


Magnetic Resonance Cholangiopancreatography Severity Predicts Disease Outcomes in Pediatric Primary Sclerosing Cholangitis: A Reliability and Validity Study

Kedar Patil,^{1,2*} Amanda Ricciuto ,^{3,4*} Alaa Alsharief,¹ Jehan Al-Rayahi,¹ Afsaneh Amirabadi,¹ Peter C. Church,^{3,4} Binita M. Kamath,^{3,4} and Mary-Louise C. Greer^{1,2}

Magnetic resonance cholangiopancreatography (MRCP) has not been assessed as a surrogate biomarker in pediatrics. We aimed to determine the inter-rater reliability, prognostic utility, and construct validity of the modified Majoie endoscopic retrograde cholangiopancreatography classification applied to MRCP in a pediatric primary sclerosing cholangitis (PSC) cohort. This single-center, retrospective, cohort study included children with PSC undergoing diagnostic MRCP between 2008 and 2016. Six variations of the Majoie classification were examined: 1) intrahepatic duct (IHD) score, 2) extrahepatic duct (EHD) score (representing the worst intrahepatic and extrahepatic regions, respectively), 3) sum IHD-EHD score, 4) average IHD score, 5) average EHD score, and 6) sum average IHD-EHD score. Inter-rater reliability was assessed using weighted kappas and intraclass correlation coefficients (ICCs). Ability to predict time to PSC-related complications (ascites, esophageal varices, variceal bleed, liver transplant [LT], or cholangiocarcinoma) (primary outcome) and LT (secondary outcome) was assessed with Harrell's concordance statistic (c-statistic) and univariate/multivariable survival analysis. Construct validity was further assessed with Spearman correlations. Forty-five children were included (67% boys; median, 13.6 years). The inter-rater reliability of MRCP scores was substantial to excellent (kappas/ICCs, 0.78-0.82). The sum IHD-EHD score had the best predictive ability for time to PSC complication and LT (c-statistic, 0.80 and SE, 0.06; and c-statistic, 0.97 and SE, 0.01, respectively). Higher MRCP scores were independently associated with a higher rate of PSC-related complications, even after adjusting for the PSC Mayo risk score (hazard ratio, 1.74; 95% confidence interval, 1.14-2.). MRCP sum scores correlated significantly with METAVIR fibrosis stage, total bilirubin, and platelets ($r = 0.42$, $r = 0.33$, $r = -0.31$, respectively; $P < 0.05$). **Conclusion:** An MRCP score incorporating the worst affected intrahepatic and extrahepatic regions is reliable and predicts meaningful outcomes in pediatric PSC. Next steps include prospective validation and responsiveness assessment. (*Hepatology Communications* 2020;4:208-218).

Primarily sclerosing cholangitis (PSC) is a chronic cholestatic liver disease that is characterized by inflammation and fibrosis of the intrahepatic and/or extrahepatic biliary ducts and is commonly associated with inflammatory bowel disease (IBD), primarily ulcerative colitis (UC).⁽¹⁾ The natural history of PSC is progression to advanced fibrosis and end-stage liver disease requiring liver transplantation (LT).

Abbreviations: 3D, three-dimensional; ALP, alkaline phosphatase; ALT, alanine aminotransferase; ASC, autoimmune sclerosing cholangitis; AST, aspartate aminotransferase; bSSFP, balanced steady-state free precession; CBD, common bile duct; CD, Crohn's disease; CHD, common hepatic duct; CI, confidence interval; c-statistic, concordance statistic; EHD, extrahepatic duct; ERCP, endoscopic retrograde cholangiopancreatography; FA, flip angle; FOV, field of view; GGT, gamma-glutamyl transferase; HR, hazard ratio; IBD, inflammatory bowel disease; IBD-U, IBD-unclassified; ICC, intraclass correlation coefficient; IHD, intrahepatic duct; IQR, interquartile range; LT, liver transplantation; MRCP, magnetic resonance cholangiopancreatography; MRS, Mayo risk score; PH, portal hypertension; PSC, primary sclerosing cholangitis; ST, slice thickness; T, Tesla; TE, echo time; TR, repetition time; TSE, turbo spin echo; UC, ulcerative colitis.

Received August 28, 2019; accepted November 11, 2019.

Additional Supporting Information may be found at onlinelibrary.wiley.com/doi/10.1002/hep4.1454/supinfo.

In children, the 10-year LT-free survival is 70%,⁽²⁾ and in adults, the median time to LT is 13–15 years at tertiary centers.^(3,4) PSC is also associated with an increased risk of hepatobiliary and colorectal cancer, particularly in adults.⁽¹⁾ For these reasons, PSC represents an important source of morbidity and mortality in adult and pediatric populations.

To date, no medical therapy has been shown to alter the natural history of PSC. PSC is both rare and slowly progressive. As a result, therapeutic clinical trials with sufficiently long follow-up to capture “hard” outcomes, like LT or death, present feasibility challenges, especially in children. Surrogate markers are therefore critical. One of the greatest limitations in the study of PSC is precisely the lack of surrogate markers that correlate closely with clinically meaningful outcomes. Several biomarkers and clinical scores have been evaluated, but all thus far have important limitations. Serum alkaline phosphatase (ALP) is widely used in adults but is neither sensitive nor specific. The PSC Mayo risk score (MRS), a composite of laboratory parameters, age, and history of variceal bleeding, is also frequently used in adult studies but has been validated only for short-term outcomes. Furthermore, both ALP and the MRS have been observed to correlate poorly with outcomes in a randomized controlled trial setting.⁽⁵⁾ In children, ALP is unreliable given fluctuations with bone growth. Gamma-glutamyl transferase (GGT) has shown potential promise in the pediatric

setting, with normalization at 1 year being associated with favorable outcomes,⁽⁶⁾ but these findings await prospective validation. Liver histology has been found to predict survival⁽⁷⁾ but requires a liver biopsy, which is invasive. Given the above, the identification of adequate surrogate biomarkers in PSC is recognized to be a priority area of research.

Imaging is central to PSC diagnosis, but its potential as a surrogate and prognostic marker has been little explored. Endoscopic retrograde cholangiopancreatography (ERCP) was previously the diagnostic procedure of choice for PSC. In 1991, Majoie et al.⁽⁸⁾ proposed an ERCP classification system for grading the severity of biliary involvement in adults with PSC based on an earlier classification developed by Chen and Goldberg.⁽⁹⁾ Ponsioen et al.^(10,11) later validated this ERCP classification in an independent adult cohort. However, over the past decade, magnetic resonance cholangiopancreatography (MRCP) has largely supplanted ERCP as it is less invasive and free of ionizing radiation, which makes it particularly attractive for use in children. This change is clearly reflected in clinical practice guidelines.⁽¹²⁾ However, there is a paucity of data on standardized MRCP scoring systems to predict PSC outcomes. No classification has been systematically assessed in children. Furthermore, reliability data are lacking in children and adults.

Given the above, we aimed to establish the inter-rater reliability of the modified Majoie classification

**These authors contributed equally to this work.*

© 2019 The Authors. *Hepatology Communications* published by Wiley Periodicals, Inc., on behalf of the American Association for the Study of Liver Diseases. This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

View this article online at wileyonlinelibrary.com.

DOI 10.1002/hep4.1454

Potential conflict of interest: Dr. Kamath consults for and has received grants from Mirum. Dr. Greer has received grants from AbbVie. The other authors have nothing to report.

ARTICLE INFORMATION:

From the ¹Department of Diagnostic Imaging, The Hospital for Sick Children, Toronto, Canada; ²Department of Medical Imaging, University of Toronto, Toronto, Canada; ³Division of Gastroenterology, Hepatology, and Nutrition, The Hospital for Sick Children, Toronto, Canada; ⁴Department of Paediatrics, University of Toronto, Toronto, Canada.

ADDRESS CORRESPONDENCE AND REPRINT REQUESTS TO:

Amanda Ricciuto, M.D., Ph.D.
Division of Gastroenterology, Hepatology and Nutrition
The Hospital for Sick Children
555 University Avenue

Toronto, ON M5G 1X8, Canada
E-mail: amanda.ricciuto@sickkids.ca
Tel.: +1-416-813-1500 Ext. 222886

applied to MRCP in pediatric PSC and to generate initial validation data by examining its construct validity and ability to predict clinical outcomes. We hypothesized that a sum score, incorporating intrahepatic and extrahepatic involvement, would perform best.

Patients and Methods

DESIGN AND SETTING

This was a single-center, retrospective, cohort study performed at the Hospital for Sick Children (SickKids) in Toronto, Canada. The institutional review board approved the study. Patient consent was waived given the study's retrospective nature.

PATIENT POPULATION

Children (<18 years) undergoing MRCP between January 2008 and September 2016 were identified by manually searching a diagnostic imaging database (ISYS Search Software Pty. Ltd., Lexmark International) for MRCP studies containing the keywords "primary sclerosing cholangitis" (or variations, such as "PSC"). This was cross-referenced with a retrospective database of children diagnosed with PSC between 2000 and 2018 (compiled based on pathology and hepatology clinic lists) to ensure no patients were missed. Only patients with liver biopsies supporting a diagnosis of PSC were included.⁽¹⁾ As is standard in pediatrics at SickKids, children with suspected PSC systematically undergo liver biopsy. This is due to the high rate (around 30%) of autoimmune sclerosing cholangitis (ASC),⁽²⁾ a variant of PSC with overlapping features of autoimmune hepatitis (AIH). As has been standard in pediatric studies, these patients were eligible for inclusion. For each patient, the diagnostic MRCP (closest to diagnostic liver biopsy) was reviewed. A minimum follow-up duration of 3 months was required. MRCPs of insufficient quality to allow interpretation were excluded.

DATA COLLECTION AND DEFINITIONS

We evaluated a modified version of the Majoie classification as operationalized by Ferrara et al.⁽¹³⁾ in their study of MRCP for diagnosing PSC in children (Table 1) because this was the only study to date to apply a standard radiographic scoring system to

TABLE 1. MODIFIED MAJOIE CLASSIFICATION APPLIED TO MRCPs (FROM FERRARA ET AL.,⁽¹³⁾ ADAPTED FROM MAJOIE ET AL.⁽⁸⁾)

Score	Definition
IHDs	
0	No abnormalities
1	Minimum stenosis with biliary ducts of regular diameter or minimally dilated
2	Multiple stenosis and saccular dilations with reduction of intraparenchymal arborization (aspect as "bare tree")
3	Closed stenosis to carrefour with obstruction or lack of visualization of one of the main hepatic ducts
EHDs	
0	No abnormalities
1	Wall irregularity in absence of significant stenosis
2	Segmental stenosis
3	Entire stenosis of the CBD
4	Irregularities in diameter, nodularity, and pseudodiverticular formations

MRCPs in pediatric PSC. The Majoie classification typically assigns a single overall score to the intrahepatic ducts (IHDs) and a single overall score to the extrahepatic ducts (EHDs). However, to promote greater consistency, we individually scored each of the liver segments according to Couinaud anatomy (left-sided segments: 1, 2, 3, 4a, 4b; right-sided segments: 5, 6, 7, 8) and the right and left IHD (all of which comprise the intrahepatic biliary tree) as well as the common hepatic duct (CHD) and common bile duct (CBD) (extrahepatic biliary tree). For segments with variable disease, the worst affected region was scored. We investigated six variations of the modified Majoie classification, as defined in Table 2. These included: (1) IHD score, which corresponded to the worst individual intrahepatic score, (2) EHD score, which corresponded to the worst individual extrahepatic score, (3) sum IHD and EHD (sum IHD-EHD) score, (4) average IHD score, (5) average EHD score, and (6) sum average IHD-EHD score. Hepatic parenchyma abnormalities and pancreatic duct involvement were assessed as well. Each MRCP was independently visually evaluated by two senior pediatric radiology fellows (K.P., A.A.), blinded to MRCP indication and clinical information. Discrepancies between the two raters were resolved by a third blinded radiologist (M.G.) with 18 years of experience reading pediatric MRCPs. The final consensus score was used in analyses.

Pertinent clinical and laboratory data were extracted from electronic medical records using a standard case report form; extracted data included

TABLE 2. VARIATIONS OF THE MODIFIED MAJOIE CLASSIFICATION EXAMINED

Cholangiographic Variation	Definition	Data Type (Range)
IHD score	Modified Majoie classification (as per Table 1) applied individually to segment 1, 2, 3, 4a, 4b, 5, 6, 7, and 8 of the liver, right intrahepatic duct, and left intrahepatic duct; the worst score taken	Ordinal (0-3)
EHD score	Modified Majoie classification (as per Table 1) applied individually to the CHD and the CBD; the worst score taken	Ordinal (0-4)
Sum IHD-EHD score	Sum of IHD and EHD scores	Ordinal (0-7)
Average IHD score	Modified Majoie classification (as per Table 1) applied individually to segment 1, 2, 3, 4a, 4b, 5, 6, 7, and 8 of the liver, right IHD, and left IHD; the average taken	Noninteger (0-3)
Average EHD score	Modified Majoie classification (as per Table 1) applied individually to the CHD and the CBD; the average taken	Noninteger (0-4)
Sum average IHD-EHD score	Sum of average IHD and average EHD scores	Noninteger (0-7)

patient demographics, PSC type (PSC vs. ASC), IBD type, hepatic biochemistry closest to MRCP date, and diagnostic liver biopsy findings. A label of ASC was applied to patients who satisfied criteria for PSC and displayed histologic features of AIH on liver biopsy (e.g., lymphoplasmacytic infiltrate, interface hepatitis). Small-duct PSC was defined as a normal cholangiogram, as reported in the medical record. IBD and its subtypes (UC, Crohn's disease [CD], IBD-unclassified [IBD-U]) were defined as per standard endoscopic and histopathologic criteria.⁽¹⁴⁾ Portal hypertension (PH) was defined as splenomegaly, thrombocytopenia, or a PH-related complication (ascites, esophageal varices, variceal bleed). The degree of fibrosis on diagnostic liver biopsies was categorized according to the METAVIR system (F0 [none] to F4 [cirrhosis]), as per reviewed pathology reports. Of note, it is convention at our center for pathologists to routinely report the METAVIR fibrosis stage. The primary outcome against which the predictive ability of MRCP was evaluated was PSC-related complication. PSC-related complications included ascites, endoscopically confirmed esophageal varices or variceal bleed, LT, and cholangiocarcinoma (CCA). LT served as a secondary outcome. The date of diagnostic MRCP was used as the baseline for survival analyses, except for analyses examining liver fibrosis in which case the diagnostic liver biopsy date was used as the baseline. Date of PSC diagnosis was defined as the date of diagnostic MRCP or liver biopsy (whichever occurred first). Patients were censored at last visit or transition to adult care, which occurs at age 18 years at our center.

MRCP IMAGING TECHNIQUE

All MRCP examinations were acquired on either a 1.5-Tesla (T) magnetic resonance unit (Siemens AVANTO;

Siemens Medical Solutions, Erlangen, Germany) or 3.0-T magnetic resonance unit (Philips ACHIEVA; Phillips Medical Systems, Best, the Netherlands) using body and surface coils dependent on patient size (generally torso phased-array coils). The institutional MRCP protocol included the following sequences, without intravenous contrast: axial balanced steady-state free precession (bSSFP), coronal three-dimensional (3D) heavy T2-weighted turbo spin echo (TSE) MRCP and radial coronal single-shot T2 sequences, thin axial and coronal single-shot T2-weighted sequences, axial T2-weighted TSE sequence with fat suppression, and axial 3D T1-weighted gradient echo imaging without fat suppression, with an approximate scan time of 20 minutes. Two key sequences were used for analysis: (1) coronal 3D T2 TSE for qualitatively grading bile duct abnormalities and (2) axial bSSFP to define their segmental anatomy. Approximate average sample parameters for the key sequences acquired at 1.5T and 3T are as follows:

1.5T

1. Coronal 3D T2 TSE: echo time (TE), 659 milliseconds; repetition time (TR), 2,922.9 milliseconds; 150-degree flip angle (FA); slice thickness (ST), 1.2 mm; gap, 1.1 mm; acquisition matrix, 256 × 256; field of view (FOV), 300 × 300 mm; and associated maximum intensity projections (MIPs).
2. Axial bSSFP : TE, 2.1 milliseconds; TR, 4.2 milliseconds; 68-degree FA; ST, 6 mm; gap, 1.8 mm; acquisition matrix, 320 × 194; FOV, 225 × 300 mm.

3.0T

1. Coronal 3D T2 TSE : TE, 800 milliseconds; TR, 3,235.6 milliseconds; 90-degree FA; ST, 2.2 mm;

- gap, 1.1 mm; acquisition matrix, 272 × 274; FOV, 300 × 300 mm; and associated MIPs.
2. Axial bSSFP : TE, 2.08 milliseconds; TR, 4.15 milliseconds; FOV, 29 cm; 45-degree FA; ST, 4 mm; gap, 0.5 mm; acquisition matrix, 208 × 208; FOV, 250 × 250 mm.

STATISTICAL METHODS

Continuous variables were summarized with medians (interquartile range [IQR]) and compared with the Mann-Whitney U test. Categorical variables were summarized with frequencies (proportions) and compared with the chi-square or Fisher exact test, as appropriate. Inter-rater reliability was assessed between the two raters using weighted kappas for ordinal scores⁽¹⁵⁾ and intraclass correlation coefficients (ICCs; two-way mixed, agreement, single measures) for noninteger scores (Table 2).⁽¹⁶⁾ The time-dependent discriminative ability of the various MRCP scores, fibrosis, and laboratory markers was assessed using Harrell's concordance statistic (c-statistic). Survival analysis, including Kaplan-Meier curves (log-rank test) and univariate and multivariable Cox proportional hazards regression, was used to further examine the predictive ability of MRCP. The multivariable Cox model was built using the change in estimate approach; covariates resulting in a change >10% in the MRCP score-point estimate were retained. Variables were selected *a priori* for inclusion based on clinical relevance. The construct validity of the score with the best predictive ability was further examined by comparing scores between patients with PSC with low and high MRCP scores and determining Spearman correlations between MRCP scores, hepatic biochemistry, and fibrosis. A sensitivity analysis restricted to patients with large-duct PSC and follow-up duration ≥6 months was performed as well.

Significance was defined as two-sided $P < 0.05$. Analyses were performed with SAS University Edition (version 3.4; SAS Institute, Cary, NC) and SPSS (version 23.0; IBM Corp., Armonk, NY).

Results

PATIENT CHARACTERISTICS AND CLINICAL OUTCOMES

Forty-five children with PSC were included. No examination was excluded for nondiagnostic quality.

Patient characteristics and clinical outcomes are shown in the first column of Table 3. The median age at PSC diagnosis was 13.6 (IQR, 10.3-15.2) years, 67% of patients were male, 89% had IBD (78% UC/IBD-U), 27% had ASC, and 9% had small-duct PSC. The median follow-up duration after PSC diagnosis was 3.4 (IQR, 2.4-4.4) years. Two patients had <6 months follow-up. The median interval between PSC diagnosis and the MRCP examined in this study was 35 (IQR, 0-380) days. Ten children (22%) developed a PSC-related complication at a median of 1.1 (IQR, 0.02-2.7) years, and 5 of these children (11%) progressed to LT at a median of 3.5 (IQR, 1.5-4.2) years. Complication-free and LT-free survival at 4 years were 76% (SE, 0.07) and 86% (SE, 0.07), respectively. There was one CCA, resulting in the one death in the cohort. One child with small-duct PSC developed a PSC-related complication, but none required LT.

MRCP SCORES

The breakdown of the various MRCP scores is shown in the first column of Table 3 and Supporting Fig. S1. For the IHD score, 9%, 44%, 38%, and 9% of patients had a score of 0, 1, 2 and 3, respectively, while 40%, 28%, 21%, and 12% had an EHD score of 0, 1, 2, and 3, respectively. No child had an EHD score of 4. MRCP images from 2 patients in the cohort are shown in Fig. 1. No patient displayed pancreatic duct involvement. On MRI, periportal edema was observed in 24% of patients and imaging features of cirrhosis in 18%. While numerically more patients with UC/IBD-U (vs. those with CD or no IBD) had IHD and EHD scores ≥2 (51% vs. 30% for IHD; 36% vs. 20% for EHD), these differences were not statistically significant ($P > 0.05$). Similarly, while numerically more patients with UC/IBD-U had sum IHD-EHD scores ≥4 (24% vs. 10% with CD or no IBD), the difference was not statistically significant ($P > 0.05$). In addition, MRCP severity did not differ by hepatic lobe (IHD scores ≥2 were in 42% of left lobes vs. 33% of right lobes; $P > 0.05$).

RELIABILITY

The weighted kappa statistics and ICCs for the various MRCP scores are shown in Table 4. All were similar, with weighted kappas ranging from 0.78-0.81

TABLE 3. PATIENT CHARACTERISTICS, MRCP SCORES, AND CLINICAL OUTCOMES

n (%) or Median (IQR)	All PSC	Sum IHD-EHD Score ≥ 4	Sum IHD-EHD Score < 4	P Value
	(n = 45)	(n = 10)	(n = 35)	
Male	30 (67%)	6 (60%)	24 (69%)	0.71
Age at diagnosis (years)				
PSC	13.6 (10.3-15.2)	12.4 (7.8-16.0)	13.6 (10.3-15.2)	0.73
IBD	13.5 (10.1-15.5)	11.2 (4.7-15.7)	13.5 (10.8-15.5)	0.38
PSC follow-up duration (years)	3.4 (2.4-4.4)	2.8 (2.4-5.3)	3.5 (2.1-4.3)	0.99
Ulcerative colitis/IBD-U	35 (78%)	9 (90%)	26 (74%)	0.39
Crohn's disease	5 (11%)	1 (10%)	4 (11%)	
No IBD	5 (11%)	0	5 (14%)	
ASC	12 (27%)	2 (20%)	10 (29%)	0.71
Small-duct PSC	4 (9%)	10 (100%)	4 (11%)	0.56
Biochemistry at time of MRCP				
ALT (U/L)	75 (46-166)	128 (50-210)	73 (44-159)	0.37
AST (U/L)	70 (40-201)	177 (56-371)	59 (37-142)	0.09
ALP (U/L)	361 (169-685)	642 (260-1003)	332 (149-565)	0.05
GGT (U/L)	189 (73-403)	331 (94-495)	188 (55-356)	0.20
Total bilirubin ($\mu\text{mol/L}$)	9 (6-19)	20 (7-45)	9 (6-15)	0.14
Albumin (g/L)	42 (40-45)	42 (38-45)	42 (40-45)	0.46
Platelets ($\times 10^9/\text{L}$)	311 (231-424)	309 (199-479)	311 (234-424)	0.92
METAVIR fibrosis closest to MRCP	2 (1-3)	2.5 (2-4)	1 (1-2)	0.02
F0	6 (13%)	0	6 (17%)	
F1	15 (33%)	1 (10%)	14 (40%)	
F2	12 (27%)	4 (40%)	8 (23%)	
F3	7 (16%)	2 (20%)	5 (14%)	
F4	5 (11%)	3 (30%)	2 (6%)	
PSC MRS	-0.97 (-1.8--0.33)	-0.09 (-1.8-0.85)	-1.3 (-1.8--0.47)	0.087
IHD score		-	-	-
0	4 (9%)			
1	20 (44%)			
2	17 (38%)			
3	4 (9%)			
EHD score*		-	-	-
0	17 (40%)			
1	12 (28%)			
2	9 (21%)			
3	5 (12%)			
4	0			
Sum IHD-EHD score		-	-	-
0	4 (9%)			
1	7 (16%)			
2	15 (33%)			
3	9 (20%)			
4	5 (11%)			
5	3 (7%)			
6	2 (4%)			
7	0			
Average IHD score	1 (1-1.6)	-	-	-
Average EHD score	0.5 (0-1.5)	-	-	-

TABLE 3. Continued

n (%) or Median (IQR)	All PSC	Sum IHD-EHD Score ≥ 4	Sum IHD-EHD Score < 4	PValue
	(n = 45)	(n = 10)	(n = 35)	
Sum average IHD-EHD score	1.9 (1-2.5)	–	–	–
Portal hypertension	15 (33%)	7 (70%)	8 (23%)	0.009
Ascites	5 (11%)	3 (30%)	2 (6%)	0.065
Esophageal varices	8 (18%)	4 (40%)	4 (11%)	0.059
Variceal bleed	4 (9%)	2 (20%)	2 (6%)	0.21
Cholangiocarcinoma	1 (2%)	1 (10%)	0	0.22
Liver transplant	5 (11%)	5 (50%)	0	< 0.001
PSC-related complication	10 (22%)	6 (60%)	4 (11%)	0.004

*Two MRCPs could not be examined for extrahepatic involvement.

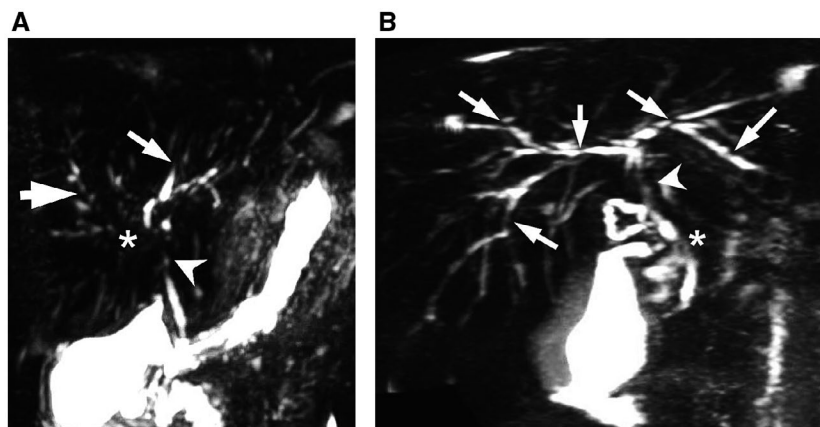


FIG. 1. Comparison of patients with UC. (A) An 11-year-old female patient shows severe pruning of the left IHD, grade 3 (thin arrow) and lack of visualization of the right main IHD, grade 3 (asterisk); multiple strictures/saccular dilations of the right IHD, grade 2 (thick arrow); stricture of the CHD, grade 3 (arrowhead) MRCP (scale 3:1, CBD 4 mm). (B) A 17-year-old male patient shows multiple strictures/saccular dilations of all IHD, grade 2 (arrows); stricture of the CHD, grade 3 (arrowhead); and segmental stricture of the proximal CBD (asterisk) MRCP (scale 2.5:1, CBD 4 mm).

and ICCs from 0.79-0.81, indicating substantial to excellent agreement.

PREDICTIVE ABILITY AND VALIDITY

The c-statistics summarizing the ability of the various MRCP scores to predict progression to a PSC-related complication in a time-dependent manner are listed in Table 5. As shown, the IHD-EHD sum score (derivation shown in Supporting Fig. S1) had the highest c-statistic and therefore the best discriminative ability. This was also the case for the secondary outcome of LT. Furthermore, the sum IHD-EHD score outperformed all the laboratory parameters, MRS, and fibrosis stage,

except for total bilirubin, which had a similar c-statistic for the primary outcome (Table 5). Parenchymal findings on MRI, including imaging features consistent with cirrhosis and periportal edema, had inferior predictive capacity (c-statistic, 0.75 and SE, 0.08; and c-statistic, 0.64 and SE, 0.04, respectively).

The predictive ability of the sum IHD-EHD score was further examined using a Kaplan-Meier curve (Fig. 2A), which depicts time to a PSC-related complication. The significant log-rank *P* value indicated a significant difference in risk across strata. The independent association between the sum IHD-EHD score and time to a PSC-related complication was then assessed using Cox proportional hazards regression. Sex, age, IBD, ASC, large-duct PSC, and MRS were selected for examination

TABLE 4. INTER-RATER RELIABILITY

MRCP Score	Weighted* Kappa (95% CI)	ICC (95% CI)
IHD score	0.81 (0.68-0.95)	–
EHD score	0.79 (0.63-0.94)	–
Sum IHD-EHD score	0.78 (0.64-0.91)	–
Average IHD score	–	0.81 (0.69-0.89)
Average EHD score	–	0.82 (0.69-0.90)
Sum average IHD-EHD score	–	0.79 (0.66-0.88)

*Using Cicchetti-Allison weights, as per SAS default. (https://support.sas.com/documentation/cdl/en/statug/63347/HTML/default/viewer.htm#statug_freq_a0000000665.htm).

TABLE 5. HARRELL'S C-STATISTICS REFLECTING DISCRIMINATIVE ABILITY OF MRCP TO PREDICT PROGRESSION TO PSC-RELATED COMPLICATION AND LIVER TRANSPLANT

MRCP Score	C-Statistic for PSC Complication (SE)	C-Statistic for LT (SE)
IHD score	0.73 (0.05)	0.79 (0.06)
EHD score	0.73 (0.10)	0.93 (0.03)
Sum IHD-EHD score	0.80 (0.06)	0.97 (0.01)
Average IHD score	0.77 (0.06)	0.77 (0.06)
Average EHD score	0.68 (0.10)	0.88 (0.07)
Sum average IHD-EHD score	0.75 (0.06)	0.90 (0.03)
ALT (U/L)	0.49 (0.09)	0.32 (0.10)
AST (U/L)	0.44 (0.10)	0.75 (0.10)
ALP (U/L)	0.65 (0.11)	0.81 (0.09)
GGT (U/L)	0.64 (0.10)	0.80 (0.09)
Total bilirubin ($\mu\text{mol/L}$)	0.83 (0.07)	0.83 (0.14)
Albumin (g/L)	0.62 (0.10)	0.63 (0.11)
Platelets ($10^9/\text{L}$)	0.65 (0.13)	0.56 (0.20)
METAVIR fibrosis stage	0.77 (0.08)	0.71 (0.15)
PSC MRS	0.75 (0.08)	0.79 (0.12)

based on clinical relevance. In univariate analysis, only sum IHD-EHD score (hazard ratio [HR], 2.00; 95% confidence interval [CI], 1.32-3.04) and MRS (HR, 1.95; 95% CI, 1.21-3.15) showed a significant association with the outcome. Using the change in estimate approach, a final model that included sum IHD-EHD score and MRS was constructed. The IHD-EHD sum score (but not MRS) retained its significant association with the outcomes (HR, 1.74; 95% CI, 1.14-2.64) (Table 6), indicating that higher MRCP sum scores are independently associated with higher rates of progression to PSC complications.

Sum IHD-EHD scores were dichotomized as ≥ 4 versus < 4 based on visual inspection of Fig. 2A to

allow comparisons between patients with PSC with high versus low MRCP severity. The fibrosis stage was higher and PSC-related complications were more frequent in patients with high versus low MRCP sum scores (Table 1, second and third columns). In addition, there was a trend toward higher ALP in the high MRCP severity group, and the sum IHD-EHD score correlated significantly with fibrosis stage (Spearman $r = 0.42$; $P = 0.004$). The relationship between sum IHD-EHD scores and fibrosis stage is further illustrated in Supporting Fig. S2; higher MRCP scores were observed in patients with more advanced fibrosis ($P = 0.019$). MRCP sum IHD-EHD scores were significantly positively correlated with total bilirubin (Spearman $r = 0.33$; $P = 0.03$) and negatively correlated with platelets (Spearman $r = -0.31$; $P = 0.04$). We did not find a significant correlation with MRS (Spearman $r = 0.18$; $P = 0.24$). Correlations with other biochemical parameters were not significant ($P > 0.05$ for alanine aminotransferase [ALT], aspartate aminotransferase [AST], ALP, GGT, albumin).

A simplified version of the sum IHD-EHD score based on visual inspection of Fig. 2A was generated by collapsing the original eight levels to three levels: level 1 (scores 0-1); level 2 (scores 2-3); and level 3 (scores 4-7). The Kaplan-Meier curve for this simplified sum score is illustrated in Fig. 2B. The overall significant log-rank test indicated a significant difference in risk across strata. Using a series of pairwise comparisons, children in the highest risk group were found to have a significantly higher rate of progression to PSC-related complications than children in the lowest risk group ($P < 0.001$) and children in the intermediate risk group ($P = 0.02$). The inter-rater reliability and predictive ability of this simplified sum score were similar to those of the original sum IHD-EHD score (weighted kappa, 0.80; 95% CI, 0.67-0.94; c-statistic, 0.77 and SE 0.06, for PSC-related complication; c-statistic, 0.94 and SE 0.02, for LT).

SENSITIVITY ANALYSIS

Similar results were obtained in a sensitivity analysis restricted to children with large-duct PSC and follow-up duration ≥ 6 months ($n = 39$); the c-statistics for the sum IHD-EHD score were 0.82 (SE, 0.06) and 0.93 (SE, 0.03) for PSC-related complication and LT, respectively, and the MRCP sum score retained a significant association with time to a PSC-related complication (HR, 1.85; 95% CI, 1.17-2.94)

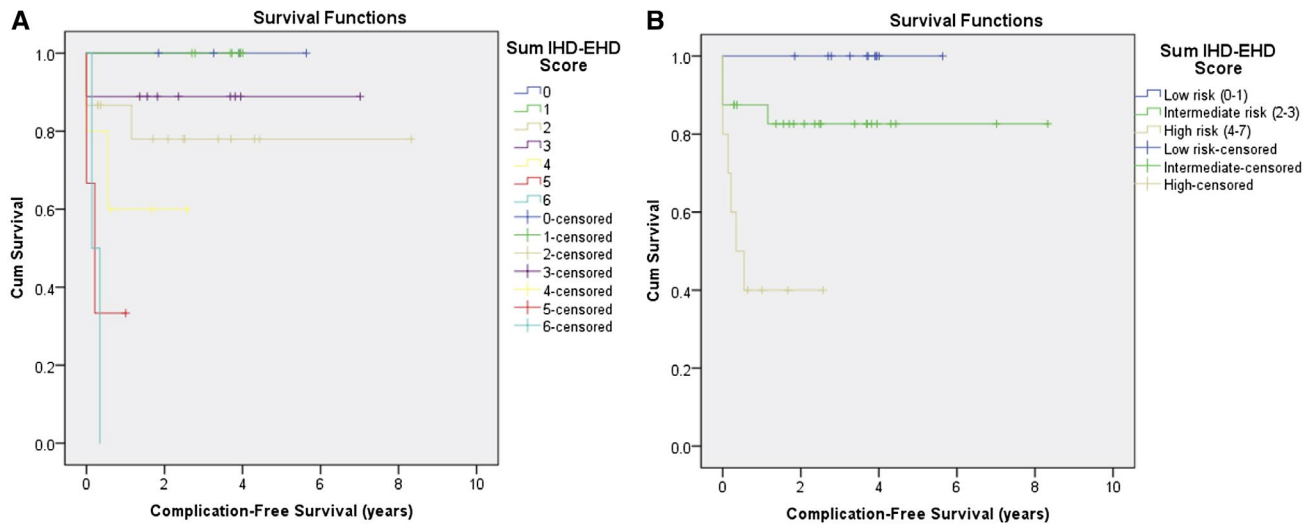


FIG. 2. Comparison of Kaplan-Meier curves for time to PSC-related complication. (A) By sum IHD-EHD score (overall log-rank, $P = 0.004$). (B) By simplified sum IHD-EHD score (overall log-rank, $P = 0.001$). Abbreviation: Cum, cumulative.

TABLE 6. UNADJUSTED AND ADJUSTED HRs FOR TIME TO PSC-RELATED COMPLICATION

Factor	Unadjusted HR (95% CI)	P Value	Adjusted HR (95% CI)	P Value
Sum IHD-EHD score	2.00 (1.32-3.04)	0.001	1.74 (1.14-2.64)	0.010
Age at PSC diagnosis (years)	0.98 (0.84-1.13)	0.76	—	—
Male	0.51 (0.15-1.75)	0.28	—	—
UC/IBD-U (vs. CD/no IBD)	1.22 (0.26-5.76)	0.80	—	—
ASC	1.28 (0.33-4.97)	0.72	—	—
Large-duct PSC	1.17 (0.15-0.93)	0.88	—	—
PSC MRS	1.95 (1.21-3.15)	0.006	1.61 (0.94-2.7)	0.084

after controlling for MRS. In this sensitivity analysis, MRS also remained significantly associated with progression to a PSC-related complication (HR, 1.81; 95% CI, 1.03-3.19).

Discussion

While MRCP is widely used to diagnose PSC, rigorous validation data are sparse in adults and the ability of MRCP to prognosticate outcomes has not been assessed in children. Reliability data are lacking as well. Here, we assess the inter-rater reliability of the modified Majoie classification applied to MRCP in a pediatric cohort with PSC and present initial validity data, including ability to predict progression to meaningful clinical outcomes and construct validity. We examined multiple variations of the Majoie

classification and found that a score derived by summing the worst affected intrahepatic and extrahepatic regions performed best at predicting PSC complications and LT, supporting that a composite of intrahepatic and extrahepatic involvement outperforms either individually and that it is the worst region of disease that is most relevant for prognostic purposes. We demonstrated substantial to excellent inter-rater reliability and supported the score's construct validity by showing correlations with other relevant biomarkers and significant differences between groups with high versus low scores. A simplified three-level version of the tool (low, intermediate, high risk) performed well.

In 2002, Ponsioen et al.⁽¹⁰⁾ investigated the prognostic value of the modified Majoie classification in 174 adult patients with PSC undergoing ERCP. They, too, found that the parameter that performed best for predicting outcomes (survival) was a sum score

incorporating intrahepatic and extrahepatic involvement. A prognostic index that included this ERCP score and additional clinical data was constructed and subsequently validated in an external adult cohort.⁽¹¹⁾ This group observed that a sum score >4 was associated with worse outcomes, a finding that has now been replicated in our pediatric cohort.

Comparable MRCP data are sparse. Two small retrospective studies have attempted to correlate MRCP severity with outcomes. Petrovic et al.⁽¹⁷⁾ found no significant association between MRCP severity and the MRS in 47 adult patients with PSC. However, this was a cross-sectional analysis. Of note, we also found no significant correlation between MRCP and MRS in a cross-sectional fashion and yet showed that both had predictive ability for progression to PSC complications in survival analysis (including independently in the sensitivity analysis). This suggests that the two (MRCP and MRS) may capture different elements of the disease that have prognostic significance and highlights that the absence of correlation between two predictive biomarkers does not exclude their independent prognostic utility. Tenca et al.⁽¹⁸⁾ observed a weak correlation between extrahepatic MRCP scores and death/LT but not between intrahepatic scores and outcomes in their study of 48 adult patients with PSC. Predictive ability through c-statistics or survival analyses was not assessed. We could identify only one pediatric study that examined this MRCP score.⁽¹⁹⁾ In this retrospective assessment of 39 children with PSC, intrahepatic and extrahepatic MRCP scores did not correlate with GGT or ALP in a cross-sectional fashion (this was also the case in our study). However, the predictive ability of MRCP was not examined.

The largest and most rigorous validation study of MRCP for predicting outcomes in PSC was recently undertaken by Lemoinne et al.⁽²⁰⁾ In contrast to the above studies, which examined the score validated by Ponsioen et al.,^(10,11) Lemoinne et al. validated the Anali score. The Anali score was developed by Ruiz et al. in 2014.⁽²¹⁾ The primary endpoint against which it was derived was radiologic course, categorized as worsening, improvement, or stabilization. Two scores were derived: one without gadolinium (including intrahepatic dilatation, liver dysmorphism, and portal hypertension) and one with gadolinium (dysmorphism and parenchymal enhancement heterogeneity). In the retrospective validation study by Lemoinne et al.,⁽²⁰⁾ which included two 119-patient validation cohorts

(internal and external), the Anali scores without and with gadolinium were found to have c-statistics of 0.89 (95% CI, 0.84-0.95) and 0.75 (95% CI, 0.64-0.87) for predicting survival without LT or cirrhosis decompensation. The score without gadolinium performed best. Similarly, our MRCP sum score was performed without gadolinium; reliability was not assessed. We opted to examine the score validated by Ponsioen et al.^(10,11) rather than the Anali score in our pediatric cohort because our goal was to examine the significance of biliary involvement for predicting outcomes, which is better captured by the former score. Validation of the Anali score in a pediatric population does, however, represent an important future endeavor.

Our study has important implications. Surrogate endpoints and prognostic markers are sorely needed to facilitate the study of PSC, particularly for therapeutic trials. Imaging has been relatively underexplored as a potential biomarker, and there is a marked paucity of data on this topic in pediatrics. While our findings require validation in a larger prospective cohort, they suggest that MRCP has prognostic significance in pediatric PSC. Potential applications might include use of MRCP as a surrogate endpoint in trials or as a parameter on which to stratify randomization. The significance of our study extends beyond the research framework to routine clinical practice. Currently, there is no standard method for reporting MRCP findings in PSC. This leads to variability in reports, which hinders communication and tracking of disease progression. The sum IHD-EHD score is reliable and easy to use as it only requires grading the worst affected intrahepatic and extrahepatic regions. Moreover, in comparison to the standard MRCP surveillance protocol for children with suspected PSC (scan time 20-30 minutes), by tailoring the MRCP protocol to the bile ducts using only a few sequences (as in our study), scan times of under 10 minutes are feasible. This could reduce the need for sedation or general anesthesia in younger patients undergoing follow-up MRCP, with the caveat that more comprehensive imaging would still be required when looking for PSC complications, such as PH.

Our study has several strengths. They include the rigorous methods employed to score MRCPs (by liver segment and by two blinded readers with disagreements resolved by a third radiologist), the assessment of reliability, the examination of construct validity, and the use of clinically relevant outcomes as endpoints.

There are limitations as well, which include the study's small size, retrospective nature (including retrospective review of liver biopsy reports), and relatively short follow-up. While we recognize that MRCP technique may vary across centers, our technical parameters were consistent with standard protocols referenced in the literature.^(22,23)

In conclusion, an MRCP score derived by summing the modified Majoie classification applied to the worst affected intrahepatic and extrahepatic regions of the biliary tree is reliable in pediatric PSC and displays construct validity and prognostic ability in a small cohort. Important next steps include prospective validation in a larger external cohort with longer follow-up.

REFERENCES

- 1) Dyson JK, Beuers U, Jones DEJ, Lohse AW, Hudson M. Primary sclerosing cholangitis. *Lancet* 2018;391:2547-2559.
- 2) Deneau MR, El-Matary W, Valentino PL, Abdou R, Alqoer K, Amin M, et al. The natural history of primary sclerosing cholangitis in 781 children: a multicenter, international collaboration. *Hepatology* 2017;66:518-527.
- 3) Boonstra K, Weersma RK, van Erpecum KJ, Rauws EA, Spanier BW, Poen AC, et al. EpiPSCPBC Study Group. Population-based epidemiology, malignancy risk, and outcome of primary sclerosing cholangitis. *Hepatology* 2013;58:2045-2055.
- 4) Weismuller TJ, Trivedi PJ, Bergquist A, Imam M, Lenzen H, Ponsioen CY, et al.; International PSC Study Group. Patient age, sex, and inflammatory bowel disease phenotype associate with course of primary sclerosing cholangitis. *Gastroenterology* 2017;152:1975-1984.e8.
- 5) Lindstrom L, Boberg KM, Wikman O, Friis-Liby I, Hulterantz R, Prytz H, et al. High dose ursodeoxycholic acid in primary sclerosing cholangitis does not prevent colorectal neoplasia. *Aliment Pharmacol Ther* 2012;35:451-457.
- 6) Deneau MR, Mack C, Abdou R, Amin M, Amir A, Auth M, et al. Gamma glutamyltransferase reduction is associated with favorable outcomes in pediatric primary sclerosing cholangitis. *Hepatol Commun* 2018;2:1369-1378.
- 7) de Vries EM, Verheij J, Hubscher SG, Leeftang MM, Boonstra K, Beuers U, et al. Applicability and prognostic value of histologic scoring systems in primary sclerosing cholangitis. *J Hepatol* 2015;63:1212-1219.
- 8) Majoie CB, Reeders JW, Sanders JB, Huibregtse K, Jansen PL. Primary sclerosing cholangitis: a modified classification of cholangiographic findings. *AJR Am J Roentgenol* 1991;157:495-497.
- 9) Chen LY, Goldberg HI. Sclerosing cholangitis: broad spectrum of radiographic features. *Gastrointest Radiol* 1984;9:39-47.
- 10) Ponsioen CY, Vrouenraets SM, Prawirodirdjo W, Rajaram R, Rauws EA, Mulder CJ, et al. Natural history of primary sclerosing

cholangitis and prognostic value of cholangiography in a Dutch population. *Gut* 2002;51:562-566.

- 11) Ponsioen CY, Reitsma JB, Boberg KM, Aabakken L, Rauws EA, Schrupf E. Validation of a cholangiographic prognostic model in primary sclerosing cholangitis. *Endoscopy* 2010;42:742-747.
- 12) Lindor KD, Kowdley KV, Harrison ME; American College of Gastroenterology. ACG clinical guideline: Primary Sclerosing Cholangitis. *Am J Gastroenterol* 2015;110:646-659.
- 13) Ferrara C, Valeri G, Salvolini L, Giovagnoni A. Magnetic resonance cholangiopancreatography in primary sclerosing cholangitis in children. *Pediatr Radiol* 2002;32:413-417.
- 14) Levine A, Koletzko S, Turner D, Escher JC, Cucchiara S, De Ridder L, et al.; European Society of Pediatric Gastroenterology, Hepatology, and Nutrition. ESPGHAN revised porto criteria for the diagnosis of inflammatory bowel disease in children and adolescents. *J Pediatr Gastroenterol Nutr* 2014;58:795-806.
- 15) Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159-174.
- 16) Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychol Assess* 1994;6:284-290.
- 17) Petrovic BD, Nikolaidis P, Hammond NA, Martin JA, Petrovic PV, Desai PM, et al. Correlation between findings on MRCP and gadolinium-enhanced MR of the liver and a survival model for primary sclerosing cholangitis. *Dig Dis Sci* 2007;52:3499-3506.
- 18) Tenca A, Mustonen H, Lind K, Lantto E, Kolho KL, 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.
- 19) Cotter JM, Browne LP, Capocelli KE, McCoy A, Mack CL. Lack of correlation of liver tests with fibrosis stage at diagnosis in pediatric primary sclerosing cholangitis. *J Pediatr Gastroenterol Nutr* 2018;66:227-233.
- 20) Lemoine S, Cazzagon N, El Mouhadi S, Trivedi PJ, Dohan A, Kengang A, et al. Simple magnetic resonance scores associate with outcomes of patients with primary sclerosing cholangitis. *Clin Gastroenterol Hepatol* 2019; <https://doi.org/10.1016/j.cgh.2019.03.013>.
- 21) Ruiz A, Lemoine 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.
- 22) Chavhan GB, Almeshdar A, Moineddin R, Gupta S, Babyn PS. Comparison of respiratory-triggered 3-D fast spin-echo and single-shot fast spin-echo radial slab MR cholangiopancreatography images in children. *Pediatr Radiol* 2013;43:1086-1092.
- 23) Zenouzi R, Welle CL, Venkatesh SK, Schramm C, Eaton JE. Magnetic resonance imaging in primary sclerosing cholangitis-current state and future directions. *Semin Liver Dis* 2019;39:369-380.

Supporting Information

Additional Supporting Information may be found at onlinelibrary.wiley.com/doi/10.1002/hep4.1454/supinfo.