Digestive and Liver Disease xxx (xxxx) xxx

Contents lists available at ScienceDirect

Digestive and Liver Disease

journal homepage: www.elsevier.com/locate/dld

Liver, Pancreas and Biliary Tract

A quantitative MRCP-derived score for medium-term outcome prediction in primary sclerosing cholangitis

Laura Cristoferi[a,b,c], Marco Porta[d], Davide Paolo Bernasconi[c], Filippo Leonardi[e], Alessio Gerussi[a,b], Giacomo Mulinacci[a,b], Andrea Palermo[a,b], Camilla Gallo[a,b], Miki Scaravaglio[a,b], Eliana Stucchi[a,b], Cesare Maino[d], Davide Ippolito[d,g], Daphne D'Amato[a,b], Carlos Ferreira[h], Alessandra Nardi[i], Rajarshi Banerjee[h], Maria Grazia Valsecchi[c], Laura Antolini[c], Rocco Corso[d], Sandro Sironi[f,g], Stefano Fagiuoli[e], Pietro Invernizzi[a,b], Marco Carbone[a,b,*]

[a] *Division of Gastroenterology, Centre for Autoimmune Liver Diseases, Department of Medicine and Surgery, University of Milano-Bicocca, Via Cadore 48, Monza, MB 20900, Italy*
[b] *European Reference Network on Hepatological Diseases (ERN RARE-LIVER), San Gerardo Hospital, Via Gian Battista Pergolesi 33, Monza 20900, Italy*
[c] *Bicocca Bioinformatics Biostatistics and Bioimaging Centre-B4, School of Medicine and Surgery, University of Milan-Bicocca, Via Cadore 48, Monza, MB 20900, Italy*
[d] *Department of Diagnostic Radiology, San Gerardo Hospital, ASST Monza, Via Gian Battista Pergolesi 33, Monza, MB 20900, Italy*
[e] *Gastroenterology Hepatology and Transplantation Unit, ASST Papa Giovanni XXIII, piazza OMS, 1, Bergamo, BG 24127, Italy*
[f] *Department of Radiology, ASST Papa Giovanni XXIII, piazza OMS, 1, Bergamo, BG 24127, Italy*
[g] *School of Medicine, University of Milano-Bicocca, via Cadore 48, Monza, MB 20900, Italy*
[h] *Perspectum Ltd, Gemini One, 5520 John Smith Dr, Oxford OX4 2LL, United Kingdom*
[i] *Department of Mathematics, University of Rome Tor Vergata, Via della Ricerca Scientifica, 1, Roma, RM 00133, Italy*

ARTICLE INFO

Article history:
Received 2 July 2022
Accepted 21 October 2022
Available online xxx

Keywords:
Artificial intelligence
MRCP
Primary sclerosing cholangitis
Prognostic score

ABSTRACT

Background: Magnetic resonance cholangiopancreatography (MRCP) is the gold standard for diagnosis of patients with primary sclerosing cholangitis (PSC). The semi-quantitative MRCP-derived Anali scores proposed for risk stratification, have poor-to-moderate inter-reader agreement.
Aims: To evaluate the prognostic performance of quantitative MRCP metrics in PSC.
Methods: This is a retrospective study of PSC patients undergoing MRCP. Images were processed using MRCP+ software (Perspectum Ltd, Oxford) that provides quantitative biliary features, semi-automatically extracted by artificial intelligence-driven analysis of MRCP-3D images. The prognostic value of biliary features has been assessed for all hepato-biliary complications.
Results: 87 PSC patients have been included in the analysis. Median follow-up from MRCP to event/censoring of 30.9 months (Q1-Q3=13.6–46.6). An adverse outcome occurred in 27 (31.0%) patients. The number of biliary strictures (HR=1.05 per unit, 95%CI 1.02–1.08, $p < 0.0001$), spleen length (HR=1.16 per cm, 95%CI 1.01–1.34, $p = 0.039$), adjusted for height, age at MRCP, and time from diagnosis to MRCP predicted higher risk of hepatobiliary complications. These were incorporated into a the quantitative MRCP-derived PSC (qMRCP-PSC) score (C-statistic=0.80). After 3-fold cross-validation, qMRCP-PSC outperformed the Anali score in our cohort (C-statistic of 0.78 vs 0.64) and enabled the discrimination of survival of PSC patients (log-rank $p < 0.0001$).

Abbreviations: AI, artificial intelligence; AIH, autoimmune hepatitis; ALP, alkaline phosphatase; AOM, Amsterdam Oxford model; CCA, cholangiocarcinoma; CI, confidence intervals; HR, hazard ratio; IBD, inflammatory bowel disease; LT, liver transplantation; MRCP, magnetic resonance cholangio-pancreatography; MRI, magnetic resonance imaging; MRS, Mayo risk score; PH, proportional hazards; PSC, primary sclerosing cholangitis; qMRCP-PSC, quantitative MRCP-derived PSC score; Q1-Q3, Interquartile range; ROC, receiver operating characteristic; TD-ROC, time dependent receiver operating characteristic.

* Corresponding author at: Center of Autoimmune Liver Disease, University of Milan Bicocca, ERN Rare Liver. Via Cadore 48, 20900, Monza, Italy. Telephone number: +390392334515.
E-mail address: marco.carbone@unimib.it (M. Carbone).

https://doi.org/10.1016/j.dld.2022.10.015
1590-8658/© 2022 The Author(s). Published by Elsevier Ltd on behalf of Editrice Gastroenterologica Italiana S.r.l. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

Please cite this article as: L. Cristoferi, M. Porta, D.P. Bernasconi et al., A quantitative MRCP-derived score for medium-term outcome prediction in primary sclerosing cholangitis, Digestive and Liver Disease, https://doi.org/10.1016/j.dld.2022.10.015

Conclusions: The qMRCP-PSC score identified patients at higher risk of hepatobiliary complications and outperformed the available radiological scores. It represents a novel quantitative biomarker for disease monitoring and a potential surrogate endpoint for clinical trials.

© 2022 The Author(s). Published by Elsevier Ltd on behalf of Editrice Gastroenterologica Italiana S.r.l.
This is an open access article under the CC BY-NC-ND license
(http://creativecommons.org/licenses/by-nc-nd/4.0/)

1. Introduction

Primary sclerosing cholangitis (PSC) is a rare, chronic immune-mediated disease of the biliary tree. The hallmark of disease is focal fibro-inflammatory strictures and secondary dilations of the biliary tree, which are detected using magnetic resonance cholangiopancreatography (MRCP). These can cause chronic cholestasis that, in the long term, may lead to biliary cirrhosis and increased risk of biliary and gallbladder neoplasia. To date, there is no disease-modifying therapy in PSC and, in case of biliary strictures not amenable to endoscopic therapy and causing recurrent bacterial cholangitis, end-stage liver disease, or intractable pruritus, the only therapeutic option is liver transplantation (LT) [1].

A major unmet need in PSC is the lack of established biomarkers of disease activity which enable risk stratification and evaluation of treatment benefit – this prevents an accurate evaluation of efficacy for several novel molecules under investigation. Prognostic models for PSC have been developed which combined markers of disease progression, but no single method can be recommended at present to predict individual patient prognosis [2]. Although alkaline phosphatase (ALP) elevation is characteristic in PSC and its reduction is often a major endpoint in clinical trials, its fluctuation over a short-to-medium time makes it an unreliable disease marker [3].

MRCP is essential for diagnosis, staging, and monitoring the disease. However, its prognostic role in PSC has been hampered by the qualitative or semi-quantitative assessment of the biliary tree, often subject to inter-observer variability. Two semi-quantitative scores based on magnetic resonance imaging (MRI) and MRCP (i.e. the Anali scores with and without gadolinium) showed good prognostic power in PSC [4]. These scores included only radiological variables selected to evaluate both biliary tree characteristics, portal hypertension signs and hepatic morphology, since biochemical markers of cholestasis (except for high serum bilirubin in Anali score without gadolinium) were not independently associated with the outcome. However, despite their good performance in outcome prediction, their use in clinical practice is hampered by a poor-to-moderate intra-observer agreement, as they were built relying on the qualitative nature of MRCP [5]. Semi-automated quantitative assessment of biliary features derived from MRCP images may be a more objective and reproducible approach to evaluate the radiological staging and progression of disease and may represent a promising non-invasive biomarker for outcome prediction in PSC.

A new software for the quantitative analysis of the biliary tree has been recently developed (MRCP+, Perspectum Ltd., Oxford) with this purpose. MRCP+ generates biliary metrics in a semi-automated way by post-processing conventional MRCP images. Artificial intelligence(AI)-driven pathfinding and tubular enhancement algorithms reconstruct a 3D model of the biliary tree and calculate a series of novel quantitative measures for the direct assessment of ductal anatomy. To date, quantitative cholangiographic metrics have never been assessed towards clinical endpoints.

The aims of this study were: (i) to identify quantitative cholangiographic metrics, obtained from routinely collected MRCP images using an AI-derived tool (MRCP+), that are predictive of clinical outcomes of PSC patients; (ii) to compare their prognostic performance with the current available MRCP-derived radiological score (Anali score).

2. Methods

2.1. Study design

This was a retrospective cohort study of PSC patients from two Italian tertiary liver referral centres (San Gerardo Hospital, Monza; ASST Papa Giovanni XXIII, Bergamo). Consecutive PSC patients with at least one MRCP with available MRCP-3D sequences performed in the referral center between Jan-2012 and Dec-2019 were enrolled in the study. The study is conformed to the ethical guidelines of the 1975 Declaration of Helsinki and the protocol was approved by the institutional review board with a waiver of written informed consent. Consent was verbally obtained from participants; however, they were not required to sign it. All radiological and clinical data have been anonymised before being analysed.

2.2. Participant selection

Inclusion criteria were the following: age at inclusion $\geq$ 16 years, diagnosis of large duct PSC according to the European guidelines [6]. The MRCP closest to the diagnosis of PSC for each patient was considered for analysis. Patients with overlap syndrome with autoimmune hepatitis (AIH) were included in the analysis. Only patients with follow up longer than 3 months after study entry were included in the study.

Exclusion criteria were as follows: small duct PSC, previous surgery on biliary tree or LT, MRCP performed on 1 Tesla machine, hepatic comorbidities (viral hepatitis, alcoholic liver disease, or non-alcoholic steatohepatitis), secondary sclerosing cholangitis, hepatic or biliary neoplasms, and decompensated cirrhosis or history of biliary complications (as defined below) at the time of MRCP acquisition.

2.3. MRCP acquisition

All the MRI exams of PSC patients performed using 1.5 or 3 Tesla MRI scanners (Ingenia 1.5T Philips, Achieva 1.5T Philips, Optima MR450w 1.5T GEM, Discovery MR750w 3T GEM) that comprises with T1 and T2-weighted and 3-dimensional isotropic MRCP sequences and good quality images, as defined by an expert biliary radiologist, were included in the analysis. All the patients attended the MRCP exam using the hospital protocol (same in both the centres) after at least 6 h fasting and administration of blueberry juice to reduce signal intensity of overlapping fluid withing the upper gastrointestinal organs. MRCP sequences of PSC patients were read by an expert biliary radiologist. For each patient, spleen maximum coronal diameter (cm) was measured, and presence or absence of gallbladder was derived.

Given the retrospective nature of the study, contrast media was different among patients (according to the radiologist preference or to patients' age), and not always administered. For this reason, Anali score without gadolinium was the only score calculated and considered for the analysis in our cohort. MRCP were scored by an

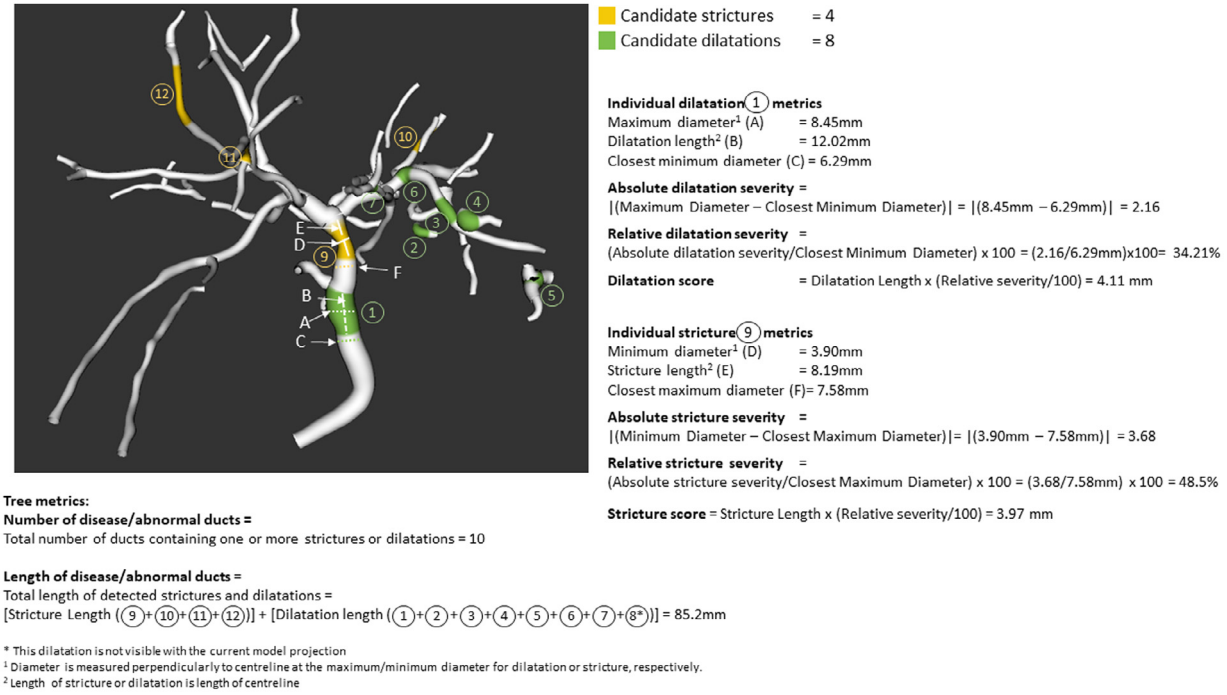

Fig. 1. Biliary tree model demonstrating candidate's stricture and dilations computed by MRCP+ software, shown as a worked example.

expert biliary radiologist blinded to clinical data according to Ruiz et al. [7].

2.4. Quantitative image analysis (MRCP+ protocol)

MRCP-3D sequences were uploaded on a dedicated online portal to Perspectum Ltd. after a complete anonymization and were processed using MRCP+ software. A detailed table of the analytic steps of MRCP+ has been provided in the supplementary materials (Suppl. Table 1). Derivation of quantitative parametric 3D models of the biliary tree was performed after enhancement of tubular structures using multi-scale hessian analysis, intelligent search, and novel duct modeling algorithms. The measurements of duct diameters, duct length, biliary tree and gallbladder volumes were obtained semi-automatically by MRCP+ [8]. Candidate strictures and dilations were defined automatically by MRCP+ based on the absolute and relative changes in diameter along duct profile according to definitions reported in Fig. 1. Metrics were derived for individual ducts and for the whole biliary tree after excluding the cystic and pancreatic ducts (Fig. 1). Potential predictor MRCP+-metrics were chosen based on earlier studies and clinical experience [9,10].

2.5. Variable collection

Clinical data were collected retrospectively from patients records in each center. For all patients the following variables have been collected: date of birth, age at diagnosis, height, age at MRCP acquisition, date of diagnosis, date of last follow-up, presence and type of IBD, diagnosis of AIH overlap, diagnosis of cirrhosis (defined on histological specimen or on typical radiological characteristics), presence of gastro-oesophageal varices and disease phenotype (i.e., classic/small duct PSC).

2.6. Endpoints

Patients were observed from the time of the first available MRCP to a clinical endpoint or censoring. We assessed the performance of the MRCP+ metrics towards all the hepato-biliary complications, defined as the earliest occurring event among the following: (1) listing for LT; (2) liver-related death; (3) diagnosis of portal hypertensive complications such as ascites requiring hospital admission or diuretic therapy and hepatic encephalopathy requiring hospital admission; (4) gastroesophageal bleeding; (5) diagnosis of biliary complications (biliary strictures requiring dilation, stenting or external drainage, or hospitalization for acute bacterial cholangitis); (6) diagnosis of cholangiocarcinoma.

2.7. Statistical analysis

Categorical data are presented as numbers (percentages) while continuous variables are expressed as medians, interquartile ranges(Q1-Q3) unless otherwise stated. Spearman correlation coefficient was used to measure association between continuous variables. All tests were 2-sided with a level of significance of 0.05 was considered.

Cox proportional hazards regression was applied to examine associations between covariates and endpoints defined above. The results were expressed as hazard ratios(HR) and 95% confidence intervals(CI).

For the multivariable model, selection of covariates was guided both by results of the univariate analysis and by known clinical importance.

Spline functions have been used to explore the shape of the HR of clinical and radiologic continuous variables. The Proportional Hazards (pH) assumption for all variables was also checked using plots of scaled Schoenfeld residuals against time. The discriminative ability of single MRCP+-derived metrics and of multivariable models to identify individuals at higher risk of developing each endpoint was assessed using Harrel's C statistic internally validated using k-fold cross-validation.

A time-dependent receiver operating characteristic (TD-ROC) curve was estimated to assess the ability of the score, derived from the multivariable Cox model, to predict events within 24 months of follow-up. The optimal cut-off was chosen maximizing the Youden Index. To visualize the discrimination ability provided by the predictive score, Kaplan-Meier survival rates were calculated

Table 1
Patients' cohort characteristics.

Demographics & clinical variables	Median or N	Q1-Q3 or %
Age at diagnosis (years)	32	23–47
Age at first MRCP (years)	40	26–52
Height (cm)	172	165–179
male gender	53	60.9
center (monza)	60	69.0
IBD	54	62.0
UC	39	72.2
CD	13	24.2
Undetermined	2	3.7
AIH overlap	15	17.2
Cirrhosis	20	23.0
GE varices	11	12.6
Event	27	11.1
Liver-related death	3	29.6
LT	8	3.7
Hepatic decompensation	1	3.7
Variceal bleeding	1	44.4
Biliary complications	12	7.4
CCA	2	
Follow up (months)	30.9	13.6–46.6
Time from diagnosis to MRCP (months)	38.0	5.7–91.0
Time from diagnosis to event/censoring (months)	78.3	37.7–132.6
Radiological variables		
MRCP with contrast medium	44	50.6
Cholecystectomy	16	18.3
Spleen length (cm)	11.9	10.5–13.3
ANALI score without gadolinium	87	
0	32	36.7
1	16	18.4
2	10	11.5
3	14	16.1
4	11	12.7
5	4	4.6

Abbreviations: MRCP: magnetic resonance cholangio-pancreatography; IBD: inflammatory bowel disease; UC: ulcerative colitis; CD: Crohn Disease; AIH: autoimmune hepatitis; GE: gastroesophageal varices; LT: Liver Transplantation; CCA: cholangiocarcinoma.

and compared among two groups defined according to the cut-off chosen using the TDROC curve. Log-Rank test has been used to compare the survival curves between the two groups. Statistical analysis was performed with R software version 4.1.2.

3. Results

3.1. Patients characteristics

One-hundred patients with PSC who performed at least one MRCP in the study period were enrolled. Among them, 13 patients were excluded due to low data quality. Therefore, 87(87.0%) patients were considered for the analysis. Demographic, clinical, and radiological characteristics of PSC patients are presented in Table 1. The median age at diagnosis and at MRCP was 32 (Q1-Q3 23–47) and 40 years (Q1-Q3 26–52), respectively, and 60.9% were male. The median time from diagnosis to first hepatobiliary complications/censoring was 78.3 months [Q1-Q3 37.7–132.6], while the median time from MRCP to first hepatobiliary complication/censoring was 30.9 months [Q1-Q3 13.6–46.6]. A hepatobiliary complication occurred in 27 patients (31.0%); of these, 8 (29.6%) underwent LT, 3 (11.1%) died due to a liver-related cause, 2 (7.4%) developed cirrhosis complications and, 14 (51.9%) had biliary complications/neoplasia (Table 1). Among patients who underwent LT, indications were in 4 jaundice secondary to end-stage biliary disease, in 3 end-stage liver disease and in 1 intractable pruritus. The quantitative cholangiographic metrics analysed and their distribution in the cohort are represented in Table 2.

3.2. Radiological features for outcome prediction

3.2.1. Prognostic performance for hepatobiliary complications

All the MRCP+ metrics had a strong correlation with each other (Spearman coefficient>0.8; Suppl. Fig. 1). Among the MRCP+ metrics, all showed a good prognostic performance at the univariable analysis when tested against all the hepato-biliary complications (Table 3). All MRCP+ metrics and available clinical variables were taken forward for multivariable modeling.

Despite the performance of all variables looking at different pathological aspects of biliary tree (e.g. biliary tree volume and number of dilations, Suppl. Table 2 and Suppl. Fig. 2), showing high degree of correlation, to avoid multicollinearity we have selected in the final model only the number of strictures. This variable was selected based on its higher accuracy in the multivariable model respect to the other quantitative biliary features, easier interpretation for clinicians and radiologists, and the robust evidence on its prognostic role [10,11].

In addition, since portal hypertension has largely been associated with outcome in PSC, we explored the prognostic role of spleen length. As this is significantly influenced by height, age and sex we adjusted the model accordingly [12]; however considering the relatively low number of events and the scarce added value in terms of accuracy of the sex, we kept only the height and age at the time of MRCP for the adjustment. Finally, given that MRCP has not been performed at diagnosis for all the patients we adjusted the model with the 'time from diagnosis to MRCP'.

The final multivariable Cox model, taking correlation structure among covariates and clinical interpretation of their effects into account, included the following variables: number of strictures in biliary tree ($p < 0.001$), spleen length ($p = 0.029$), height

Table 2
Definition and statistical description of selected MRCP+ derived variables.

MRCP + variable	Definition	Median [Q1-Q3]
Biliary tree volume	Total measured biliary tree volume (ml)	10.30 [5.20, 21.95]
Duct number	Total number of modelled ducts (n)	60.00 [31.00, 141.00]
Candidate stricture	A local minimum in modelled duct widths, whose diameter is narrowed with respect to its neighbours (in diameter), by more than 1 mm in absolute terms and more than 30% in relative terms.	
Number of strictures	The total number of areas of decreasing width (strictures)	9.00 [5.00, 17.50]
Sum of strictures' lengths (mm)	Total stricture length (mm) (E in Fig. 1)	68.38 [31.29, 122.64]
Sum of absolute strictures' severity (mm)	Sum of the absolute strictures' severity (Fig. 1)	13.66 [6.61, 26.82]
Sum of relative strictures' severity (%)	Sum of the relative strictures' severity (Fig. 1)	364.90 [206.62, 746.18]
Sum of strictures' score	Sum of severity score for strictures. Composite score are computed as the product of the relative severity (expressed as fraction) and the measured length in mm (Fig. 1)	28.07 [12.68, 51.18]
Candidate dilations	A bile duct dilation with at least 1 mm absolute increase and at least 30% relative increase in diameter, compared to its closest (in diameter) neighbouring local minimum.	
Number of dilations	The total number of areas of increasing width (candidate dilatations).	14.00 [6.00, 32.50]
Sum of dilations' lengths (mm)	The sum of the lengths dilation (B in Fig. 1)	95.04 [42.18, 197.44]
Sum of absolute dilations' severity (mm)	The sum of the absolute severity of the dilations (Fig. 1)	21.77 [9.03, 55.85]
Sum of relative dilations' severity (%)	The sum of relative dilations' severity	981.21 [379.18, 2206.01]
Sum of dilations' score	The sum of severity score for dilations. Composite score are computed as the product of the relative severity (expressed as fraction) and the measured length in mm (Fig. 1)	65.69 [30.04, 138.53]
Number of ducts with strictures AND dilations	The number of ducts containing both one or more candidate strictures and one or more candidate dilatations.	5.00 [2.00, 10.00]
Number of ducts with strictures OR dilations	The number of ducts containing one or more candidate abnormalities (strictures or dilatations)	15.00 [6.50, 32.50]
Sum of lengths of abnormal biliary tree tracts	The total length (mm) of detected candidate abnormalities (Sum of strictures' lengths + Sum of dilations' lengths	158.22 [68.22, 305.53]

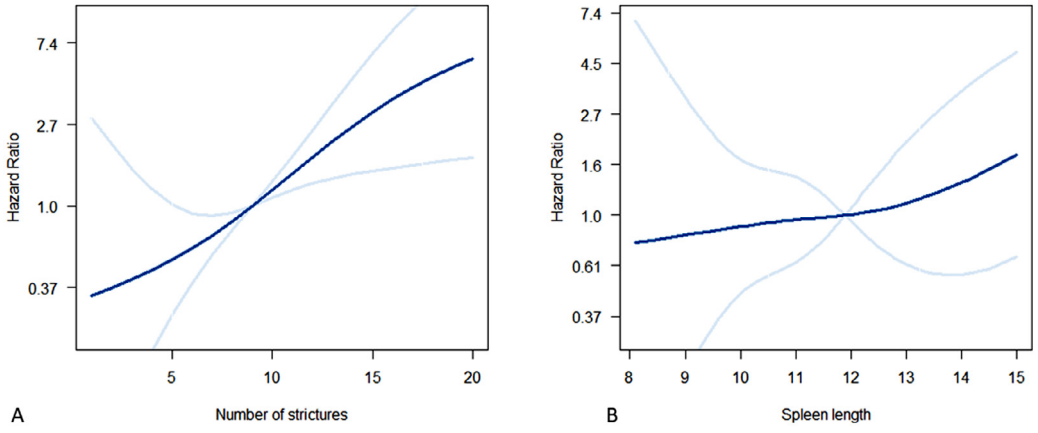

Fig. 2. Cox model HR by spline functions from the qMRCP-PSC score to explore the shape of the association between the number of strictures (A) and spleen length (B) and the endpoint. Despite the small sample size and the consequent large CI, the assumption of a positive linear association between each of the two covariates and the log-hazard of event is tenable.

($p = 0.774$), age at MRCP ($p = 0.419$) and time from diagnosis to MRCP ($p = 0.724$) (Table 3). Spline functions representing shape of the HR of number of strictures and spleen length are represented in Fig. 2. Scaled Schoenfeld residuals plots to evaluate the plausibility of pH assumption are represented in Supplementary Fig. 3.

3.2.2. Derivation of a quantitative MRCP-derived PSC score

We used the regression coefficients of the selected variables (Table 3) to develop a predictive score of all hepatobiliary complications for each patient according to the following formula:

Quantitative MRCP-derived PSC (qMRCP-PSC) score: *Spleen length (cm) * 0.1552 + number of strictures (n) * 0.0509 + age at MRCP (years) * 0.0107 + height* (10 cm) * (−0.067) + *time from diagnosis to MRCP (months) * 0.0010.*

The score demonstrated high discrimination ability with a Harrell's c of 0.80 (Table 3). It ranged from 0.74 to 4.53 with a median of 1.85, Q1-Q3 1.51–2.37, and a mean of 1.92.

To allow for a visual assessment of discriminative power of the qMRCP-PSC score across the study cohort, risk groups were built using as cut-off the one obtained by maximized Youden Index calculated on the time dependent ROC at 24months which was 2.08 (Suppl. Fig. 4). Fig. 3 offers an intuitive depiction of variation in prognosis between patients with different scores according to the derived cut-off (log-rank $p < 0.0001$).

3.3. Cross-validation of qMRCP-PSC score and comparison with Anali score

The qMRCP-PSC score was internally validated with a 3-fold cross-validation approach resulting in a C-statistic of 0.78. This

Table 3
Cox regression analysis for all hepatobiliary complications in PSC cohort.

	Univariable analyses			Multivariable analysis		
	HR	95% CI	P value	HR	95% CI	P value
Number of strictures (n)	1.05	1.03–1.08	<0.001	1.05	1.02–1.08	<0.001
Spleen length (cm)	1.16	0.99–1.26	0.074	1.16	1.01–1.34	0.039
Time from diagnosis to MRCP (months)	0.99	0.99–1.00	0.897	1.00	0.99–1.00	0.724
Height (x10cm)	1.05	0.69–1.58	0.819	0.06	0.00–9.00	0.273
Age at MRCP (years)	1.01	0.99–1.03	0.238	1.02	0.99–1.05	0.119
Age at diagnosis	1.02	0.99 −1.04	0.200			
Sex (male)	0.72	0.33–1.57	0.415			
GE varices at endoscopy	3.53	1.58–7.87	0.002			
Cirrhosis	1.71	0.78–3.77	0.178			
Diagnosis of AIH overlap	0.88	0.30–2.54	0.810			
Concomitant IBD	0.56	0.24–1.26	0.164			
ANALI w/o gadolinium >2	3.32	1.53–7.188	0.002			
Biliary tree volume (ml)	1.05	1.03–1.08	<0.001			
Duct number (n)*	1.06	1.03–1.08	<0.001			
Sum of stricture lengths (mm)*	1.08	1.04–1.12	<0.001			
Sum of absolute strictures severity (%)*	1.47	1.26–1.71	<0.001			
Sum of relative strictures severity (%)*	1.01	1.01–1.02	<0.001			
Sum of strictures scores (mm)*	1.22	1.11–1.34	<0.001			
Number of dilations (n)	1.02	1.01–1.04	<0.001			
Sum of dilations length (mm)*	1.05	1.03–1.07	<0.001			
Sum of absolute dilation severity (%)*	1.14	1.08–1.21	<0.001			
Sum of relative dilation severity (%)**	1.04	1.02–1.05	<0.001			
Sum of dilation scores (mm)*	1.07	1.04–1.10	<0.001			
Number of ducts with strictures **or** dilations (n)	1.03	1.01–1.04	<0.001			
Number of ducts with strictures **and** dilations (n)	1.10	1.05–1.14	<0.001			
Sum of lengths of abnormal biliary tree tracts (mm)*	1.03	1.02–1.05	<0.001			

Abbreviations: AIH: autoimmune hepatitis; CI: confidence interval; GE: gastroesophageal; HR: Hazard Ratio; IBD: inflammatory bowel disease. For continuous variables HR per unit increase is reported unless.
* HR per 10-unit increase is reported.
** HR per 100-unit increase is reported.

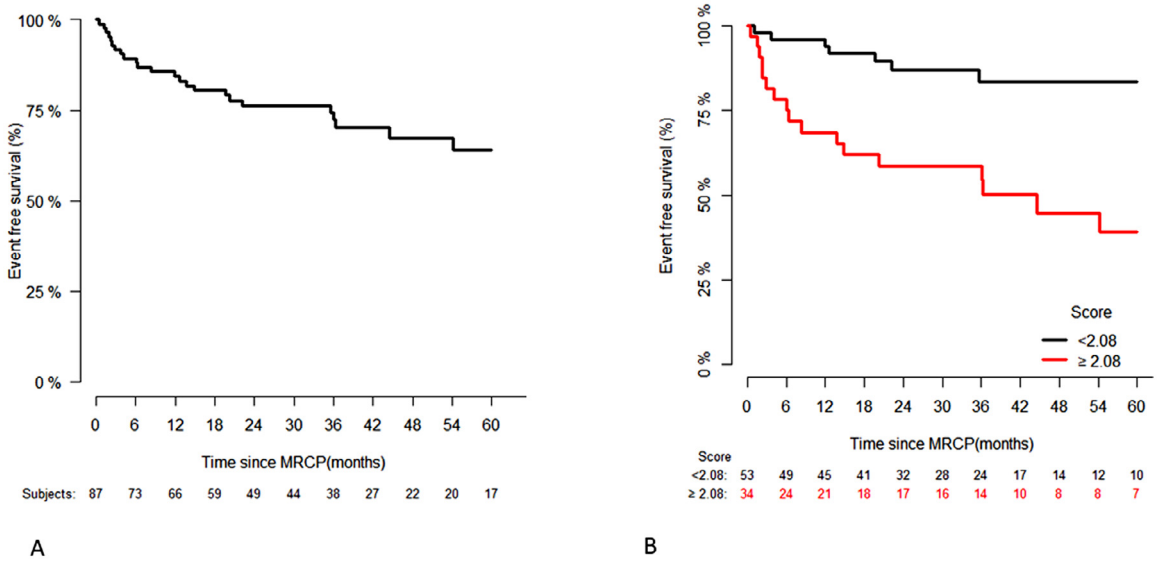

Fig. 3. Kaplan-Meier curves for hepatobiliary-complication-free survival A) in the overall cohort, B) by groups according to qMRCP-PSC score split using maximized Youden Index cut-off derived by time dependent ROC curve (Suppl. Fig. 4) (2.08). Log Rank < 0.0001.

outperformed the Anali score without gadolinium, used as dichotomous variable considering values higher than 2, demonstrating to identify high risk patients, which showed a prognostic performance against hepatobiliary complications (C-statistic=0.64, HR of 3.32 [CI 95% 1.53–7.19], $p = 0.002$) [4].

Finally, we explored the prognostic role of the qMRCP-PSC covariates also toward hard endpoints (i.e., LT, liver-related death, and hepatic decompensation) without adjusting for height, age at MRCP, and time from diagnosis to MRCP due to the lower number of events [13(14.9%)]. The number of strictures and spleen length remain both independently associated with hard outcome development with HR=1.06 ([CI 95 1.02–1.10], $p = 0.0009$) and HR=1.34 ([CI 95 1.13–1.61], $p = 0.004$) with good discriminatory power (C-statistic=0.76)(Suppl. Table 3) also after 3-fold cross validation (C-statistic=0.72). When Anali without gadolinium using values higher than two was tested towards hard outcomes, prognostic prediction was increased with a C-statistic=0.76 (HR=11.67 [CI 95% 2.58–52.69], $p = 0.001$).

4. Discussion

We developed and internally validated a quantitative MRCP-based model for medium-term prediction of hepatobiliary complications in patients with PSC. The scoring system incorpo-

rates quantitative and reproducible metrics semi-automatically extracted from MRCP-3D sequences, and spleen length, adjusted for age, height, and time from diagnosis to MRCP.

The qMRCP-PSC score was superior to the Anali without gadolinium in predicting overall hepatobiliary complications with C-statistic of 0.80 and 0.78 in the derived model and after cross-validation, respectively. Several reasons might explain its stronger performance. Radiological variables were treated as continuous, overcoming the limitation of using qualitative or semi-quantitative variables both in terms of reproducibility and intra-observer agreement [9]. The qMRCP-PSC score incorporates not only a quantitative assessment of biliary tree, but quantifies portal hypertension with a continuous variable, i.e. the spleen length, which often enlarges in PSC even before cirrhosis develops, and has previously been shown as a major prognostic parameter [13–15].

We do acknowledge that the comparison of prediction accuracy of Anali score without gadolinium with that of qMRCP-PSC score might not be entirely fair since the former has been developed toward hard events and without considering biliary complications; however, to date this is the gold-standard as radiological score. Moreover, our model is comparable to the Anali score even if only LT, liver-related death and hepatic decompensation are considered.

To our knowledge, MRCP+ is the only software available for the quantitative analysis of biliary tree, strictures and dilations, using AI-derived sequences from MRCP-3D. It has been demonstrated to be accurate in distinguishing healthy controls from PSC patients and to have a good-to-excellent interobserver agreement for strictures and dilations metrics [9]. Its scan/rescan repeatability and reproducibility across a range of MRI scanners has been previously shown [5].

There are no studies exploring the prognostic role of MRCP+ metrics toward clinical outcomes in PSC. Two studies explored the relation between MRCP+ metrics and estimates of outcomes by prognostic scores (e.g. Mayo Risk Score[MRS] and Amsterdam-Oxford model[AOM]) [11,16]. Selvaraj et al. identified a good positive correlation between the severity of intrahepatic dilations and surrogate markers of disease severity. Differently with respect to our study, the number and severity of strictures were not found to be positively correlated with disease severity nor to have significant risk stratification ability. Conversely, Ismail et al. showed a strong correlation of the total number of segmental strictures with MRS [10]. The difference with the former study may rely on the distinct analytical approach. The Oxford group has normalized the biliary metrics by the total length of the biliary tree, to account for variable lengths of biliary tree among different patients. This might have introduced a bias; indeed, the presence of a stricture could lead to a dilation of the upstream bile ducts, resulting in an increase of the whole biliary tree visualized at MRCP. Indeed, we found in our cohort a linear correlation between the sum of bile duct lengths and the number and severity of both strictures and dilations (Suppl. Fig. 5).

The key variables included in the model were the number of strictures and spleen length. The former is a marker of the biliary damage, which shows a strong correlation with the other biliary metrics explored by MRCP+, reflecting that they quantify all the same pathological process. Similarly to our findings, in AOM, which is based on biliary tree aspect visualised during endoscopic cholangiography, the presence of strictures of intrahepatic and extrahepatic ducts were associated with a worst prognosis over time [11]. Unfortunately, its application on MRCP images showed only moderate agreement [17]; however, this difference between ERCP and MRCP results might be overcome by a more precise strictures' detection method as an AI-driven software.

Spleen length may represent a surrogate marker of disease severity, it might be not necessarily associated with cirrhosis development in PSC [13–15]. In our cohort, the spleen length is independently associated with outcome (Table 3) and it remains significant also in multivariable models using other MRCP-derived metrics as co-variates (Suppl. Table 2, Suppl. Fig. 2).

Patients with PSC tend to develop end-stage liver disease within 10–20 years from diagnosis [6]. The long time for development of hard outcomes (i.e. LT or liver-related mortality) hampers drug development. In any clinical trial with a therapy aimed at modifying the clinical course of PSC, intermediate events need to be included [18]. Biliary complications requiring endoscopic treatment and recurrent biliary infections requiring hospitalization compete with liver fibrosis-related complications in determining future liver outcomes. In addition, patients at early stage of disease may present with an insignificant amount of fibrosis respect to the evidence of advanced biliary strictures and dilations at MRCP. Therefore, in this study we chose to explore together with hard endpoints (i) portal hypertensive complications and (ii) biliary complications. By using only MRCP-derived variables, it is possible to stratify patients into high and low risk of development of all hepatobiliary complications in a medium to short-term follow-up (Fig. 3).

Finally, MRCP+ might have a role also in the early detection of CCA, one of the most worrisome complications of PSC. Unfortunately, only 2 patients in our cohort developed CCA (one intrahepatic and the other peri–hilar). Both showed a qMRCP-PSC score higher than 2.08 (3.21 and 2.23), however, considering the heterogeneity of the disease and of CCA phenotype and presentation in PSC patients, a more appropriate study, designed with this aim, might be more helpful in evaluating this potential application of the software.

Our study has some limitations. The median follow-up from the MRCP to event/censoring is relatively short with 30.9 months (Table 1). However, considering that biliary complications occur earlier in PSC natural history and tend to re-occur becoming a LT indication, a score estimating medium-term outcomes might be useful both in clinical practice and clinical trials to predict intermediate event. Our study is retrospective and performed in two hospitals using two different MRI scanners and, despite the same acquisition protocols, this might have introduced a bias. However, the MRCP-derived metrics of the two groups when compared did not reveals significant differences (data not shown). Indeed, it might be seen as MRCP+ software strength which could analyze images performed from different MRI scans and, thus, more adaptable to clinical practise. We did not include biochemical parameters to adjust our model since, given the retrospective nature of the study, many of the laboratory investigations were not exactly correspondent to the time of the scanning. However, considering the weakness of liver function tests in prediction in PSC and their fluctuation over time [3], we anticipate that they would not have a major impact on the discriminative power of the MRCP-based model. Finally, our study does not have an external validation cohort; however, by using a cross-validation approach we are confirm the qMRCP-PSC demonstrates promising out-of-sample generalisability.

In conclusion, the qMRCP PSC model, which includes the number of strictures and the spleen length, can identify PSC patients at higher risk of adverse outcome and outperforms the available radiological score. This represents a novel biomarker for disease monitoring and a potential surrogate endpoint for clinical trials. Further studies, with prospective, standardized data collection and larger patient's cohorts, are needed for confirmation.

Grant support

This research was partially supported by the Italian Ministry of University and Research (MIUR) - Department of Excellence project PREMIA (PREcision MedIcine Approach: bringing biomarker research to clinic) and PRIN 2020 DEMO:Diseases modeling and

early therapeutic tools in chronic inflammatory cholangiopathies and primary liver cholangiocarcinoma.

Declaration of Competing Interest

RB is CEO at Perspectum Ltd, CF is employed by Perspectum Ltd. MC is a consultant without fee at Perspetum Ltd. LC, MP, DPB, FL, GM, AP, CG, MS, ES, CM, DI, DD, AN, MGV, LA, RC, SS, SF and PI have nothing to disclose.

CRediT authorship contribution statement

Laura Cristoferi: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Visualization, Writing – original draft. **Marco Porta:** Conceptualization, Data curation, Writing – review & editing. **Davide Paolo Bernasconi:** Formal analysis, Methodology, Visualization, Supervision, Writing – review & editing. **Filippo Leonardi:** Resources. **Alessio Gerussi:** Data curation, Writing – review & editing. **Giacomo Mulinacci:** Data curation, Writing – review & editing. **Andrea Palermo:** Data curation, Writing – review & editing. **Camilla Gallo:** Data curation, Writing – review & editing. **Miki Scaravaglio:** Data curation, Writing – review & editing. **Eliana Stucchi:** Data curation, Writing – review & editing. **Cesare Maino:** Data curation, Writing – review & editing. **Davide Ippolito:** Data curation, Writing – review & editing. **Daphne D'Amato:** Data curation, Writing – review & editing. **Carlos Ferreira:** Software. **Alessandra Nardi:** Supervision, Formal analysis. **Rajarshi Banerjee:** Software. **Maria Grazia Valsecchi:** Supervision, Writing – review & editing. **Laura Antolini:** Supervision, Writing – review & editing. **Rocco Corso:** Supervision, Writing – review & editing. **Sandro Sironi:** Supervision, Writing – review & editing. **Marco Carbone:** Conceptualization, Formal analysis, Investigation, Methodology, Visualization, Supervision, Writing – review & editing.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.dld.2022.10.015.

References

[1] Hirschfield GM, Karlsen TH, Lindor KD, et al. Primary sclerosing cholangitis. Lancet 2013;382:1587–99. doi:10.1016/S0140-6736(13)60096-3.
[2] Chapman MH, Thorburn D, Hirschfield GM, et al. British society of gastroenterology and UK-PSC guidelines for the diagnosis and management of primary sclerosing cholangitis. Gut 2019;68:1356–78. doi:10.1136/flgastro-2019-101343.
[3] Trivedi PJ, Muir AJ, Levy C, et al. Inter- and Intra-individual variation, and limited prognostic utility, of serum alkaline phosphatase in a trial of patients with primary sclerosing cholangitis. Clin Gastroenterol Hepatol 2020. doi:10.1016/j.cgh.2020.07.032.
[4] Lemoinne S, Cazzagon N, El Mouhadi S, et al. Simple magnetic resonance scores associate with outcomes of patients with primary sclerosing cholangitis. Clin Gastroenterol Hepatol 2019;17:2785–92 e3. doi:10.1016/j.cgh.2019.03.013.
[5] Grigoriadis A, Ringe KI, Andersson M, et al. Assessment of prognostic value and interreader agreement of ANALI scores in patients with primary sclerosing cholangitis. Eur J Radiol 2021;142:109884 Sep. doi:10.1016/j.cgh.2019.03.013.
[6] European Association for the Study of the LiverEASL clinical practice guidelines: management of cholestatic liver diseases. J Hepatol 2009;51:237–67. doi:10.1016/j.jhep.2009.04.009.
[7] Ruiz A, Lemoinne S, Carrat F, et al. Radiologic course of primary sclerosing cholangitis: assessment by three-dimensional magnetic resonance cholangiography and predictive features of progression. Hepatology 2014;59:242–50. doi:10.1002/hep.26620.
[8] Goldfinger MH, Ridgway GR, Ferreira C, et al. Quantitative MRCP imaging: accuracy, repeatability, reproducibility, and cohort-derived normative ranges. J Magn Reson Imaging 2020;52:807–20. doi:10.1002/jmri.27113.
[9] Selvaraj EA, Ba-Ssalamah A, Poetter-Lang S, et al. A quantitative magnetic resonance cholangiopancreatography metric of intrahepatic biliary dilatation severity detects high-risk primary sclerosing cholangitis. Hepatol Commun 2021. doi:10.1002/hep4.1860.
[10] Ismail MF, Hirschfield GM, Hansen B, et al. Evaluation of quantitative MRCP (MRCP+) for risk stratification of primary sclerosing cholangitis: comparison with morphological MRCP, MR elastography, and biochemical risk scores. Eur Radiol 2022;32:67–77. doi:10.1007/s00330-021-08142-y.
[11] Ponsioen CY, Vrouenraets SME, Prawirodirdjo W, et al. Natural history of primary sclerosing cholangitis and prognostic value of cholangiography in a Dutch population. Gut 2002;51:562–6. doi:10.1136/gut.51.4.562.
[12] Chow KU, Luxembourg B, Seifried E, et al. Spleen size is significantly influenced by body height and sex: establishment of normal values for spleen size at us with a cohort of 1200 healthy individuals. Radiology 2016;279:306–13. doi:10.1148/radiol.2015150887.
[13] Jung F, Cazzagon N, Vettorazzi E, et al. Rate of spleen length progression is a marker of outcome in patients with primary sclerosing cholangitis. Clin Gastroenterol Hepatol 2019;17:2613–15. doi:10.1016/j.cgh.2018.12.033.
[14] Ehlken H, Wroblewski R, Corpechot C, et al. Spleen size for the prediction of clinical outcome in patients with primary sclerosing cholangitis. Gut 2016;65:1230–2. doi:10.1136/gutjnl-2016-311452.
[15] Eaton JE, Welle CL, Monahan H, et al. Comparative performance of quantitative and qualitative magnetic resonance imaging metrics in primary sclerosing cholangitis. Gastro Hep Adv 2022;1:287–95. doi:10.1016/j.gastha.2022.
[16] Kim WR, Therneau TM, Wiesner RH, et al. A revised natural history model for primary sclerosing cholangitis. Mayo Clin Proc 2000;75:688–94. doi:10.4065/75.7.688.
[17] Tenca A, Mustonen H, Lind K, et al. The role of magnetic resonance imaging and endoscopic retrograde cholangiography in the evaluation of disease activity and severity in primary sclerosing cholangitis. Liver Int 2018;38:2329–39. doi:10.1111/liv.13899.
[18] Ponsioen CY, Lindor KD, Mehta R, et al. Design and endpoints for clinical trials in primary sclerosing cholangitis. Hepatology 2018;68:1174–88. doi:10.1002/hep.29882.