A Quantitative Magnetic Resonance Cholangiopancreatography Metric of Intrahepatic Biliary Dilatation Severity Detects High-Risk Primary Sclerosing Cholangitis

Emmanuel A. Selvaraj,[1-4] Ahmed Ba-Ssalamah,[5] Sarah Poetter-Lang,[5] Gerard R. Ridgway,[6] J. Michael Brady,[6,7] Jane Collier,[2] Emma L. Culver,[2-4] Adam Bailey,[2-4] and Michael Pavlides[1-4]

Magnetic resonance imaging with magnetic resonance cholangiopancreatography (MRI-MRCP) in primary sclerosing cholangitis (PSC) is currently based on qualitative assessment and has high interobserver variability. We investigated the utility and performance of quantitative metrics derived from a three-dimensional biliary analysis tool in adult patients with PSC. MRI-MRCP, blood-based biomarkers, and FibroScan were prospectively performed in 80 participants with large-duct PSC and 20 healthy participants. Quantitative analysis was performed using MRCP+ (Perspectum Ltd., United Kingdom), and qualitative reads were performed by radiologists. Inter-reader agreements were compared. Patients were classified into high risk or low risk for disease progression, using Mayo risk score (MRS), Amsterdam-Oxford model (AOM), upper limit of normal (ULN) alkaline phosphatase (ALP), disease distribution, and presence of dominant stricture. Performance of noninvasive tools was assessed using binomial logistic regressions and receiver operating characteristic curve analyses. Quantitative biliary metrics performed well to distinguish abnormal from normal bile ducts ($P < 0.0001$). Interobserver agreements for MRCP+ dilatation metrics (intraclass correlation coefficient, 0.90-0.96) were superior to modified Amsterdam intrahepatic stricture severity score ($\kappa = 0.74$) and Anali score ($\kappa = 0.38$). MRCP+ intrahepatic dilatation severity showed excellent performance to classify patients into high-risk and low-risk groups, using predictors of disease severity as the reference (MRS, $P < 0.0001$; AOM, $P = 0.0017$; $2.2 \times$ ULN ALP, $P = 0.0007$; $1.5 \times$ ULN ALP, $P = 0.0225$; extrahepatic disease, $P = 0.0331$; dominant stricture, $P = 0.0019$). MRCP+ intrahepatic dilatation severity was an independent predictor of MRS >0 (odds ratio, 31.3; $P = 0.035$) in the multivariate analysis. *Conclusion:* Intrahepatic biliary dilatation severity calculated using MRCP+ is elevated in patients with high-risk PSC and may be used as an adjunct for risk stratification in PSC. This exploratory study has provided the groundwork for examining the utility of novel quantitative biliary metrics in multicenter studies. (*Hepatology Communications* 2022;6:795-808).

Primary sclerosing cholangitis (PSC) is a chronic immune-mediated liver disease characterized by intrahepatic and extrahepatic bile duct inflammation and fibrosis leading to multifocal biliary strictures.[1] PSC is insidious, with nearly half of patients being asymptomatic at diagnosis.[2] The rate

of disease progression and development of complications is variable.[3] In the absence of effective medical therapies, liver transplantation is the only proven life-extending intervention for patients with end-stage liver disease.

There are currently no established tools that reliably estimate prognosis in the individual patient. Two prognostic risk models, the revised Mayo risk score (MRS) and Amsterdam-Oxford model (AOM), have shown reasonable discriminatory performance and predictive accuracy estimates of survival.[4,5] Noninvasive surrogate markers of disease severity and progression that are commonly used in clinical practice and trial endpoints include serum alkaline phosphatase (ALP), FibroScan liver stiffness (LS), and the enhanced liver fibrosis (ELF) test. Repeated staging with invasive liver biopsy or endoscopic retrograde cholangiopancreatography (ERCP) is generally avoided given the associated risk and discomfort to the patient.

Magnetic resonance imaging with magnetic resonance cholangiopancreatography (MRI-MRCP) is the imaging modality of choice in establishing the diagnosis, staging cholangiopathy, and surveillance for biliary complications in PSC.[6] This supports the use of MRI-MRCP features as predictors of disease severity and progression in PSC. If such predictive features could be found, the noninvasive nature of MRI-MRCP would make it a safe and attractive option for the development of imaging biomarkers that can be used in clinical practice and interventional trials.

Current definitions of strictures and dilatations in PSC rely on recognition of morphologic changes on MRI-MRCP and radiologists' judgements of the severity of changes. Reliance on qualitative descriptors leads to high interobserver variability in interpretation, even among experts in PSC.[7] This makes developing MRI-based biomarkers using currently accepted standards particularly challenging. Two qualitative MRI-MRCP scores have been proposed. The Amsterdam cholangiographic classification was originally designed for ERCP interpretation and predicted transplant-free survival in two cohorts.[8,9] A modified Amsterdam stricture severity score has been adapted to MRCP interpretation.[10] The Anali score, which incorporates hepatic dysmorphy, portal hypertension, and severity of intrahepatic bile duct dilatation, predicted 4-year radiologic progression from baseline with good accuracy, and its prognostic value was subsequently validated in a retrospective multicenter cohort study.[11,12] The severity of intrahepatic bile duct dilatation was assessed using a categorical grading scale based on the

ARTICLE INFORMATION:

From the ¹Oxford Centre for Clinical Magnetic Resonance Research, Radcliffe Department of Medicine, University of Oxford, Oxford, United Kingdom; ²Translational Gastroenterology Unit, Nuffield Department of Medicine, University of Oxford, Oxford, United Kingdom; ³National Institute for Health Research Oxford Biomedical Research Centre, University of Oxford, Oxford, United Kingdom; ⁴Oxford University Hospitals National Health Service Foundation Trust, Oxford, United Kingdom; ⁵Medical University of Vienna, Department of Biomedical Imaging and Image-Guided Therapy, General Hospital of Vienna, Vienna, Austria; ⁶Perspectum Ltd., Oxford, United Kingdom; ⁷Department of Oncology, Medical Sciences Division, University of Oxford, Oxford, United Kingdom.

ADDRESS CORRESPONDENCE AND REPRINT REQUESTS TO:

Michael Pavlides, D.Phil
Oxford Centre for Clinical Magnetic Resonance Research
Radcliffe Department of Medicine, University of Oxford
Headington, Oxford OX3 9DU, United Kingdom
E-mail: michael.pavlides@cardiov.ox.ac.uk
Tel.: +44 1865 234577

maximum diameter of an intrahepatic bile duct. This does not consider the continuous nature of diameter measurements or the potential multiplicity of ductal strictures or dilatations in a patient. Furthermore, interobserver variability was not assessed.

Quantitative MRI-MRCP techniques and automated derivation of bile duct diameter profiles could potentially provide a more objective and nuanced assessment of the biliary tree measurements. To this end, a quantitative biliary tree analysis software (MRCP+; Perspectum Ltd., Oxford, United Kingdom) has recently been developed. This tool generates a multiplicity of measurements of the biliary tree derived from three-dimensional (3D) MRCP images, using advanced image processing and computational modeling techniques.[13] The intraindividual repeatability, reproducibility across different scanners, and interobserver and intraobserver agreements of MRCP+ metrics have been tested in healthy volunteers and a small population of patients with unselected biliary diseases.[13] MRCP+ performance to distinguish between PSC and autoimmune sclerosing cholangitis has been assessed in a pediatric population.[14]

In this proof-of-concept study, we evaluated the (i) difference in MRCP+ metrics between participants with PSC and healthy volunteers; (ii) association between MRCP+ metrics and noninvasive covariates in PSC; (iii) interobserver agreement for MRCP+ metrics in PSC; (iv) performance of MRCP+ metrics against LS, ELF, the modified Amsterdam stricture severity score, and the Anali score in identifying the high-risk group defined by the MRS and AOM; and (v) difference in MRCP+ metrics stratified by upper limit of normal (ULN) ALP, disease distribution, and presence of a dominant stricture.

Patients and Methods

STUDY DESIGN

We conducted a prospective, single-center, cross-sectional study at a tertiary nontransplant center in the United Kingdom (Oxford University Hospitals National Health Service Foundation Trust) between September 2018 and November 2020. The study protocol conformed to the ethical guidelines of the 1975 Declaration of Helsinki and the principles of good clinical practice. All participants provided written informed consent. Ethical approval was granted by the UK Research Ethics Service (18/SC/0367) and by the local Research and Development Department.

PARTICIPANT SELECTION

Adult patients (≥18 years) with a known diagnosis of large-duct PSC were consecutively recruited from outpatient clinics. Large-duct PSC was defined according to the American Association for the Study of Liver Diseases (AASLD) guidelines as the association of chronic cholestasis, typical features on MRCP or ERCP, and no cause of secondary sclerosing cholangitis. Patients were excluded if they had contraindications to MRI scanning, other proven or suspected coexisting cholangiopathy, small-duct PSC, overlap with primary biliary cholangiopathy, previous choledochojejunostomy, previous liver transplant, cholangiocarcinoma, hepatocellular carcinoma, cirrhosis decompensation at the time of inclusion, clinical or laboratory evidence of a liver diagnosis other than PSC, or consumed more alcohol than the current limit recommended by the UK Department of Health (14 units/week or 16-24 g/day).

Given the absence at the time of study conception of normal values for MRCP+, matched healthy volunteers were recruited for reference. Adults (≥18 years) were considered healthy if they had normal liver biochemistry, no previous self-reported history of liver or biliary disease or intervention, alcohol consumption within the current recommended limit as above, and body mass index not greater than 25 kg/m^2.

MRI-MRCP ACQUISITION

All participants attended a study visit after fasting for at least 4 hours. T2-weighted 3D MRCP, T1 and T2* mapping, as well as 3D volume interpolated breath-hold examination scans were acquired at the University of Oxford Center for Clinical Magnetic Resonance Research on a 3T scanner (Siemens Magnetom Prisma, Erlangen, Germany). All participants consumed 200 mL of pineapple juice 20 minutes before the scan to reduce the signal intensity of overlapping fluid within the stomach and duodenum.[15] A heavily T2-weighted, navigator-based, respiratory-triggered, coronal 3D MRCP was acquired during the quiescent portion of end expiration by using an 18-channel

phased-array abdominal coil. Details of the acquisition protocol are reported in Supporting Table S1.

MRCP+ ANALYSIS

MRCP images were uploaded and transferred by a dedicated online portal to Perspectum Ltd. to be processed using MRCP+ by trained radiographers. Quantitative parametric 3D models of the biliary tree and pancreatic ducts were derived (Fig. 1). Automatic measurements of duct diameters, duct length, as well as the biliary tree and gallbladder volumes were obtained. Details of the analysis steps are included in Supporting Table S2. The threshold for classification of a candidate stricture and dilatation is automatic and was already predefined in the MRCP+ algorithm to meet both an absolute and relative change in bile duct diameter criteria as summarized in Table 1 and Supporting Table S3. These thresholds were based on
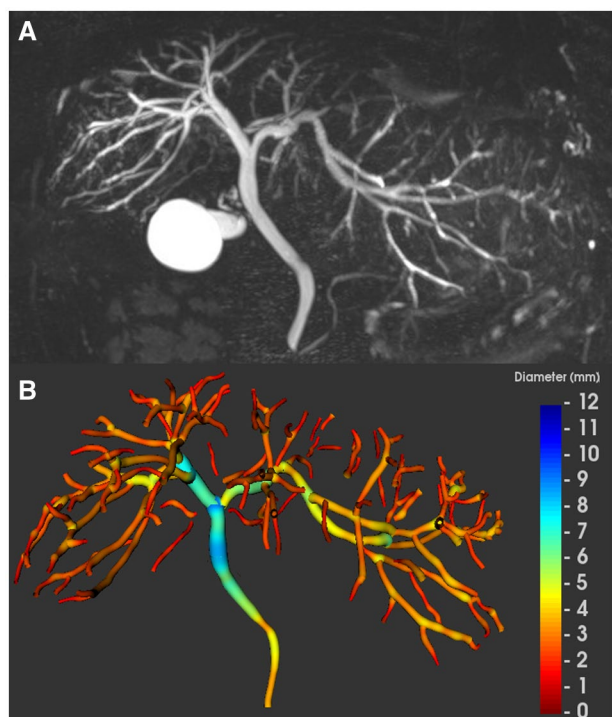
internal performance testing against synthetic digital phantom instances and scans of a 3D-printed phantom that demonstrated the sensitivity was 77% for stricture detection and 86% for dilatation detection with no false-positive detections.[13] Metrics were derived for individual ducts, for the whole biliary tree after excluding the cystic and pancreatic ducts, for the extrahepatic segment (common bile duct, common hepatic duct, and the main right and left hepatic ducts up to their first notable branches), and for the intrahepatic segment (starting from the first branches of the main right and left hepatic ducts). All variables relating to the whole biliary tree (e.g., total number of dilatations) were normalized and reported as per meter of biliary tree to allow accurate comparisons between participants. These normalizations were not carried out to variables relating to defined anatomical structures (e.g., gallbladder volume, number of dilatations in the common bile duct). The quantitative metrics derivation from MRCP+ analysis used in this study is annotated in Fig. 2. For interobserver reliability testing, MRCP images were independently reanalyzed by two trained radiographers using MRCP+ blinded to clinical and qualitative MRI-MRCP data.

MRI-MRCP INTERPRETATION

Both MRI and MRCP sequences of patients with PSC were read by two expert hepatopancreatobiliary radiologists with a subspecialist interest in PSC, (A.B. and S.P.L., with 25 and 8 years of experience, respectively). Both radiologists independently scored the components of the modified Amsterdam stricture severity scores (Supporting Table S4)[10] and Anali score (Supporting Table S5).[11] A dominant stricture was defined as a stricture <1.5 mm diameter in the common bile duct or <1 mm in the left or right main hepatic ducts, as per AASLD guidelines. Radiologists were blinded to clinical and MRCP+ data.

CLINICAL AND LABORATORY DATA

Before the MRI-MRCP scan, fasting blood samples were taken for ALP, aspartate aminotransferase (AST), alanine aminotransferase, gamma-glutamyltransferase, bilirubin, albumin, platelet count, prothrombin time, and ELF (Siemens Healthineers,

FIG. 1. MRCP and MRCP+ of a 32-year-old man with extrahepatic and intrahepatic PSC. (A) Maximum intensity projection image and (B) the corresponding parametric biliary 3D tree model generated from MRCP+, color coded according to duct diameter. The gallbladder has been automatically segmented and its volume presented separately. We excluded the cystic and pancreatic ducts for the biliary analysis in the PSC reported here.

TABLE 1. DEFINITIONS OF QUANTITATIVE MRCP+ BILIARY DILATATION METRICS

Quantitative MRCP Metrics	Definition
Individual dilatation metrics	
Candidate dilatation	Bile duct dilatation with at least 1 mm absolute increase and at least 30% relative increase in diameter compared to its closest (in diameter) neighboring local minimum
Relative severity of candidate dilatation	$\dfrac{\text{Maximum dilatation diameter} - \text{Nearest local minimum diameter}}{\text{Nearest local minimum diameter}}$
Length of candidate dilatation (mm)	Full width at half maximum of candidate dilatation diameter profile
Dilatation score (mm)	Length of candidate dilatation $\times$ Relative severity of dilatation
Tree dilatation metrics*	
Biliary tree length (m)	Total length of the biliary tree modeled by MRCP+
Number of dilatations, DilatNum (m^{-1})	$\dfrac{\text{Total number of candidate dilatations}}{\text{Total length of measured biliary tree in meters}}$
Sum of relative severity of dilatations, SumRelSevDilat (m^{-1})	$\dfrac{\text{Sum of relative severity of all candidate dilatations}}{\text{Total length of measured biliary tree in meters}}$
Proportion of tree dilated, DilatTreeProp (%)	$\dfrac{\text{Sum of lengths of candidate dilatations in meters}}{\text{Total length of measured biliary tree in meters}} \times 100$
Tree dilatation score, TreeDilatScore	$\dfrac{\text{Sum of candidate dilatation scores in meters}}{\text{Total length of measured biliary tree in meters}}$
Volume metrics	
Biliary tree volume,* TreeVol (mL m^{-1})	$\dfrac{\text{Total measured biliary tree volume}}{\text{Total length of measured biliary tree in meters}}$
Gallbladder volume (mL)	Fasted volume of the segmented gallbladder

*These metrics were normalized to the total length of the measured biliary tree to account for variable lengths of the biliary tree modeled on MRCP+ among participants.
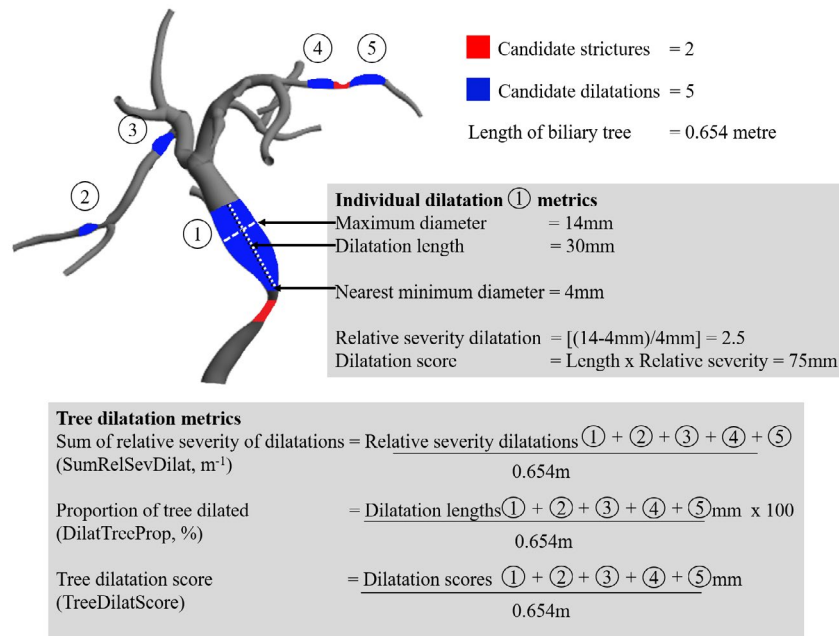

FIG. 2. Biliary tree model demonstrating candidate strictures and dilatations computed by MRCP+ analysis. Dilatation metrics derived for an individual candidate dilatation and the whole tree are shown as a worked example. Tree metrics have been normalized to the total length of the biliary tree in meters.

Erlangen, Germany). Transient elastography LS was performed using FibroScan 502 Touch (Echosens, Paris) by trained operators using the automatic probe selection tool on the same day. For a successful measurement, 10 valid readings were required, and as per recommended guidelines, unreliable readings were defined as having an interquartile range (IQR)/median ratio >0.3. Electronic medical records were searched to obtain relevant clinical data (age at diagnosis, presence of inflammatory bowel disease [IBD], medication history, history of variceal bleeding, and distribution of disease).

RISK CLASSIFICATION

Surrogate predictive markers of disease progression were used as the reference in this study. Patients with PSC were classified into high-risk or low-risk groups according to published criteria as described below.

Prognostic Risk Models

MRS (based on patient age, bilirubin, albumin, AST, and history of variceal bleeding) provides 4-year mortality risk of patients with PSC.[16] Patients were categorized into (1) low risk, ≤0; (2) intermediate risk, >0-2; and (3) high risk, >2. For the analysis, the intermediate- and high-risk categories were grouped together as in other studies of similar size[17,18]; we defined MRS > 0 as high risk to avoid small numbers in a group.

The AOM (based on PSC subtype, age, albumin, ALP, AST, bilirubin, and platelet count) predicts the long-term risk of PSC-related death and/or liver transplantation.[19] Although initially modeled with parameters at the time of diagnosis, a subsequent validation study has shown similar accuracy when the score was recalculated at different time points during follow-up and a cut-off score of 2 separated the low-risk from the high-risk group.[4] We used this cutoff to define high-risk (AOM > 2) and low-risk (AOM ≤ 2) groups in our study.

Serum ALP

While historically a cut-off value of 1.5 × ULN has been reported to have prognostic implications,[20] higher cutoffs of ≥2.2 × ULN have been proposed to have a better predictive ability of liver transplantation

and survival in a large cohort of patients with PSC in the United Kingdom.[21] We examined both cutoffs.

Liver Fibrosis Markers

High-risk PSC in our cohort was defined as LS by transient elastography >9.6 kPa[22] or ELF score > 9.8[23,24] as these cutoffs correlate with histologically assessed fibrosis and clinical outcomes.

Disease Distribution and Dominant Stricture

The presence of extrahepatic biliary disease was associated with adverse outcomes in a large UK multicenter PSC research cohort.[21] The same disease distribution classification was applied to our study. The presence of dominant strictures was associated with a high risk of developing cholangiocarcinoma and worse outcomes.[25]

STATISTICAL ANALYSIS

Descriptive statistics were used to summarize participant characteristics. Associations between variables were tested using Spearman's correlation coefficient. Interobserver and intraobserver variability of MRCP+ metrics in PSC was assessed by the intraclass correlation coefficient (ICC) based on a two-way mixed-effects model and strength classified as reported.[26] Interobserver variability in qualitative MRI-MRCP interpretation was evaluated using Cohen's κ coefficient, and strength was classified as reported.[27]

Receiver operating characteristic curve analyses were performed to assess the diagnostic performance of MRCP+ metrics and other noninvasive risk stratification tools to discriminate high- and low-risk PSC according to the risk classification described above. Area under the curve (AUC) was computed, and sensitivity and specificity were recorded at the cut-off point that maximized the Youden index. For MRCP+ analysis, we examined whole-biliary tree metrics and metrics relating to the intrahepatic ducts only as intrahepatic duct dilatation had prognostic significance in the Anali score.[12] For the qualitative MRI-MRCP scores, results from both readers are presented, and where appropriate, we used results from reader 1 (A.B.), selected based on their prior experience, for the main analysis.

Multivariate logistic regression analyses were performed to identify the best predictive variables of

high-risk PSC as stratified by MRS and AOM. Covariates with $P < 0.10$ in the univariate analysis were entered into the multivariate analysis. Variables that contributed to MRS and AOM scores were excluded. Linear variables were entered as continuous variables. Potential multicollinearity between MRCP+ metrics was assessed using thresholds of Spearman's rank correlation coefficient >0.80 and a variance inflation factor <1 or >10.

Statistical significance was set at $P < 0.05$, and 95% confidence intervals (CIs) were calculated where appropriate. Analysis was performed using GraphPad Prism version 9.0 for Windows (San Diego, CA) and IBM SPSS (v25; IBM Corp., Armonk, NY).

Results

COHORT CHARACTERISTICS

We recruited 80 patients with large-duct PSC to the study. MRCP+ analysis was rejected in 4 patients due to low data quality from the original image acquisitions (Supporting Table S6), leaving 76 for the final analysis. Radiologist reads of MRI-MRCP were available in all 80 patients, but only the 76 with corresponding MRCP+ data were included in the comparative analysis. The outcome of MRI-MRCP reads is summarized in Supporting Table S7. The median age of patients with PSC was 44 years (IQR, 31-56 years), and 67% were men. The median disease duration was 8 years (IQR, 4-12 years), 55% were on ursodeoxycholic acid therapy, 75% had concomitant IBD (55% ulcerative colitis), and 54% had both intrahepatic and extrahepatic disease distribution. The clinical and biochemical parameters for patients with PSC are summarized in Table 2.

Normal reference values were established from 20 age- and sex-matched healthy volunteers after excluding one poor quality MRCP acquisition due to motion artifact. The median age was 35 years (IQR, 31-39 years; $P = 0.07$, compared to patients), and 65% were men ($P = 0.80$, compared to patients). FibroScan was successful in all participants.

COMPARISON OF MRCP+ METRICS BETWEEN PATIENTS WITH PSC AND HEALTHY VOLUNTEERS

Biliary tree models representative of both cohorts are shown in Fig. 3. All MRCP+ metrics were compared between patients with PSC and healthy

TABLE 2. BASELINE CHARACTERISTICS OF 76 PATIENTS WITH PSC INCLUDED IN THE STUDY

Characteristic	PSC (n = 76)	
Male sex	51	(67)
Age (years)	44	(31-56)
PSC duration (years)	8	(4-12)
PSC disease distribution		
Intrahepatic + extrahepatic	41	(54)
Intrahepatic only	35	(46)
On UDCA therapy	42	(55)
IBD present	57	(75)
IBD phenotype		
Ulcerative colitis	42	(55)
Crohn's	8	(11)
Unspecified	7	(9)
Laboratory parameters		
Total bilirubin (µmol/L)	13	(10-20)
ALT (IU/l)	42	(27-84)
AST (IU/l)	38	(27-62)
ALP (IU/l)	150	(101-234)
×ULN ALP	1.2	(0.8-1.8)
GGT (IU/L)	123	(49-306)
Albumin (g/L)	40	(38-42)
Prothrombin time (seconds)	10.6	(10.3-11.0)
Platelet count (×10⁹/L)	254	(193-302)
Liver fibrosis markers		
Liver stiffness (kPa)	6.8	(5.3-10.0)
ELF score	9.4	(8.7-9.9)
Prognostic risk models		
MRS	0.1	(–0.2 to 0.5)
AOM score	1.6	(1.3-2.1)

Continuous variables are expressed as median (IQR) and nominal variables as absolute number (%).
Abbreviations: ALT, alanine aminotransferase; GGT, gamma-glutamyltransferase; INR, international normalized ratio; UDCA, ursodeoxycholic acid.

controls (Supporting Table S8). Gallbladder volume ($P < 0.0001$) and total biliary tree volume ($P = 0.0020$) were higher in patients with PSC. The sum of relative severity of candidate dilatations (SumRelSevDilat) within the biliary tree had the highest performance to differentiate PSC from no PSC (AUC, 0.87; 95% CI, 0.77-0.96; $P < 0.0001$; sensitivity 86%; specificity 85%).

COMPARISON OF INTEROBSERVER AGREEMENT BETWEEN MRCP+ METRICS AND QUALITATIVE MRI-MRCP SCORES

Dilatation metrics had good to excellent interobserver agreements (ICC, 0.90-0.96), while stricture metrics had
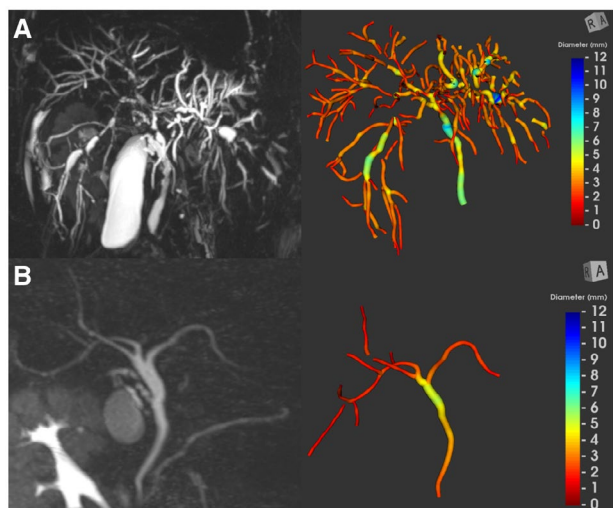
FIG. 3. Maximum intensity projection image and the corresponding color-coded, 3D, parametric biliary tree model. (A) 33-year-old patient with extrahepatic and intrahepatic PSC (candidate strictures, 20; candidate dilatations, 42). (B) 40-year-old healthy volunteer (candidate strictures, 0; candidate dilatations, 0).

poor to good interobserver agreements (ICC, 0.64-0.84) (Supporting Table S9). Intraobserver agreements were excellent for dilatation metrics (ICC, 0.93-0.97) and good for stricture metrics (ICC, 0.81-0.88). Gallbladder and biliary tree volume metrics had the highest interobserver and intraobserver agreements (ICC, 0.95-1.00).

The interobserver agreement for total stricture severity score was substantial ($\kappa = 0.66$), with agreement for extrahepatic stricture severity score (ESSS) better than intrahepatic stricture severity score (ISSS) (Supporting Table S10). Interobserver agreement for the Anali score was only fair ($\kappa = 0.38$), with severity classification for maximum intrahepatic bile duct diameter having the lowest agreement among the three variables included in the Anali score. Interobserver agreement for the presence of a dominant stricture was excellent ($\kappa = 0.81$).

The ICC for the measurement of maximum duct diameter was significantly better with MRCP+ than conventional MRI-MRCP measurement for all biliary segments (Supporting Table S11).

PERFORMANCE OF MRCP+ METRICS TO DETECT HIGH-RISK PSC

Correlations between all MRCP+ metrics and study variables are presented in Supporting Tables S12 and S13. Dilatation metrics had significant correlations with noninvasive markers and were higher in the high-risk PSC group compared to the low-risk group (Supporting Tables S14-S17). Correlations with stricture metrics were poor, and there were no significant differences between high-risk and low-risk PSC groups except when stratified by disease distribution and presence of dominant stricture. Intrahepatic dilatation metrics had better risk stratification performance compared to the whole-tree dilatation metrics. The SumRelSevDilat per meter of the biliary tree in the intrahepatic segment demonstrated the strongest correlation with noninvasive markers and had better interobserver agreement compared to qualitative MRI-MRCP scores (Table 3). The discriminatory performance of this candidate metric to classify patients into high-risk and low-risk groups is specifically reported here.

Prognostic Risk Models

Intrahepatic SumRelSevDilat demonstrated a positive correlation with the two risk models (MRS $r_s = 0.51$, $P < 0.0001$; AOM $r_s = 0.41$, $P < 0.0001$). Patients at high risk (MRS > 0 or AOM > 2) had higher intrahepatic SumRelSevDilat (MRS 8.2 m^{-1} vs. 5.6 m^{-1}, $P < 0.0001$; AOM 9.1 m^{-1} vs. 6.3 m^{-1}, $P = 0.0017$) compared to those at low risk (Table 4). Intrahepatic SumRelSevDilat (AUC, 0.77; 95% CI, 0.66-0.88) and ELF (AUC, 0.77; 95% CI, 0.67-0.88) showed the highest performance to diagnose MRS > 0 (Fig. 4; Supporting Table S18). The performance of LS, qualitative MRI-MRCP scores, and ULN ALP were numerically lower (LS AUC, 0.73; 95% CI, 0.61-0.84; Anali score AUC, 0.70, 95% CI, 0.58-0.81; ISSS AUC, 0.69; 95% CI, 0.56-0.81; ALP × ULN AUC, 0.67; 95% CI, 0.54-0.79).

When stratified by AOM, LS (AUC, 0.83; 95% CI, 0.74-0.95) numerically had the best performance to detect AOM > 2 (Fig. 4, Supporting Table S18) and was followed by ELF (AUC, 0.78; 95% CI, 0.65-0.92). The discriminatory performance of the Anali score (AUC, 0.77; 95% CI, 0.67-0.91) was numerically better than intrahepatic SumRelSevDilat (AUC, 0.74; 95% CI, 0.60-0.85) and ISSS (AUC, 0.64; 95% CI, 0.50-0.78).

Serum ALP

Intrahepatic SumRelSevDilat was positively correlated with ALP ($r_s = 0.40$; $P = 0.0003$) and was

TABLE 3. ASSOCIATION BETWEEN INTRAHEPATIC SUMRELSEVDILAT AND QUALITATIVE MRI-MRCP SCORES WITH NONINVASIVE SURROGATE MARKERS OF DISEASE SEVERITY AND PROGRESSION IN PSC

		Interobserver Variability	MRS Correlation Coefficient	MRS P Value	AOM Correlation Coefficient	AOM P Value	ALP Correlation Coefficient	ALP P Value	LS Correlation Coefficient	LS P Value	ELF Correlation Coefficient	ELF P Value
MRCP+												
SumRelSevDilat (m^{-1})	Intrahepatic	ICC = 0.90	0.51	<0.0001	0.41	0.0002	0.40	0.0003	0.46	<0.0001	0.30	0.0085
	Whole tree	ICC = 0.90	0.45	<0.0001	0.38	0.0006	0.38	0.0007	0.36	0.0015	0.24	0.0300
MRI-MRCP												
Modified Amsterdam stricture severity score	ISSS	κ = 0.74	0.39	0.0005	0.39	0.0005	0.37	0.0011	0.34	0.0029	0.16	0.1787
	TSSS	κ = 0.66	0.33	0.0033	0.31	0.0061	0.34	0.0030	0.41	0.0003	0.13	0.2673
Anali classification	Intrahepatic dilatation	κ = 0.50	0.43	<0.0001	0.31	0.0061	0.42	0.0002	0.37	0.0009	0.29	0.0110
	Anali score	κ = 0.38	0.39	0.0006	0.48	<0.0001	0.29	0.0118	0.53	<0.0001	0.28	0.0141

Abbreviation: TSSS, total stricture severity score.

TABLE 4. COMPARISON OF INTRAHEPATIC SUMRELSEVDILAT IN 76 PATIENTS WITH LARGE-DUCT PSC CLASSIFIED INTO HIGH-RISK AND LOW-RISK GROUPS

	SumRelSevDilat (m^{-1})		P Value
Prognostic Risk Models			
MRS > 0 (n = 43)	8.2	(6.8-9.8)	<0.0001
MRS ≤ 0 (n = 33)	5.6	(3.4-6.9)	
AOM > 2 (n = 23)	9.1	(7.4-10.1)	0.0017
AOM ≤ 2 (n = 53)	6.3	(4.8-8.4)	
Serum ALP			
ALP > 2.2 × ULN (n = 13)	8.9	(8.1-11.6)	0.0007
ALP ≤ 2.2 × ULN (n = 63)	6.7	(4.8-9.0)	
ALP > 1.5 × ULN (n = 25)	8.2	(6.6-9.5)	0.0225
ALP ≤ 1.5 × ULN (n = 51)	6.7	(4.5-9.1)	
Liver fibrosis markers			
LS > 9.6 kPa (n = 21)	9.1	(7.0-11.5)	0.0022
LS ≤ 9.6 kPa (n = 55)	6.3	(4.8-8.3)	
ELF > 9.8 (n = 23)	8.1	(6.9-11.4)	0.0189
ELF ≤ 9.8 (n = 53)	6.6	(5.0-8.6)	
MRCP features			
Extrahepatic disease (n = 41)	8.1	(5.7-9.6)	0.0331
No extrahepatic disease (n = 35)	6.6	(4.4-8.2)	
Dominant stricture (n = 25)	8.3	(6.8-10.0)	0.0019
No dominant stricture (n = 51)	6.3	(4.7-8.6)	

All values are presented as median (IQR).

higher in the high-risk group stratified by the 2.2 × ULN ALP threshold (8.9 m^{-1} vs. 6.7 m^{-1}, $P = 0.0007$; Table 4). Intrahepatic SumRelSevDilat (AUC, 0.79; 95% CI, 0.68-0.90) had the best diagnostic performance to detect the >2.2 × ULN ALP group and numerically performed better than others (LS AUC, 0.77; 95% CI, 0.65-0.89; ELF AUC, 0.72; 95% CI, 0.57-0.87; Anali score AUC, 0.71; 95% CI, 0.57-0.84; ISSS AUC, 0.72; 95% CI, 0.60-0.85; Supporting Table S18). The performance of all noninvasive markers was weak at the 1.5 × ULN threshold.

Liver Fibrosis Markers

Intrahepatic SumRelSevDilat correlated with LS ($r_s = 0.46$, $P < 0.001$) and ELF ($r_s = 0.30$, $P = 0.0085$). Patients had higher intrahepatic SumRelSevDilat when stratified by LS > 9.6 kPa (9.1 m^{-1} vs. 6.3 m^{-1}, $P = 0.0022$) and ELF > 9.8 (8.1 m^{-1} vs. 6.6 m^{-1}, $P = 0.0189$) (Table 4). The Anali score diagnosed high-risk fibrosis (LS AUC, 0.76; 95% CI, 0.66-0.88;
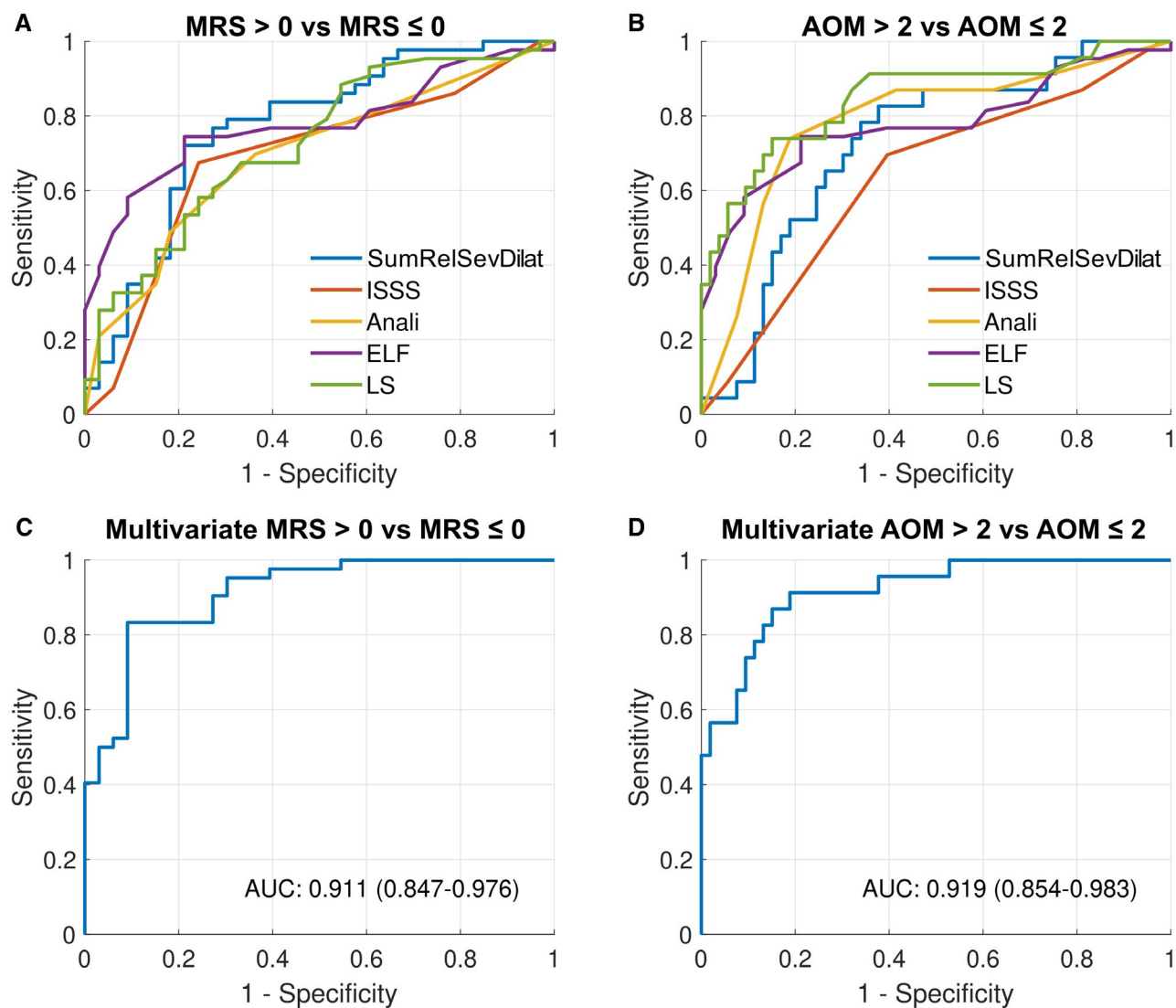
FIG. 4. Diagnostic performance of noninvasive markers to detect the high-risk group. (A) High-risk group stratified by MRS (intrahepatic SumRelSevDilat [AUC, 0.77; 95% CI, 0.66-0.88], ISSS [AUC, 0.69; 95% CI, 0.56-0.81], Anali [AUC, 0.70; 95% CI, 0.58-0.81], ELF [AUC, 0.77; 95% CI, 0.67-0.88], and LS [AUC, 0.73; 95% CI, 0.61-0.84]). (B) High-risk group stratified by AOM (intrahepatic SumRelSevDilat [AUC, 0.74; 95% CI, 0.60-0.85], ISSS [AUC, 0.64; 95% CI, 0.50-0.78], Anali [AUC, 0.77; 95% CI, 0.67-0.91], ELF [AUC, 0.78; 95% CI, 0.65-0.92], and LS [AUC, 0.83; 95% CI, 0.74-0.95]). (C,D) Multiple logistic regression model performance to detect high-risk PSC stratified by (C) MRS with independent predictors intrahepatic SumRelSevDilat (OR, 31.3; $P = 0.035$) and ELF (OR, 3.5; $P = 0.042$) and (D) AOM with LS (OR, 1.3; $P = 0.033$) as the only independent predictor.

ELF AUC, 0.71; 95% CI, 0.58-0.83) and was numerically better than intrahepatic SumRelSevDilat (LS AUC, 0.73; 95% CI, 0.60-0.85; ELF AUC, 0.68; 95% CI, 0.55-0.81; Supporting Table S18). The performance of ISSS was numerically lower with no significant difference between the high-risk and low-risk groups.

Disease Distribution and Dominant Stricture

In patients who have both extrahepatic and intrahepatic disease, the intrahepatic SumRelSevDilat was higher compared to patients with only intrahepatic disease distribution (8.1 m^{-1} vs. 6.6 m^{-1}, $P = 0.0331$;

Table 4). When a dominant stricture was present, the intrahepatic SumRelSevDilat was higher (8.3 m^{-1} vs. 6.3 m^{-1}, P = 0.0019; Table 4). Extrahepatic stricture metrics were significantly higher in patients with extrahepatic involvement compared to those with only intrahepatic PSC. Intrahepatic stricture metrics were significantly higher when a dominant stricture was present.

INDEPENDENT PREDICTORS OF DISEASE SEVERITY

In the multivariate logistic regression model, intrahepatic SumRelSevDilat (odds ratio [OR], 29.2; P = 0.038) and ELF (OR, 3.5; P = 0.041) was associated with an increased likelihood of having high-risk PSC when stratified by MRS. The model correctly classified 83% of cases (AUC, 0.92; 95% CI, 0.85-0.98) (Fig. 4). LS (OR, 1.3; P = 0.033) was the only covariate associated with increased likelihood of having high-risk PSC when stratified by AOM (Supporting Tables S19-S21).

Discussion

This is the first study to evaluate the utility and performance of a range of quantitative metrics of the biliary tree derived from a 3D-biliary analysis tool in a well-characterized adult cohort with large-duct PSC. Our study assessed these metrics against reference normal values, compared their interobserver variability against qualitative MRI-MRCP interpretation, evaluated their correlation with noninvasive markers, and evaluated their performance to classify patients into high-risk and low-risk groups by using predictors of disease severity as the reference. Intrahepatic tree dilatation metrics were found to have promising risk-stratification ability and better interobserver agreement compared to qualitative MRI-MRCP interpretation. In particular, the intrahepatic SumRelSevDilat demonstrated positive correlations with important biomarkers in PSC and was consistently higher in the high-risk group. This metric was an independent predictor of high-risk PSC defined by MRS and will require external validation and longitudinal follow-up for correlation with clinical outcomes. This exploratory study has provided the groundwork for examining the utility of novel quantitative metrics in larger cohorts of patients in multicenter studies, for example, through the International PSC Study Group.

While there has been significant progress made in quantitative imaging techniques of the liver parenchyma in PSC, there has been surprisingly little in the way of quantitative techniques to assess the primary insult, the biliary component of the disease.[28] To our knowledge, MRCP+ is the only tool at present that allows quantitative interrogation of the biliary tree. The first step in biomarker identification is to test if these quantitative metrics can differentiate a normal from an abnormal biliary tree. We compared biliary metrics in patients with large-duct PSC and healthy volunteers and demonstrated changes in gallbladder volume and biliary dilatations. We observed an increased gallbladder volume in patients with PSC compared with matched healthy volunteers, a phenomenon that has been described in PSC[29] and may assist in the development of a radiologic diagnostic score for the detection of PSC.[6] We also observed higher dilatation metrics in patients with PSC than in matched healthy volunteers, with the severity of biliary dilatations showing the highest discriminatory performance between abnormal and normal biliary trees.

Given the pivotal role of MRI-MRCP in the follow-up of patients with PSC, there has been much recent interest in exploring the potential applications of MRI techniques to risk stratify patients for the prediction of clinical outcomes. However, biliary features on their own consistently fall short of parenchymal features in existing studies using conventional MRI-MRCP. One of the drawbacks of conventional MRCP is the inherent technical limitation resulting in poor depiction of anatomy and strictures in an underdistended biliary tree, making it subject to both false-positive and false-negative calls. Data interpretation using qualitative features of disease further compounds the matter due to its susceptibility to high interobserver variability.[7,30] In a study among experts in PSC, the interobserver agreement for biliary changes on a single-time-point follow-up MRI-MRCP was poor, with κ values <0.2.[7] Similarly, interobserver agreement for bile duct changes between two serial follow-up MRI-MRCP examinations was also poor, with ICC values ranging between –0.45 and 0.22.[30] It is therefore unsurprising that there is significant variability on the reported prognostic ability of biliary features in predicting clinical outcomes in PSC.

A study that used a modified Amsterdam cholangiographic classification for MRCP interpretation found weak performance for prediction of clinical endpoints.[10] Another study found that biliary stricture severity assessed using a slightly different standardized interpretation model had poor risk stratification ability on its own compared to when combined with LS measurements using magnetic resonance elastography (MRE).[17] The Anali score has shown promising prognostic ability in a multicenter study, albeit in a retrospective cohort where images were reviewed centrally by two radiologists without assessment of interobserver agreement.[12] The score is weighted more toward parenchymal changes, which tend to appear in more advanced disease.

The interobserver agreement for the Anali score is yet to be reported. We compared interobserver variability between Anali score, modified Amsterdam stricture severity score, and MRCP+ metrics. While the agreement for binary classifications, such as hepatic dysmorphy and portal hypertension, was good to excellent, interobserver agreement for the Anali score was only fair. This is likely due to greater discrepancies in classifying the intrahepatic bile duct dilatation component of the Anali score into categorical variables with millimeter margins of error. We also observed that the ICC for maximum duct diameter measurements from MRCP+ decreases gradually moving proximally from the common bile duct, suggesting a greater room for error in manual measurements of intrahepatic duct diameter compared to extrahepatic ducts. The agreement for ISSS and ESSS using the modified Amsterdam classification was significantly better than the Anali score, probably as specialist radiologists are already experienced with the well-established ductal features described in the model. This is reflected in the better consistency between the two readers in our study for the diagnostic performance of ISSS to detect the high-risk PSC group compared to the Anali score. Future radiologic scores could consider including MRCP+-derived metrics to avoid variations in interpretations of continuous variables, such as bile duct diameter, and provide a more standardized and objective assessment of strictures and dilatations.

In our study, the severity of intrahepatic dilatations showed the best correlation with surrogate markers of disease severity. The metric SumRelSevDilat was consistently higher in patients classified as high risk. This supports previous work that has shown that the severity of biliary dilatation, particularly that of intrahepatic bile ducts, was an independent predictive feature for disease progression.[11] A dilatation severity score for each dilatation is computed by MRCP+ as the product of the length of the dilatation and its relative severity. However, the sum of all the dilatation scores within the biliary tree (TreeDilatScore) was numerically inferior to SumRelSevDilat to diagnose the high-risk group. This would indicate that the length of the dilatation is of lesser importance than the severity, which is supported by another study.[11] The performance of intrahepatic SumRelSevDilat was mixed compared to other noninvasive risk-stratification tools. It had the best performance to diagnose high-risk PSC stratified by MRS. Intrahepatic SumRelSevDilat and ELF were the only two noninvasive markers that were associated with increased odds of having high-risk PSC when stratified by MRS in the multivariate analysis. When stratified by AOM, however, LS numerically performed better than intrahepatic SumRelSevDilat to identify the AOM > 2 group and was the only independent predictor of high-risk PSC as defined by AOM in the multivariate analysis.

Biliary strictures are thought to be the primary pathology that drives fibrosis of the liver parenchyma in PSC. Therefore, severity of candidate strictures would be expected to correlate with disease severity, but we found no positive correlation or significant risk-stratification ability of any of the stricture metrics in our study. This was similarly observed in a study that found no association between the severity of the biliary component of a score from standardized qualitative interpretation of MRI-MRCP and advanced stages of liver fibrosis using the Nakanuma histologic classification system.[31] Another study showed a weak positive correlation between MRE LS and intrahepatic stricture severity, no correlation with extrahepatic stricture severity, and poor ICC of LS values on MRE among the different grades of intrahepatic stricture severity.[17] These findings may be partially explained by the poor reliability in assessing strictures as was seen in the observer-dependent analysis in our study where agreement for stricture metrics was not as good as dilatations. MRCP images are often noisy, with bile ducts ranging from several millimeters to below the image resolution of the MR scanner. Characterization of strictures is more likely to be affected by these limitations than dilatations. However, MRCP+-derived

stricture metrics appear to be more sensitive in quantitatively monitoring the extrahepatic biliary tree changes and dominant strictures.

Our study has several limitations. Clinical outcomes are slow to develop in PSC, therefore risk stratification against other biomarkers that have been validated against clinical endpoints were used in this study. However, these surrogate biomarkers themselves are imperfect, and none have achieved level three validation.[5] Second, the directionality of the occurrence of stricture(s) and dilatation(s) is not currently defined in the MRCP+ algorithm; therefore, it is assumed that dilatations are upstream to strictures. Third, our study was a small, single, tertiary-center study to identify candidate metrics. Given the heterogeneity of the disease, multicenter studies with larger cohorts of patients with PSC are needed to validate the metrics presented here. Although we have assessed the interobserver and intraobserver variability in PSC, we have not conducted a scan–rescan reproducibility study specifically in PSC; the data supporting reproducibility come from 6 patients with PSC,[13] and further validation would be needed. Fourth, we used FibroScan, which samples a lower volume of the liver compared to MRE, as a reference to stage liver fibrosis. This could be relevant in assessing a disease with heterogeneous distribution of fibrosis, such as in PSC. Fifth, evaluating cost effectiveness of off-site data processing and capital expenditure of MRCP+ versus the conventional qualitative approach to image analysis was beyond the scope of this study. The benefit of having an expert radiologist to make a judgement call on a poor quality MRI-MRCP acquisition in the right clinical context cannot be underestimated, as shown by the fact that all 80 scans could be read by the radiologists while only 76 scans could be processed by MRCP+.

In conclusion, MRCP+ allows unique quantitative interrogation of the biliary tree, and MRCP+ metrics related to biliary dilatation severity have promising risk-stratification ability and better interobserver agreement compared to qualitative MRI-MRCP interpretation in adult patients with PSC. Longitudinal studies with repeat measurements in the same patient will be required and validated first before it can be considered for use in clinical trials.

REFERENCES

1) Hirschfield GM, Karlsen TH, Lindor KD, Adams DH. Primary sclerosing cholangitis. Lancet 2013;382:1587-1599.

2) Lazaridis KN, LaRusso NF. Primary sclerosing cholangitis. N Engl J Med 2016;375:1161-1170.

3) Trivedi PJ, Corpechot C, Pares A, Hirschfield GM. Risk stratification in autoimmune cholestatic liver diseases: opportunities for clinicians and trialists. Hepatology 2016;63:644-659.

4) Goet JC, Floreani A, Verhelst X, Cazzagon N, Perini L, Lammers WJ, et al. Validation, clinical utility and limitations of the Amsterdam-Oxford model for primary sclerosing cholangitis. J Hepatol 2019;71:992-999.

5) Trivedi PJ. Risk stratification in primary sclerosing cholangitis: It's time to move on from replicating imperfection and break the glass ceiling. J Hepatol 2019;71:867-870.

6) Schramm C, Eaton J, Ringe KI, Venkatesh S, Yamamura J; MRI working group of the IPSCSG. Recommendations on the use of magnetic resonance imaging in PSC-A position statement from the International PSC Study Group. Hepatology 2017;66:1675-1688.

7) Zenouzi R, Liwinski T, Yamamura J, Weiler-Normann C, Sebode M, Keller S, et al.; International PSC Study Group (IPSCSG). Follow-up magnetic resonance imaging/3D-magnetic resonance cholangiopancreatography in patients with primary sclerosing cholangitis: challenging for experts to interpret. Aliment Pharmacol Ther 2018;48:169-178.

8) Ponsioen CY, Vrouenraets SM, Prawirodirdjo W, Rajaram R, Rauws EAJ, Mulder CJJ, et al. Natural history of primary sclerosing cholangitis and prognostic value of cholangiography in a Dutch population. Gut 2002;51:562-566.

9) Ponsioen C, Reitsma J, Boberg K, Aabakken L, Rauws E, Schrumpf E. Validation of a cholangiographic prognostic model in primary sclerosing cholangitis. Endoscopy 2010;42:742-747.

10) Tenca A, Mustonen H, Lind K, Lantto E, Kolho K-L, Boyd S, 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;38:2329-2339.

11) Ruiz A, Lemoinne S, Carrat F, Corpechot C, Chazouillères O, Arrivé L. Radiologic course of primary sclerosing cholangitis: assessment by three-dimensional magnetic resonance cholangiography and predictive features of progression. Hepatology 2014;59:242-250.

12) Lemoinne S, Cazzagon N, El Mouhadi S, Trivedi PJ, Dohan A, Kemgang A, et al. Simple magnetic resonance scores associate with outcomes of patients with primary sclerosing cholangitis. Clin Gastroenterol Hepatol 2019;17:2785-2792.e3.

13) Goldfinger MH, Ridgway GR, Ferreira C, Langford CR, Cheng L, Kazimianec A, et al. Quantitative MRCP imaging: accuracy, repeatability, reproducibility, and cohort-derived normative ranges. J Magn Reson Imaging 2020;52:807-820.

14) Gilligan LA, Trout AT, Lam S, Singh R, Tkach JA, Serai SD, et al. Differentiating pediatric autoimmune liver diseases by quantitative magnetic resonance cholangiopancreatography. Abdom Radiol (NY) 2020;45:168-176.

15) Arrive L, Coudray C, Azizi L, Lewin M, Hoeffel C, Monnier-Cholley L, et al. Pineapple juice as a negative oral contrast agent in magnetic resonance cholangiopancreatography. [in French] J Radiol 2007;88:1689-1694.

16) Kim WR, Therneau TM, Wiesner RH, Poterucha JJ, Benson JT, Malinchoc M, et al. A revised natural history model for primary sclerosing cholangitis. Mayo Clin Proc 2000;75:688-694.

17) Tafur M, Cheung A, Menezes RJ, Feld J, Janssen H, Hirschfield GM, et al. Risk stratification in primary sclerosing cholangitis: comparison of biliary stricture severity on MRCP versus liver stiffness by MR elastography and vibration-controlled transient elastography. Eur Radiol 2020;30:3735-3747.

18) Ni Mhuircheartaigh JM, Lee KS, Curry MP, Pedrosa I, Mortele KJ. Early peribiliary hyperenhancement on MRI in patients with primary sclerosing cholangitis: significance and association with the Mayo risk score. Abdom Radiol (NY) 2017;42: 152-158.

19) de Vries EM, Wang J, Williamson KD, Leeflang MM, Boonstra K, Weersma RK, et al. A novel prognostic model for transplant-free survival in primary sclerosing cholangitis. Gut 2018;67:1864-1869.

20) Al Mamari S, Djordjevic J, Halliday JS, Chapman RW. Improvement of serum alkaline phosphatase to <1.5 upper limit of normal predicts better outcome and reduced risk of cholangio-carcinoma in primary sclerosing cholangitis. J Hepatol 2013;58: 329-334.

21) Goode EC, Clark AB, Mells GF, Srivastava B, Spiess K, Gelson WTH, et al.; UK-PSC Consortium. Factors associated with outcomes of patients with primary sclerosing cholangitis and development and validation of a risk scoring system. Hepatology 2019;69:2120-2135.

22) Corpechot C, Gaouar F, El Naggar A, Kemgang A, Wendum D, Poupon R, et al. Baseline values and changes in liver stiffness measured by transient elastography are associated with severity of fibrosis and outcomes of patients with primary sclerosing cholangitis. Gastroenterology 2014;146:970-979.

23) de Vries EMG, Färkkilä M, Milkiewicz P, Hov JR, Eksteen B, Thorburn D, et al. Enhanced liver fibrosis test predicts transplant-free survival in primary sclerosing cholangitis, a multicentre study. Liver Int 2017;37:1554-1561.

24) Vesterhus M, Hov JR, Holm A, Schrumpf E, Nygård S, Godang K, et al. Enhanced liver fibrosis score predicts transplant-free survival in primary sclerosing cholangitis. Hepatology 2015;62:188-197.

25) Chapman MH, Webster GJM, Bannoo S, Johnson GJ, Wittmann J, Pereira SP. Cholangiocarcinoma and dominant strictures in patients with primary sclerosing cholangitis: a 25-year single-centre experience. Eur J Gastroenterol Hepatol 2012;24:1051-1058.

26) Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 2016;15:155-163. Erratum in: J Chiropr Med 2017;16:346.

27) Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159-174.

28) Selvaraj EA, Culver EL, Bungay H, Bailey A, Chapman RW, Pavlides M. Evolving role of magnetic resonance techniques in primary sclerosing cholangitis. World J Gastroenterol 2019;25: 644-658.

29) van de Meeberg PC, Portincasa P, Wolfhagen FH, van Erpecum KJ, VanBerge-Henegouwen GP. Increased gall bladder volume in primary sclerosing cholangitis. Gut 1996;39:594-599.

30) Grigoriadis A, Morsbach F, Voulgarakis N, Said K, Bergquist A, Kartalis N. Inter-reader agreement of interpretation of radiological course of bile duct changes between serial follow-up magnetic resonance imaging/3D magnetic resonance cholangiopancreatography of patients with primary sclerosing cholangitis. Scand J Gastroenterol 2020;55:228-235.

31) Song C, Lewis S, Kamath A, Hectors S, Putra J, Kihira S, et al. Primary sclerosing cholangitis: diagnostic performance of MRI compared to blood tests and clinical scoring systems for the evaluation of histopathological severity of disease. Abdom Radiol (NY) 2020;45:354-364.

Supporting Information

Additional Supporting Information may be found at onlinelibrary.wiley.com/doi/10.1002/hep4.1860/suppinfo.