced, were defined by normal bilirubin and albumin, abnormal bilirubin or albumin, and abnormal bilirubin and albumin, respectively. As per standard of care, the following laboratory parameters were collected every 6–12 months: total bilirubin, ALP, albumin, aspartate aminotransferase, alanine aminotransferase, and platelet count (8,11). Histological data obtained from liver biopsies were staged according to the criteria of Ludwig et al. and Scheuer (12,13).

Statistical analysis

The primary endpoint was a composite of LT and all-cause mortality. Survival was defined as an absence in LT and all-cause mortality. Patients without an event at the end of follow-up or who were lost to follow-up were censored at their last visit. The survival rates across quartiles corresponding to each cohort were estimated with a Kaplan-Meier curve and compared with a logrank test. Multivariable Cox proportional hazards' regression (hazard ratio [HR] with 95% CI) analyses were performed to adjust for potential confounding variables.

To test the hypothesis of a threshold and to determine the optimal threshold for bilirubin within the normal range, 2 approaches were followed: (i) bilirubin at baseline and 1 year were dichotomized according to various thresholds ranging from 0.3 to 0.9 × ULN in 0.01 increments. Multivariable Cox proportional hazards' regression analyses were used to estimate the risk for LT or death associated with each threshold. The C-statistic was calculated to evaluate the performance of each threshold in predicting survival and the threshold with the best performance was determined by the highest C-statistic. (ii) To assess bilirubin on a continuous spectrum and test the hypothesis that the risk for LT or death increases at the predetermined bilirubin threshold, bilirubin was analyzed as a restricted cubic spline function with 4 knots. Patients with abnormal bilirubin were included to illustrate their risk for a poor prognosis relative to those with normal bilirubin. The restricted spline function was repeated with crude bilirubin levels (mg/dL).

Sensitivity analyses of the predetermined bilirubin threshold by multivariable Cox regression were performed in subgroups: bilirubin ULN (75th percentile: 1.2 mg/dL [21.0 μmol/L]), age at study entry (≤45, 46–55, 56–65, and >65 years), sex, UDCA treatment, histologic stage (I/II and III/IV), and ALP (≤1.67 × ULN and >1.67 × ULN). Of note, sensitivity analyses according to histologic stage were conducted in a subset of patients with available histologic stage at baseline. Furthermore, sensitivity analyses were performed for bilirubin at 2–5 years after the start of follow-up.

Kaplan-Meier analyses were conducted to illustrate the survival rates associated with bilirubin at baseline and 1 year (normal bilirubin [≤/> the threshold] and abnormal bilirubin [reference purposes]). The distribution of the clinical events (LT, liver-related death, or liver-unrelated death) at 10 years within each bilirubin group was also evaluated. Of interest was the impact of bilirubin change in UDCA-treated patients with normal bilirubin at baseline, which was assessed after 1 year.

Table 1. Characteristics of patients with PBC in each normal bilirubin cohort

Parameter	Time zero cohort[a,b] (n = 2,281)	1-year cohort (n = 2,555)
Follow-up time, y, median (IQR)	7.9 (4.3–12.7)	7.3 (3.7–11.5)
Age at study entry, mean ± SD	55.3 ± 12.0	54.6 ± 11.8
Female, no. (%)	2,086 (91.5)	2,354 (92.1)
AMA-positive, no. (%)	2,036/2,222 (91.6)	2,273/2,485 (91.5)
Year of diagnosis, median (range)[c]	1998 (1961–2014)	1997 (1961–2013)
UDCA-treated, no. (%)	1,979/2,223 (89.0)	2,345/2,523 (92.9)
Laboratory parameters, median (IQR)[d]		
Total bilirubin, ×ULN	0.53 (0.40–0.70)	0.50 (0.38–0.67)
ALP, ×ULN	1.99 (1.27–3.32)	1.26 (0.88–1.96)
Albumin, ×LLN	1.17 (1.09–1.26)	1.17 (1.09–1.26)
AST, ×ULN	1.30 (0.93–1.93)	0.87 (0.65–1.20)
ALT, ×ULN	1.51 (0.98–2.35)	0.83 (0.58–1.33)
Platelet count, $\times 10^9$/L	255 (207–308)	250 (202–304)
Bilirubin ULN (mg/dL), median (IQR)[e]	1.1 (1.0–1.2)	1.17 (1.0–1.2)

AMA, antimitochondrial antibody; ALP, alkaline phosphatase; ALT, alanine aminotransferase; AST, aspartate aminotransferase; IQR, interquartile range; LLN, lower limit of normal; PBC, primary biliary cholangitis; UDCA, ursodeoxycholic acid; ULN, upper limit of normal.
[a]Histological disease stage at study entry available for 898 patients (39.4%)—stage I/II: 686 patients (76.4%), stage III/IV: 212 patients (23.6%).
[b]Biochemical disease stage at study entry available for 1,809 patients (79.3%)—mild: 1,670 (92.3%), moderate: 139 (7.7%), and advanced: 0 patients.
[c]Year of diagnosis was available for 2,279 in the time zero cohort and 2,554 patients in the 1-year cohort.
[d]Laboratory parameters other than bilirubin were not available for all patients:
Time zero cohort: ALP (n = 2,119), albumin (n = 1,809), AST (n = 2,023), ALT (n = 1937), platelet count (n = 1,627).
1-year cohort: ALP (n = 2,413), albumin (n = 1,678), AST (n = 2,138), ALT (n = 2,142), and platelet count (n = 1,201).
[e]The upper limit of normal for bilirubin was variable per center.

The pattern of bilirubin (mean and 95% CI) over the first 5 years was evaluated in patients with normal bilirubin at time zero and stratified on whether they experienced a late clinical event (LT or death from 5 to 10 years) or no clinical event in the first 10 years of follow-up.

The association of ALP with survival was assessed in a subgroup of UDCA-treated patients with ALP $\leq 1.67 \times$ ULN at 1 year by dichotomization according to thresholds ranging from 0.5 to 1.6 × ULN. ALP was assessed at 1 year to maximize the number of patients and emphasize the development of thresholds as treatment targets. The risk for LT and death for each threshold was estimated with Cox regression analyses. Within the normal range of ALP, the survival associated with quartiles was also emphasized.

Multiple imputation was conducted by using the Markov chain Monte Carlo method for missing data with Statistical Analysis System (SAS) version 9.4 (SAS Institute, Cary, NC). Ten imputed data sets for missing biochemical values were generated to reduce sampling variability of the laboratory results (14). Imputation was performed based on the assumption that data were missing at random, in which variables predicting outcomes and outcomes themselves were included in the imputation model. Rubin's rules were used to estimate the parameter and SE (15,16).

All multivariable analyses were adjusted for age at study entry, sex, year of diagnosis, UDCA therapy, ALP, and geographical region. Biochemical markers that were not normally distributed were log transformed. A P value less than 0.05 was considered statistically significant. All analyses were 2-sided and were performed using IBM SPSS Statistics for Windows, version 25.0 (IBM Corp., Armonk, NY) and SAS version 9.4 (SAS Institute).

RESULTS

Study population characteristics

At baseline, there were 2,545 adult PBC patients with normal bilirubin, of whom 264 met exclusion criteria (54 for autoimmune hepatitis overlap, 76 for UDCA discontinuation, and 134 for short follow-up). Similarly, from 2,797 adult patients with PBC with normal bilirubin at 1 year, 242 were excluded (64 for autoimmune hepatitis overlap, 47 for UDCA discontinuation, and 131 for short follow-up). Overall, a total of 3,059 patients with normal bilirubin at baseline or one year were included and 2 cohorts were constructed: time zero cohort (n = 2,281) and 1-year cohort (n = 2,555). An overlap of 1,777 patients exists between these cohorts. There were 297 and 344 primary endpoints according to each respective cohort. Patient characteristics per cohort are presented in Table 1.

Normal bilirubin quartiles are associated with survival

In patients with normal bilirubin at time zero, the cumulative 10-year survival rate decreased with higher bilirubin quartiles and was 93.3%, 89.9%, 87.7%, and 81.3% from quartiles 1–4 (Q1–Q4), respectively (Figure 1). In pairwise comparisons, Q4 was significantly different from Q1–Q3 (all $P < 0.005$). In addition, Q1 was significantly different from Q3 ($P = 0.04$). Similar results were obtained in the Kaplan-Meier analysis of the 1-year cohort, in which the 10-year survival rates with increasing bilirubin quartiles were 92.0%, 92.3%, 86.1%, and 78.2%. Q3 and Q4 were significantly different from one another and from the remaining quartiles (all $P < 0.01$). In multivariable Cox regression analyses of the time zero cohort, the risk for LT or death increased with higher bilirubin quartiles: Q1 (reference), Q2 (HR 1.12, 95% CI 0.73–1.72, $P = 0.61$), Q3 (HR 1.34, 95% CI 0.89–2.01, $P = 0.16$),

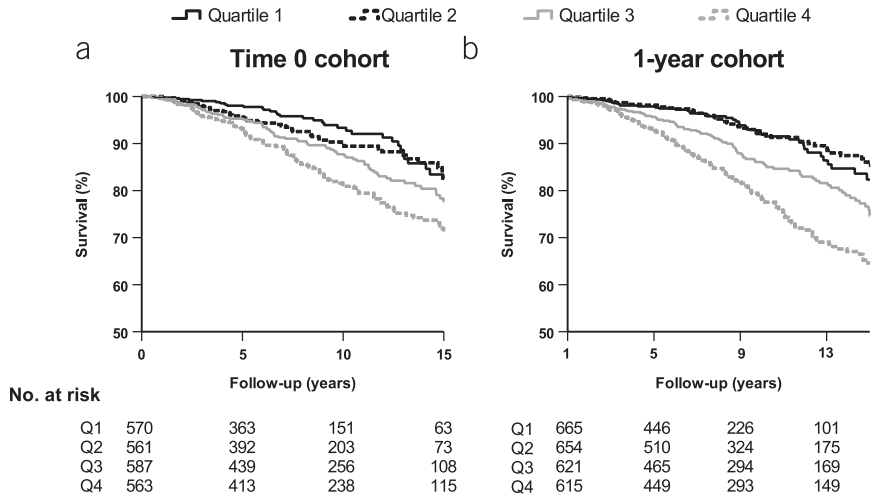

Figure 1. Survival estimates of bilirubin quartiles in patients with normal bilirubin. Kaplan-Meier estimates of bilirubin quartiles from (**a**) time zero and (**b**) 1 year.

and Q4 (HR 1.83, 95% CI 1.24–2.71, $P = 0.003$). A similar trend was observed in the 1-year cohort: Q1 (reference), Q2 (HR 0.97, 95% CI 0.65–1.45, $P = 0.88$), Q3 (HR 1.46, 95% CI 1.02–2.10, $P = 0.04$), and Q4 (HR 2.20, 95% CI 1.56–3.10, $P < 0.001$).

Bilirubin threshold within the normal range to predict survival

On exploration of the optimal threshold of bilirubin within the normal range at 1 year, all bilirubin thresholds (0.3–0.9 × ULN) were significant predictors of survival in that patients with bilirubin above each threshold had an increased risk for LT or death (see Table 1, Supplementary Digital Content 1, http://links.lww.com/AJG/B408). The bilirubin threshold at 1 year with the highest ability to predict LT or death was 0.6 × ULN (C-statistic 0.7429, 95% CI 0.7144–0.7713). The 10-year survival of patients with normal bilirubin ≤0.6 × ULN, normal bilirubin >0.6×ULN, and abnormal bilirubin at 1 year were 91.3%, 79.2%, and 37.3%, respectively ($P < 0.001$) (Figure 2a). At baseline, the 10-year survival rates were 91.7%, 85.6%, and 49.5% ($P < 0.001$). Discordant survival rates were also observed at 15 years (see Table 2, Supplementary Digital Content 1, http://links.lww.com/AJG/B408). We evaluated the distribution of clinical events from the 10-year survival rates associated with each bilirubin group. Clinical events in patients with bilirubin from 0.6 to 1.0 × ULN were characterized by a significantly increased proportion of LT and liver-related deaths, alongside a decreased proportion of liver-unrelated deaths as compared to patients with bilirubin ≤0.6 × ULN ($P < 0.001$) (see Figure 1, Supplementary Digital Content 1, http://links.lww.com/AJG/B408). Of 1,934 UDCA-treated patients with normal baseline bilirubin, 1-year bilirubin was available for 1,644 (85%). A UDCA-induced reduction in bilirubin (≤0.6 × ULN) at 1 year was associated with a 10.5% improvement in 10-year survival (93.7% vs 83.2%) and 17.1% in 15-year survival (86.5% vs 69.4%) as compared to bilirubin that remained above the threshold (Figure 2b).

The threshold was evaluated in various subgroups of patients who had normal bilirubin at 1 year, all of which confirmed that patients with bilirubin >0.6 × ULN have an increased risk for LT or death (Figure 3). Surpassing the bilirubin threshold of 0.6 × ULN was associated with an increased risk for LT or death according to histologic stage but demonstrated a trend for significance (stage I/II: $P = 0.05$; stage III/IV:$P = .07$) that may be because of the limited availability of baseline histologic stages in our cohort. The association remained when patients in whom the ULN was defined as ≥1.2 mg/dL (21.0 μmol/L) were excluded from the analyses (HR 2.10, 95% CI 1.54–2.85, $P < 0.001$). In addition, the threshold was a significant factor in a subgroup (n = 495) with known UDCA dosage ≥13 mg/kg (HR 1.85, 95% CI 1.02–3.34, $P = 0.04$). There were no significant interactions between the bilirubin threshold and the variables explored in subgroup analyses (Figure 3).

The risk for LT or death increases at bilirubin levels of 0.6 × ULN

We assessed bilirubin on a continuous spectrum with a restricted spline function to evaluate whether the predetermined threshold is the point at which the HR for LT or death increases. The reference in each cohort was the predetermined threshold of 0.6 × ULN. In both cohorts, the risk for LT or death remained stable below 0.6 × ULN (Figure 4). However, beyond this threshold, a linear relationship was observed between bilirubin and the risk for LT or death that continued past the normal range. The test for curvature that establishes a significant deviation from a linear relationship was significantly different for the time zero ($P = 0.03$) and 1-year cohorts ($P = 0.05$). As sensitivity analyses, the restricted spline function was repeated using crude bilirubin levels (mg/dL) and with normal bilirubin levels from 2 to 5 years (see Figures 2 and 3, Supplementary Digital Content 1, http://links.lww.com/AJG/B408).

Patients who remain below 0.6 × ULN over time have good long-term prognosis

To assess how the trajectory of bilirubin over time may be related with the development of LT or death, bilirubin was evaluated over 5 years in patients with normal bilirubin at time zero. Patients were stratified according to whether they developed a late clinical event from 5 to 10 years (n = 103) or did not develop a clinical event in the first 10 years of follow-up (n = 848). Patients who had no clinical event within 10 years presented with a mean bilirubin level of 0.55 × ULN (95% CI 0.54–0.56) that remained stable (below 0.6 × ULN) over 5 years (see Figure 4, Supplementary Digital Content 1, http://links.lww.com/AJG/B408). By contrast, patients who reached a clinical endpoint presented with slightly

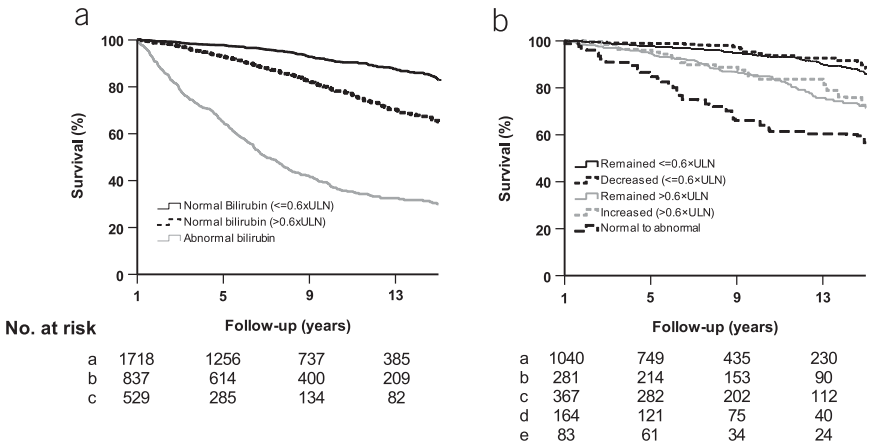

Figure 2. Survival estimates in patients with normal bilirubin (stratified by 0.6 × ULN threshold) and abnormal bilirubin. (**a**) Kaplan-Meier estimates of survival rates in patients with normal bilirubin (stratified by 0.6 × ULN threshold) and abnormal bilirubin at 1 year. (**b**) Additional analysis of the survival rates in ursodeoxycholic acid-treated patients with normal bilirubin levels at baseline and stratified according to the change in bilirubin from baseline to 1 year. ULN, upper limit of normal.

higher mean bilirubin levels (0.61 × ULN, 95% CI 0.57–0.65, P = 0.01) and exhibited a gradual increase within the normal range that precluded the occurrence LT or death.

ALP normalization is associated with improved survival

In a subgroup of patients with ALP ≤1.67 × ULN from the normal bilirubin cohort at 1 year (n = 1,523), the optimal ALP threshold was 1.0 × ULN (C-statistic 0.7552, 95% CI 0.7151–0.7953). The HR for LT or death was 1.44 in those with ALP >1 × ULN (95% CI 1.04–2.00, P = 0.03). Patients with ALP ≤1 × ULN had the highest survival rate (93.2% at 10 years and 84.1% at 15 years), which was significantly different from those with ALP between 1.0 and 1.67 × ULN (86.1% at 10 years and 76.4% at 15 years) and ALP >1.67 × ULN (85.4% at 10 years and 73.8% at 15 years), P < 0.005 (see Figure 5, Supplementary Digital Content 1, http://links.lww.com/AJG/B408). Survival between the group with ALP 1.0–1.67 × ULN and that with ALP >1.67 × ULN was similar (P = 0.64). Quartiles within the normal range of ALP (n = 773) were not associated with survival.

ALP normalization and bilirubin levels below 0.6 × ULN

Implementing both ALP and bilirubin thresholds established, the prognosis of patients with bilirubin >0.6 × ULN was dependent on ALP normalization (see Figure 6, Supplementary Digital Content 1, http://links.lww.com/AJG/B408). Given normal ALP levels, their survival rates were similar to those with bilirubin ≤0.6 × ULN; however, if ALP was 1.0–1.67 × ULN, their survival was diminished to 74.2% at 10 years and 63.4% at 15 years (P < 0.001 compared with the remaining groups).

DISCUSSION

This study reports that bilirubin levels within the normal range are associated with the risk for LT or death in patients with PBC. We demonstrated that bilirubin levels ≤0.6 × ULN at baseline and 1 year were associated with a decreased risk for LT or death as compared to patients with bilirubin above this threshold. Although the risk for LT or death was stable when bilirubin levels were below 0.6 × ULN, beyond this threshold, a positive linear relationship was observed between bilirubin and the risk for a clinical event. In patients with already normal bilirubin at baseline, a reduction below this threshold after 1 year of UDCA therapy was associated with an 11% improvement in 10-year survival. These results were confirmed in several subgroups of patients. These findings suggest that the interpretation of not being at risk if bilirubin is within the normal range needs to be revised. In addition, ALP normalization was also associated with prolonged survival. This might have implications in the number of patients eligible for inclusion in clinical trials for novel second-line therapies because ALP >1.67 × ULN/abnormal bilirubin are common eligibility requirements. These results are in line with the change in criteria for disease remission in autoimmune hepatitis. Previously, one of the requirements for disease remission was transaminase levels below twice the ULN. However, the definition for disease remission now includes normal transaminases (17,18) because failure to normalize these liver enzymes is associated with an increased risk for relapse after treatment withdrawal (19,20), histological worsening or progression to cirrhosis (19,21), and poor clinical outcomes (21–23).

Although the ULN of bilirubin is reported as the most predictive for survival in patients with PBC (2), we found that the risk for LT or death is already increased when bilirubin is above 0.6 × ULN. The optimum bilirubin cutoff associated with survival has been previously suggested to be lower than the ULN (24). The current ULN of bilirubin represents the 97.5 percentile cutoff in the general population, yet this may not be the best approach to determine an optimal threshold because levels below this threshold are not reflective of an absence of increased risk (25). This may partly be explained by the high percentage (3%–10%) of individuals with Gilbert's syndrome in the general population (26). In addition, the current ULN of bilirubin may be suboptimal for risk stratification in PBC because of the female predominance of the disease, whereas sex differences in bilirubin are present in the general population (27). An American study based on the Third National Health and Nutrition Examination Survey assessed the serum bilirubin levels in 16,865 adults from the general population and reported that the mean serum bilirubin levels are significantly lower in women (0.52 mg/dL ± 0.003) than in men (0.72 mg/dL ± 0.004) (25). Consequently, the 97.5 percentile cutoff was 0.5 mg/dL higher in men. Other studies have reported similar sex differences in bilirubin levels in the general population (28,29). Thus, the overall ULN of bilirubin may be skewed higher because of the inclusion of both men and women.

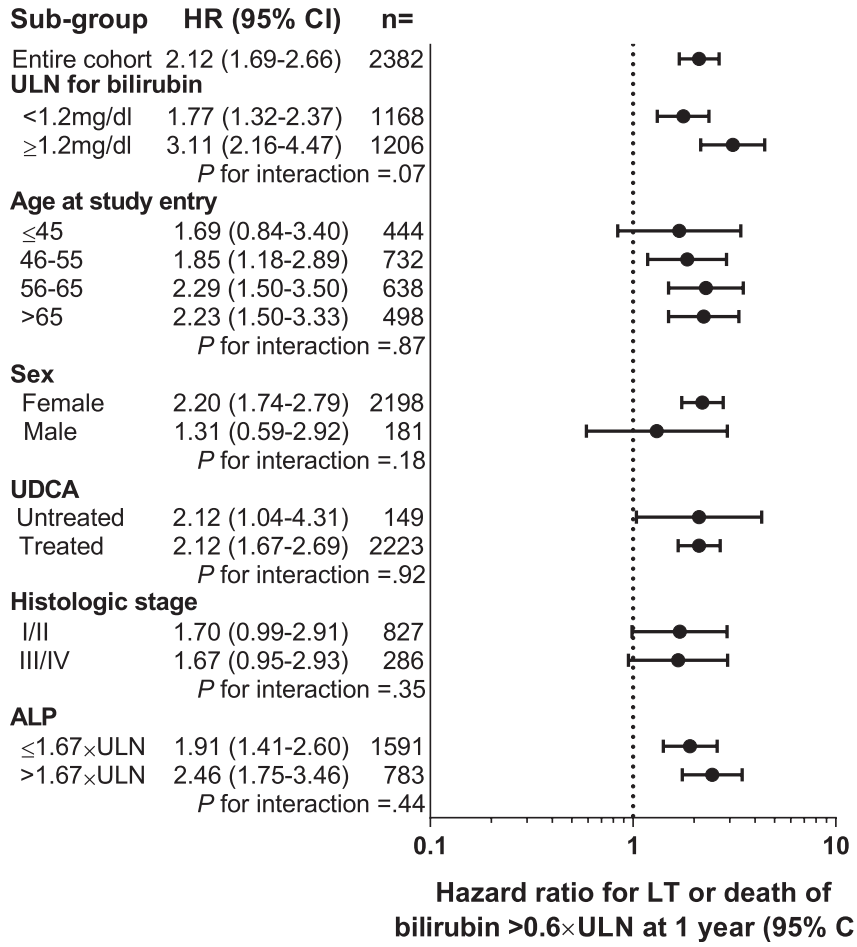

Figure 3. Subgroup analyses based on the bilirubin threshold of 0.6 × ULN in patients with normal bilirubin at 1 year. HR for LT or death (95% CI) obtained from multivariable Cox proportional hazards regression analyses in patients with normal bilirubin in various subgroups. The hazard ratios correspond to bilirubin levels >0.6 × ULN (vs bilirubin ≤0.6 × ULN). The *P* value for interaction corresponds to an interaction between the bilirubin threshold and associated variable. ALP, alkaline phosphatase; CI, confidence interval; HR, hazard ratio; LT, liver transplantation; UDCA, ursodeoxycholic acid; and ULN, upper limit of normal.

These considerations suggest that the ULN for bilirubin may need to be stratified by sex, as has been previously implemented for aspartate aminotransferase (30,31).

We found that the predictive value of the bilirubin threshold of 0.6 × ULN was irrespective of age, UDCA treatment, and ALP levels. Importantly, it remained significantly predictive at various independent time points. Furthermore, in UDCA-treated patients with normal bilirubin levels above 0.6 × ULN at treatment initiation, a reduction below 0.6 × ULN was associated with prolonged survival. This suggests that besides the predictive value of bilirubin within the normal range, a treatment-induced reduction of bilirubin within the current normal range is beneficial for long-term prognosis, which could have important implications for current patient care, but also for the design and interpretation of future clinical trials of potential second-line therapies in PBC. Although recent clinical trials have often included normalization of bilirubin as a primary endpoint, it might be preferable to aim for lower bilirubin levels (32,33).

The pattern of bilirubin over time may also be relevant because there was an overall increase of 0.47 × ULN in the mean bilirubin after 5 years in patients who reached a clinical endpoint after an extended follow-up. Although rapid increases in bilirubin have been shown to preclude death in untreated patients, these results suggest that there is an association between the trajectory of bilirubin and clinical outcomes, even within the normal range (3). The finding that mean bilirubin levels of patients who did not experience a clinical event remained below 0.6 × ULN over time further supports an incentive to aim for bilirubin levels below our proposed threshold of 0.6 × ULN and emphasizes the importance of continuous bilirubin evaluation even in early stage disease.

A robust analysis of the predictive value of bilirubin within the normal range would not be possible without the large number of patients and extended follow-ups available from the Global PBC Study Group cohort. Furthermore, bilirubin was assessed at multiple independent time points to confirm that bilirubin levels obtained during a random follow-up assessment could also be used for risk stratification. Nonetheless, some study limitations should be noted. Whereas total serum bilirubin levels in healthy patients are primarily composed of unconjugated bilirubin, it is predominantly of conjugated form in patients with PBC (34). Therefore, evaluating the influence of solely the conjugated form could be of interest. However, the real-world nature of our cohort only allowed us to evaluate total bilirubin, since independent measurements of the conjugated and unconjugated forms are not part of routine standard of care in most laboratories. The methodological limitations to determining accurate conjugated bilirubin measurements need consideration because direct bilirubin also measures delta bilirubin.

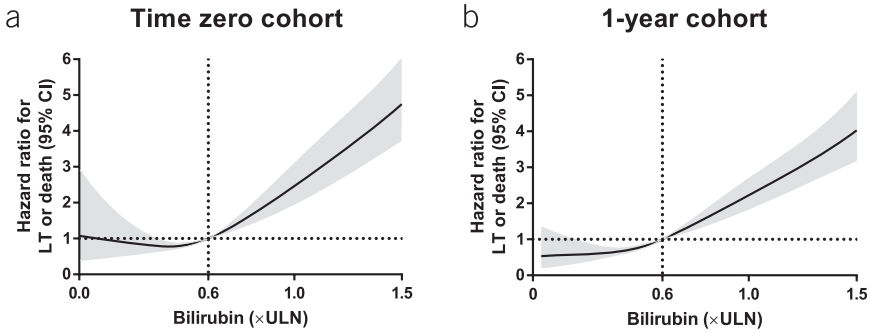

Figure 4. The association between bilirubin levels (×ULN) and risk for LT or death. Hazard ratios and 95% CI were estimated by a restricted cubic spline function in (**a**) the time zero cohort and (**b**) the 1-year cohort. The bilirubin reference in each cohort is 0.6 × ULN. CI, confidence interval; LT, liver transplantation; and ULN, upper limit of normal.

In addition, some patients may also be affected by Gilbert's syndrome and can potentially lead to an underestimation of the association of bilirubin above 0.6 × UN with survival, albeit this is not expected to play a major role because of the large sample size.

Although bilirubin was analyzed based on the ULN defined by local centers, which ranged from 0.6 to 1.7 mg/dL, sensitivity analyses were performed to address this. The analyses with crude bilirubin levels (mg/dL) and the one excluding patients with an ULN above 1.2 mg/dL (21.0 μmol/L) confirmed our initial findings and exclude the possibility that patients with bilirubin levels above 0.6 × ULN have worse survival because of the utilization of high ULNs.

In this multicenter international follow-up study of patients with PBC, bilirubin levels below the current ULN were shown to be predictive of survival and 0.6 × ULN was established as the threshold from which point on the risk for LT or death increases. In addition, treatment-induced reduction of normal bilirubin below 0.6 × ULN was associated with prolonged survival. By contrast, ALP normalization was the established threshold for improved survival. Our proposed thresholds of 0.6 × ULN for bilirubin and normalization for ALP represent a refinement of previous criteria with an aim to optimize survival and identify patients at risk for poor outcome. Moreover, their implementation can broaden the patient population included in intervention studies who may benefit from therapeutic agents.

ACKNOWLEDGEMENTS

This study was performed on behalf of the *Global PBC Study Group* and was supported by unrestricted grants from CymaBay Therapeutics Inc., Intercept Pharmaceuticals, and previously from Zambon Nederland BV, and was funded by the Toronto General & Western Hospital Foundation (a not-for-profit organization) in Toronto, Canada, and the Foundation for Liver and Gastrointestinal Research (a not-for-profit organization) in Rotterdam, the Netherlands.

We would like to acknowledge the participating centers in this study: Toronto Centre for Liver Disease, Toronto General Hospital; Erasmus University Medical Center; Arizona State University; University of Birmingham; Hôpital Saint-Antoine; Academic Medical Center; University of Milan-Bicocca; University of Alberta; UT Southwestern Medical Center; Università degli Studi di Milano; University of Padua; Humanitas University; University Hospitals Leuven; Swedish Medical Center; University of Jena; University of Thessaly; The Royal Free Hospital; Ghent University Hospital; and University of Barcelona.

CONFLICTS OF INTEREST

Guarantor of the article: Carla F. Murillo Perez, MSc and Bettina E. Hansen, PhD had full access to all data in the study and take responsibility for the integrity of the data and the accuracy of data analyses.

Specific author contributions: Study concept and design: All authors. Acquisition of data: All authors. Analysis and interpretation of data: C.F.M.P., M.H.H., and B.E.H. Drafting of the manuscript: C.F.M.P., M.H.H., K.D.L., H.R.v.B., G.M.H., C.C., A.J.v.d.M., J.F., H.L.A.J., and B.E.H. Critical revision of the manuscript for important intellectual content: C.F.M.P., M.H.H., K.D.L., H.R.v.B., G.M.H., C.C., A.J.v.d.M., J.J.F., H.L.A.J., and B.E.H. Statistical analysis: C.F.M.P., and B.E.H. Obtained funding: H.R.v.B., and B.E.H. Study supervision: H.R.v.B., C.C., D.T., H.L.A.J., K.D.L., G.M.H., A.P., A.F., A.L., M.J.M., P.I., P.M.B., F.N., C.Y.P., A.L.M., K.V.K., and B.E.H.

Financial support: This investigator-initiated study was supported by unrestricted grants from CymaBay Therapeutics Inc., Intercept Pharmaceuticals, and previously from Zambon Nederland BV, and was funded by the Toronto General & Western Hospital Foundation (a not-for-profit foundation) in Toronto, Canada, and the Foundation for Liver and Gastrointestinal Research (a not-for-profit foundation in Rotterdam, the Netherlands. The supporting parties had no influence on the study design, data collection and analyses, writing of the manuscript, or on the decision to submit the manuscript for publication.

Potential competing interests: The following authors declared that they have no conflicts of interest: P.I., P.M.B., M.C., G.N.D., N.G., X.V.. C.F.M.P. reports speaker fee from Merck & Co. M.H.H. reports speaker fees from Zambon Nederland B.V. K.D.L. reports that he is an unpaid advisor for Intercept Pharmaceuticals and Shire. H.R.v.B. is a consultant for Intercept Pharma Benelux and received unrestricted research grants from Intercept Pharmaceuticals and from Zambon Nederland B.V. G.M.H. reports advisory services for Intercept Pharmaceuticals, Novartis, CymaBay Therapeutics and GlaxoSmithKline Pharmaceuticals. C.C. reports receiving consulting fees from Intercept Pharmaceuticals and Inventiva, grant support from Arrow and Intercept Pharmaceuticals, and fees for teaching from GlaxoSmithKline. A.J.v.d.M. reports speaker fees from Gilead Sciences, AbbVie Pharmaceuticals and Zambon Nederland B.V., received an unrestricted grant from Gilead Sciences, and reports travel expenses covered by Dr. Falk Pharma. J.JF. received personal fees for scientific consulting from AbbVie Pharmaceuticals , Gilead, ContraVir, Enanta, and Merck; and reports receiving grants for research from AbbVie, Gilead, Janssen, Merck, and Wako. A.G.

reports advisory for Intercept Pharmaceuticals and AbbVie . W.J.L. reports consulting services for Intercept Pharmaceuticals. C.Y.P. has received grant support form Takeda, speaker's fees from AbbVie , Takeda, and Dr Falk Pharma, and served as consultant for Takeda. A.L.M. reports advisory services for Intercept Pharmaceuticals, AbbVie, and Novartis; and research funding resources from the Canadian Institutes of Health Research, Canadian Liver Foundation, American Kennel Club, Intercept Pharmaceuticals Inc., AbbVie and Gilead Sciences. M.J.M. reports being on advisory committees or review panels for GSK; grant/research support from Gilead, CymaBay, Intercept, Mallinckrodt, Novartis, Target, GSK, and GENFIT. A.F. reports consulting activities for Intercept Pharmaceuticals. A.L. has served as a speaker for AbbVie, MSD, Gilead, and Intercept Pharmaceuticals. F.N.: Advisory boards for Astellas, Janssen-Cilag, AbbVie, Gilead, CAF, Intercept, Gore, BMS, Novartis, MSD, Janssen-Cilag, Promethera Biosciences, Ono Pharma, Durect, Roche, Ferring. Research grants from Roche, Ferring, and Novartis. K.V.K. reports personal fees from Gilead Sciences, Intercept Pharmaceuticals and Novartis; and grants from Gilead Sciences and Intercept Pharmaceuticals. T.B. has received honoraria from Intercept Pharmaceuticals, Falk Foundation, AbbVie, and Norgine, and travel expenses from Gilead. D.T. reports consulting activities for Intercept Pharmaceuticals. R.P. received consulting fees from Intercept Pharmaceuticals. P.J.T. receives institutional salary support from the NIHR Birmingham Liver Biomedical Research Centre. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health. P.J.T. receives research grant funding from the Wellcome trust, the Core Digestive Diseases Charity, Intercept Pharmaceuticals and PSC Support. A.P. reports consulting services for Intercept Pharmaceuticals and Novartis Pharma. H.L.A.J. reports grants from and consulting work for AbbVie Pharmaceuticals, Bristol-Myers Squibb, Gilead Sciences, Innogenetics, Merck, Novartis, Roche, Intercept Pharmaceuticals and Janssen. B.E.H. reports grants from Intercept Pharmaceuticals, CymaBay Therapeutics, and Zambon Nederland B.V. and consulting work for Intercept Pharmaceuticals, CymaBay Therapeutics, Albireo AB, and Novartis.

Study Highlights

WHAT IS KNOWN

- ✓ Bilirubin and ALP are surrogate markers of survival in PBC.
- ✓ Bilirubin is believed to be a poor predictor of prognosis in early stages of disease.
- ✓ Current treatment targets include normal bilirubin and ALP below 1.67 × ULN.

WHAT IS NEW HERE

- ✓ Bilirubin below the normal threshold and ALP below 1.67 × ULN were associated with improved survival.
- ✓ Attaining bilirubin levels of 0.6 × ULN and normalization of ALP are ideal.
- ✓ An UDCA-induced decrease in normal bilirubin below 0.6 × ULN was associated with an 11% improvement in 10-year survival.
- ✓ Treatment goals in PBC should emphasize normalization of these surrogate markers.

REFERENCES

1. Kaplan MM, Gershwin ME. Primary biliary cirrhosis. N Engl J Med 2005; 353:1261–73.
2. Lammers WJ, van Buuren HR, Hirschfield GM, et al. Levels of alkaline phosphatase and bilirubin are surrogate end points of outcomes of patients with primary biliary cirrhosis: An international follow-up study. Gastroenterology 2014;147:1338–49.e5.
3. Shapiro JM, Smith H, Schaffner F. Serum bilirubin: A prognostic factor in primary biliary cirrhosis. Gut 1979;20:137–40.
4. Bonnand AM, Heathcote EJ, Lindor KD, et al. Clinical significance of serum bilirubin levels under ursodeoxycholic acid therapy in patients with primary biliary cirrhosis. Hepatology 1999;29:39–43.
5. ter Borg PCJ, Schalm SW, Hansen BE, et al. Prognosis of ursodeoxycholic acid-treated patients with primary biliary cirrhosis. Results of a 10-yr Cohort Study involving 297 patients. Am J Gastroenterol 2006;101:2044–50.
6. Corpechot C, Abenavoli L, Rabahi N, et al. Biochemical response to ursodeoxycholic acid and long-term prognosis in primary biliary cirrhosis. Hepatology 2008;48:871–7.
7. Corpechot C, Chazouillères O, Poupon R. Early primary biliary cirrhosis: Biochemical response to treatment and prediction of long-term outcome. J Hepatol 2011;55:1361–7.
8. Hirschfield GM, Beuers U, Corpechot C, et al. EASL Clinical Practice Guidelines: The diagnosis and management of patients with primary biliary cholangitis. J Hepatol 2017;67:145–72.
9. Murillo Perez CF, Goet JC, Lammers WJ, et al. Milder disease stage in patients with primary biliary cholangitis over a 44-year period: A changing natural history. Hepatology 2018;67:1920–30.
10. Reshetnyak VI. Primary biliary cirrhosis: Clinical and laboratory criteria for its diagnosis. World J Gastroenterol 2015;21:7683–708.
11. Lindor K, Bowlus C, Boyer J, et al. Primary Biliary Cholangitis: 2018 Practice Guidance from the American Association for the Study of Liver Diseases. Hepatology 2019;69:394–419.
12. Ludwig J, Dickson ER, McDonald GSA. Staging of chronic nonsuppurative destructive cholangitis (Syndrome of primary biliary cirrhosis). Virchows Arch A Pathol Anat Histol 1978;379:103–12.
13. Scheuer P. Primary biliary cirrhosis. Proc R Soc Med 1967;60:1257–60.
14. Sterne J, White I, Carlin J, et al. Multiple imputation for missing data in epidemiological and clinical research: Potential and pitfalls. BMJ 2009; 338:b2393.
15. Rubin DB. Multiple imputation after 18+ years. J Am Statitical Assoc 1996;91:473–89.
16. Rubin DB. Multiple Imputation for Nonresponse in Surveys. New York: Wiley; 1987.
17. Manns M, Czaja A, Gorham J, et al. Diagnosis and management of autoimmune hepatitis. Hepatology 2010;51:2193–213.
18. European Association for the Study of the Liver. EASL clinical practice guidelines: Autoimmune hepatitis. J Hepatol 2015;63:971–1004.
19. Verma S, Gunuwan B, Mendler M, et al. Factors predicting relapse and poor outcome in type I autoimmune hepatitis: Role of cirrhosis development, patterns of transaminases during remission and plasma cell activity in the liver biopsy. Am J Gastroenterol 2004;99:1510–6.
20. Montano-Loza A, Carpenter H, Czaja A. Improving the end point of corticosteroid therapy in type 1 autoimmune hepatitis to reduce the frequency of relapse. Am J Gastroenterol 2007;102:1005–12.
21. Muratori L, Muratori P, Lanzoni G, et al. Application of the 2010 American Association for the study of liver diseases criteria of remission to a cohort of Italian patients with autoimmune hepatitis. Hepatology 2010;52:1857.
22. Miyake Y, Iwasaki Y, Terada R, et al. Persistent normalization of serum alanine aminotransferase levels improves the prognosis of type 1 autoimmune hepatitis. J Hepatol 2005;43:951–7.
23. Hoeroldt B, McFarlane E, Dube A, et al. Long-term outcomes of patients with autoimmune hepatitis managed at a nontransplant center. Gastroenterology 2011;140:1980–9.
24. Trivedi PJ, Bruns T, Cheung A, et al. Optimising risk stratification in primary biliary cirrhosis: AST/platelet ratio index predicts outcome independent of ursodeoxycholic acid response. J Hepatol 2014;60:1249–58.
25. Zucker SD, Horn PS, Sherman KE. Serum bilirubin levels in the U.S. population: Gender effect and inverse correlation with colorectal cancer. Hepatology 2004;40:827–35.
26. Bosma PJ, Roy Chowdhury J, Bakker C, et al. The genetic basis of the reduced expression of bilirubin UDP-glucuronosyltransferase 1 in gilbert's syndrome. N Engl J Med 1995;333:1171–5.
27. Podda M, Selmi C, Lleo A, et al. The limitations and hidden gems of the epidemiology of primary biliary cirrhosis. J Autoimmun 2013;46:81–7.

28. White GL, Nelson JA, Pedersen DM, et al. Fasting and gender (and altitude?) influence reference intervals for serum bilirubin in healthy adults. Clin Chem 1981;27:1140–2.
29. Rosenthal P, Pincus M, Fink D. Sex- and age-related differences in bilirubin concentrations in serum. Clin Chem 1984;30:1380–2.
30. Terrault NA, Bzowej NH, Chang K, et al. AASLD guidelines for treatment of chronic hepatitis B. Hepatology 2016;63:261–83.
31. Prati D, Taioli E, Zanella A, et al. Updated definitions of healthy ranges for serum alanine aminotransferase levels. Ann Intern Med 2002;137:1–10.
32. Nevens F, Andreone P, Mazzella G, et al. A placebo-controlled trial of obeticholic acid in primary biliary cholangitis. N Engl J Med 2016;375:631–43.
33. Corpechot C, Chazouillères O, Rousseau A, et al. A placebo-controlled trial of bezafibrate in primary biliary cholangitis. N Engl J Med 2018;378:2171–81.
34. Levitt DG, Levitt MD. Quantitative assessment of the multiple processes responsible for bilirubin homeostasis in health and disease. Clin Exp Gastroenterol 2014;7:307–28.