

Alma Mater Studiorum Università di Bologna
Archivio istituzionale della ricerca

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

This is the final peer-reviewed author's accepted manuscript (postprint) of the following publication:

Published Version:

Carbone, M., Nardi, A., Flack, S., Carpino, G., Varvaropoulou, N., Gavrilu, C., et al. (2018). Pretreatment prediction of response to ursodeoxycholic acid in primary biliary cholangitis: development and validation of the UDCA Response Score. THE LANCET. GASTROENTEROLOGY & HEPATOLOGY, 3(9), 626-634 [10.1016/S2468-1253(18)30163-8].

Availability:

This version is available at: <https://hdl.handle.net/11585/678099> since: 2019-02-28

Published:

DOI: [http://doi.org/10.1016/S2468-1253\(18\)30163-8](http://doi.org/10.1016/S2468-1253(18)30163-8)

Terms of use:

Some rights reserved. The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

This item was downloaded from IRIS Università di Bologna (<https://cris.unibo.it/>).
When citing, please refer to the published version.

(Article begins on next page)

Published in final edited form as:

Lancet Gastroenterol Hepatol. 2018 September 01; 3(9): 626–634. doi:10.1016/S2468-1253(18)30163-8.

Pre-treatment prediction of response to ursodeoxycholic acid (UDCA) in primary biliary cholangitis: the UDCA Response Score

M Carbone^{1,2,*}, A Nardi^{3,*}, S Flack¹, G Carpino⁴, N Varvaropoulou¹, C Gavrila³, A Spicer¹, J Badrock¹, F Bernuzzi², V Cardinale⁶, HF Ainsworth⁵, MA Heneghan⁷, D Thorburn⁸, A Bathgate⁹, R Jones¹⁰, JM Neuberger¹¹, PM Battezzati¹², M Zuin¹², S Taylor-Robinson¹³, M Donato¹⁴, J Kirby¹⁵, R Mitchell-Thain¹⁶, A Floreani¹⁷, F Sampaziotis¹⁸, L Muratori¹⁹, D Alvaro⁶, M Marzoni²⁰, L Miele²¹, F Marra²², E Giannini²³, E Gaudio⁴, V Ronca¹², G Bonato², L Cristofori², F Malinverno², A Gerussi², DD Stocken⁵, HJ Cordell²⁴, GM Hirschfield²⁵, GJ Alexander²⁶, RN Sandford¹, DE Jones²⁷, P Invernizzi^{2,†}, GF Mells^{1,†} for the Italian PBC Study Group & UK-PBC Consortium.

¹Academic Department of Medical Genetics, University of Cambridge, Cambridge, United Kingdom (UK)

²Division of Gastroenterology and Hepatology, Department of Medicine and Surgery, University of Milan Bicocca, Milan, Italy

³Department of Mathematics, Tor Vergata University of Rome, Rome, Italy

⁴Division of Health Sciences, Department of Movement, Human and Health Sciences, University of Rome "Foro Italico," Rome, Italy

⁵Institute of Health & Society, Newcastle University, Newcastle, UK

⁶Department of Medico-Surgical Sciences and Biotechnologies, Polo Pontino, Sapienza University of Rome, Rome, Italy

⁷Institute of Liver Studies, King's College Hospital NHS Foundation Trust, London, UK

⁸Sheila Sherlock Liver Center, The Royal Free London NHS Foundation Trust, London, UK

⁹Scottish Liver Transplant Unit, Royal Infirmary of Edinburgh, Edinburgh, UK

¹⁰Liver Unit, St James's University Hospital, Leeds, United Kingdom

¹¹Liver Unit, Queen Elizabeth Hospital, Birmingham, UK

Corresponding Author: Dr. Marco Carbone, University Hospital of Cambridge.

*Joint first authors

†Joint last authors

Author's contribution

Study concept and design: MC, AN, GFM

Acquisition of data: MC, SF, NV, AS, JB

Analysis and interpretation of data: MC, AN, DEJ, PI, GFM

Drafting of the manuscript: MC, AN, GFM

Critical revision of the manuscript for important intellectual content: all Authors

Statistical analysis: AN

Obtained funding: DEJ, PI, GFM

Technical, or material support: AN, GC, CG

Study supervision: DEJ, RS, PI, GFM

Conflict of interest statements: Nothing to declare.

- ¹²Division of Internal Medicine and Liver Unit, Ospedale San Paolo, Milan, Italy
- ¹³Liver Unit, Division of Diabetes, Endocrinology and Metabolism, Department of Medicine, Imperial College London, London, UK
- ¹⁴Gastroenterology and Hepatology Unit, Fondazione IRCCS Ca' Granda, Ospedale Maggiore Policlinico, Milan, Italy
- ¹⁵Applied Immunobiology and Transplantation Research Group, Institute of Cellular Medicine, Faculty of Medical Sciences, University of Newcastle, Newcastle upon Tyne, UK
- ¹⁶PBC Foundation, UK
- ¹⁷Department of Surgery, Oncology and Gastroenterology, University of Padova, Italy
- ¹⁸Department of Surgery, Wellcome Trust - Medical Research Council Stem Cell Institute, Anne McLaren Laboratory, University of Cambridge, Cambridge, UK
- ¹⁹Liver Unit, Policlinico di Sant'Orsola-Malpighi, Bologna, Italy
- ²⁰Clinic of Gastroenterology and Hepatology, Ospedali Riuniti University Hospital, Ancona, Italy
- ²¹Department of Internal Medicine and Gastroenterology, Gemelli University Hospital, Rome, Italy
- ²²Department of Clinical and experimental Medicine, University of Florence, Florence, Italy
- ²³Division of Gastroenterology, Department of Internal Medicine, IRCCS-Azienda Ospedaliera Universitaria San Martino-IST, Genova, Italy
- ²⁴Institute of Genetic Medicine, Newcastle University, Newcastle upon Tyne, UK
- ²⁵Center for Liver Research and NIHR Biomedical Research Unit, University of Birmingham, Birmingham, UK
- ²⁶Department of Medicine, University of Cambridge, Cambridge, UK
- ²⁶Institute of Cellular Medicine, Newcastle University, Newcastle upon Tyne, UK

Abstract

Background—Current guidelines advocate a step-wise approach to disease-modifying treatment of primary biliary cholangitis (PBC): all patients begin treatment with ursodeoxycholic acid (UDCA) monotherapy – and those with inadequate biochemical response to UDCA are subsequently considered for second-line therapies, the conventional period to demonstrate inadequate UDCA response being 12 months. A potential limitation with this approach, however, is that patients at highest risk end up waiting longest for effective treatment. In this study, we sought to determine whether UDCA response can be accurately predicted using pre-treatment clinical parameters, so that alternative approaches to treatment stratification might be explored.

Methods—We undertook logistic regression analysis of pre-treatment variables in 2,703 UDCA-treated patients to derive the best-fitting model of UDCA response, defined as $ALP_{T12} < 1.67 \times ULN$. We validated the model in an external PBC cohort from Italy ($n=460$). Finally, we explored the biological plausibility of the model by looking for correlation between model

predictions and key histological features on PBC liver biopsies ($n = 20$), such as biliary injury and fibrosis.

Findings—The following pre-treatment parameters were associated with lower probability of UDCA response: higher ALP ($p < 0.0001$), higher bilirubin ($p = 0.0003$), lower transaminases (TA, $p = 0.0012$), younger age ($p < 0.0001$), longer interval from diagnosis to the start of UDCA (treatment time lag, $p < 0.0001$), and worsening of ALP from diagnosis (ALP, $p < 0.0001$). Based on these variables, we derived a predictive score of UDCA response:

$$\begin{aligned} \text{UDCA response score (URS)} \\ = 0.77 + 0.60 \times \left(\sqrt{TB_{diag}} \right)^{-1} - 2.73 \times \ln(ALP_{diag}) + 0.35 \\ \times \ln(TA_{diag}) + 0.03 \times age - 0.15 \times (treatment\ time\ lag) - 0.56 \\ \times \Delta ALP \end{aligned}$$

In external validation, the AUROC was 0.83 (0.79-0.87). In PBC liver biopsies, the URS was associated with ductular reaction (DR) and intermediate hepatocytes (IH).

Interpretation—We have derived and externally validated a model based on pre-treatment variables that accurately predicts UDCA response. Association with DR and IH provides face validity. Thus, this model provides a basis to explore alternative approaches to treatment stratification in PBC.

Fundings—Medical Research Council (MR/L001489/1); University of Milan-Bicocca. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Keywords

risk stratification; UDCA response; predictive score; primary biliary cholangitis; patient selection

Background

Primary biliary cholangitis (PBC) is an autoimmune liver disease characterized by destructive cholangitis affecting the small intra-hepatic bile ducts, leading to chronic cholestasis and progressive fibrosis.¹ A substantial proportion of patients eventually develop end-stage liver disease with attendant need for liver transplantation (LT).² First-line treatment for PBC is with ursodeoxycholic acid (UDCA), a hydrophilic bile acid that has been shown to improve the liver biochemistry, delay histological progression and improve LT-free survival.^{3–5} It is well-established that the biochemical response to treatment with UDCA (so-called ‘UDCA response’) strongly predicts long-term outcome. Thus, patients with normal or near-normal liver biochemistry on UDCA have LT-free survival comparable to that of the general population, whereas LT-free survival is significantly reduced in those with abnormal liver biochemistry in spite of treatment.⁶

The increased risk of progressive liver disease in patients with inadequate UDCA response has prompted the development of novel, second-line therapies. The first of these, obeticholic acid, has already entered clinical practice. Others will follow. The most recent guidelines

therefore recommend that PBC patients with inadequate UDCA response be sought and considered for second-line therapies, the conventional period to demonstrate inadequate UDCA response being 12 months.⁷ A potential limitation with this step-up approach, however, is that patients at highest risk of disease progression (that is, those with active disease that does not subsequently respond to UDCA) end up waiting longest for effective treatment.

At present, there are no clinical tools enabling pre-treatment identification of patients who are unlikely to respond to UDCA, who might therefore benefit from early introduction of second-line therapy. Thus, in the current study, we set out to determine whether it is possible to predict inadequate UDCA response using pre-treatment 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 correlation between model predictions and key histological features on PBC liver biopsies such as biliary injury and fibrosis.

Materials and Methods

Participants

For derivation, we used data from the UK-PBC Research Cohort, part of the ongoing UK-PBC project (see Supplementary Methods).⁹ In the discovery cohort, we included only those participants who were diagnosed with PBC between 01.01.1998 and 31.05.2015, with follow-up data until 31.05.2016. We restricted the analysis to this time period to ensure that everyone in the derivation cohort had equal access to UDCA following diagnosis, the medication having been registered in 1997.¹⁰ For external validation, we used data from a well-characterized cohort of PBC patients recruited by the Italian PBC Study Group (see Supplementary Methods).¹¹ In the validation cohort, to replicate a real world setting, we included UDCA-treated patients diagnosed before or after 1998, with follow-up data until 31.05.2016.

Study definitions

In the current study, PBC and 'definite' PBC – AI overlap syndrome were defined according to EASL guidelines.⁷ 'Probable' PBC – AI overlap syndrome was defined as the combination of pre-treatment immunoglobulin G (IgG) > 2×ULN and transaminases (TA) > 5×ULN. The date of diagnosis of PBC was the date of detection of anti-mitochondrial antibodies (AMA) or the date of the diagnostic liver biopsy, whichever occurred first. Baseline (time T0) data were those immediately before starting UDCA therapy. The end-point was UDCA response, defined as ALP < 1.67×ULN measured after 12 months of treatment with UDCA (ALP_{T12}). Recognizing on-going debate about the optimal ALP_{T12} cut-off to define UDCA response, we modelled three other cut-offs (ALP_{T12} ≤ 1×ULN; ALP_{T12} < 1.5×ULN; and ALP_{T12} < 2×ULN) and present these data in the Supplementary Table S1.

Statistical analysis

Continuous variables were described by median, first and third quartiles because most showed a skewed distribution with significant departure from the normal density. To account for inter-laboratory variability, the ALP, alanine transaminase (ALT), aspartate transaminase (AST) and total bilirubin (TB) were expressed as a multiple of their respective ULNs. We used a composite variable, TA, that was the ALT when available; otherwise, the AST. Categorical variables were described by absolute frequencies and percentages. To compare groups, we used the χ^2 test for categorical variables (or Fisher exact test in the case of sparse tables) and student t-test for continuous variables (or Wilcoxon rank-sum test when a significant departure from normality was detected). The Spearman's correlation was used to measure the strength and direction of monotonic association between two ranked variables.

Multivariate analysis was undertaken using logistic regression. Variable selection was based on non-automated backward selection, taking correlation structure among covariates and clinical interpretation of their effects into account. Different parametric transformations were considered to model the effect of continuous covariates, including first and second degree fractional polynomials. Influence analysis was performed, and overly influential observations were underweighted according to Huber weights to limit the risk of local overfitting. Poorly predicted observations were identified by the standardized deviance residuals. In both internal and external validation, model calibration and predictive ability were evaluated using calibration belts¹² and receiver under operator characteristic (ROC) curves. Non-parametric stratified bootstrapping was used to compute confidence bands for ROC curves. For further details, please see Supplementary Methods.

All analyses were undertaken using SAS version 9.4 (SAS Institute, Cary NC) and R version 3.4 (R Core Team (2012)).

Correlation with histological features

We evaluated formalin fixed and paraffin embedded liver biopsies from PBC patients at the Department of Clinical Medicine, Sapienza University of Rome. Biopsies were performed at the time of diagnosis in treatment-naïve patients with serological or biochemical suspicion of PBC. They were collected consecutively from 01.01.1996 until 31.12.2006, when the unit policy changed, and liver biopsies were no longer routinely performed in PBC. Biopsies with less than nine complete portal tracts were excluded. Portal inflammation, interface hepatitis, focal necrosis, apoptosis and lobular inflammation were graded using the Ishak system.¹³ Specimens were staged according to the Ludwig criteria.¹ Automated, quantitative assessment of fibrosis was undertaken in Sirius Red stained sections using an image analysis algorithm.¹⁴ Ductular reaction (DR) and intermediate hepatocytes (IH, previously known as biliary metaplasia) were evaluated using cytokeratin 7 (CK7) immunoreactivity (Supplementary Figure S1).

Results

Characteristics of the UK-PBC derivation cohort

We identified 3,073 UDCA-treated participants diagnosed with PBC between 01.01.1998 and 31.05.2015. Of these, we excluded 330 participants because the ALP_{T12} was not available; 25 participants because treatment with UDCA lasted less than nine months, and 15 participants because they started UDCA after LT. No participants had definite or probable PBC-AI overlap syndrome, as defined above. The derivation cohort therefore consisted of 2,703 participants.

Characteristics of the derivation cohort are reported in Table 1. The median age at diagnosis was 56.8 years; 89.7% were female. The median liver biochemistries at diagnosis were TB_{diag} 0.53×ULN, TA_{diag} 1.40×ULN, ALP_{diag} 1.85×ULN and albumin (ALB_{diag}) 41 g/L. The median platelet count at diagnosis (PLT_{diag}) was 272 × 10³/μL. The median time from diagnosis to the start of treatment (treatment time lag) was 75 days (interquartile range [IQR] 0-258 days). As expected, the treatment time lag was longer in participants diagnosed with PBC in earlier compared to later eras (Supplementary Figure S2). We also observed that the proportion of patients with increase in the ALP from diagnosis to the start of treatment – and the magnitude of this change – was greater in those with longer treatment time lag (Supplementary Figure S3). Overall, 1,902 of 2,703 participants (70.4%) achieved the end-point, ALP_{T12} < 1.67×ULN, measured at median 13.4 months after the start of treatment (IQR 11.8-16.9 months). For further details, please see Supplementary Results.

Identification of variables that predict UDCA response

We undertook logistic regression analysis of diverse explanatory variables to derive the best-fitting model of UDCA response. The following variables were excluded owing to missingness >5%: splenomegaly (15.9% missing data), ascites (15.6%), immunoglobulins (29.3%) and INR (23.5%). The remaining variables (see Table 1) were taken forward for multivariable analysis. Of these, maximum missingness was 4.8% for the platelet count at diagnosis (PLT_{diag}).

The best-fitting logistic regression model included five variables: ALP_{diag} ($p < 0.0001$), TB_{diag} ($p = 0.0003$), TA_{diag} ($p = 0.0012$), age at diagnosis ($p < 0.0001$), treatment time lag ($p < 0.0001$) and change in the ALP from the time of diagnosis to the start of treatment (ALP, $p < 0.0001$). Log transformation was preferred for the ALP_{diag} and TA_{diag}; the inverse of the squared root for TB_{diag}. A linear effect was confirmed for the treatment time lag and ALP. 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 88.9% of observations, Huber weights were about 1; in the remaining cases, weights ranged from 0.20 to 0.99 with a median value of 0.71.

Parameter estimates are reported in Table 2. Figure 1 shows the effect of the selected variables on the probability of response. Higher ALP, ALP and TB were associated with lower likelihood of UDCA response (Figure 1, a – c). Unexpectedly, higher TA_{diag} was

Parameter estimates are reported in Table 2. Figure 1 shows the effect of the selected variables on the probability of response. Higher ALP, ALP and TB were associated with lower likelihood of UDCA response (Figure 1, a – c). Unexpectedly, higher TA_{diag} was

associated with higher likelihood of UDCA response (Figure 1d). Older age at diagnosis predicted UDCA response, as did a shorter treatment time lag (Figure 1, e & f). Using the Hosmer and Lemeshow test, there was no evidence of lack of fit to the observed data ($p=0.4967$). Inspection of residuals identified 138 (5.2%) poorly predicted observations, consistent with the expected percentage of 5%. Thirty-six of these outliers achieved UDCA response despite predicted probability <0.21 , while 102 failed to achieve UDCA response despite predicted probability >0.80 . None of the study characteristics distinguished these outliers from the remainder of the study cohort. We did not identify statistical interactions in the final model but, as expected from a multivariable model, we did observe that the effect of one variable on the probability of UDCA response is related to the levels of other variables in the model (Table 3 and Supplementary figure S4).

Development of a predictive model of UDCA-response

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

$$\begin{aligned} \text{UDCA response score(URS)} &= 0.77 + 0.60 \times \left(\sqrt{TB_{diag}} \right)^{-1} - 2.73 \times \ln(ALP_{diag}) + 0.35 \\ &\times \ln(TA_{diag}) + 0.03 \times \text{age} - 0.15 \times (\text{treatment time lag}) - 0.56 \\ &\times \Delta \text{ ALP} \end{aligned}$$

Based on URS values, the predicted probability of response can be estimated as:

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

Internal validation of the model demonstrated high discrimination ability with an AUROC of 0.87 (0.86 - 0.89) (Figure 2a). Calibration of the model in the derivation cohort showed that observed event rates were correctly estimated by the predicted probabilities except at very extreme values, where there was a slight tendency to underestimate the observed proportion (Figure 2b). Table 3 shows different clinical scenarios that highlight the impact of each variable on the probability of UDCA response. An online calculator based on the URS is available at <http://www.mat.uniroma2.it/~alenardi/URS10b.html>.

External validation of the URS

We identified 984 UDCA-treated PBC patients from the Italian PBC Study Group. Variables available for these patients included demographic characteristics; the liver biochemistry at diagnosis; and the liver biochemistry on treatment. Data on the dose of UDCA and the liver biochemistry at the start of treatment (T0) were not available. Application of the proposed score requires the ALP_{T0} therefore we included in the validation cohort only those patients who had started UDCA within one year of diagnosis ($n=460$) and fixed the ALP to zero. No participants had definite or probable PBC-AIH overlap, as defined above.

Characteristics of the study cohort were as follows: the median age at diagnosis was 52.0 years; 92.0% were female; the median ALP_{diag} was $1.78 \times ULN$, TB_{diag} $0.60 \times ULN$; TA_{diag} $1.38 \times ULN$; ALB_{diag} 41 g/L and PLT_{diag} $237 \times 10^3/\mu L$. The rate of response based on the end-point, $ALP_{T12} < 1.67 \times ULN$, was 72.8%. The validation cohort was younger than the model derivation cohort with slightly lower PLT_{diag} , ALP_{T12} and TA_{T12} , and slightly higher TB_{diag} and TB_{T12} (Supplementary table S3). The AUROC for the URS in the Italian cohort was 0.83 (0.79 - 0.87), confirming high discrimination ability (Figure 2c). The calibration plot in Figure 2d shows no significant departure between the observed response rate and the predicted probability of response, confirming that the URS is well-calibrated.

Recognizing on-going debate about the optimal ALP_{T12} cut-off to define treatment response, we fitted models using three other cut-offs ($ALP_{T12} = 1 \times ULN$, $ALP_{T12} < 1.5 \times ULN$ and $ALP_{T12} < 2 \times ULN$). All models included the same variables, with the size and direction of effect of each variable comparable across all models. For further details, please see Supplementary Tables S1, a - c.

The URS was derived using a composite variable, TA, that was the ALT where available, otherwise the AST. To evaluate potential bias resulting from use of the AST as a surrogate for the ALT, we re-fitted the model in a subgroup of 2,319 participants from the derivation cohort for whom ALT_{diag} values were available. We found that parameter estimates in the re-fitted model were similar to those in the original model. Notably, the parameter estimate for $\ln(ALT_{diag})$ was 0.359 (standard error [s.e.] = 0.114, $p = 0.0017$), comparable to the parameter estimate for $\ln(TA_{diag})$ in the whole derivation cohort, which was 0.350 (s.e. = 0.108, $p = 0.0012$). For further details, please see Supplementary Table S2.

Relationship with histological features

Liver biopsies from 20 PBC patients were suitable for analysis. There was no correlation between the URS and the Ishak grade or Ludwig stage of disease. There was, however, statistically significant correlation between the URS and extent of DR (Figure 3a) as well as the extent of fibrosis (Supplementary Table S4). The URS was also associated with the presence of IH, with median probability of response 0.90 in biopsies with absent or minimal IH, compared to median probability of response 0.51 in those with clustered or diffuse IH (Figure 3b). Moreover, there was correlation between the extent of DR and the ALP at diagnosis, ALP at 12 months after treatment with UDCA, Ludwig stage of disease, interface hepatitis, portal inflammation, and the extent of fibrosis (Supplementary Table S4).

Discussion

In the current study, we have shown that in PBC, 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 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. Notably, one of the parameters associated with higher risk of inadequate UDCA response was delay in starting UDCA therapy, suggesting that in PBC, delay to optimal treatment may reduce the likelihood of responding to it.

The strongest predictor of UDCA response was the baseline ALP, the probability of response declining sharply as the ALP increased. This strong inverse relationship suggests that – at least in the context of untreated PBC – the ALP accurately reflects the severity of the biliary injury, apparently a key determinant of whether choleretic therapy will be effective. Consistent with this, UDCA response was less likely if the ALP increased between diagnosis and the start of treatment (possibly reflecting progression of the biliary injury) and if treatment was delayed (possibly because this allows the biliary injury to progress). The latter observation especially has implications for the timing of second-line therapy.

Having excluded from analysis patients with definite or probable AIH overlap, finding that higher transaminases were associated with higher likelihood of UDCA response was unexpected. One possibility is that elevated transaminases identify a hepatic phenotype of PBC that is more responsive to choleretic treatment. Alternatively, elevated transaminases may identify an early, hepatic stage of the PBC disease process, when choleretic treatment is conceivably more effective. Either way, the observation is important because it emphasizes that in treatment-naïve PBC patients, elevated transaminases do not invariably signify AIH overlap – and additional evidence is essential to justify immunosuppression. It is perhaps no surprise that higher bilirubin was associated with lower likelihood of UDCA response: in PBC, elevated bilirubin may reflect advanced ductopenia or ESLD; it is plausible that cholestasis should be less effective in either setting.

We have previously shown that younger age at diagnosis predicts inadequate UDCA response.⁹ We confirmed this finding in the expanded UK-PBC Research Cohort that was the basis of the current study. The relationship between age at diagnosis and likelihood of UDCA response may be explicable by the effect of hormones, such that high estrogen levels increase resistance to treatment, which is lost when patients present after menopause¹⁵. Immune senescence, however, may also be important, T-cell exhaustion having been shown to play a central role in determining outcome in autoimmune disease.¹⁶

We did not identify statistical interactions in the final model – but we did observe that the effect of one variable on the probability of UDCA response is related to the levels of other variables in the model. For example, at lower ALP levels, when the estimated probability of UDCA response is above 0.9, the transaminases and bilirubin have minimal effect. At higher ALP levels, however, when the estimated probability decreases, the effect of the transaminases is pronounced. Furthermore, the effect of bilirubin on the probability of UDCA response is more pronounced at higher ALP levels and lower transaminases, while the effects of delayed treatment and worsening of ALP from diagnosis to the start of treatment are marked in younger, high risk patients but not in older, low risk patients. These seemingly 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 elevated transaminases, jaundice may be attributable to hepatic activity, amenable to treatment. Conversely, if the transaminases are not elevated, jaundice may reflect ductopenia or ESLD, less amenable to treatment. Taking the combination of different factors into account is what makes multivariable models so valuable for precision medicine.

Only 20 biopsies were available for analysis; we acknowledge that this is a limitation of the study. Nevertheless, we identified correlation between the URS and the extent of DR. Ductular reaction represents a trans-amplifying population consisting of strings of cells with irregular lumens and a highly variable phenotypic profile.¹⁷ The origin of DR is debated; it is nevertheless a hallmark of severe biliary injury.^{18–20} In the current study, DR was also strongly correlated with the ALP_{diag} and the observed as well as predicted UDCA treatment response. These observations emphasize the value of the ALP as a biomarker for biliary injury in PBC and suggest that the severity of biliary injury is a major determinant of responsiveness to choleretic treatment. Evaluation of additional pre-treatment biopsies is necessary, however, before conclusions may be drawn.

In this work, we report results for the cut-off, $ALP_{T12} < 1.67 \times ULN$, 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 agents. This cut-off is, however, debatable: Lammers *et al.*²¹ showed that $ALP_{T12} < 2.0 \times ULN$ best discriminates positive versus negative outcomes in PBC, while EASL suggests that $ALP_{T12} > 1.5 \times ULN$ is the threshold at which long-term risk becomes clinically meaningful.⁷ Given the strong correlation between ALP and histological features of biliary injury, it might be argued that the threshold should be $ALP_{T12} = 1 \times ULN$ (that is, biochemical remission). Recognizing this on-going debate, we provide results for all these cut-offs, and show that the respective models include the same variables and are comparable in performance.

Since 2016, regulatory authorities have approved OCA for use in PBC patients with inadequate response to, or intolerance of, UDCA. More recently, Corpechot *et al.*²² presented data from the BEZURSO trial, a phase III trial of bezafibrate or placebo in combination with UDCA, in which normalization of ALP occurred in 67% of patients on bezafibrate versus 0% on placebo. Several novel agents for PBC are currently in phase II or III evaluation, such as Seladelpar (a PPAR δ agonist),²³ Elafibranor (a PPAR α / δ agonist), and the Novartis molecule, LJN452 (a non-bile acid FXR agonist). The current approach to management of PBC is to initiate 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 (that is, 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 an evolved treatment strategy (for example, early addition of second-line treatment). In this study, we present such a model. We recognize that the variables, ALP and treatment time lag, would be redundant in clinical practice – but we retain them in the current model to emphasize the potential importance of delaying effective treatment.

In conclusion, we have developed an accurate model enabling patients unlikely to respond to UDCA monotherapy to be identified at baseline. We believe this model (or an iteration of it) could inform future treatment stratification in PBC.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Bibliography

1. Ludwig J, Dickson ER, McDonald GS. Staging of chronic nonsuppurative destructive cholangitis (syndrome of primary biliary cirrhosis). *Virchows Archiv A, Pathological anatomy and histology*. 1978; 379:103–12. [PubMed: 150690]
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. *Journal of Hepatology*. 2013; 59:67–73. [PubMed: 23466308]
3. Poupon RE, Balkau B, Eschwege E, et al. A Multicenter, Controlled Trial of Ursodiol for the Treatment of Primary Biliary Cirrhosis. *New England Journal of Medicine*. 1991; 324:1548–1554. [PubMed: 1674105]
4. Angulo P, Batts KP, Therneau TM, et al. Long-term ursodeoxycholic acid delays histological progression in primary biliary cirrhosis. *Hepatology (Baltimore, Md)*. 1999; 29:644–7.
5. Poupon RE, Poupon R, Balkau B. Ursodiol for the Long-Term Treatment of Primary Biliary Cirrhosis. *New England Journal of Medicine*. 1994; 330:1342–1347. [PubMed: 8152446]
6. 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 (Baltimore, Md)*. 2016; 63:930–50.
7. EASL Clinical Practice Guidelines. The diagnosis and management of patients with primary biliary cholangitis. *Journal of Hepatology*. 2017; 67:145–172. [PubMed: 28427765]
8. Corpechot C, Carrat F, Bonnand AM, et al. The effect of ursodeoxycholic acid therapy on liver fibrosis progression in primary biliary cirrhosis. *Hepatology (Baltimore, Md)*. 2000; 32:1196–9.
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; 144:560–569–4. [PubMed: 23246637]
10. fda. https://www.accessdata.fda.gov/drugsatfda_docs/label/2009/020675s0171bl.pdf.
11. Liu X, Invernizzi P, Lu Y, et al. Genome-wide meta-analyses identify three loci associated with primary biliary cirrhosis. *Nature genetics*. 2010; 42:658–60. [PubMed: 20639880]
12. 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 Jun 30; 33(14):2390–407. [PubMed: 24497413]
13. Ishak K, Baptista A, Bianchi L, et al. Histological grading and staging of chronic hepatitis. *Journal of hepatology*. 1995; 22:696–9. [PubMed: 7560864]
14. 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 international : official journal of the International Association for the Study of the Liver*. 2013; 33:1249–56. [PubMed: 23617278]
15. Alvaro D, Invernizzi P, Onori P, et al. Estrogen receptors in cholangiocytes and the progression of primary biliary cirrhosis. *J Hepatol*. 2004; 41:905–912. [PubMed: 15645536]
16. McKinney EF, Lee JC, Jayne DRW, et al. T-cell exhaustion, co-stimulation and clinical outcome in autoimmunity and infection. *Nature*. 2015; 523:612–6. [PubMed: 26123020]
17. 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; 64:277–286. [PubMed: 26524612]
18. Lu W-Y, Bird TG, Boulter L, et al. Hepatic progenitor cells of biliary origin with liver repopulation capacity. *Nature Cell Biology*. 2015; 17:971–983. [PubMed: 26192438]
19. Kopp JL, Grompe M, Sander M. Stem cells versus plasticity in liver and pancreas regeneration. *Nature cell biology*. 2016; 18:238–45. [PubMed: 26911907]

20. Williams MJ, Clouston AD, Forbes SJ. Links Between Hepatic Fibrosis, Ductular Reaction, and Progenitor Cell Expansion. *Gastroenterology*. 2014; 146:349–356. [PubMed: 24315991]
21. Lammers WJ, van Buuren HR, Hirschfield GM, et al. Levels of alkaline phosphatase and bilirubin are surrogate end points of outcomes of patients with primary biliary cirrhosis: an international follow-up study. *Gastroenterology*. 2014; 147:1338–49.e5. [PubMed: 25160979]
22. Corpechot C, Chazouillères O, Rousseau A, et al. A 2-year multicenter, double-blind, randomized, placebo-controlled study of bezafibrate for the treatment of primary biliary cholangitis in patients with inadequate biochemical response to ursodeoxycholic acid therapy (Bezurso) [abstract]. *Journal of Hepatology*. 2017; 66(Supplement):S89.
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. *The Lancet Gastroenterology & Hepatology*. 2017; 2:716–726. [PubMed: 28818518]



Research in context

Evidence before this study

Novel second-line therapies are now available for patients with primary biliary cholangitis (PBC) and inadequate response to ursodeoxycholic acid (UDCA). Current guidelines therefore recommend that PBC patients with inadequate UDCA response be sought and considered for second-line therapy. The conventional period to demonstrate inadequate UDCA response is 12 months. A potential limitation with this step-up approach, however, is that patients at highest risk of disease progression (that is those with active disease that does not subsequently respond to UDCA) end up waiting longest for effective treatment. At present, there are no clinical tools enabling pre-treatment identification of patients who are unlikely to respond to UDCA, who might benefit from early introduction of second-line therapy.

Added value of this study

In this study, we have derived a model based on pre-treatment clinical variables that accurately predicts future UDCA response, with an AUROC 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 CK7+ intermediate hepatocytes, providing face validity. The model consists of readily available parameters, such as the alkaline phosphatase, bilirubin and transaminases; the patient's age at the time of diagnosis; and the interval from diagnosis to the start of treatment. Finding that delayed initiation of UDCA reduced the probability of response highlights the importance of early, effective therapy.

Implication of all the available evidence

We show that future UDCA response can be predicted. This provides a basis to explore alternative approaches to treatment stratification in PBC, such as earlier introduction of second-line therapy. The model might even now be useful in Precision Medicine initiatives aimed at the identification of predictive biomarkers for treatment or risk stratification in PBC.

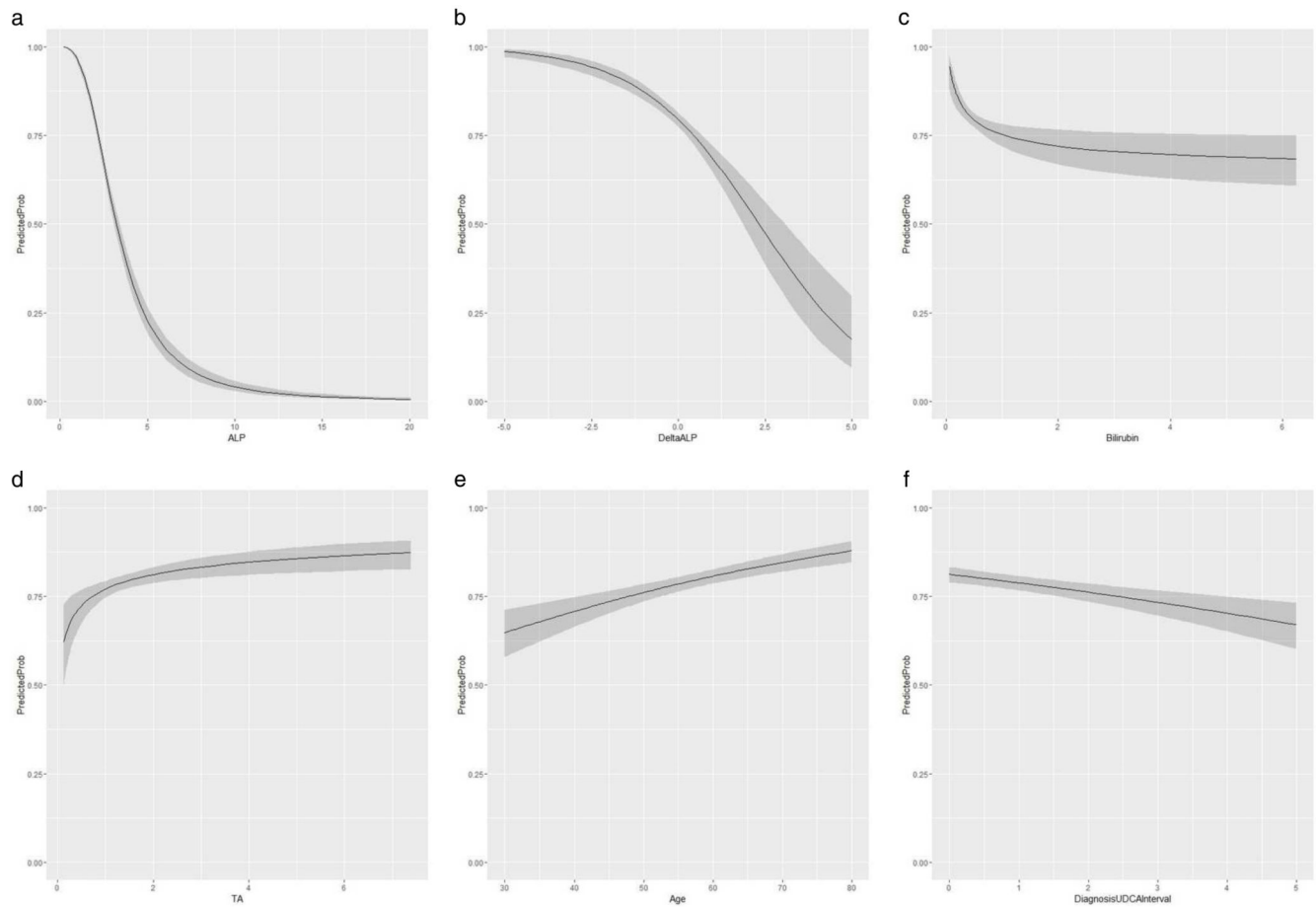


Figure 1.

a-f: Relationship between the variables selected in the best-fitting logistic regression model and the probability of UDCA response.

Abbreviations: ALP_{diag}, alkaline phosphatase (ALP) at diagnosis; TA_{diag}, transaminases at diagnosis; TB_{diag}, total bilirubin at diagnosis.

Note that the ALP, TA and TB are expressed as multiples of their respective upper limits of normal.

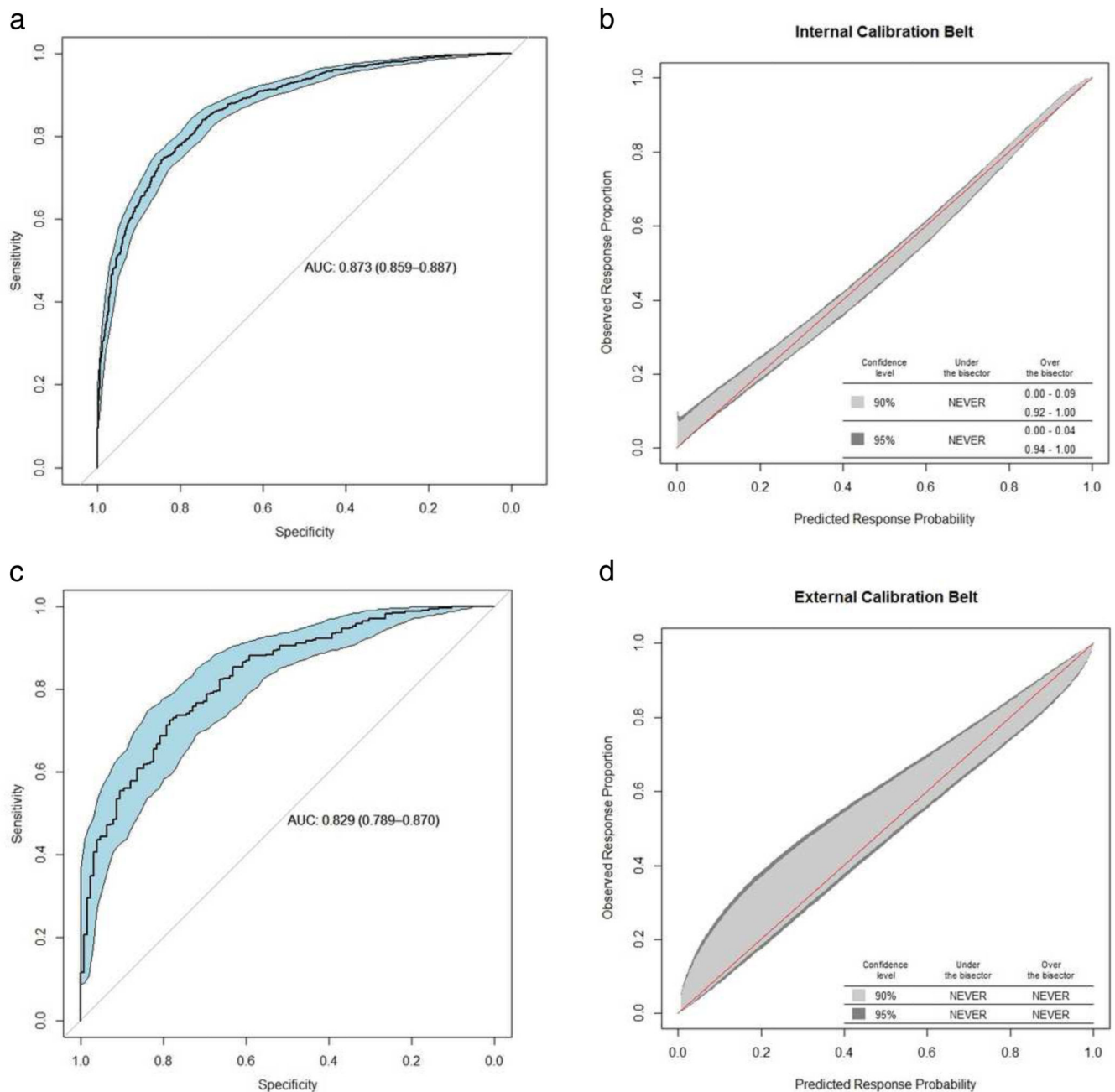


Figure 2.

a-d: (a) Area under the receiver operating characteristic (AUROC) curve for the prediction of response to ursodeoxycholic acid (UDCA) calculated by UDCA Response Score (URS) in the derivation cohort, using stratified bootstrapping to estimate the confidence interval (AUROC = 0.87, CI 0.86 - 0.89); (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 URS in the external validation cohort

(AUROC = 0.83 CI 0.79 - 0.87); **(d)** The predicted versus observed probability of response in the external validation cohort.

Abbreviations: AUC, area under the receiver operating characteristic curve

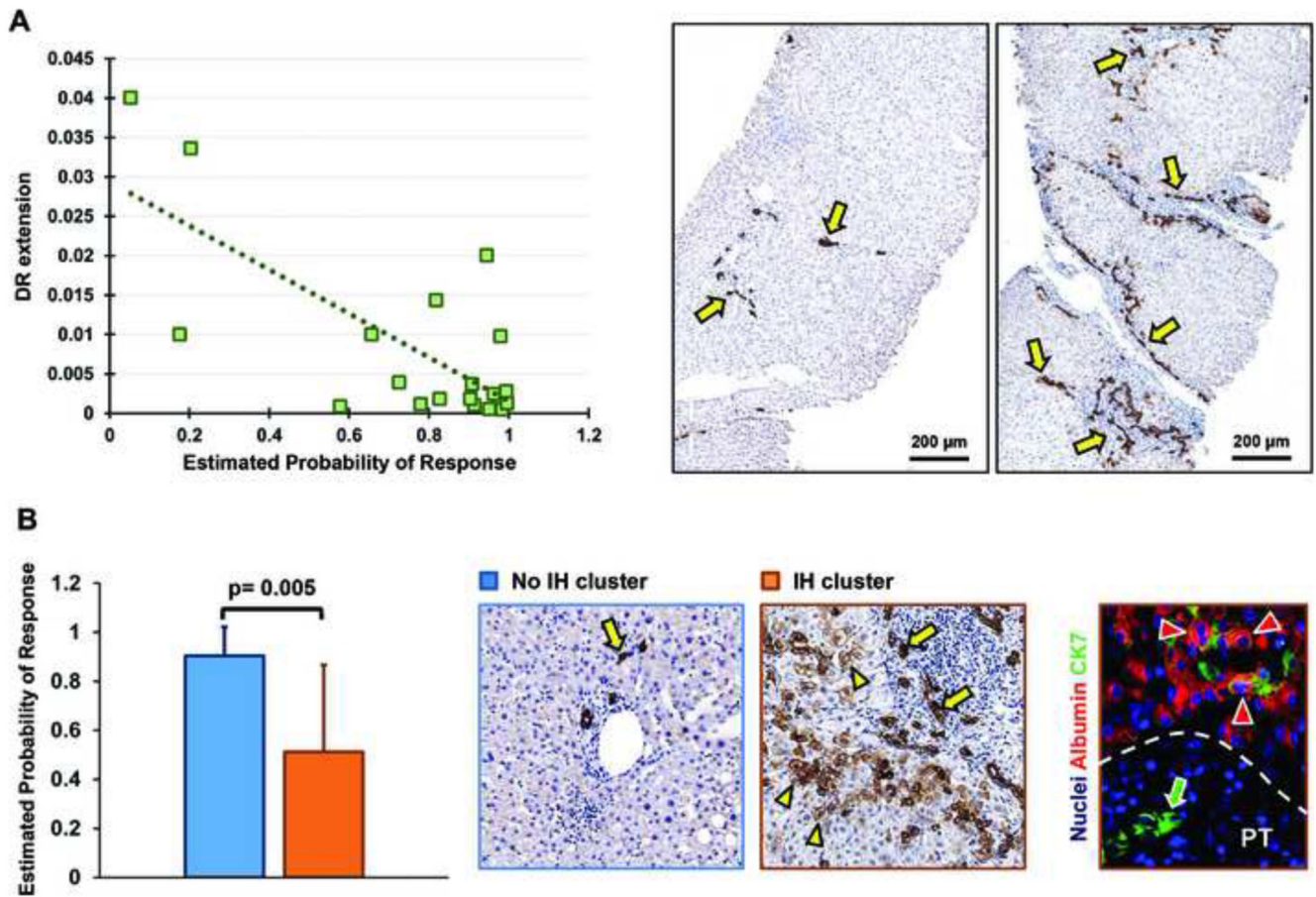


Figure 3.

a-b: **(a)** Correlation of estimated probabilities of response to ursodeoxycholic acid (UDCA) based on the UDCA response score (URS) with the extent of ductular reaction; **(b)** Association of estimated probabilities of UDCA response based on the URS with the presence of intermediate hepatocytes.

Abbreviations: DR, ductular reaction; IH, intermediate hepatocytes.

Table 1
Characteristics of the derivation cohort (n=2,703)

	Median or n	IQR or %
Age, y	56.80	49.52 - 64.16
Female	2,409	89.7%
Treatment time-lag* (days)	75	0 - 258
ALP _{diag} (×ULN)	1.85	1.21 - 3.25
TA _{diag} (×ULN)	1.40	0.90 - 2.25
TB _{diag} (×ULN)	0.53	0.37 - 0.76
PLT _{diag} , x 10 ³ /μL	272	225 - 324
ALB _{diag} (g/L)	41	38 - 44
Creatinine (μmol/l)	76	67 - 86
Sodium (mEq/L)	139	138 - 141
Splenomegaly	263	11.5%
Ascites	48	2.1%
ALP _{T0} (×ULN)	1.91	1.25 - 3.32
TA _{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
TA _{T12} (×ULN)	0.78	0.54 - 1.23
TB _{T12} (×ULN)	0.48	0.35 - 0.65

Abbreviations: ALB_{diag}, albumin at diagnosis; ALP_{diag}, alkaline phosphatase (ALP) at diagnosis; ALP_{T12}, ALP after 12 months of treatment with ursodeoxycholic acid (UDCA) ; PLT_{diag}, platelet count at diagnosis; TA_{diag}, transaminases at diagnosis; TA_{T12}, transaminases after 12 months treatment with UDCA; TB_{diag}, total bilirubin at diagnosis; TB_{T12}, total bilirubin after 12 months of treatment with UDCA.

Note: ALP, TA and TB (at diagnosis, time 0 and time 12) are expressed as multiples of their respective upper limits of normal.

Table 2
Estimated parameters for UDCA response in the model derivation cohort: results from the logistic model based on baseline characteristics (original cohort n=2703, used observations n=2640, missing 2.3%)

Variable	Parameter Estimate	Standard error	Wald statistic	p-value
Intercept	0.774	0.425		
Ln(ALP_{diag} (xULN))	-2.730	0.138	-19.765	<0.0001
1/ TB_{diag} (xULN))	0.600	0.165	3.637	0.0003
Ln(TA_{diag} (xULN))	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
ALP (xULN)	-0.557	0.073	-7.588	<0.0001

Abbreviations: Ln, natural logarithm

Table 3
Clinical scenarios that highlight the impact of each variable on the estimated probability of UDCA response.

Recipient age	ALP _{diag}	TB _{diag}	TA _{diag}	ALP	Treatment time-lag	Estimated probability of UDCA response	95% CI
50	2	0.5	0.5	0	0	0.71	(0.65, 0.76)
50	2	0.5	3	0	0	0.82	(0.79, 0.85)
50	2	1	0.5	0	0	0.65	(0.58, 0.73)
50	2	1	3	0	0	0.78	(0.74, 0.82)
50	2	2	0.5	0	0	0.61	(0.52, 0.70)
50	2	2	3	0	0	0.75	(0.69, 0.80)
50	2	3	0.5	0	0	0.59	(0.49, 0.69)
50	2	3	3	0	0	0.73	(0.67, 0.79)
50	3	2	0.5	0	0	0.34	(0.25, 0.44)
50	3	2	3	0	0	0.49	(0.43, 0.55)
50	4	2	0.5	0	0	0.19	(0.13, 0.27)
50	4	2	3	0	0	0.30	(0.26, 0.36)
50	1	0.5	0.5	0	0	0.94	(0.92, 0.96)
50	1	0.5	0.5	1	2	0.87	(0.83, 0.90)
50	1	0.5	0.5	2	3	0.77	(0.60, 0.83)
50	2	1	1	0	0	0.71	(0.65, 0.75)
50	2	1	1	1	2	0.50	(0.44, 0.57)
50	2	1	1	2	3	0.33	(0.25, 0.52)
30	2	1	1	2	3	0.14	(0.07, 0.23)
70	2	1	1	2	3	0.32	(0.21, 0.46)

Abbreviations: ALP_{diag}, alkaline phosphatase (ALP) at diagnosis; CI, confidence interval; TA_{diag}, transaminases at diagnosis; TB_{diag}, total bilirubin at diagnosis.

Note: ALP is the change in the level of ALP from diagnosis to the start of treatment. Treatment time-lag is the time from diagnosis to the start of treatment. Statistically significant p-values are reported in bold.

The table provides estimates of the probability of UDCA response in hypothetical patients based on the baseline variables. Clearly the effect of one variable on the probability of UDCA response is related to levels of other variables in the model. For example, at lower ALP levels, TA and TB have minimal effect on the probability of UDCA response. At higher ALP levels, however, the effect of TA and TB on the probability of UDCA response is pronounced. Furthermore, the negative effect of TB on the probability of UDCA response is more pronounced at lower TA levels, and less pronounced at higher TA levels.