Levels of alkaline phosphatase and bilirubin are surrogate end points of outcomes of patients with primary biliary cirrhosis

Lammers, Willem J; van Buuren, Henk R; Hirschfield, Gideon M; Janssen, Harry L A; Invernizzi, Pietro; Mason, Andrew L; Ponsioen, Cyriel Y; Floreani, Annarosa; Corpechot, Christophe; Mayo, Marlyn J; Battezzati, Pier M; Parés, Albert; Nevens, Frederik; Burroughs, Andrew K; Kowdley, Kris V; Trivedi, Palak J; Kumagi, Teru; Cheung, Angela; Lleo, Ana; Imam, Mohamad H

DOI:
10.1053/j.gastro.2014.08.029

Accepted Manuscript

Levels of Alkaline Phosphatase and Bilirubin are Surrogate Endpoints of Outcomes of Patients with Primary Biliary Cirrhosis – an International Follow-up Study

Willem J. Lammers, Henk R. van Buuren, Gideon M. Hirschfield, Harry L.A. Janssen, Pietro Invernizzi, Andrew L. Mason, Cyriel Y. Ponsioen, Annarosa Floreani, Christophe Corpechot, Marlyn J. Mayo, Pier M. Battezzati, Albert Parés, Frederik Nevens, Andrew K. Burroughs, Kris V. Kowdley, Palak J. Trivedi, Teru Kumagi, Angela Cheung, Ana Lleo, Mohamad H. Imam, Kirsten Boonstra, Nora Cazzagon, Irene Franceschet, Raoul Poupon, Llorenç Caballeria, Giulia Pieri, Pushpjeet S. Kanwar, Keith D. Lindor, Bettina E. Hansen

PII: S0016-5085(14)01040-3
DOI: 10.1053/j.gastro.2014.08.029
Reference: YGAST 59312

To appear in: *Gastroenterology*
Accepted Date: 20 August 2014

Please cite this article as: Lammers WJ, van Buuren HR, Hirschfield GM, Janssen HLA, Invernizzi P, Mason AL, Ponsioen CY, Floreani A, Corpechot C, Mayo MJ, Battezzati PM, Parés A, Nevens F, Burroughs AK, Kowdley KV, Trivedi PJ, Kumagi T, Cheung A, Lleo A, Imam MH, Boonstra K, Cazzagon N, Franceschet I, Poupon R, Caballeria L, Pieri G, Kanwar PS, Lindor KD, Hansen BE, on behalf of the Global PBC Study Group, Levels of Alkaline Phosphatase and Bilirubin are Surrogate Endpoints of Outcomes of Patients with Primary Biliary Cirrhosis – an International Follow-up Study, *Gastroenterology* (2014), doi: 10.1053/j.gastro.2014.08.029.

Levels of Alkaline Phosphatase and Bilirubin are Surrogate Endpoints of Outcomes of Patients with Primary Biliary Cirrhosis – an International Follow-up Study

Short title:

Alkaline phosphatase and bilirubin as surrogate endpoints in PBC

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], Bettina E. Hansen[1] – **on behalf of the Global PBC Study Group**

[1]Dept 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, UK;[3]Liver Clinic, Toronto Western & General Hospital, University Health Network, Toronto, ON, Canada; [4]Liver Unit and Center for Autoimmune Liver Diseases, Humanitas Clinical and Research Center, Rozzano (MI), Italy; [5]Divison of Gastroenterology and Hepatology, University of Alberta, Edmonton, AB, Canada; [6]Dept 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, UT Southwestern Medical Center, Dallas, TX, USA; [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]Dept of Hepatology, University

Hospitals Leuven, KULeuven, Leuven, Belgium; [13]The Sheila Sherlock Liver Centre, The Royal Free Hospital, London, United Kingdom; [14]Liver Center of Excellence, Digestive Disease Institute, Virginia Mason Medical Center, Seattle, WA, USA;[15]Department of Gastroenterology and Metabology, Ehime University graduate School of Medicine, Ehime, Japan; [16]Dept Gastroenterology and Hepatology, Mayo Clinic, Rochester, MN, USA; [17]Arizona State University, Phoenix, AZ, USA

*Deceased

Funding

This investigator-initiated study was supported by unrestricted grants from Intercept Pharmaceuticals and Zambon Nederland B.V., 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 on the decision to submit the manuscript for publication.

List of abbreviations

HR, hazard ratio; IQR, interquartile range; MELD, model of end stage liver disease; PBC, primary biliary cirrhosis; UDCA, ursodeoxycholic acid; ULN, upper limit of normal

Corresponding author

Bettina E. Hansen, PhD

Erasmus MC, University Medical Center

Department of Gastroenterology and Hepatology

's-Gravendijkwal 230, Room Ha 203

3015 CE Rotterdam, the Netherlands

Tel: +31 10 703 8465 / +31 10 704 4215

Fax: +31 10 436 5916

Email: b.hansen@erasmusmc.nl

Disclosure of potential conflicts of interest

The following authors declared that they have no conflicts of interest **:** W. Lammers, R. Poupon, C. Corpechot, A. Floreani, M. Imam, N. Cazzagon, K. Puhpjeet, K. Boonstra, A. Lleo, G. Pieri, L. Caballeria, I. Franceschet, P. Battezzati, A. Cheung and B. Hansen

H. van Buuren received unrestricted grants from Intercept Pharmaceuticals and Zambon Nederland B.V. G. Hirschfield is study Investigator for Intercept, FalkPharma, Gilead, Lumena, FFP Pharma, Janssen, and he is unpaid consultant for Intercept, Lumena and FFP Pharma. He is supported in part by the UK Medical Research Council Stratified Medicine Grant, UK-PBC. H. Janssen received grants from and is consultant for: Bristol Myers Squibb, Gilead Sciences, Novartis, Roche, Merck, Innogenetics, Abbott, Santaris, Medtronic, Tibotec. P. Invernizzi is consultant for Menarini Diagnostics, Instrumentation laboratories, Medigene. A. Mason received funding from the Canadian Institutes for Health Research, Canadian Liver Foundation, Alberta Heritage Foundation for Medical research, and Alberta Cancer Foundation. Abbvie and Gilead have provided medications for ongoing clinical trials. C. Ponsioen received unrestricted grants from Abbvie, Dr. Falk Pharma, and Takeda. He received consultancy fees from Abbvie, GSK, and Takeda, and speaker's fees from Abbvie, MSD, Takeda, and Ferring. M. Mayo received Grant support in the form of carrying out sponsor-initiated clinical trials: Intercept, Gilead, Salix, Lumena, NGM. A. Pares is consultant for Lumena Pharmaceuticals, Inc. K. Kowdley received grants and research support from AbbVie, Beckman, BMS, Boeringer Ingelheim, Gilead, Ikaria, Intercept, Janssen, Merck, Mochida, Vertex; He received consultancy fees from Novartis and Tekmira (honorarium payable to Institution); and he participated in advisory boards for AbbVie , Boeringer Ingelheim, Gilead,

Ikaria , Janssen , Merck , Trio Health (honorarium payable to Institution). P. Trivedi is the recipient of a Wellcome Trust Clinical Research Fellowship. T. Kumagi received fellowship support from an unrestricted grant from Axcan Pharma. K. Lindor is affliliated as AASLD Governing Board member and primary author PSC guidelines (to be completed this year) and he is unpaid advisor of Intercept Pharmaceuticals and Lumena Pharmaceuticals.

Authors contributions

Willem J. Lammers, MD and Bettina E. Hansen, PhD, had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Study concept and design: Lammers, van Buuren, Hirschfield, Janssen, Invernizzi, Mason, Ponsioen, Floreani, Corpechot, Mayo, Battezzati, Parés, Nevens, Burroughs, Kowdley, Trivedi, Kumagi, Cheung, Lleo, Imam, Boonstra, Cazzagon, Franceschet, Poupon, Caballeria, Pieri, Kanwar, Lindor, Hansen

Acquisition of data: Lammers, van Buuren, Hirschfield, Janssen, Invernizzi, Mason, Ponsioen, Floreani, Corpechot, Mayo, Battezzati, Parés, Nevens, Burroughs, Kowdley, Trivedi, Kumagi, Cheung, Lleo, Imam, Boonstra, Cazzagon, Franceschet, Poupon, Caballeria, Pieri, Kanwar, Lindor, Hansen

Analysis and interpretation of data: Lammers, van Buuren, Hirschfield, Invernizzi, Mason, Hansen

Drafting of the manuscript: Lammers, van Buuren, Hirschfield, Invernizzi, Mason, Hansen

Critical revision of the manuscript for important intellectual content: Lammers, van Buuren, Hirschfield, Janssen, Invernizzi, Mason, Ponsioen, Floreani, Corpechot, Mayo, Battezzati, Parés, Nevens, Burroughs, Kowdley, Trivedi, Kumagi, Cheung, Lleo, Imam, Boonstra, Cazzagon, Franceschet, Poupon, Caballeria, Pieri, Kanwar, Lindor, Hansen

Statistical analysis: Lammers, Hansen

Administrative, technical or material support: NA

Obtained funding: van Buuren, Hansen

Study supervision Lammers, van Buuren, Hirschfield, Janssen, Invernizzi, Mason, Ponsioen, Floreani, Corpechot, Mayo, Battezzati, Parés, Nevens, Burroughs, Kowdley, Trivedi, Kumagi, Cheung, Lleo, Imam, Boonstra, Cazzagon, Franceschet, Poupon, Caballeria, Pieri, Kanwar, Lindor, Hansen

ABSTRACT

Background & Aims: Non-invasive surrogate endpoints 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 of cohort studies to evaluate whether patient levels of alkaline phosphatase and bilirubin correlate with their outcomes and can be used as surrogate endpoints.

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 sub-populations, at different time points relative to the clinical endpoint (liver transplantation or death).

Results: Of the 4845 patients, 1118 reached a clinical endpoint. The median follow-up time 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 values associated with longer transplant-free survival). At 1 year after study enrollment, levels of alkaline phosphatase 2.0-fold 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 ≤2.0-fold the ULN, 84% survived for 10 years, compared to 62% of those with levels >2.0-fold 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-fold the ULN best predicted patient transplant-free survival (C statistic, 0.79). Of patients with bilirubin levels ≤1.0-fold the ULN, 86% survived for 10 years after study

enrollment, compared with 41% of those with levels >1.0-fold 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 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 endpoints in therapy trials.

KEYWORDS: autoimmune liver disease; response to treatment; biomarker; new therapies.

INTRODUCTION

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

A major challenge for patients, healthcare providers and drug developers, however, is the slowly progressive nature of the disease which effectively precludes the evaluation of classical clinical outcome, such as transplant-free survival. The low prevalence of PBC also represents a significant barrier to the conduct of large controlled clinical outcome trials in PBC. Hence while clinical endpoints such as liver transplantation and death have been evaluated in an early primary interventional trial in PBC with UDCA,[9] the majority of patients are now diagnosed at an earlier stage of disease and started on UDCA therapy shortly after diagnosis, further impacting 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 datasets that are representative of the spectrum of disease globally, and sufficiently powered through size, duration of follow-up and numbers of clinical events in order to refine the scientific validity of specific biochemical surrogates.

Serum bilirubin is well established as 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 those patients with relatively advanced disease are likely to demonstrate meaningful changes in bilirubin values that are stratifying. A biochemical variable and potentially more broadly applicable surrogate endpoint is alkaline phosphatase, an isoenzyme involved in dephosphorylation.[17] Elevated 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 values and the risk for adverse outcome 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 the use in PBC clinical studies.

We sought to investigate how serum alkaline phosphatase and bilirubin values, 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 endpoints 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.

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 with UDCA treatment.

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

Data collection and quality assessment

Collected clinical and laboratory data included gender; age; PBC diagnosis; liver histology; treatment (type of medication, dosage and duration); duration and last date of follow-up; baseline antimitochondrial antibody (AMA) status; baseline and yearly laboratory values (serum alkaline phosphatase, total bilirubin, albumin, AST, ALT, γ-GT and platelets); outcomes (death and cause of death, liver transplantation, hepatocellular carcinoma, ascites and variceal bleeding).

Liver histology performed within one year of study entry, or documented cirrhosis prior to study entry, were 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 values. According to this biochemical classification early stage was defined if both bilirubin and albumin were normal; moderately advanced disease if one of these criteria were abnormal and advanced disease if both were abnormal.[22]

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

Statistical analysis

The study entry was defined as start date of UDCA therapy or date of first center visit for UDCA non-treated patients. The primary endpoint was defined as a composite of either liver transplantation or death. Patients without documented events during the course of the follow-up were censored at their last follow-up visit.

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

To find an optimal threshold for each variable, alkaline phosphatase and bilirubin values at one year follow-up were categorized according to multiple thresholds ranging from 1.0-3.0x the upper limit of normal (ULN), in steps of 0.1 (and also including 1.67xULN[7] for alkaline phosphatase values). The C statistic was calculated for each of these thresholds to evaluate their ability to predict liver transplant-free survival. Accompanying HR were calculated for each threshold using 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 values 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 values of alkaline phosphatase, the percentage change in alkaline phosphatase values[4] from baseline to one year follow-up was evaluated using a 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 treatment and gender.

To investigate if alkaline phosphatase and bilirubin are meaningful surrogate endpoints the association with the clinical endpoint 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 analysed were baseline and yearly, up to five years follow-up. Given the nature of this study, alkaline phosphatase or bilirubin 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 endpoints at baseline and several intervals thereafter, up to five years. Subgroups were defined by treatment (UDCA treated and non-treated patients); baseline alkaline phosphatase values (>2.0xULN and >4.0xULN); baseline bilirubin values (>1xULN and >3xULN); PBC disease state based on both histology and biochemistry; age at time of diagnosis (<45 years and ≥45 years)[23]; gender; and date of diagnosis (<1990, 1990-1999, 2000-2009).

Normally distributed data was presented as mean ± standard deviation and skewed distributed data as median and interquartile range (IQR). All analyses were two-sided. A p-value <0.05 was considered statistically significant. Statistical analyses were performed with IBM SPSS Statistics 22.0 (SPSS Inc, Chicago, IL, USA) and SAS 9.3.

RESULTS

Baseline data

Data were obtained from a total of 6191 individuals 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 patients had been diagnosed with PBC ranged from 1959 to 2012. The diagnosis was established after 1990 for 79% of patients, while the median year of diagnosis was 1998 (IQR 1991-2004). The median follow-up was 7.3 (IQR 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. The majority, 4119 (85%) patients, were treated with UDCA at a median dosage of 12.3 mg/kg/day (9.4-14.6). Histological stage of disease was available for 76% of patients who had undergone a liver biopsy; the majority having been diagnosed with early disease (stage I or II).

During follow-up 1118 patients reached a clinical endpoint; 389 underwent liver transplantation and 729 died, of which 358 (49%) patients from liver-related causes and 245 (34%) from other causes. The cause of death was unknown for 126 (17%) patients. In the total cohort, the 5-year transplant-free survival was 88%, 10-year survival was 77% and 15-year survival was 63%. For UDCA treated patients, this was 90%, 78% and 66% respectively and for non-treated patients 79%, 59% and 32% respectively (treated versus non-treated, $P<0.0001$).

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

The association between alkaline phosphatase and bilirubin values and the risk of liver transplantation or death

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

Optimal threshold for alkaline phosphatase and bilirubin values and the risk of liver transplantation and death

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

After one year follow-up, while all thresholds were predictive of outcomes, a threshold of 2.0xULN for alkaline phosphatase values was found to have the highest predictive ability (C statistic, 0.71; 95% CI 0.69-0.73). Notably, this threshold was not a significantly better predictor than the other thresholds, such as 1.5xULN[8], 1.67xULN[7, 19], or 3.0xULN[5] (Table 2 and Supplementary Figure 2). Similarly, all assessed bilirubin thresholds were predictive of outcomes. For bilirubin a threshold of 1.0xULN 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 for patients with alkaline phosphatase values ≤2.0xULN was 94%, 84% and 73% respectively and for patients with alkaline phosphatase values >2.0xULN this was 81%, 62% and 50% respectively (P<0.0001), as shown in Figure 3A.

The accompanying 5-, 10- and 15-year transplant-free survival for patients with normal bilirubin after one year follow-up was 95%, 86% and 74% respectively and for patients with abnormal bilirubin values 65%, 41% and 30% respectively (P<0.0001) (Figure 3B).

The prognostic information provided by alkaline phosphatase remained important in addition to bilirubin; the risk of liver transplantation or death of patients with alkaline phosphatase >2.0xULN was significantly higher in both, those patients with normal (≤1xULN) and abnormal bilirubin (>1xULN). The 5-, 10- and 15-year transplant-free survival in the normal bilirubin group for patients with alkaline phosphatase values ≤2.0xULN was 97%, 89% and 79% respectively, and for patients with alkaline phosphatase values >2.0xULN this was 95%, 82% and 68% respectively (P<0.0001). In the abnormal bilirubin group this was 74%, 51% and 39% respectively for patients with alkaline phosphatase values ≤2.0xULN and 63%, 34% and 24% respectively for patients with alkaline phosphatase values >2.0xULN (P<0.0001) (Figure 3C).

An alkaline phosphatase threshold of 2.0xULN was also predictive in addition to other bilirubin thresholds between 1.0xULN and 3.0xULN, but was not predictive in addition to bilirubin values >3xULN (HR 0.71; 95% CI 0.39-1.32; P=0.29).

Comparable results were found for other alkaline phosphatase thresholds (e.g. 1.5xULN and 1.67xULN in combination with normal or abnormal bilirubin (data not shown).

The predictive value of percentage changes in alkaline phosphatase values at one year follow-up

A prior study showed that patients who achieved a normal alkaline phosphatase or had >40% decrease in alkaline phosphatase values after UDCA treatment had a normal prognosis.[4] In line with this study, the percentage change in alkaline phosphatase values from baseline to one year follow-up was predictive of outcome; the greater the percentage alkaline phosphatase decrease, the better the transplant-free survival (HR per 10% alkaline phosphatase change 0.98 [95% CI 0.96-0.99], P<0.01).

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

However, the percentage decrease in alkaline phosphatase values did not add prognostic information to absolute alkaline phosphatase values after one year follow-up (HR per 10% alkaline phosphatase change: 1.00 [95% CI 0.99-1.02], P=0.72), apart from very high alkaline phosphatase values (>5.0xULN) (HR per 10% alkaline phosphatase change: 0.86 [95% CI 0.76-0.96], P<0.005).

Predictive ability of alkaline phosphatase and bilirubin values across subgroups

To assess if alkaline phosphatase can be used as a predictor independent of patient characteristics, the above analyses were repeated for a range of subgroups (Figure 4A). Of note, using an alkaline phosphatase threshold of 2.0xULN after one year of follow-up, was not only predictive for UDCA-treated patients, but also for non-treated patients. Similar results were seen in patients with baseline alkaline phosphatase values >2.0xULN and >4.0xULN; patients with histologically early and late disease; patients with biochemically early and moderately advanced disease; patients ≤45 years old at diagnosis and >45 years at diagnosis; males and females; and regardless of year of diagnosis. alkaline phosphatase values were not predictive for patients with biochemical advanced disease, i.e. those patients with both abnormal bilirubin and albumin. A bilirubin threshold of 1.0xULN after one year follow-up, was also predictive in several subsets of patients (Figure 4B).

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

Translation into clinical practice

For illustrative purposes the above findings were translated into a practical example (Figure 5) to demonstrate the association of a composite surrogate endpoint (bilirubin <1xULN and alkaline phosphatase <threshold) on 5-year liver transplantation-free survival in different settings. Three groups of high-risk PBC patients diagnosed after 1990 and initiated on UDCA treatment were defined at two different time points, at baseline (upper panels) and after one year of UDCA treatment (lower panels). The subgroups were defined as follows 1). all PBC patients, 2). patients meeting the inclusion criteria of a recent clinical trial: bilirubin <2xULN and (alkaline phosphatase >1.67xULN or bilirubin >1xULN)[24] and 3). patients with bilirubin <3xULN and (alkaline phosphatase >2xULN or bilirubin >1xULN). The surrogate endpoint was determined one year after inclusion. Figure 5 shows the proportion of patients reaching the surrogate endpoint (left panels) and accompanying transplant-free survival (right panels).
If bilirubin <1xULN and alkaline phosphatase <2xULN is used as a surrogate endpoint in high-risk PBC population 3 (light grey) and if patients are already treated with UDCA for a year (lower panels) the proportion of patients reaching the surrogate endpoint after one additional year of UDCA treatment is 18% (lower left panel) with an accompanying 5-year transplantation-free survival of 92% (lower right panel). The 5-year transplantation-free survival for patients not reaching the surrogate endpoint was 75%.

In summary, using higher alkaline phosphatase thresholds resulted in a lower proportion of patients not reaching the surrogate endpoint with a poorer corresponding 5-year transplant-free survival. The 5-year liver transplant-free survival following continued UDCA treatment 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 PBC individual data meta-analysis assembled to date. We unequivocally demonstrate that both increased serum alkaline phosphatase and bilirubin values are strongly associated with reduced transplant-free survival in 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. This data support that both, alkaline phosphatase and bilirubin, provide meaningful surrogate endpoints in PBC that can reasonably be used in clinical trials.

Prior studies have demonstrated an association between normalisation, percentage decreases or absolute decreases of alkaline phosphatase values and improved prognosis upon treatment with UDCA.[4, 5, 7, 8, 18, 19] The present study reports for the first time a near log-linear association between alkaline phosphatase values 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 values during follow-up but also to baseline values, irrespective of subsequent UDCA treatment. The suitability of alkaline phosphatase as a surrogate endpoint for clinical benefit is further supported by the finding that the prognostic information provided by alkaline phosphatase values was confirmed across a wide range of subgroups such as non-UDCA treated patients, relatively young patients and patients with histologically early and late disease. This finding is of considerable clinical significance as alkaline phosphatase constitutes one of the three potential diagnostic criteria and is used routinely to assess disease activity.

Our study additionally confirms that as baseline bilirubin values or bilirubin values over time increase, the likelihood of survival correspondingly decreases.[10] The predictive ability of alkaline phosphatase values was demonstrated in addition to bilirubin to discriminate high and low-risk patients.

This is an important observation because bilirubin on its own is unsuitable as surrogate endpoint in clinical trials as it is typically elevated only when the disease has progressed to the stage where liver function becomes impaired. Most patients likely to be included in such studies will have normal values, 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 endpoint, might be a meta-analysis of clinical trials of one or more interventions.[25] A four-level hierarchy of evidence to consider the validation of surrogate endpoints was proposed.[26] '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.'

The particular challenges in PBC to confirm biomarkers as surrogate endpoints is that there is just one approved treatment, and previous meta-analyses on 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 collected from existing cohorts of patients at centers across North America and Europe with long-term follow-up data of large numbers of PBC patients. This design has sufficient power to intensively study alkaline phosphatase and bilirubin as potential surrogate endpoints in different settings, sub-populations and at different time points. Based on these current results we postulate that alkaline phosphatase and bilirubin 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 PBC. Non-invasive 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 PBC patients as 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 PT, IgM and IgG values) in the individual databases varied considerably. In many cases, in particular when databases contained data of patients entered more than 10-20 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 was not always available at the fixed time points during follow-up. This was mainly encountered when the original data had been collected more than twenty years ago. Dose changes or interruption of UDCA treatment 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 non-treated patients, the substantial incidence of clinical endpoints and the duration of follow-up. Additionally, adjusting for missing data by multiple imputation of the data the results did not actually change (Supplementary Table 5).

Based on our present results any decrease in alkaline phosphatase or bilirubin values translates into improved prognosis; lower values are clearly associated with better transplant-free survival. In our population the most discriminative alkaline phosphatase threshold after one year follow-up was 2.0xULN, an earlier proposed threshold,[18] although an alkaline phosphatase threshold of 1.5xULN,[8] 1.67xULN,[7, 19] or 3.0xULN will all work well as a surrogate endpoint 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 <1.0xULN is reasonable. However, designing clinical trials implies the a priori requirement to estimate the quantitative effect of an experimental intervention on a given endpoint. Based on the current study we are not able to translate this data into a specific threshold for a clinical trial in general.

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

REFERENCES

1. Kaplan MM, Gershwin ME. Primary biliary cirrhosis. N Engl J Med 2005;353:1261-73.
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 L. EASL Clinical Practice Guidelines: management of cholestatic liver diseases. J Hepatol 2009;51:237-67.
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;130:715-20.
5. Corpechot C, Abenavoli L, Rabahi N, et al. Biochemical response to ursodeoxycholic acid and long-term prognosis in primary biliary cirrhosis. Hepatology 2008;48:871-7.
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;136:1281-7.
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;105:2186-94.
8. Corpechot C, Chazouilleres O, Poupon R. Early primary biliary cirrhosis: biochemical response to treatment and prediction of long-term outcome. J Hepatol 2011;55:1361-7.
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;330:1342-7.
10. Shapiro JM, Smith H, Schaffner F. Serum bilirubin: a prognostic factor in primary biliary cirrhosis. Gut 1979;20:137-40.
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;30:865-9.
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;60:646-9.
15. Kamath PS, Wiesner RH, Malinchoc M, et al. A model to predict survival in patients with end-stage liver disease. Hepatology 2001;33:464-70.
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;13:926-37.
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;19:115-21.
19. Momah N, Silveira MG, Jorgensen R, et al. Optimizing biochemical markers as endpoints for clinical trials in primary biliary cirrhosis. Liver Int 2012;32:790-5.
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;379:103-12.
21. Scheuer P. Primary biliary cirrhosis. Proc R Soc Med 1967;60:1257-60.
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;101:2044-50.

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 e13-4.
24. http://clinicaltrials.gov/ct2/show/NCT01473524.
25. Fleming TR, Powers JH. Biomarkers and surrogate endpoints in clinical trials. Stat Med 2012;31:2973-84.
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;354:1053-60.
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;101:1529-38.
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;113:884-90.
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;49:1017-44.
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;41:310-7.

Author names in bold designate shared co-first authors

FIGURE LEGENDS

Figure 1. Flow chart

Flow diagram of patients included in this study.

Figure 2. Hazard of liver transplantation or death for alkaline phosphatase and bilirubin values

The hazard of liver transplantation or death for (A) alkaline phosphatase values and (B) bilirubin values at different time points estimated with cublic spline function.

Figure 3. Liver transplantation-free survival

(A) Transplant-free survival of patients with alkaline phosphatase values ≤2.0xULN versus >2.0xULN at one year follow-up. (B) Transplant-free survival of patients with bilirubin values ≤1.0xULN versus >1.0xULN at one year follow-up. (C) Transplant-free survival of patients with alkaline phosphatase ≤2.0xULN versus >2.0xULN at one year follow-up within both, patients with bilirubin values ≤1xULN and >1xULN.

Figure 4. Subgroup analyses of alkaline phosphatase and bilirubin values

Hazard ratio of liver transplantation or death for (A) alkaline phosphatase values >2.0xULN versus ≤2.0xULN and (B) bilirubin values >1.0xULN versus ≤1.0xULN at one year follow-up for different subgroups.

Figure 5. Translation into clinical practice

The association of a surrogate endpoint, defined as alkaline phosphatase <threshold and bilirubin <1xULN, on the 5-year liver transplantation-free survival in different settings. Inclusion (diagnosed

>1990 and initiated on UDCA) was made at baseline (upper panels) and after one year on UDCA treatment (lower panels). Three high risk groups were defined as follows: 1) all patients (black lines), 2) bilirubin <2xULN and (alkaline phosphatase >1.67xULN or bilirubin >1xULN)[24] (dark grey lines), and 3) bilirubin <3xULN and (alkaline phosphatase >2xULN or bilirubin >1xULN) (light grey lines). The full lines represent the patients who reached the surrogate endpoint and the dotted lines those who did not. The left panels show the proportion of patients reaching the surrogate endpoint 1 year after inclusion and the right panels the corresponding 5-year transplant-free survival.

TABLES

Table 1. **Baseline patient characteristics**

	Total cohort (n = 4845)
Age at entry (years)[a]	54.5 (12.0)
Female, n (%)	4348 (90%)
AMA+, n (%)	4280 (88%)
Year of diagnosis[b]	1998 (1991-2004)
Year of diagnosis, time frame	1959-2012
Histological disease stage, n (%)[c]	
Stage I	1017 (27%)
Stage II	862 (23%)
Stage III	483 (13%)
Stage IV	454 (12%)
Not available	953 (25%)
Biochemical disease stage, n (%)[d]	
Early	2040 (42%)
Moderately advanced	730 (15%)
Advanced	259 (5%)
Not available	1816 (38%)
Baseline alkaline phosphatase values, n (%)	
>2.0xULN	1931 (52%)
>4.0xULN	816 (22%)
Not available	1140 (24%)
UDCA treated patients, n (%)[e]	4119 (85%)
Laboratory data at entry[b]	
Serum bilirubin (xULN)	0.67 (0.45-1.06)
Not available	1118 (23%)
Serum alkaline phosphatase (xULN)	2.10 (1.31-3.72)
Not available	1140 (24%)

Abbreviations: AMA, antimitochondrial antibody; UDCA, ursodeoxycholic acid; ULN, upper limit of normal.

[a]Data is expressed as mean and standard deviation

[b]Data is expressed as median and interquartile range

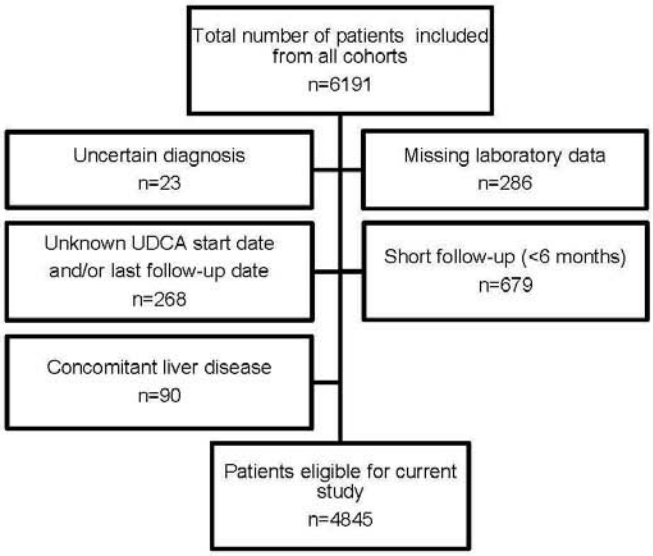
[c]Histological disease stage according to Ludwig and Scheuer's classification

[d]Biochemical disease stage according to Rotterdam criteria (using albumin and bilirubin)[22]

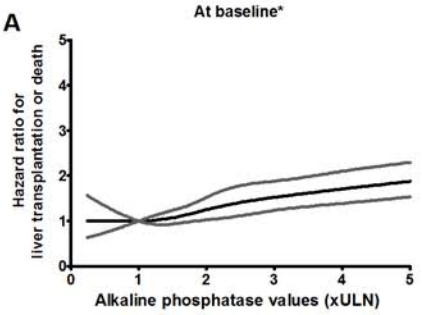
[e]640 subjects were non-treated and 86 subjects were without definitive information on UDCA use

Table 2. Performance of different alkaline phosphatase and bilirubin thresholds at one year follow-up for prediction liver transplantation or death

	Alkaline phosphatase (n=3710)				Bilirubin (n=3681)			
Thresholds	**C statistic (95% CI)**	**Hazard ratio (95% CI)**	**P-value**	**Number of patients ≤/> threshold**	**C statistic (95% C()**	**Hazard ratio (95% CI)**	**P-value**	**Number of patients ≤/> threshold**
1.0xULN	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**
1.1xULN	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
1.2xULN	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
1.3xULN	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
1.4xULN	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
1.5xULN	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
1.6xULN	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
1.67xULN	0.70 (0.69-0.72)	2.18 (1.88-2.53)	<.001	2231/1479				
1.7xULN	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
1.8xULN	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
1.9xULN	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
2.0xULN	**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
2.1xULN	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
2.2xULN	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
2.3xULN	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
2.4xULN	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
2.5xULN	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
2.6xULN	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
2.7xULN	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
2.8xULN	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
2.9xULN	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
3.0xULN	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

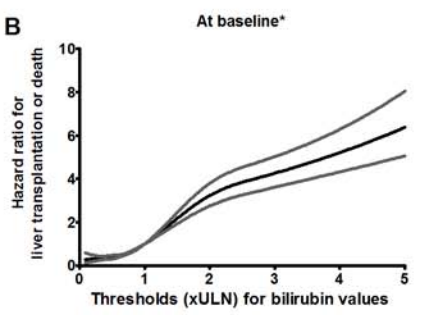

Total number of patients included from all cohorts
n=6191
Uncertain diagnosis
n=23
Missing laboratory data
n=286
Unknown UDCA start date and/or last follow-up date
n=268
Short follow-up (<6 months)
n=679
Concomitant liver disease
n=90
Patients eligible for current study
n=4845

A

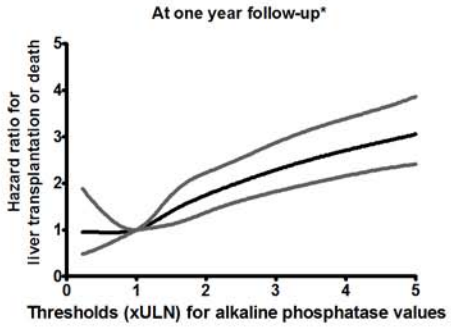


*3705/4845 patients were included for this analysis

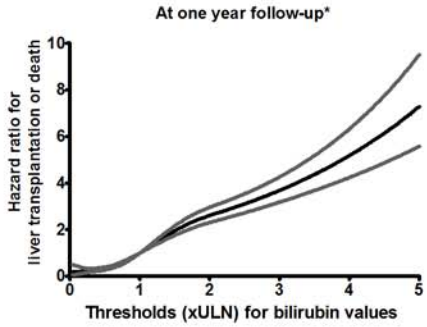
B



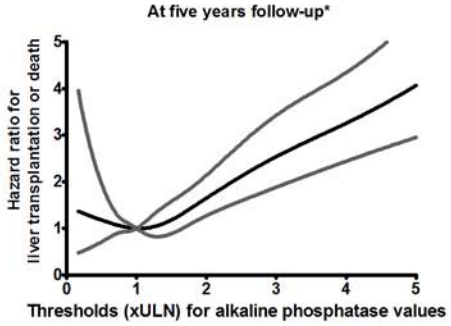
*3727/4845 patients were included for this analysis



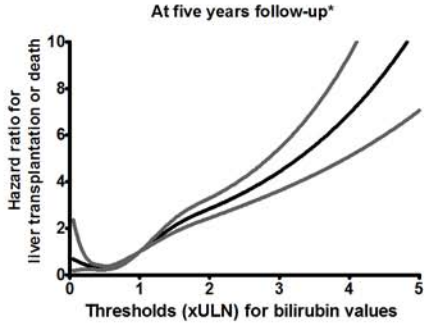
*3710/4635 patients were included for this analysis



*3681/4635 patients were included for this analysis

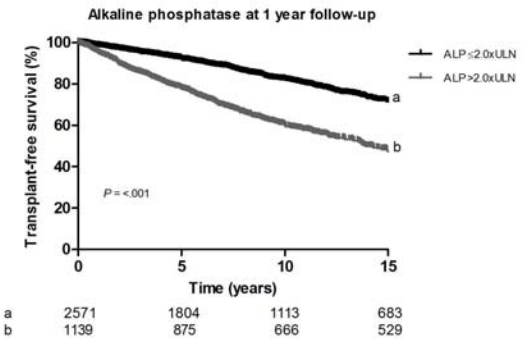


*2203/3161 patients were included for this analysis



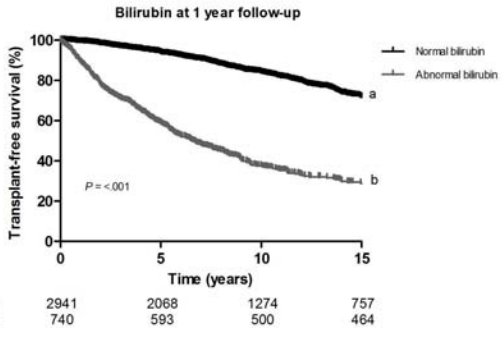
*2109/3161 patients were included for this analysis

A



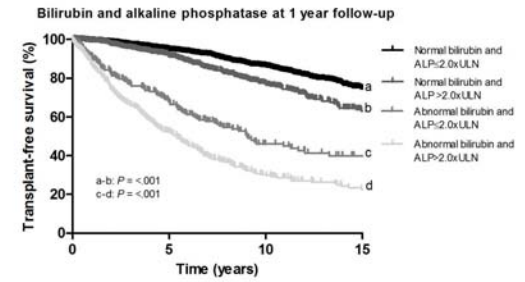
a	2571	1804	1113	683
b	1139	875	666	529

B



a	2941	2068	1274	757
b	740	593	500	464

C

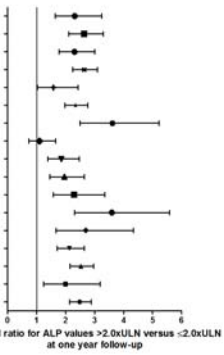


a	2112	1482	887	504
b	681	489	337	228
c	271	193	153	137
d	400	345	302	283

abbreviation: ALP, alkaline phosphatase

A

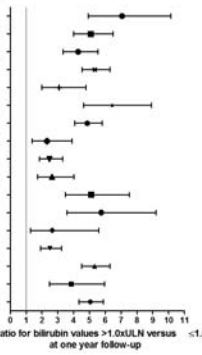
Subgroups	N=	Hazard ratio (95% CI)	*P* - value
Year of diagnosis 2000 -2008	1463	2.32 (1.65-3.25)	< .0001
Year of diagnosis 1990 -1999	1364	2.64 (2.11-3.31)	< .0001
Year of diagnosis <1990	708	2.31 (1.78-3.01)	< .0001
Female	3342	2.64 (2.25-3.10)	< .0001
Male	368	1.58 (1.03-2.44)	.04
Age at diagnosis >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	582	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 values >4.0xULN	704	2.70 (1.68-4.34)	< .0001
Baseline alkaline phosphatase values >2.0xULN	1669	2.13 (1.72-2.65)	< .0001
UDCA treated	3249	2.53 (2.15-2.97)	< .0001
UDCA non-treated	418	2.00 (1.25-3.19)	.004
Entire cohort	3710	2.49 (2.14-2.89)	< .0001



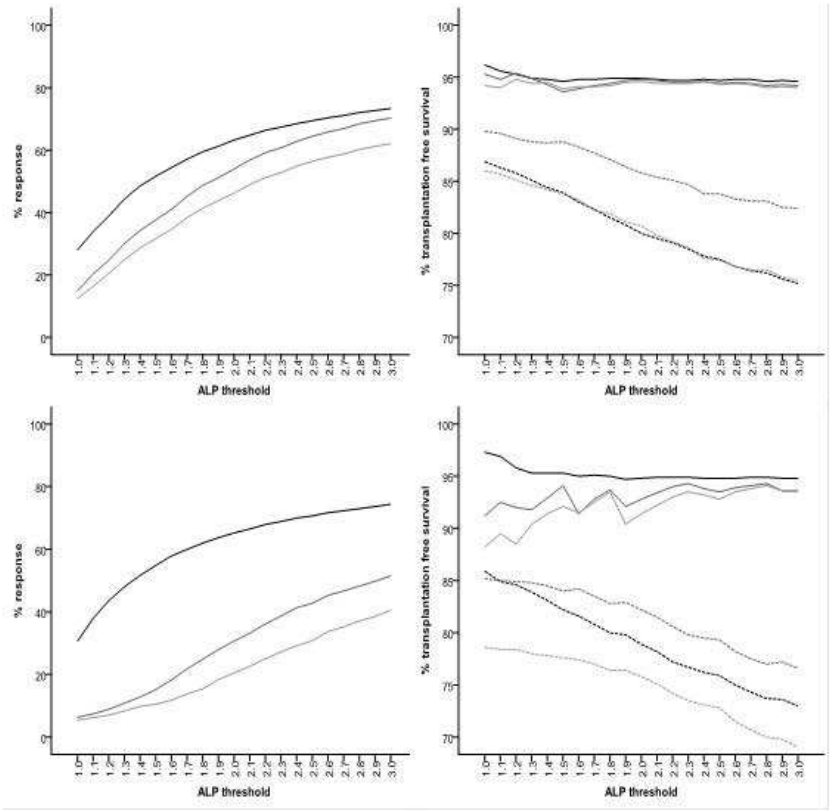
Hazard ratio for ALP values >2.0xULN versus ≤2.0xULN at one year follow-up

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	1378	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 values >3.0xULN	178	2.65 (1.26 -5.58)	.01
Baseline bili rubin values >1.0xULN	819	2.50 (1.92 -3.25)	< .0001
UDCA treated	3181	5.32 (4.50 -6.29)	< .0001
UDCA non-treated	459	3.84 (2.47 -5.96)	< .0001
Entire cohort	3681	5.06 (4.34 -5.89)	< .0001



Hazard ratio for bilirubin values >1.0xULN versus ≤1.0xULN at one year follow-up


% response
ALP threshold
% transplantation free survival
ALP threshold
% response
ALP threshold
% transplantation free survival
ALP threshold

SUPPLEMENTARY MATERIAL

Supplementary figures

Supplementary Figure 1. Hazard of liver transplantation or death for alkaline phosphatase and bilirubin values.

Supplementary Figure 2. Performance of thresholds for alkaline phosphatase values.

Supplementary Figure 3. Liver transplant-free survival for percent decrease of alkaline phosphatase values.

Supplementary tables

Supplementary Table 1. Center specific characteristics of the study population.

Supplementary Table 2. Univariable and multivariable analysis showing the effects of variables at baseline predictive for liver transplantation and death.

Supplementary Table 3. Hazard ratio of percentage change of alkaline phosphatase values from baseline to one year follow-up.

Supplementary Table 4. Hazard ratio of liver transplantation or death for (A) alkaline phosphatase values >2.0xULN versus ≤2.0xULN and (B) bilirubin values >1.0xULN versus ≤1.0xULN at baseline and two years follow-up for different subgroups.

Supplementary Table 5. Multivariate analysis of treated and non-treated patients following multiple imputation to correct for missing data values.

Supplementary figures

Supplementary Figure 1. Hazard of liver transplantation or death for alkaline phosphatase and bilirubin values.

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

Supplementary Figure 2. **Performance of alkaline phosphatase thresholds.**

C statistic was performed for different thresholds for alkaline phosphatase values at one year 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. Liver transplant-free survival for percent decrease of alkaline phosphatase values.

Transplant-free survival for percent decrease of alkaline phosphatase values at one year follow-up. (A) Transplant-free survival of patients with >40% decrease of alkaline phosphatase values, (B) Transplant-free survival of patients with 0-40% decrease of alkaline phosphatase values and (C) Transplant-free survival for patients with no decrease of alkaline phosphatase values.

Supplementary tables

Supplementary Table 1. Center specific characteristics of the study population.

Center		Year of diagnosis		Follow-up (in years)		UDCA		Alkaline phosphatase at entry	Bilirubin at entry
Country (city)	N	median	Time frame	Median (IQR)	Range	N	%	Median (IQR)	Median (IQR)
USA, (*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)
UK, (*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)
French, (*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)
USA, (*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)
UK, (*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)

Supplementary Table 2 (continued). Center specific characteristics of the study population.

USA, (*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)

Supplementary Table 2. Univariable and multivariable analysis showing the effects of variables at baseline predictive for liver transplantation and death.

	Univariable analysis		Multivariable analysis	
	Hazard ratio (95% CI)	***P*-value**	**Hazard ratio (95% CI)**	***P*-value**
Year of diagnosis (year)	0.95 (0.94-0.95)	<0.001	0.95 (0.94-0.96)	<0.001
Age at study entry (year)	1.04 (1.03-1.04)	<0.001	1.03 (1.03-1.04)	<0.001
UDCA treatment	0.59 (0.50-0.71)	<0.001	0.61 (0.51-0.74)	<0.001
Male gender	1.52 (1.28-1.80)	<0.001	1.46 (1.22-1.75)	<0.001

Supplementary Table 3. Hazard ratio for predicting liver transplantation or death for percentage change of alkaline phosphatase values from baseline to one year follow-up.

Percentage reduction of alkaline phosphatase	Hazard Ratio (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. Hazard ratio of liver transplantation or death for (A) alkaline phosphatase values >2.0xULN versus ≤2.0xULN and (B) bilirubin values >1.0xULN versus ≤1.0xULN at baseline and two years follow-up for different subgroups.

	Alkaline phosphatase >2.0xULN versus ≤2.0xULN At baseline			**Alkaline phosphatase >2.0xULN versus ≤2.0xULN At 2 years 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 <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 >45 years	2847	<.0001	1.84 (1.56-2.17)	2264	<.0001	2.35 (1.93-2.87)
Age at diagnosis ≤45 years	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 values >4.0xULN				557	<.0001	2.87 (1.80-4.58)
Baseline alkaline phosphatase values >2.0xULN				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 non-treated	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)

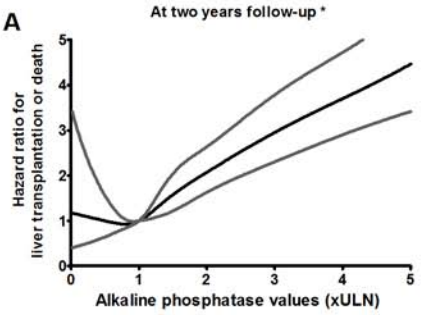
Supplementary Table 4 (continued). Hazard ratio of liver transplantation or death for (A) alkaline phosphatase values >2.0xULN versus ≤2.0xULN and (B) bilirubin values >1.0xULN versus ≤1.0xULN at baseline and two years follow-up for different subgroups.

	Bilirubin >1.0xULN versus ≤1.0xULN At baseline			**Bilirubin >1.0xULN versus ≤1.0xULN At 2 years 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 <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 >45 years	901	<.0001	4.27 (3.64-5.01)	2225	<.0001	4.23 (3.44-5.21)
Age at diagnosis ≤45 years	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 values >3.0xULN				116	.35	1.51 (0.64-3.53)
Baseline bilirubin values >1.0xULN				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 non-treated	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 5. Multivariate analysis of treated and non-treated patients following multiple imputation to correct for missing data values.

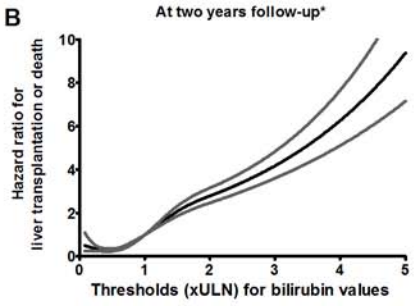
	Alkaline phosphatase >2.0xULN			**Bilirubin >1.0xULN**		
Cohorts	Hazard ratio	95% CI	P-value	Hazard ratio	95% CI	*P*-value
Entire cohort	2.46	2.16-2.80	<.0001	4.80	4.13-5.57	<.0001
UDCA treated patients	2.49	2.15-2.88	<.0001	4.95	4.21-5.83	<.0001
UDCA untreated patients	2.07	1.45-2.95	<.0001	3.79	2.58-5.59	<.0001

A

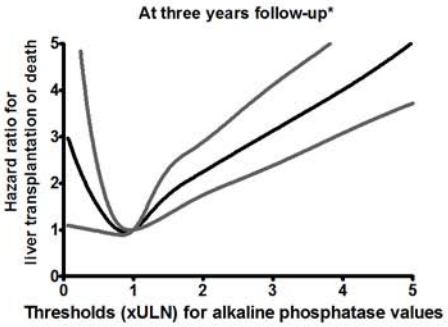


*3018/4257 patients were included for this analysis

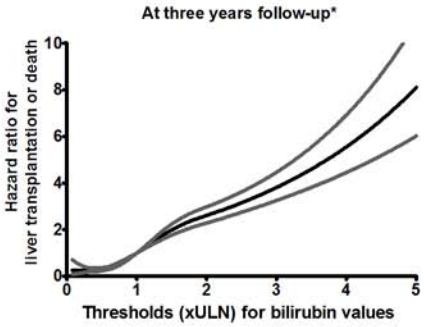
B



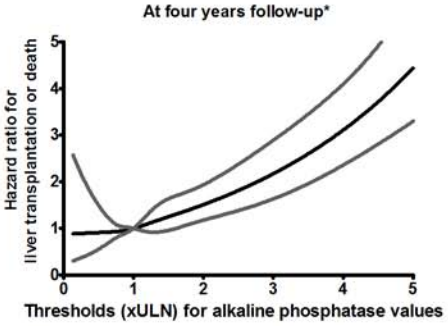
*2994/4257 patients were included for this analysis



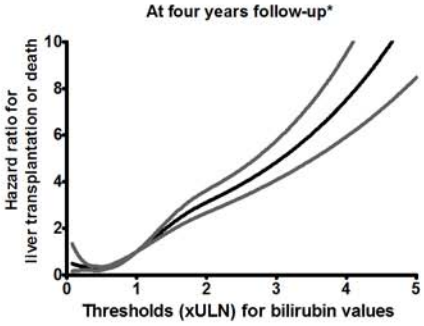
*2573/3865 patients were included for this analysis



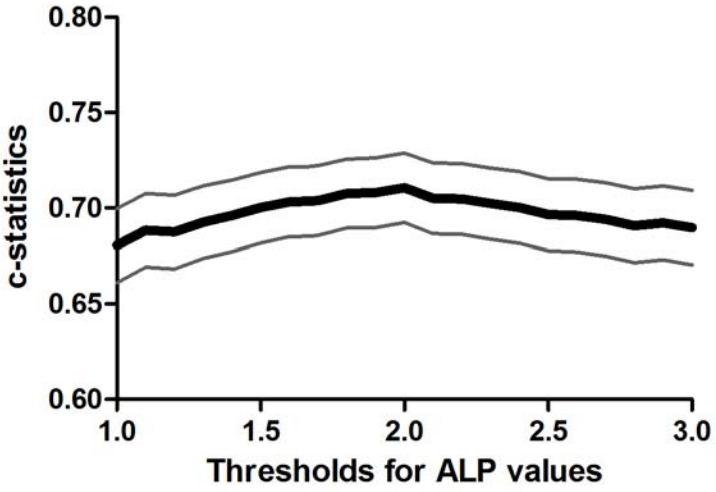
*2522/3865 patients were included for this analysis

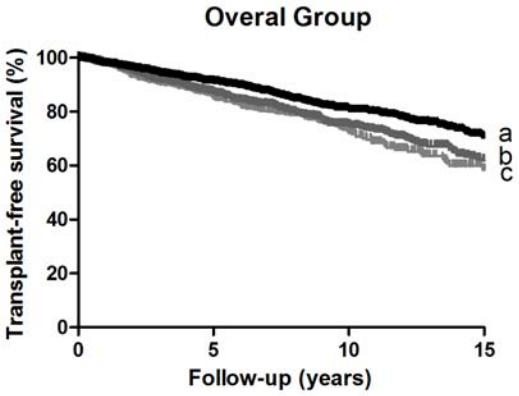


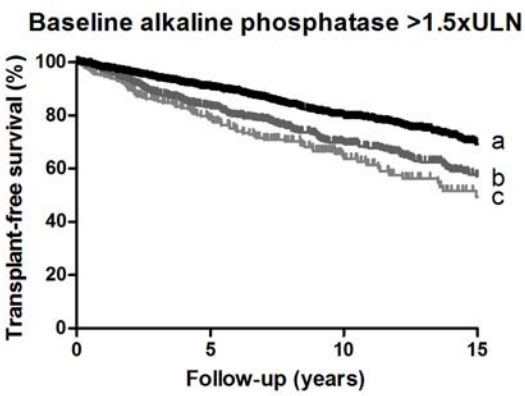
*2424/3495 patients were included for this analysis

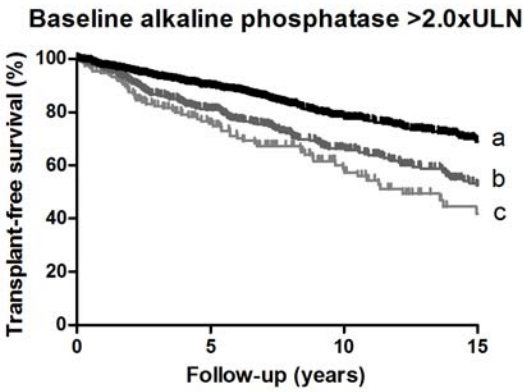


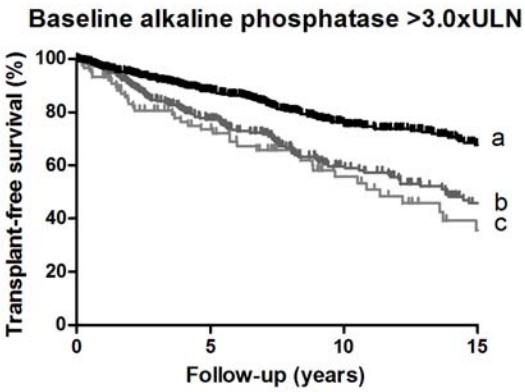
*2360/3495 patients were included for this analysis




Overal Group
Transplant-free survival (%)
100
80
60
40
20
0
0
5
10
15
Follow-up (years)
a
b
c


Baseline alkaline phosphatase >1.5xULN
Transplant-free survival (%)
100
80
60
40
20
0
0
5
10
15
Follow-up (years)
a
b
c


Baseline alkaline phosphatase >2.0xULN
Transplant-free survival (%)
100
80
60
40
20
0
0
5
10
15
Follow-up (years)
a
b
c


Baseline alkaline phosphatase >3.0xULN
Transplant-free survival (%)
100
80
60
40
20
0
0
5
10
15
Follow-up (years)
a
b
c