Pretreatment prediction of response to ursodeoxycholic acid in primary biliary cholangitis: development and validation of the UDCA Response Score

Marco Carbone, Alessandra Nardi*, Steve Flack, Guido Carpino, Nikoletta Varvaropoulou, Caius Gavrila, Ann Spicer, Jonathan Badrock, Francesca Bernuzzi, Vincenzo Cardinale, Holly F Ainsworth, Michael A Heneghan, Douglas Thorburn, Andrew Bathgate, Rebecca Jones, James M Neuberger, Pier Maria Battezzati, Massimo Zuin, Simon Taylor-Robinson, Maria F Donato, John Kirby, Robert Mitchell-Thain, Annarosa Floreani, Fotios Sampaziotis, Luigi Muratori, Domenico Alvaro, Marco Marzioni, Luca Miele, Fabio Marra, Edoardo Giannini, Eugenio Gaudio, Vincenzo Ronca, Giulia Bonato, Laura Cristoferi, Federica Malinverno, Alessio Gerussi, Deborah D Stocken, Heather J Cordell, Gideon M Hirschfield, Graeme J Alexander, Richard N Sandford, David E Jones, Pietro Invernizzi†, George F Mells†, for the Italian PBC Study Group and the UK–PBC Consortium*

Lancet Gastroenterol Hepatol 2018; 3: 626–34

Published Online July 12, 2018 http://dx.doi.org/10.1016/S2468-1253(18)30163-8

See Comment page 591

*Joint first authors

†Joint last authors

Academic Department of Medical Genetics (M Carbone MD, S Flack, N Varvaropoulou, A Spicer, J Badrock BSc, R N Sandford MD, G F Mells MD), **Department of Surgery, Wellcome Trust–Medical Research Council Stem Cell Institute, Anne McLaren Laboratory** (F Sampaziotis MD), **Department of Medicine** (Prof G J Alexander MD), **University of Cambridge, Cambridge, UK; Division of Gastroenterology and Hepatology, Department of Medicine and Surgery, University of Milan Bicocca, Milan, Italy** (M Carbone, F Bernuzzi, G Bonato MD, L Cristoferi MD, F Malinverno MD, A Gerussi MD, Prof P Invernizzi MD); **Department of Mathematics, University of Rome Tor Vergata, Rome, Italy** (Prof A Nardi PhD); **Department of Movement, Human and Health Sciences, University of Rome "Foro Italico", Rome, Italy** (Prof G Carpino MD); **Department of Anatomy, Histology, Legal Medicine, and Orthopedics, Sapienza University of Rome, Rome, Italy** (Prof E Gaudio MD); **IC Antonio Rosmini, Rome, Italy** (C Gavrila PhD); **Department of Medico-Surgical Sciences and Biotechnologies, Polo Pontino, Sapienza University of Rome, Rome, Italy** (V Cardinale MD,

Summary

Background Treatment guidelines recommend a stepwise approach to primary biliary cholangitis: all patients begin treatment with ursodeoxycholic acid (UDCA) monotherapy and those with an inadequate biochemical response after 12 months are subsequently considered for second-line therapies. However, as a result, patients at the highest risk can wait the longest for effective treatment. We determined whether UDCA response can be accurately predicted using pretreatment clinical parameters.

Methods We did logistic regression analysis of pretreatment variables in a discovery cohort of patients in the UK with primary biliary cholangitis to derive the best-fitting model of UDCA response, defined as alkaline phosphatase less than 1·67 times the upper limit of normal (ULN), measured after 12 months of treatment with UDCA. We validated the model in an external cohort of patients with primary biliary cholangitis and treated with UDCA in Italy. Additionally, we assessed correlations between model predictions and key histological features, such as biliary injury and fibrosis, on liver biopsy samples.

Findings 2703 participants diagnosed with primary biliary cholangitis between Jan 1, 1998, and May 31, 2015, were included in the UK-PBC cohort for derivation of the model. The following pretreatment parameters were associated with lower probability of UDCA response: higher alkaline phosphatase concentration (p<0·0001), higher total bilirubin concentration (p=0·0003), lower aminotransferase concentration (p=0·0012), younger age (p<0·0001), longer interval from diagnosis to the start of UDCA treatment (treatment time lag, p<0·0001), and worsening of alkaline phosphatase concentration from diagnosis (p<0·0001). Based on these variables, we derived a predictive score of UDCA response. In the external validation cohort, 460 patients diagnosed with primary biliary cholangitis were treated with UDCA, with follow-up data until May 31, 2016. In this validation cohort, the area under the receiver operating characteristic curve for the score was 0·83 (95% CI 0·79–0·87). In 20 liver biopsy samples from patients with primary biliary cholangitis, the UDCA response score was associated with ductular reaction (r=–0·556, p=0·0130) and intermediate hepatocytes (probability of response was 0·90 if intermediate hepatocytes were absent *vs* 0·51 if present).

Interpretation We have derived and externally validated a model based on pretreatment variables that accurately predicts UDCA response. Association with histological features provides face validity. This model provides a basis to explore alternative approaches to treatment stratification in patients with primary biliary cholangitis.

Funding UK Medical Research Council and University of Milan-Bicocca.

Introduction

Primary biliary cholangitis is an autoimmune liver disease characterised by destructive cholangitis affecting the small intrahepatic bile ducts, leading to chronic cholestasis and progressive fibrosis.[1] Many patients eventually develop end-stage liver disease with attendant need for liver transplantation.[2] First-line treatment for primary biliary cholangitis is with ursodeoxycholic acid (UDCA), a hydrophilic bile acid that improves liver biochemistry, delays histological progression, and improves liver-transplantation-free survival.[3–6] The biochemical response to treatment with UDCA (the UDCA response) strongly predicts long-term outcome. Thus, liver-transplantation-free survival is similar between patients with normal or near-normal liver biochemistry on UDCA and the general population, but is significantly reduced in those with abnormal liver biochemistry on treatment.[7]

Research in context

Evidence before this study

We searched PubMed for the term primary biliary cholangitis from 1998 to 2017. Treatment guidelines recommend that patients with primary biliary cholangitis who have an inadequate response to ursodeoxycholic acid (UDCA) after 12 months consider second-line therapy. However, with this approach many patients at highest risk of disease progression (ie, those with active disease that does not respond to UDCA) wait the longest for effective treatment. At present, there are no reliable means to identify patients before treatment who are unlikely to respond to UDCA, and so might benefit from early introduction of second-line therapy.

Added value of this study

We have derived a model based on pretreatment clinical variables that accurately predicts response to UDCA, with an area under the receiver operating characteristic curve of 0·83 (0·79–0·87) in external validation. We observed correlation between model predictions and key pathological features, such as the extent of fibrosis, ductular reaction, and cytokeratin 7 intermediate hepatocytes, providing face validity. The model uses readily available parameters, such as alkaline phosphatase, bilirubin, aminotransferases, patient's age at diagnosis, and the interval from diagnosis to the start of treatment. That delayed initiation of UDCA reduced the probability of response shows the importance of early, effective therapy.

Implications of all the available evidence

Future response to UDCA treatment can be predicted in patients with primary biliary cholangitis. This provides a basis to explore alternative approaches to treatment stratification, such as earlier introduction of second-line therapy. The model might even be useful in precision medicine initiatives to identify predictive biomarkers for treatment or risk stratification in primary biliary cholangitis.

The increased risk of progressive liver disease in patients with inadequate UDCA response has prompted the development of second-line therapies. Obeticholic acid is already used in clinical practice, and others will follow. The most recent guidelines therefore recommend that patients with inadequate UDCA response be considered for second-line therapies; the conventional period to demonstrate inadequate UDCA response is 12 months.[8] However, with this approach, patients at the highest risk of disease progression (ie, those with active disease that does not respond to UDCA) wait the longest for effective treatment.

At present, there are no clinical means to identify, before treatment, patients who are unlikely to respond to UDCA and who therefore might benefit from early introduction of second-line therapy. We aimed to determine whether inadequate UDCA response can be predicted using pretreatment clinical parameters; to understand the nature of those parameters; and to develop a predictive model that would enable accurate identification of patients unlikely to respond to UDCA, in whom alternative approaches to treatment stratification might be explored. Finally, we sought to test the biological plausibility of the model by looking for correlations between model predictions and key histological features, such as biliary injury and fibrosis, on liver biopsy samples from patients with primary biliary cholangitis.

Methods

Study design and participants

For derivation, we used data from the UK-PBC Research Cohort, part of the UK-PBC project.[9] In the discovery cohort, we included only those participants who were diagnosed with primary biliary cholangitis between Jan 1, 1998, and May 31, 2015, with follow-up data until May 31, 2016. We restricted the analysis to this period to ensure that all patients in the derivation cohort had equal access to UDCA following diagnosis, as UDCA was been registered in 1997.[10] For external validation, we used data from a well characterised cohort of patients recruited by the Italian PBC Study Group (appendix p 6).[11] In the validation cohort, to replicate real-world conditions, we included patients treated with UDCA diagnosed before or after 1998, with follow-up data until May 31, 2016.

We defined primary biliary cholangitis and definite primary biliary cholangitis–autoimmune hepatitis overlap syndrome according to EASL guidelines.[8] We defined probable primary biliary cholangitis–autoimmune hepatitis overlap syndrome as the combination of pretreatment immunoglobulin G more than twice the upper limit of normal (ULN) and aminotransferases more than five time the ULN. The date of diagnosis of primary biliary cholangitis was the date of detection of antimitochondrial antibodies or the date of the diagnostic liver biopsy, whichever occurred first.

The UK-PBC project and Italian PBC Study were done in accordance with the Declaration of Helsinki and the principles of good clinical practice. In both studies, all participants provided written informed consent. The UK-PBC project was approved by the Oxford C research ethics committee (07/H0606/96) and by each collaborating hospital. The Italian PBC Study was approved by the research ethics committee of the University of Milan Bicocca (ICH/232/11) and by each collaborating hospital.

Procedures

Baseline (T_0) data were those immediately before starting UDCA therapy. Clinical data and laboratory data were collected at diagnosis (alkaline phosphatase, aminotransferases, bilirubin, albumin, platelet count, creatinine,

Prof D Alvaro MD); **Institute of Health & Society** (H F Ainsworth PhD, Prof D D Stocken PhD), **Applied Immunobiology and Transplantation Research Group, Institute of Cellular Medicine, Faculty of Medical Sciences** (Prof J Kirby PhD), **Institute of Genetic Medicine** (Prof H J Cordell PhD), **Institute of Cellular Medicine** (Prof D E Jones MD), **Newcastle University, Newcastle-upon-Tyne, UK; Institute of Liver Studies, King's College Hospital NHS Foundation Trust, London, UK** (Prof M A Heneghan MD); **Sheila Sherlock Liver Centre, The Royal Free London NHS Foundation Trust, London, UK** (D Thorburn MD); **Scottish Liver Transplant Unit, Royal Infirmary of Edinburgh, Edinburgh, UK** (A Bathgate MD); **Liver Unit, St James's University Hospital, Leeds, UK** (R Jones MD); **Liver Unit, Queen Elizabeth Hospital, Birmingham, UK** (Prof J M Neuberger MD); **Division of Internal Medicine and Liver Unit, Ospedale San Paolo, Milan, Italy** (Prof P M Battezzati MD, Prof M Zuin MD, V Ronca MD); **Liver Unit, Division of Diabetes, Endocrinology and Metabolism, Department of Medicine, Imperial College London, London, UK** (Prof S Taylor-Robinson MD); **CRC "AM e A Migliavacca" Center for the Study of Liver Disease, Division of Gastroenterology and Hepatology, Fondazione IRCCS Cà Granda Ospedale Maggiore Policlinico, Università degli Studi di Milano, Milan, Italy** (M F Donato MD); **PBC Foundation, Edinburgh, UK** (R Mitchell-Thain); **Department of Surgery, Oncology and Gastroenterology, University of Padova, Padua, Italy** (Prof A Floreani MD); **Liver Unit, Policlinico di Sant'Orsola-Malpighi, Bologna, Italy** (Prof L Muratori MD); **Division of Gastroenterology and Hepatology, Ospedali Riuniti University Hospital, Ancona, Italy** (Prof M Marzioni MD); **Department of Internal Medicine and Gastroenterology, Gemelli University Hospital, Rome, Italy** (L Miele MD); **Department of Clinical and Experimental**

Medicine, University of Florence, Florence, Italy (Prof F Marra MD); **Division of Gastroenterology, Department of Internal Medicine, IRCCS-Azienda Ospedaliera Universitaria San Martino-IST, Genoa, Italy** (Prof E Giannini MD); **NIHR Birmingham Biomedical Research Centre, University of Birmingham, Birmingham, UK** (Prof G M Hirschfield MD); **University Hospitals Birmingham NHS Foundation Trust, Birmingham, UK** (Prof G M Hirschfield); **and Institute of Immunology and Immunotherapy, University of Birmingham, Birmingham, UK** (Prof G M Hirschfield)

Correspondence to:
Dr Marco Carbone, Division of Gastroenterology and Hepatology, Department of Medicine and Surgery, University of Milan Bicocca, 20126, Milan, Italy
marco.carbone@unimib.it

See Online for appendix

sodium, splenomegaly, and ascites), immediately before starting UDCA therapy (alkaline phosphatase, aminotransferases, bilirubin), and after 12 months of treatment with UDCA (alkaline phosphatase, aminotransferases, bilirubin). The endpoint was UDCA response, defined as alkaline phosphatase less than 1·67 times the ULN,[12] measured after 12 months of treatment with UDCA (ALP_{T12}). Because of controversy about the best ALP_{T12} cutoff to define UDCA response, we modelled three other cutoffs (ALP_{T12} ≤1×ULN; ALP_{T12} <1·5×ULN; and ALP_{T12} <2×ULN; appendix pp 13–15).

To account for interlaboratory variability, the alkaline phosphatase, alanine aminotransferase, aspartate aminotransferase, and total bilirubin were expressed as a multiple of their respective ULNs. We used a composite variable, aminotransferases, which was the alanine aminotransferase concentration when available; otherwise, the aspartate aminotransferase concentration was used.

To assess correlations with histological features, we evaluated formalin-fixed and paraffin-embedded liver biopsy samples from patients with primary biliary cholangitis at the Department of Clinical Medicine, Sapienza University of Rome (Rome, Italy). Biopsies were done at the time of diagnosis in treatment-naive patients with serological or biochemical suspicion of primary biliary cholangitis. Biopsy samples were collected consecutively from Jan 1, 1996, until Dec 31, 2006, when the unit policy changed and liver biopsies were no longer routinely done for patients with primary biliary cholangitis. Biopsies with fewer than nine complete portal tracts were excluded. Automated quantitative assessment of fibrosis was done in Sirius Red stained sections with an image analysis algorithm.[13] We evaluated ductular reaction and intermediate hepatocytes (previously known as biliary metaplasia) by cytokeratin 7 immunoreactivity (appendix p 9).[14–16] Histological evaluation was done independently by two authors blinded to the clinical data. In the case of disagreement, consensus was obtained by joint review of sections. Consensus on the grade and stage of the biopsy sample was reached in all cases.

	Derivation cohort (n=2703)	Validation cohort (n=460)
Age (years)	56·80 (49·52–64·16)	52·0 years (44·00–60·00)
Female	2409/2703 (89·1%)	423/460 (92·0%)
Treatment time lag* (days)	75 (0–258)	..
ALP_{diag} (×ULN)	1·85 (1·21–3·25)	1·78 (1·18–3·06)
AT_{diag} (×ULN)	1·40 (0·90–2·25)	1·38 (0·85–2·27)
TB_{diag} (×ULN)	0·53 (0·37–0·76)	0·60 (0·44–0·82)
PLT_{diag} (×10^3 per μL)	272 (225–324)	237 (199–286)
ALB_{diag} (g/L)	41 (38–44)	41 (38–43)
Creatinine (μmol/L)	76 (67–86)	..
Sodium (mEq/L)	139 (138–141)	..
Splenomegaly	263/2287 (11·5%)	..
Ascites	48/2285 (2·1%)	..
ALP_{T0} (×ULN)	1·91 (1·25–3·32)	..
AT_{T0} (×ULN)	1·42 (0·92–2·25)	..
TB_{T0} (×ULN)	0·53 (0·37–0·76)	..
ALP_{T12} (×ULN)	1·22 (0·88–1·88)	1·12 (0·77–1·76)
AT_{T12} (×ULN)	0·78 (0·54–1·23)	0·76 (0·52–1·08)
TB_{T12} (×ULN)	0·48 (0·35–0·65)	0·57 (0·41–0·73)

Data are median (IQR) or n (%). ALB_{diag}=albumin at diagnosis. ALP_{diag}=alkaline phosphatase at diagnosis. ALP_{T12}=alkaline phosphatase after 12 months of treatment with ursodeoxycholic acid (UDCA). PLT_{diag}=platelet count at diagnosis. AT_{diag}=aminotransferases at diagnosis. AT_{T12}=aminotransferases after 12 months of treatment with UDCA. TB_{diag}=total bilirubin at diagnosis. TB_{T12}=total bilirubin after 12 months of treatment with UDCA. ULN=upper limit of normal.

Table 1: **Characteristics of the derivation and validation cohorts**

	Parameter estimate	SE	Wald statistic	p value
Intercept	0·774	0·425	..	..
$\ln(ALP_{diag}$ [×ULN])	−2·730	0·138	−19·765	<0·0001
$1/\sqrt{TB_{diag}}$ [×ULN])	0·600	0·165	3·637	0·0003
$\ln(AT_{diag}$ [×ULN])	0·350	0·108	3·236	0·0012
Age (years)	0·028	0·006	5·074	<0·0001
Treatment time lag (years)	−0·154	0·035	−4·362	<0·0001
ΔALP (×ULN)	−0·557	0·073	−7·588	<0·0001

The results are from the logistic model based on baseline characteristics. Derivation cohort n=2703, used observations n=2640, missing 2·3%. ALP_{diag}=alkaline phosphatase concentration at diagnosis. TB_{diag}=total bilirubin concentration at diagnosis. AT_{diag}=aminotransferase concentration at diagnosis. ΔALP=change in alkaline phosphatase concentration from diagnosis to start of treatment.

Table 2: **Estimated parameters for UDCA response in the model derivation cohort**

Statistical analysis

We described continuous variables by median and IQR because most were not normally distributed. Categorical variables were described by absolute numbers and percentages. To compare groups, we used the χ^2 test for categorical variables (or Fisher's exact test in the case of sparse tables) and Student's *t* test for continuous variables (or Wilcoxon rank-sum test when a significant departure from normality was detected). We used Spearman's correlation to measure the strength and direction of monotonic association between two ranked variables.

We did multivariate analysis using logistic regression. We selected variables on the basis of non-automated backward selection, taking correlation structure among covariates and clinical interpretation of their effects into account. We explored possible interactions following a clinically driven approach. Different parametric transformations were considered to model the effect of continuous covariates, including first-degree and second-degree fractional polynomials. We did influence analysis, and underweighted overly influential observations according to Huber weights to limit the risk of local overfitting. We identified poorly predicted observations by the standardised deviance

residuals. In both internal and external validation, we evaluated model calibration and predictive ability using calibration belts[17] and area under the receiver operating characteristic curve (AUROC). We used non-parametric stratified bootstrapping to compute confidence bands for AUROC. The appendix (pp 7–8) contains further details of the statistical analyses.

We did all analyses using SAS (version 9.4) and R (version 3.4).

Role of the funding source

The funders had no role in study design, data collection, analysis, or interpretation, preparation of the report, or the decision to publish. The corresponding author had full access to the raw data and had final responsibility for the decision to submit for publication.

Results

For the UK-PBC derivation cohort, we identified 3073 UDCA-treated participants diagnosed with primary biliary cholangitis between Jan 1, 1998, and 31 May, 2015. We excluded 330 participants because ALP_{T12} was not available, 25 participants because treatment with UDCA lasted less than 9 months, and 15 participants because they started UDCA after liver transplantation. No participants had definite or probable primary biliary cholangitis–autoimmune hepatitis overlap syndrome. The derivation cohort therefore consisted of 2703 participants.

Median age at diagnosis was 56·80 years (IQR 49·52–64·16) and 2409 (89·7%) of participants were female (table 1). The median time from diagnosis to the start of treatment (the treatment time lag) was 75 days (IQR 0–258). As expected, the treatment time lag was longer in participants with primary biliary cholangitis diagnosed at the start of the study period than later (appendix p 10). The proportion of patients whose alkaline phosphatase increased between diagnosis and the start of treatment—and the size of this change—was greater in those with a longer treatment time lag (appendix p 10). Overall, 1902 (70·4%) of 2703 participants achieved the endpoint, ALP_{T12} less than 1·67×ULN, measured at a median of 13·4 months (IQR 11·8–16·9) after the start of treatment.

We did logistic regression analysis of explanatory variables to derive the best-fitting model to predict UDCA response. The following variables were excluded because more than 5% of data were missing: splenomegaly (429 [16%] of 2703 had missing data), ascites (422 [16%]), immunoglobulins (791 [29%]), and international normalised ratio (635 [23%]). The remaining variables were used in multivariable analysis. Of these, platelet count at diagnosis had the most missing data (130 [5%]).

The best-fitting logistic regression model included five variables: alkaline phosphatase at diagnosis (ALP_{diag}; p<0·0001), total bilirubin at diagnosis (TB_{diag}; p=0·0003), aminotransferases at diagnosis (AT_{diag}; p=0·0012), age at diagnosis (p<0·0001), treatment time lag (p<0·0001), and change in alkaline phosphatase concentration from the time of diagnosis to the start of treatment (ΔALP, p<0·0001; table 2). Log transformation was preferred for ALP_{diag} and AT_{diag}; the inverse of the squared root for TB_{diag}. A linear effect was confirmed for the treatment time lag and ΔALP. Overall, 63 observations were excluded from final fitting because of incomplete data in one or more of the selected variables. Influence analysis identified 29 observations as highly influential in the parameter estimates of the final model. Thus, to avoid potential model instability owing to local overfitting, observations were weighted according to Huber weights: in 2349 of 2641 of observations, Huber weights were equal to 1; in the remaining 292, weights had a median value of 0·71 (IQR 0·54–0·85)

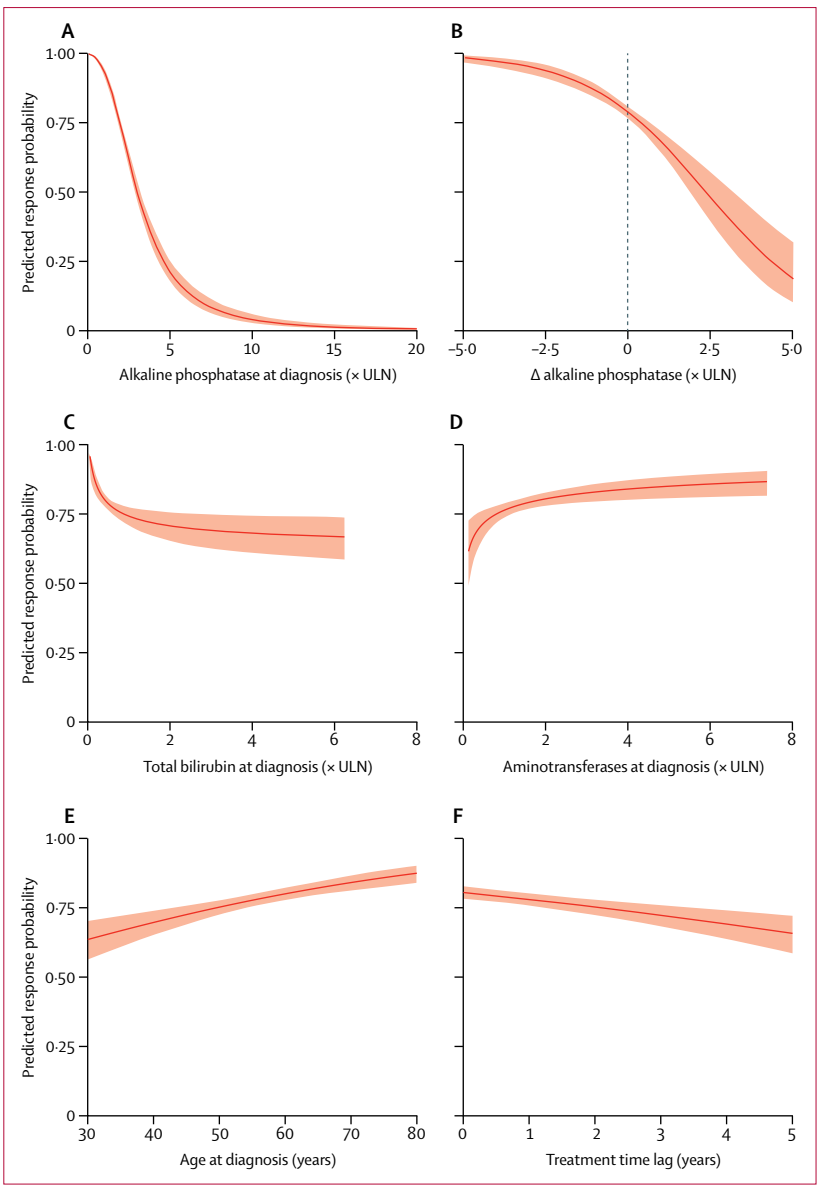

Figure 1: **Relationship between the variables selected in the best-fitting logistic regression model and the probability of UDCA response**
In each plot the remaining variables were set to their mean values.

	AT_{diag}	TB_{diag}	Recipient age (years)	ALP_{diag}	ΔALP	Treatment time lag	Estimated probability of UDCA response (95% CI)	Relative probability change
Varying aminotransferases								
High probability of response								
Low aminotransferases	0·5	0·5	60	2	0	0	0·76 (0·71–0·81)	ref
High aminotransferases	3	0·5	60	2	0	0	0·86 (0·83–0·89)	13%
Low probability of response								
Low aminotransferases	0·5	2	40	3	1	1	0·16 (0·09–0·23)	ref
High aminotransferases	3	2	40	3	1	1	0·27 (0·20–0·33)	40%
Varying time lag								
High probability of response								
Short time lag and low ΔALP	2	0·5	60	2	0	0·1	0·84 (0·81–0·86)	ref
Long time lag and high ΔALP	2	0·5	60	2	1	2	0·68 (0·64–0·73)	–19%
Low probability of response								
Short time lag and low ΔALP	1	2	40	3	0	0·1	0·33 (0·25–0·41)	ref
Long time lag and high ΔALP	1	2	40	3	1	2	0·17 (0·12–0·23)	–48%

The main effect of each variable in the model is less pronounced when the estimated probability of response is high. For example, if the AT_{diag} is increased from 0·5 × ULN to 3 × ULN, the relative change in the probability of response is 13% when the estimated probability of response is high (76% and 86%, respectively), whereas the relative change in the probability of response is 40% when the estimated probability of response is low (16% and 27%, respectively). Similarly, if the treatment time lag is increased from 0·1 (1 month) to 2 years, the relative change in the probability of response is –19% when the estimated probability of response is high (84% and 68%, respectively), whereas the relative change in the probability of response is –48% when the estimated probability of response is low (33% and 17%, respectively). ALP_{diag}=alkaline phosphatase concentration at diagnosis. AT_{diag}=aminotransferase concentration at diagnosis. TB_{diag}=total bilirubin concentration at diagnosis. ULN=upper limit of normal. ΔALP=change in ALP concentration from diagnosis to the start of treatment. Treatment time lag=time from diagnosis to the start of treatment.

***Table 3*: Clinical scenarios showing the effect of aminotransferase concentration and time lag on the estimated probability of UDCA response**

Higher ALP_{diag}, ΔALP, and TB_{diag} were associated with lower likelihood of UDCA response (figure 1A–C). Unexpectedly, higher AT_{diag} was associated with higher likelihood of UDCA response (figure 1D). Older age at diagnosis was associated with UDCA response, as was a shorter treatment time lag (figure 1E, 1F). The Hosmer-Lemeshow test showed no evidence of lack of fit to the data (p=0·4967). Inspection of residuals identified 138 (5·2%) of 2703 poorly predicted observations, consistent with the expected percentage of 5%. For 36 of these outliers, participants responded to UDCA despite predicted probability of less than 0·21, whereas 102 did not respond to UDCA despite a predicted probability of more than 0·80. None of the study characteristics distinguished these outliers from the remainder of the study cohort. We did not identify statistically significant interactions in the final model. However, we found that the main effect of the selected variables was less pronounced when the estimated probability of response was high (table 3, appendix p 12); this effect depends on the features of the logistic model.

We used the regression coefficients of the selected variables (table 2) to develop a predictive score of UDCA response for each patient according to the following formula:

$$\text{UDCA response score} = 0\cdot77 + 0\cdot60 \times (\sqrt{}TB_{diag})^{-1} - 2\cdot73 \times \ln(ALP_{diag}) + 0\cdot35 \times \ln(AT_{diag}) + 0\cdot03 \times \text{age} - 0\cdot15 \times (\text{treatment time lag}) - 0\cdot56 \times \Delta ALP$$

Based on the UDCA Response Score (URS), the predicted probability of response can be estimated as:

$$\text{Probability (response)} = \frac{\text{Exp(URS)}}{1 + \text{Exp(URS)}}$$

Internal validation of the URS demonstrated high discrimination ability with an AUROC of 0·87 (95% CI 0·86–0·89; figure 2A). Calibration of the model in the derivation cohort showed that event rates were correctly estimated by the predicted probabilities except at very extreme values, for which there was a slight tendency to underestimate the proportion (figure 2B). Table 3 shows how each variable affects the probability of UDCA response in different clinical scenarios. A calculator based on the URS is available online.

In clinical practice, information at the time of diagnosis might be missing. We therefore tested the URS by substituting measurements from the treatment start date (T_0) for those from the date of diagnosis (eg, TB_{T0} in place of TB_{diag}), fixing the treatment time lag to 0. Using T_0 measurements, the URS still had high discrimination ability, with an AUROC of 0·87 (95% CI 0·86–0·89). Additionally, we developed an alternative model, URS_{T0}, fitted in the derivation cohort using only data from the treatment start date (appendix p 17). The URS_{T0} contained the same variables as the URS, with similar parameter estimates. The URS may therefore be used with measurements from the treatment start date, setting and the treatment time lag to 0.

For the **UDCA Response Score calculator** see http://www.mat.uniroma2.it/~alenardi/URS.html

We validated the URS in an external population of 984 patients treated with UDCA from the Italian PBC Study Group, diagnosed before or after 1998, with follow-up data until May 31, 2016. Variables available for these patients included demographic characteristics, liver biochemistry at diagnosis, and the liver biochemistry during treatment. Data on the dose of UDCA and liver biochemistry at the start of treatment were not available. Application of the proposed score requires the ALP_{T0} to calculate the change; therefore, we included in the validation cohort only those patients who had started UDCA within 1 year of diagnosis (n=460) and fixed the ΔALP to zero. No participants had definite or probable primary biliary cholangitis–autoimmune hepatitis overlap syndrome.

Median age at diagnosis in the validation cohort was 52·0 years (IQR 44·00–60·00), and 423 (92·0%) of 460 were female. 335 (73%) of 460 participants responded to treatment. The validation cohort was younger than the derivation cohort with slightly lower PLT_{diag}, ALP_{T12}, and AT_{T12}, and slightly higher TB_{diag} and TB_{T12} (table 1, appendix p 19). The AUROC for the URS in the Italian cohort was 0·83 (95% CI 0·79–0·87), confirming a high ability to discriminate (figure 2C). The calibration plot showed no significant departure between the observed response rate and the predicted probability of response, confirming that the URS is well calibrated (figure 2D).

To confirm that the Italian cohort was suitable for validation even without ΔALP measurements, we tested the URS in a subgroup of 615 patients from the validation cohort with treatment time lag greater than 0 but less than 10 years, still setting the ΔALP and the treatment time lag to zero (effectively excluding these two variables). The model performed reasonably well in this subgroup, with an AUROC of 0·81 (95% CI 0·77–0·84), although less well than the proposed model with all variables. This finding shows that the URS has high predictive performance, even without information on ΔALP and treatment time lag. ALP_{diag}, TB_{diag}, and AT_{diag} are the variables with strongest effects in the URS, whereas the effect of the treatment time lag is limited (figure 1F). Based on these observations, we believe that the external validation was done appropriately.

When we fitted models using three other cutoffs for ALP_{T12}, all models included the same variables, with the size and direction of effect of each variable similar across all models (appendix pp 13–15).

To evaluate potential bias resulting from use of the aspartate aminotransferase as a surrogate for the alanine aminotransferase in the composite aminotransferase variable, we refitted the model in a subgroup of 2319 participants from the derivation cohort for whom ALT_{diag} values were available. We found that parameter estimates in the refitted model were similar to those in the original model: the parameter estimate for $\ln(ALT_{diag})$ was 0·359 (SE 0·114, p=0·0017; appendix p 18), similar to the parameter estimate for $\ln(AT_{diag})$ in the whole derivation cohort, which was 0·350 (SE 0·108, p=0·0012; table 2).

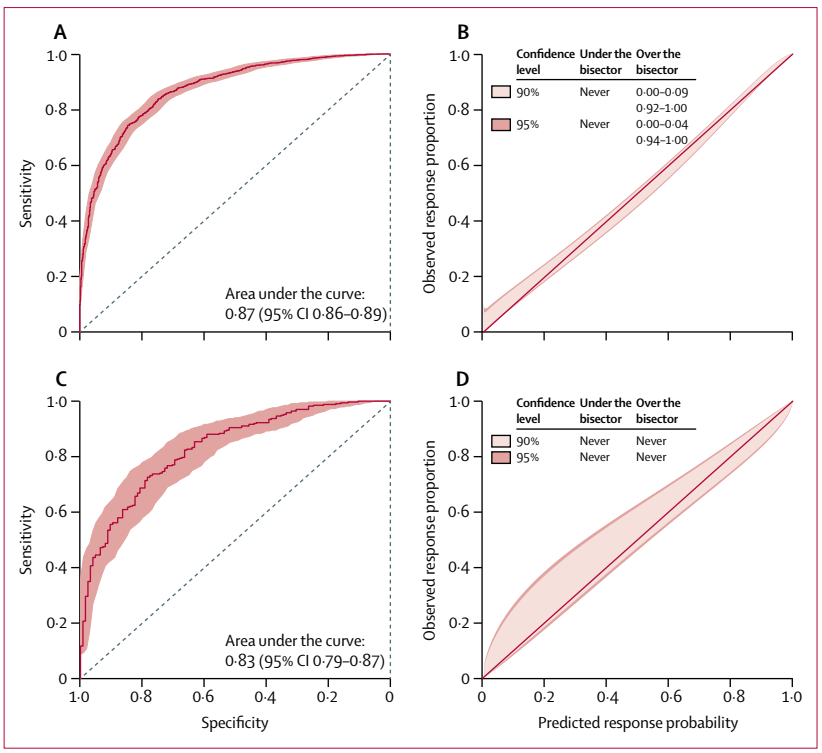

Figure 2: **Performance of the model**
(A) AUROC for the prediction of response to ursodeoxycholic acid (UDCA) calculated by UDCA Response Score in the derivation cohort, using stratified bootstrapping to estimate the confidence interval. (B) The predicted versus observed probability of response in the derivation cohort. "Never" means that the calibration belt never goes above or under the bisector; this suggests no evidence of miscalibration. (C) AUROC curve for the prediction of treatment response calculated by UDCA Response Score in the external validation cohort. (D) The predicted versus observed probability of response in the external validation cohort. AUROC=area under the receiver operating characteristic curve.

Using liver biopsy samples from 20 patients to assess the relationship with histological features, we found no correlation between the URS and the Ishak grade or Ludwig stage of disease (appendix p 20). There was, however, significant correlation of the URS with extent of ductular reaction (figure 3A), and extent of fibrosis (appendix p 20). The URS was also associated with the presence of intermediate hepatocytes, with median probability of response of 0·90 in patients with absent or minimal intermediate hepatocytes, compared with a median probability of response of 0·51 in those with clustered or diffuse intermediate hepatocytes (figure 3B). Moreover, there was correlation between the extent of ductular reaction and the ALP_{diag}, ALP_{T12}, Ludwig stage, interface hepatitis, portal inflammation, and the extent of fibrosis (appendix p 20).

Discussion

We have shown that, in patients with primary biliary cholangitis, the state of disease at baseline has a significant impact on the likelihood of response to UDCA, and that parameters associated with inadequate UDCA response

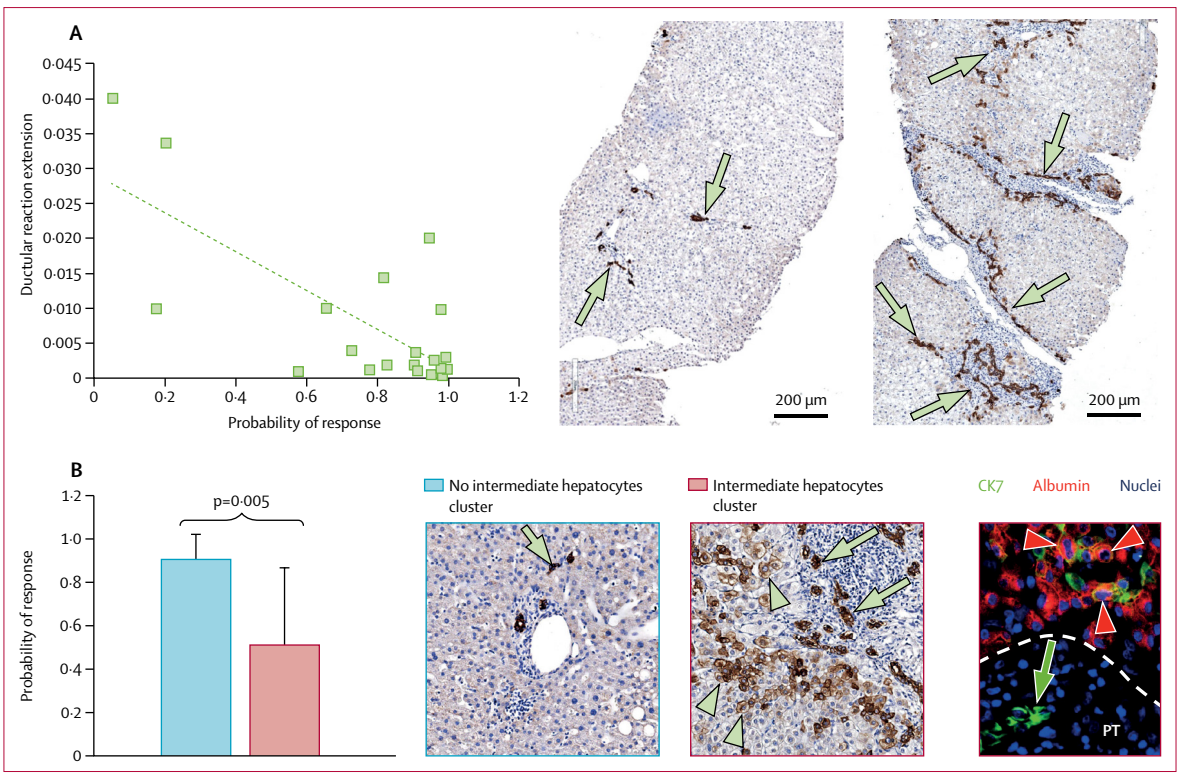

Figure 3: **Associations between UCDA Response Score and histological features of disease**
(A) Correlation of estimated probabilities of response to UDCA based on the UDCA Response Score with the extent of ductular reaction expressed as volume fraction of liver parenchyma. The panel on the right shows immunohistochemical staining for CK7 in liver biopsy samples. A different extension of CK7-positive ductular reaction is present (arrows). (B) Association of estimated probability of UDCA response with the presence of intermediate hepatocytes. The middle panels show representative images of immunohistochemical staining for CK7 in biopsy samples with or without intermediate hepatocytes (arrows). In the panel on the right, intermediate hepatocytes are confirmed in double immunofluorescence stains by the coexpression of albumin (arrowheads). Arrows indicate ductular reaction in the portal tract and at the interface with liver parenchyma. CK7=cytokeratin 7. DR=ductular reaction. IH=intermediate hepatocytes. UDCA=ursodeoxycholic acid. PT=portal tract.

can be integrated into an accurate predictive model, which we validated in an external cohort. Estimates from the model correlated with tissue-based markers of disease severity, providing face validity. Delay in starting UDCA therapy was associated with a higher risk of inadequate UDCA response, suggesting that delay to optimal treatment might reduce the likelihood of response.

The strongest predictor of UDCA response was the baseline alkaline phosphatase concentration; the probability of response declined sharply as the alkaline phosphatase increased. This strong inverse relationship suggests that—at least in patients with untreated primary biliary cholangitis—the alkaline phosphatase concentration accurately reflects the severity of biliary injury, apparently a key determinant of whether choleretic therapy will be effective. Consistent with this, UDCA response was less likely if the alkaline phosphatase concentration increased between diagnosis and the start of treatment (possibly reflecting progression of biliary injury) and if treatment was delayed (possibly because this allows the biliary injury to progress). The latter observation has implications for the timing of second-line therapy—ie, if a patient is unlikely to respond to UDCA they could be switched to second-line therapy sooner, with the hope of improving their prognosis.

Having excluded patients with definite or probable autoimmune hepatitis overlap from the analysis, the finding that higher concentrations of aminotransferases were associated with higher likelihood of UDCA response was unexpected. One possibility is that high aminotransferase concentrations are associated with a hepatitic phenotype of primary biliary cholangitis that is more responsive to choleretic treatment. Alternatively, aminotransferase concentrations might be high at an early, hepatitic stage of the disease process, when choleretic treatment might be more effective. Either way, the finding is important because it emphasises that, in treatment-naive patients with primary biliary cholangitis, high aminotransferase concentrations do not invariably signify autoimmune hepatitis overlap, and additional evidence is needed to justify immunosuppression. It is perhaps no surprise that high bilirubin was associated with low likelihood of UDCA response: in patients with primary biliary cholangitis, high bilirubin might reflect advanced ductopenia or end-stage liver disease, and it is plausible that choleresis is less effective in either setting.

We have previously shown that younger age at diagnosis predicts inadequate UDCA response.[9] We confirmed this finding in this study. The relationship between age at diagnosis and likelihood of UDCA response might be caused by the effect of hormones, such that high oestrogen levels increase resistance to treatment, an effect that is lost when patients present after menopause.[18] Immune senescence may also be important; T-cell exhaustion has a central role in determining outcome in autoimmune disease.[19]

We explored interactions that we considered biologically plausible but did not detect statistically significant effects. However, we showed that the effect of each variable in the final model depends on the estimated probability of response. For example, changes in AT_{diag} or the treatment time lag have relatively small effects when the estimated probability of UDCA response is high, but relatively large effects when the estimated probability of UDCA response is low. The same is true for the ALP_{diag}, TB_{diag}, ΔALP, and age of the patient (data not shown). This finding shows that, in patients already at risk, no variable can be taken for granted, particularly the treatment time lag, which is the only variable that the clinician can influence. These apparently differential effects are explained by the logistic link between the effect of covariates and the probability of response, and the different weights of the selected variables in the fitted model. They are biologically plausible; for example, in patients with high aminotransferases, jaundice may be attributable to hepatitic activity, which is amenable to treatment. Conversely, if the aminotransferases are not high, jaundice might indicate ductopenia or end-stage liver disease, which is less amenable to treatment. Taking the combination of different factors into account is what makes multivariable models so valuable for precision medicine.

We identified a correlation between the URS and the extent of ductular reaction. Ductular reaction represents a trans-amplifying population of cells, consisting of strings of cells with irregular lumens and a highly variable phenotypic profile.[20] The origin of ductular reaction is debated, but it is a hallmark of severe biliary injury. In our study, ductular reaction was also strongly correlated with ALP_{diag} and the observed as well as predicted UDCA treatment response. These observations emphasise the value of alkaline phosphatase concentration as a biomarker for biliary injury in patients with primary biliary cholangitis and suggest that the severity of biliary injury is a major determinant of responsiveness to choleretic treatment. However, only 20 biopsy samples were available for analysis, which is a limitation of the study. More pretreatment biopsy samples need to be assessed before conclusions can be drawn. Histological grading and staging in this study should ideally have been undertaken using the Nakanuma system,[21] which includes cholangitis and chronic cholestasis. However, this was not possible because the orcein stain was not routinely done in our pathology centre at the time the biopsy samples were collected and the remaining tissue was insufficient to obtain additional sections for research. Rather than using different grading and staging systems, we opted to use the well established Ishak system to grade the samples (evaluating interface hepatitis, focal necrosis, lobular inflammation, and portal inflammation) and the Ludwig system to stage them.[1] We recommend that future studies use systems that score necroinflammatory activity and chronic cholestasis.

In this work, we mainly used a cutoff of ALP_{T12} less than 1·67×ULN to define response, because this is how UDCA response has been defined in clinical trials of second-line agents and, as the recent industry standard, it will probably be used to decide which patients should receive second-line drugs. However, this cutoff is debated: Lammers and colleagues[12] showed that an ALP_{T12} less than 2·0×ULN best discriminates positive and negative outcomes in primary biliary cholangitis, whereas the European Association for the Study of the Liver suggests that an ALP_{T12} more than 1·5×ULN is the threshold at which long-term risk of death or liver transplantation becomes higher than that in a sex-matched and age-matched healthy population.[8] Given the strong correlation between alkaline phosphatase concentration and histological features of biliary injury, it might be argued that the threshold should be less than or equal to 1×ULN (ie, biochemical remission). When we repeated our analysis using these alternative cutoffs the respective models included the same variables and were similar in performance.

Since 2016, regulatory authorities have approved obeticholic acid for use in patients with primary biliary cholangitis with inadequate response to, or intolerance of, UDCA. More recently, Corpechot and colleagues[22] presented data from the BEZURSO trial, a phase 3 trial of bezafibrate versus placebo in combination with UDCA, in which normalisation of alkaline phophatase occurred in 67 (67%) of 100 patients on bezafibrate versus none on placebo. Several novel agents for primary biliary cholangitis are in phase 2 or 3 trials, such as Seladelpar (a PPARδ agonist),[23] Elafibranor (a PPARαδ agonist), and LJN452 (a non-bile acid FXR agonist). The current approach to management of primary biliary cholangitis is to start treatment with an optimal dose of UDCA in all patients, risk-stratify after 12 months of treatment using any of several binary or continuous scoring systems, then offer second-line therapy to high-risk patients (ie, those with abnormal liver biochemistry despite UDCA). Given the current and forthcoming availability of more efficacious disease-modifying treatments, now may be an appropriate time to review this approach. A predictive model enabling baseline identification of patients likely to need enhanced therapy could inform a novel treatment strategy (eg, early addition of second-line treatment). In this study, we present such a model. We recognise that ΔALP and treatment time lag would be redundant in clinical practice, but we retain them in the model to emphasise the importance of delaying effective treatment.

In conclusion, we have developed an accurate model to identify patients unlikely to respond to UDCA monotherapy at baseline. This model (or an iteration of it) could inform future treatment stratification in patients with primary biliary cholangitis.

Contributors

MC, AN, GMH, DEJ, RNS, and GFM designed the study. DEJ, RNS, PI, and GFM supervised the study. MC, SF, NV, AS, and JB collected data. AN, GC, and CG helped develop the risk score calculator. MC, AN, CG, DEJ, PI, and GFM analysed and interpreted data. AN did the statistical analysis. MC, AN, and GFM wrote the first draft of the report. All authors revised the report.

Declaration of interests

MC was a Sheila Sherlock Fellow of the European Association for the Study of the Liver and has received personal fees from Intercept Pharmaceuticals. JMN has received personal fees from Intercept Pharmaceuticals. MFD has received personal fees from Abbvie and Gilead, and travel grants from Kedrion. FM has received personal fees from Abbvie, Allergan, AstraZeneca, Bayer Gilead, Intercept, Menarini, and Novo Nordisk, and non-financial support from Alfasigma. HJC has received a Medical Research Council (MRC) Stratified Medicine award. GMH has received an MRC Stratified Medicine award; grants from the National Institute for Health Research; grants and personal fees from Intercept Pharmaceuticals; and personal fees from GSK, Cymabay, Novartis, and Falk Pharma. RNS has received salary support from the National Health Service (NHS) in the East of England through the Clinical Academic Reserve and from a MRC Stratified Medicine award. DEJ has received salary support from an MRC Stratified Medicine award; grants, personal fees, and other from Intercept Pharmaceuticals, GSK, Novartis, Cymabay, FFPharma, Shire, and grants and other from Pfizer. PI has received grant support and personal fees from Intercept Pharmaceuticals. GFM was a post-doctoral fellow of the NIHR Rare Diseases–Translational Research Collaboration; has received salary support from an MRC Stratified Medicine award; and has received personal fees from Intercept Pharmaceuticals. All other authors declare no competing interests.

Acknowledgments

This paper presents independent research supported by Medical Research Council (grant reference: MR/L001489/1), Isaac Newton Trust, Addenbrooke's Charitable Trust, PBC Foundation, National Institute for Health Research (NIHR) Birmingham Biomedical Research Centre (BRC), NIHR Cambridge BRC, NIHR Newcastle BRC, NIHR Nottingham BRC, NIHR Imperial BRC, NIHR University College London BRC, Wellcome Trust (grant reference: 085925), University of Milan-Bicocca, and Intercept Pharmaceuticals. The views expressed are those of the authors and not necessarily those of the NHS, NIHR or Department of Health in the UK, or the Sistema Sanitario Nazionale or Ministero dell'istruzione, dell'università e della ricerca in Italy.

References

1 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.

2 Pells G, Mells GF, Carbone M, et al. The impact of liver transplantation on the phenotype of primary biliary cirrhosis patients in the UK-PBC cohort. *J Hepatol* 2013; **59:** 67–73.

3 Poupon RE, Balkau B, Eschwège E, Poupon R. A multicenter, controlled trial of ursodiol for the treatment of primary biliary cirrhosis. *N Engl J Med* 1991; **324:** 1548–54.

4 Angulo P, Batts KP, Therneau TM, Jorgensen RA, Rolland Dickson E, Lindor KD. Long-term ursodeoxycholic acid delays histological progression in primary biliary cirrhosis. *Hepatology* 1999; **29:** 644–47.

5 Poupon RE, Poupon R, Balkau B. Ursodiol for the long-term treatment of primary biliary cirrhosis. *N Engl J Med* 1994; **330:** 1342–47.

6 Corpechot C, Carrat F, Bonnand AM, Poupon RE, Poupon R. The effect of ursodeoxycholic acid therapy on liver fibrosis progression in primary biliary cirrhosis. *Hepatology* 2000; **32:** 1196–99.

7 Carbone M, Sharp SJ, Flack S, et al. The UK-PBC risk scores: derivation and validation of a scoring system for long-term prediction of end-stage liver disease in primary biliary cholangitis. *Hepatology* 2016; **63:** 930–50.

8 EASL Clinical Practice Guidelines: The diagnosis and management of patients with primary biliary cholangitis. *J Hepatol*. 2017; **145:** 167–72.

9 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:** 560–69.

10 https://www.accessdata.fda.gov/drugsatfda_docs/label/2009/020675s017lbl.pdf (accessed May 30, 2018).

11 Liu X, Invernizzi P, Lu Y, et al. Genome-wide meta-analyses identify three loci associated with primary biliary cirrhosis. *Nat Genet* 2010; **42:** 658–60.

12 Lammers WJ, Van Buuren HR, Hirschfield GM, et al. Levels of alkaline phosphatase and bilirubin are surrogate end points of outcomes of patients with primary biliary cirrhosis: an international follow-up study. *Gastroenterology* 2014; **147:** 1338–1349.e5.

13 Huang Y, de Boer WB, Adams LA, et al. Image analysis of liver collagen using sirius red is more accurate and correlates better with serum fibrosis markers than trichrome. *Liver Int* 2013; **33:** 1249–56.

14 Carpino G, Cardinale V, Folseraas T, et al. Hepatic stem/progenitor cell activation differs between primary sclerosing and primary biliary cholangitis. *Am J Pathol* 2018; **188:** 627–39.

15 Semeraro R, Carpino G, Cardinale V, et al. Multipotent stem/progenitor cells in the human foetal biliary tree. *J Hepatol* 2012; **57:** 987–94.

16 Cardinale V, Wang Y, Carpino G, et al. The biliary tree—a reservoir of multipotent stem cells. *Nat Rev Gastroenterol Hepatol* 2012; **9:** 231–40.

17 Nattino G, Finazzi S, Bertolini G. A new calibration test and a reappraisal of the calibration belt for the assessment of prediction models based on dichotomous outcomes. *Stat Med* 2014; **33:** 2390–407.

18 Alvaro D, Invernizzi P, Onori P, et al. Estrogen receptors in cholangiocytes and the progression of primary biliary cirrhosis. *J Hepatol* 2004; **41:** 905–12.

19 McKinney EF, Lee JC, Jayne DRW, Lyons PA, Smith KGC. T-cell exhaustion, co-stimulation and clinical outcome in autoimmunity and infection. *Nature* 2015; **523:** 612–16.

20 Lanzoni G, Cardinale V, Carpino G. The hepatic, biliary, and pancreatic network of stem/progenitor cell niches in humans: a new reference frame for disease and regeneration. *Hepatology* 2016; **64:** 277–86.

21 Nakanuma Y, Zen Y, Harada K, et al. Application of a new histological staging and grading system for primary biliary cirrhosis to liver biopsy specimens: interobserver agreement. *Pathol Int* 2010; **60:** 167–74.

22 Corpechot C, Chazouillères O, Rousseau A, et al. A placebo-controlled trial of bezafibrate in primary biliary cholangitis. *NEJM* 2018; **378:** 2171–81.

23 Jones D, Boudes PF, Swain MG, et al. Seladelpar (MBX-8025), a selective PPAR-δ agonist, in patients with primary biliary cholangitis with an inadequate response to ursodeoxycholic acid: a double-blind, randomised, placebo-controlled, phase 2, proof-of-concept study. *Lancet Gastroenterol Hepatol* 2017; **2:** 716–26.