Gastroenterology 2014;147:1338–1349

CLINICAL—BILIARY

Levels of Alkaline Phosphatase and Bilirubin Are Surrogate End Points of Outcomes of Patients With Primary Biliary Cirrhosis: An International Follow-up Study

Willem J. Lammers,[1] Henk R. van Buuren,[1] Gideon M. Hirschfield,[2] Harry L. A. Janssen,[3] Pietro Invernizzi,[4] Andrew L. Mason,[5] Cyriel Y. Ponsioen,[6] Annarosa Floreani,[7] Christophe Corpechot,[8] Marlyn J. Mayo,[9] Pier M. Battezzati,[10] Albert Parés,[11] Frederik Nevens,[12] Andrew K. Burroughs,[13,†] Kris V. Kowdley,[14] Palak J. Trivedi,[2] Teru Kumagi,[3,15] Angela Cheung,[3] Ana Lleo,[4] Mohamad H. Imam,[16] Kirsten Boonstra,[6] Nora Cazzagon,[7] Irene Franceschet,[7] Raoul Poupon,[8] Llorenç Caballeria,[11] Giulia Pieri,[13] Pushpjeet S. Kanwar,[14] Keith D. Lindor,[16,17] and Bettina E. Hansen,[1] on behalf of the Global PBC Study Group

[1]Department of Gastroenterology and Hepatology, Erasmus University Medical Centre, Rotterdam, The Netherlands; [2]NIHR Biomedical Research Unit and Centre for Liver Research, University of Birmingham, Birmingham, England; [3]Liver Clinic, Toronto Western & General Hospital, University Health Network, Toronto, Ontario, Canada; [4]Liver Unit and Center for Autoimmune Liver Diseases, Humanitas Clinical and Research Center, Rozzano, Italy; [5]Divison of Gastroenterology and Hepatology, University of Alberta, Edmonton, Alberta, Canada; [6]Department of Gastroenterology and Hepatology, Academic Medical Center, Amsterdam, The Netherlands; [7]Department of Surgery, Oncology and Gastroenterology, University of Padua, Padua, Italy; [8]Centre de Référence des Maladies Inflammatoires des VoiesBiliaires, Hôpital Saint-Antoine, APHP, Paris, France; [9]Digestive and Liver Diseases, University of Texas Southwestern Medical Center, Dallas, Texas; [10]Department of Health Sciences, Università degli Studi di Milano, Milan, Italy; [11]Liver Unit, Hospital Clínic, CIBERehd, IDIBAPS, University of Barcelona, Barcelona, Spain; [12]Department of Hepatology, University Hospitals Leuven, KU Leuven, Leuven, Belgium; [13]The Sheila Sherlock Liver Centre, The Royal Free Hospital, London, England; [14]Liver Center of Excellence, Digestive Disease Institute, Virginia Mason Medical Center, Seattle, Washington; [15]Department of Gastroenterology and Metabology, Ehime University Graduate School of Medicine, Ehime, Japan; [16]Department of Gastroenterology and Hepatology, Mayo Clinic, Rochester, Minnesota; and [17]College of Health Solutions, Arizona State University, Phoenix, Arizona

This article has an accompanying continuing medical education activity on page e15. Learning Objective: Upon completion of this CME exam, successful learners will understand the concept and importance of surrogate end points in medicine and be aware of the proposed four-level hierarchy of evidence for validation. Further, they will be able to apply this knowledge to a particular disease, namely primary biliary cirrhosis.

See Covering the Cover synopsis on page 1194.

BACKGROUND & AIMS: Noninvasive surrogate end points of long-term outcomes of patients with primary biliary cirrhosis (PBC) are needed to monitor disease progression and evaluate potential treatments. We performed a meta-analysis of individual patient data from cohort studies to evaluate whether patients' levels of alkaline phosphatase and bilirubin correlate with their outcomes and can be used as surrogate end points. **METHODS:** We performed a meta-analysis of data from 4845 patients included in 15 North American and European long-term follow-up cohort studies. Levels of alkaline phosphatase and bilirubin were analyzed in different settings and subpopulations at different time points relative to the clinical end point (liver transplantation or death). **RESULTS:** Of the 4845 patients, 1118 reached a clinical end point. The median follow-up period was 7.3 years; 77% survived for 10 years after study enrollment. Levels of alkaline phosphatase and bilirubin measured at study enrollment (baseline) and each year for 5 years were strongly associated with clinical outcomes (lower levels were associated with longer transplant-free survival). At 1 year after study enrollment, levels of alkaline phosphatase that were 2.0 times the upper limit of normal (ULN) best predicted patient outcome (C statistic, 0.71) but not significantly better than other thresholds. Of patients with alkaline phosphatase levels $\leq$2.0 times the ULN, 84% survived for 10 years compared with 62% of those with levels >2.0 times the ULN ($P < .0001$). Absolute levels of alkaline phosphatase 1 year after study enrollment predicted patient outcomes better than percentage change in level. One year after study enrollment, a bilirubin level 1.0 times the ULN best predicted patient transplant-free survival (C statistic, 0.79). Of patients with bilirubin levels $\leq$1.0 times the ULN, 86% survived for 10 years after study enrollment compared with 41% of those with levels >1.0 times the ULN ($P < .0001$). Combining levels of alkaline phosphatase and bilirubin increased the ability to predict patient survival times. We confirmed the predictive value of alkaline phosphatase and

†Deceased.

Abbreviations used in this paper: CI, confidence interval; HR, hazard ratio; PBC, primary biliary cirrhosis; UDCA, ursodeoxycholic acid; ULN, upper limit of normal.

0016-5085/$36.00
http://dx.doi.org/10.1053/j.gastro.2014.08.029

bilirubin levels in multiple subgroups, such as patients who had not received treatment with ursodeoxycholic acid, and at different time points after study enrollment. **CONCLUSIONS:** Levels of alkaline phosphatase and bilirubin can predict outcomes (liver transplantation or death) of patients with PBC and might be used as surrogate end points in therapy trials.

Keywords: Autoimmune Liver Disease; Response To Treatment; Biomarker; New Therapies.

Primary biliary cirrhosis (PBC) is a rare, chronic, and slowly progressive autoimmune hepatobiliary disease. PBC typically progresses to cirrhosis, which may lead to complications from liver failure and premature death.[1] Presently, most patients with PBC are treated with ursodeoxycholic acid (UDCA), the only approved therapy for PBC, which is in keeping with treatment guidelines.[2,3] Although UDCA therapy has a marked impact on clinical outcomes in patients with PBC, up to 40% of patients have an insufficient response to this treatment and accordingly have a significantly increased risk of developing an adverse outcome, such as liver transplantation or death.[4–8] Therefore, there is a pressing unmet medical need for better therapies for this serious disease.

A major challenge for patients, health care providers, and drug developers is the slowly progressive nature of PBC, which effectively precludes the evaluation of classic clinical outcomes such as transplant-free survival. The low prevalence of PBC also represents a significant barrier to conducting large controlled clinical outcome trials in patients with this disease. Clinical end points such as liver transplantation and death were evaluated in an early primary interventional trial of UDCA in patients with PBC,[9] but most cases of PBC are now diagnosed at an earlier stage of disease and UDCA therapy is initiated shortly after diagnosis, further affecting the ability to assess the clinical benefit of new PBC therapies in a timely and realistic manner. Thus, the evaluation of scientifically valid surrogate parameters for clinical outcomes is inevitable at least at some stage in the development pathway. Further evaluation of possible surrogates for clinical benefit are needed, particularly with a focus on using large data sets that are representative of the spectrum of disease globally and sufficiently powered through size, duration of follow-up, and numbers of clinical events to refine the scientific validity of specific biochemical surrogates.

Serum bilirubin is well established as an independent predictor of prognosis in PBC, regardless of treatment.[10–12] In addition, bilirubin has previously been shown to be predictive of clinical outcomes across other liver diseases and is incorporated in several commonly used prognostic scoring models, such as the Child–Turcotte–Pugh score,[13,14] the Model of End-Stage Liver Disease (MELD),[15] and, specifically in PBC, the Mayo PBC score.[16] However, despite the proven prognostic value of bilirubin, only patients with relatively advanced disease are likely to show meaningful changes in bilirubin levels that are stratifying. A biochemical variable and potentially more broadly applicable surrogate end point is alkaline phosphatase, an isoenzyme involved in dephosphorylation.[17] An elevated level of alkaline phosphatase, a marker of cholestasis, is typically seen across the spectrum of PBC disease severity and is a key component of the diagnosis of PBC in both the American and European guidelines.[2,3] The relationship between alkaline phosphatase levels and the risk of adverse outcomes in PBC has been extensively documented in several relatively small studies,[4,5,7,8,18,19] but no systematic effort has been reported to date using a pooled meta-analysis approach to validate a biochemical surrogate for use in clinical studies of PBC.

We sought to investigate how serum alkaline phosphatase and bilirubin levels individually and in combination, correlate with transplant-free survival to determine the prognostic significance of these biochemical variables and hence their utility as robust surrogate end points for therapeutic PBC trials. To do so, we assembled a large, international, observational PBC database, allowing for a robust individual patient–level meta-analysis, to ensure both a rigorous statistical assessment and widespread applicability.

Patients and Methods

Study Design and Study Population

This study was a meta-analysis performed by the Global PBC Study Group, an international and multicenter collaboration between 15 liver centers in 8 North American and European countries, which combined individual patient data from major long-term follow-up cohorts. Most individual databases contained prospectively collected follow-up data on patients starting UDCA therapy.

This study was conducted in accordance with the protocol and the principles of the Declaration of Helsinki. The protocol was reviewed and approved by the institutional research board of the corresponding center and at each participating center in accordance with local regulations.

Both UDCA-treated and nontreated patients with an established diagnosis of PBC in accordance with European and American guidelines were eligible for inclusion in this study.[2,3] Patients were excluded from analysis if follow-up data were insufficient or unavailable, the start date of treatment or the exact date of major clinical events was unknown, or they had concomitant liver disease.

Data Collection and Quality Assessment

Collected clinical and laboratory data included sex, age, diagnosis of PBC, liver histology, treatment (type of medication, dosage, and duration), duration and last date of follow-up, baseline antimitochondrial antibody status, baseline and yearly laboratory levels (serum alkaline phosphatase, total bilirubin, albumin, aspartate aminotransferase, alanine aminotransferase, γ-glutamyl transpeptidase, and platelets), and outcomes (death and cause of death, liver transplantation, hepatocellular carcinoma, ascites, and variceal bleeding).

Liver histology performed within 1 year of study entry or documented cirrhosis before study entry was classified as a baseline biopsy. Histological data was assessed for severity according to Ludwig[20] and Scheuer's[21] classification. Disease stage was classified histologically as early (stage I and II) or late

(stage III or IV) and biochemically using serum albumin and bilirubin levels. According to this biochemical classification, early stage was defined by normal bilirubin and albumin levels, moderately advanced disease was defined by an abnormal bilirubin or albumin level, and advanced disease was defined by abnormal bilirubin and albumin levels.[22]

Completeness, plausibility, and validity of the data were carefully verified. Extensive efforts, including site visits with review of medical charts, were undertaken to retrieve missing data. Data of the original cohorts were collected through the end of December 2012.

Statistical Analysis

Study entry was defined as the start date of UDCA therapy or the date of the first center visit for patients not treated with UDCA. The primary end point was defined as a composite of either liver transplantation or death. Patients without documented events during follow-up were censored at their last follow-up visit.

To study the association between the absolute alkaline phosphatase and bilirubin levels, the hazard ratios (HRs) of liver transplantation or death were estimated by applying a cubic spline function of alkaline phosphatase and bilirubin at baseline and yearly up to 5 years of follow-up.

To find an optimal threshold for each variable, alkaline phosphatase and bilirubin levels at 1 year of follow-up were categorized according to multiple thresholds ranging from 1.0 to 3.0 times the upper limit of normal (ULN) in steps of 0.1 (including 1.67 times the ULN[7] for alkaline phosphatase levels). The C statistic was calculated for each of these thresholds to evaluate their ability to predict liver transplant-free survival. Accompanying HRs were calculated for each threshold by using the Cox proportional hazard regression model. The log-likelihood test was used to assess significance. Transplant-free survival was assessed for the peak thresholds of alkaline phosphatase and bilirubin levels and for a combination of both by Kaplan–Meier estimates. Log-rank test was used for comparisons between groups.

In addition to the predictive ability of absolute levels of alkaline phosphatase, the percentage change in alkaline phosphatase levels[4] from baseline to 1-year follow-up was evaluated using the same approach.

All analyses were stratified by center to account for possible heterogeneity across center populations. The effects of alkaline phosphatase and bilirubin were adjusted for year of diagnosis, age at study entry, UDCA therapy, and sex.

To investigate if alkaline phosphatase and bilirubin levels are meaningful surrogate end points, the association with the clinical end point must hold true independent of time and specific patient subgroups. Therefore, the survival analyses were repeated at different time points and for multiple subgroups of patients. The time points analyzed were baseline and yearly up to 5 years of follow-up. Given the nature of this study, alkaline phosphatase or bilirubin levels were not always available for every patient at these time points. Accordingly, we aimed for the optimal use of the available data by assessing the association with hard clinical end points at baseline and several intervals thereafter up to 5 years. Subgroups were defined by treatment (UDCA-treated and nontreated patients), baseline alkaline phosphatase levels (>2.0 times the ULN and >4.0 times the ULN), baseline bilirubin levels (>1 times the ULN and >3 times the ULN), PBC disease state based on both histology and biochemistry, age at time of diagnosis (younger than 45 years and 45 years or older),[23] sex, and date of diagnosis (before 1990, 1990–1999, and 2000–2009).

Normally distributed data are presented as mean ± SD and skewed distributed data as median and interquartile range. All analyses were 2 sided. $P < .05$ was considered statistically significant. Statistical analyses were performed with IBM SPSS Statistics 22.0 (SPSS Inc, Chicago, IL) and SAS 9.3 (SAS institute, Cary, NC).

Results

Baseline Data

Data were obtained from 6191 patients with PBC, of whom 4845 met the inclusion criteria (Figure 1). A total of 65,642 patient visits and a mean of 11 visits per patient were reported across the entire cohort, with a median of 132 elapsing days between visits. Cohort characteristics per center are summarized in Supplementary Table 1. The year in which PBC was diagnosed ranged from 1959 to 2012. The diagnosis was established after 1990 for 79% of patients, and the median year of diagnosis was 1998 (interquartile range, 1991–2004). The median follow-up period was 7.3 years (interquartile range, 3.6–11.5 years) for the cohort, ranging from 6 months to 34 years.

Clinical and biochemical patient characteristics are shown in Table 1. Overall, the demographics were consistent with previous reports of PBC disease epidemiology. Most patients (4119 [85%]) were treated with UDCA at a median dosage of 12.3 mg · kg^{-1} · day^{-1} (interquartile range

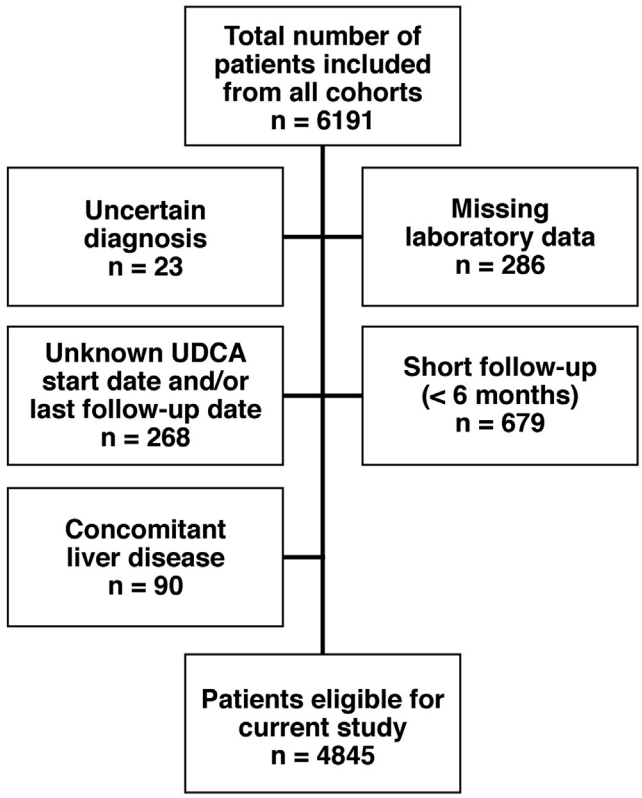


Figure 1. Flowchart of patients included in this study.

Table 1. Baseline Patient Characteristics

	Total cohort (N = 4845)
Age at entry (*y*)	54.5 ± 12.0
Female	4348 (90)
Antimitochondrial antibody positive	4280 (88)
Year of diagnosis	1998 (1991–2004)
Year of diagnosis, time frame	1959–2012
Histological disease stage[a]	
Stage I	1017 (27)
Stage II	862 (23)
Stage III	483 (13)
Stage IV	454 (12)
Not available	953 (25)
Biochemical disease stage[b]	
Early	2040 (42)
Moderately advanced	730 (15)
Advanced	259 (5)
Not available	1816 (38)
Baseline alkaline phosphatase levels	
>2.0× ULN	1931 (52)
>4.0× ULN	816 (22)
Not available	1140 (24)
Treated with UDCA[c]	4119 (85)
Laboratory data at entry	
Serum bilirubin (×ULN)	0.67 (0.45–1.06)
Not available	1118 (23)
Serum alkaline phosphatase (×ULN)	2.10 (1.31–3.72)
Not available	1140 (24)

Data are mean ± SD, n (%), and median (interquartile range).
[a]Histological disease stage according to Ludwig and Scheuer's classification.
[b]Biochemical disease stage according to Rotterdam criteria (using albumin and bilirubin).[22]
[c]640 subjects were nontreated and 86 subjects did not have definitive information on UDCA therapy.

9.4–14.6 mg · kg^{-1} · day^{-1}). Histological stage of disease was available for 76% of patients who had undergone a liver biopsy; most had a diagnosis of early disease (stage I or II).

During follow-up, 1118 patients reached a clinical end point; 389 underwent liver transplantation and 729 died; 358 (49%) died of liver-related causes, 245 patients (34%) died of other causes, and the cause of death was unknown for 126 patients (17%). In the total cohort, 5-year transplant-free survival was 88%, 10-year survival was 77%, and 15-year survival was 63%; in UDCA-treated patients, these findings were 90%, 78%, and 66%, respectively, and in nontreated patients 79%, 59%, and 32%, respectively, (treated vs nontreated, $P < .0001$).

The effects of factors adjusted for in further analyses are shown in Supplementary Table 2.

Association Between Alkaline Phosphatase and Bilirubin Levels and the Risk of Liver Transplantation or Death

A log-linear association was observed between alkaline phosphatase levels and the risk of liver transplantation and death after 1 year and up to 5 years of follow-up, whereby higher alkaline phosphatase levels were associated with reduced transplant-free survival. This association was also found for baseline alkaline phosphatase levels, thus irrespective of subsequent UDCA therapy (Figure 2*A* and Supplementary Figure 1*A*). Abnormal bilirubin levels were even more strongly associated with poor clinical outcome at baseline and up to 5 years of follow-up (Figure 2*B* and Supplementary Figure 1*B*).

Optimal Threshold for Alkaline Phosphatase and Bilirubin Levels and the Risk of Liver Transplantation and Death

The study population was analyzed according to a multitude of thresholds for alkaline phosphatase levels at 1 year of follow-up. This analysis consistently showed that patients with alkaline phosphatase levels below any of these thresholds had significantly improved transplant-free survival compared with patients with alkaline phosphatase levels above the thresholds (Table 2).

After 1 year of follow-up, while all thresholds were predictive of outcomes, a threshold of 2.0 times the ULN for alkaline phosphatase levels was found to have the highest predictive ability (C statistic, 0.71; 95% confidence interval [CI], 0.69–0.73). Notably, this threshold was not a significantly better predictor than the other thresholds, such as 1.5 times the ULN,[8] 1.67 times the ULN,[7,19] or 3.0 times the ULN[5] (Table 2 and Supplementary Figure 2). Similarly, all assessed bilirubin thresholds were predictive of outcomes. For bilirubin, a threshold of 1.0 times the ULN had the highest predictive ability (C statistic, 0.79; 95% CI, 0.77–0.80) (Table 2).

The 5-, 10-, and 15-year transplant-free survival rates for patients with alkaline phosphatase levels ≤2.0 times the ULN were 94%, 84%, and 73%, respectively; for patients with alkaline phosphatase levels >2.0 times the ULN, these rates were 81%, 62%, and 50%, respectively ($P < .0001$), as shown in Figure 3*A*. The accompanying 5-, 10-, and 15-year transplant-free survival rates for patients with normal bilirubin levels after 1 year of follow-up were 95%, 86%, and 74%, respectively; for patients with abnormal bilirubin levels these rates were 65%, 41%, and 30%, respectively ($P < .0001$) (Figure 3*B*).

The prognostic information provided by alkaline phosphatase levels remained important in addition to bilirubin levels; the risk of liver transplantation or death of patients with alkaline phosphatase levels >2.0 times the ULN was significantly higher in both those patients with normal (≤1 times the ULN) and abnormal bilirubin (>1 times the ULN) levels. The 5-, 10-, and 15-year transplant-free survival rates in the normal bilirubin group for patients with alkaline phosphatase levels ≤2.0 times the ULN were 97%, 89%, and 79%, respectively; for patients with alkaline phosphatase levels >2.0 times the ULN, these rates were 95%, 82%, and 68%, respectively ($P < .0001$). In the abnormal bilirubin group, these rates were 74%, 51%, and 39%, respectively, for patients with alkaline phosphatase levels ≤2.0 times the ULN and 63%, 34%, and 24%, respectively, for patients with alkaline phosphatase levels >2.0 times the ULN ($P < .0001$) (Figure 3*C*).

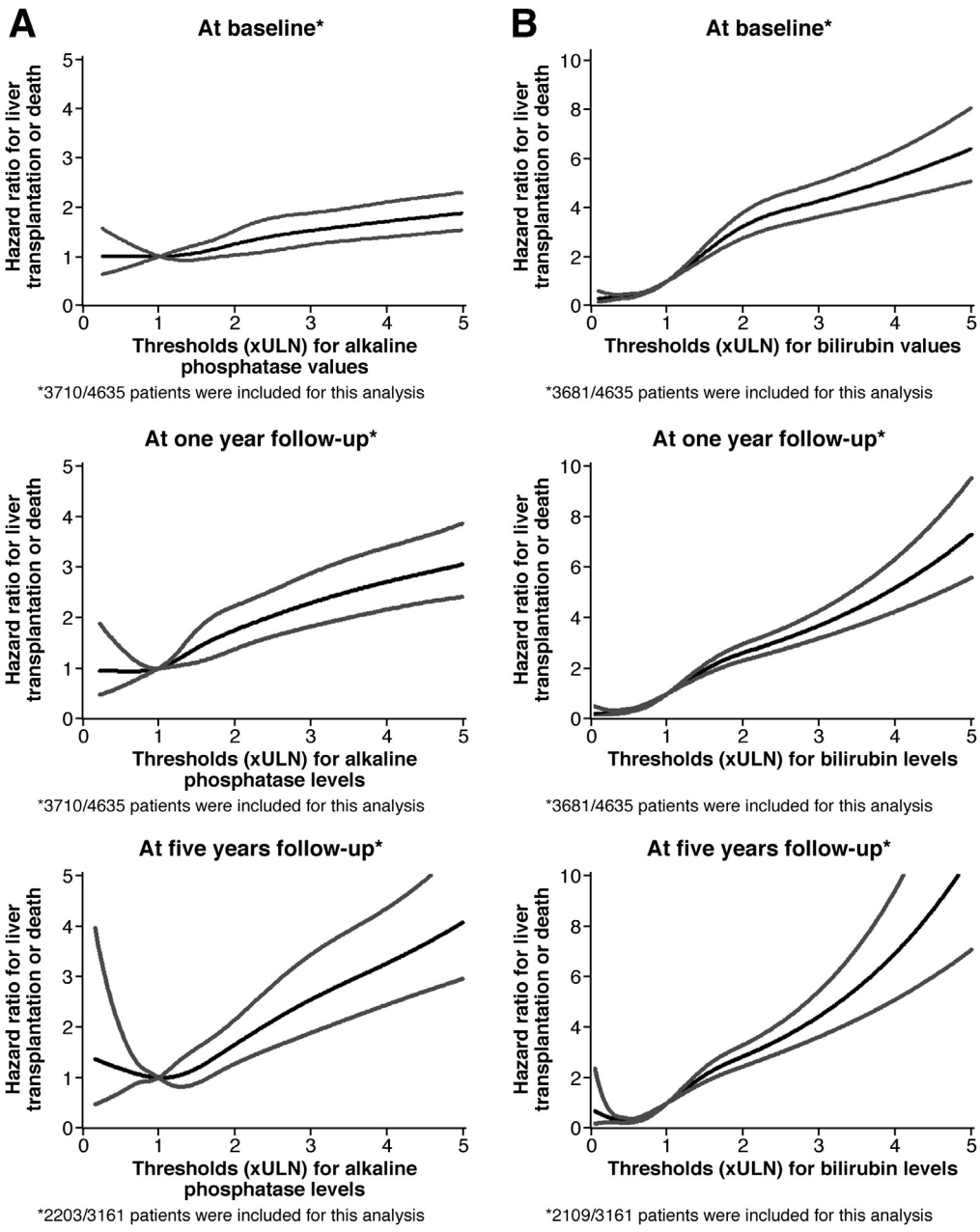


Figure 2. The hazard of liver transplantation or death for (*A*) alkaline phosphatase levels and (*B*) bilirubin levels at different time points estimated with cubic spline function.

An alkaline phosphatase threshold of 2.0 times the ULN was also predictive in addition to other bilirubin thresholds between 1.0 times and 3.0 times the ULN but was not predictive in addition to bilirubin levels >3 times the ULN (HR, 0.71; 95% CI, 0.39–1.32; $P = .29$). Comparable results were found for other alkaline phosphatase thresholds (eg, 1.5 times the ULN and 1.67 times the ULN in combination with normal or abnormal bilirubin levels) (data not shown).

Predictive Value of Percentage Changes in Alkaline Phosphatase Levels at 1-Year Follow-up

A prior study showed that patients who achieved a normal alkaline phosphatase value or had a >40% decrease in alkaline phosphatase levels after UDCA therapy had a normal prognosis.[4] In line with this study, the percentage change in alkaline phosphatase levels from baseline to 1-year follow-up was predictive of outcome; the greater the percentage decrease in alkaline phosphatase levels, the better the transplant-free survival (HR per 10% change in alkaline phosphatase levels 0.98; 95% CI, 0.96–0.99; $P < .01$).

A >40% decrease in alkaline phosphatase levels was found to be significant in predicting outcome (Supplementary Table 3). The predictive value of percentage decrease of alkaline phosphatase levels with UDCA therapy was independent of the baseline alkaline phosphatase levels in patients with a decrease between 0–40% and >40% compared with patients without any decrease (Supplementary Figure 3).

Table 2. Performance of Different Alkaline Phosphatase and Bilirubin Thresholds At 1-Year Follow-up for Prediction of Liver Transplantation or Death

	Alkaline phosphatase (n = 3710)				Bilirubin (n = 3681)			
Thresholds	C statistic (95% CI)	HR (95% CI)	*P* value	No. of patients ≤/> threshold	C statistic (95% CI)	HR (95% CI)	*P* value	No. of patients ≤/> threshold
1.0× ULN	0.68 (0.66–0.70)	2.06 (1.69–2.52)	<.001	1071/2639	0.79 (0.77–0.80)	5.06 (4.34–5.89)	<.001	2941/740
1.1× ULN	0.69 (0.67–0.71)	2.14 (1.79–2.57)	<.001	1306/2404	0.78 (0.77–0.80)	5.22 (4.48–6.08)	<.001	3019/662
1.2× ULN	0.69 (0.67–0.71)	1.97 (1.66–2.33)	<.001	1515/2195	0.78 (0.76–0.80)	5.95 (5.09–6.94)	<.001	3108/573
1.3× ULN	0.69 (0.67–0.71)	2.02 (1.72–2.37)	<.001	1727/1983	0.78 (0.76–0.80)	6.58 (5.61–7.72)	<.001	3172/509
1.4× ULN	0.70 (0.68–0.71)	2.05 (1.75–2.39)	<.001	1900/1810	0.78 (0.76–0.80)	6.87 (5.84–8.09)	<.001	3219/462
1.5× ULN	0.70 (0.68–0.72)	2.14 (1.84–2.50)	<.001	2030/1680	0.77 (0.76–0.79)	7.68 (6.47–9.12)	<.001	3271/410
1.6× ULN	0.70 (0.69–0.72)	2.23 (1.92–2.60)	<.001	2158/1552	0.77 (0.75–0.79)	8.32 (6.99–9.91)	<.001	3297/384
1.67× ULN	0.70 (0.69–0.72)	2.18 (1.88–2.53)	<.001	2231/1479				
1.7× ULN	0.70 (0.69–0.72)	2.22 (1.91–2.57)	<.001	2274/1436	0.77 (0.75–0.79)	8.99 (7.53–10.74)	<.001	3323/358
1.8× ULN	0.71 (0.69–0.73)	2.31 (1.99–2.68)	<.001	2393/1317	0.77 (0.75–0.78)	9.04 (7.53–10.84)	<.001	3346/335
1.9× ULN	0.71 (0.69–0.73)	2.37 (2.04–2.75)	<.001	2466/1244	0.76 (0.75–0.78)	9.30 (7.73–11.20)	<.001	3368/313
2.0× ULN	0.71 (0.69–0.73)	2.49 (2.14–2.89)	<.001	2571/1139	0.76 (0.74–0.78)	10.33 (8.50–12.54)	<.001	3404/277
2.1× ULN	0.71 (0.69–0.72)	2.41 (2.07–2.80)	<.001	2648/1062	0.76 (0.74–0.78)	10.66 (8.76–12.97)	<.001	3417/264
2.2× ULN	0.70 (0.69–0.72)	2.38 (2.05–2.77)	<.001	2714/996	0.75 (0.73–0.77)	10.31 (8.43–12.62)	<.001	3439/242
2.3× ULN	0.70 (0.68–0.72)	2.37 (2.04–2.76)	<.001	2774/936	0.75 (0.73–0.77)	9.98 (8.13–12.24)	<.001	3449/232
2.4× ULN	0.70 (0.68–0.72)	2.37 (2.04–2.77)	<.001	2831/879	0.74 (0.73–0.76)	10.43 (8.46–12.86)	<.001	3461/220
2.5× ULN	0.70 (0.68–0.72)	2.31 (1.98–2.70)	<.001	2885/825	0.74 (0.72–0.76)	10.08 (8.14–12.50)	<.001	3473/208
2.6× ULN	0.70 (0.68–0.72)	2.40 (2.05–2.81)	<.001	2934/776	0.73 (0.71–0.75)	9.81 (7.89–12.21)	<.001	3482/199
2.7× ULN	0.69 (0.67–0.71)	2.38 (2.04–2.79)	<.001	2983/727	0.73 (0.71–0.75)	10.08 (8.07–12.59)	<.001	3487/194
2.8× ULN	0.69 (0.67–0.71)	2.26 (1.92–2.65)	<.001	3036/674	0.73 (0.71–0.75)	9.80 (7.85–12.24)	<.001	3495/186
2.9× ULN	0.69 (0.67–0.71)	2.32 (1.97–2.73)	<.001	3072/638	0.73 (0.71–0.75)	9.49 (7.57–11.91)	<.001	3507/174
3.0× ULN	0.69 (0.67–0.71)	2.31 (1.96–2.73)	<.001	3104/606	0.72 (0.70–0.74)	9.10 (7.23–11.45)	<.001	3516/165

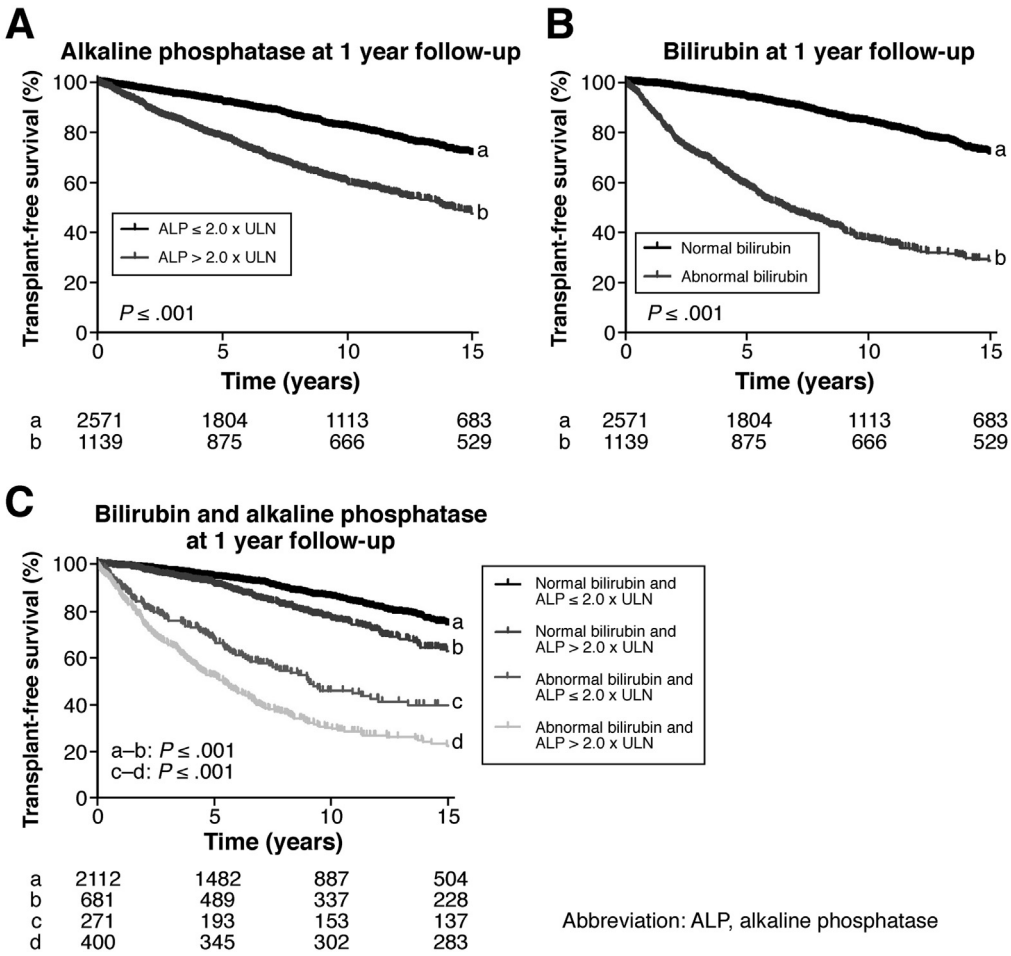


Figure 3. (*A*) Transplant-free survival of patients with alkaline phosphatase levels ≤2.0 times the ULN versus >2.0 times the ULN at 1-year follow-up. (*B*) Transplant-free survival of patients with bilirubin levels ≤1.0 times the ULN versus >1.0 times the ULN at 1-year follow-up. (*C*) Transplant-free survival of patients with alkaline phosphatase levels ≤2.0 times the ULN versus >2.0 times the ULN at 1-year follow-up in both patients with bilirubin levels ≤1 times the ULN and >1 times the ULN.

However, the percentage decrease in alkaline phosphatase levels did not add prognostic information to absolute alkaline phosphatase levels after 1 year of follow-up (HR per 10% change in alkaline phosphatase levels 1.00; 95% CI, 0.99–1.02; $P = .72$), apart from very high alkaline phosphatase levels (>5.0 times the ULN) (HR per 10% change in alkaline phosphatase level, 0.86; 95% CI, 0.76–0.96; $P < .005$).

Predictive Ability of Alkaline Phosphatase and Bilirubin Levels Across Subgroups

To assess if alkaline phosphatase can be used as a predictor independent of patient characteristics, the previously described analyses were repeated for a range of subgroups (Figure 4*A*). Of note, using an alkaline phosphatase threshold of 2.0 times the ULN after 1 year of follow-up was not only predictive for UDCA-treated patients but also for nontreated patients. Similar results were seen in patients with baseline alkaline phosphatase levels >2.0 times the ULN and >4.0 times the ULN, patients with histologically early and late disease, patients with biochemically early and moderately advanced disease, patients 45 years of age or younger at diagnosis and older than 45 years of age at diagnosis, male and female patients, and regardless of the year of diagnosis. Alkaline phosphatase levels were not predictive for patients with advanced biochemical disease (ie, patients with both abnormal bilirubin and albumin levels). A bilirubin threshold of 1.0 times the ULN after 1 year of follow-up was also predictive in several subsets of patients (Figure 4*B*).

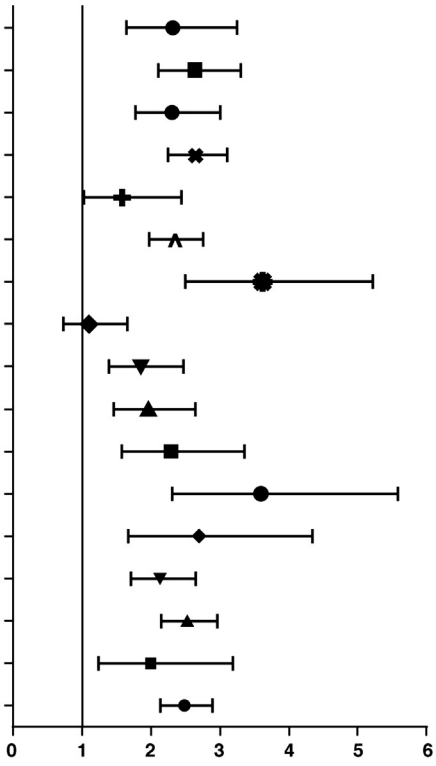
Comparable results were found for alkaline phosphatase and bilirubin at other time points among almost all subgroups (Supplementary Tables 4 and 5).

Translation Into Clinical Practice

For illustrative purposes, the preceding findings were translated into a practical example (Figure 5) to show the association of a composite surrogate end point (bilirubin value <1 times the ULN and alkaline phosphatase value less than the threshold) on 5-year transplant-free survival in different settings. Three groups of high-risk patients with PBC diagnosed after 1990 and treated with UDCA were defined at 2 different time points: baseline (upper panels) and after 1 year of UDCA therapy (lower panels). The subgroups were defined as follows: (1) all patients with PBC, (2) patients meeting the inclusion criteria of a recent clinical

A

Subgroups	N=	Hazard ratio (95% CI)	*P* value
Year of diagnosis 2000–2009	1463	2.32 (1.65–3.25)	< .0001
Year of diagnosis 1990–1999	1364	2.64 (2.11–3.31)	< .0001
Year of diagnosis < 1990	708	2.31 (1.78–3.01)	< .0001
Female	3342	2.64 (2.25–3.10)	< .0001
Male	368	1.58 (1.03–2.44)	.04
Age at diagnosis > 45 years	2800	2.34 (1.98–2.77)	< .0001
Age at diagnosis ≤ 45 years	910	3.62 (2.50–5.23)	< .0001
Advanced stage	211	1.10 (0.73–1.66)	.64
Moderately advanced stage	592	1.85 (1.39–2.48)	< .0001
Early stage	1691	1.97 (1.46–2.65)	< .0001
Histological stage 3 or 4	459	2.46 (1.73–3.51)	< .0001
Histological stage 1 or 2	1042	3.48 (2.32–5.22)	< .0001
Baseline alkaline phosphatase levels > 4.0 x ULN	704	2.70 (1.68–4.34)	< .0001
Baseline alkaline phosphatase levels > 2.0 x ULN	1669	2.13 (1.72–2.65)	< .0001
UDCA treated	3249	2.53 (2.15–2.97)	< .0001
UDCA nontreated	418	2.00 (1.25–3.19)	.004
Entire cohort	3710	2.49 (2.14–2.89)	< .0001



B

Subgroups	N=	Hazard ratio (95% CI)	*P* value
Year of diagnosis 2000–2009	1353	7.06 (4.92–10.15)	< .0001
Year of diagnosis 1990–1999	1376	5.09 (4.00–6.48)	< .0001
Year of diagnosis < 1990	795	4.29 (3.33–5.53)	< .0001
Female	3314	5.34 (4.53–6.29)	< .0001
Male	367	3.09 (2.00–4.77)	< .0001
Age at diagnosis > 45 years	2750	6.44 (4.64–8.93)	< .0001
Age at diagnosis ≤ 45 years	931	4.86 (4.07–5.80)	< .0001
Advanced stage	221	2.32 (1.39–3.89)	.001
Moderately advanced stage	628	2.48 (1.85–3.32)	< .0001
Early stage	1723	2.63 (1.72–4.02)	< .0001
Histological stage 3 or 4	436	5.11 (3.48–7.51)	< .0001
Histological stage 1 or 2	969	5.75 (3.59–9.21)	< .0001
Baseline bilirubin levels > 3.0 x ULN	178	2.65 (1.26–5.58)	.01
Baseline bilirubin levels > 1.0 x ULN	819	2.50 (1.92–3.25)	< .0001
UDCA treated	3181	5.32 (4.50–6.29)	< .0001
UDCA nontreated	459	3.84 (2.47–5.96)	< .0001
Entire cohort	3681	5.06 (4.34–5.89)	< .0001

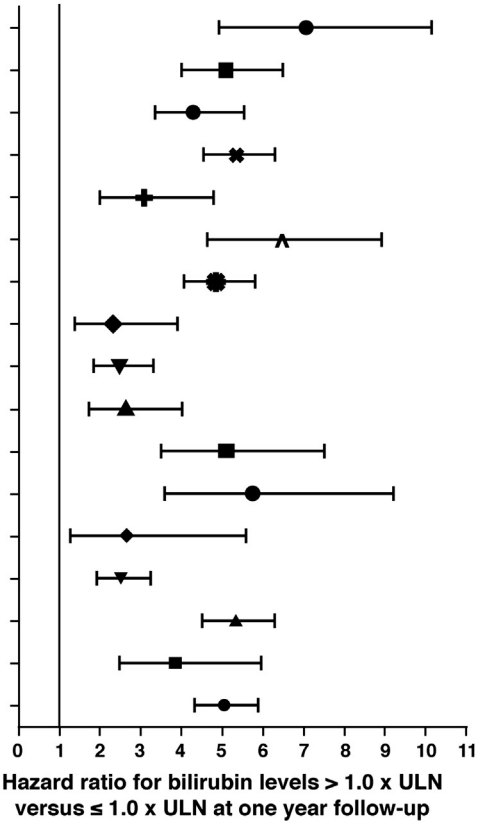


Figure 4. Subgroup analyses of alkaline phosphatase and bilirubin levels. HRs of liver transplantation or death for (*A*) alkaline phosphatase levels >2.0 times the ULN versus ≤2.0 times the ULN and (*B*) bilirubin levels >1.0 times the ULN versus ≤1.0 times the ULN at 1-year follow-up for different subgroups.

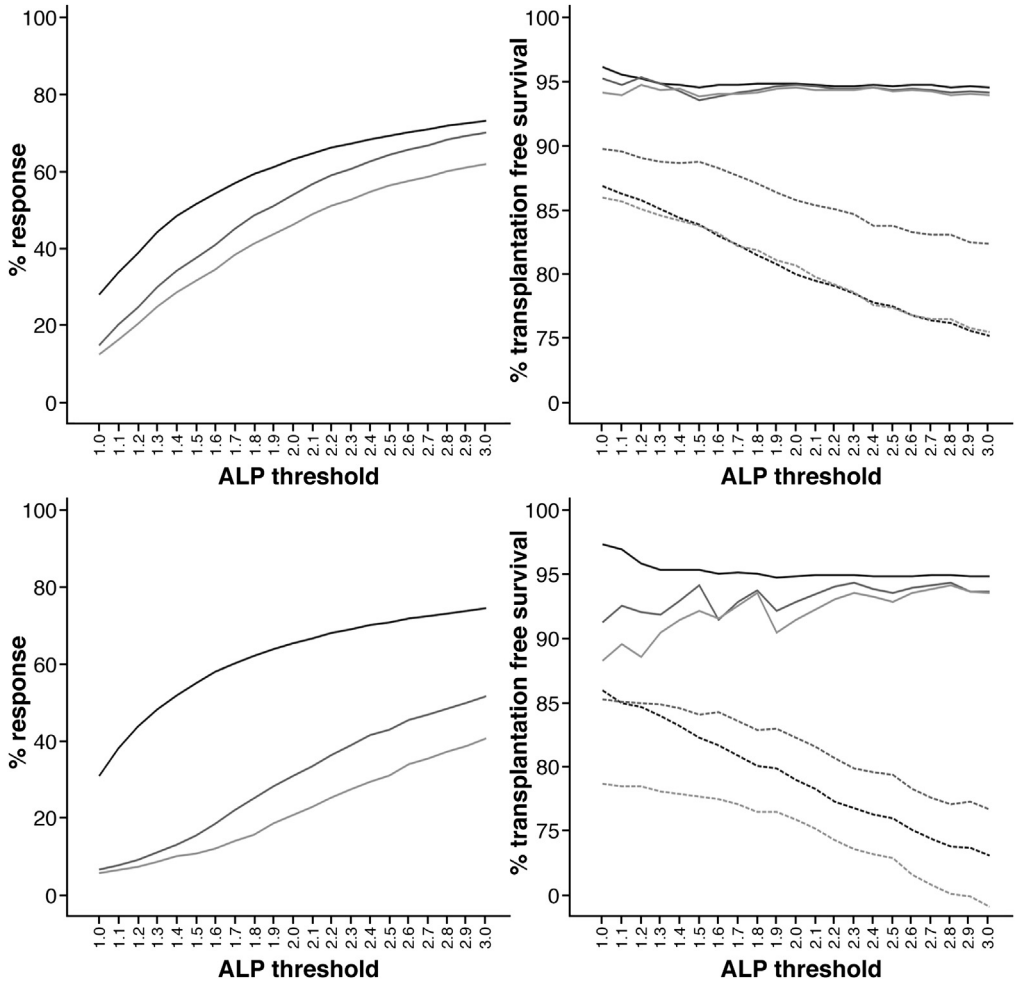


Figure 5. Translation into clinical practice. The association of a surrogate end point, defined as alkaline phosphatase levels less than the threshold and bilirubin levels? less than 1 time the ULN, on 5-year transplant-free survival in different settings. Inclusion (diagnosis after 1990 and UDCA therapy initiated) was made at baseline (*upper panels*) and after 1 year of UDCA therapy (*lower panels*). Three high-risk groups were defined as follows: (1) all patients (*black lines*), (2) bilirubin levels less than 2 times the ULN and either alkaline phosphatase levels >1.67 times the ULN or bilirubin levels >1 time the ULN[24] (*dark gray lines*), and (3) bilirubin levels <3 times the ULN and either alkaline phosphatase levels >2 times the ULN or bilirubin levels >1 time the ULN) (*light gray lines*). The *solid lines* represent the patients who reached the surrogate end point, and the *dotted lines* represent those who did not. The *left panels* show the proportion of patients reaching the surrogate end point 1 year after inclusion, and the *right panels* show the corresponding 5-year transplant-free survival.

trial: bilirubin value <2 times the ULN and either alkaline phosphatase value >1.67 times the ULN or bilirubin value >1 times the ULN,[24] and (3) patients with a bilirubin value <3 times the ULN and either alkaline phosphatase value >2 times the ULN or bilirubin value >1 times the ULN. The surrogate end point was determined 1 year after inclusion. Figure 5 shows the proportion of patients reaching the surrogate end point (left panels) and accompanying transplant-free survival (right panels).

If a bilirubin level <1 times the ULN and alkaline phosphatase level <2 times the ULN is used as a surrogate end point in high-risk PBC population 3 (light gray) and if patients are already treated with UDCA for 1 year (lower panels), the proportion of patients reaching the surrogate end point after an additional year of UDCA therapy is 18% (lower left panel), with an accompanying 5-year transplantation-free survival rate of 92% (lower right panel). The 5-year transplantation-free survival rate for patients not reaching the surrogate end point was 75%.

In summary, using higher alkaline phosphatase thresholds resulted in a lower proportion of patients not reaching the surrogate end point, with a poorer corresponding 5-year transplant-free survival. The 5-year transplant-free survival after continued UDCA therapy is irrespective of the chosen alkaline phosphatase threshold and risk population.

Discussion

This study reports a robust and uniquely powered, independent evaluation of the largest meta-analysis of individual data on PBC to date. We unequivocally show that both increased serum alkaline phosphatase and bilirubin

levels are strongly associated with reduced transplant-free survival in patients with PBC and that a combination of both variables improves prognostic prediction for patients. These associations are independent of use of UDCA and follow-up time and held for multiple subgroups. These data support that both alkaline phosphatase and bilirubin provide meaningful surrogate end points in PBC that can reasonably be used in clinical trials.

Prior studies have shown an association between normalization, percentage decreases or absolute decreases of alkaline phosphatase levels and improved prognosis with UDCA therapy.[4,5,7,8,18,19] The present study reports for the first time a near log-linear association between alkaline phosphatase levels and transplant-free survival and clearly shows that the lower the alkaline phosphatase value, the greater the transplant-free survival time. This applied not only to alkaline phosphatase levels during follow-up but also to baseline levels irrespective of subsequent UDCA therapy. The suitability of alkaline phosphatase as a surrogate end point for clinical benefit is further supported by the finding that the prognostic information provided by alkaline phosphatase levels was confirmed across a wide range of subgroups such as nontreated patients, relatively young patients, and patients with histologically early and late disease. This finding is of considerable clinical significance because alkaline phosphatase constitutes one of the 3 potential diagnostic criteria and is used routinely to assess disease activity.

Our study additionally confirms that as baseline bilirubin levels or bilirubin levels increase over time, the likelihood of survival correspondingly decreases.[10] The predictive ability of alkaline phosphatase levels was shown in addition to bilirubin to discriminate high-risk and low-risk patients. This is an important observation because bilirubin on its own is unsuitable as a surrogate end point in clinical trials because it is typically elevated only when the disease has progressed to the stage at which liver function becomes impaired. Most patients likely to be included in such studies will have normal levels precluding the possibility of observing potential beneficial treatment effects based on bilirubin alone.

It has been suggested that the best way to evaluate the utility of a biomarker as a good surrogate end point may be a meta-analysis of clinical trials of one or more interventions.[25] A 4-level hierarchy of evidence to consider the validation of surrogate end points has been proposed:

> *Level 1: a true clinical-efficacy measure; Level 2: a validated surrogate endpoint (for a specific disease setting and class of interventions); Level 3: a non-validated surrogate endpoint, yet one established to be "reasonably likely to predict clinical benefit" (for a specific disease setting and class of interventions); Level 4: a correlate that is a measure of biological activity but that has not been established to be a higher level.*[26]

The particular challenge of confirming biomarkers as surrogate end points in PBC is that there is only one approved treatment, and previous meta-analyses of published clinical trials that have been conducted in PBC have been interpreted in conflicting ways.[27–29] Interpretation of the data is compromised due to design issues, such as a lack of consistent long-term follow-up.[29,30] Our approach was therefore to conduct a more rigorous patient-level meta-analysis of existing cohorts of patients at centers across North America and Europe with long-term follow-up data of large numbers of patients with PBC. This design has sufficient power to intensively study alkaline phosphatase and bilirubin as potential surrogate end points in different settings, subpopulations, and time points. Based on these current results, we postulate that alkaline phosphatase and bilirubin levels are "reasonably likely to predict clinical benefit" in PBC.[26] This is of relevance to future trial design for new therapeutic agents.

Alternative surrogates have been suggested, such as liver histology,[31] which may provide key information on treatment effects in PBC. However, liver biopsy is not routinely conducted in patients with PBC. Given its invasive nature and small but well-recognized risks,[32] liver histology, with its added inherent sampling variability, is not an ideal surrogate for widespread use in patients with PBC. Noninvasive elastography-based assessment of liver fibrosis may potentially be used as a reliable alternative in the prediction of fibrosis[33]; however, further long-term evaluation is required in PBC. Similarly other biochemical surrogates have been suggested[5,6,34] but as of yet are not widely studied. We focused on the routine biochemical measurements that have been used for many years in both the diagnosis and management of patients with PBC, because this approach is likely to be the most easily applied in practice.

There are some limitations to our study. The availability of some clinical data (such as ascites, edema, pruritus, fatigue, or use of diuretics) and laboratory data (including prothrombin time, immunoglobulin M, and immunoglobulin G levels) in the individual databases varied considerably. In many cases, in particular when databases contained data of patients entered more than 10 to 20 years ago, it was not possible to collect these data consistently in a reliable way. Further, no uniform or generally accepted or validated methods had been used in the contributing centers to quantify subjective signs and symptoms. As a consequence, within the context of this study, we were unfortunately unable to include this type of information in our analyses and, in particular, were not able to calculate the Mayo risk score[16] and to compare the prognostic information provided by this established prediction tool with that provided by alkaline phosphatase and bilirubin.

Due to the nature of our study, biochemical data were not always available at the fixed time points during follow-up. This was mainly encountered when the original data had been collected more than 20 years ago. Data on dose changes or interruption of UDCA therapy was also not uniformly available. However, we believe that these limitations had no major impact on the reliability of the results, considering the unique large size of the study population, the prospective nature of most of the data, the inclusion of both UDCA-treated and nontreated patients, the substantial

incidence of clinical end points, and the duration of follow-up. Additionally, adjusting for missing data by multiple imputations of the data, the results did not change (Supplementary Table 6).

Based on our present results, any decrease in alkaline phosphatase or bilirubin levels translates into improved prognosis; lower levels are clearly associated with better transplant-free survival. In our population, the most discriminative alkaline phosphatase threshold after 1 year of follow-up was 2.0 times the ULN, which is an earlier proposed threshold,[18] although an alkaline phosphatase threshold of 1.5 times the ULN,[8] 1.67 times the ULN,[7,19] or 3.0 times ULN would all work well as a surrogate end point in a clinical trial setting. For bilirubin, the choice of threshold is even clearer; the spline plots (Figure 1) suggest that a choice of bilirubin <1.0 times the ULN is reasonable. However, designing clinical trials implies the a priori requirement to estimate the quantitative effect of an experimental intervention on a given end point. Based on the current study, we are not able to translate these data into a specific threshold for a clinical trial in general.

In conclusion, our study shows that alkaline phosphatase and bilirubin levels strongly correlate with the ultimate outcomes of death and liver transplantation in patients with PBC; the lower the alkaline phosphatase and bilirubin levels the better the transplant-free survival times. This robust analysis suggests that these variables can reasonably be regarded as useful surrogate end points in clinical trials. There is a high unmet medical need for new therapies for this rare autoimmune liver disease, and this study provides an important impetus for the selection of appropriate end points and to facilitate the conduct of meaningful therapeutic intervention trials in the absence of long-term outcome studies.

Supplementary Material

Note: To access the supplementary material accompanying this article, visit the online version of *Gastroenterology* at www.gastrojournal.org, and at http://dx.doi.org/10.1053/j.gastro.2014.08.029.

References

1. Kaplan MM, Gershwin ME. Primary biliary cirrhosis. N Engl J Med 2005;353:1261–1273.
2. Lindor KD, Gershwin ME, Poupon R, et al. Primary biliary cirrhosis. Hepatology 2009;50:291–308.
3. European Association for the Study of the Liver. EASL clinical practice guidelines: management of cholestatic liver diseases. J Hepatol 2009;51:237–267.
4. Pares A, Caballeria L, Rodes J. Excellent long-term survival in patients with primary biliary cirrhosis and biochemical response to ursodeoxycholic acid. Gastroenterology 2006;130:715–720.
5. Corpechot C, Abenavoli L, Rabahi N, et al. Biochemical response to ursodeoxycholic acid and long-term prognosis in primary biliary cirrhosis. Hepatology 2008;48:871–877.
6. Kuiper EM, Hansen BE, de Vries RA, et al. Improved prognosis of patients with primary biliary cirrhosis that have a biochemical response to ursodeoxycholic acid. Gastroenterology 2009;136:1281–1287.
7. Kumagi T, Guindi M, Fischer SE, et al. Baseline ductopenia and treatment response predict long-term histological progression in primary biliary cirrhosis. Am J Gastroenterol 2010;105:2186–2194.
8. Corpechot C, Chazouilleres O, Poupon R. Early primary biliary cirrhosis: biochemical response to treatment and prediction of long-term outcome. J Hepatol 2011;55:1361–1367.
9. Poupon RE, Poupon R, Balkau B. Ursodiol for the long-term treatment of primary biliary cirrhosis. The UDCA-PBC Study Group. N Engl J Med 1994;330:1342–1347.
10. Shapiro JM, Smith H, Schaffner F. Serum bilirubin: a prognostic factor in primary biliary cirrhosis. Gut 1979;20:137–140.
11. 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.
12. Krzeski P, Zych W, Kraszewska E, et al. Is serum bilirubin concentration the only valid prognostic marker in primary biliary cirrhosis? Hepatology 1999;30:865–869.
13. Child CG, Turcotte JG. Surgery and portal hypertension. Major Probl Clin Surg 1964;1:1–85.
14. Pugh RN, Murray-Lyon IM, Dawson JL, et al. Transection of the oesophagus for bleeding oesophageal varices. Br J Surg 1973;60:646–649.
15. Kamath PS, Wiesner RH, Malinchoc M, et al. A model to predict survival in patients with end-stage liver disease. Hepatology 2001;33:464–470.
16. Dickson ER, Grambsch PM, Fleming TR, et al. Prognosis in primary biliary cirrhosis: model for decision making. Hepatology 1989;10:1–7.
17. Warnes TW. Alkaline phosphatase. Gut 1972;13:926–937.
18. Angulo P, Lindor KD, Therneau TM, et al. Utilization of the Mayo risk score in patients with primary biliary cirrhosis receiving ursodeoxycholic acid. Liver 1999;19:115–121.
19. Momah N, Silveira MG, Jorgensen R, et al. Optimizing biochemical markers as endpoints for clinical trials in primary biliary cirrhosis. Liver Int 2012;32:790–795.
20. Ludwig J, Dickson ER, McDonald GS. Staging of chronic nonsuppurative destructive cholangitis (syndrome of primary biliary cirrhosis). Virchows Arch A Pathol Anat Histol 1978;379:103–112.
21. Scheuer P. Primary biliary cirrhosis. Proc R Soc Med 1967;60:1257–1260.
22. ter Borg PC, 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;101:2044–2050.
23. **Carbone M, Mells GF**, Pells G, 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; quiz e13–14.
24. Phase 3 study of obeticholic acid in patients with primary biliary cirrhosis. http://clinicaltrials.gov/ct2/show/NCT01473524.
25. Fleming TR, Powers JH. Biomarkers and surrogate endpoints in clinical trials. Stat Med 2012;31:2973–2984.
26. Fleming TR. Surrogate endpoints and FDA's accelerated approval process. Health Aff (Millwood) 2005;24:67–78.
27. Goulis J, Leandro G, Burroughs AK. Randomised controlled trials of ursodeoxycholic-acid therapy for primary biliary cirrhosis: a meta-analysis. Lancet 1999;354:1053–1060.
28. Rudic JS, Poropat G, Krstic MN, et al. Ursodeoxycholic acid for primary biliary cirrhosis. Cochrane Database Syst Rev 2012;12:CD000551.
29. **Shi J, Wu C, Lin Y**, et al. Long-term effects of mid-dose ursodeoxycholic acid in primary biliary cirrhosis: a meta-analysis of randomized controlled trials. Am J Gastroenterol 2006;101:1529–1538.
30. Poupon RE, Lindor KD, Cauch-Dudek K, et al. Combined analysis of randomized controlled trials of ursodeoxycholic acid in primary biliary cirrhosis. Gastroenterology 1997;113:884–890.
31. Corpechot C, Carrat F, Bahr A, et al. The effect of ursodeoxycholic acid therapy on the natural course of primary biliary cirrhosis. Gastroenterology 2005;128:297–303.
32. Rockey DC, Caldwell SH, Goodman ZD, et al. Liver biopsy. Hepatology 2009;49:1017–1044.
33. Corpechot C, Carrat F, Poujol-Robert A, et al. Noninvasive elastography-based assessment of liver fibrosis progression and prognosis in primary biliary cirrhosis. Hepatology 2012;56:198–208.
34. Azemoto N, Kumagi T, Abe M, et al. Biochemical response to ursodeoxycholic acid predicts long-term outcome in Japanese patients with primary biliary cirrhosis. Hepatol Res 2011;41:310–317.

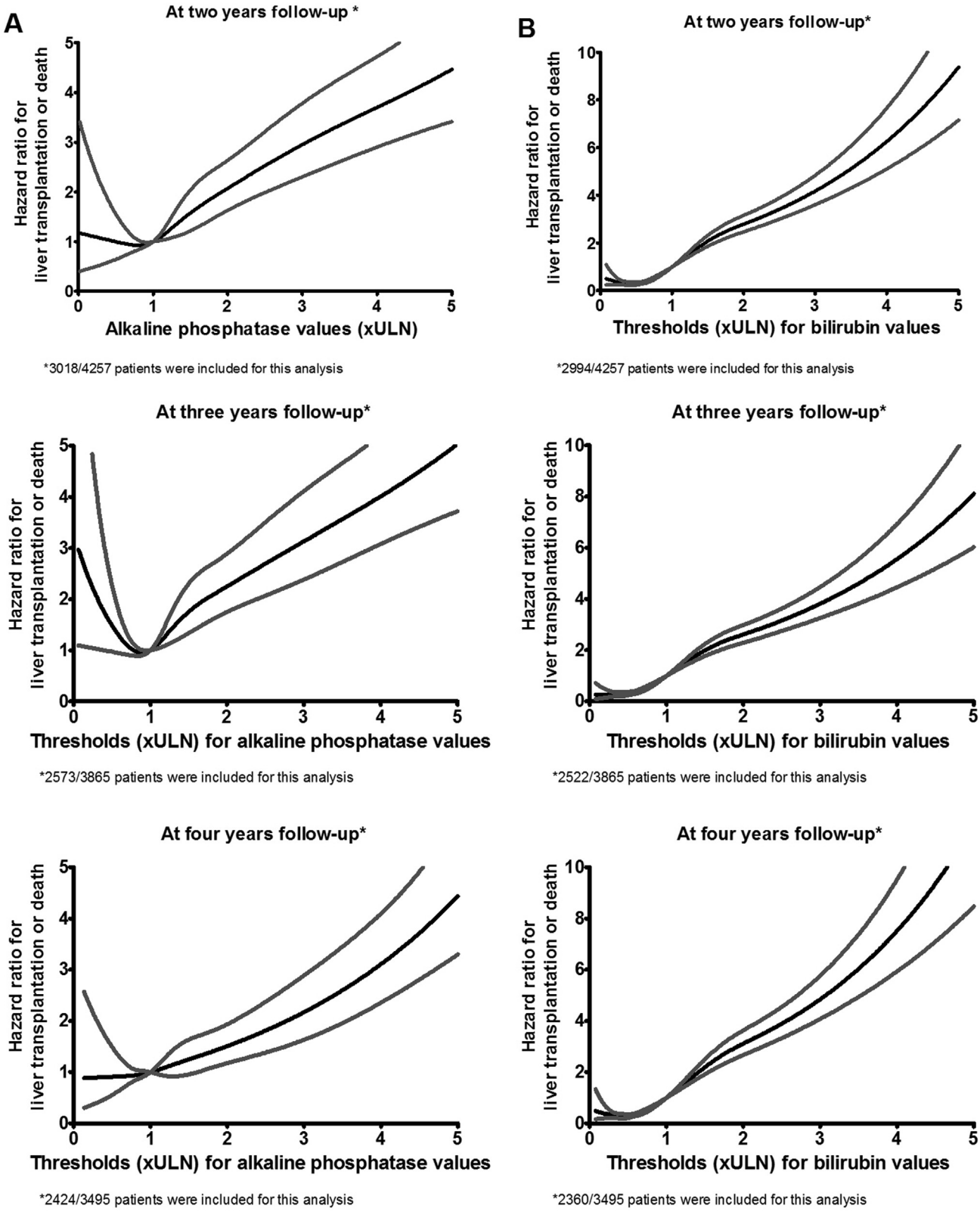
Author names in bold designate shared co-first authorship.

Received April 17, 2014. Accepted August 20, 2014.

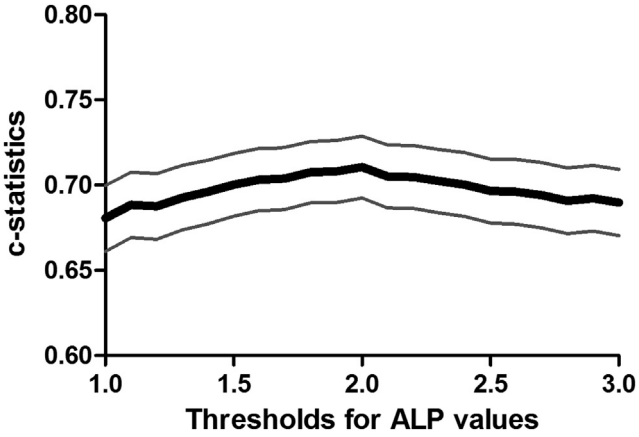
Reprint requests
Address requests for reprints to: Bettina E. Hansen, PhD, Department of Gastroenterology and Hepatology, Erasmus University Medical Centre, 's-Gravendijkwal 230, Room Ha 203, 3015 CE Rotterdam, The Netherlands. e-mail: b.hansen@erasmusmc.nl; fax: (31) 10 436 5916.

Conflicts of interest
The authors disclose the following: H.R.V. has received unrestricted grants from Intercept Pharmaceuticals and Zambon Nederland BV. G.M.H. is a study investigator for Intercept Pharmaceuticals, Dr Falk Pharma, Gilead Sciences, Lumena Pharmaceuticals, FFP Pharma, and Janssen and is an unpaid consultant for Intercept, Lumena Pharmaceuticals, and FFP Pharma. H.L.A.J. has received grants from and is a consultant for Bristol-Myers Squibb, Gilead Sciences, Novartis, Roche, Merck, Innogenetics, Abbott, Santaris, Medtronic, and Tibotec. P.I. is a consultant for Menarini Diagnostics, Instrumentation Laboratories, and Medigene. A.L.M. has received funding from the Canadian Institutes for Health Research, Canadian Liver Foundation, Alberta Heritage Foundation for Medical Research, and Alberta Cancer Foundation, and Abbvie and Gilead Sciences have provided medications for ongoing clinical trials. C.Y.P. has received unrestricted grants from Abbvie, Dr Falk Pharma, and Takeda; consultancy fees from Abbvie, GlaxoSmithKline, and Takeda; and speaker's fees from Abbvie, MSD, Takeda, and Ferring Pharmaceuticals. M.J.M. has received grant support in the form of conducting sponsor-initiated clinical trials from Intercept Pharmaceuticals, Gilead Sciences, Salix, Lumena Pharmaceuticals, and NGM. A.P. is a consultant for Lumena Pharmaceuticals, Inc. K.V.K. has received grants and research support from AbbVie, Beckman, Bristol-Myers Squibb, Boehringer Ingelheim, Gilead Sciences, Ikaria, Intercept Pharmaceuticals, Janssen, Merck, Mochida, and Vertex; received consultancy fees from Novartis and Tekmira Pharmaceuticals (honorarium payable to institution); and served as an advisory board member for AbbVie, Boehringer Ingelheim, Gilead Sciences, Ikaria, Janssen, Merck, and Trio Health (honorarium payable to institution). P.J.T. is the recipient of a Wellcome Trust Clinical Research Fellowship. T.K. has received fellowship support from an unrestricted grant from Axcan Pharma. K.D.L. is a member of the American Association for the Study of Liver Diseases Governing Board and primary author of guidelines on primary sclerosis cholangitis (to be completed this year) and is an unpaid advisor for Intercept Pharmaceuticals and Lumena Pharmaceuticals. The remaining authors disclose no conflicts.

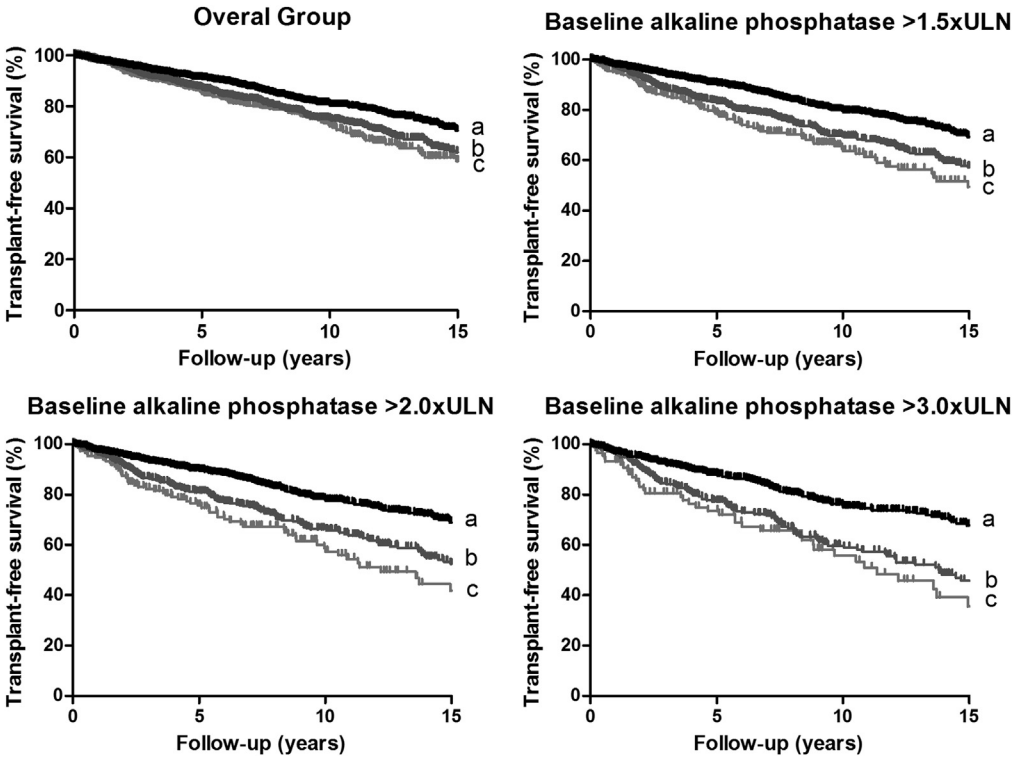
Funding
This investigator-initiated study was supported by unrestricted grants from Intercept Pharmaceuticals and Zambon Nederland BV and was funded by the Foundation for Liver and Gastrointestinal Research (a not-for-profit foundation) in Rotterdam, The Netherlands. The supporting parties had no influence on the study design, data collection and analyses, writing of the manuscript, or the decision to submit the manuscript for publication.



Supplementary Figure 1. The hazard of liver transplantation or death for (*A*) alkaline phosphatase levels and (*B*) bilirubin levels at different time points estimated with cubic spline function.



Supplementary Figure 2. Performance of alkaline phosphatase (ALP) thresholds. The C statistic was performed for different thresholds for alkaline phosphatase levels at 1 year of follow-up. The C statistic reflects the predictive accuracy of alkaline phosphatase thresholds to distinguish patients with a high risk of liver transplantation or death from patients with a low risk.



Supplementary Figure 3. Transplant-free survival for percent decrease of alkaline phosphatase levels at 1 year of follow-up. Transplant-free survival of patients with (*A*) >40% decrease of alkaline phosphatase levels (*B*) 0–40% decrease of alkaline phosphatase levels and (*C*) no decrease of alkaline phosphatase levels.

Supplementary Table 1. Center-Specific Characteristics of the Study Population

Center		Year of diagnosis		Follow-up (*y*)		UDCA therapy		Alkaline phosphatase level at study entry	Bilirubin level at study entry
Country (city)	N	Median	Time frame	Median (IQR)	Range	n	%	Median (IQR)	Median (IQR)
United States (Rochester)	857	2002	1970–2012	4.7 (2.3–9.0)	0.5–18.1	590	69	1.63 (1.07–2.59)	0.80 (0.50–1.40)
The Netherlands (nationwide cohort)	838	1999	1961–2012	8.9 (4.8–14.2)	0.5–24.2	838	100	2.10 (1.39–3.63)	0.61 (0.44–0.90)
Canada (Toronto)	628	1999	1974–2010	7.3 (4.0–11.4)	0.5–34.3	529	84	2.50 (1.66–4.60)	0.55 (0.40–0.82)
Italy (Padua)	544	1989	1959–2005	7.1 (3.6–12.0)	0.5–25.8	386	71	2.56 (1.50–4.29)	0.80 (0.54–1.38)
England (Birmingham)	363	2003	1972–2011	6.0 (3.3–9.4)	0.6–16.7	285	79	1.91 (1.16–3.20)	0.52 (0.38–1.10)
France (Paris)	348	1988	1974–2001	5.9 (2.1–8.9)	0.5–22.5	348	100	3.00 (1.90–5.30)	0.67 (0.43–1.17)
United States (Dallas)	326	1993	1977–2011	8.8 (6.9–11.6)	0.8–23.7	326	100	2.67 (1.54–3.86)	0.46 (0.31–0.67)
Italy (Milan, 2 centers)	289	1999	1972–2012	7.2 (3.4–13.3)	0.5–23.8	289	96	1.74 (1.05–3.26)	0.67 (0.48–1.00)
Spain (Barcelona)	273	1995	1971–2005	12.1 (7.7–16.3)	0.6–23.8	266	97	1.89 (1.24–3.32)	0.67 (0.50–1.00)
Belgium (Leuven)	150	2000	1974–2011	6.8 (3.4–12.8)	0.6–28.8	136	91	2.75 (1.66–4.58)	0.72 (0.47–1.18)
England (London)	138	1994	1972–2007	8.5 (5.1–12.1)	0.5–22.5	56	41	1.83 (1.14–3.09)	0.53 (0.41–0.71)
Canada (Edmonton)	57	2003	1989–2008	5.9 (4.0–8.3)	0.7–18.4	53	93	3.13 (2.10–5.57)	0.82 (0.57–1.24)
United States (Seattle)	34	2008	1995–2012	6.0 (3.3–9.4)	0.5–16.7	30	88	1.69 (1.11–2.68)	0.42 (0.33–0.50)
Total	4845	1998	1959–2012	7.3 (3.6–11.5)	0.5–34.3	4119	85	2.10 (1.31–3.72)	0.67 (0.45–1.06)

IQR, interquartile range.

Supplementary Table 2. Univariable and Multivariable Analysis Showing the Effects of Variables at Baseline Predictive of Liver Transplantation and Death

	Univariable analysis		Multivariable analysis	
	HR (95% CI)	*P* value	HR (95% CI)	*P* value
Year of diagnosis	0.95 (0.94–0.95)	<.001	0.95 (0.94–0.96)	<.001
Age at study entry	1.04 (1.03–1.04)	<.001	1.03 (1.03–1.04)	<.001
UDCA therapy	0.59 (0.50–0.71)	<.001	0.61 (0.51–0.74)	<.001
Male sex	1.52 (1.28–1.80)	<.001	1.46 (1.22–1.75)	<.001

Supplementary Table 3. HRs for Predicting Liver Transplantation or Death for Percentage Change of Alkaline Phosphatase Levels From Baseline to 1-Year Follow-up

Percentage reduction of alkaline phosphatase	HR (95% CI)	*P* value
No reduction	1	
0–10%	0.88 (0.63–1.23)	.45
10–20%	0.85 (0.60–1.20)	.36
20–30%	0.67 (0.48–0.95)	.03
30–40%	0.84 (0.61–1.15)	.23
40–50%	0.70 (0.51–0.96)	.03
50–60%	0.59 (0.42–0.83)	.003
>60%	0.62 (0.44–0.86)	.005

Supplementary Table 4. HRs of Liver Transplantation or Death for Alkaline Phosphatase Levels at Baseline and 2-Year Follow-up for Different Subgroups

	Alkaline phosphatase levels >2.0× ULN vs ≤2.0× ULN					
	At baseline			At 2-y follow-up		
	n	*P* value	HR (95% CI)	n	*P* value	HR (95% CI)
Year of diagnosis 2000–2009	1479	<.0001	2.23 (1.65–3.02)	1170	<.0001	2.50 (1.61–3.88)
Year of diagnosis 1990–1999	1298	<.0001	2.05 (1.62–2.60)	1176	<.0001	2.41 (1.87–3.11)
Year of diagnosis before 1990	754	.002	1.54 (1.17–2.04)	579	<.0001	2.41 (1.78–3.25)
Female	3320	<.0001	1.94 (1.65–2.29)	2717	<.0001	2.75 (2.27–3.32)
Male	385	.01	1.64 (1.11–2.42)	301	.64	1.13 (0.67–1.91)
Age at diagnosis older than 45 y	2847	<.0001	1.84 (1.56–2.17)	2264	<.0001	2.35 (1.93–2.87)
Age at diagnosis 45 y or younger	858	<.0001	2.52 (1.65–3.84)	754	<.0001	3.11 (2.08–4.64)
Advanced stage	239	.44	1.17 (0.78–1.75)	152	.57	1.17 (0.68–2.01)
Moderately advanced stage	667	.79	1.04 (0.79–1.37)	453	.01	1.56 (1.11–2.19)
Early stage	1905	.002	1.52 (1.16–2.00)	1347	<.0001	2.02 (1.42–2.87)
Histological stage 3 or 4	449	.54	0.88 (0.59–1.32)	396	.0008	2.26 (1.40–3.64)
Histological stage 1 or 2	1013	<.0001	2.63 (1.73–3.99)	866	<.0001	4.15 (2.50–6.87)
Baseline alkaline phosphatase levels >4.0× ULN				557	<.0001	2.87 (1.80–4.58)
Baseline alkaline phosphatase levels >2.0× ULN				1342	<.0001	2.22 (1.74–2.83)
UDCA treated	3090	<.0001	2.01 (1.68–2.39)	2719	<.0001	2.68 (2.22–3.23)
UDCA nontreated	537	.003	1.68 (1.19–2.37)	265	.06	1.67 (0.97–2.86)
Entire cohort	3705	<.0001	1.87 (1.61–2.18)	3018	<.0001	2.49 (2.09–2.96)

Supplementary Table 5. HRs of Liver Transplantation or Death for Bilirubin Levels at Baseline and 2-Year Follow-up for Different Subgroups

	Bilirubin value >1.0× ULN vs ≤1.0× ULN					
	At baseline			At 2-y follow-up		
	n	*P* value	HR (95% CI)	n	*P* value	HR (95% CI)
Year of diagnosis 2000–2009	1409	<.0001	4.73 (3.45–6.48)	1054	<.0001	6.55 (4.09–10.48)
Year of diagnosis 1990–1999	1312	<.0001	5.50 (4.36–6.93)	1194	<.0001	4.54 (3.46–5.96)
Year of diagnosis before 1990	846	<.0001	4.00 (3.16–5.05)	665	<.0001	4.33 (3.22–5.82)
Female	3332	<.0001	4.94 (4.24–5.75)	2699	<.0001	4.92 (4.06–5.97)
Male	395	<.0001	3.50 (2.36–5.19)	295	<.0001	4.34 (2.55–7.38)
Age at diagnosis older than 45 y	901	<.0001	4.27 (3.64–5.01)	2225	<.0001	4.23 (3.44–5.21)
Age at diagnosis 45 y or younger	2826	<.0001	7.25 (5.12–10.26)	769	<.0001	8.62 (5.79–12.83)
Advanced stage				161	.01	2.33 (1.20–4.51)
Moderately advanced stage				477	<.0001	2.49 (1.74–3.57)
Early stage				1379	<.0001	3.70 (2.42–5.67)
Histological stage 3 or 4	429	<.0001	3.80 (2.54–5.70)	370	<.0001	4.20 (2.73–6.48)
Histological stage 1 or 2	946	<.0001	8.98 (5.63–14.32)	814	<.0001	5.19 (3.08–8.74)
Baseline bilirubin levels >3.0× ULN				116	.35	1.51 (0.64–3.53)
Baseline bilirubin levels >1.0× LN				630	<.0001	2.09 (1.54–2.84)
UDCA treated	3069	<.0001	5.28 (4.50–6.20)	2662	<.0001	5.05 (4.17–6.13)
UDCA nontreated	596	<.0001	3.41 (2.42–4.81)	301	<.0001	3.53 (1.96–6.35)
Entire cohort	3727	<.0001	4.74 (4.12–5.46)	2994	<.0001	4.87 (4.07–5.83)

Supplementary Table 6. Multivariate Analysis of Treated and Nontreated Patients After Multiple Imputation to Correct for Missing Data Values

	Alkaline phosphatase >2.0× ULN vs <=2.0×ULN after one year follow-up			Bilirubin >1.0× ULN vs <=1.0×ULN after one year follow-up		
Cohorts	HR	95% CI	*P* value	HR	95% CI	*P* value
Entire cohort	2.46	2.16–2.80	<.0001	4.80	4.13–5.57	<.0001
UDCA-treated patients	2.49	2.15–2.88	<.0001	4.95	4.21–5.83	<.0001
UDCA-nontreated patients	2.07	1.45–2.95	<.0001	3.79	2.58–5.59	<.0001