

Quantitative magnetic resonance imaging for the assessment of patients with primary sclerosing cholangitis

Doctoral thesis for the award of DPhil in Medical Sciences

Emmanuel Anandraj Selvaraj MBBS MRCP

Hertford College

October 2021

Candidate number: 1028321

Supervisors

Dr Michael Pavlides BSc MBBS MRCP DPhil

Dr Adam Bailey MB ChB FRCP

Prof Sir Michael Brady PhD FMedSci FREng FRS

I dedicate this thesis to my son, Ezra, who was stillborn on 19th November 2020.

The little time I had holding you in my arms has taught me so much more…

Contents

Acknowledgements 10
Declaration 12
Funding 12
Published abstracts 13
Publications arising 13
Other publications 14
Abbreviations 15
Index of tables 18
Index of figures 21
Thesis abstract 24
1 Introduction 26
1.1 Epidemiology of primary sclerosing cholangitis 26
1.2 Diagnosis of PSC 26
1.3 MRI-MRCP features in PSC 28
1.4 Complications of PSC 31
1.4.1 Biliary complications 31
1.4.2 Liver complications 32
1.4.3 Extrahepatic complications 33
1.4.4 Survival 33
1.4.5 Patient experience 34
1.5 Disease severity and prognosis 34
1.5.1 Serum markers 35
1.5.2 Prognostic risk models 36
1.5.3 Liver histology 38
1.5.4 Transient elastography 39
1.5.5 ERCP 39
1.5.6 Qualitative MRI and MRCP scores and features 41
1.6 Quantitative MR techniques 45
1.6.1 Magnetic resonance elastography 45
1.6.2 Diffusion-weighted MRI 45
1.6.3 Dynamic contrast-enhanced MRI 45
1.6.4 Spleen and liver size 48
1.6.5 Machine learning 48
1.7 Limitations of MRI-MRCP 49
1.8 Quantitative MR techniques assessed in this thesis 49

1.8.1 Liver*MultiScan*™ 49
1.8.2 *MRCP*+™ 50
1.9 Principles of MRI 50
1.9.1 MRI and relaxation times 50
1.9.2 T2-weighted MRCP 52
1.10 General aims 53
2 General methods 54
2.1 Study design 54
2.2 Eligibility criteria 54
2.2.1 General consideration 54
2.2.2 PSC 54
2.2.3 Healthy controls 55
2.3 MRI acquisition protocol 55
2.3.1 3D MRCP 56
2.3.2 Liver MRI 57
2.3.2.1 T1 mapping and T2* mapping 57
2.3.2.2 Volumetric interpolated breath hold examination 57
2.4 Quantitative MRI analysis 60
2.4.1 *MRCP*+ analysis 60
2.4.2 Liver*MultiScan* analysis 63
2.4.2.1 ROI-placement method 63
2.4.2.2 Liver-segmentation method 65
2.4.2.3 Superpixel analysis 67
2.5 Qualitative MRI analysis 69
2.6 Clinical and laboratory data 71
2.7 Transient elastography 71
2.8 Patient reported outcomes 72
2.9 PSC risk classification 73
2.9.1 Prognostic risk models 73
2.9.2 Cholestasis marker 74
2.9.3 Liver fibrosis markers 74
2.9.4 Qualitative MRI-MRCP scores 74
2.9.5 Disease distribution 75
2.9.6 Dominant stricture 75
2.10 Sample size calculation 76
2.11 Statistical analysis 77
3 Validation of quantitative image analysis methods 79

3.1 Chapter summary 79
3.2 Introduction 82
3.3 Variability in *MRCP+* metrics between methods of MRCP acquisition protocols 83
3.3.1 Background, aim and hypothesis 83
3.3.2 Methods 83
3.3.3 Results 84
3.3.4 Discussion 88
3.4 Repeatability and intra-observer reliability of *MRCP+* metrics in healthy volunteers 89
3.4.1 Background, aim and hypothesis 89
3.4.2 Methods 89
3.4.3 Results 90
3.4.4 Discussion 93
3.5 Variability in biliary tree visualisation between MRCP and *MRCP+* 94
3.5.1 Background, aim and hypothesis 94
3.5.2 Methods 94
3.5.3 Results 95
3.5.4 Discussion 97
3.6 Inter-observer and intra-observer variability in *MRCP+* metrics in PSC 98
3.6.1 Background, aim and hypothesis 98
3.6.2 Methods 98
3.6.3 Results 98
3.6.4 Discussion 101
3.7 Effect of pineapple juice consumption on liver cT1 measurements 102
3.7.1 Background, aim and hypothesis 102
3.7.2 Methods 102
3.7.3 Results 102
3.7.4 Discussion 103
3.8 Variability between liver T1 and iron corrected T1 103
3.8.1 Background, aim and hypothesis 103
3.8.2 Methods 104
3.8.3 Results 104
3.8.4 Discussion 106
3.9 Variability between methods of data processing for liver cT1 measurements in PSC 106
3.9.1 Background, aim and hypothesis 106
3.9.2 Methods 107
3.9.3 Results 107

3.9.4 Discussion 109
3.10 Conclusions 110
4 Quantitative assessment of cholangiopathy in PSC 111
4.1 Abstract 111
4.2 Introduction 113
4.3 Hypotheses 114
4.4 Aims 115
4.5 Participants and methods 116
4.5.1 Participants 116
4.5.2 Clinical and laboratory data 116
4.5.3 Transient elastography 116
4.5.4 Risk stratification 116
4.5.5 MRI acquisition 116
4.5.6 Qualitative image analysis 116
4.5.7 Quantitative image analysis 116
4.5.8 Normalisation to the length of the biliary tree 117
4.5.9 Statistical analysis 120
4.6 Results 121
4.6.1 Cohort characteristics 121
4.6.2 Comparison of *MRCP*+ metrics between PSC and healthy 125
4.6.3 Inter-observer agreement of MRI-MRCP scores and features 128
4.6.4 Inter-observer and intra-observer agreement of *MRCP*+ candidate stricture and dilatation metrics 128
4.6.5 Association of *MRCP*+ metrics with study variables 131
4.6.6 Performance of all MRCP+ metrics to discriminate high-risk vs low-risk PSC 138
4.6.7 The performance of intrahepatic sum of relative severity of dilatations (SumRelSevDilat) to detect high-risk PSC 145
4.6.7.1 Prognostic risk models 145
4.6.7.2 Cholestasis markers 146
4.6.7.3 Liver fibrosis markers 147
4.6.7.4 Qualitative MRI-MRCP scores 148
4.6.7.5 Presence of extrahepatic disease and dominant stricture 149
4.6.7.6 Independent predictors of high-risk PSC 151
4.6.8 Association between *MRCP*+ metrics, other study variables and patient-reported outcomes in PSC 155
4.7 Discussion 159
5 Quantitative assessment of liver parenchyma in PSC 165

5.1 Abstract 165
5.2 Introduction 167
5.3 Hypotheses 170
5.4 Aims 170
5.5 Participants and methods 171
5.5.1 Participants 171
5.5.2 Clinical, laboratory and liver stiffness data 173
5.5.3 Risk stratification 173
5.5.4 MRI acquisition 173
5.5.5 Quantitative image analysis 173
5.5.6 Qualitative image analysis 176
5.5.7 Statistical analysis 176
5.6 Results 177
5.6.1 Derivation cohort: MoBILISE study 177
5.6.1.1 Cohort characteristics 177
5.6.1.2 Comparison of quantitative metrics between PSC and healthy controls 180
5.6.1.3 Qualitative MRI data in PSC 182
5.6.1.4 Association between liver cT1 metrics and non-invasive markers 183
5.6.1.5 Risk stratification by liver fibrosis severity 186
5.6.1.6 Risk stratification by serum alkaline phosphatase 189
5.6.1.7 Risk stratification by prognostic risk models 191
5.6.1.8 Risk stratification by morphological features on liver MRI 193
5.6.1.9 Risk stratification by morphological features on MRCP 198
5.6.1.10 Independent predictors of high-risk PSC 201
5.6.1.11 Association between liver cT1 metrics and patient-reported outcomes in PSC 207
5.6.1.12 Association between liver fibrosis markers and patient-reported outcomes in PSC 207
5.6.2 Validation cohort: RIAL, NICOLA, and MAPLE studies 209
5.6.2.1 Cohort characteristics 209
5.6.2.2 Comparison of candidate liver cT1 metrics between liver diseases 211
5.6.2.3 Liver cT1 metrics and severity of liver fibrosis 214
5.7 Discussion 220
6 Assessment of periductal abnormalities 227
6.1 Abstract 227
6.2 Introduction 228
6.3 Hypotheses 230

6.4 Aims 231
6.5 Methods 232
6.5.1 Co-registration of axial liver cT1 maps and biliary tree model 232
6.5.2 Periductal cT1 quantification in 3D space 232
6.5.3 Partial volume effect 235
6.5.4 Relationship between periductal cT1 and other quantitative MRI and MRCP markers 237
6.5.5 Healthy volunteer validation cohort 238
6.5.6 Statistical analysis 238
6.6 Results 239
6.6.1 Cohort characteristics 239
6.6.2 Periductal mean cT1 in healthy controls 242
6.6.3 Validation of periductal cT1 in R&R study healthy volunteers 242
6.6.4 Periductal mean cT1 in patients with PSC 242
6.6.5 Influence of oblique angle bile ducts on periductal mean cT1 245
6.6.6 Association of periductal mean cT1 with liver fibrosis and inflammation 246
6.6.7 Risk stratification 247
6.6.8 Periductal cT1 stratified by quantitative MRI-derived marker of parenchymal heterogeneity 250
6.6.9 Periductal cT1 stratified by quantitative MRCP-derived marker of intrahepatic duct dilatation severity 250
6.6.10 Association between mean periductal cT1 and patient-reported outcomes in PSC 250
6.7 Discussion 251
7 Temporal relationship in candidate quantitative metrics 257
7.1 Abstract 257
7.2 Introduction 259
7.3 Hypothesis 260
7.4 Aims 261
7.5 Methods 262
7.5.1 Follow-up assessments 262
7.5.2 Statistical analysis 262
7.6 Results 263
7.6.1 Cohort characteristics 263
7.6.2 Within-subject temporal relationship of candidate MRI metrics and other markers 265
7.6.3 Between-subject and within-subject effect interaction on candidate MRI metrics 267

7.6.3.1 Intrahepatic SumRelSevDilat 267
7.6.3.2 IQR cT1 270
7.6.3.3 Mean superpixel variance cT1 273
7.6.3.4 Mean periductal cT1 1.5-3.5mm 276
7.6.4 Comparison of the diagnostic performance of candidate MRI metrics at baseline and their temporal change from baseline to detect the high-risk PSC group 279
7.6.5 Differences in patient reported outcomes between the two study visits 283
7.7 Discussion 285
8 General discussion and thesis conclusion 289
8.1 Summary of main findings 289
8.2 Implications of thesis findings 292
8.3 Limitations 293
8.4 Suggested further work 295
8.5 Conclusion 296
References 297

Acknowledgements

First and foremost, I am grateful to all the participants and their families who took part in this study. They have kindly sacrificed their time and in some cases travelled a long distance whilst fasting to attend scan appointments in the morning.

The decision to embark on this DPhil journey originated following several discussions on future career paths with Adam Bailey during our endoscopy sessions, and subsequently with Michael Pavlides, who inspired me to follow in his footsteps into academia. They introduced me to Mike Brady, a world-renowned expert in medical imaging. Together, they helped shape this project and I had the privilege of being supervised by three DPhil supervisors. I am indebted to both Michael and Adam for taking me under their wings, challenging me to think bigger and aim higher, keeping their doors always open despite their busy schedules, and supporting me through a particularly difficult life event that occurred during this DPhil. They have become both my mentors and friends.

This thesis would not have been possible without the support from the incredible team at the Oxford Centre for Clinical Magnetic Resonance Research (OCMR), Radcliffe Department of Medicine. I would like to thank Rebecca Mills, Fiona McMillan, Joana Pelado, Raj Soundaralingam, and Cassandra Baker for their teaching and advice with MRI scanning, Catherine Krasopoulos and Judith Delos-Santos for teaching me centrifugation, preparation, and storage of aliquot samples, Mike MacDonald for his expertise with IT, Kathryn Lacey for addressing logistical problems within the department, Heidi Crook, Charlotte Rush, Emma Grant, Sharon Dash and Cate Wheatley for their administrative help and for managing the study finances, Karen Whatley and Ella May in OCMR reception for organising study participant transport and courier services, Hanan Lamlum and Philippa Kemp for their help with gaining regulatory approvals and study monitoring.

During my time at OCMR, I had the opportunity to discuss, exchange ideas, and collaborate with other research fellows who have influenced this work through informal interactions or provided moral support. I would like to thank Ferenc Mozes, Arjun Jayaswal, Christina Levick, Liam Young, Gabriela Belsley, Ines Abdesselam, Moritz Hundertmark and Andrew Apps. I was also lucky to have rubbed shoulders with some of the bright minds in the field during my time there. I would like to thank Stefan Neubauer, Ellie Barnes, Stefan Piechnik, Vanessa Ferreira, and Elizabeth Tunnicliffe for their encouraging feedback and pointers during departmental presentations of my work.

The work carried out in this thesis would not have been possible without close collaborations with Perspectum. I would like to thank Ged Ridgway, Bruno Paun, Carlos Ferreira, Mariana Marieiro, Marc Goldfinger, Marija Mavar, Andrea Borghetto, Alexander Bagur, Neelam Hassanali, Helena Thomaides-Brears, and Matt Kelly for teaching me how to analyse my scans, engaging in discussions for improvement, and working collaboratively to enable independent data analysis of my scans. I would like to specifically take the opportunity to thank Ged Ridgway, whom I worked closely with on this project. Our discussions on medical imaging helped spark ideas and concepts that led to further exploration in this thesis.

The incredible support I have received from the Hepatology clinical and research teams of the Translational Gastroenterology Unit, Nuffield Department of Medicine deserves a mention. I would like to thank my colleagues Jane Collier, Emma Culver, and Jeremy Cobbold for their intellectual input into the study through discussions and their help with recruitment in clinic, Justyna Szczurkowska and Lou Holland for performing FibroScan on my study participants, Loren Smith, and Abbie Martin for their help with patient identification and coordinating with me to recruit patients into other PSC studies. A special mention to Emma Culver who introduced me to the International PSC Study Group MRI working party and provided scientific and clinical input to the project. As part of the study, the help of two external radiologists were enlisted to report the study scans. I would like to thank Ahmed Bassalamah and Sarah Potter-Lang for helping with this piece of work.

I would like to express my gratitude to my parents Selvaraj Sandoshom and Devathan Jamal Lily for their sacrifice and hardship that they endured to provide me with the best education and supporting my dreams of studying in the UK. I would not be where I am today without their prayers. I would also like to thank my sisters Priscilla Shanthi, Rebecca Jeyanthi, Deborah Vasanthi, and Beulah Malathi for their regular words of encouragement, and my close friends Viyaasan Mahalingasivam, Jerny Dela Cruz, Pairaw Kader, and Maygol Dadgar for supporting me through a tragic life event during my DPhil.

The biggest pillar of strength that enabled me to complete my work is my wife, Ashton Burden-Selvaraj. She is the rock upon which I stand and the sea upon which I float. Thank you for putting your career plans on hold to enable me to pursue this DPhil, sharing my joys and laughter on good days, and raising my spirits with your words of wisdom on those bad days. I could not have done this without you by my side.

Declaration

I certify that the work presented in this thesis is the result of my own study and research and has been written entirely by me. No part of my thesis has been accepted or is currently being submitted for any degree, diploma or certificate or other qualification in this University or elsewhere.

The MoBILISE study protocol (chapter 2) and technical validation experiments (chapter 3) were designed by me with supervision from Dr Michael Pavlides, Dr Adam Bailey and Prof Michael Brady. I screened, recruited, collected clinical data, and scanned all the patients and healthy volunteers recruited to this study. I was trained to perform the *MRCP*+ image analysis and conducted the analysis for all the technical validation data (chapter 3) and baseline data (chapter 4). Liver*MultiScan* image analysis (chapter 5) and co-registration of *MRCP*+ and Liver*MultiScan* images (chapter 6) were performed by imaging scientists at Perspectum blinded to the clinical data. I performed the quality check for the automated *LiverMultiScan* analysis which involved manual inspection and re-contouring of the liver mask. Due to Covid-19 restrictions, the follow-up *MRCP*+ image analysis (chapter 7) was performed by Perspectum. All statistical analysis and data interpretation were done by me.

Funding

This MoBILISE study was funded by the NIHR Oxford Biomedical Research Centre.

Published abstracts

1. Selvaraj EA, Ridgway G, Culver EL, et al. Combination of quantitative 3D MRCP and multi-parametric liver MRI demonstrates increased periductal iron-corrected T1 in large-duct primary sclerosing cholangitis. Journal of Hepatology August 2020; 73: S86

2. Selvaraj EA, Ridgway G, Brady JM, et al. Quantitative 3D MRCP-derived dilatation score in the assessment of primary sclerosing cholangitis. Hepatology October 2019; 70(S1):188-1382.

3. Selvaraj EA, Ridgway G, Brady JM, et al. Quantifying the heterogeneity of liver inflammation and fibrosis in primary sclerosing cholangitis using the interquartile range of iron-corrected T1 from multi-parametric MRI. Hepatology October 2019; 70(S1):188-1382.

Publications arising

1. 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(6):644-658.

2. Selvaraj EA, Ba-Ssalamah A, Poetter-Lang S, Ridgway GR, Brady JM, Collier J, Culver EL, Bailey A, Pavlides M. A quantitative MRCP metric of intrahepatic biliary dilatation severity detects high-risk primary sclerosing cholangitis. Hepatol Commun. 2021 Nov 21. doi: 10.1002/hep4.1860. Epub ahead of print. PMID: 34802195.

Other publications

During my DPhil, I also undertook two other projects that resulted in publications.

1. Selvaraj EA*, Mozes FE*, Jayaswal ANA*, et al. Diagnostic accuracy of elastography, and magnetic resonance imaging in patients with NAFLD: a systematic review and meta-analysis. J Hepatol. 2021 May;75(4):770-785.

2. Mozes FE, Lee J, Selvaraj EA, et al. Diagnostic accuracy of non-invasive tests for advanced fibrosis in patients with NAFLD – An individual patient data meta-analysis. Gut. 2021 May 17: gutjnl-2021-324243.

3. Radford -Smith DE*, Selvaraj EA*, Peters R, et al. A novel serum metabolomic panel distinguishes IgG4-related sclerosing cholangitis from primary sclerosing cholangitis. Liver Int. 2022 Feb 7. doi: 10.1111/liv.15192. Epub ahead of print. PMID: 35129255.

* joint first authors

Abbreviations

2D	Two dimensional
3D	Three dimensional
AASLD	American Association for the Study of Liver Diseases
ADC	Apparent diffusion coefficient
AIH	Autoimmune hepatitis
ALP	Alkaline phosphatase
AOM	Amsterdam-Oxford model
APRI	AST to Platelet Ratio Index
ArLD	Alcohol-related liver disease
AST	Aspartate aminotransferase
AUC	Area under the curve
BMI	Body mass index
BTvol	Biliary tree volume
CA19-9	Carbohydrate antigen 19-9
CBD	Common bile duct
CCA	Cholangiocarcinoma
CD	Cystic duct
CHD	Common hepatic duct
CI	Confidence interval
CoV	Coefficient of variation
cT1	Iron-corrrected T1 relaxation time
DCE	Dynamic contrast enhanced
DilatNum	Number of candidate dilatations
DilatTreeProp	Proportion of tree dilated
DS	Dominant stricture
DWI	Diffusion weighted imaging
EASL	European Association for the Study of the Liver
ECG	Electrocardiogram
EHD	Extrahepatic disease
ELF	Enhanced liver fibrosis
ERCP	Endoscopic retrograde cholangiopancreatography
ESSS	Extrahepatic stricture severity score
FIB-4	Fibrosis-4
FOV	Field of view
GBvol	Gallbladder volume
GGT	Gamma-glutamyl transferase
HA	Hyaluronic acid
HBV	Hepatitis B virus
HCV	Hepatitis C virus
HRQoL	Health-related quality of life
IBD	Inflammatory bowel disease
ICC	Intraclass correlation coefficient
IgG4	Immunoglobulin G-subtype 4

IgG4-SC	Immunoglobulin G-subtype 4 related sclerosing cholangitis
IHBD	Intrahepatic bile duct dilatation
INR	International normalised ratio
IPSCSG	International Primary Sclerosing Cholangitis Study Group
IQR	Interquartile range
ISSS	Intrahepatic stricture severity score
IU	International unit
LD-PSC	Large-duct primary sclerosing cholangitis
LHD	Left hepatic duct
LoA	Limits of agreement
LS	Liver stiffness
LTX	Liver transplantation
MELD	Model for End-stage Liver Disease
MIP	Maximum intensity projection
MoBILISE	Multiparametric magnetic resonance imaging of biliary diseases
MOLLI	Modified Look-Locker inversion recovery
MRCP	Magnetic resonance cholangiopancreatography
MRE	Magnetic resonance elastogram
MRI	Magnetic resonance imaging
MRS	Mayo risk score
NAFLD	Non-alcoholic fatty liver disease
NPV	Negative predictive value
OR	Odds ratio
PBC	Primary biliary cholangitis
PD	Pancreatic duct
PIIINP	Type III procollagen peptide
PPV	Positive predictive value
PSC	Primary sclerosing cholangitis
PT	Prothrombin time
PTC	Percutaneous transhepatic cholangiogram
QoL	Quality of life
RC	Repeatability coefficient
RHD	Right hepatic duct
RLE	Relative liver enhancement
ROC	Receiver operating characteristic curve
ROI	Region of interest
SD	Standard deviation
SEM	Standard error of measurement
shMOLLI	Shortened Modified Look-Locker inversion recovery
SI	Signal intensity
StricNum	Number of candidate strictures
StricTreeProp	Proportion of tree strictured
SumRelSevDilat	Sum of relative severity of candidate dilatations
SumRelSevStric	Sum of relative severity of candidate strictures
T1	T1 relaxation time

T2	T2 relaxation time
TE	Echo time
TIMP	Tissue inhibitor of matrix metalloproteinases
TR	Repetition time
TSSS	Total stricture severity score
TreeDilatScore	Tree dilatation score
TreeStricScore	Tree stricture score
UC	Ulcerative colitis
UDCA	Ursodeoxycholic acid
ULN	Upper limit normal
VCTE	Vibration controlled transient elastography
VIBE	Volumetric interpolated breath hold examination

Index of tables

Table 1. Morphological features of PSC on MRI-MRCP 30
Table 2. Detailed acquisition parameters for MRI. 58
Table 3. The original and modified score from the Amsterdam cholangiographic classification system. 69
Table 4. Classification of components for the derivation of the Anali score. 70
Table 5. Other morphological features assessed on MRI-MRCP interpretation. 70
Table 6. Surrogate prognostic biomarkers in PSC that was selected for use in this thesis. . 75
Table 7. Tree and individual duct metrics variability between protocol 1 (Perspectum) and protocol 2 (Siemens) in 10 participants. 85
Table 8. Tree metrics variability between protocol 1 (Perspectum) and protocol 2 (Siemens) stratified by cohort. 86
Table 9. Repeatability of tree metrics in 10 healthy volunteers. 91
Table 10. Repeatability of extrahepatic bile duct diameter measurements in 10 healthy volunteers. 91
Table 11. Intra-observer variability of *MRCP*+ metrics assessed by Bland-Altman analyses and intraclass correlation coefficients in 10 healthy volunteers. 92
Table 12. The proportion of the biliary tree segment modelled without a break in duct continuity in patients with PSC and healthy volunteers. 97
Table 13. Intra- and inter-observer variability of *MRCP*+ metrics assessed by Bland-Altman analyses and intraclass correlation coefficients in 30 patients with large-duct PSC. 100
Table 14. Clinical details of the four patients with PSC that demonstrated a difference in T1 and cT1 measurements. 105
Table 15. Quantitative metrics derived from *MRCP*+ analysis and their respective definitions in this study. 118
Table 16. Baseline characteristics of participants included for MRCP+ analysis. 123
Table 17. Characteristics of the four patients excluded from the study. 124
Table 18. Diagnostic performance of *MRCP*+ whole tree metrics to distinguish between patients with PSC and healthy controls. 127
Table 19. Qualitative interpretation of MRCP for morphological features and their corresponding inter-observer agreements. 129
Table 20. Manual measurements of bile ducts by MRI-MRCP readers and their corresponding inter-observer agreements. 130
Table 21. Inter- and intra-observer agreement for *MRCP*+ dilatation and stricture metrics. 130
Table 22. Automatic measurements of bile duct diameter by *MRCP*+ and their corresponding inter-observer agreements. 130
Table 23. Whole tree *MRCP*+ metrics correlation with blood tests. 132
Table 24. Whole tree *MRCP*+ metrics correlation with prognostic risk models and liver fibrosis markers. 133
Table 25. Whole tree *MRCP*+ metrics correlation with qualitative MRCP scores. 134
Table 26. Intrahepatic tree *MRCP*+ metrics correlation with blood tests. 135
Table 27. Intrahepatic *MRCP*+ metrics correlation with prognostic risk models and liver fibrosis markers. 136
Table 28. Intrahepatic tree *MRCP*+ metrics correlation with qualitative MRCP scores. 137
Table 29. *MRCP*+ metrics stratified into high-risk and low-risk groups by Mayo Risk Score (MRS > 0 or ≤ 0) and Amsterdam-Oxford Model (AOM > 2 or ≤ 2). 139
Table 30. *MRCP*+ metrics stratified into high-risk and low-risk groups by times upper limit of normal serum alkaline phosphatase (xULN ALP > 2.2 or ≤ 2.2) and (xULN ALP > 1.5 or ≤ 1.5). 140

Table 31. *MRCP+* metrics stratified into high-risk and low-risk groups by transient elastography liver stiffness (LS > 9.6kPa or ≤ 9.6kPa) and enhanced liver fibrosis score (ELF > 9.8 or ≤ 9.8)..... 141
Table 32. *MRCP+* metrics stratified into high-risk and low-risk groups by modified Amsterdam stricture severity score (intrahepatic (ISSS): > 2 or ≤ 2, whole tree (TSSS): > 4 or ≤ 4) and Anali score (> 2 or ≤ 2). 142
Table 33. *MRCP+* metrics stratified into high-risk and low-risk groups by the presence of extrahepatic PSC and dominant stricture (present vs absent)..... 143
Table 34. Association between intrahepatic sum of relative severity of dilatations (SumRelSevDilat) and qualitative MRI-MRCP scores with non-invasive surrogate markers of disease severity and progression in PSC. 144
Table 35. Performance of non-invasive markers to detect the high-risk PSC group stratified by non-invasive risk stratification tools..... 150
Table 36. Univariate and multiple logistic regression analysis for the detection of high-risk PSC defined by Mayo Risk Score and Amsterdam-Oxford Model score. 153
Table 37. Multivariate logistic regression model parameters to detect Mayo Risk Score (MRS) > 0 group. 154
Table 38. Multivariate logistic regression model parameters to detect Amsterdam-Oxford Model (AOM) > 2 group. 154
Table 39. Correlation between SF-36 and PSC-PRO domains with the intrahepatic sum of relative severity of candidate dilatations metric (SumRelSevDilat)..... 158
Table 40. Correlation between SF-36 and PSC-PRO domains and serum ALP. 158
Table 41. Baseline characteristics of participants included in the MoBILISE study. 179
Table 42. Diagnostic performance of LiverMultiScan liver cT1 metrics to distinguish between patients with PSC and healthy controls. 181
Table 43. Qualitative interpretation of MRI for morphological features and their corresponding inter-observer agreements. 182
Table 44. Correlation between quantitative liver MRI metrics and study variables. 184
Table 45. Performances of liver cT1 metrics to discriminate between high-risk and low-risk PSC stratified by transient elastography liver stiffness (LS) and enhanced liver fibrosis score (ELF) thresholds. 187
Table 46. Performances of liver cT1 metrics to discriminate between high-risk and low-risk PSC stratified by upper limit of normal (ULN) serum alkaline phosphatase thresholds..... 190
Table 47. Performances of liver cT1 metrics to discriminate between high-risk and low-risk PSC stratified by Amsterdam-Oxford Model score (AOM) and Mayo Risk Score (MRS) thresholds. 192
Table 48. Liver cT1 metrics in high-risk and low-risk PSC stratified by Anali score and morphological features on MRI. 196
Table 49. Performances of liver cT1 metrics to discriminate between high-risk and low-risk PSC stratified by Anali score, presence of hepatic dysmorphy, portal hypertension, and cirrhosis..... 197
Table 50. Liver cT1 metrics in high-risk and low-risk PSC stratified by stricture severity scores and presence of dominant stricture. 200
Table 51. Performance of IQR cT1 to discriminate between high-risk and low-risk PSC groups stratified by modified Amsterdam severity scores and the presence of a dominant stricture. 201
Table 52. Performance of non-invasive markers to discriminate high-risk vs low-risk PSC group stratified by Amsterdam-Oxford Model and Mayo Risk Score thresholds..... 203
Table 53. Univariate and multiple logistic regression analysis for the detection of high-risk PSC defined by Amsterdam-Oxford Model score and Mayo Risk Score thresholds. 204

Table 54. Multivariate logistic regression model parameters to detect Amsterdam-Oxford Model (AOM) > 2 group. 205
Table 55. Multivariate logistic regression model parameters to detect Mayo Risk Score (MRS) > 0 group. 206
Table 56. Correlation between SF-36 domains and median cT1. 208
Table 57. Correlation between SF-36 and PSC-PRO domains and transient elastography liver stiffness. 208
Table 58. Baseline characteristics of all participants included in the final analysis. 211
Table 59. Baseline characteristics of 73 patients with PSC included in the final analysis. . 241
Table 60. Correlation between SF-36 domains and mean periductal cT1 1.5–3.5mm. 251
Table 61. Characteristics of patients with large-duct PSC at baseline and follow-up visits. 264
Table 62. Summary results of candidate MRI metrics and other markers at visits 1 and 2. 265
Table 63. Summary of mixed-model ANOVA results assessing the interaction of intrahepatic SumRelSevDilat with time and risk categorisation stratified by non-invasive surrogate markers of disease severity. 269
Table 64. Summary of mixed-model ANOVA results assessing the interaction of IQR cT1 with time point and risk categorisation stratified by non-invasive surrogate markers of disease severity. 272
Table 65. Summary of mixed-model ANOVA results assessing the interaction of mean superpixel variance cT1 with time point and risk categorisation stratified by non-invasive surrogate markers of disease severity. 275
Table 66. Summary of mixed-model ANOVA results assessing the interaction of mean periductal cT1 1.5–3.5mm with time point and risk categorisation stratified by non-invasive surrogate markers of disease severity. 278
Table 67. Performance of intrahepatic SumRelSevDilat and IQR cT1 at baseline and their changes from baseline to discriminate high-risk vs low-risk PSC groups stratified by non-invasive surrogate markers. 280
Table 68. Performance of mean superpixel variance cT1 and periductal cT1 1.5–3.5mm at baseline and their changes from baseline to discriminate high-risk vs low-risk PSC groups stratified by non-invasive surrogate markers. 281
Table 69. Performance of transient elastography liver stiffness (LS), enhanced liver fibrosis score (ELF), and serum alkaline phosphatase (ALP) at baseline and their changes from baseline to discriminate high-risk vs low-risk PSC groups stratified by non-invasive surrogate markers. 282

Index of figures

Figure 1. Typical features of large-duct PSC on MRCP maximum intensity projection (MIP) images. 29
Figure 2. Summary of complications resulting from disease progression in PSC..... 31
Figure 3. An example of maximum intensity projection (MIP) image from 3D MRCP acquisition. 56
Figure 4. MRI transverse slice through the liver and spleen. 59
Figure 5. *MRCP*+ analysis.. 62
Figure 6. Region of interests (ROIs)-placement method.. 64
Figure 7. Liver segmentation method. 66
Figure 8. Superpixel analysis of a single transverse MRI slice. 68
Figure 9. Domains in the quality of life questionnaires. 73
Figure 10. Bland-Altman plots in 6 patients with PSC that showed a difference between two MRCP acquisition protocols. 87
Figure 11. Bland-Altman plot for biliary tree length modelled between the first and second MRCP acquisitions. 91
Figure 12. Bland-Altman plots of intra-observer variability in whole tree and individual duct MRCP+ metrics in 10 healthy volunteers. 92
Figure 13. Schematic diagram of the biliary tree divided into extrahepatic, hilar, and intrahepatic segments. 95
Figure 14. Mean ± SEM visualisation score comparison between MRCP and *MRCP*+ in patients with PSC and healthy volunteers. 96
Figure 15. Inter-observer and intra-observer variability in biliary tree length and number of bile ducts modelled by *MRCP*+ in patients with large-duct PSC..... 99
Figure 16. Mean liver cT1 measurements in healthy volunteers before and after the consumption of pineapple juice..... 103
Figure 17. Comparisons between T1 measurements with and without iron-correction in patients with PSC and healthy controls. 105
Figure 18. Variability in cT1 between manual ROI-placement and semi-automated multi-slice liver segmentation methods. 108
Figure 19. Variability in cT1 between multi-slice semi-automated and single-slice automated liver segmentation methods. 109
Figure 20. Biliary tree model demonstrating candidate strictures (red) and dilatations (blue) computed by *MRCP*+ analysis. 119
Figure 21. Study flow chart for *MRCP*+ analysis. 122
Figure 22. Maximum intensity projection (MIP) image and the corresponding 3D parametric biliary tree model..... 126
Figure 23. Intrahepatic sum of relative severity of dilatations (SumRelSevDilat) in high-risk vs low-risk groups in patients with large-duct PSC stratified by Mayo Risk Score (MRS) and Amsterdam-Oxford Model score (AOM). 145
Figure 24. Intrahepatic sum of relative severity of dilatations (SumRelSevDilat) in high-risk vs low-risk groups in patients with large-duct PSC stratified by 2.2x and 1.5x the upper limit of normal serum ALP. 146
Figure 25. Intrahepatic sum of relative severity of dilatations (SumRelSevDilat) in high-risk vs low-risk groups in patients with large-duct PSC stratified by transient elastography liver stiffness (LS) and enhanced liver fibrosis score (ELF). 147
Figure 26. Intrahepatic sum of relative severity of dilatations (SumRelSevDilat) in high-risk vs low-risk groups in patients with large-duct PSC stratified by modified Amsterdam intrahepatic stricture severity score (ISSS) and Anali score..... 148

Figure 27. Intrahepatic sum of relative severity of dilatations (SumRelSevDilat) in high-risk vs low-risk groups in patients with large-duct PSC stratified by the presence of extrahepatic disease and dominant stricture. 149
Figure 28. ROC curves summarising the individual high-risk vs low-risk discriminatory performance of non-invasive risk stratification tools and the multivariable logistic regression models for identifying the high-risk PSC group. 152
Figure 29. Summary of SF-36 mean domain and summary scores in patients with PSC. 157
Figure 30. Summary of PSC-PRO mean domain scores in patients with PSC..... 157
Figure 31. Schematic diagram explanation for skewness and kurtosis of a distribution. 175
Figure 32. Study flow diagram. 178
Figure 33. Correlation between IQR cT1 and mean superpixel variance cT1 with transient elastography liver stiffness in patients with PSC. 186
Figure 34. Comparisons of IQR cT1 and mean superpixel variance cT1 in patients with PSC stratified into high-risk and low-risk groups by transient elastography liver stiffness (LS) and enhanced liver fibrosis score (ELF) thresholds. 188
Figure 35. ROC curve of IQR cT1 for detecting advanced fibrosis (≥F3) defined by transient elastography liver stiffness measurement > 9.6kPa in patients with large-duct PSC. 188
Figure 36. Comparisons of IQR cT1 and mean superpixel variance cT1 in patients with PSC stratified into high-risk and low-risk groups by 2.2x and 1.5x upper limit of normal serum ALP thresholds. 189
Figure 37. Comparisons of IQR cT1 and mean superpixel variance cT1 in patients with PSC stratified into high-risk and low-risk groups by Amsterdam-Oxford Model score (AOM) and Mayo Risk Score (MRS) thresholds. 191
Figure 38. Comparisons of IQR cT1 and mean superpixel variance cT1 in patients with PSC stratified into high-risk and low-risk groups by Anali score threshold and the presence of hepatic dysmorphy on MRI. 194
Figure 39. Comparisons of IQR cT1 and mean superpixel variance cT1 in patients with PSC stratified into high-risk and low-risk groups by the presence of portal hypertension and cirrhosis on MRI. 195
Figure 40. Comparisons of IQR cT1 and mean superpixel variance cT1 in patients with PSC stratified into high-risk and low-risk groups by modified Amsterdam intrahepatic stricture severity score (ISSS) threshold and the presence of dominant stricture. 199
Figure 41. ROC curves summarising the individual discriminatory performance of non-invasive risk stratification tools and the multivariable logistic regression models for identifying the high-risk PSC group.. 202
Figure 42. Breakdown of the pooled participants included in the validation cohort from the RIAL, NICOLA and MAPLE studies. 210
Figure 43. IQR cT1 in chronic liver diseases and healthy volunteers. 212
Figure 44. Mean superpixel variance cT1 in chronic liver diseases and healthy volunteers. 213
Figure 45. IQR cT1 and mean superpixel variance cT1 stratified by Ishak's fibrosis stages in patients with mixed aetiology chronic liver diseases. 215
Figure 46. Median cT1 stratified by Ishak's fibrosis stages in patients with mixed aetiology chronic liver diseases. 215
Figure 47. IQR cT1 stratified by Ishak's fibrosis stages in patients with mixed aetiology chronic liver diseases. 217
Figure 48. Median cT1 stratified by Ishak's fibrosis stages in patients with mixed aetiology chronic liver diseases. 217
Figure 49. Mean superpixel variance cT1 stratified by Ishak's fibrosis stages in patients with mixed aetiology chronic liver diseases. 218

Figure 50. Difference in liver cT1 metrics between compensated and decompensated cirrhosis groups in the MAPLE study. 219
Figure 51. Histopathology of periductal fibrosis. 229
Figure 52. Co-registration of two quantitative MRI techniques for periductal cT1 quantification in a patient with large-duct PSC. 233
Figure 53. Periductal distance map from distance transform. 234
Figure 54. Schematic diagram representing the boundaries of the five regions of interest for periductal cT1 quantification. 235
Figure 55. A schematic diagram representing a cross section of a bile duct intersecting with an axial segmented liver cT1 slice. 236
Figure 56. A schematic diagram representing partial volume effect exerted by a single bile duct intersecting an axial segmented liver cT1 slice at different angles. 237
Figure 57. Study flow diagram. 240
Figure 58. Mean ± SEM cT1 in periductal region of interests (ROIs) at fixed radial distances from the bile ducts in comparison to each other and the whole liver in healthy controls in the MoBILISE study. 243
Figure 59. Mean ± SEM cT1 in periductal region of interests (ROIs) at fixed radial distances from the bile ducts in comparison to each other and the whole liver in the R&R study validation cohort of healthy volunteers. 243
Figure 60. Mean ± SEM cT1 in periductal region of interests (ROIs) at fixed radial distances from the bile ducts in comparison to each other and the whole liver in patients with PSC. 244
Figure 61. Mean ± SEM cT1 comparisons between ROI 1 (1.5–5.5mm), ROI 2 (5.5–9.5mm), and ROI 3 (whole liver) stratified according to angle of intersection of bile ducts with the slice normal. 246
Figure 62. Correlation between transient elastography liver stiffness and periductal mean cT1 in ROI (1.5-3.5mm) in patients with PSC. 247
Figure 63. Receiver operator curve for the identification of patients with advanced fibrosis defined as transient elastography liver stiffness greater than 9.6kPa using periductal mean cT1 1.5–3.5mm. 248
Figure 64. Receiver operator curve to detect patients with high-risk for death and/or liver transplantation stratified by Amsterdam-Oxford Model score > 2 using periductal mean cT1 1.5–3.5mm. 249
Figure 65. Summary of 2 x 2 mixed-effects ANOVA. 262
Figure 66. Study flow chart for follow-up visit. 263
Figure 67. Differences in mean superpixel variance cT1 between study visits 1 and 2 in high-risk and low-risk group stratified by non-invasive surrogate markers. 266
Figure 68. Differences in mean intrahepatic SumRelSevDilat between study visits 1 and 2 according to high-risk or low-risk group stratified by non-invasive surrogate markers. 268
Figure 69. Differences in mean IQR cT1 between study visits 1 and 2 according to high-risk or low-risk group stratified by non-invasive surrogate markers. 271
Figure 70. Differences in mean superpixel variance cT1 between study visits 1 and 2 according to high-risk or low-risk group stratified by non-invasive surrogate markers. 274
Figure 71. Differences in mean periductal cT1 1.5-3.5mm between study visits 1 and 2 according to high-risk or low-risk group stratified by non-invasive surrogate markers. 277
Figure 72. Differences in mean domain scores of SF-36 and PSC-PRO between visits 1 and 2. 284
Figure 73. Changes in the mean score of physical functioning domain of SF-36 and emotional impact domain of PSC-PRO between the two study visits stratified by fibrosis severity. 285

Thesis abstract

Primary sclerosing cholangitis (PSC) affects the biliary tree and liver. Magnetic resonance imaging (MRI) with cholangiopancreatography (MRCP) is used to diagnose and follow-up patients. Its current use and applications, however, are not standardised and relies on qualitative assessments susceptible to inter-observer variability. The aim of this thesis was to identify standardised and objective imaging metrics derived from quantitative MRI techniques that could potentially be used in future clinical assessment of patients with PSC, longitudinal studies, and interventional drug trials.

The quantitative techniques used in this thesis include (i) *MRCP+*™, a three-dimensional (3D) biliary tree model computed from MRCP data to derive biliary metrics; (ii) T1 and T2* mapping to compute the iron-corrected T1 (cT1) using Liver*MultiScan*™; and (iii) co-registration of *MRCP+* biliary tree model with Liver*MultiScan* cT1 maps to calculate periductal cT1. Patients with PSC were categorised into either high-risk or low-risk group for disease progression based on published thresholds of revised Mayo Risk Score (MRS), Amsterdam-Oxford Model (AOM), upper limit of normal (ULN) alkaline phosphatase (ALP), transient elastography liver stiffness (LS), enhanced liver fibrosis score (ELF), modified Amsterdam intrahepatic stricture severity score (ISSS), Anali score, presence of extrahepatic disease, and presence of dominant stricture. Patient reported outcomes were collected using generic SF-36 and disease-specific PSC-PRO questionnaires.

The intrahepatic sum of the relative severity of dilatations (SumRelSevDilat) was consistently higher in the high-risk compared to the low-risk PSC group (MRS: $p<0.0001$; AOM: $p=0.0017$; 2.2x ULN ALP: $p=0.0007$; 1.5x ULN ALP: $p=0.0025$; LS: $p=0.0022$; ELF: $p=0.0189$; ISSS: $p<0.0001$; Anali: $p=0.0299$). It identified the high-risk MRS > 0 group (AUC=0.77, 95% CI 0.66–0.88) numerically better than other non-invasive markers and was

an independent predictor of this high-risk group (OR: 29.2, p=0.038). SF-36 bodily pain (r_s=-0.26, p=0.025) and PSC-PRO activities of daily living (r_s=0.29, p=0.014) showed the strongest correlations with intrahepatic SumRelSevDilat.

Two liver cT1 metrics, interquartile range of cT1 (IQR cT1) and mean superpixel variance cT1, were identified as potential markers of liver parenchymal heterogeneity in PSC. IQR cT1 had the best performance to detect the high-risk PSC group stratified by LS (AUC=0.84, 95% CI 0.75–0.93, p<0.0001). Mean superpixel variance cT1 differentiated PSC from every other liver disease (p<0.0001) but did not differentiate between the other liver diseases. Median cT1 correlated with all the domains of SF-36 with the strongest being mental health (r_s=-0.37, p=0.001) and social functioning: r_s=-0.32, p=0.005).

The periductal mean cT1 1.5–3.5mm surrounding the bile ducts were higher than regions further away from the bile duct (p<0.0001) in PSC but not in healthy controls. Patients with LS > 9.6kPa (AUC=0.74, 95% CI 0.60–0.88, p=0.0021) and AOM > 2 (AUC=0.72, 95% CI 0.59–0.85, p=0.0031) had higher values than those with LS ≤ 9.6kPa and AOM ≤ 2, respectively. The mental health (r_s=0.38, p=0.002) and role emotional (r_s=0.37, p=0.010) components of SF-36 showed the strongest correlations with periductal mean cT1.

Temporal changes in mean values of IQR cT1 showed a significant interaction with time and baseline risk categorisation by AOM (p<0.001), LS (p=0.010), ELF (p=0.004), ISSS (p=0.039) and Anali (p<0.001). The diagnostic performances of single baseline measurements were numerically better than temporal changes to detect the high-risk group.

Validation against clinical endpoints is warranted before these metrics can be tested in multicentre prospective studies and considered as surrogate endpoints in clinical trials.

1 Introduction[1]

1.1 Epidemiology of primary sclerosing cholangitis

Primary sclerosing cholangitis (PSC) is a rare, chronic, immune-mediated, cholestatic liver disease. It is characterised by intrahepatic and extrahepatic bile duct inflammation with development of multifocal biliary strictures, biliary fibrosis and chronic cholestasis, often leading to biliary cirrhosis and end-stage liver disease[1]. It has a male preponderance, with a mean age at diagnosis of 40 years, and a strong association with concomitant inflammatory bowel disease (IBD)[2]. In a recent systematic review of 17 population-based studies, the incidence ranged from 0 to 1.6 per 100,000 persons annually, and the prevalence ranged from 6 to 7.4 per 100,000 persons[3].

1.2 Diagnosis of PSC

A clinical definition for PSC was based upon three landmark papers in the 1980s from the United States (US), United Kingdom (UK) and Norway[4-6]. Subsequently, three sub-types of PSC were described. 'Classical' large-duct PSC (LD-PSC), accounting for 90% of patients, involves either the intrahepatic bile ducts, extra hepatic bile ducts, or both. It is usually diagnosed based on characteristic changes in the bile ducts on cholangiography with or without cholestatic liver biochemistry. Magnetic resonance imaging with magnetic resonance cholangiopancreatography (MRI-MRCP) is the standard imaging modality to confirm a diagnosis of LD-PSC[7]. 'Small duct' PSC that has normal cholangiography but affects only the small intrahepatic bile ducts on liver histology, accounts for 5% of patients[8]. PSC with 'autoimmune hepatitis (AIH) overlap', confirmed histologically in those with elevated transaminases and/or immunoglobulin G levels and an abnormal cholangiogram, presents in 5% of patients[9]. Lastly, 'PSC with high immunoglobulin G subclass 4 (IgG4) levels' in the

[1] The contents of this introduction chapter have been previously published in my review article Selvaraj EA, Culver EL, Bungay H et al. Evolving role of magnetic resonance techniques in primary sclerosing cholangitis. World J Gastroenterol 2019; 25:644-658.

serum and/or tissue, has been reported in 10-18% of LD-PSC patients, with mixed evidence of whether it is a distinct clinical phenotype and natural course [10-12].

PSC is insidious, with nearly half of patients being asymptomatic at diagnosis, identified after the investigation of abnormal liver biochemistry[13]. When present, symptoms can be non-specific and may include abdominal pain, pruritus, and fatigue[14]. Cholangiography is required to make a diagnosis of LD-PSC. Endoscopic retrograde cholangiopancreatography (ERCP) is invasive with potential serious complications including pancreatitis, cholangitis, perforation, and bleeding[15]. Over the last twenty years, MRCP has replaced ERCP as the first line imaging method for the diagnosis of PSC. A meta-analysis that included 189 patients with PSC compared the diagnostic accuracy of MRCP against combined clinical, biochemical, and ERCP or PTC endpoint as the reference standard for establishing a diagnosis of PSC. The study concluded that MRCP has high diagnostic accuracy for the diagnosis of PSC with an area under the curve (AUC) of 0.91, sensitivity 86% (95% confidence interval (CI): 80-90%), and specificity 94% (95% CI: 86–98%)[16]. Advancement in imaging techniques with higher quality images and spatial resolution since that study are likely to have increased the diagnostic accuracy further. Both the European Association for the Study of the Liver (EASL) and American Association for the Study of Liver Diseases (AASLD) recommend MRCP as the first-choice imaging modality in PSC[7, 17]. Performing MRCP first — before ERCP — has been shown to be a more cost-effective strategy[18, 19]. Moreover, incomplete biliary tract distension mimicking the ductal irregularities of PSC can give rise to false-positive diagnosis on ERCP cholangiogram and false-negative diagnosis if a high-grade stricture causes inadequate opacification of the intrahepatic ducts[20, 21]. Nevertheless, ERCP is still performed when diagnostic doubt exists after MRCP.

1.3 MRI-MRCP features in PSC

Whilst MRI of the liver was being performed for liver disease, it was not until 1991 that the first “MR cholangiography” sequence was performed by Wallner et al.[22]. They developed a T2-weighted rapid sequential gradient-echo two-dimensional (2D) acquisition and a three-dimensional (3D) post-processing technique to produce coronal and sagittal images, without the need for ionising radiation or intravenous biliary contrast. MRCP has undergone significant advancement since its first description.

Identification of multifocal fibrotic strictures and areas of dilatation and ductal wall thickening of the intrahepatic or extrahepatic biliary systems, or both, underpins the diagnosis of LD-PSC. Most patients have involvement of both the intra- and extrahepatic bile ducts, with less than 25% with intrahepatic disease only[17]. However, more recent population-based data from UK show that only 33-48% of patients have extrahepatic ductal involvement[23]. Exclusive involvement of the extrahepatic duct is uncommon (less than 5%) and should prompt a search for an alternative cause[1]. Figure 1 illustrates typical MRCP features of LD-PSC. The MRI-MRCP features of PSC that have been reported in the literature are summarised in Table 1.

As for any other qualitative imaging modality, MRI-MRCP is subject to inter- and intra-observer variability. The cholangiographic features of PSC alone do not necessarily distinguish PSC from secondary sclerosing cholangitis, particularly in the absence of IBD diagnosis. Immunoglobulin G4-related sclerosing cholangitis (IgG4-SC) can often mimic PSC on MRCP. Features that support a diagnosis of IgG4-SC over PSC include dilatation proximal to confluent stricture, symmetrical bile duct wall thickening with smoother outer and inner margins, presence of continuous as opposed to skip disease in the bile ducts, common bile duct thickness greater than 2.5mm, gallbladder and pancreatic involvement[24, 25]. Other mimickers include ischaemic cholangiopathy, AIDS-related cholangiopathy, secondary sclerosing cholangitis after repeated ascending cholangitis and portal biliopathy.

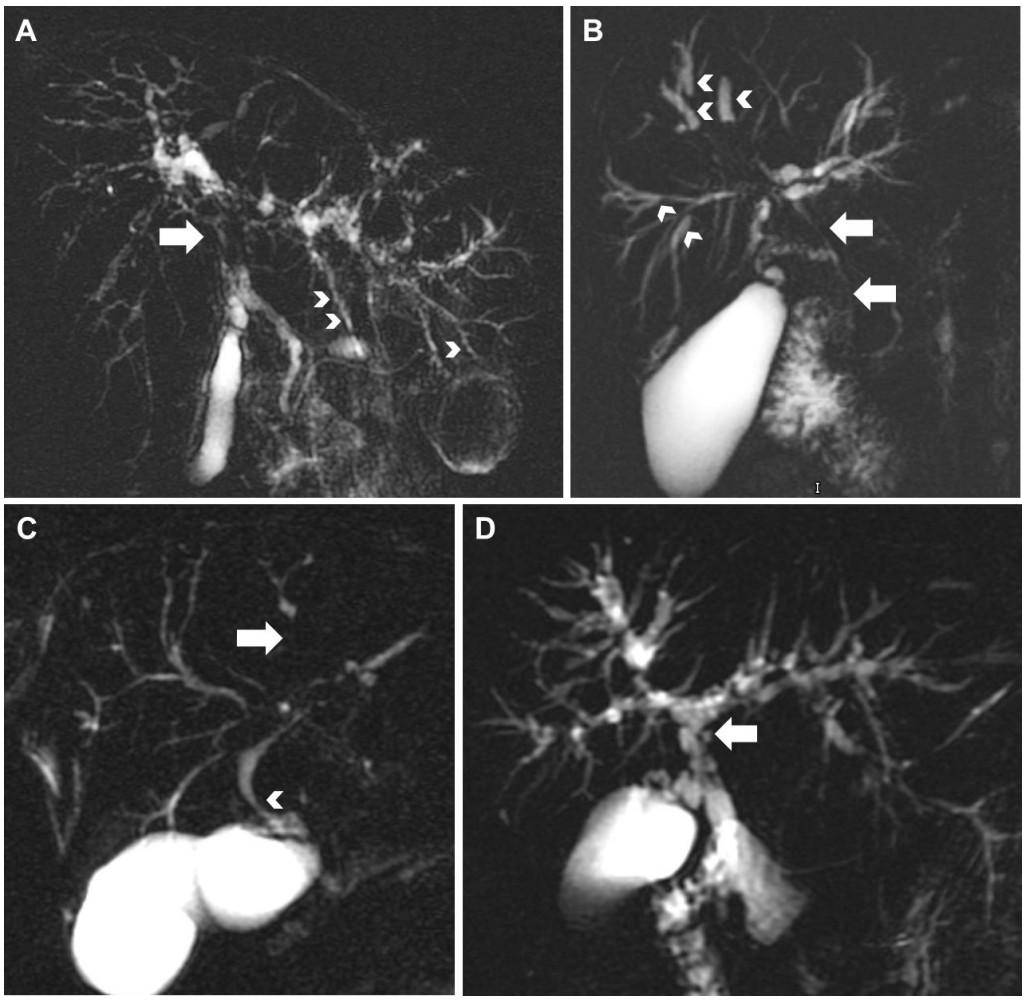


Figure 1. Typical features of large-duct PSC on MRCP maximum intensity projection (MIP) images. A) Common hepatic duct stricture (arrow) and intrahepatic duct beading (arrowheads); B) Dominant stricture (arrow) and dilated proximal intrahepatic ducts secondary to distal strictures (arrowheads); C) Area of non-filling of intrahepatic ducts indicating tight stricture (arrow) and common hepatic duct stricture (arrowhead); D) Pseudodiverticula.

Reprinted with copyright permission from Selvaraj EA, Culver EL, Bungay H et al. Evolving role of magnetic resonance techniques in primary sclerosing cholangitis. *World J Gastroenterol* 2019; 25:644-658. © 2019 Baishideng Publishing Group Inc.

Table 1. Morphological features of PSC on MRI-MRCP

	Description[26-29].
Bile duct	• Multiple annular or short segmental strictures (1-2mm) with slightly dilated ducts among them: “beaded” appearance • Obliteration of small peripheral ducts “pruned tree” • Periductal inflammation • Thickening of walls of large ducts • Strictures seen at bile duct bifurcation • Angles between peripheral and central bile ducts become obtuse • Exclusive involvement of extrahepatic bile duct is infrequent • Bile duct dilatations are usually subtle • Retraction of papilla • Webs, diverticula, and pigmented stones
Liver parenchyma	• Segmental or lobular atrophy with compensatory hypertrophy attributed to chronic biliary obstruction • Patchy areas of peripheral parenchymal enhancement • Caudate lobe hypertrophy† • Spherical liver shape‡ • Peripheral wedge-shaped areas with focal increased signal intensity on T2-weighted images • T2-weighted hyperintensity around portal vein branches
Regional	• Enlarged regional lymph nodes • Signs of portal hypertension including splenomegaly and collateral vessels • Gallbladder enlargement

†Autoimmune process spares bile ducts in caudate lobe that results in compensatory hypertrophy. This feature is also seen in other cirrhotic livers because the caudate lobe has its own venous, lymphatic, and biliary drainage.
‡This is due to atrophy of left lateral segments and posterior segments of the right lobe.

1.4 Complications of PSC

Disease progression in PSC can affect the biliary tree and/or the liver. A third domain of disease complications or progression includes extra-hepatic manifestations and symptoms, which can be independent of the stage of the biliary or liver disease. The complications of disease progression in PSC are summarised in Figure 2.

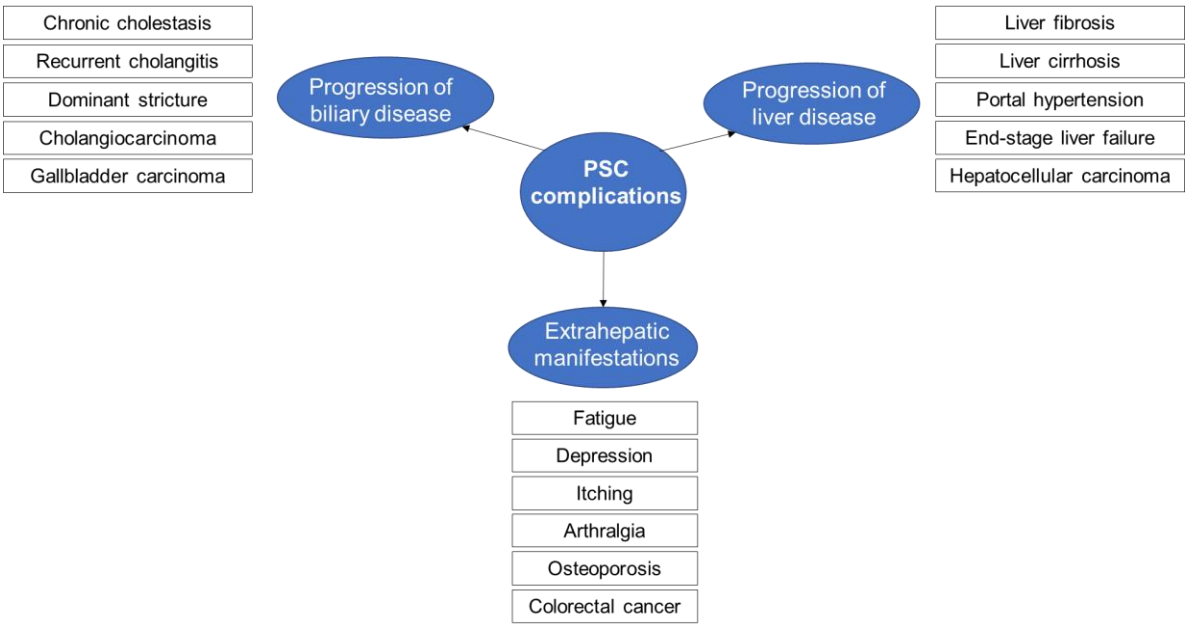


Figure 2. Summary of complications resulting from disease progression in PSC.

1.4.1 Biliary complications

Progressive biliary inflammation and strictures lead to chronic cholestasis and predisposes patients to recurrent cholangitis. A stricture less than 1.5 mm diameter in the common bile duct, and/or less than 1 mm in the left or right main hepatic ducts, within 2cm of the bifurcation on cholangiography, has traditionally been defined as a dominant stricture[30]. A more pragmatic definition has been proposed in the UK guidelines defining a dominant stricture as functional stenoses of the major bile ducts with signs of biliary obstruction shown by worsening liver biochemistry and/or proximal biliary dilatation or symptoms of cholestasis[31]. The presence of a dominant stricture either at diagnosis or follow-up is associated with an increased risk of developing cholangiocarcinoma and mortality[32, 33].

Deterioration in clinical and biochemical parameters prompts evaluation for a dominant stricture and/or cholangiocarcinoma on MRI-MRCP. However, there are currently no validated criteria for MR definition of a dominant stricture given the spatial resolution of MRI in extrahepatic bile ducts is lower than ERCP and MRI lacks the filling pressure of bile ducts seen at ERCP.

Cholangiocarcinoma is one of the feared complications of PSC. Patients with PSC have an increased risk of death and hepatopancreatobiliary malignancy, particularly cholangiocarcinoma, when compared to patients with IBD alone and the general population[34-36]. In one UK study, the incidence of cholangiocarcinoma was estimated at 8.1 per 1000 person-years[35]. Over one third of cholangiocarcinoma cases were detected within the first year following PSC diagnosis in a retrospective, international, observational cohort study involving 594 PSC patients[37]. This is likely due to length-time bias and detection of cholangiocarcinoma only when it becomes clinically overt. Serum tumour marker carbohydrate antigen (CA)19-9 is widely used in surveillance strategy, but it lacks both sensitivity and specificity for the detection of cholangiocarcinoma[38, 39]. The cholangiocarcinoma is usually too advanced for curable treatment by the time CA19-9 becomes persistently elevated[33]. Annual MRI-MRCP as a surveillance strategy is often performed[40]. In a large retrospective observational study, patients who underwent imaging surveillance were more likely to receive liver transplantation as a curative treatment and had better survival than those who did not have surveillance[41]. However, recommendations for surveillance between the UK and USA guidelines are mixed[17, 31, 42]. There is an unmet need for earlier detection of cholangiocarcinoma and closer monitoring of newly diagnosed patients.

1.4.2 Liver complications

Progressive multifocal biliary inflammation and fibrosis leads to accumulation of fibrosis in the liver parenchyma of patients. Liver cirrhosis and portal hypertension ensues leading to end-stage liver disease and its associated complications including ascites, hepatic

encephalopathy, variceal bleeding, and hepatocellular carcinoma as with other chronic liver diseases. In one UK study, the crude incidence of developing cirrhosis over time was estimated at 18.6, portal hypertension 14.5, and liver transplantation 10.9 per 1000 person-years[35]. Unlike in primary biliary cholangitis (PBC), data for the use of ursodeoxycholic acid (UDCA) in PSC remains inconclusive due to limitations in trial designs. Low-dose UDCA has been shown to improve levels of serum alkaline phosphatase (ALP) and surrogate markers of prognosis but is yet to demonstrate a survival benefit[43].

1.4.3 Extrahepatic complications

A third domain of disease complications includes extra-hepatic manifestations and symptoms which can be independent of the stage of the biliary or liver disease. Patients with PSC and ulcerative colitis (UC) have a 9-fold increase in developing colorectal cancer compared to matched general population and 10-fold increase compared to UC controls[2]. Gallbladder complications include dilatation, cholecystitis, gallstones, polyp, and cancer.

1.4.4 Survival

In the absence of effective medical therapies to date, the natural history of PSC is progressive. Liver transplantation is the only proven life-extending intervention with transplant indication for PSC rising in several European countries[3]. Transplant-free survival in patients with PSC depends on whether data was collected in transplant or non-transplant centres. Median survival ranged between 9.6–13.2 years in transplant centres and 18.0–21.3 years in non-transplant centres[2, 44-47]. In a UK study, the 5-year and 10-year survival was 75% and 64%, respectively[35]. PSC accounts for 10-15% of all liver transplant activity in Europe and the median transplant-free survival in the most recent multicentre tertiary hospitals and transplant units within the International PSC Study Group in 7121 patients with PSC was 14.5 years[37].

The survival of patients also depends on the PSC disease subtype, presence, and phenotype of IBD. Patients with small-duct PSC have longer transplant free survival

compared to patients with large-duct PSC and tend not to develop cholangiocarcinoma unless progression to large-duct PSC has occurred[2, 8, 37, 48]. Survival of patients with PSC-AIH overlap seems to be similar to large-duct PSC[2, 37], although their risk of hepatobiliary malignancies appears to be lower[49]. Survival may also differ within the same PSC disease subtype, adding to the heterogeneity and complexity of disease progression in PSC. Patients with PSC and concomitant IBD have an increased risk of cancers and death. Age at diagnosis, sex and race have also recently been identified as a risk factor for PSC-related death and incidence of liver transplantation[36]. A diagnosis of UC is associated with a greater risk of developing hepatopancreatobiliary malignancy and lower transplant-free survival than Crohn's irrespective of the influence exerted by PSC phenotype and sex[37].

1.4.5 Patient experience

Development of non-invasive methods to risk-stratify patients with PSC and predict clinical events was identified as a research priority in a 2016 survey by PSC Support, a registered UK charity[14]. It highlighted that patients experience significant anxiety due to the uncertainty about the future of their disease, in particular the risk of disease progression, malignancy, and liver transplantation. A recent systematic review[50] that included 28 heterogeneous studies found a poorer quality of life (QoL) for people with PSC compared to healthy controls, similar QoL to other chronic conditions, and mixed results compared to the general population. The factors that were associated with impaired QoL were itch, pain, jaundice, severity of IBD, liver cirrhosis, and large-duct PSC. The review also found that surrogate markers of disease progression or prognostic risk scores do not reflect the experiences of patients[50].

1.5 Disease severity and prognosis

Predicting which patients are at risk of developing disease progression and complications is challenging in clinical practice given the heterogeneity in the natural course of the disease between patients. In the absence of effective medical therapies, there is a need to develop risk-stratification methods to identify surrogate markers of disease severity, progression, and

clinical endpoints, both for clinical use and interventional trials. Development of clinical outcomes in PSC is slow. Therefore, several prognostic surrogate biomarkers have been proposed to date and are discussed below. A robust surrogate endpoint is one that measures a true clinical efficacy (level 1) or is a validated surrogate endpoint (level 2) according to Fleming's hierarchy of endpoints[51]. However, none of the current surrogate biomarkers in PSC has achieved Level 2 yet[52].

1.5.1 Serum markers

Serum ALP has been the most widely studied serum biomarker in PSC. ALP levels can fluctuate throughout the disease course, but persistently low or normalisation of ALP has been shown to correlate with better clinical outcomes[53]. A cut-off value of 1.5x the upper limit of normal (ULN) has been demonstrated in several patient cohorts to have good prognostic implications[54-58]. However, patients can have normal serum ALP with advanced liver disease[59]. In one study with 10-year follow-up, 62% of patients did not experience any liver-related endpoints despite having a serum ALP that did not improve to levels less than 1.5x ULN[58]. In the UK-PSC Consortium retrospective study involving 1001 patients with PSC, higher thresholds of ALP at 2.2x and 2.4x ULN have been suggested to have better predictive ability of liver transplantation and survival[23]. Despite this, a more recent prospective study analysed data from 234 patients with large-duct PSC in a phase II Simtuzumab trial (NCT01672853) and reported large variations in serum ALP within and among patients. Serum ALP was not associated with disease progression, albeit only over a 2-year follow-up period[60].

The enhanced liver fibrosis score (ELF®) is based on three direct markers of fibrogenesis: hyaluronic acid, tissue inhibitor of metalloproteinases-1 (TIMP-1) and procollagen III N-terminal peptide (PIIINP). In a retrospective study of 167 derivation and 138 validation cohorts of patients with PSC, ELF was reported to be a strong predictor of transplant-free survival with AUC of 0.81 (95% CI: 0.73–0.87) and was independent of other risk factors or prognostic scores that predicted outcomes[61]. This was subsequently validated in an external,

multicentre, retrospective cohort of 534 patients with PSC[62]. ELF showed borderline significance for the prediction of PSC-related liver events in the Simtuzumab clinical trial (hazard ratio 1.26; 95% CI 0.98−1.61, p = 0.073)[63]. The C-statistic for ELF 0.71 (95% CI: 0.63−0.79) was similar to Ishak fibrosis stage 0.69 (95% CI: 0.61−0.77) to predict clinical events in the trial. Higher ELF scores were seen in patients with cholangiocarcinoma[64]. However, it is worth noting that ELF can also be elevated in acute inflammation and biliary obstruction.

Total bilirubin, albumin, and prothrombin time have been incorporated into several PSC-specific prognostic risk models[65-68]. Persistently elevated bilirubin level defined as greater than three months was reported in one study to have a median survival of 30 months[47]. Measurements of bilirubin, however, may be confounded by occurrence of cholangitis or biliary obstruction in PSC. Gamma-glutamyltranspeptidase (GGT) levels have not been widely studied in adults but has been shown to have favourable outcomes in paediatric PSC[69]. Fibrosis-4 (FIB-4) and aspartate aminotransferase (AST)-to-platelet ratio index (APRI) scores have been studied in other chronic liver diseases as a marker of liver fibrosis, but their roles in PSC have not been reported to date.

1.5.2 Prognostic risk models

Several clinical risk scores have been developed to predict disease progression and clinical outcomes in PSC. The revised Mayo Risk Score (MRS), based on age, bilirubin, albumin, AST, and variceal bleeding, is the most used clinical risk model[68]. The model was developed using multi-centre data of 405 patients and was subsequently validated in a separate cohort of 124 patients. The risk score provides survival estimates up to four years follow-up but does not include other important outcomes such as liver transplantation, decompensation, or development of cholangiocarcinoma. Given that it is made up of markers predictive of advanced disease, it probably has insufficient power and not clinically useful in discriminating and predicting the clinical course of early disease. However, it remains an important model to which newer scores are compared to in clinical studies.

The Amsterdam-Oxford Model (AOM) score based on seven variables (PSC subtype, age at diagnosis, albumin, platelet, AST, ALP and bilirubin) predicted liver-related death or liver transplantation in a derivation cohort of 692 patients with a median follow-up of 9.2 years, and was externally validated in 264 patients with a median follow-up of 8.6 years[65]. These were predominantly population-based cohorts. The C-statistic of the model at diagnosis was moderate at 0.68 (95% CI: 0.51–0.85) and was similar when recalculated using data collected during the first three years of follow-up for the overall discriminatory performance. Predicted rate of events by the model were similar to actuarial survival estimated by Kaplan-Meier analysis at 5-year, 10-year, and 15-year probabilities. A subsequent validation study involving 534 patients in three liver transplant centres showed similar accuracy when the score was re-calculated at different time points during follow-up[70]. The C-statistic ranged from 0.67 at diagnosis to 0.75 at five years. This study also provided a direct comparison of AOM with MRS and ALP. MRS exhibited a greater discriminatory value at baseline (C-statistic: 0.73) and at five years (C-statistic: 0.79). However, the risk of future events was overestimated by MRS. The C-statistic of ALP was marginal at 0.52–0.63 during the first five years since diagnosis for the liver transplant-free survival prediction.

The PSC risk estimate tool (PREsTo) was recently developed using machine learning techniques. It consists of nine variables (bilirubin, albumin, ALP xULN, platelet, AST, haemoglobin, sodium, age, and number of years since diagnosis)[67]. The model was derived using 509 patients and validated in an international multicentre cohort of 278 patients who did not have markers of advanced disease. It accurately predicted the 5-year risk of liver decompensation with a C-statistic of 0.90 (95% CI: 0.85–0.99) and performed better than MRS and Model for End-stage Liver Disease (MELD) scores. The derivation formula is yet to be published and it cannot be used for patients who have portal hypertension or platelet count less than 150 x 10^9/L at baseline.

The latest prognostic model to be published is the UK-PSC score[23] derived from 1001 patients. Data was collected retrospectively from 101 non-transplant and 7 transplant centres in the UK. The 2-year and 10-year outcome prediction of clinical variables were assessed. Serum ALP > 2.4x ULN at 1 year after diagnosis and the presence of extrahepatic disease on MRCP at diagnosis predicted the 10-year outcome. The final UK-PSC score made up of six variables (age, bilirubin, ALP, albumin, platelets, presence of extrahepatic biliary disease, and variceal haemorrhage), predicted the 2-year (C-statistic: 0.81) and 10-year (C-statistic: 0.80) outcome of all-cause mortality and liver transplantation with good discrimination. The score was validated in two external cohorts of 352 patients from the UK and 99 patients from an international cohort. It outperformed MRS and APRI scores. This was the first prognostic model to incorporate a radiological feature. However, it relies on obtaining accurate variables at diagnosis and 2-years after diagnosis, which may not be useful at different time points during follow-up.

1.5.3 Liver histology

Liver biopsy, assessed using the Ludwig staging system, has been shown to be an independent predictor of survival in PSC[71]. In a study of 64 patients with PSC, the prognostic value of three separate histological scoring systems (Nakanuma, Ishak and Ludwig) were shown to be strongly associated with transplant-free survival and the time to liver transplant, but not with liver decompensation[72]. The Nakanuma scoring system had the highest hazard ratios amongst the three. All three classification systems were only weakly correlated with biochemical parameters. The findings were validated in a multicentre study involving 119 patients with PSC[73]. However, liver biopsy is not recommended in the current EASL and AASLD guidelines for the diagnosis and follow-up of PSC due to its invasive nature and risk of complications[7, 17]. Moreover, distribution of disease in PSC is patchy and liver biopsy is prone to sampling variability[74]. A liver biopsy is usually performed when MRCP is normal to diagnose small-duct PSC and/or there is suspicion of an autoimmune overlap syndrome or IgG4-SC.

1.5.4 Transient elastography

Liver stiffness has been shown to be a surrogate marker for liver fibrosis in chronic liver disease[75, 76]. Liver stiffness can be measured using shear-wave-based technology such as vibration controlled transient elastography (VCTE) or magnetic resonance elastography (MRE). Two studies involving 28 and 73 patients with PSC demonstrated independent association between liver stiffness measurements and stages of liver fibrosis assessed by the METAVIR system[75, 76]. Liver stiffness showed high diagnostic accuracy for detecting severe fibrosis and cirrhosis (F3–4). In 168 patients with PSC (n = 141 large-duct), both baseline and changes in liver stiffness values correlated with adverse clinical outcomes (death, liver transplantation, decompensation, hepatocellular carcinoma, or cholangiocarcinoma) over a mean follow-up period of 4 years[76]. Bilirubin, baseline liver stiffness, and change in liver stiffness were all independently associated with survival. Another study replicated the diagnostic accuracy of VCTE in a cohort of 62 patients with PSC and its prognostic value was confirmed in a multicentre cohort of 130 patients[77]. Although widely available with relatively low-cost, false positive elevation of VCTE measurements can be caused by biliary obstruction, elevated total bilirubin and/or transaminases levels, and presence of acute cholangitis that can intermittently occur in PSC[78, 79]. Technical failures and unreliable results have been reported to be as high as 10% in PSC patients[76]. More recently, liver stiffness measurements obtained from other ultrasound-based shear-wave elastography techniques have been shown to correlate with ELF[80].

1.5.5 ERCP

Craig et al.[81] retrospectively reviewed the first available cholangiograms of a cohort of 129 patients with relatively advanced disease using an interpretation model that they had previously developed based on grade, length, and extent of strictures and dilatations[82]. Grade of strictures were defined by visual assessment of percentage narrowing of duct diameter by 25% quartiles into grades 1−4; stricture length classified as band (1−2mm), segmental (3−10mm) or confluent (>10mm); and stricture extent classified as localised

(≤25% of ducts involved) or diffused (>25% of ducts involved). Severity of dilatations were assessed as either none, mild or marked with varying diameter thresholds set for the extrahepatic ducts. The intrahepatic duct dilatations were defined as none (<4mm), mild (4mm) and marked (≥5mm). In that study, both high-grade intrahepatic duct strictures and diffuse intrahepatic duct strictures were associated with a lower 3-year survival. However, the grade, length, and extent of extrahepatic strictures did not correlate with survival. Severity of intrahepatic bile duct dilatations showed a trend towards a significant difference in survival between patients in that study. Similarly, Olsson et al.[83] concluded that high-grade intrahepatic strictures predicted shorter survival in a study involving 94 patients and proposed that the grade of intrahepatic stricture should be incorporated into survival models in PSC.

The Amsterdam cholangiographic classification system was developed by Ponsioen et al.[46], incorporating the previously reported classifications by Majoie et al.[84] and Chen-Goldberg[85]. It assigned scores based on qualitative interpretation of the severity of intrahepatic and extrahepatic strictures and dilatations on cholangiograms. In a large multicentre study with a long follow-up period, cholangiograms of 174 patients were reviewed. Intrahepatic and extrahepatic bile duct findings were combined with age to derive the Amsterdam score. The Amsterdam scores was able to differentiate between different risk groups and were inversely correlated to survival[46]. By combining both intrahepatic and extrahepatic findings, the study has assumed similar prognostic implications of both, despite earlier studies demonstrating only intrahepatic findings conferred prognostic significance. Furthermore, inter- and intra-observer variability was not assessed. The cholangiographic model was subsequently modified and externally validated[86]. Despite the better resolution of image with ERCP, its invasive nature makes staging disease severity and prognosis with ERCP unfeasible.

1.5.6 Qualitative MRI and MRCP scores and features

MRI-MRCP presents a more favourable option than ERCP as a marker of disease activity and prognosis in PSC as it allows co-assessment of liver parenchyma and biliary abnormalities.

Bader et al.[87] investigated the correlation of retrospective single time point analysis of MRI changes in the liver and disease severity in 52 patients with PSC. They examined imaging pattern of the liver parenchymal, presence and grade of intrahepatic biliary ductal dilatation, and presence of areas of parenchymal atrophy or abnormal signal intensity and/or gadolinium enhancement. None of these features showed correlation with Childs-Pugh or MELD scores.

Petrovic et al.[88] retrospectively examined the relationship between the first available MRI-MRCP findings and survival as predicted by the MRS. The severity of biliary stricture was graded using the Amsterdam cholangiographic score proposed by Ponsioen et al[46]. In this study involving 47 patients with PSC, delayed (3 minutes post contrast) peribiliary hyperenhancement in the liver parenchyma using extracellular gadolinium contrast, showed weak correlation with the MRS. There was no correlation with peribiliary hyperenhancement at 2 minutes post contrast, intrahepatic or extrahepatic duct grading of strictures. Extrapolating ERCP-based cholangiographic findings to MR-based cholangiography cast doubts on the reproducibility of stricture grading, particularly with variability in contrast injection technique and volume used during ERCP.

Tenca et al.[89] compared ERCP cholangiogram and MRI-MRCP changes using a modified Amsterdam cholangiographic classification in 48 patients with PSC. Despite the high agreement between ERCP and MRCP in detecting any PSC changes in the intrahepatic and extrahepatic bile ducts, there was only moderate agreement when the modified Amsterdam score was applied to assess severity of disease. The difference was due to an overestimation of the severity of extrahepatic changes in MRI-MRCP compared to ERCP, which may be due to difference in the bile duct physiological state with contrast injection

during ERCP compared to the resting state in MRI-MRCP. Nevertheless, weak correlations were demonstrated between severity of extrahepatic biliary changes and serum ALP as well as clinical endpoints defined by liver transplantation or death.

Whilst delayed phase peribiliary hyperenhancement showed weak correlation with MRS, Ni Mhuircheataigh et al.[90] reported that the presence and extent of early arterial phase peribiliary hyperenhancement on MRI was associated with a higher MRS in a cohort of 62 patients with PSC. The inter-observer agreement for the presence of peribiliary hyperenhancement was good, but only fair for the extent of it. The presence or absence of arterial phase peribiliary hyperenhancement was associated with a significant difference in mean transplant-free survival. Although an overall difference between the three risk groups of MRS was observed, there was no difference between the intermediate- and high-risk groups. They postulated that arterial peribiliary hyperenhancement could potentially be a marker of active biliary inflammation and poorer prognosis.

Change in the morphological appearance of the biliary tree and liver on interval MRI-MRCP is often used to comment on whether the disease is stable or has progressed. Ruiz et al.[91] designed the first MRI-based score to determine radiological disease progression on follow-up MRI-MRCP of 64 well-characterised cohort of patients with large-duct PSC. They designed an interpretation standard model that allowed them to systematically review the bile ducts and liver parenchyma. This model was based on Craig et al.[81] classification of cholangiographic findings and Petrovic et al.[88] biliary wall enhancement after contrast administration.

Two MRI progression risk scores, Anali without and with gadolinium scores, were developed using independent factors that predicted radiological progression between two successive MRIs. These were intrahepatic bile duct dilatations, hepatic dysmorphy, portal hypertension, and heterogeneous parenchymal enhancement. This study demonstrated radiological progression in 58% of patients (n = 37) over a 4-year follow-up period. Both scores had AUC

of 0.80 and 0.83, respectively for predicting radiological progression. The prognostic value of radiological progression and Anali scores were not assessed. Furthermore, inter- and intra-observer variability were not examined. The prognostic ability of Anali scores were subsequently validated in a multicentre study that included 119 internal validation and 119 external validation cohorts. Anali scores without and with gadolinium identified transplant-free survival or cirrhosis decompensation with a C-statistic of 0.89 (95% CI: 0.84–0.95) and 0.75 (95% CI: 0.64–0.87), respectively. Cazzagon et al.[92] also demonstrated the complementary prognostic value of combining Anali score and transient elastography liver stiffness measurement for risk stratification in a retrospective multicentre study of 162 patients with PSC. Of note, it is worth noting that the Anali score found intrahepatic biliary dilatation — as opposed to extrahepatic biliary disease — to be an important prognostic feature whereas the UK-PSC score found that patients with extrahepatic biliary disease to have poorer outcomes. It is possible that both were measuring the same effect as the presence of extrahepatic disease is more likely to be associated with intrahepatic dilatations.

Muir et al[93] proposed an MRI risk score using the baseline MRI-MRCP data from the Simtuzumab trial. The score, based on portal hypertension, hepatic dysmorphy, and enlarged perihepatic nodes, was associated with the development of PSC-related clinical events that included 57% cholangitis, 27% jaundice, 13% liver decompensation. However, follow-up was limited to only two years. The full analysis of this work is yet to be published.

Kitzing et al.[94] retrospectively examined serial MRI-MRCP of 53 patients with PSC and reported that liver morphological changes, specifically liver atrophy, was associated with adverse clinical outcome and shorter transplant-free survival over a mean intervening period of 5 years. The authors, however, did not report on biliary changes.

Bookwalter et al.[95] retrospectively reviewed MRI that included dynamic contrast enhanced (DCE) sequences and diffusion-weighted imaging (DWI), MRCP and MRE of 55 patients with PSC to examine the relationship between liver parenchymal changes, biliary features

assessed using a simplified version of Ruiz et al. model[91], and liver stiffness at a segmental, lobar and global level. They found weak correlation at segmental level between liver stiffness on MRE and liver parenchymal changes (T2 hyperintensity, DWI hyperintensity, post-contrast hyperenhacement) as well as segmental strictures. However, neither parenchymal nor biliary changes were associated with MRS, Childs-Pugh, or MELD scores.

Tafur et al.[96] retrospectively examined the risk stratification ability of biliary stricture severity on MRCP, liver stiffness on VCTE and MRE, with MRS and AOM as the reference standard in 87 patients with predominantly large-duct PSC (98%). Biliary stricture severity was assessed using a standardised model modified from ERCP classification systems. MRE demonstrated good risk stratification ability when stratified by MRS and AOM scores, but performance was poor for VCTE and intrahepatic stricture severity. Combination of MRE and intrahepatic and extrahepatic stricture severity improved its risk stratification ability. Extrahepatic stricture severity had poor correlation with liver stiffness and prognostic scores and did not show any risk stratification ability. This study did not assess the inter-observer variability in MRI-MRCP interpretation.

Song et al.[97] evaluated the diagnostic performance of MRI-MRCP, blood tests and clinical risk scores against liver histology severity assessed using modified Nakanuma scoring system in a single-centre retrospective study of 51 patients with large-duct PSC. Whilst several morphological MRI features, blood tests and clinical risk scores had good performance to identify advanced histopathologic stage of PSC, severity of biliary stricture on MRCP showed poor performance. Histology in over half of patients were from liver explant (49%) or liver resection for cholangiocarcinoma (6%) with over two-thirds of patients in stage 3–4 PSC, indicating a proportional bias of including more advanced liver disease patients. It is also worth reflecting that the qualitative assessment of stricture severity was probably oversimplified in this study.

1.6 Quantitative MR techniques

1.6.1 Magnetic resonance elastography

MRE has been shown to predict liver decompensation (HR 1.55, 95% CI 1.41−1.70) in a large single-centre retrospective study involving 266 patients with median follow-up of 2 years[98]. In a small (n = 20) sub analysis of this study of patients who had a liver biopsy, MRE was also shown to correlate with fibrosis stage and had excellent diagnostic performance (AUC 0.97−0.99) to detect stages of liver fibrosis . Serum ALP < 1.5x ULN normal excluded the presence of advanced liver stiffness in this study but was not associated with the clinical outcome. MRE has also been shown to correlate better with MRS than VCTE and was an independent predictor of worse MRS in one study[99]. Hoodeshenas et al.[100] assessed the inter-observer agreement for MRE in 79 patients with PSC and demonstrated excellent intraclass correlation coefficients (ICC) between 0.97 and 0.99. It is worth noting that all the MRE studies were performed by the same group and single centre in USA over a period of five years. MRE is currently not widely available outside research settings and requires additional external equipment to MRI scanner.

1.6.2 Diffusion-weighted MRI

DW-MRI manipulates the altered diffusion properties of water protons in fibrotic tissue and allows assessment of liver fibrosis[101]. Addition of the short sequence to routine MRI-MRCP enables whole liver assessment of fibrosis distribution, particularly useful in a patchy disease such as PSC[28, 102]. In a prospective study involving 47 patients with PSC, DWI-MRI performed better than DCE-MRI in detecting and staging liver fibrosis using VCTE as the reference standard[103]. Since VCTE and MRE have shown better diagnostic accuracy for the staging of liver fibrosis[102, 104], DWI has fallen out of favour but is still performed for better characterisation of lesions involving the liver parenchyma.

1.6.3 Dynamic contrast-enhanced MRI

Administration of hepatocyte-specific contrast agents such as Gd-BOPTA (gadobenate dimeglumine) and Gd-EOB-DTPA (gadoxetate disodium) allows assessment of liver function

by analysing the liver uptake and elimination of contrast. Several studies have attempted to quantify liver parenchymal changes with the administration of a contrast agent.

In a retrospective study, Ringe et al.[105] demonstrated that hepatobiliary excretion of hepatocyte-specific contrast is significantly reduced in 111 patients with PSC when compared to 61 controls and correlated with bilirubin and ALP levels in PSC. Noren et al.[106] quantitatively compared hepatocyte-specific contrast uptake with histopathological stage of fibrosis in a prospective study involving 38 patients with compensated chronic liver diseases of varying aetiology. They demonstrated that quantitative measurement of signal intensities using three DCE-MRI parameters was able to distinguish advanced fibrosis (F3–4) from no and moderate fibrosis (F0–2) with AUC ranging from 0.71−0.78. Nilsson et al.[107] developed a non-invasive imaging-based method using DCE-MRI to assess liver function at the segmental and global level and showed significant heterogeneity in the liver parenchyma of 12 patients with mild PSC compared to 20 healthy controls. This small study reported significant correlation (r_s = -0.85) between MRI-derived liver function indices and disease severity using MRS. Segmental liver function correlated with level of downstream biliary obstruction.

Hinrichs et al.[108] reported that reduction in T1 relaxation time after hepatocyte-specific contrast administration showed significant correlation with liver biochemistry tests (AST: r = -0.60, ALP: r = -0.51, bilirubin: r = 0.64), MELD (r = -0.63), and MRS (r = -0.64) in 61 patients with PSC. They proposed that global liver function could be non-invasively assessed using this specific T1 mapping sequence technique in PSC. Nolz et al.[109] performed retrospective quantitative analysis of T1-weighted MRI-MRCP liver parenchymal enhancement in 20 patients with PSC and calculated the difference in signal intensity (SI) ratio between the hepatobiliary phase and unenhanced parenchyma. This difference was termed relative liver enhancement (RLE). They demonstrated significant reduction in RLE in localised areas of

impaired liver parenchyma in comparison to normal areas, thus allowing regional functional assessment.

Keller et al.[110] adapted the above technique to MRI scans performed with extracellular gadolinium-based instead of hepatocyte-specific contrast. They retrospectively reviewed scans and liver biopsies of 40 patients with PSC to evaluate the utility of T2 signal, contrast enhancement, and RLE as markers of liver inflammation and fibrosis. Liver biopsy was assessed using Ishak modified hepatic activity index (mHAI) and Scheuer fibrosis staging. Regression models demonstrated significant association between RLE in the delayed phase of T1-weighted MRI and stages of liver fibrosis (stage 2: $p = 0.007$; stage 3/4: $p < 0.001$). A borderline association ($p = 0.04$) between RLE in the portal phase and mHAI grade ≥ 7 was also observed. The same group also demonstrated an increased RLE within T2 hyperintense areas in the liver parenchyma on T2-weighted MRI of 29 patients with PSC and 19 controls. They postulated that this could be early changes of patchy inflammation[111].

Schulze et al.[112] calculated RLE in the hepatobiliary MRI phase in a prospective study using hepatocyte-specific contrast agent and evaluated its role as a prognostic marker in 111 patients with PSC. They reported a negative correlation with ALP ($r = -0.64$), bilirubin ($r = -0.65$), MELD ($r = -0.59$), MRS ($r = -0.54$), AOM ($r = -0.60$), and positive correlation with albumin ($r = 0.54$) and international normalised ratio (INR; $r = 0.46$). RLE correlated with clinical endpoints defined as development of cholangiocarcinoma, liver transplantation, liver-related mortality, and survival. They proposed a cut-off RLE value that predicted clinical endpoints (AUC 0.90; likelihood ratio 10.3, $p < 0.0001$) with reasonable sensitivity (74%) and excellent specificity (94%), but no information on the independent predictive value of RLE for the development of clinical outcomes was provided. This study was validated with similar results by another group in a retrospective study of 104 patients with PSC[113].

Evrimler et al.[114] assessed the relationship between pre- and post-contrast T1 relaxation time, extracellular volume fraction and severity of PSC in a retrospective study involving 93

patients with PSC and 66 normal liver controls. Both metrics were observed to be higher in PSC than controls and showed weak to moderate correlation with VCTE, FIB-4, and MRS.

1.6.4 Spleen and liver size

Ultrasound-based spleen measurements have been reported and shown to correlate with outcomes in PSC[77, 115]. Khoshpouri et al.[116] investigated the value of change and ratios of spleen and liver volumes on MRI in distinguishing those with adverse outcome (liver transplantation, transplant waiting list, and death) from those without in 89 patients. Spleen volume and left lobe-to-total liver volume were significantly different between the two groups and showed moderate risk stratification performance. Idilman et al.[117], however, showed that the performance of MRE for prediction of disease severity and liver decompensation was better than liver and spleen volume measurements in 266 patients with PSC. Changes in spleen length has also been shown to predict disease progression and warrants further validation[118].

1.6.5 Machine learning

Ringe et al.[119] has reported the first fully automated detection of PSC-compatible bile duct changes using a deep machine learning algorithm. Retrospective 3D MRCP from 205 patients with PSC and 223 non-PSC patients were randomly assigned to training (n = 386) and validation groups (n = 42). The number of images available for the training (n = 7720) and validation (n = 840) databases were increased by uniformly rotating axial MRCP images to generate 20 additional 2D projection images to boost the inference performance. They reported sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) of ≥ 91%. The results need to be replicated in other cohorts as deep neural networks tend to overfit the characteristics of the dataset. Nonetheless, this result looks promising and may help in reducing the significantly high inter-observer variability observed interpreting MRI-MRCP in PSC.

1.7 Limitations of MRI-MRCP

First, MRI-MRCP acquisition and 3D image reconstruction protocols still vary significantly across centres. Moreover, some centres are still performing thick and thin slabs 2D MRCP sequences. Second, there is currently no standard model or consensus for interpreting MRI-MRCP data. The MRI working group of the International PSC Study Group (IPSCSG) are currently working on guidelines to inform this. Current techniques allow only qualitative assessment and is susceptible to subjectivity in interpretation. Zenouzi et al[120] reported significant variability in MRI-MRCP interpretation even among experts. Third, assessment of the distal common bile duct and subtle changes in the smaller peripheral intrahepatic bile ducts remains a challenge despite the use of modern 3 Tesla (T) MRI scanners[29]. Fourth, MRCP lacks the depth perception that is provided by ERCP. Interpreting MRCP data using models designed for ERCP cholangiogram interpretation may overestimate findings as discussed above. Recognising these limitations, the MRI working group of the IPSCSG has outlined recommendations to standardise imaging techniques in PSC based on consensus statements and identified future research priorities[29]. Some of these unmet needs were addressed in this thesis.

1.8 Quantitative MR techniques assessed in this thesis

1.8.1 Liver*MultiScan*™

T1 mapping quantifies the T1 relaxation time (T1) based on the inherent property of tissues that largely reflect water content. It is increased with congestion, inflammation, and fibrosis. Further explanation of MRI relaxation times will be discussed in section 1.9. Several MRI techniques have been developed to measure T1. One of the methods is the modified Look-Locker inversion recovery (MOLLI) sequence that was originally proposed for measuring T1 values of the myocardium in cardiac imaging. A modification to this sequence was developed in our centre called the shortened modified Look-Locker inversion recovery (shMOLLI)[121]. Liver iron content can falsely decrease T1 measurement and can be estimated from the measured T2* relaxation time. A model that corrects this confounding

effect from iron was developed to return the measured liver T1 to the value it would be had the liver iron content been normal[122]. This algorithm was incorporated into what is known as the Liver*MultiScan* (Perspectum, Oxford, UK). Iron-corrected T1 (cT1) measured by Liver*MultiScan* has shown good correlation with histologically assessed liver fibrosis in other chronic liver diseases[123-126] and predicted outcomes in two studies[127, 128]. There are no published studies that have reported on the performance of Liver*MultiScan* specifically in PSC.

1.8.2 *MRCP+*™

A quantitative biliary analysis and advanced visualisation software have been developed called *MRCP+* (Perspectum, Oxford, UK). 3D MRCP images are processed using advanced image analysis techniques to produce a 3D model of the biliary tree. It enables measurement of bile duct diameter, automatic detection of regions of variation in diameter, interactive segmentation, and labelling of the biliary structures, and allows volumetric analysis of gallbladder and the biliary tree. Its utility in PSC remains to be established.

1.9 Principles of MRI

1.9.1 MRI and relaxation times

In MRI, images are acquired by exploiting the properties of protons in tissue. In their resting state, protons in tissue spin around their axis, exerting a magnetic moment. However, protons in tissue spin in different directions, therefore the direction of the magnetic field at any given time is random with no net magnetisation effect. When protons in tissue encounter an external strong magnetic field (B_0) such as when a person lies in an MRI scanner, low energy protons will align in the direction of the external magnetic field whereas high energy protons will align against the direction of the magnetic field. The revolving motion of protons around their axis in a magnetic field is termed precession. More protons precess in a low energy resting state, creating a small net magnetisation effect. This is not detectable as the magnetisation effect is in the same longitudinal direction of the much stronger external

magnetic field (B_0). A radiofrequency (RF) pulse is then applied at an angle to B_0 which has two main effect on the protons. These two effects contribute to the measurement of relaxation times in MRI.

T2 relaxation time*

Application of an external RF pulse has the effect of making the protons precess together in a synchronised fashion to produce a magnetic field at 90° to B_0. This transverse magnetisation component is now in a different direction to the main magnetic field, hence a signal that can be recorded by the scanner is generated. The coherence of the magnetic precession by the protons reduces over time with decay of the transverse magnetisation signal. The time taken for this signal to disappear is termed the T2 relaxation time. Magnetic field inhomogeneity influences the rate at which the protons precess and is taken into account to derive the T2* relaxation time. Iron has paramagnetic properties and liver iron generates small magnetic field inhomogeneities that influences T2* measurements.

T1 relaxation time

The RF pulse also causes some protons to enter a high energy state, thereby switching their alignment to the opposite direction to B_0, generating a magnetic field 180° to B_0. The time taken for this longitudinal magnetisation to return to 0° or low energy state following a 180° RF pulse application is termed the T1 relaxation time. Hydrogen protons are the most common type of proton in the body. The T1 and T2* relaxation times are influenced by the size and motion of the molecule in which the hydrogen proton is contained within. Hydrogen protons in small and rapidly rotating molecules such as in blood and bile has longer T1 relaxation time compared to hydrogen protons contained in dense fatty acid hydrocarbon chains such as liver fat.

T1 and T2 mapping*

The signal generated by the decay in longitudinal (T1) and transverse (T2*) magnetisation of a tissue can be registered onto pixelwise illustrations based on T1 and T2* signal intensities termed “mapping”. This allows direct T1 and T2* quantification.

1.9.2 T2-weighted MRCP

Static fluid-filled structures in the abdomen have long T2 relaxation time in comparison to adjacent tissue. MRCP exploits these differences by using heavily T2-weighted sequences that depicts higher signal intensity (white) of slow-moving or static fluid within the bile and pancreatic ducts in comparison to lower signal intensity (grey/black) of adjacent solid structures. Specific image acquisition sequences ensure that flowing blood has minimal or no measurable signal in order not to mistake blood vessels for bile or pancreatic ducts. Acquisition is performed in a fasted state (often for at least four hours) to reduce signal overlap from fluid in the surrounding stomach and duodenum, reduce peristalsis and promote gallbladder distension. Some centres use a negative oral contrast agent, such as 200-400ml of pineapple juice, 20-30 minutes prior to MRCP. The high concentration of manganese in pineapple juice is thought to have a paramagnetic effect, suppressing signal from overlapping fluid in the stomach and duodenum[129]. 2D and 3D MRCP has undergone significant advances since its first description with shorter acquisition time, better image quality and improved reconstruction algorithms. Optimal imaging protocol depends on the specific scanner used and parameters including field strength (e.g., 1.5 or 3 T), the manufacturer, institutional preference, and experience. Good quality 3D MRCP acquisition also depends on good navigator-based respiratory-triggering of the diaphragm. 3 T scanners have a higher spatial resolution and are preferred over 1.5 T scanners[29].

1.10 General aims

The overall aim of my thesis was to contribute to the on-going search in the field for non-invasive imaging biomarker that could be used as a surrogate marker of endpoints in future longitudinal studies and/or interventional drug trials. It aims to establish candidate metrics for risk stratification in PSC derived from two standardised quantitative MRI techniques; *MRCP+* for biliary tree assessment and Liver*MultiScan* for liver parenchymal assessment.

The specific aims of my thesis were to:

1. Validate *MRCP+* and Liver*MultiScan* analysis methods in PSC

2. Define and assess the severity of biliary tree strictures and dilatations using *MRCP+*

3. Assess liver parenchymal heterogeneity using Liver*MultiScan*

4. Explore periductal changes in the liver parenchyma by combining *MRCP+* and Liver*MultiScan* methods

5. Determine the temporal relationship in *MRCP+* and Liver*MultiScan* metrics

6. Explore the relationship of patient-reported outcomes with MRI metrics and other non-invasive tools

2 General methods

2.1 Study design

The multiparametric magnetic resonance imaging of biliary diseases (MoBILISE) study was a prospective, single-centre, cross-sectional study at a UK tertiary non-transplant centre (Oxford University Hospitals NHS 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. Study assessments were conducted at baseline (visit 1) and follow-up (visit 2) at least 12 months apart.

Follow-up study visits were temporarily halted between March and July 2020 due to the Covid-19 pandemic restrictions. Study assessments resumed in August 2020 and were again halted in November 2020 due to the second wave of the pandemic. Data was censored in January 2021 to allow completion of this thesis.

I conducted a patient and public involvement workshop at the conception of this study to obtain feedback on study agenda and assessments.

2.2 Eligibility criteria

2.2.1 General consideration

All participants were aged ≥18 years, were able to provide informed written consent for participation in the study, and had no contraindications to MRI (e.g., pacemaker, pregnancy, ferromagnetic implant, shrapnel injury, severe claustrophobia).

2.2.2 PSC

Patients with an established diagnosis of large-duct PSC were consecutively recruited from the outpatient specialist PSC and general hepatology clinics. Large-duct PSC was defined

according to EASL and AASLD guidelines as the association of chronic cholestasis, typical features on MRCP or ERCP, and no cause of secondary sclerosing cholangitis[7, 17]. Patients were excluded if they had co-existing cholangiopathy, small-duct PSC, overlap with primary biliary cholangitis (PBC), 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, including chronic viral hepatitis, non-alcoholic fatty liver disease (NAFLD), Wilson's disease, hereditary haemochromatosis or alpha-1-antitrypsin deficiency, or consumed more alcohol than the current limit recommended by the UK Department of Health (14 units/week).

2.2.3 Healthy controls

Age and sex-matched healthy volunteers were recruited by poster advertisement and by word of mouth, for reference given the absence of normal values for *MRCP+* technique. Participants 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 (BMI) not greater than 25kg/m^2.

2.3 MRI acquisition protocol

All participants attended a study visit after fasting for at least four hours. MRI and MRCP sequences were acquired at the University of Oxford Centre for Clinical Magnetic Resonance Research (OCMR) on a 3 T scanner (Siemens Magnetom Prisma, Erlangen, Germany). An 18-channel phased-array torso coil was used. All participants consumed 200ml of pineapple juice 20 minutes prior to the scan to reduce the signal intensity of overlapping fluid within the stomach and duodenum for the MRCP portion of the scan protocol[129]. The scan protocol included 3D MRCP, T1 mapping, T2* mapping, and volumetric interpolated breath hold examination (VIBE).

2.3.1 3D MRCP

A heavily T2-weighted navigator-based respiratory-triggered coronal 3D MRCP with maximum intensity projection (MIP) reconstruction was acquired during the quiescent portion of end expiration as shown in Figure 3. The sequence parameters were field of view (FOV)280 x 280mm, acquisition matrix 256 x 256mm, voxel size 0.5 x 0.5 x 1.1mm, 1.1mm slice thickness, 60 slices, in-plane image resolution 1.1 x 1.1mm, echo time (TE) 604, repetitition time (TR) 3000ms (variable depending on the respiratory rate), 120° flip angle, GRAPPA acceleration factor 2 with 24 reference lines integrated mode and strong fat suppression. The acquisition time was 5 minutes. Two images were acquired with opposing coronal oblique orientations, one to cover the largest extent of the biliary tree, the other to cover the full extent of the gallbladder.



Figure 3. An example of maximum intensity projection (MIP) image from 3D MRCP acquisition.

Reprinted with copyright permission from Selvaraj EA, Culver EL, Bungay H et al. Evolving role of magnetic resonance techniques in primary sclerosing cholangitis. *World J Gastroenterol* 2019; 25:644-658. © 2019 Baishideng Publishing Group Inc.

2.3.2 Liver MRI

2.3.2.1 T1 mapping and T2* mapping

T1 maps were acquired using the shMOLLI sequence described in chapter 1 in an axial plane through the abdomen, obtaining good visual representations of the liver and spleen. The sequence used balanced steady state free precession to measure the signal intensity at different times throughout T1 recovery. These signals are then modelled to a T1 recovery curve to derive the measured T1. The details of this method have been described previously[121, 122] and applied to the liver[123]. The sequence parameters were 440 x 330mm FOV; 8mm slice thickness, 4 slices, voxel resolution 2.2 x 2.2 x 8.0mm (interpolated to 1.1 x 1.1 x 8.0), acquisition matrix 192 x 144mm, TE 1.05; TR 2.43ms, 35° flip angle, GRAPPA acceleration factor 2 with 24 reference lines. The acquisition time was approximately 9 seconds.

T2* maps were acquired using a multi-echo gradient echo (GRE) sequence DIXON method for the quantification of liver iron concentration. The same slice position and FOV used for the T1 mapping was used. The other parameters were 6mm slice thickness, 1 slice, voxel resolution 3.1 x 3.1 x 6.0mm (interpolated to 1.6 x 1.6 x 6.0), acquisition matrix 208 x 256mm, TE 1.23 with echo separation 1.23ms, TR 22.2ms, 20° flip angle, GRAPPA acceleration factor 2 with 35 reference lines. The acquisition time was approximately 11 seconds.

Both sequences were acquired with ECG or peripheral pulse gating in expiration breath-hold.

2.3.2.2 Volumetric interpolated breath hold examination

3D VIBE was acquired using double echo fast low angle shot (FLASH) sequence in the axial plane covering the whole liver and spleen. The sequence parameters were FOV 309 x 380mm (optimised to the patient), 3mm slice thickness, 72 slices, voxel resolution 1.2 x 1.2 x 3.0mmm, acquisition matrix 195 x 320, TE 1.29 and 2.52, TR 3.97ms, 9° flip angle,

CAIPIRINHA acceleration factor 2 with 24 reference lines in both phase encoding and 3D. The acquisition time was approximately 11 seconds.

Details of acquisition parameters are summarised in Table 2 and an example of each sequence in Figure 4.

Table 2. Detailed acquisition parameters for MRI.

Parameter	T1	T2*	VIBE
Sequence	Shortened modified Look- Locker Inversion (shMOLLI)	Multi-echo gradient echo (GRE)	Double echo fast low angle shot (FLASH)
Plane	Axial	Axial	Axial
Triggering	Breath-hold, pulse or ECG-gated	Breath-hold, pulse or ECG-gated	Breath-hold, not triggered
Number of slices	5	1	72
Slice thickness, mm	8	6	3
Echo time (TE), ms	1.05	1.23 with echo separation 1.23ms	1.29, 2.52
Repetition time (TR), ms	2.43	22.2	3.97
Acquisition matrix	192 x 144	208 x 256	195 x 320
FOV, mm	440 x 330 (75% phase FOV)	400 x 325 (81% phase FOV)	309 x 380
Flip angle (degrees)	35	20	9
iPAT factor	GRAPPA 2-fold with 24 reference lines	GRAPPA 2-fold with 35 reference lines	CAIPIRINHA 2-fold with 24 reference lines in both phase encoding and 3D
Voxel resolution, mm	2.2 x 2.2 x 8.0 (interpolated to 1.1 x 1.1 x 8.0)	3.1 x 3.1 x 6.0 (interpolated to 1.6 x 1.6 x 6.0)	1.2 x 1.2 x 3.0
Number of averages	1.0	1.0	1.0
Fat suppression	None	None	None

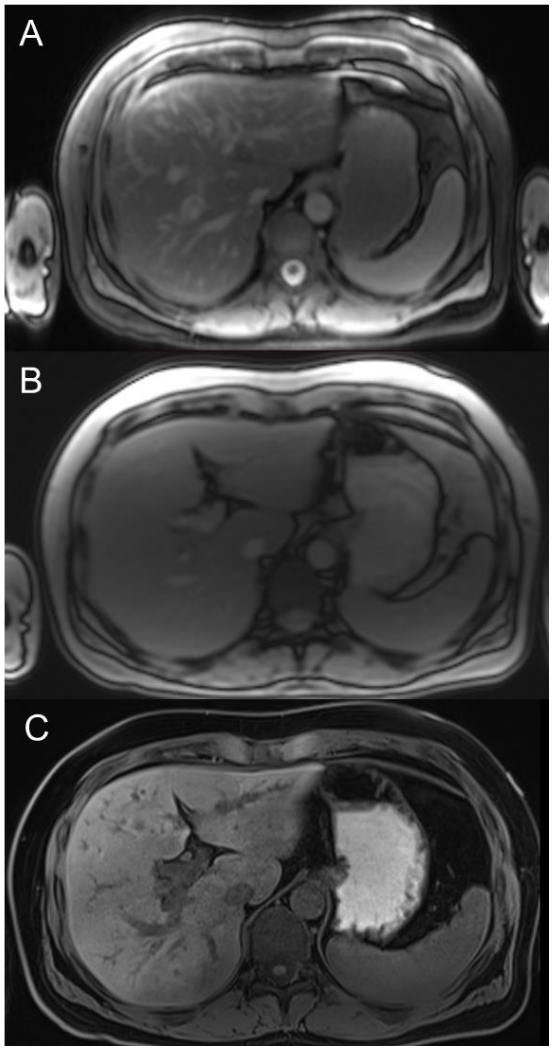


Figure 4. MRI transverse slice through the liver and spleen with an example of (A) T1 mapping; (B) T2* mapping; and (C) volumetric interpolated breath-hold examination (VIBE) sequences.

2.4 Quantitative MRI analysis

2.4.1 *MRCP+* analysis

3D MRCP data was uploaded and transferred via a dedicated online portal to Perspectum to be processed using the *MRCP+* software. A quantitative parametric 3D model of the biliary tree and pancreatic ducts was computed as shown in Figure 5. The first step involved automatic representation of tubular structures by the algorithm on the *MRCP+* interface. The algorithm used advanced image analysis techniques to enhance tubular structures and identify connected components to the tubular structures. This was performed automatically using an intelligent path search algorithm that was able to systematically follow branching points from the initial path until the whole biliary tree was traversed. The gallbladder was identified and segmented automatically when present. Next, manual visual identification and discrimination of the biliary tree and pancreatic ducts from neighbouring tubular structures (e.g., duodenum, blood vessels) was performed. Raw MRCP images were available on the interface alongside the 3D model and were interrogated to ensure accuracy in modelling of the biliary tree The diameter of each modelled duct is automatically determined, and a parametric 3D model of the biliary tree colour-coded by variation in duct diameter is automatically generated. The following ducts were identified and labelled by the trained operator: common bile duct (CBD), common hepatic duct (CHD), main right hepatic duct (RHD), main left hepatic duct (LHD), cystic duct (CD) and pancreatic duct (PD). The point of insertion of the cystic duct marked the distal border of the CHD. The first proximal bifurcation of the CHD into RHD and LHD marked the proximal border of the CHD. The proximal extent of the RHD and LHD was the first point of bifurcation of RHD and LHD into second-order duct. If a particular duct was disconnected in its course, ducts were labelled as duct segment 1, 2, 3 and so on. Automatic measurements of duct diameters, duct length, as well as the biliary tree and gallbladder volumes were obtained.

Metrics were derived for individual ducts, for the whole biliary tree after excluding the cystic and pancreatic ducts, the extrahepatic (CBD, CHD, and the main RHD and LHD up to their first notable branches) and intrahepatic (starting from the first branches of the main RHD and LHD) segments.

MRCP+ automatically detects regions with variation in bile duct diameters. The threshold for classification of a candidate stricture and dilatation was already pre-defined in the *MRCP+* algorithm. Derivation and application of candidate stricture and dilatation metrics will be discussed in more detail in chapter 4.

The *MRCP+* metrics were:

- Total biliary tree length (mm)
- Total biliary tree volume (ml)
- Gallbladder volume (ml)
- CBD, CHD, RHD, and LHD minimum and maximum diameter (mm)
- Proportion of ducts with median diameter <3mm (%)
- Proportion of ducts with median diameter 3–5mm (%)
- Proportion of ducts with median diameter >5mm (%)
- Candidate stricture metrics (discussed in chapter 4)
- Candidate dilatation metrics (discussed in chapter 4)

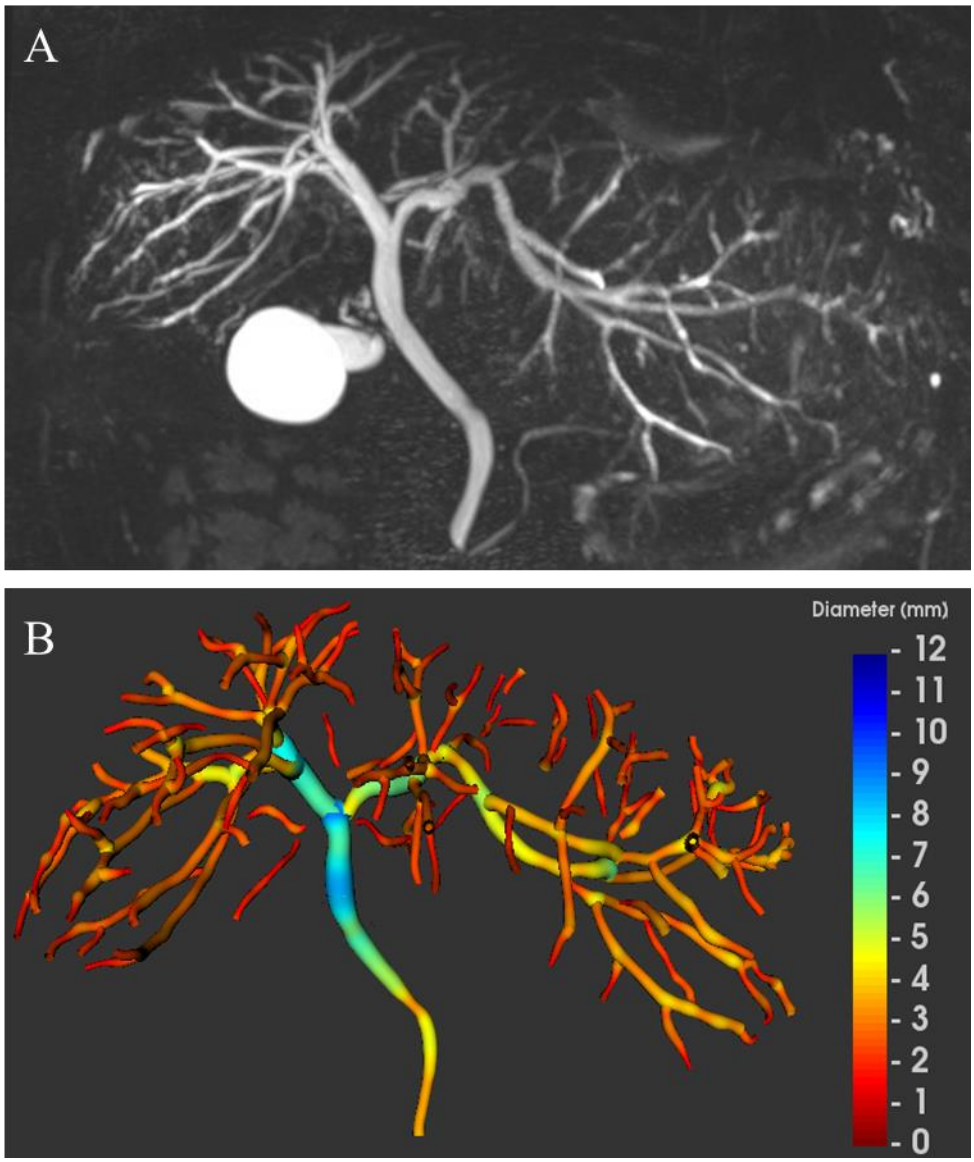


Figure 5. *MRCP+* analysis. (A) 3D MRCP maximum intensity projection (MIP) image, and (B) the corresponding parametric biliary 3D tree model generated from *MRCP+* in a 32-year-old man with PSC, colour coded according to duct diameter. The gallbladder has been automatically segmented and its volume presented separately. The cystic and pancreatic ducts were excluded for the biliary analysis in this thesis.

2.4.2 Liver*MultiScan* analysis

MRI data were uploaded and transferred via a dedicated online portal to Perspectum to be processed using the Liver*MultiScan* software by trained image analysts. Regions of interest (ROIs) were manually placed on T2* colour maps to obtain iron concentration representation within the liver. An algorithm transforms the measured T1 map into a cT1 map using iron concentration acquired from the T2* map to correct for distortions caused by excess iron in the liver parenchyma. This quantitative T1 maps were calculated by fitting T1 recovery curves to the acquired data. Liver cT1 can be quantified using two different Liver*MultiScan* processing methods: (1) ROI-placement method; and (2) liver-segmentation method.

2.4.2.1 ROI-placement method

Four circular ROIs of 15mm diameter were manually placed on an axial T1 and T2* maps to capture the distribution of image intensities within the liver, where possible on the posterior, lateral, anterior, and medial (left lobe) of the liver in that order as depicted in Figure 6. Blood vessels, organ boundaries and potential cysts or ascites were avoided. Mean cT1 pixel values within a single ROI were calculated. The mean of the four ROIs is represented as the final cT1 value for that axial slice. The MRI metrics from this analysis method were:

- Pooled ROI mean and standard deviation T1 (ms)
- Pooled ROI T2* median and interquartile range (ms)
- Pooled ROI mean and standard deviation cT1 (ms)
- Pooled ROI median and interquartile range iron (mg/g)
- Pooled ROI mean and standard deviation proton density fat fraction (%)

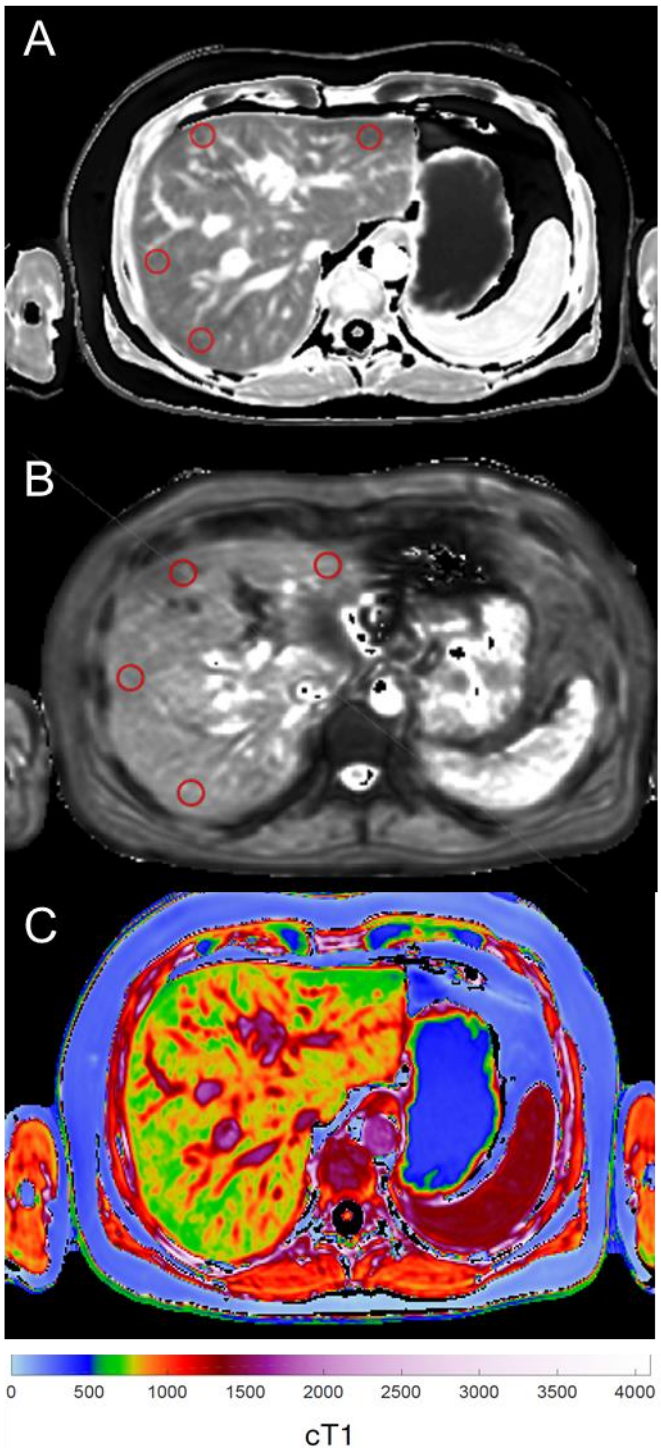


Figure 6. Region of interests (ROIs)-placement method. Liver*MultiScan* analysis of a single transverse MRI slice using in a 46-year-old man with large-duct PSC. Four ROIs (red circles) were placed on the (A) T1 map and (B) T2* map to generate (C) iron-corrected T1 (cT1) map. The colour-scale represents the distribution of cT1 within the liver. The mean/median value of from the four ROIs are computed. In this example, mean T1 = 717ms, median T2* = 26.1ms mean cT1 = 718ms, mean iron content = 0.8mg/g dry weight liver, and median PDFF = 3.6%.

2.4.2.2 Liver-segmentation method

In this method, the surface area of the whole liver is considered and cT1 is measured in each voxel of the liver acquisition rather than discrete ROIs. Liver segmentation can be performed either using a semi-automated analysis of all four liver slices with input from an image analyst or it can be a fully-automated analysis of a single representative liver slice without any input from an image analyst.

Semi-automated multi-slice liver segmentation method

The liver is segmented semi-automatically on the cT1 map of a liver slice, removing blood vessels and other abnormalities (e.g., cysts or ascites). This process is performed and checked by a trained image analyst. cT1 values from all pixels from the four segmented liver slices are combined to produce the metrics representative of the whole liver with associated histogram as shown in Figure 7. The MRI metrics from this analysis method were:

- Whole liver mode cT1 (ms)
- Whole liver mean cT1 (ms)
- Whole liver median cT1 (ms)
- Whole liver interquartile range cT1 (ms)

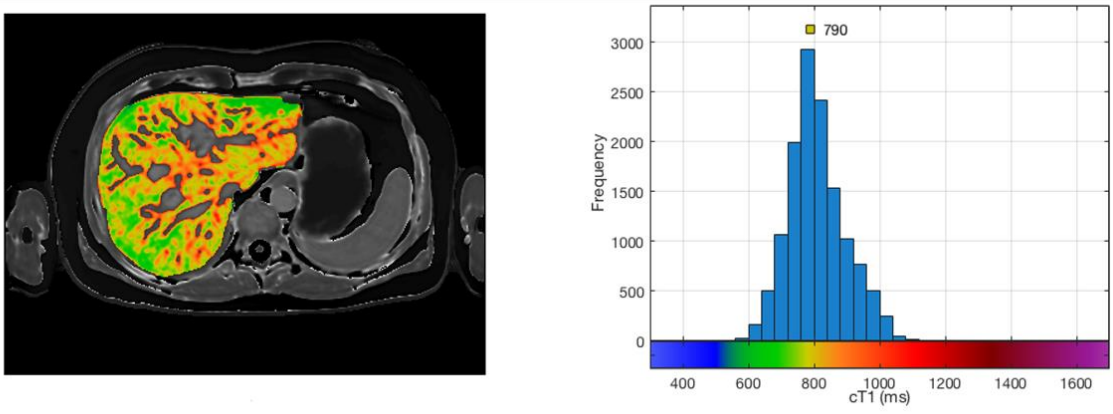
Fully-automated single-slice liver segmentation method

An automated liver segmentation approach have been developed[130]. A single axial liver T1 and T2* map with the largest area of the liver and fewest artifact exclusions was selected by the algorithm for analysis. The cT1 maps were then visually assessed for quality control. The mask was manually re-contoured where necessary to ensure only the liver parenchyma is included within the mask. This method became available after the conception of this study, therefore was not used in the main analysis of this study. It was used to standardise Liver*MultiScan* analysis across a large set of previously acquired liver MRI data as part of

other studies to validate candidate metrics described in chapter 5. The MRI metrics from this analysis method were:

- Median and IQR cT1 (ms)
- Mean and standard deviation cT1 (ms)
- Skewness and kurtosis cT1

A)



B)

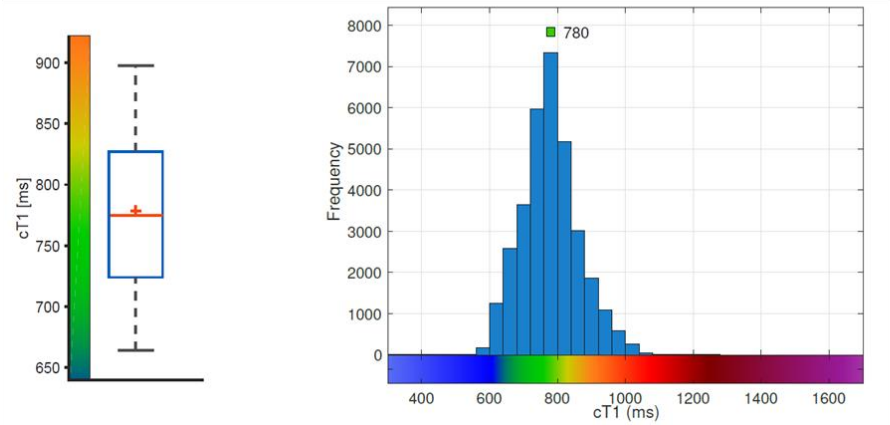


Figure 7. Liver segmentation method. (A) Single-slice iron-corrected T1 (cT1) map from Liver*MultiScan* analysis with its associated histogram depicting the distribution of cT1 within the transverse slice segmented liver in a 46-year-old man with large-duct PSC. Mode = 790ms, mean cT1 = 809ms, median cT1 = 799ms, and interquartile range cT1 = 107ms. (B) Summary combined statistics from all four transverse liver slices analysis in the same patient. Mode cT1 = 780ms, mean cT1 = 778ms, median cT1 = 775ms, and interquartile range cT1: 104ms.

2.4.2.3 Superpixel analysis

The principle of the superpixel analysis method is to partition an image into subregions within the liver according to local similarity which allows analysis of regional variance as shown in Figure 8. In this study, the maskSLIC method was used to create evenly distributed superpixels inside an irregular liver mask that is not affected by vessels and liver border transitions[131]. The same liver slice from the automated single-slice liver segmentation method was chosen to perform this analysis.

For each of the superpixels, the mean and standard deviation were computed. The MRI metrics from this analysis method were:

- Mean superpixel variance cT1 (ms)
- Skewness of superpixel cT1 distribution
- Kurtosis of superpixel cT1 distribution

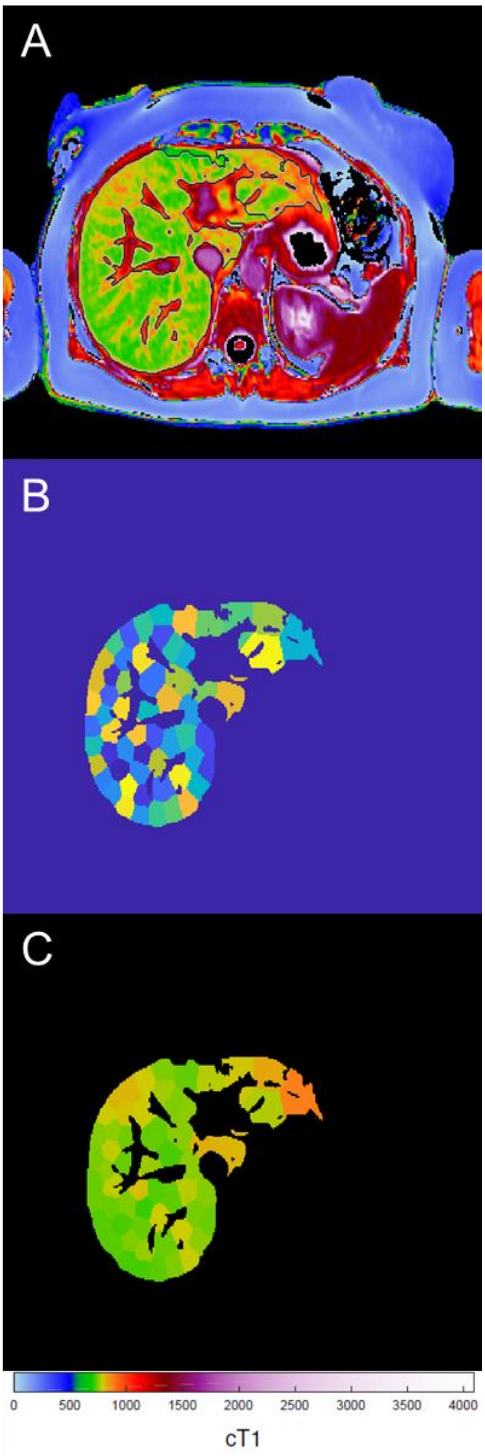


Figure 8. Superpixel analysis of a single transverse MRI slice. (A) Iron-corrected T1 (cT1) map with liver mask in a 46-year old man with large-duct PSC. (B) Segmented liver parcellated into superpixels according to local similarity in cT1. Each superpixel is denoted by a different colour to differentiate them from each other for visualisation purpose only. Variance of cT1 within each superpixel is computed. (C) Colour representation of mean cT1 for each superpixel according to cT1 values. Total number of superpixels = 59, mean superpixel variance cT1 = 54ms, skewness = 0.5, and kurtosis = 2.1.

2.5 Qualitative MRI analysis

Both MRCP and MRI sequences of patients with PSC were independently assessed by two expert hepatopancreatobiliary radiologists with a subspecialist interest in PSC, (A.B.S and S.P.L. with 25 and 8 years of experience, respectively). A standardised interpretation model was developed for this study which incorporated the following information: modified Amsterdam cholangiographic stricture severity score (Table 3)[46, 89], Anali score (Table 4)[91], and other morphological features such as the presence of liver cirrhosis, splenomegaly, dominant stricture, gallbladder enlargement, and the maximum diameter of CBD, CHD, LHD, RHD and IHD (Table 5). Information on the diagnosis of PSC was available to the radiologists but they were both blinded to clinical and quantitative MRI-MRCP data.

Table 3. The original and modified score from the Amsterdam cholangiographic classification system.

Amsterdam score[46]	Modified score[89]	Description
Intrahepatic tree		
0	0	No visible abnormalities
I	1	Ductular irregularities
I	2	Multiple calibre changes; minimal dilatation
II	3	Multiple strictures: saccular dilatations, decreased arborisation
III	4	Only central branches filled despite adequate filling pressure; severe pruning
Extrahepatic tree		
0	0	No visible abnormalities
I	1	Slight irregularities of duct contour, no stricture
II	2	Segmental stricture
III	3	Stricture of almost entire length of duct
IV	4	Extremely irregular margins, diverticulum-like outpouchings

Table 4. Classification of components for the derivation of the Anali score.

Component	Score	Description
Intrahepatic bile duct dilatation*	0	3mm or smaller
	1	4mm
	2	5mm or larger
Hepatic dysmorphy†	0	Absent
	1	Present
Portal hypertension‡	0	Absent
	1	Present

*Intrahepatic duct with the largest diameter
†Significant atrophy of either the right or left hepatic lobe and/or marked lobulations of liver surface and/or increase of the caudate/right lobe ratio
‡The presence of portosystemic shunts with or without splenomegaly

Table 5. Other morphological features assessed on MRI-MRCP interpretation.

Component	Score	Description
Dominant stricture	0	Absent
	1	Present
Liver cirrhosis	0	Absent
	1	Present
Splenomegaly	0	Absent
	1	Present
Gallbladder enlargement	0	Absent
	1	Present
Maximum diameter (mm) • CBD • CHD • RHD • LHD		

2.6 Clinical and laboratory data

Participants' height and weight were measured for calculation of the body mass index as weight (kg)/height (m)^2. Prior to MRI-MRCP acquisition, fasting blood samples were taken for ALP, AST, ALT, GGT, bilirubin and albumin. Samples for ELF (Siemens Healthineers, Erlangen, Germany) were collected, centrifuged, and stored at -80 degrees Celsius and were batch analysed and recorded once recruitment was completed. Electronic medical records were searched to obtain relevant clinical data (age at diagnosis, presence of IBD, medication history, and history of variceal bleeding). The distinction between intrahepatic and extrahepatic biliary disease was determined by review of radiographic reports of cholangiographic imaging. Involvement of first-order bile ducts (right or left main hepatic duct) and/or common bile duct was classified as extrahepatic disease, as opposed to their absence being classified as intrahepatic disease[23, 91].

2.7 Transient elastography

Vibration-controlled transient elastography using FibroScan® 502 Touch (Echosens, Paris) is a non-invasive bedside test that measures liver stiffness (LS), a surrogate markers of liver fibrosis. Liver stiffness for each participant was obtained using the automatic probe selection tool in accordance with the manufacturer's instruction. 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. Measurements were performed on the same day as blood test and MRI-MRCP.

2.8 Patient reported outcomes

All participants completed two quality of life questionnaires: SF-36[132] and PSC-PRO[133]. The SF-36 is a generic 36-item questionnaire that assesses eight domains: physical functioning, role physical, bodily pain, general health, energy/fatigue, social functioning, role emotional, and mental health (Figure 9). The physical component summary mean aggregate score is composed of the first four domains, and the mental component summary mean aggregate score consists of the last four domains. Possible scores range from 0 to 100, with higher scores indicating higher QoL.

The PSC-PRO is a disease-specific 42-item questionnaire comprised of two modules that assesses PSC symptoms (Module 1: 12 items) and impact of symptoms (Module 2: 30 items). For module 1, the item scores range from 0 to 10 to be additive with maximum score of 120. For module 2, there are seven domains: physical function, activities of daily living, work productivity, role function, emotional impact, social or leisure impact, and quality of life. The item scores range from 1 to 5 to be summed within their domains, with lower scores indicating higher QoL.

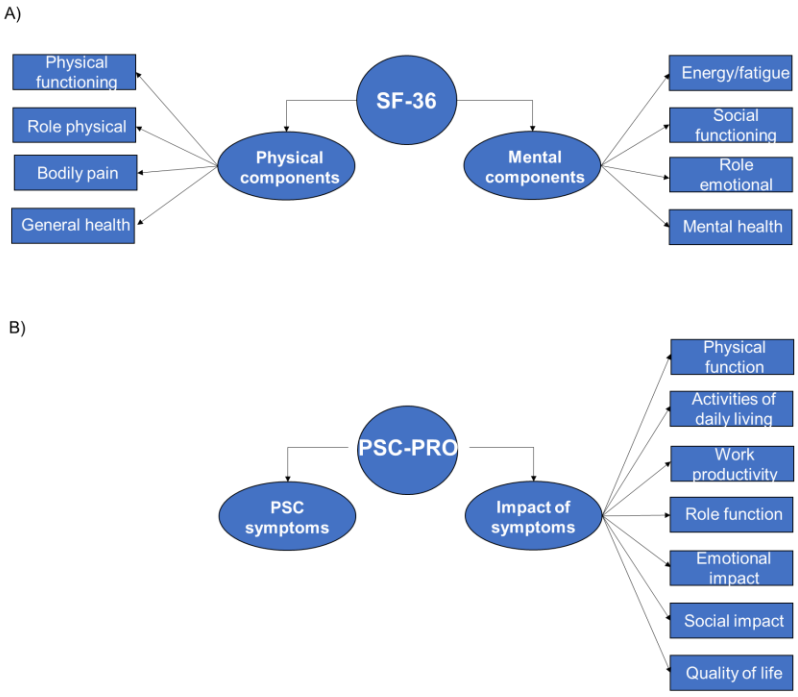


Figure 9. Domains in the quality of life questionnaires. (A) SF-36 and (B) PSC-PRO

2.9 PSC risk classification

Patients with PSC were classified into high-risk or low-risk groups according to previously published thresholds for non-invasive risk-stratification tools as summarised in Table 6.

2.9.1 Prognostic risk models

The revised Mayo Risk Score (MRS; based on patient age, bilirubin, albumin, AST, and history of variceal bleeding) has been the most cited prognostic risk models in PSC studies. It provides short-term (4-year) mortality risk of patients with PSC[68]. Patients were categorised 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[90, 96] and defined MRS >0 as high-risk to avoid small numbers in group.

The Amsterdam-Oxford Model (AOM; based on PSC subtype, age, albumin, ALP, AST, bilirubin, and platelet count) predicts long-term risk of PSC-related death and/or liver transplantation[65]. Although initially modelled 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[70]. This cut-off was used to define high-risk (AOM > 2) and low-risk (AOM ≤ 2) groups in this study.

2.9.2 Cholestasis marker

Serum ALP has been the most widely studied biomarker in PSC. Whilst historically a cut-off value of 1.5 times the upper limit of normal (xULN) has been reported to have prognostic implications[58], higher cut-offs of ≥ 2.2 xULN and ≥ 2.4 xULN have been recently proposed to have a better predictive ability of liver transplantation and survival in a large cohort of patients with PSC in the UK[23]. Both the 1.5x and 2.2x ULN cut-offs were examined. The ULN of ALP in our centre was 130 IU/L.

2.9.3 Liver fibrosis markers

High-risk PSC was defined as LS by transient elastography of > 9.6kPa[76] or ELF score > 9.8[61, 62] as these cut-offs correlate with histologically-assessed fibrosis and clinical outcomes.

2.9.4 Qualitative MRI-MRCP scores

The Amsterdam cholangiographic score was derived originally from ERCP images and predicted transplant-free survival in two cohorts[46, 86], but has been modified for MRCP interpretation in this study. The modified Amsterdam intrahepatic stricture severity score (ISSS) and extrahepatic stricture severity score (ESSS) ranged from 0 to 4, respectively, and were added together for the total stricture severity score (TSSS). A score of 0–2 was classified as mild disease (low risk) and 3–4 as severe disease (high risk)[89].

The Anali score without gadolinium was calculated as (1 x dilatation of intrahepatic bile duct) + (2 x dysmorphy) + (1 x portal hypertension)[91]. Anali score > 2 predicted adverse-outcome-free survival in a large multi-centre validation study[134], so in this study Anali score 0–2 was classified as low-risk and 3–5 as high-risk.

2.9.5 Disease distribution

The presence of extrahepatic biliary disease was associated with adverse outcomes in a large UK multicentre PSC research cohort[23]. In that study, extrahepatic disease was defined as the involvement of first-order bile ducts (right or left main hepatic duct) and/or common bile duct, as opposed to their absence being classified as intrahepatic disease. The same classification was applied here.

2.9.6 Dominant stricture

The presence of dominant strictures is associated with a high risk of developing CCA and worse outcomes[32]. In this study, a dominant stricture was defined as stricture less than 1.5mm diameter in the CBD, or less than 1mm in the left or right main hepatic ducts[7]. The presence of a dominant stricture was assessed on MRCP reviewed by specialist radiologists.

Table 6. Surrogate prognostic biomarkers in PSC that was selected for use in this thesis.

Biomarker	Thresholds	Outcome
Prognostic risk models		
MRS	> 2 (high risk)[68] 0-2 (intermediate risk < 0 (low risk)	4-year overall survival
AOM score	> 2 (high risk)[70]	Composite: Liver-related or PSC-related death including CCA, colorectal cancer[65]
Liver fibrosis markers		
VCTE LS	> 9.6kPa[75, 76]	• Advanced fibrosis (F3) • Composite: decompensation, CCA, HCC, death or liver transplant
ELF score	> 9.8[61, 62]	Transplant-free survival
Cholestasis marker		
ALP	> 1.5x ULN[54-56, 58]	• Composite: PSC-related death, liver transplant[54] • Death or liver transplant[55]

		• Composite: Liver transplant, death, CCA[56] • Composite: CCA, decompensation, liver-related death[58]
	> 2.2x ULN[23]	Liver transplant and survival
MRI-MRCP features		
Anali score	Without gadolinium > 2[134]	Composite: Transplant-free survival, liver-related death or decompensation
Modified Amsterdam score	ISSS > 2 = severe ESSS > 2 = severe	Transplant-free survival, PSC-related death
Extrahepatic disease[23]	Present	All-cause mortality and liver transplant
Dominant stricture	Present	Transplant-free survival

Abbreviations: ALP, alkaline phosphatase; CCA, cholangiocarcinoma, ULN, upper limit of normal; LS, liver stiffness; ELF, enhanced liver fibrosis; MRS, revised Mayo risk score; AOM, Amsterdam-Oxford model; ISSS, intrahepatic stricture severity score; ESSS, extrahepatic stricture severity score; TSSS, total stricture severity score.

2.10 Sample size calculation

In the absence of *a priori* data for the *MRCP+* technique, sample size calculation was based on recruitment figures in a seminal study reporting an MRI-based score that predicted radiological progression and clinical endpoints in PSC[91]. In that study, 37/64 patients had radiological evidence of progression (high risk) and 27/64 had stable disease (low risk) over a period of four years. Maintaining the same sample size will yield a power of 82% to detect a significant difference with medium effect size between the two groups assuming a non-normal distribution of *MRCP+* metrics. After accounting for 10% exclusion due to poor quality MRCP acquisition and 10% dropout rate during follow-up, the number to be recruited was determined to be 76 patients. For the healthy volunteers, a minimum of 16 participants were required to achieve a minimum of 80% power to detect a large effect size in *MRCP+* metrics between PSC and healthy cohorts in a 5:1 allocation ratio. Calculations were performed using G*Power version 3.1.

There is no published data assessing liver cT1 in PSC. The UK-PSC research cohort (n = 1001) identified a clinical event rate (transplant or death) of 36% over a cumulative follow-up period of 7,904 years; yielding an approximate average of 5 events per 100-patient years[23]. This incidence matches the rate of transplant or deathreported from the study conducted by the IPSCSG (n = 7121 patients from 39 centres)[37]. Using this figure, our intended study population of 76 patients will yield 8 events at 2 years, 20 at 5 years and 40 at 10 years.

2.11 Statistical analysis

Descriptive statistics were used to summarise participant characteristics. Normality was assessed using the Shapiro-Wilk test. Associations between continuous variables were tested using the Pearson's correlation coefficient for parametric distribution and Spearman's rank correlation coefficient for non-parametric distribution. Associations between ordinal and continuous variables were assessed using Kendall's tau if ordinal data contained < 5 categories and Spearman's rank correlation coefficient if ≥ 5 categories. Associations between binominal and continuous variables were assessed using point-biserial correlation coefficient.

Differences between two independent groups were assessed using Student's *t*-test for parametric distribution and Mann-Whitney test for non-parametric distribution. Differences between three independent groups were assessed using one-way analysis of variance (ANOVA) for parametric distribution with Tukey's multiple comparisons tests, and Kruskal-

Wallis for non-parametric distribution with Dunn's multiple comparisons tests. Receiver operating characteristic curve (ROC) analyses were used to determine diagnostic accuracy of MRI and MRCP metrics to detect the high-risk group. Area under the curve (AUC), sensitivity and specificity were recorded to maximise the Youden index.

Statistical significance was set at $p < 0.05$ and 95% confidence intervals (CI) were calculated where appropriate. Statistical analysis was carried out using GraphPad Prism version 9.0 for Windows (California, USA) and IBM SPSS (v25, Armonk NY, IBM Corp).

In the next chapter, I will discuss technical development work undertaken to validate some of the methods that were used in this thesis.

3 Validation of quantitative image analysis methods

3.1 Chapter summary

Background: The quantitative MRCP method used in this thesis was novel. The influence of acquisition protocols, internal measurement error, and observer variability on derivation of *MRCP+* biliary metrics were investigated to inform and interpret the use of these metrics in the thesis. Liver*MultiScan*-derived cT1 method was more established but its role and performance specifically in PSC is unknown. The effect of two confounding factors on cT1 measurement (liver iron and consumption of pineapple juice) were explored and the difference between the three Liver*MultiScan* analysis methods were compared to inform data analysis method in this thesis.

Methods: Seven experiments (exp) were performed in total. Healthy volunteers and/or patients with PSC were recruited. MRCP and MRI data were analysed with *MRCP+* and Liver*MultiScan* software, respectively, to compute biliary *MRCP+* metrics and liver cT1. For *MRCP+*, exp 1 assessed the influence of two different MRCP acquisition protocols, exp 2 the repeatability and intra-observer variability in healthy volunteers, exp 3 compared biliary tree visualisation between MRCP and *MRCP+*, and exp 4 assessed the inter- and intra-observer variability in PSC. For Liver*MultiScan*, exp 5 explored the effect of pineapple juice consumption on cT1, exp 6 the variability between T1 and cT1, and exp 7 the variability in the different cT1 processing methods. Paired t-tests, Bland-Altman analysis with mean difference and 95% limits of agreement (LoA), linear regression, and/or Pearson's correlation coefficient were used to assess the relationship between two acquisitions techniques, measurements, or interpretation methods. Repeatability was assessed using repeatability coefficient (RC). Intra- and inter-observer variability was assessed using intraclass correlation coefficient (ICC) and Bland-Altman plots.

Results:

Exp 1: In healthy volunteers (n = 4), there was no difference in *MRCP+* metrics between two different MRCP acquisition protocols. In patients with PSC (n = 6), there was a difference in the proportion of ducts with median diameter < 3mm (63% vs 83%, p = 0.0217) and 3–5mm (28% vs 16%, p = 0.0220).

Exp 2: For repeatability testing, biliary tree length (RC = 395.6mm) showed significant variability in healthy volunteers (n = 10) compared to other *MRCP+* metrics but had excellent intra-observer agreement (ICC = 0.96, 95% CI 0.88–0.99).

Exp 3: All segments of the biliary tree were visualised better on *MRCP+* than conventional MRCP images with numerically higher mean visualisation score, but only CBD (PSC: n = 20; p = 0.0020, healthy: n = 10; p = 0.0313) and LHD for PSC (p = 0.0117) reached statistical significance.

Exp 4: The inter-observer variability in biliary tree length (bias: -256.8 ± 413.4mm, 95% LoA -1067.0–553.4mm), and number of modelled ducts (bias: -19.9 ± 27.2, 95% LoA -73.2–33.4) between the two *MRCP+* analysts were judged to be clinically significant in the context of PSC (n = 30).

Exp 5: Mean cT1 measured before and after ingestion of pineapple juice did not differ significantly (n = 8; 772.8 ± 48.2ms vs 775.3 ± 51.5ms, p = 0.4623).

Exp 6: There was a difference between the paired measurements of mean T1 and cT1 in patients with PSC (n = 80; T1: 769.1ms vs cT1: 773.7ms, p < 0.0001) but no difference observed in healthy volunteers (n = 20; T1: 717.5ms vs cT1: 721.3ms, p = 0.2293).

Exp 7: There was a strong association between the ROI-placement and semi-automated multi-slice liver segmentation methods (n = 80; median cT1: r = 0.93; R^2 = 0.87; p < 0.0001; IQR cT1: r = 0.91, R^2 = 0.82; p < 0.0001). The association between automated single-slice and semi-automated multi-slice methods were also strong but weaker for IQR cT1 (n = 80; median cT1: r = 0.95, R^2 = 0.90, p < 0.0001; IQR cT1: r = 0.80, R^2 = 0.62, p < 0.0001).

Conclusions:

Exp 1: MRCP acquisition protocols and difference in acquisition resolution may have an impact on computation of *MRCP+* metrics, particularly when involving smaller calibre ducts.

Exp 2: The significant variability in the *MRCP+* biliary tree length with repeated acquisitions in healthy volunteers but excellent intra-observer agreement for tree length suggests the variability is due to raw MRI acquisition rather than image analysis factors. It emphasises the need to normalise whole tree metrics to the tree length for within-subject and between-subject comparisons.

Exp 3: Visualisation of the biliary tree is better with *MRCP+* than MRCP when MRI signal is present, particularly in the extrahepatic tree.

Exp 4: When the same scans were analysed by two *MRCP+* analysts, there was significant difference in the biliary tree length and number of ducts modelled in patients with PSC. The inter- and intra-observer agreement for *MRCP+* metrics were either good or excellent.

Exp 5: Liver cT1 measurement is not affected by pineapple juice consumption 20 minutes prior to scan.

Exp 6: Liver T1 and cT1 can be different in patients with PSC due to iron content, therefore cT1 is preferred for standardisation.

Exp 7: Liver segmentation method is preferred to ROI-placement method in PSC due to variability in IQR cT1 quantification between the two methods. Automated Liver*MultiScan* analysis is a feasible method to analyse large dataset despite only analysing a single representative slice of the liver.

3.2 Introduction

Quantitative techniques of the biliary tree have not been widely studied. Manual measurements of bile ducts are time consuming, susceptible to inter-observer variability, and are applied predominantly to the extrahepatic biliary tree. The *MRCP+* analysis method used in this thesis is novel. To the best of my knowledge, it was the only quantitative biliary analysis tool that was available to quantitatively assess the biliary tree at the time of writing. The effect of different acquisition protocols, the reliability of *MRCP+* metrics, and how it compared with conventional MRCP was unknown at the start of this study. Therefore, I investigated the (i) effect of two different MRCP acquisition protocols on *MRCP+* metrics; (ii) repeatability and intra-observer variability of *MRCP+* metrics in healthy volunteers; (iii) difference in visualisation quality between MRCP acquisition and *MRCP+* model; and (iv) inter- and intra-observer agreements of *MRCP+* metrics in PSC.

On the contrary, iron-corrected T1 (cT1) derived from Liver*MultiScan* analysis has gained momentum over the last seven years since its original description in 2014. The performance of cT1 has been evaluated in healthy volunteers and several chronic liver diseases in single-centre and multi-centre studies. The reproducibility of cT1 measurements have been reported in five studies and the coefficient of variation (CoV) for cT1 varied between 1.3% and 3.3%, with one study reporting an ICC of 0.93 between readers[123, 126, 135-137]. These studies included patients with mixed aetiology liver diseases and healthy volunteers. To this end, there has been no published studies examining the performance of cT1 specifically in a heterogeneous chronic liver disease such as PSC. Three different Liver*MultiScan* analysis techniques have been developed since the original ROI-placement method, but the utility of these techniques has not been examined specifically in PSC. Whilst liver iron content is a confounder in cT1 measurement, the role of correcting liver iron content in PSC is debatable compared to liver diseases known to be associated with iron accumulation in the liver such as non-alcoholic fatty liver disease (NAFLD) and alcohol-related liver disease (ArLD). Fluid

ingestion is another known confounder to cT1 measurement, and my study involves administration of pineapple juice prior to T1 mapping. Therefore, I evaluated the following: (i) the effect of pineapple juice consumption to optimise MRCP acquisition on cT1 measurement; (ii) the need for iron correction method in PSC; and (iii) the variability in cT1 measurements across three different data processing methods.

3.3 Variability in *MRCP+* metrics between methods of MRCP acquisition protocols

3.3.1 Background, aim and hypothesis

MRCP acquisition protocols are usually optimised by scanner manufacturers and the sequences are loaded onto scanner consoles. Two different MRCP acquisition protocols were available at the conception of this study. The aim of this experiment was to investigate whether there was any variability in quantification of biliary metrics between the two acquisition protocols. The hypothesis was that there would be no difference in *MRCP+*-derived metrics between the two protocols.

3.3.2 Methods

Protocol 1 was optimised and standardised for use in all MRI scanner types and field strength by our collaborator, Perspectum, who developed the *MRCP+* software. The parameters of this protocol were summarised in chapter 2. Protocol 2 was optimised by Siemens, the MRI scanner manufacturer. The Siemens SPACE sequence parameters were FOV 380 x 380mm, acquisition matrix 384 x 384mm, voxel size 0.5 x 0.5 x 1.0mm, 1mm slice thickness, in-plane image resolution 1.0 x 1.0mm, 64 slices, TE 701ms, TR 2500ms (variable depending on respiratory rate), 140° flip angle, GRAPPA acceleration factor 3 with 24 reference lines. Both protocols were acquired in the first ten MOBILISE study participants. Protocol 1 (Perspectum) was acquired first, followed immediately by protocol 2 (Siemens) without moving the patient out of the scanner. *MRCP+* analysis was performed on both sets of acquisitions. Tree metrics (gallbladder volume, biliary tree volume, proportion of

ducts with median diameter < 3mm, 3–5mm, and > 5mm) and individual duct metrics (maximum and minimum diameter of CBD, RHD, and LHD) were compared between the two protocols. Paired *MRCP+* metrics were compared as a combined cohort and separately for healthy volunteers and patients with PSC.

3.3.3 Results

Four healthy volunteers and six patients with PSC were included. The average acquisition time for protocol 1 was 5 minutes and for protocol 2 was 9 minutes. Variability in the tree and individual duct metrics from both protocols for all participants are summarised in Table 7. There was a significant difference in the biliary tree volume (4.5ml vs 3.7ml, p = 0.0117), proportion of ducts with median diameter < 3mm (78% vs 83%, p = 0.0205), and proportion of ducts with median diameter 3–5mm (19% vs 15%, p = 0.0235) between protocol 1 (Perspectum) and protocol 2 (Siemens). For individual duct metrics, there was no difference in the maximum and minimum diameters of either CBD, RHD or LHD.

When analysed separately, there were no differences in either the tree metrics or individual duct metrics between the two MRCP protocols in healthy volunteers as summarised in Table 8. In patients with PSC, however, there were borderline differences in the biliary tree volume (13.7ml vs 10.4ml, p = 0.0492) and the proportion of ducts with median diameter < 3mm (63% vs 83%, p = 0.0217). The proportion of ducts with median diameter 3–5mm (28% vs 16%, p = 0.0220) was significantly higher in protocol 1 (Perspectum) compared to protocol 2 (Siemens). However, all measurements were within the 95% limits of agreements from the mean differences as shown in the Bland-Altman plots in Figure 10. There was no difference in the proportion of ducts with median diameter > 5mm (9% vs 6%, p = 0.2181). The difference in the proportion of ducts < 3mm and 3–5mm was not observed in healthy volunteers, despite the higher relative frequency of these ducts in healthy volunteers (<3mm: 84% vs 90%, p = 0.4634; 3–5mm: 16% vs 9%, p = 0.2864).

Table 7. Tree and individual duct metrics variability between protocol 1 (Perspectum) and protocol 2 (Siemens) in 10 participants.

	All participants (n = 10)				
	Protocol 1 (Perspectum)	Protocol 2 (Siemens)	p-value	Bias (SD)	95% LoA
Tree metrics					
Biliary tree volume (ml)	4.5 (2.9–17.0)	3.7 (2.0–12.2)	0.0117	2.1 (2.7)	-3.2–7.4
Gallbladder volume† (ml)	21.0 (16.1–35.3)	23.7 (15.3–34.7)	0.2188	0.7 (1.9)	-2.9–4.4
Proportion of ducts median diameter <3mm (%)	78.0 (47.3–82.8)	83.0 (79.0–92.5)	0.0205	-15.2 (17.1)	-48.8–18.4
Proportion of ducts median diameter 3–5mm (%)	19.0 (14.5–38.3)	15.0 (3.8–18.0)	0.0235	12.6 (14.6)	-16.1–41.3
Proportion of ducts median diameter >5mm (%)	5.5 (1.5-15.0)	1.0 (0.0–7.8)	0.1351	3.1 (6.0)	-8.6–14.8
Individual duct metrics					
CBD diameter (mm)					
Maximum	5.8 (5.1–8.4)	5.0 (3.9–8.1)	0.1395	0.4 (0.8)	-1.1–1.9
Minimum	2.9 (2.5–3.5)	2.2(1.9–3.0)	0.2266	0.4 (0.8)	-1.2–1.9
Median	3.8 (3.6–5.2)	3.6 (3.1–5.5)	0.1055	0.3 (0.4)	-0.6–1.1
RHD diameter (mm)					
Maximum	4.0 (3.1–7.0)	4.4 (2.9–6.1)	0.7773	0.1 (1.4)	-2.5–2.8
Minimum	2.2 (1.8–3.9)	2.2 (1.6–3.9)	0.4844	0.1 (0.3)	-0.4–0.6
Median	2.6 (2.4–5.1)	3.1 (2.3–5.1)	0.6406	-0.1 (0.4)	-0.8 (0.7)
LHD diameter (mm)					
Maximum	4.3 (3.4–6.7)	3.6 (3.3–5.7)	0.1992	0.9 (1.6)	-2.3–4.1
Minimum	2.1 (1.7–4.0)	2.0 (1.5–3.8)	0.5078	0.1 (0.4)	-0.6–0.8
Median	3.4 (2.7–4.9)	3.1 (2.4–4.9)	0.6094	0.1 (0.5)	-0.9–1.0

Values are presented as median (interquartile range). †Three patients with PSC had a cholecystectomy. SD, standard deviation; LoA, limits of agreement.

Table 8. Tree metrics variability between protocol 1 (Perspectum) and protocol 2 (Siemens) stratified by cohort.

	Healthy (n = 4)					PSC (n = 6)				
	Protocol 1 (Perspectum)	Protocol 2 (Siemens)	p-value	Bias (SD)	95% LoA	Protocol 1 (Perspectum)	Protocol 2 (Siemens)	p-value	Bias (SD)	95% LoA
Biliary tree volume (ml)	2.9 (1.6–3.7)	2.2 (1.4–3.4)	0.1337	0.4 (0.4)	-0.4–1.2	12.6 (4.5–22.0)	10.4 (3.2–16.1)	0.0492	3.2 (3.0)	-2.7–9.1
Gallbladder volume† (ml)	17.5 (10.6–23.0)	16.6 (9.5–22.4)	0.0465	0.9 (0.4)	0.2–1.6	35.3 (21.0–89.7)	34.7 (23.7–86.0)	0.8000	0.5 (3.2)	-5.7–6.8
Proportion of ducts median diameter <3mm (%)	84.5 (75.0–90.3)	89.5 (79.0–100.0)	0.5149	-6.3 (17.0)	-39.5–27.0	63.0 (37.8–79.3)	83.0 (74.3–85.5)	0.0217	-21.2 (15.8)	-52.0–9.7
Proportion of ducts median diameter 3–5mm (%)	16.0 (6.3–19.8)	8.0 (0.0–19.8)	0.5579	4.8 (14.5)	-23.6–33.1	28.0 (17.3–46.5)	15.5 (6.5–18.8)	0.0220	17.8 (13.3)	-8.3–44.0
Proportion of ducts median diameter >5mm (%)	2.0 (0.0–6.3)	0.0 (0.0–3.8)	0.4765	1.5 (3.7)	-5.7–8.7	10.5 (4.3–15.5)	4.5 (0.0–11.5)	0.2182	4.2 (7.3)	-10.0–18.4

†Three patients with PSC had a cholecystectomy. Values presented as median (interquartile range). SD, standard deviation; LoA, limits of agreement.

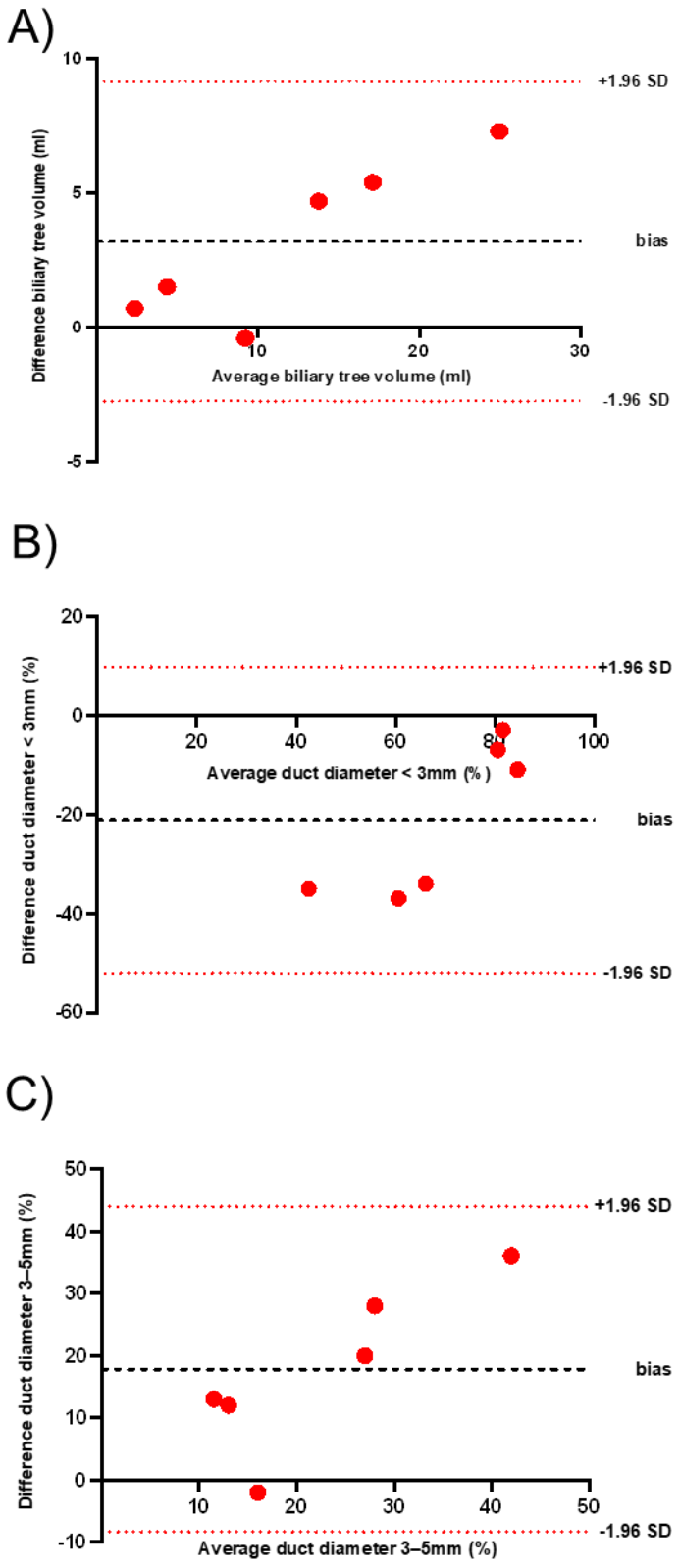


Figure 10. Bland-Altman plots in 6 patients with PSC that showed a difference between two MRCP acquisition protocols. (A) Biliary tree volume, (B) proportion of ducts with median diameter < 3mm, and (C) proportion of ducts with median diameter 3–5mm.

3.3.4 Discussion

In this method comparison study, the differences could be due to inherent variability in each of the acquisition methods (systematic bias) and/or bias between the measurements from both methods (measurement errors). Protocol 2 (Siemens) had additional slice acquisitions with thinner slices compared to protocol 1 (Perspectum), therefore technically a slightly better image resolution with higher signal-to-noise ratio. This may explain why protocol 2 contains a higher proportion of MRCP signals from smaller ducts compared to protocol 1, albeit this experiment was performed in a small sample size. This relative advantage in image resolution may be lost with increasing duct diameter. The gallbladder and extrahepatic ducts have defined anatomical margins and can therefore be directly compared between two acquisitions regardless of the MRCP protocol used. When comparing metrics relating to the whole tree, however, this experiment has shown that the proportion of ducts modelled on *MRCP+* maybe influenced by the scan parameters at acquisition. The key message that was learnt early on through this experiment was that it was important to quantify the length of the biliary tree and number of bile ducts modelled with *MRCP+* to enable accurate interpretation of metrics relating to the whole tree within and between study participants. For the MoBILISE study, I chose protocol 1 (Perspectum) as the standardised method for MRCP acquisition given that the acquisition time is almost halved compared to protocol 2, considering two MRCP images will be required in each study participant. Furthermore, protocol 1 has been standardised for use in different MRI scanners and field strengths by Perspectum. This would allow direct comparisons of results from this study to be made with other datasets acquired on different scanners and/or field strengths.

3.4 Repeatability and intra-observer reliability of *MRCP+* metrics in healthy volunteers

3.4.1 Background, aim and hypothesis

At the time of conception of this study, repeatability and intra-observer variability data was not available for *MRCP+* metrics. This experiment was designed to investigate both in healthy volunteers. Repeatability was defined as the precision of *MRCP+* measurements in the same scanner by the same operator. Intra-observer reliability was defined as the measurement of the degree of agreement by the same *MRCP+* analyst. The hypothesis was that there will be good repeatability and intra-observer reliability of biliary metrics.

3.4.2 Methods

Ten healthy volunteers were first scanned using Perspectum's MRCP protocol. Then, the volunteers were asked to get off the scanner for five minutes before getting back onto the scanner to be re-scanned by the same operator. Both sets of MRCP data were analysed with *MRCP+* for whole tree metrics and individual duct metrics. In addition to the metrics described in the above experiment in section 3.3, the length of the biliary tree and number of bile ducts modelled on *MRCP+* were now available on the software for this experiment. Repeatability was calculated as the difference in *MRCP+* metrics between the two scans on the same subject. The mean difference (bias), standard deviation, and 95% limits of agreement (LoA) from the Bland-Altman analysis were recorded. The repeatability coefficient (RC) is the threshold under which the difference between any two repeat measurements on the same participant acquired under identical conditions should fall with 95% probability[138]. In other words, the maximum difference that will be exceeded by only 5% of measurement pairs. It is interpreted in the same unit as the measurements. RC was calculated using the standard formula:[139]

$$RC = 1.96 \times \sqrt{2} \times \text{standard deviation within-participants } (SD_{\omega}) = 2.77 \times SD_{\omega}$$

Unlike CoV, RC can be interpreted in the same unit of measurement and does not rely on having a parametric distribution. For the intra-observer variability analysis, the first set of MRCP acquisitions were re-analysed with *MRCP+* a week later. Intra-observer variability was assessed using intraclass correlation coefficient (ICC) and Bland-Altman analysis of biliary metrics from the temporal analysis. ICC was defined as excellent (>0.90), good (0.75–0.90), moderate (0.50–0.75), and poor (<0.50)[138].

3.4.3 Results

Repeatability results for tree metrics are summarised in Table 9 and for individual duct metrics in

Table *10*. The most striking finding from this analysis is the high mean difference, wide 95% LoA, and high RC for biliary tree length (bias 34.1 ± 212.8mm, 95% LoA -382.9–451.2mm, RC = 395.6mm). Figure 11 shows the Bland-Altman plot for difference in biliary tree length between the two acquisitions. Repeatability for volume metrics were good with a tight 95% LoA and for the classification of duct diameter proportions, they were moderate with a wider 95% LoA. The repeatability and 95% LoA for maximum and minimum diameter measurements of extrahepatic bile ducts were good.

The intra-observer variability results are reported in Table 11. Bland-Altman plots for the tree metrics are presented in Figure 12. Overall, the mean difference was not significantly different from zero, indicating good intra-observer agreements for both the whole tree and individual duct *MRCP+* metrics. The ICCs (0.90–1.00) were excellent for all metrics. Gallbladder volume showed perfect intra-observer agreement.

Table 9. Repeatability of tree metrics in 10 healthy volunteers.

MRCP+ tree metric	Bias (SD)	95% LoA	RC
Biliary tree length (mm)	34.1 (212.8)	-382.9–451.2	±395.6
Gallbladder volume (ml)	1.2 (2.5)	-3.7–6.1	±5.2
Biliary tree volume (ml)	0.2 (1.1)	-1.9–2.3	±2.0
Number of bile ducts	0.8 (7.4)	-13.7–15.2	±13.6
Median diameter < 3mm (%)	0.1 (8.0)	-15.6–15.8	±14.7
Median diameter 3–5mm (%)	1.1 (7.5)	-13.6–15.8	±13.9
Median diameter > 5mm (%)	-1.2 (4.7)	-10.4–7.9	±12.5

SD, standard deviation; LoA, limits of agreement; RC, repeatability coefficient

Table 10. Repeatability of extrahepatic bile duct diameter measurements in 10 healthy volunteers.

Bile duct	*MRCP*+ diameter (mm)	Bias (SD)	95% LoA	RC
CBD	Maximum	0.1 (0.8)	-1.5–1.8	1.6
	Minimum	0.4 (1.0)	-1.6–2.4	2.0
CHD	Maximum	-0.3 (0.8)	-1.8–1.2	1.5
	Minimum	0.2 (1.2)	-2.2–2.6	2.3
RHD	Maximum	0.4 (0.6)	-0.8–1.6	1.4
	Minimum	0.0 (0.5)	-1.0–1.0	0.9
LHD	Maximum	-0.2 (0.7)	-1.6–1.3	1.3
	Minimum	0.0 (0.5)	-0.5–0.5	0.5

SD, standard deviation; LoA, limits of agreement; RC, repeatability coefficient

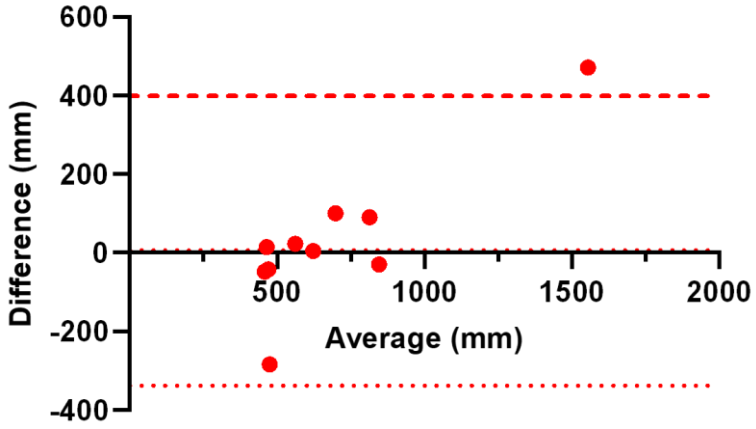


Figure 11. Bland-Altman plot for biliary tree length modelled between the first and second MRCP acquisitions. Mean difference 34.1 ± 212.8mm, 95% limits of agreement -382.9–451.2mm, repeatability coefficient = 395.6mm.

Table 11. Intra-observer variability of *MRCP*+ metrics assessed by Bland-Altman analyses and intraclass correlation coefficients in 10 healthy volunteers.

MRCP+ metrics	Bias (SD)	95% LoA	ICC	95% CI
Whole tree metrics				
Gallbladder volume (ml)	0.0 (0.0)	0.0–0.0	1.00	1.00–1.00
Biliary tree length (mm)	-95.9 (54.6)	-202.8–11.1	0.96	0.88–0.99
Number of bile ducts	-3.4 (2.6)	-8.5–1.7	0.95	0.85–0.99
Biliary tree volume (ml)	-0.7 (0.6)	-1.9–0.4	0.96	0.89–0.99
Median diameter < 3mm (%)	-0.2 (3.6)	-7.3–6.9	0.98	0.90–0.99
Median diameter 3–5mm (%)	0.3 (3.3)	-6.2–6.8	0.98	0.92–1.00
Median diameter > 5mm (%)	0.0 (0.5)	-0.9–1.0	0.99	0.97–1.00
Individual duct metrics				
CBD maximum diameter (mm)	-0.2 (0.5)	-1.1–0.7	0.97	0.90–0.99
CBD minimum diameter (mm)	-0.2 (0.6)	-1.4–1.1	0.90	0.86–0.95
CHD maximum diameter (mm)	-0.4 (0.6)	-1.5–0.8	0.96	0.83–0.99
CHD minimum diameter (mm)	-0.2 (0.4)	-1.0–0.7	0.97	0.88–0.99
RHD maximum diameter (mm)	-0.2 (0.3)	-0.8–0.5	0.97	0.89–0.99
RHD minimum diameter (mm)	-0.2 (0.3)	-0.8–0.4	0.97	0.84–0.99
LHD maximum diameter (mm)	0.4 (0.7)	-0.9–1.0	0.85	0.42–0.96
LHD minimum diameter (mm)	-0.1 (0.4)	-0.9–0.7	0.94	0.80–0.99

SD; standard deviation, LoA; limits of agreement, ICC; intraclass correlation coefficient.

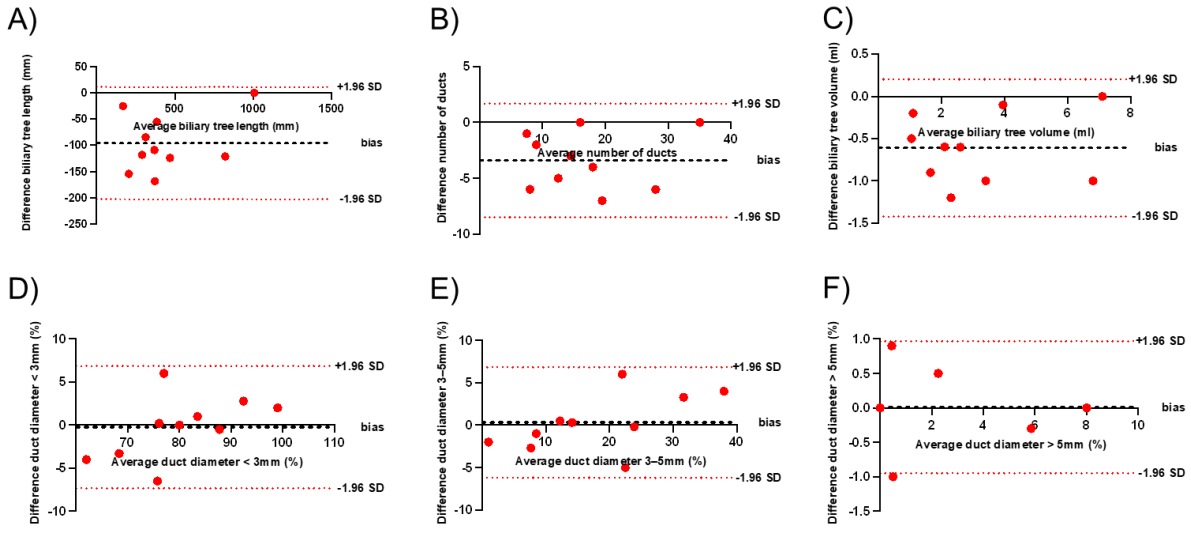


Figure 12. Bland-Altman plots of intra-observer variability in whole tree and individual duct *MRCP*+ metrics in 10 healthy volunteers. (A) Biliary tree length; (B) number of bile ducts modelled; (C) biliary tree volume; (D-F) proportion of ducts with median diameter < 3mm, 3–5mm, and > 5mm, respectively.

3.4.4 Discussion

This experiment has demonstrated that the repeatability for biliary tree length is not reliable in healthy volunteers. The agreement for one of the volunteers was outside the limits of agreement and occurred at the higher length of the biliary tree. This could be clinically significant given that the length of the biliary tree imaged on MRCP in patients with PSC are expected to be higher if they have significant multifocal distal strictures causing proximal tree dilatation. A difference of up to 40cm of biliary tree model within the same individual suggests that *MRCP+* is sensitive to how much of the raw biliary tree data from MRCP acquisition is 'apparent' to the algorithm. One of the main reasons for this discrepancy is the acquisition quality which is dependent on participant factors (breathing pattern, artifact from adjacent structures) and technical factors (planning of acquisition matrix, placement of navigator). Whilst the agreement for extrahepatic ducts metrics were good, there was greater variability in the minimum diameter of ducts than maximum diameter measurements. This is expected as minimum measurements are more sensitive to signal loss from poor MRCP image resolution and artifacts. This may have implications in quantifying strictures in PSC.

When MRCP data for the same subject was re-analysed by the same *MRCP+* analyst, there was good agreement for all metrics. However, there was greater variability in the biliary tree length and number of ducts modelled compared to the other metrics. This effect is likely to be amplified when a larger extent of the biliary tree is acquired on MRCP either due to better acquisition techniques or the disease state itself (e.g., higher signal intensities in the smaller intrahepatic bile ducts due to downstream tight strictures).

Data from this experiment led me to consider normalising the tree metrics to the length of their corresponding biliary tree, particularly when comparing data between participants. This may be particularly important in stricturing diseases of the biliary tree such as PSC where the tree length 'apparent' to the algorithm may be expected to be variable depending on severity of downstream strictures that would make proximal ducts more visible.

Since the conception of this study, Perspectum had conducted and published their own internal performance testing data of *MRCP+* on 20 healthy volunteers and 20 self-reported mixed aetiology liver diseases (10 'biliary disease' and 10 'liver disease' patients)[140]. Their repeatability and intra-observer variability data were either comparable or slightly worse compared to results obtained here. However, data on biliary tree length and number of ducts modelled were not published to compare and comment.

3.5 Variability in biliary tree visualisation between MRCP and *MRCP+*

3.5.1 Background, aim and hypothesis

Early work revealed occurrence of occasional discontinuity in bile ducts in the *MRCP+* models. This was hypothesised to be due to poor quality MRCP acquisition or presence of artifact rather than algorithmic failure of modelling the duct. To investigate this, I felt that it would be important to perform direct comparisons of *MRCP+* models and MRCP acquisitions, the currently accepted gold standard. This experiment investigated whether there were differences in the quality of the biliary tree modelled with *MRCP+* compared to conventional MRCP acquisitions.

3.5.2 Methods

The biliary tree was divided into three segments: (1) extrahepatic (CBD, CHD, junction of CBD and CHD); (2) hilar region (hilum bifurcation, main left hepatic duct, main right hepatic duct); and (3) intrahepatic (1st and 2nd bifurcation of the main right and left hepatic ducts) as shown in Figure 13. Each segment of the biliary tree on *MRCP+* and conventional MRCP was scored according to an ordinal scale of increasing visibility of bile ducts: 0 – not seen; 1 – faint; 2 – adequate; 3 – good; and 4 – excellent with no duct discontinuity. Duct discontinuity for each biliary segment on *MRCP+* was assessed as a binary classification: 0 – absent and 1 – present. For the conventional MRCP scoring, I reviewed the continuous image series, and the final maximum intensity projection (MIP) image. For *MRCP+* scoring, I interrogated the final 3D model on the dedicated *MRCP+* analysis interface. Formal blinding of data was not performed. However, consecutive MRCP and *MRCP+* interpretation from the

same subject was avoided to minimise reporting bias. The mean visualisation scores obtained for each biliary segment from both MRCP and *MRCP+* were compared in the first 30 participants of the MoBILISE study.

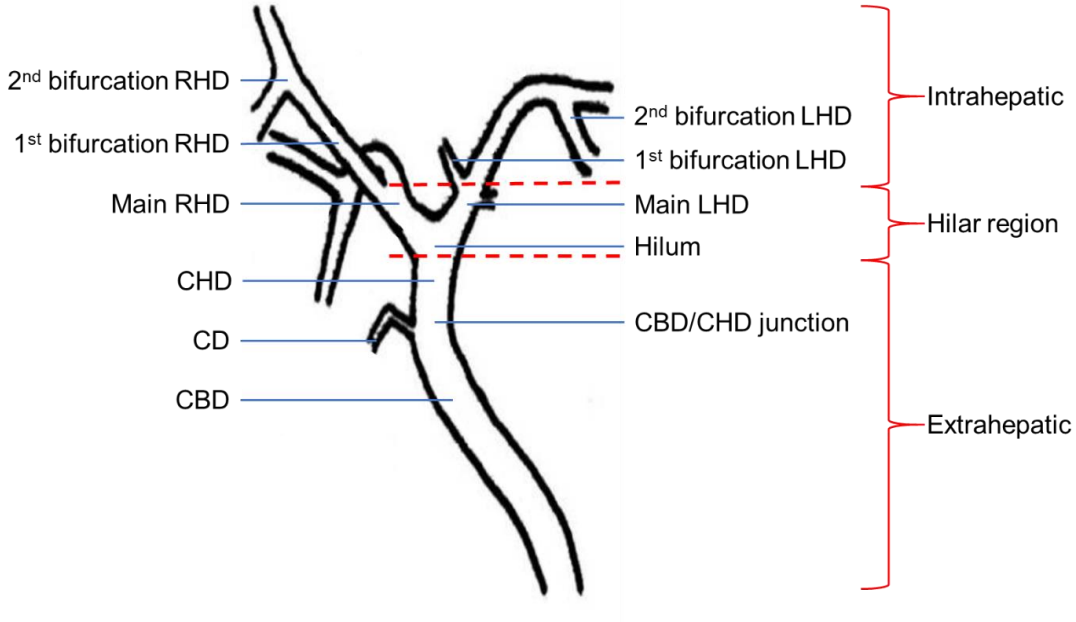


Figure 13. Schematic diagram of the biliary tree divided into extrahepatic, hilar, and intrahepatic segments. CBD; common bile duct, CHD; common hepatic duct; CD; cystic duct, RHD; right hepatic duct, LHD; left hepatic duct

3.5.3 Results

There were 20 patients with PSC and 10 healthy volunteers. One patient with PSC and one healthy volunteer were excluded from analysis due to poor quality MRCP acquisition from motion artifact (both were excluded from the final study analysis). All segments of the biliary tree were visualised better on *MRCP+* than MRCP with numerically higher mean visualisation score values in patients with PSC and healthy volunteers as shown in Figure 14. However, statistical significance was reached only for CBD (PSC: $p = 0.0020$, healthy: $p = 0.0313$) and LHD for PSC ($p = 0.0117$). Mean visualisation scores showed a decreasing trend moving proximally from the extrahepatic segments into the intrahepatic segments in both MRCP and *MRCP+*. The proportion of tree modelled without a break was higher in the extrahepatic than intrahepatic segments as summarised in Table 12.

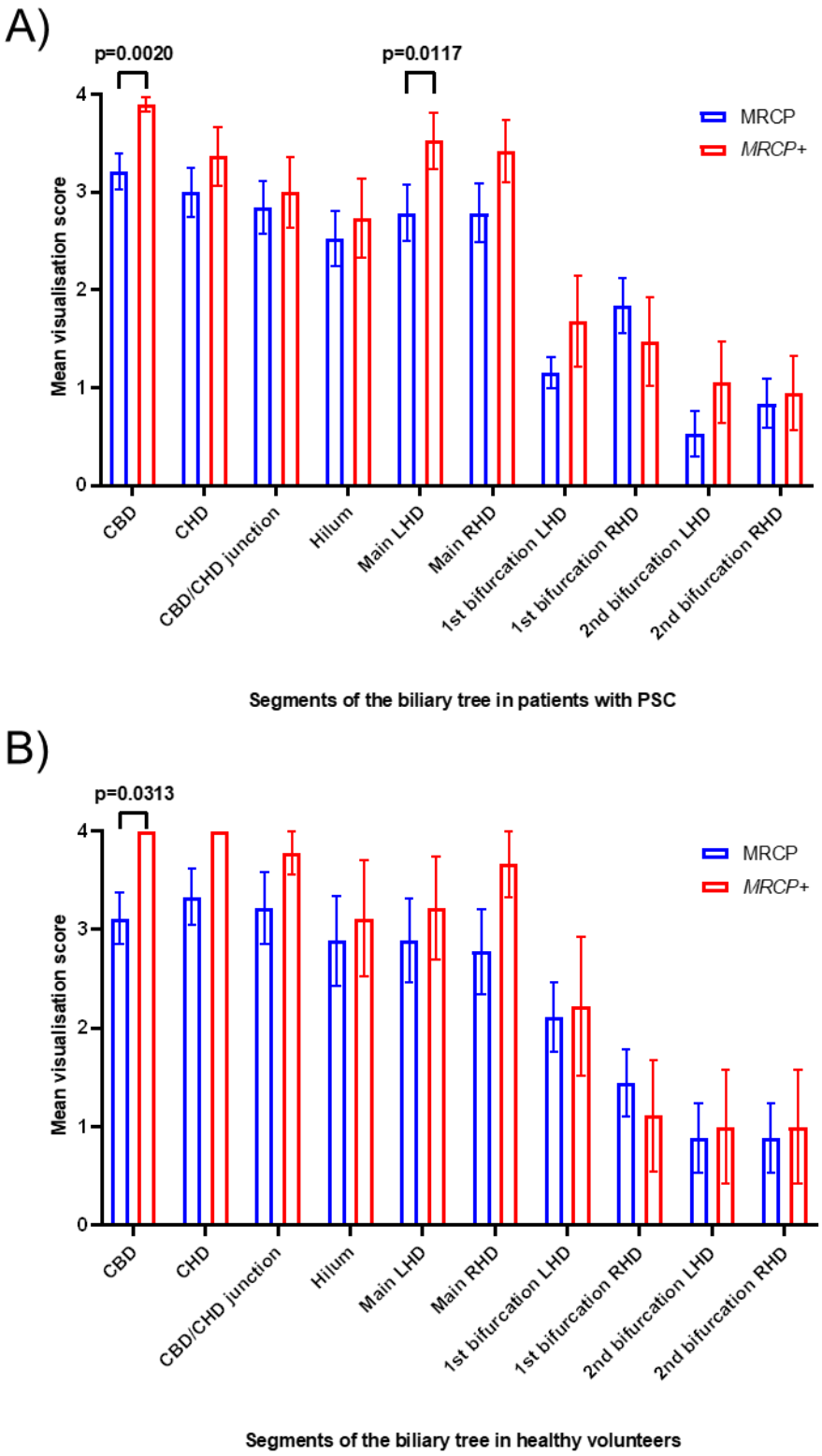


Figure 14. Mean ± SEM visualisation score comparison between MRCP and *MRCP+* in (A) patients with PSC (n = 19), and (B) healthy volunteers (n = 9). Scores ranging from 0 (not seen) to 4 (excellent visualisation) assessed for each segment of the biliary tree. A significant difference in score is highlighted with a p-value.

Table 12. The proportion of the biliary tree segment modelled without a break in duct continuity in patients with PSC and healthy volunteers.

	Proportion without a break in duct continuity (%)	
Biliary tree segment	PSC (n = 19)	Healthy (n = 9)
Extrahepatic		
CBD	100	100
CHD	84	100
CBD/CHD junction	74	89
Hilar region		
Hilum	68	78
Main LHD	89	78
Main RHD	74	89
Intrahepatic		
LHD 1st bifurcation	42	56
RHD 1st bifurcation	37	22
LHD 2nd bifurcation	26	22
RHD 2nd bifurcation	26	22

3.5.4 Discussion

When MRI signal was present, visualisation of bile ducts was better with *MRCP+* than conventional MRCP when reviewed by a non-radiologist, although statistically significance was reached only for the visualisation of the CBD. This difference was more prominent in the extrahepatic than intrahepatic segment of the biliary tree. Apart from the CBD, all other biliary segments were modelled with a certain degree of discontinuity in duct profile. When there was MRI signal loss causing discontinuity in bile ducts in the original MRCP acquisition, *MRCP+* accurately reflects these as breaks in the model. Signal void on MRCP acquisition can occur due to several factors such as (i) anatomical (e.g., extrinsic vascular compression, overlapping intestinal fluid); (ii) technical (e.g., poor acquisition from motion artifact, maximum intensity projection (MIP) reconstruction artifact); and/or (iii) true pathology (e.g., stricture, calculi, pneumobilia). Discontinuity in ducts is more likely to affect modelling of small diameter ducts and strictures. This should be kept in mind when interpreting tree stricture metrics from *MRCP+*, particularly in PSC.

3.6 Inter-observer and intra-observer variability in *MRCP*+ metrics in PSC

3.6.1 Background, aim and hypothesis

The internal performance testing by Perspectum reported the inter- and intra-observer variability in a heterogeneous group of patients with self-reported 'biliary disease' and 'liver disease'[140]. In the biliary disease group, there were six self-reported PSC with subtype unknown and four PBC. Results were pooled and reported as a combined cohort with the liver disease group. The aim of this experiment was to investigate the intra- and inter-observer variability in *MRCP*+ metrics specifically in a well-characterised cohort of patients with large-duct PSC. The agreement for *MRCP*+ metrics were hypothesised to be good.

3.6.2 Methods

30 MRCP acquisitions from the MoBILISE study were analysed by two trained *MRCP*+ analysts independently (M.M. radiographer for 1 year, *MRCP*+ experience 1 year; C.F. abdominal radiographer for 5 years, *MRCP*+ experience 2 years). Checks were performed to ensure that both analysts were analysing the same scan series. The scans were re-analysed by M.M after a week for intra-observer testing. Inter-observer and intra-observer variability of *MRCP*+ metrics in PSC were assessed by intraclass correlation coefficient (ICC) based on a two-way mixed-effects model and strength classified as previously reported[138].

3.6.3 Results

Despite an excellent intraclass correlation (ICC = 0.95), there was a significant difference in the length of biliary tree modelled between the two *MRCP*+ analysts (p = 0.002), therefore strictly inter-observer agreement should not be assessed. There was no significant difference in the intra-observer variability (p = 0.057) for tree length, albeit with a borderline p value. This suggests that there may also be a significant difference in tree length when *MRCP*+ analysis was performed by the same analyst a week apart. Biliary tree length mean difference ± standard deviation of -256.8 ± 413.4mm (95% LoA -1067.0–553.4mm), and number of modelled ducts mean difference ± standard deviation of -19.9 ± 27.2 (95% LoA -

73.2–33.4) between the two analysts were judged to be clinically significant in the context of PSC. Results are summarised in Figure 15 and Table 13.

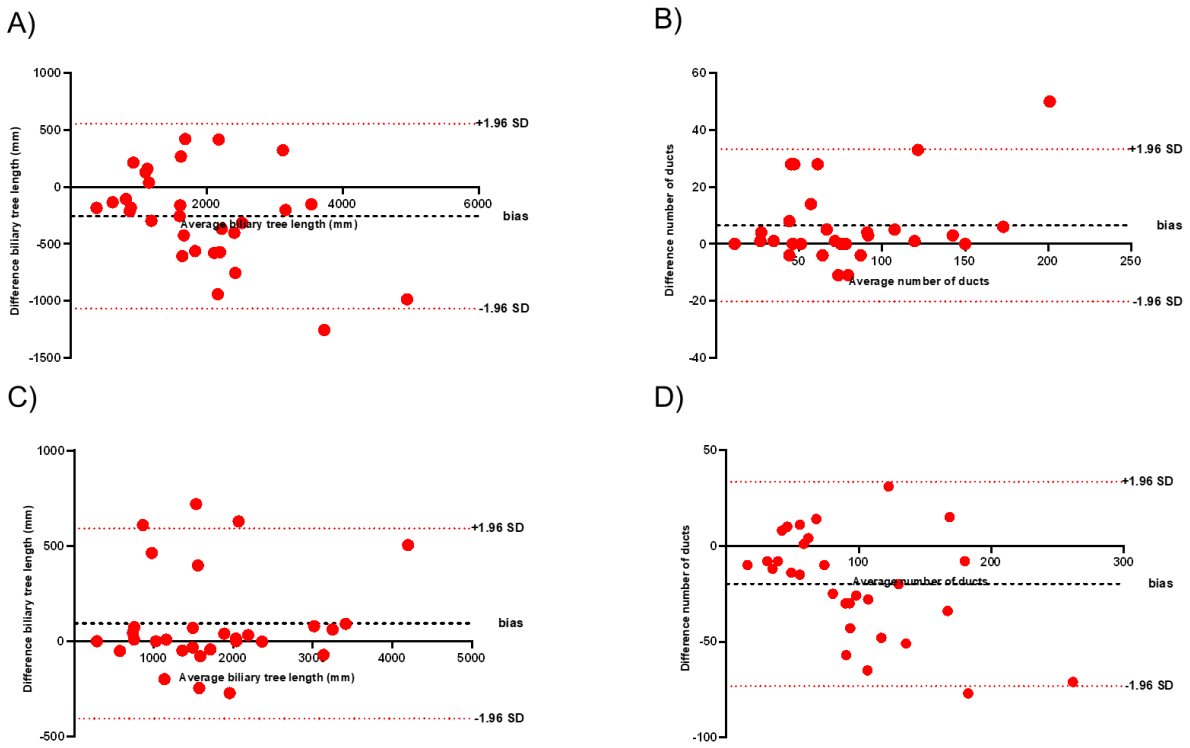


Figure 15. Inter-observer (A-B) and intra-observer (C-D) variability in biliary tree length and number of bile ducts modelled by *MRCP+* in 30 patients with large-duct PSC.

Table 13. Intra- and inter-observer variability of *MRCP+* metrics assessed by Bland-Altman analyses and intraclass correlation coefficients in 30 patients with large-duct PSC.

	Inter-observer variability				Intra-observer variability			
MRCP+ metrics	Bias (SD)	95% LoA	ICC	95% CI	Bias (SD)	95% LoA	ICC	95% CI
Whole tree metrics								
Biliary tree length	-256.8 (413.4)	-1067.0–553.4	0.95	(0.83–0.98)	93.4 (254.4)	-405.2–591.9	0.98	(0.96–0.99)
Number of bile ducts	-19.9 (27.2)	-73.2–33.4	0.91	(0.68–0.97)	6.5 (13.7)	-20.2–33.2	0.98	(0.94–0.99)
Gallbladder volume	0.3 (2.7)	-4.9–5.6	1.00	(1.00–1.00)	0.2 (0.7)	-1.2–1.6	1.00	(1.00–1.00)
Biliary tree volume	-2.2 (5.6)	-13.1–8.8	0.93	(0.84–0.97)	0.5 (3.0)	-5.4–6.4	0.98	(0.95–0.99)
Median diameter < 3mm (%)	-4.1 (5.0)	-13.9–5.6	0.88	(0.50–0.96)	0.3 (3.9)	-7.3–8.0	0.97	(0.93–0.98)
Median diameter 3–5mm (%)	3.2 (4.6)	-5.8–12.3	0.95	(0.88–0.98)	-0.1 (3.4)	-6.9–6.6	0.98	(0.97–0.99)
Median diameter 5–7mm (%)	0.4 (0.9)	-1.4–2.2	0.93	(0.82–0.97)	-0.2 (0.9)	-2.0–1.7	0.95	(0.89–0.97)
Median diameter > 7mm (%)	0.4 (1.0)	-1.5–2.4	0.95	(0.88–0.98)	0.0 (0.6)	-1.2–1.2	0.99	(0.97–0.99)
Individual duct metrics								
CBD maximum diameter (mm)	0.1 (0.9)	-1.6–1.8	0.98	(0.96–0.99)	0.2 (0.7)	-1.1–1.5	0.99	(0.98–1.00)
CBD minimum diameter (mm)	0.6 (1.7)	-2.6–3.9	0.68	(0.34–0.85)	0.6 (1.3)	-2.0–3.2	0.85	(0.67–0.93)
CHD maximum diameter (mm)	0.4 (0.8)	-1.3–2.0	0.98	(0.95–0.99)	0.2 (0.9)	-1.6–2.0	0.99	(0.97–1.00)
CHD minimum diameter (mm)	-0.3 (1.8)	-3.8–3.3	0.87	(0.73–0.94)	-0.2 (1.8)	-3.7–3.4	0.88	(0.73–0.94)
RHD maximum diameter (mm)	0.3 (1.9)	-3.5–4.1	0.81	(0.51–0.91)	0.6 (1.7)	-2.6–3.9	0.81	(0.56–0.92)
RHD minimum diameter (mm)	0.0 (0.6)	-1.2–1.2	0.94	(0.87–0.98)	0.1 (0.6)	-1.1–1.3	0.94	(0.86–0.97)
LHD maximum diameter (mm)	-0.5 (2.0)	-4.4–3.4	0.82	(0.62–0.92)	0.2 (1.1)	-1.9–2.3	0.94	(0.87–0.97)
LHD minimum diameter (mm)	-0.1 (0.6)	-1.3–1.1	0.98	(0.95–0.99)	0.2 (0.7)	-1.2–1.6	0.97	(0.94–0.99)

CBD, common bile duct; CHD, common hepatic duct; ICC, intraclass correlation coefficient; LHD, main left hepatic duct; LoA, limits of agreement; RHD, main right hepatic duct; SD, standard deviation.

3.6.4 Discussion

This difference in variable proportion of the biliary tree modelled by *MRCP*+ needs to be considered when comparing metrics between study participants. Whilst the difference in biliary tree lengths does not affect comparison of metric related to a defined anatomical duct (e.g., maximum diameter of CBD, difference in CHD minimum diameter over two time points), it is likely to affect metrics that are related to the whole tree (e.g., biliary tree volume, number of candidate strictures and dilatations). Although a standardised processing and reporting steps were followed for *MRCP*+ analysis, this difference may be due to the diligence and experience of each analyst in thresholding and modelling the same biliary tree, particularly in cases where the acquisition quality was not optimum. This data supports the reason for choosing to normalise all metrics relating to the whole tree before comparisons were made between participants and within the same participant at different time points.

The intra- and inter-observer agreement for the other tree metrics and individual duct metrics were deemed to be clinically acceptable. The 95% LoA was wider for the minimum diameter measurements of CBD and CHD in both the intra- and inter-observer analysis, suggesting that minimum diameter measurements within the extrahepatic tree could be more prone to signal loss from artifact and/or gastroduodenal noise contamination. The distal CBD has traditionally thought to be rather ambiguous on MRCP and perhaps will continue to remain so until better MRCP techniques and image resolution can be achieved.

3.7 Effect of pineapple juice consumption on liver cT1 measurements

3.7.1 Background, aim and hypothesis

Food and liquid consumptions have a confounding effect on T1 measurements[141]. In the MoBILISE study, participants were given 200ml of pineapple juice prior to entering the scanner to optimise the acquisition of 3D MRCP. The effect of pineapple juice on cT1 measurements is unknown. A study on 90 healthy individuals found that the mean ± standard deviation of half gastric emptying time and the time of fastest gastric emptying of a fluid test meal was 80.5 ± 22.1 minutes and 40.3 ± 10.2 minutes, respectively[142]. The liver MRI sequence was performed before the 3D MRCP sequence and was completed on average within 20 minutes of pineapple juice ingestion. Therefore, the hypothesis was that 200ml of pineapple juice ingested immediately before the MRI will not have a significant effect on cT1 measurements as the MRI data would be acquired before the liquid is absorbed into the portal blood.

3.7.2 Methods

T1 mapping was first acquired in eight healthy volunteers after a period of 4-hour fast. They were asked to come off the scanner and consume 200ml of pineapple juice before returning to the scanner. MRCP data was acquired first before T1 mapping to mimic MoBILISE study protocol. T1 data were analysed using Liver*MultiScan* analysis. Paired measurements of mean cT1 were compared using t-test.

3.7.3 Results

Mean cT1 ± standard deviation measured before and after ingestion of pineapple juice did not differ significantly (772.8 ± 48.2ms vs 775.3 ± 51.5ms, p = 0.4623; Figure 16).

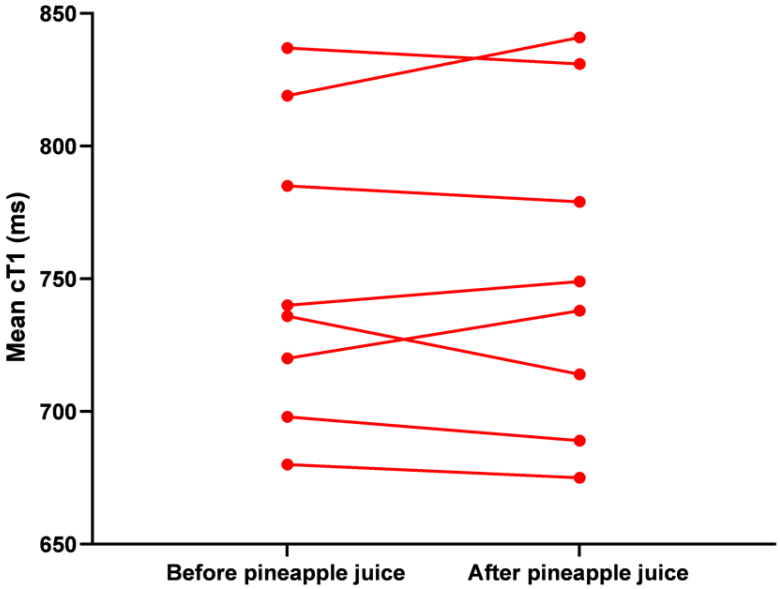


Figure 16. Mean liver cT1 measurements in healthy volunteers (n = 8) before and after the consumption of 200ml of pineapple juice.

3.7.4 Discussion

Although a small sample size, this result is in keeping with published literature on cT1 measurements. One study found no difference in the cT1 measurements of 22 healthy volunteers before and after eating, with a CoV of 1.4% between the two measurements[126]. Therefore, I concluded that consumption of pineapple juice before T1 mapping in the MoBILISE study will not interfere with cT1 measurements.

3.8 Variability between liver T1 and iron corrected T1

3.8.1 Background, aim and hypothesis

The liver plays an important role in iron metabolism. Chronic liver diseases affect the synthetic function of the liver, including the production of hepcidin, a key protein in maintaining iron homeostasis in the body. Low levels of hepcidin leads to iron accumulation in the liver due to excess iron[143]. Liver iron accumulation whilst is well-described in liver diseases such as NAFLD, ArLD, and viral hepatitis, but is typically not a feature in PSC. Iron correction does not alter measured T1 values of patients with normal

liver iron concentration[122]. One study, however, has demonstrated lower levels of hepcidin in patients with autoimmune hepatitis (AIH), primary biliary cholangitis (PBC) and PSC compared to other liver diseases, suggesting a similar liver iron overload phenomenon[144]. Nevertheless, this raises the question as to whether there is a need to perform iron-correction in T1 relaxation time measurement in a study involving patients with PSC and healthy volunteers. The hypothesis was that there will be no difference between T1 and cT1 measurements in patients with PSC and healthy volunteers.

3.8.2 Methods

Paired T1 and cT1 measurements of all participants in the MoBILISE study were compared using paired t-test for parametric distribution and Wilcoxon matched-pairs test for non-parametric distribution.

3.8.3 Results

Data from 80 patients with PSC and 20 healthy volunteers were available for analysis. There was a difference between the paired measurements in patients with PSC (mean T1: 769.1ms vs mean cT1: 773.7ms, $p < 0.0001$) but no difference observed in healthy volunteers (mean T1: 717.5ms vs mean cT1: 721.3ms, $p = 0.2293$; Figure 17). In 76 patients, there was clinically negligible difference between T1 and cT1 measurements at levels below 10ms. The remaining four patients with PSC were identified to have a lower T1 compared to cT1 measurements. Their clinical details are summarised in Table 14. These patients had liver iron levels above the upper limit of normal of 1.50mg/g which has confounded the T1 measurement. None were obese and all four had clinical cirrhosis with portal hypertension.

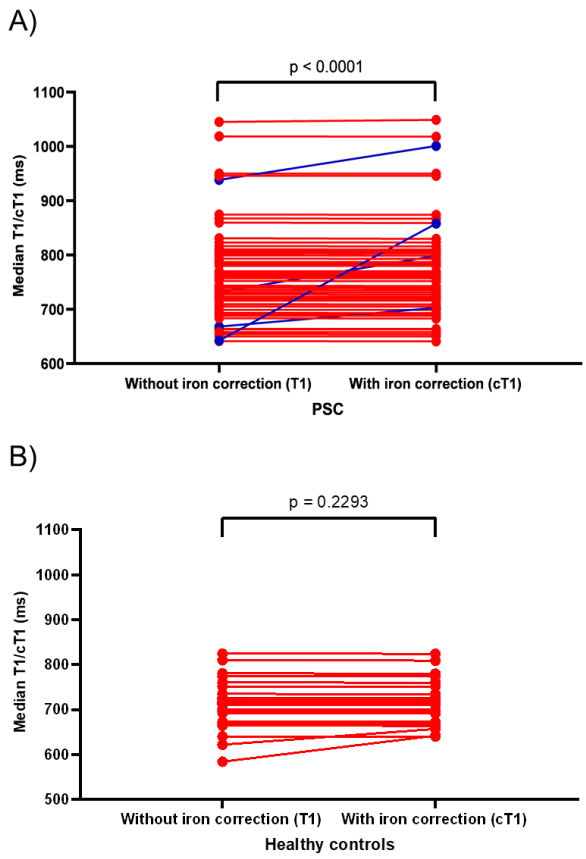


Figure 17. Comparisons between T1 measurements with and without iron-correction in (A) patients with PSC (n = 80) and (B) healthy controls (n = 20). The four patients with PSC that had a significant difference between the two measurements are denoted in blue.

Table 14. Clinical details of the four patients with PSC that demonstrated a difference in T1 and cT1 measurements.

Age	Sex	T1 (ms)	cT1 (ms)	Liver iron (mg/g)	Liver fat (%)	BMI (kg/m^2)	Presence of cirrhosis
32	Male	743.1	811.8	1.71	6.0	27.9	Yes
64	Male	696.3	732.3	1.54	2.3	20.6	Yes
32	Male	946.1	1009.4	1.59	3.0	27.5	Yes
75	Male	651.3	869.6	2.87	3.9	24.4	Yes

3.8.4 Discussion

The increased liver iron content in the participants could be due to possible undiagnosed fatty liver in one of the participants and advanced liver disease in all four. Therefore, the hypothesis was rejected, and it was decided that a standardised iron-correction method should be applied for all participants to allow a fairer comparison between participants and within participants over two time points in the present study. A standardised image acquisition and cT1 analysis approach could also allow comparisons to be made with cT1 obtained from other liver disease cohorts (as reported in chapter 6) and future studies using Liver*MultiScan*.

3.9 Variability between methods of data processing for liver cT1 measurements in PSC

3.9.1 Background, aim and hypothesis

Liver*MultiScan* analysis was originally developed and performed using the manual ROI-placement method. The mean cT1 using this method was shown to correlate well with histologically assessed liver fibrosis stages in predominantly NAFLD, ArLD, hepatitis C (HCV), and hepatitis (HBV). Since then, a semi-automated multi-slice liver segmentation method have been developed to enable assessment of larger areas of the liver. This method has been validated against the ROI method in healthy volunteers and other liver diseases in work carried out internally in OCMR and Perspectum, but not specifically in PSC. More recently, an automated liver segmentation method using a single slice has been developed using machine learning techniques to enable processing of large datasets. The association between the ROI and liver segmentation methods as well as between the two liver segmentations methods were compared in patients with PSC. The hypothesis was that there will be no difference in cT1 measurements.

3.9.2 Methods

Data from all patients with PSC in the MoBILISE study were included. Median and IQR cT1 were computed using Liver*MultiScan* using manual ROI-placement and the two liver segmentation methods (semi-automated multi-slice and automated single-slice). Linear regression and Pearson's correlation coefficient were used to assess the association between the two methods. The mean difference (bias), standard deviation, and 95% limits of agreement (LoA) from the Bland-Altman analysis were recorded.

3.9.3 Results

Manual ROI-placement vs semi-automated multi-slice segmentation

There was a strong association ($r = 0.93$; $R^2 = 0.87$; $p < 0.0001$) between median cT1 measurements obtained from both methods as depicted in Figure 18. The mean difference ± standard deviation between median cT1 measurements was -0.4 ± 28.3ms (95% LoA -55.9–55.1ms). The bias is not significantly different from zero, indicating good agreement between the two methods. The association between IQR cT1 measurements between the two methods were also good ($r = 0.91$, $R^2 = 0.82$; $p < 0.0001$). However, a mean difference ± standard deviation of 20.2 ± 14.0ms (95% LoA -7.2–47.6ms) was deemed to be clinically significant for measuring dispersion of cT1 data within the liver parenchyma.

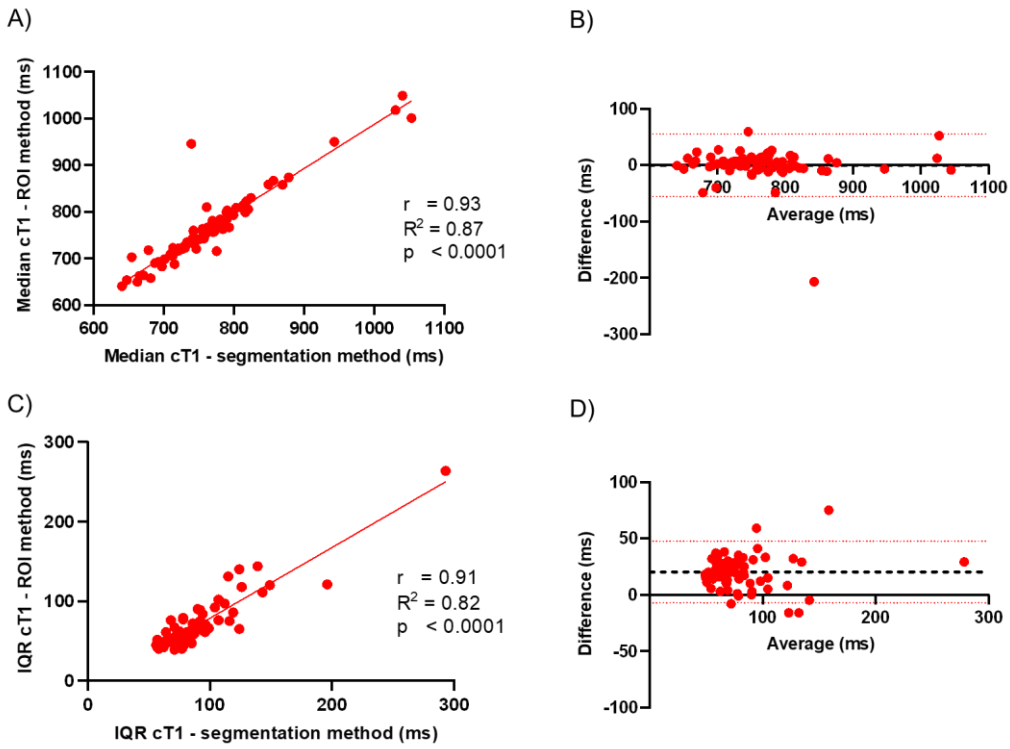


Figure 18. Variability in cT1 between manual ROI-placement and semi-automated multi-slice liver segmentation methods. (A) Correlation and (B) Bland-Altman plot for median cT1 measurements. (C) Correlation and (D) Bland-Altman plot for IQR cT1 measurements.

Semi-automated multi-slice segmentation vs automated single-slice segmentation

There was a strong association between median cT1 quantified using both liver segmentations methods ($r = 0.95$, $R^2 = 0.90$, $p < 0.0001$) as shown in Figure 19. The mean difference ± standard deviation between median cT1 measurements was -6.2 ± 26.2ms (95% LoA -57.5–45.1ms). The association between IQR cT1, however, was good but weaker than median cT1 ($r = 0.80$, $R^2 = 0.62$, $p < 0.0001$). The mean difference ± standard deviation between IQR cT1 measurements was -1.0 ± 20.4ms (95% LoA -41.0–38.9ms). The bias for median and IQR cT1 is not significantly different from zero, indicating good agreement between the two methods.

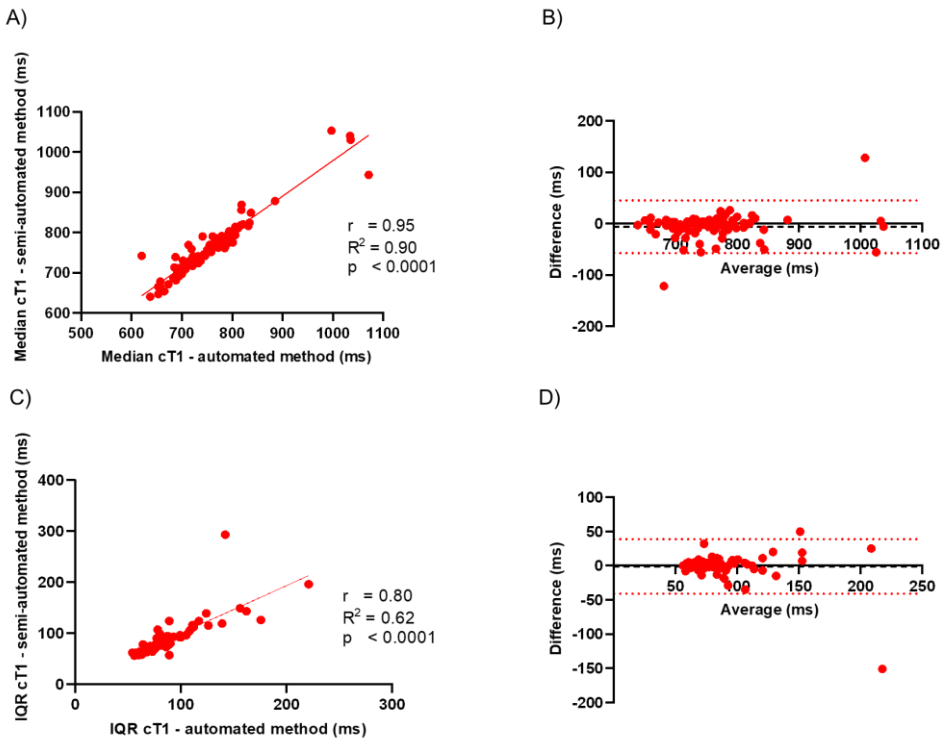


Figure 19. Variability in cT1 between multi-slice semi-automated and single-slice automated liver segmentation methods. (A) Correlation and (B) Bland-Altman plot for median cT1 measurements. (C) Correlation and (D) Bland-Altman plot for IQR cT1 measurements.

3.9.4 Discussion

Given that the liver parenchyma in patients with PSC is thought to be heterogeneous due to patchy distribution of inflammation and fibrosis [145], I decided that discrete circular ROIs of 20mm diameter may not accurately reflect these variability. Therefore, the main data analysis for the MoBILISE study used the semi-automatic multi-slice liver segmentation method, which allowed interrogation of a larger representative area of the liver. The good agreement for median and IQR cT1 between the semi-automated and automated liver segmentation methods, deemed it acceptable to use the automated method to analyse large datasets for median and IQR cT1 quantification in chapter 5.

3.10 Conclusions

In summary, work presented in this chapter have highlighted a few key points for data analysis in my thesis. These include:

- MRCP acquisition protocols and difference in resolution may have an impact on computation of *MRCP+* metrics, particularly when involving smaller calibre bile ducts
- There is significant variability in biliary tree length modelled between acquisitions
- Biliary visualisation is better with *MRCP+* than MRCP when MRI signal is present
- *MRCP+* metrics have excellent intra-observer and inter-observer agreement in healthy volunteers and patients with PSC
- Liver cT1 is not affected by pineapple juice consumption prior to scan
- Liver T1 and cT1 can be different in patients with PSC due to iron content, therefore cT1 is preferred for standardisation
- Liver segmentation method is preferred to ROI-placement method in PSC due to variability in IQR cT1 quantification between the two methods
- Automated Liver*MultiScan* analysis is a feasible method to analyse large dataset despite only analysing a single representative slice of the liver

In the next chapter, I will be discussing the assessment of biliary stricture and dilatations seen in PSC using *MRCP+*.

4 Quantitative assessment of cholangiopathy in PSC

4.1 Abstract

Background: MRI-MRCP in PSC is currently based on qualitative assessment and has high inter-observer variability. This study investigated the risk-stratification performance of *MRCP+* metrics against established prognostic models and non-invasive markers to propose candidate imaging biomarkers for assessment in future longitudinal studies.

Methods: MRCP data was analysed using *MRCP+* and were also reviewed by two radiologists for calculation of qualitative scores (Anali score and modified Amsterdam intrahepatic stricture severity score (ISSS)). Inter-reader agreement was assessed using intraclass correlation coefficient (ICC) for *MRCP+* and Cohen's kappa (κ) for MRCP. Patients were classified into high-risk or low-risk group to develop clinical endpoints using two prognostic models as the main reference standard, Mayo Risk Score (MRS: > 0 or ≤ 0) and Amsterdam-Oxford Model score (AOM: > 2 or ≤ 2). Published thresholds of severe disease for alkaline phosphatase, transient elastography liver stiffness (LS), and enhanced liver fibrosis score (ELF) were also examined. Diagnostic performance was assessed using receiver operating characteristic curve (ROC) analyses and binomial logistic regression models. Correlations of *MRCP+* metrics with quality of life assessed by SF-36 and PSC-PRO were performed.

Results: Data from 76 patients and 20 healthy controls were included. Inter-observer agreements for *MRCP+* dilatation metrics in PSC were better (ICC = 0.90–0.96) than ISSS (κ = 0.74) and Anali score (κ = 0.38). The sum of the relative severity of biliary dilatations compared to its neighbouring ducts (SumRelSevDilat) was higher in the high-risk compared to the low-risk group ($8.2m^{-1}$ vs $5.6m^{-1}$, $p < 0.0001$), and discriminated MRS > 0 vs ≤ 0 (AUC = 0.77, 95% CI 0.66–0.88) numerically better than other non-invasive markers. However, it was numerically lower to LS to differentiate AOM > 2 vs AOM ≤ 2 (SumRelSevDilat: AUC = 0.72, 95% CI 0.60–0.85; LS: AUC = 0.85, 95% CI

0.74–0.95). *MRCP+* intrahepatic dilatation severity was an independent predictor of MRS > 0 (OR: 29.2, p = 0.038). Dilatation metrics showed significant correlations with several domains of the patient-reported outcomes. SF-36 bodily pain (r_s=-0.26, p=0.025) and PSC-PRO activities of daily living (r_s=0.29, p=0.014) showed the strongest correlations with intrahepatic SumRelSevDilat.

Conclusion: Dilatation metrics derived from *MRCP+* are elevated in patients with high-risk PSC and shows promising risk stratification ability to identify patients with more severe disease.

4.2 Introduction

MRI-MRCP is the imaging modality of choice in establishing the diagnosis, staging cholangiopathy, and surveillance for biliary complications in PSC[29]. Multifocal biliary strictures and consequent dilatations underpin the pathological process in PSC. Current definition of strictures and dilatations rely on recognition of morphological abnormalities in a duct compared to adjacent ducts on MRI-MRCP and a judgement call of the relative severity of changes in diameter. Reliance on qualitative descriptors leads to high inter-observer variability in interpretation even among experts in PSC[120], and makes developing MRI-based biomarkers using current standards particularly challenging. Standardised interpretation models have been developed to interpret bile duct abnormalities. The Amsterdam cholangiographic classification was originally designed for ERCP interpretation and predicted transplant-free survival in two cohorts[46, 86]. A modified version has been adapted to MRCP interpretation[89]. In a landmark retrospective study, a standardised interpretation model specific to MRI-MRCP data was proposed, where the presence of marked intrahepatic bile duct dilatation, hepatic dysmorphy, and features of portal hypertension on unenhanced scans, predicted radiological progression from baseline with good accuracy (AUC = 0.80) during a mean follow-up of 4 years[91]. Using a combination of these three features, the “Anali score” was developed and its prognostic value was validated in an international retrospective multicentre cohort study[134]. In this score, the severity of intrahepatic bile duct dilatation is assessed using a categorical grading scale based on the 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 and could introduce observer dependent bias.

Quantitative MRCP techniques and automated derivation of bile duct diameter profiles potentially provide more objective and nuanced assessment of the biliary tree measurements. Such a quantitative approach could be useful in the search for non-invasive surrogate endpoints for PSC. *MRCP+* generates a multiplicity of measurements

of the biliary tree derived from MRCP images and its utility in assessing the biliary tree abnormalities in PSC remains to be tested. *MRCP+* automatically detects regions with variation in bile duct diameters. The threshold for classification of a candidate stricture and dilatation was pre-defined in the *MRCP+* algorithm to meet both an absolute change of 1mm and a minimum 30% relative change in bile duct diameter criteria. These thresholds were decided based on their internal performance testing against synthetic digital phantom instances and 3D-printed phantom[146]. Across all models of both digital and 3D-printed phantoms, the sensitivity was 76.6% for stricture detection and 85.9% for dilatation detection[146]. Importantly, there were no false positive detections. This chapter will specifically focus on how the candidate stricture and dilatation metrics were derived from *MRCP+* and their application in the assessment of patients with PSC by comparing against published thresholds of existing non-invasive markers.

4.3 Hypotheses

Stricture metrics are likely to be more susceptible to technical limitations of MRCP and *MRCP+* as discussed in chapter 3. This coupled with the fact that the sensitivity for the detection of dilatations is better than strictures in phantom studies, it could be expected that candidate dilatation metrics will perform better than stricture metrics in PSC. My hypotheses for this chapter were that dilatation metrics will be:

- higher in PSC than normal biliary tree
- correlated with surrogate markers of disease severity in PSC
- higher in more severe PSC compared to mild disease
- associated with worse QoL

4.4 Aims

This chapter aims to:

- explore the use of *MRCP+* stricture, dilatation, and volume metrics in disease assessment
- investigate their performance to discriminate patients with PSC from healthy volunteers
- assess their relationship with non-invasive markers of PSC disease severity
- assess their relationship with morphological features on MRI
- compare their performance against qualitative MRI-MRCP scores to classify patients into high-risk and low-risk groups based on surrogate risk-stratification tools
- assess whether they are independent predictors of high-risk PSC using prognostic risk models as the reference standard
- examine the relationship between *MRCP+* metrics and QoL responses

4.5 Participants and methods

4.5.1 Participants
Patients with large-duct PSC and healthy controls were recruited as reported in section 2.2 chapter 2.

4.5.2 Clinical and laboratory data
Clinical and biochemical data was obtained according to methods reported in section 2.6 chapter 2.

4.5.3 Transient elastography
Liver stiffness was measured using FibroScan as reported in section 2.7 chapter 2.

4.5.4 Risk stratification
Patients with PSC were classified into high-risk and low-risk groups according to published thresholds of risk scores reported earlier in this thesis in Table 6 section 2.9 chapter 2. To recap, these are MRS > 0 or ≤ 0, AOM > 2 or ≤ 2, ALP > 2.2x ULN or ≤ 2.2x ULN, ALP > 1.5x ULN or ≤ 1.5x ULN, LS > 9.6kPa or ≤ 9.6kPa, ELF > 9.8 or ≤ 9.8, ISSS > 2 or ≤ 2, Anali score > 2 or ≤ 2, presence of EHD, and presence of DS.

4.5.5 MRI acquisition
MRI-MRCP acquisition followed the methods described in section 2.3 chapter 2.

4.5.6 Qualitative image analysis
MRI-MRCP scans of patients with PSC were interpreted independently by two specialist radiologists using a standardised interpretation model to compute the modified Amsterdam stricture severity and Anali scores as described in section 2.5 chapter 2.

4.5.7 Quantitative image analysis
Image analysis was performed following the steps described in section 2.4 chapter 2. Metrics relating to the relative severity and length of each candidate stricture and dilatation were computed. A composite dilatation score was derived as the product of the relative severity and length of each candidate stricture or dilatation. Cumulative tree metrics for all candidate strictures and dilatations within the intrahepatic and whole biliary tree were calculated.

4.5.8 Normalisation to the length of the biliary tree

MRCP images are often noisy, with bile ducts ranging from several millimetres to below the image resolution of the MR scanner. The limited imaging resolution plus relatively low signal-to-noise ratio in the original MRCP data mean that only a proportion of the biliary tree can be detected and reconstructed by *MRCP+*, as demonstrated in early technical development data discussed in chapter 3. In general, it is difficult to determine how much of the actual biliary tree is reconstructed and represented by *MRCP+* as there is no gold standard anatomical comparator (i.e., in one case 60% of the total biliary tree may be reconstructed, but in another case maybe only 40% of the biliary tree is reconstructed). To account for this, all variables relating to the whole biliary tree (e.g., total number of dilatations) were normalised to their corresponding biliary tree length and reported as per metre of biliary tree, to allow comparisons. These normalisations were not carried out to variables relating to defined anatomical structures (e.g., gallbladder volume, number of dilatations in the common bile duct). The *MRCP+* metrics assessed in this chapter and their abbreviated nomenclatures were:

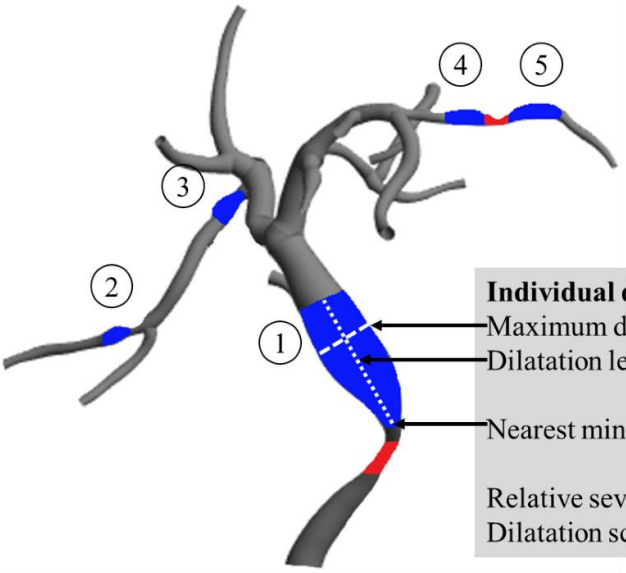
- Biliary tree length (TreeLength; m)
- Number of candidate dilatations (DilatNum; m^{-1})
- Number of candidate strictures (StricNum; m^{-1})
- Sum of relative severity of candidate dilatations (SumRelSevDilat; m^{-1})
- Sum of relative severity of candidate strictures (SumRelSevStric; m^{-1})
- Proportion of tree dilated (DilatTreeProp; %)
- Proportion of tree strictured (StricTreeProp; %)
- Tree dilatation score (TreeDilatScore; no unit)
- Tree stricture score (TreeStricScore; no unit)
- Biliary tree volume (BTvol; mlm^{-1})
- Gallbladder volume (GBvol; ml)

The definitions of individual and whole tree candidate stricture and dilatation metrics are summarised in Table 15. A worked example of how the candidate dilatation metrics were computed is shown in Figure 20.

Table 15. Quantitative metrics derived from *MRCP*+ analysis and their respective definitions in this study.

MRCP+ metrics	Definition
Individual dilatation metrics	
Candidate dilatation	A bile duct dilatation with at least 1mm absolute increase and at least 30% relative increase in diameter, compared to its closest (in diameter) neighbouring local minimum
Relative severity of candidate dilatation	$\frac{\text{Maximum dilatation diameter - Nearest local minimum diameter}}{\text{Nearest local minimum diameter}}$
Length of candidate dilatation (mm)	Full width at half maximum of candidate dilatation diameter profile
Dilatation score (mm)	Length of candidate dilatation x Relative severity of dilatation
Whole tree dilatation metrics†	
Number of dilatations, DilatNum (m^{-1})	$\frac{\text{Total number of candidate dilatations}}{\text{Total length of measured biliary tree in metres}}$
Sum of relative severity of dilatations, SumRelSevDilat (m^{-1})	$\frac{\text{Sum of relative severity of all candidate dilatations}}{\text{Total length of measured biliary tree in metres}}$
Proportion of tree dilated, DilatTreeProp (%)	$\frac{\text{Sum of lengths of candidate dilatations in metres}}{\text{Total length of measured biliary tree in metres}} \times 100$
Tree dilatation score, TreeDilatScore	$\frac{\text{Sum of candidate dilatation scores in metres}}{\text{Total length of measured biliary tree in metres}}$
Individual stricture metrics	
Candidate stricture	A bile duct narrowing with at least 1mm absolute decrease and at least 30% relative decrease in diameter, compared to its closest (in diameter) neighbouring local maximum
Relative severity of candidate stricture	$\frac{\text{Nearest local maximum diameter - Minimum stricture diameter}}{\text{Nearest local maximum diameter}}$
Length of candidate stricture (mm)	Half the distance between the minimum stricture diameter and nearest local maximum
Stricture score (mm)	Length of candidate stricture x Relative severity of stricture
Whole tree stricture metrics†	
Number of candidate strictures, StricNum (m^{-1})	$\frac{\text{Total number of all candidate strictures}}{\text{Total length of measured biliary tree in metres}}$
Sum of relative severity of candidate strictures, SumRelSevStric (m^{-1})	$\frac{\text{Sum of the relative severity of all candidate strictures}}{\text{Total length of measured biliary tree}}$
Proportion of tree strictured, StricTreeProp (%)	$\frac{\text{Sum of candidate stricture lengths}}{\text{Total length of measured biliary tree}}$
Tree stricture score, TreeStricScore	$\frac{\text{Sum of candidate stricture scores}}{\text{Total length of measured biliary tree}}$

†These metrics were normalised to the total length of the measured biliary tree to account for variable lengths of the biliary tree modelled on *MRCP*+ between participants.



Candidate strictures = 2

Candidate dilatations = 5

Length of biliary tree = 0.654 metre

Individual dilatation ① metrics

Maximum diameter = 14mm

Dilatation length = 30mm

Nearest minimum diameter = 4mm

Relative severity dilatation = [(14-4mm)/4mm] = 2.5

Dilatation score = Length x Relative severity = 75mm

Tree dilatation metrics

Sum of relative severity of dilatations (SumRelSevDilat, m^{-1}) = $\frac{\text{Relative severity dilatations } ① + ② + ③ + ④ + ⑤}{0.654\text{m}}$

Proportion of tree dilated (DilatTreeProp, %) = $\frac{\text{Dilatation lengths } ① + ② + ③ + ④ + ⑤ \text{mm}}{0.654\text{m}} \times 100$

Tree dilatation score (TreeDilatScore) = $\frac{\text{Dilatation scores } ① + ② + ③ + ④ + ⑤ \text{mm}}{0.654\text{m}}$

Figure 20. Biliary tree model demonstrating candidate strictures (red) and dilatations (blue) 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 normalised to the total length of the biliary tree in metres.

4.5.9 Statistical analysis

Inter-observer and intra-observer variability of *MRCP+* metrics in PSC were assessed visually using Bland-Altman plots and by the intraclass correlation coefficient (ICC) using a two-way mixed-effects model based on absolute agreement and average measurements. ICC values were classified as follows[138]: >0.90 excellent, 0.75-0.90 good, 0.50-0.75 moderate, and <0.50 poor. Inter-observer variability in qualitative MRI-MRCP interpretation was evaluated using Cohen's kappa (κ) coefficient for qualitative data and ICC as above for quantitative data. κ values were classified as follows[147]: >0.80 excellent, 0.79-0.60 substantial, 0.59-0.40 moderate, 0.39-0.20 fair, and <0.2 poor.

Student's *t*-tests were used to compare differences in *MRCP+* metrics between groups. ROC analyses were performed to assess the performance of *MRCP+* metrics, modified Amsterdam cholangiographic scores and Anali score, respectively, to discriminate high- and low-risk PSC according to published thresholds of non-invasive risk stratification tools. AUC, sensitivity, and specificity were calculated. For *MRCP+* analysis, whole biliary tree metrics and metrics relating to the intrahepatic ducts only were examined as intrahepatic duct dilatation had prognostic significance in the recently validated Anali score.[134] For the qualitative MRI-MRCP scores, results from both readers are presented and where appropriate, results from reader 1 (A.B.S), selected based on their prior experience, was used for the main analysis. Consensus between both readers were not sought in cases of disagreement.

Multivariate logistic regression analyses were performed to identify the best predictive variables of high-risk PSC as stratified by MRS and AOM, as these scores have reasonable discriminatory performance and predictive accuracy estimates of survival[52, 70]. Covariates with $p < 0.1$ in the univariate analysis were entered into the multivariate analysis. Variables that contributed to MRS and AOM scores were excluded. *MRCP+* metrics with multicollinearity were excluded if Spearman's rank correlation coefficient > 0.80 and variance inflation factor < 1 or > 10. Odd ratios (OR) for covariates that

remained significant in the multivariate model were reported. Spearman's correlation coefficient was used to examine the strength of association between QoL data and study variables.

4.6 Results

4.6.1 Cohort characteristics

Recruitment to the study is summarised in Figure 21. Eighty-one patients with large-duct PSC consented to the study. One patient was excluded due to claustrophobia and four patients were excluded due to poor quality MRCP acquisition. Seventy-six patients were included in the final analysis. The median age of patients with PSC was 44 years (IQR: 31–56) and 67% were male. The median disease duration was 8 years (IQR: 4–12), 55% were on UDCA 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 summarised in Table 16. Although *MRCP+* analysis was possible in the four excluded patients, their *MRCP+* model was judged to be of poor quality as they did not meet the required quality criteria, therefore were excluded. Characteristics of the four patients excluded from the study is summarised in Table 17. There were no technical failures for transient elastography.

Normal reference values were established from 20 age- and sex-matched healthy volunteers after excluding one poor quality MRCP acquisition due to motion artifact and contamination from gastrointestinal tract. Transient elastography was successful in all volunteers. The median age was 35 years (IQR: 31–39; p = 0.07 compared to patients) and 65% were male (p = 0.80 compared to patients).

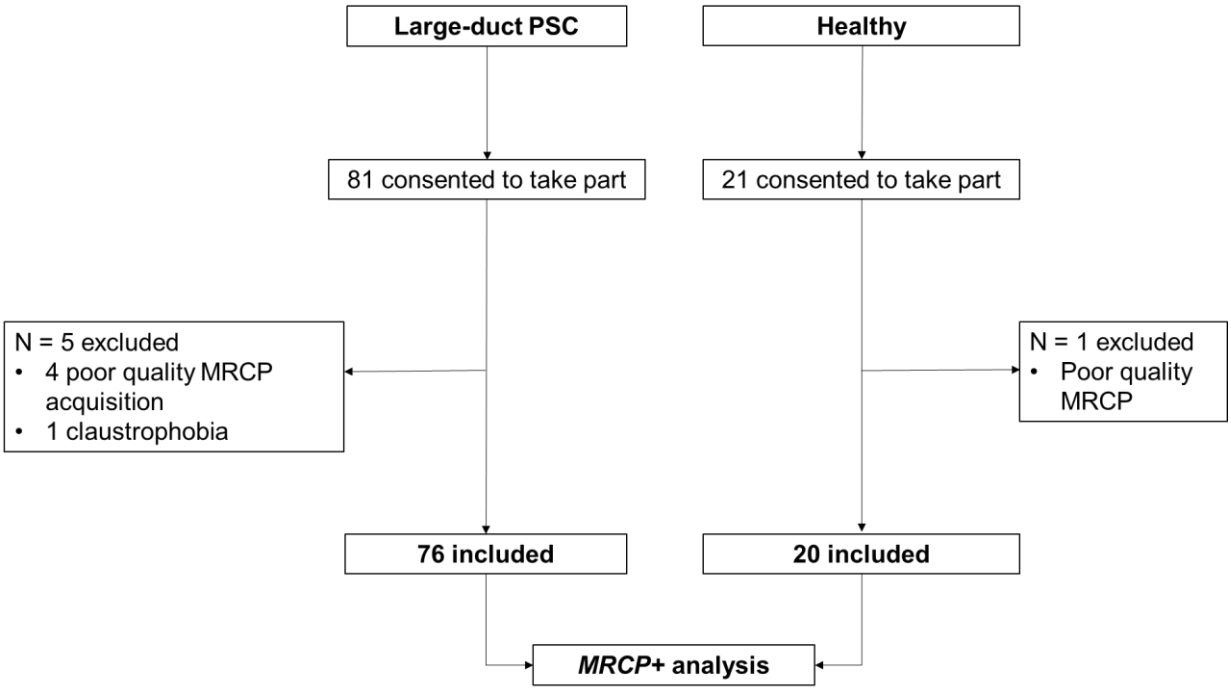


Figure 21. Study flow chart for *MRCP+* analysis.

Table 16. Baseline characteristics of participants included for *MRCP+* analysis.

	PSC (n = 76)	Healthy (n = 20)
Demographics		
Male	51 (67)	13 (65)
Age (years)	44 (31–56)	35 (31–39)
BMI (kg/m^2)	24.5 (22.2-27.6)	22.7 (21.2–24.8)
Laboratory parameters		
Total bilirubin, µmol/L	13 (10–20)	10 (8–13)
ALT (IU/l)	42 (27–84)	20 (13–26)
AST (IU/l)	38 (27–62)	22 (19–24)
ALP (IU/l)	150 (101–234)	58 (51–70)
xULN ALP†	1.2 (0.8–1.8)	0.4 (0.4–0.5)
GGT (IU/l)	123 (49–306)	17 (14–20)
Albumin (g/l)	40 (38–42)	42 (41–43)
Prothrombin time (s)	10.6 (10.3–11.0)	10.5 (10.5-11.0)
Platelet count (x 10^9/l)	254 (193–302)	256 (234–270)
Clinical data		
PSC duration (years)	8 (4–12)	-
PSC localisation at recruitment		
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)	-
Liver fibrosis markers		
LS (kPa)	6.8 (5.3–10.0)	4.7 (4.1–5.3)
ELF	9.4 (8.7–9.9)	8.6 (8.0–8.9)
Prognostic risk models		
Mayo Risk Score	0.1 (-0.2–0.5)	-
Amsterdam-Oxford Model score	1.6 (1.3–2.1)	-

Continuous variables are expressed as median (interquartile range) and nominal variables as absolute number (percentage). †ULN of ALP was 130 IU/L.
Abbreviations: ALP; alkaline phosphatase, ALT; alanine aminotransferase, AST; aspartate aminotransferase, BMI; body mass index, ELF; enhanced liver fibrosis score, GGT; gamma-glutamyltransferase, IBD; inflammatory bowel disease, LS; transient elastography liver stiffness, UDCA; ursodeoxycholic acid, xULN; times upper limit of normal.

Table 17. Characteristics of the four patients excluded from the study.

Age (years)	Sex	PSC duration (years)	*MRCP*+ analysis outcome		Risk stratification markers				
			Acquisition quality	Reason for rejection	xULN ALP	LS	ELF	MRS	AOM	Qualitative MRCP
32	Male	15	Original image with gastrointestinal contamination and low signal-to-noise ratio with poor biliary contrast	Unreliable trajectory and duct widths	2.6	14.7	9.7	0.44	1.8	ISSS = 4, ESSS = 3, TSSS = 7 Anali = 5 Cirrhosis = Yes DS = No
19	Male	7	Original image with low fat suppression and suboptimal field of view positioning	Low data quality	1.6	45.5	12.6	1.29	4.3	ISSS = 4, ESSS = 3, TSSS = 7 Anali = 3 (reader 1) or 4 (reader 2) Cirrhosis = Yes DS = No
32	Male	2	Original image with suboptimal biliary contrast and field of view positioning	Low data quality	0.9	33.2	11.7	3.02	3.2	ISSS = 4, ESSS = 1, TSSS = 5 Anali = 3 Cirrhosis = Yes DS = No
20	Female	7	Original image with suboptimal field of view positioning	Low data quality	0.8	5.4	8.1	1.13	0.2	ISSS = 1, ESSS = 1, TSSS = 2 Anali = 0 Cirrhosis = No DS = No

ALP; alkaline phosphatase, AOM; Amsterdam-Oxford Model score, DS; dominant stricture, ELF; enhanced liver fibrosis score, ESSS; extrahepatic stricture severity score, ISSS; intrahepatic stricture severity score, LS; transient elastography liver stiffness, MRS; Mayo Risk Score, TSSS; total stricture severity score

4.6.2 Comparison of *MRCP+* metrics between PSC and healthy

Biliary tree models representative of both cohorts are shown in Figure 22. All *MRCP+* tree metrics were compared between PSC and healthy volunteers as summarised in Table 18. Gallbladder volume ($p < 0.0001$) and total biliary tree volume ($p = 0.0020$) were higher in PSC. Dilatation metrics and gallbladder volumes had AUC between 0.83–0.87 to differentiate PSC from healthy biliary tree. The sum of relative severity of candidate dilatations within the biliary tree (SumRelSevDilat), had numerically the highest performance to differentiate PSC from healthy (AUC = 0.87, 95% CI 0.77–0.96, $p < 0.0001$, sensitivity 86%, specificity 85%). Amongst the non-invasive markers, serum ALP (AUC = 0.93, 95% CI 0.88–0.98, $p < 0.0001$, sensitivity 90%, specificity 88%) had numerically the highest discriminatory performance followed by LS (AUC = 0.84, 95% 0.76–0.92, $p < 0.0001$, sensitivity 100%, specificity 61%), and ELF (AUC = 0.79, 95% CI 0.70–0.89, $p < 0.0001$, sensitivity 81%, specificity 66%).

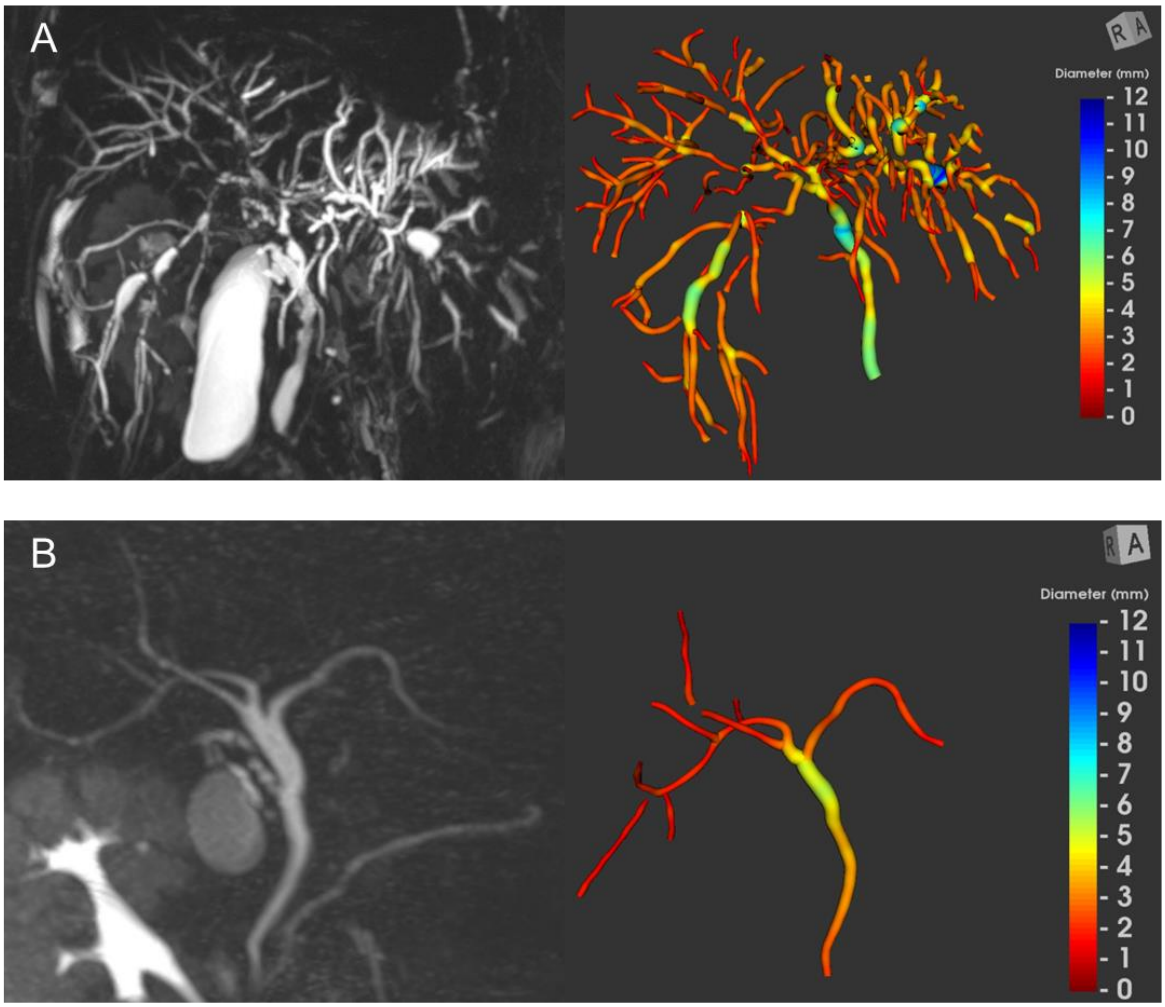


Figure 22. Maximum intensity projection (MIP) image (left) and the corresponding 3D parametric biliary tree model (right). (A) 33-year-old patient with extrahepatic and intrahepatic PSC (strictures = 20, dilatations = 42), and (B) 40-year-old healthy volunteer (strictures = 0, dilatations = 0). The gallbladders have been segmented out in both cases.

Table 18. Diagnostic performance of *MRCP+* whole tree metrics to distinguish between patients with PSC and healthy controls.

	PSC (n=76)		Healthy (n=20)		Cut-off	AUC (95% CI)	Sens (%)	Spec (%)	p-value
Tree metrics									
TreeLength (m)	1.2	(0.6–2.0)	0.4	(0.4–0.7)	1.1	0.78 (0.69–0.87)	100	53	<0.0001
GallbladderVol (ml)	39.3‡	(21.5–62.8)	16.0	(10.4–18.9)	27.1	0.85 (0.77–0.94)	83	70	<0.0001
TreeVol† (ml^{-1})	8.7	(7.5–10.1)	7.2	(6.6–8.1)	7.3	0.72 (0.59–0.85)	60	81	0.0020
Tree dilatation metrics†									
DilatNum (m^{-1})	10.5	(7.3–14.0)	2.4	(0.0–6.6)	3.8	0.86 (0.78–0.96)	71	92	<0.0001
SumRelSevDilat (m^{-1})	6.7	(4.7–8.8)	0.2	(0.0–0.4)	3.8	0.87 (0.77–0.96)	86	85	<0.0001
DilatTreeProp (%)	6.9	(4.4–8.3)	1.0	(0.0–4.1)	1.6	0.85 (0.73–0.95)	67	93	<0.0001
TreeDilatScore	42.1	(29.1–56.1)	6.9	(0.0–26.5)	19.1	0.83 (0.71–0.94)	76	90	<0.0001
Tree stricture metrics†									
StricNum (m^{-1})	5.8	(3.5–8.1)	2.4	(0.0–4.5)	4.7	0.76 (0.65–0.88)	81	64	0.0002
SumRelSevStric (m^{-1})	2.4	(1.5–3.4)	0.9	(0.0–2.0)	1.1	0.79 (0.68–0.79)	67	83	0.0037
StricTreeProp (%)	4.8	(2.6–6.1)	2.0	(0.0–5.0)	2.2	0.70 (0.55–0.84)	57	83	0.0052
TreeStricScore (%)	19.6	(10.9–26.6)	8.7	(0.0–18.9)	4.0	0.71 (0.58–0.85)	48	92	0.0029
Other non-invasive markers									
ALP (IU/l)	150.0	(101.3–234.8)	58.0	(50.5–70.0)	80.5	0.93 (0.88–0.98)	90	88	<0.0001
LS (kPa)	6.8	(5.3–10.0)	4.7	(4.1–5.3)	6.1	0.84 (0.76–0.92)	100	61	<0.0001
ELF	9.4	(8.7–9.9)	8.6	(8.0–8.9)	9.1	0.79 (0.70–0.89)	81	66	<0.0001

†These parameters have been normalised to the corresponding total length of the biliary tree. All values are presented as median (interquartile range).
‡n=57. ALP, alkaline phosphatase; LS, transient elastography liver stiffness; ELF, enhanced liver fibrosis score; Sens, sensitivity; Spec; specificity, BiliaryTreeLength; total length of the biliary tree, GallbladderVol; gallbladder volume, TreeVol; biliary tree volume, DilatNum; number of candidate dilatations per metre, SumRelSevDilat; sum of relative severity of candidate dilatations per metre, DilatTreeProp; proportion of tree dilated, TreeDilatScore; sum of tree dilatation scores, StricNum; number of candidate strictures per metre, SumRelSevStric; sum of relative severity of candidate strictures per metre, StricTreeProp; proportion of tree strictured, TreeStricScore; sum of tree stricture scores.

4.6.3 Inter-observer agreement of MRI-MRCP scores and features

Radiologist reads of MRI-MRCP were possible in all 80 patients but only the 76 patients with corresponding *MRCP+* data were included in the analysis. The outcomes from both readers are summarised in Table 19. For qualitative MRI-MRCP interpretation, the inter-observer agreement for TSSS was substantial (κ = 0.66), with agreement for ESSS (κ = 0.82) better than ISSS (κ = 0.74). The inter-observer agreement for the Anali score was fair (κ = 0.38) and for the classification of maximum intrahepatic bile duct diameter was moderate (κ = 0.50). There was excellent agreement for the presence of a dominant stricture (κ = 0.81) and gallbladder enlargement (κ = 0.90). The inter-observer agreement for CBD and CHD diameter measurements were better than RHD and LHD (Table 20).

4.6.4 Inter-observer and intra-observer agreement of *MRCP+* candidate stricture and dilatation metrics

Dilatation metrics had good to excellent inter-observer agreements (ICC = 0.90–0.96) whilst stricture metrics had poor to good inter-observer agreements (ICC = 0.64–0.84; Table 21). Intra-observer agreements were excellent for dilatation metrics (ICC = 0.93–0.97) and good for stricture metrics (ICC = 0.81–0.88; Table 21). The ICC for the measurement of maximum duct diameter was significantly better with *MRCP+* than manual measurements on MRI-MRCP for all biliary segments (Table 22).

Table 19. Qualitative interpretation of MRCP for morphological features and their corresponding inter-observer agreements.

MRI-MRCP features	Number of cases		Cases in agreement	Cohen's kappa, κ
	Reader 1	Reader 2		
Intrahepatic stricture severity score (ISSS)			61/76 (80%)	0.74
0	1	4	1	
1	12	17	8	
2	26	16	16	
3	32	35	32	
4	5	4	4	
Extrahepatic stricture severity score (ESSS)			67/76 (88%)	0.82
0	16	12	11	
1	31	34	29	
2	25	27	24	
3	3	2	2	
4	1	1	1	
Total stricture severity score (TSSS)			54/76 (71%)	0.66
0	1	4	1	
1	8	6	3	
2	11	11	5	
3	15	12	8	
4	12	15	11	
5	24	24	22	
6	4	3	3	
7	1	1	1	
8	0	0	0	
Anali score			37/76 (49%)	0.38
0	23	27	17	
1	11	11	5	
2	15	12	4	
3	7	10	3	
4	10	12	5	
5	10	4	3	
Intrahepatic bile duct dilatation classification			51/76 (67%)	0.50
≤3mm	32	37	27	
4mm	19	20	10	
≥5mm	25	19	14	
Dominant stricture			69/76 (91%)	0.81
Absent	51	46	45	
Present	25	30	24	
Gallbladder enlargement			73/76 (96%)	0.90
Absent	37	36	35	
Present	39	40	38	

Table 20. Manual measurements of bile ducts by MRI-MRCP readers and their corresponding inter-observer agreements.

	MRI-MRCP maximum diameter measurement† (mm)		
Individual duct	Reader 1	Reader 2	ICC
CBD	6.0 (4.0–6.0)	4.5 (3.0–6.0)	0.78
CHD	4.0 (3.0–6.0)	5.0 (3.0–7.0)	0.83
Main RHD	3.0 (2.0–3.0)	2.0 (1.5–4.0)	0.69
Main LHD	2.0 (2.0–3.0)	2.5 (1.5–3.5)	0.40

†Measurements presented as median (interquartile range). ICC, intraclass correlation coefficient; CBD, common bile duct; CHD, common hepatic duct; RHD, right hepatic duct; LHD, left hepatic duct.

Table 21. Inter- and intra-observer agreement for *MRCP+* dilatation and stricture metrics.

	Intraclass correlation coefficient (ICC)	
Whole tree metrics	Intra-observer agreement (analyst 1)	Inter-observer agreement (analyst 1 vs 2)
DilatNum (m^{-1})	0.95	0.93
SumRelSevDilat (m^{-1})	0.95	0.90
DilatTreeProp (%)	0.97	0.96
TreeDilatScore	0.93	0.92
StricNum (m^{-1})	0.85	0.84
SumRelSevStric (m^{-1})	0.86	0.85
StricTreeProp (%)	0.81	0.64
TreeStricScore	0.83	0.68

DilatNum, number of candidate dilatations per metre; SumRelSevDilat, sum of relative severity of candidate dilatations per metre; DilatTreeProp. proportion of tree dilated; TreeDilatScore, sum of tree dilatation scores; StricNum, number of candidate strictures per metre; SumRelSevStric, sum of relative severity of candidate strictures per metre; StricTreeProp, proportion of tree strictured; TreeStricScore, sum of tree stricture scores.

Table 22. Automatic measurements of bile duct diameter by *MRCP+* and their corresponding inter-observer agreements.

	MRCP+ maximum diameter measurement† (mm)		
Individual duct	Reader 1	Reader 2	ICC
CBD	7.0 (5.8–8.7)	7.0 (5.5–8.8)	0.98
CHD	7.5 (5.7–8.9)	6.4 (5.3–8.8)	0.98
Main RHD	5.4 (4.2–6.0)	4.8 (4.2–6.2)	0.81
Main LHD	5.3 (4.0–7.0)	5.5 (4.6–7.0)	0.82

†Measurements presented as median (interquartile range). ICC, intraclass correlation coefficient; CBD, common bile duct; CHD, common hepatic duct; RHD, right hepatic duct; LHD, left hepatic duct.

4.6.5 Association of *MRCP+* metrics with study variables

The correlations between study variables and all *MRCP+* metrics for the whole biliary tree (Tables Table *23*Table 24Table 25) and specifically for the intrahepatic tree (Tables Table *26*, Table *27*, Table *28*) are presented. In summary, dilatation metrics had weak to moderate strength but statistically significant correlations with non-invasive markers. Correlations for stricture metrics were poor and not statistically significant. Correlations were better for the intrahepatic tree metrics than whole tree metrics.

Table 23. Whole tree *MRCP+* metrics correlation with blood tests.

Variable	Correlation	TreeVol (mlm^{-1})	DilatNum (m^{-1})	SumRelSevDilat (m^{-1})	DilatTreeProp (%)	TreeDilatScore	StricNum (m^{-1})	SumRelSevStric (m^{-1})	StricTreeProp (%)	TreeStricScore
Total bilirubin (µmol/l)	coefficient	-0.050	0.185	0.169	0.144	0.103	0.023	0.031	-0.007	-0.022
	p-value	0.670	0.110	0.145	0.214	0.374	0.843	0.789	0.950	0.851
ALT (IU/l)	coefficient	-0.040	0.132	0.136	0.027	0.005	0.179	0.182	0.178	0.177
	p-value	0.729	0.257	0.241	0.816	0.963	0.121	0.116	0.123	0.126
AST (IU/l)	coefficient	-0.046	0.229	0.243	0.079	0.071	0.231	0.230	0.223	0.225
	p-value	0.695	0.046	**0.034**	0.495	0.543	**0.044**	0.046	0.053	0.050
ALP (IU/l)	coefficient	0.022	0.371	0.380	0.263	0.265	0.250	0.256	0.104	0.123
	p-value	0.852	**0.001**	**0.001**	**0.022**	**0.021**	**0.029**	**0.026**	0.372	0.291
GGT (IU/l)	coefficient	0.062	0.299	0.320	0.207	0.188	0.222	0.217	0.182	0.177
	p-value	0.597	**0.009**	**0.005**	0.074	0.106	0.055	0.062	0.119	0.130
Albumin (g/l)	coefficient	-0.168	-0.288	-0.294	-0.292	-0.257	-0.140	-0.138	-0.077	-0.075
	p-value	0.148	**0.012**	**0.010**	**0.010**	**0.025**	0.229	0.235	0.507	0.518
Prothrombin time (s)	coefficient	-0.100	-0.121	-0.127	-0.203	-0.168	-0.063	-0.055	0.036	0.042
	p-value	0.389	0.297	0.274	0.079	0.148	0.591	0.637	0.754	0.716
Platelet (x 10^9/l)	coefficient	0.232	-0.043	-0.039	-0.016	0.002	-0.063	-0.083	-0.182	-0.191
	p-value	**0.043**	0.714	0.741	0.889	0.983	0.591	0.476	0.116	0.098

ALP, alkaline phosphatase; ALT, alanine aminotransferase; AST, aspartate aminotransferase; GGT, gamma-glutamyltransferase; xULN, times upper limit normal; TreeVol; biliary tree volume, DilatNum; number of candidate dilatations per metre, SumRelSevDilat; sum of relative severity of candidate dilatations per metre, DilatTreeProp; proportion of tree dilated, TreeDilatScore; sum of tree dilatation scores, StricNum; number of candidate strictures per metre, SumRelSevStric; sum of relative severity of candidate strictures per metre, StricTreeProp; proportion of tree strictured, TreeStricScore; sum of tree stricture scores.

Table 24. Whole tree *MRCP+* metrics correlation with prognostic risk models and liver fibrosis markers.

Variable	Correlation	TreeVol (mlm^{-1})	DilatNum (m^{-1})	SumRelSevDilat (m^{-1})	DilatTreeProp (%)	TreeDilatScore	StricNum (m^{-1})	SumRelSevStric (m^{-1})	StricTreeProp (%)	TreeStricScore
MRS	coefficient	0.022	0.371	0.380	0.263	0.265	0.250	0.256	0.104	0.123
	p-value	0.852	**0.001**	**0.001**	**0.022**	**0.021**	**0.029**	**0.026**	0.372	0.291
AOM	coefficient	0.385	0.401	0.445	0.465	0.455	0.291	0.279	0.154	0.144
	p-value	**0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**0.011**	**0.015**	0.185	0.215
LS (kPa)	coefficient	-0.049	0.312	0.358	0.188	0.199	0.256	0.265	0.238	0.240
	p-value	0.672	**0.006**	**0.001**	0.104	0.085	**0.025**	**0.021**	**0.039**	**0.037**
ELF	coefficient	0.221	0.208	0.242	0.126	0.102	0.175	0.154	0.134	0.099
	p-value	0.055	0.071	**0.035**	0.277	0.383	0.131	0.183	0.249	0.393

AOM, Amsterdam-Oxford Model score; ELF, enhanced liver fibrosis score; LS, transient elastography liver stiffness; MRS, Mayo Risk Score; TreeVol; biliary tree volume, DilatNum; number of candidate dilatations per metre, SumRelSevDilat; sum of relative severity of candidate dilatations per metre, DilatTreeProp; proportion of tree dilated, TreeDilatScore; sum of tree dilatation scores, StricNum; number of candidate strictures per metre, SumRelSevStric; sum of relative severity of candidate strictures per metre, StricTreeProp; proportion of tree strictured, TreeStricScore; sum of tree stricture scores.

Table 25. Whole tree *MRCP+* metrics correlation with qualitative MRCP scores.

Variable	Correlation	TreeVol (mlm^{-1})	DilatNum (m^{-1})	SumRelSevDilat (m^{-1})	DilatTreeProp (%)	TreeDilatScore	StricNum (m^{-1})	SumRelSevStric (m^{-1})	StricTreeProp (%)	TreeStricScore
ISSS (reader 1)	coefficient	0.129	0.412	0.495	0.423	0.455	0.171	0.170	0.160	0.162
	p-value	0.267	**<0.001**	**<0.001**	**<0.001**	**<0.001**	0.140	0.143	0.167	0.162
ESSS (reader 1)	coefficient	0.025	0.375	0.454	0.377	0.429	0.244	0.248	0.225	0.255
	p-value	0.831	**0.001**	**<0.001**	**0.001**	**<0.001**	**0.034**	**0.031**	0.051	**0.026**
TSSS (reader 1)	coefficient	0.098	0.430	0.518	0.437	0.482	0.229	0.230	0.212	0.230
	p-value	0.399	**<0.001**	**<0.001**	**<0.001**	**<0.001**	0.047	0.045	0.066	0.045
Anali score (reader 1)	coefficient	-0.001	0.270	0.344	0.247	0.299	0.137	0.149	0.091	0.107
	p-value	0.994	**0.018**	**0.002**	**0.031**	**0.009**	0.238	0.198	0.435	0.356

ESSS, modified Amsterdam extrahepatic stricture severity score; ISSS, modified Amsterdam intrahepatic stricture severity score, TSSS, modified Amsterdam total stricture severity score; TreeVol; biliary tree volume, DilatNum; number of candidate dilatations per metre, SumRelSevDilat; sum of relative severity of candidate dilatations per metre, DilatTreeProp; proportion of tree dilated, TreeDilatScore; sum of tree dilatation scores, StricNum; number of candidate strictures per metre, SumRelSevStric; sum of relative severity of candidate strictures per metre, StricTreeProp; proportion of tree strictured, TreeStricScore; sum of tree stricture scores.

Table 26. Intrahepatic tree *MRCP+* metrics correlation with blood tests.

Variables	Correlation	TreeVol (mlm^{-1})	DilatNum (m^{-1})	SumRelSevDilat (m^{-1})	DilatTreeProp (%)	TreeDilatScore	StricNum (m^{-1})	SumRelSevStric (m^{-1})	StricTreeProp (%)	TreeStricScore
Total bilirubin (µmol/l)	coefficient	-0.019	0.224	0.212	0.188	0.156	0.059	0.077	0.027	0.041
	p-value	0.869	0.052	0.066	0.105	0.180	0.612	0.507	0.817	0.728
ALT (IU/l)	coefficient	0.074	0.095	0.124	0.038	0.046	0.146	0.128	0.120	0.113
	p-value	0.527	0.416	0.284	0.742	0.696	0.209	0.271	0.301	0.330
AST (IU/l)	coefficient	0.175	0.218	0.263	0.113	0.128	0.241	0.222	0.208	0.202
	p-value	0.131	0.058	**0.021**	0.330	0.269	**0.036**	0.054	0.071	0.080
ALP (IU/l)	coefficient	0.296	0.360	0.403	0.301	0.319	0.256	0.258	0.115	0.124
	p-value	**0.009**	**0.001**	**<0.001**	**0.008**	**0.005**	**0.026**	**0.024**	0.322	0.288
GGT (IU/l)	coefficient	0.306	0.287	0.333	0.227	0.242	0.203	0.176	0.132	0.116
	p-value	**0.008**	**0.013**	**0.003**	0.050	**0.036**	0.081	0.131	0.259	0.322
Albumin (g/l)	coefficient	-0.401	-0.316	-0.292	-0.235	-0.211	-0.128	-0.120	0.015	0.007
	p-value	**<0.001**	**0.005**	**0.010**	**0.041**	0.067	0.271	0.301	0.897	0.953
Prothrombin time (s)	coefficient	-0.192	-0.109	-0.097	-0.183	-0.186	-0.080	-0.070	-0.018	-0.004
	p-value	0.096	0.349	0.406	0.114	0.108	0.492	0.549	0.878	0.970
Platelet (x 10^9/l)	coefficient	0.143	-0.025	-0.066	0.030	0.050	-0.008	-0.014	-0.122	-0.130
	p-value	0.218	0.830	0.571	0.796	0.669	0.948	0.906	0.295	0.264

ALP, alkaline phosphatase; ALT, alanine aminotransferase; AST, aspartate aminotransferase; GGT, gamma-glutamyltransferase; xULN, times upper limit normal; TreeVol; biliary tree volume, DilatNum; number of candidate dilatations per metre, SumRelSevDilat; sum of relative severity of candidate dilatations per metre, DilatTreeProp; proportion of tree dilated, TreeDilatScore; sum of tree dilatation scores, StricNum; number of candidate strictures per metre, SumRelSevStric; sum of relative severity of candidate strictures per metre, StricTreeProp; proportion of tree strictured, TreeStricScore; sum of tree stricture scores.

Table 27. Intrahepatic *MRCP+* metrics correlation with prognostic risk models and liver fibrosis markers.

Variables	Correlation	TreeVol (mlm^{-1})	DilatNum (m^{-1})	SumRelSevDilat (m^{-1})	DilatTreeProp (%)	TreeDilatScore	StricNum (m^{-1})	SumRelSevStric (m^{-1})	StricTreeProp (%)	TreeStricScore
MRS	coefficient	0.577	0.467	0.507	0.478	0.478	0.361	0.352	0.130	0.132
	p-value	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**0.001**	**0.002**	0.262	0.256
AOM	coefficient	0.364	0.358	0.411	0.327	0.304	0.248	0.242	0.191	0.186
	p-value	**0.001**	**0.001**	**<0.001**	**0.004**	**0.008**	**0.031**	**0.035**	0.098	0.107
LS (kPa)	coefficient	0.304	0.340	0.458	0.267	0.284	0.249	0.264	0.165	0.178
	p-value	**0.008**	**0.003**	**<0.001**	**0.020**	**0.013**	**0.030**	**0.021**	0.154	0.124
ELF	coefficient	0.226	0.254	0.300	0.251	0.226	0.201	0.187	0.050	0.053
	p-value	**0.049**	**0.027**	**0.008**	**0.029**	0.050	0.081	0.105	0.670	0.649

AOM, Amsterdam-Oxford Model score; ELF, enhanced liver fibrosis score; LS, transient elastography liver stiffness; MRS, Mayo Risk Score; TreeVol; biliary tree volume, DilatNum; number of candidate dilatations per metre, SumRelSevDilat; sum of relative severity of candidate dilatations per metre, DilatTreeProp; proportion of tree dilated, TreeDilatScore; sum of tree dilatation scores, StricNum; number of candidate strictures per metre, SumRelSevStric; sum of relative severity of candidate strictures per metre, StricTreeProp; proportion of tree strictured, TreeStricScore; sum of tree stricture scores.

Table 28. Intrahepatic tree *MRCP+* metrics correlation with qualitative MRCP scores.

Variables	Correlation	TreeVol (mlm^{-1})	DilatNum (m^{-1})	SumRelSevDilat (m^{-1})	DilatTreeProp (%)	TreeDilatScore	StricNum (m^{-1})	SumRelSevStric (m^{-1})	StricTreeProp (%)	TreeStricScore
ISSS (reader 1)	coefficient	0.460	0.371	0.427	0.325	0.351	0.138	0.109	0.135	0.129
	p-value	**<0.001**	**0.001**	**<0.001**	**0.004**	**0.002**	0.235	0.349	0.244	0.267
ESSS (reader 1)	coefficient	0.452	0.332	0.429	0.263	0.340	0.252	0.247	0.245	0.249
	p-value	**<0.001**	**0.003**	**<0.001**	**0.022**	**0.003**	**0.028**	**0.032**	**0.033**	**0.030**
TSSS (reader 1)	coefficient	0.492	0.380	0.463	0.312	0.370	0.216	0.196	0.217	0.214
	p-value	**<0.001**	**0.001**	**<0.001**	**0.006**	**0.001**	0.061	0.091	0.060	0.063
Anali score (reader 1)	coefficient	0.357	0.308	0.422	0.317	0.369	0.217	0.241	0.133	0.158
	p-value	**0.002**	**0.007**	**<0.001**	**0.005**	**0.001**	0.060	**0.036**	0.250	0.172

ESSS, modified Amsterdam extrahepatic stricture severity score; ISSS, modified Amsterdam intrahepatic stricture severity score, TSSS, modified Amsterdam total stricture severity score; TreeVol; biliary tree volume, DilatNum; number of candidate dilatations per metre, SumRelSevDilat; sum of relative severity of candidate dilatations per metre, DilatTreeProp; proportion of tree dilated, TreeDilatScore; sum of tree dilatation scores, StricNum; number of candidate strictures per metre, SumRelSevStric; sum of relative severity of candidate strictures per metre, StricTreeProp; proportion of tree strictured, TreeStricScore; sum of tree stricture scores.

4.6.6 Performance of all *MRCP+* metrics to discriminate high-risk vs low-risk PSC

The performance of *MRCP+* metrics were compared between the high-risk and low-risk PSC groups stratified by prognostic risk models (Table 29), serum ALP (Table 30), liver fibrosis markers (Table 31), qualitative MRI-MRCP scores (Table 32), and presence of extrahepatic disease or dominant stricture (Table 33). In summary, dilatation metrics were significantly higher in the high-risk PSC group compared to the low-risk group when stratified by all non-invasive risk stratification tools. Stricture metrics, however, were no different between the high-risk and low-risk PSC groups, except when stratified by the presence of extrahepatic disease.

Candidate MRCP+ metric

Intrahepatic dilatation metrics performed better than whole tree dilatation metrics. The intrahepatic sum of relative severity of dilatations per metre of the biliary tree (SumRelSevDilat; m^{-1}) demonstrated the strongest correlation with non-invasive markers and the best risk stratification ability amongst all the dilatation metrics when stratified by non-invasive tools. The inter-observer agreement for this *MRCP+* metric was better compared to qualitative MRI-MRCP scores (Table 34). The performance of this metric is specifically reported herein.

Table 29. *MRCP+* metrics stratified into high-risk and low-risk groups by Mayo Risk Score (MRS > 0 or ≤ 0) and Amsterdam-Oxford Model (AOM > 2 or ≤ 2).

	Mayo Risk Score (MRS)					Amsterdam-Oxford Model (AOM)				
	MRS > 0 (n = 43)		MRS ≤ 0 (n = 33)		p-value	AOM > 2 (n = 23)		AOM ≤ 2 (n = 53)		p-value
Intrahepatic tree										
TreeVol (mlm^{-1})	6.9	(5.5-8.2)	5.5	(4.8-5.8)	<0.0001	6.4	(5.5-7.7)	5.7	(5.0-7.7)	0.0685
Dilatation metrics										
DilatNum (m^{-1})	12.6	(9.9-15.6)	8.3	(6.0-11.5)	0.0002	14.1	(11.3-15.6)	9.7	(7.6-13.1)	0.0046
SumRelSevDilat (m^{-1})	8.2	(6.8-9.8)	5.6	(3.8-6.9)	<0.0001	9.1	(7.4-10.1)	6.3	(4.8-8.4)	0.0017
DilatTreeProp (%)	8.3	(6.3-10.2)	5.7	(3.6-7.1)	<0.0001	8.3	(6.4-9.8)	6.3	(4.8-8.4)	0.0195
TreeDilatScore	55.7	(40.9-65.2)	37.5	(20.0-48.7)	<0.0001	55.7	(38.9-64.2)	41.0	(27.1-57.8)	0.0398
Stricture metrics										
StricNum (m^{-1})	5.9	(4.0-8.4)	4.4	(1.6-6.5)	0.0035	5.7	(4.5-8.5)	4.8	(2.9-7.7)	0.1144
SumRelSevStric (m^{-1})	2.5	(1.7-3.5)	1.8	(6.5-2.9)	0.0051	2.4	(2.1-3.7)	1.9	(1.3-3.3)	0.0605
StricTreeProp (%)	3.8	(3.0-5.4)	4.2	(1.6-5.5)	0.2904	4.1	(2.6-7.2)	4.1	(2.1-5.3)	0.3881
TreeStricScore	15.6	(11.9-24.1)	17.7	(6.4-23.7)	0.2671	17.9	(11.2-30.5)	15.8	(8.3-23.2)	0.2565
Whole tree										
TreeVol (mlm^{-1})	8.1	(7.1-9.6)	6.7	(5.8-9.0)	0.0027	7.1	(6.4-8.4)	7.8	(6.3-9.3)	0.5287
Dilatation metrics										
DilatNum (m^{-1})	11.6	(8.2-15.0)	7.8	(5.5-11.3)	0.0029	12.0	(10.3-14.9)	8.5	(6.6-12.9)	0.0130
SumRelSevDilat (m^{-1})	7.6	(6.1-9.3)	5.0	(3.4-7.2)	0.0008	8.2	(6.8-9.6)	5.7	(3.9-7.9)	0.0062
DilatTreeProp (%)	7.5	(6.4-9.3)	5.3	(3.3-7.5)	0.0030	7.6	(6.5-8.9)	6.3	(4.0-7.8)	0.0340
TreeDilatScore	51.7	(40.9-63.9)	33.0	19.8-50.7)	0.0005	50.7	(38.2-61.2)	40.0	(24.6-55.3)	0.0482
Stricture metrics										
StricNum (m^{-1})	6.7	(4.5-9.5)	4.7	(2.3-7.5)	0.0206	6.5	(5.0-9.8)	5.3	(3.0-7.8)	0.0488
SumRelSevStric (m^{-1})	2.9	(1.8-3.8)	2.0	(1.0-3.3)	0.0319	2.9	(2.1-4.4)	2.2	(1.2-3.3)	0.0278
StricTreeProp (%)	4.7	(3.0-7.3)	5.0	(2.0-5.8)	0.4917	5.2	(3.3-7.6)	4.4	(2.4-5.8)	0.0835
TreeStricScore	19.6	(13.0-32.1)	19.7	(8.3-24.8)	0.3851	22.9	(13.7-32.1)	17.5	(9.4-24.1)	0.0610

Table 30. *MRCP+* metrics stratified into high-risk and low-risk groups by times upper limit of normal serum alkaline phosphatase (xULN ALP > 2.2 or ≤ 2.2) and (xULN ALP > 1.5 or ≤ 1.5).

	Upper limit of normal (ULN) serum ALP					Upper limit of normal (ULN) serum ALP				
	> 2.2 x ULN (n = 13)		≤ 2.2 x ULN (n = 63)		p-value	> 1.5 x ULN (n = 25)		≤ 1.5 x ULN (n = 51)		p-value
Intrahepatic tree										
TreeVol (mlm^{-1})	7.4	(6.4-7.8)	5.7	(5.0-7.6)	0.0023	6.7	(5.6-7.8)	5.6	(5.0-7.6)	0.0322
Dilatation metrics										
DilatNum (m^{-1})	15.4	(12.7-17.1)	10.1	(7.6-13.3)	0.0008	12.4	(9.1-15.3)	9.5	(7.5-14.2)	0.0383
SumRelSevDilat (m^{-1})	8.9	(8.1-11.6)	6.7	(4.8-9.0)	0.0007	8.2	(6.6-9.5)	6.6	(4.5-9.1)	0.0225
DilatTreeProp (%)	8.3	(7.7-10.0)	6.4	(4.4-9.0)	0.0117	7.5	(6.4-9.2)	6.4	(4.4-9.0)	0.1015
TreeDilatScore	62.1	(44.7-64.2)	40.9	(29.2-60.5)	0.0076	51.0	(39.7-62.3)	40.4	(26.5-63.4)	0.0709
Stricture metrics										
StricNum (m^{-1})	7.2	(4.6-9.6)	4.8	(3.0-7.4)	0.0458	5.7	(4.1-8.1)	4.8	(3.0-7.5)	0.2352
SumRelSevStric (m^{-1})	2.9	(1.8-4.2)	2.2	(1.3-3.2)	0.0768	2.4	(1.8-3.5)	2.0	(1.3-3.2)	0.2060
StricTreeProp (%)	4.3	(3.1-6.6)	4.1	(1.9-5.3)	0.3229	3.8	(2.3-5.3)	4.2	(1.7-6.0)	0.8713
TreeStricScore	17.2	(12.4-29.2)	15.8	(8.6-23.5)	0.4063	15.6	(9.5-23.0)	16	(9.3-23.9)	0.8756
Whole tree										
TreeVol (mlm^{-1})	7.8	(6.8-8.5)	7.5	(6.2-9.2)	0.3596	7.7	(6.5-8.5)	7.5	(6.3-9.2)	0.9213
Dilatation metrics										
DilatNum (m^{-1})	14.5	(12.3-15.9)	9.0	(6.6-12.5)	0.0026	11.8	(8.4-14.9)	8.7	(6.0-13.5)	0.0379
SumRelSevDilat (m^{-1})	8.5	(7.8-10.9)	6.2	(4.0-8.2)	0.0007	7.5	(6.1-9.2)	5.7	(4.0-8.6)	0.0470
DilatTreeProp (%)	7.9	(7.6-9.7)	6.4	(4.1-8.2)	0.0079	7.2	(5.6-8.5)	6.8	(3.7-8.4)	0.2737
TreeDilatScore	56.3	(41.5-62.4)	40.3	(25.2-54.0)	0.0161	47.1	(38.5-58.9)	41.1	(24.6-54.8)	0.2430
Stricture metrics										
StricNum (m^{-1})	7.2	(5.2-9.7)	5.5	(3.0-7.9)	0.0511	6.4	(4.5-9.2)	5.5	(3.1-8.1)	0.3243
SumRelSevStric (m^{-1})	2.9	(2.1-4.2)	2.4	(1.3-3.4)	0.2928	2.7	(1.9-3.9)	2.4	(1.3-3.4)	0.3071
StricTreeProp (%)	4.9	(3.8-6.7)	4.7	(2.5-6.1)	0.2928	4.7	(3.1-6.9)	4.8	(2.6-6.1)	0.7354
TreeStricScore	20.1	(15.6-30.5)	19.6	(10.2-25.9)	0.3968	19.7	(13.5-28.8)	19.5	(10.8-25.9)	0.7171

Table 31. *MRCP+* metrics stratified into high-risk and low-risk groups by transient elastography liver stiffness (LS > 9.6kPa or ≤ 9.6kPa) and enhanced liver fibrosis score (ELF > 9.8 or ≤ 9.8).

	Transient elastography liver stiffness (LS)					Enhanced liver fibrosis score (ELF)				
	LS > 9.6kPa (n = 21)		LS ≤ 9.6 (n = 55)		p-value	ELF > 9.8 (n = 23)		ELF ≤ 9.8 (n = 53)		p-value
Intrahepatic tree only										
TreeVol (mlm^{-1})	6.2	(5.5-7.7)	5.7	(5.0-7.6)	0.0969	6.3	(5.4-7.8)	5.8	(5.0-7.5)	0.1620
Dilatation metrics										
DilatNum (m^{-1})	13.6	(10.7-16.8)	9.9	(7.9-13.2)	0.0061	12.4	(10.3-15.6)	9.9	(7.8-13.6)	0.0463
SumRelSevDilat (m^{-1})	9.1	(7.0-11.5)	6.3	(4.8-8.3)	0.0022	8.1	(6.9-11.4)	6.6	(5.0-8.6)	0.0189
DilatTreeProp (%)	8.3	(6.0-9.8)	6.4	(4.4-9.0)	0.0448	7.6	(5.6-10.3)	6.4	(4.5-8.4)	0.0556
TreeDilatScore	49.5	(39.1-64.1)	42.4	(26.9-62.1)	0.0675	50.4	(38.9-68.9)	42.4	(27.5-57.8)	0.0403
Stricture metrics										
StricNum (m^{-1})	5.6	(4.1-9.5)	4.8	(3.0-7.5)	0.1152	5.5	(3.9-8.5)	4.8	(3.0-7.6)	0.0995
SumRelSevStric (m^{-1})	2.4	(2.0-4.1)	2.0	(1.3-3.2)	0.0758	2.3	(1.6-3.6)	2.0	(1.3-3.2)	0.0862
StricTreeProp (%)	4.7	(2.1-6.8)	3.8	(2.1-5.3)	0.4978	3.8	(1.8-6.1)	4.1	(2.7-5.4)	0.7639
TreeStricScore	19.1	(10.4-29.3)	15.5	(8.5-23.0)	0.3754	15.5	(9.4-26.8)	16.7	(9.1-23.6)	0.6231
Whole tree										
TreeVol (mlm^{-1})	7.1	(6.0-8.4)	7.8	(6.4-9.3)	0.5958	7.8	(6.4-10.2)	7.3	(6.3-8.5)	0.1997
Dilatation metrics										
DilatNum (m^{-1})	12.3	(9.4-15.5)	8.7	(7.1-12.8)	0.0366	11.4	(8.1-14.6)	9.2	(7.0-12.7)	0.2320
SumRelSevDilat (m^{-1})	7.6	(6.5-10.0)	5.7	(4.0-8.2)	0.0126	7.0	(5.4-9.4)	6.1	(4.2-8.2)	0.1259
DilatTreeProp (%)	7.1	(5.9-9.1)	6.6	(3.9-8.2)	0.3538	7.1	(6.0-8.6)	6.6	(4.3-8.3)	0.5548
TreeDilatScore	47.1	(38.4-59.3)	41.1	(24.6-55.8)	0.3802	45.5	(36.9-56.3)	40.4	(25.0-57.2)	0.5032
Stricture metrics										
StricNum (m^{-1})	6.4	(4.6-9.7)	5.4	(3.5-7.8)	0.2018	6.6	(4.4-10.4)	5.3	(3.3-7.8)	0.1815
SumRelSevStric (m^{-1})	2.9	(1.9-4.1)	2.3	(1.3-3.3)	0.2023	2.7	(1.8-4.4)	2.3	(1.3-3.3)	0.1770
StricTreeProp (%)	5.6	(2.9-7.5)	4.6	(2.6-5.7)	0.2209	4.8	(3.1-7.6)	4.8	(2.5-5.8)	0.2554
TreeStricScore	24.2	(11.5-31.5)	18	(10.9-23.8)	0.2200	19.6	(12.1-34.4)	19.6	(10.7-23.9)	0.3030

Table 32. *MRCP+* metrics stratified into high-risk and low-risk groups by modified Amsterdam stricture severity score (intrahepatic (ISSS): > 2 or ≤ 2, whole tree (TSSS): > 4 or ≤ 4) and Anali score (> 2 or ≤ 2).

	Modified Amsterdam stricture severity score					Anali score				
Intrahepatic tree	ISSS > 2 (n = 37)		ISSS ≤ 2 (n = 39)		p-value	> 2 (n = 27)		≤ 2 (n = 49)		p-value
TreeVol (mlm^{-1})	7.2	(5.7-8.4)	5.4	(4.9-5.9)	<0.0001	6.1	(5.4-7.8)	5.7	(5.0-7.3)	0.1160
Dilatation metrics										
DilatNum (m^{-1})	12.6	(10.1-15.8)	8.6	(6.9-12.7)	0.0012	12.7	(8.0-15.6)	10.1	(8.0-13.1)	0.1470
SumRelSevDilat (m^{-1})	8.3	(6.7-10.3)	5.8	(4.5-7.8)	<0.0001	8.3	(6.6-9.8)	6.6	(5.0-8.4)	0.0299
DilatTreeProp (%)	8.3	(6.3-9.8)	5.9	(3.7-7.5)	0.0008	7.8	(5.4-9.8)	6.4	(4.6-8.5)	0.2046
TreeDilatScore	55.8	(41.7-64.8)	38.1	(24.1-49.1)	0.0003	55.3	(38.1-64.3)	41.0	(29.0-59.0)	0.0693
Stricture metrics										
StricNum (m^{-1})	5.5	(4.0-7.9)	4.4	(2.5-7.5)	0.1212	5.6	(3.9-7.8)	4.8	(3.1-8.0)	0.3135
SumRelSevStric (m^{-1})	2.3	(1.7-3.2)	1.9	(0.9-3.0)	0.2088	2.3	(1.6-3.2)	2.0	(1.2-3.1)	0.2006
StricTreeProp (%)	4.3	(2.9-5.3)	3.7	(1.7-5.6)	0.3140	4.3	(2.8-5.3)	3.7	(2.0-5.4)	0.5155
TreeStricScore	16.2	(11.8-23.1)	15.8	(7.6-24.1)	0.3721	18.9	(11.5-23.5)	15.8	(8.4-23.3)	0.3801
Whole tree	TSSS > 4 (n = 29)		TSSS ≤ 4 (n = 47)		p-value	> 2 (n = 27)		≤ 2 (n = 49)		p-value
TreeVol (mlm^{-1})	7.8	(6.9-8.7)	7.4	(6.4-9.6)	0.5804	7.6	(6.6-8.7)	7.7	(6.4-9.7)	0.9141
Dilatation metrics										
DilatNum (m^{-1})	12.0	(9.6-15.5)	8.5	(5.5-11.7)	0.0041	11.1	(7.0-15.0)	9.6	(6.8-12.5)	0.2652
SumRelSevDilat (m^{-1})	7.8	(6.6-9.9)	5.1	(3.5-8.0)	0.0006	8.0	(4.9-9.3)	6.6	(4.2-8.1)	0.0992
DilatTreeProp (%)	7.7	(6.7-9.6)	5.7	(3.5-8.0)	0.0026	7.3	(5.1-9.3)	6.7	(4.4-8.5)	0.4674
TreeDilatScore	54.1	(45.4-64.9)	38.9	(20.9-54.4)	0.0015	53.5	(26.8-63.7)	42.1	(26.8-57.5)	0.2216
Stricture metrics										
StricNum (m^{-1})	5.8	(4.5-8.2)	5.7	(2.6-7.5)	0.1865	5.8	(3.4-8.1)	5.6	(3.6-7.8)	0.8498
SumRelSevStric (m^{-1})	2.3	(1.8-3.4)	2.3	(1.2-3.3)	0.2140	2.6	(1.5-3.4)	2.2	(1.4-3.3)	0.7264
StricTreeProp (%)	4.9	(3.7-6.3)	4.6	(2.2-6.1)	0.1344	4.9	(2.6-6.4)	4.7	(2.7-6.1)	0.8273
TreeStricScore	21.9	(14.9-27.0)	18.5	(8.5-26.3)	0.1268	20.3	(10.2-27.9)	18.5	(10.4-26.5)	0.7058

Table 33. *MRCP+* metrics stratified into high-risk and low-risk groups by the presence of extrahepatic PSC and dominant stricture (present vs absent).

	Presence of extrahepatic PSC					Presence of dominant stricture				
	Yes (n = 41)		No (n = 35)		p-value	Yes (n = 25)		No (n = 51)		p-value
Intrahepatic tree										
TreeVol (mlm^{-1})	6.4	(5.5-7.7)	5.7	(6.4-7.8)	0.0956	7.3	(5.5-8.6)	5.6	(5.0-6.8)	0.0017
Dilatation metrics										
DilatNum (m^{-1})	12.5	(8.9-15.5)	8.9	(7.5-12.7)	0.0180	12.6	(8.3-15.8)	10.1	(7.9-14.1)	0.0861
SumRelSevDilat (m^{-1})	8.1	(5.7-9.6)	6.6	(4.4-8.2)	0.0323	8.3	(6.8-10.0)	6.3	(4.7-8.6)	0.0019
DilatTreeProp (%)	7.1	(5.5-9.2)	6.4	(4.0-8.3)	0.1876	7.8	(5.5-9.4)	6.4	(4.4-8.3)	0.1415
TreeDilatScore	49.2	(36.7-63.0)	42.4	(26.5-62.4)	0.3203	57.6	(38.8-66.8)	40.9	(26.1-55.7)	0.0254
Stricture metrics										
StricNum (m^{-1})	6.2	(3.9-8.6)	4.0	(2.3-5.6)	0.0172	5.9	(4.0-8.3)	4.8	(2.5-7.7)	0.0782
SumRelSevStric (m^{-1})	2.7	(1.7-3.5)	1.7	(0.9-2.4)	0.0310	2.7	(1.9-3.5)	1.9	(0.9-3.0)	0.0212
StricTreeProp (%)	4.6	(3.1-6.2)	2.9	(1.7-5.1)	0.0183	4.6	(3.3-5.4)	3.3	(1.9-5.3)	0.0801
TreeStricScore	18.9	(12.3-25.5)	13.0	(8.2-23.5)	0.0267	20.3	(13.0-25.4)	14.4	(7.6-22.0)	0.0445
Whole tree										
TreeVol (mlm^{-1})	7.6	(6.4-8.4)	7.4	(6.2-9.3)	0.7481	7.8	(6.7-9.2)	7.4	(6.5-9.4)	0.3730
Dilatation metrics										
DilatNum (m^{-1})	11.9	(8.3-14.9)	8.3	(5.9-11.9)	0.0150	11.7	(7.9-15.5)	9.6	(6.2-12.4)	0.0527
SumRelSevDilat (m^{-1})	7.5	(5.4-9.2)	5.3	(3.9-7.9)	0.0280	7.8	(6.1-9.7)	6.4	(4.1-8.2)	0.0072
DilatTreeProp (%)	7.1	(5.4-9.0)	6.6	(3.7-7.7)	0.0935	7.3	(5.6-9.5)	6.8	(4.1-8.4)	0.1418
TreeDilatScore	47.4	(36.3-52.1)	38.4	(24.1-54.4)	0.1669	54.1	(41.8-64.9)	40.9	(23.7-55.4)	0.0371
Stricture metrics										
StricNum (m^{-1})	6.9	(5.3-9.3)	4.5	(2.7-7.7)	0.0040	6.9	(4.7-7.9)	5.5	(2.6-8.0)	0.1093
SumRelSevStric (m^{-1})	3.0	(2.2-3.7)	1.8	(1.0-3.3)	0.0044	2.9	(1.9-3.7)	2.2	(1.1-3.3)	0.0502
StricTreeProp (%)	5.0	(3.5-7.5)	3.6	(1.8-5.1)	0.0113	4.9	(3.6-6.1)	4.6	(2.2-6.4)	0.2720
TreeStricScore	20.5	(15.1-31.5)	13.8	(7.7-22.5)	0.0054	22.1	(14.9-26.5)	18.5	(10.0-27.2)	0.1453

Table 34. Association between intrahepatic sum of relative severity of dilatations (SumRelSevDilat) and qualitative MRI-MRCP scores with non-invasive surrogate markers of disease severity and progression in PSC.

			MRS		AOM		ALP		LS		ELF	
		Inter-observer variability	Correlation coefficient	p-value	Correlation coefficient	p-value	Correlation coefficient	p-value	Correlation coefficient	p-value	Correlation coefficient	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 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

AOM, Amsterdam-Oxford Model score; ICC, intra class correlation coefficient; SumRelSevDilat, sum of the relative severity of candidate dilatations; ELF, enhanced liver fibrosis score, ISSS, intrahepatic stricture severity score; LS, transient elastography liver stiffness; MRS, Mayo Risk Score; TSSS, total stricture severity score; xULN, times upper limit normal.

4.6.7 The performance of intrahepatic sum of relative severity of dilatations (SumRelSevDilat) to detect high-risk PSC

4.6.7.1 Prognostic risk models

Intrahepatic SumRelSevDilat demonstrated a positive correlation with the two risk models (MRS: $r_s = 0.51$, $p < 0.0001$; AOM: $r_s = 0.41$, $p < 0.0001$). High-risk patients (MRS > 0 or AOM > 2) had higher intrahepatic SumRelSevDilat (MRS: $8.2m^{-1}$ vs $5.6m^{-1}$, $p < 0.0001$; AOM: $9.1m^{-1}$ vs $6.3m^{-1}$, $p = 0.0017$) compared to the low-risk patients (Figure 23). Intrahepatic SumRelSevDilat (AUC = 0.77, 95% CI 0.66–0.88) and ELF (AUC = 0.78, 95% CI 0.67–0.88) showed the highest performance to discriminate MRS > 0 vs ≤ 0 (Table 35). The performance of qualitative MRI-MRCP scores were numerically lower (Anali score: AUC = 0.70, 95% CI 0.58–0.81; ISSS: AUC = 0.69, 95% CI 0.56–0.81).

When stratified by AOM, LS (AUC = 0.85, 95% CI 0.74–0.95) had the best performance to discriminate AOM > 2 vs ≤ 2 (Figure 23). The discriminatory performance of Anali score (AUC = 0.79, 95% CI 0.67–0.91) was numerically better than intrahepatic SumRelSevDilat (AUC = 0.72, 95% CI 0.60–0.85) and ISSS (AUC = 0.64, 95% CI 0.50–0.78; Table 35).

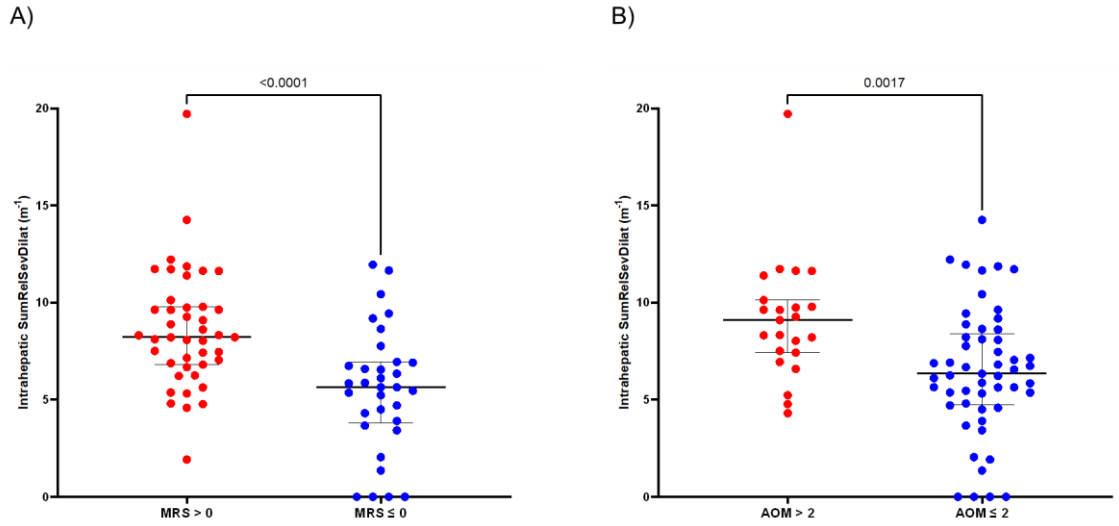


Figure 23. Intrahepatic sum of relative severity of dilatations (SumRelSevDilat) in high-risk vs low-risk groups in 76 patients with large-duct PSC stratified by (A) Mayo Risk Score (MRS); and (B) Amsterdam-Oxford Model score (AOM). Lines and error bars indicate median and interquartile range.

4.6.7.2 Cholestasis markers

Intrahepatic SumRelSevDilat was positively correlated with ALP (r_s = 0.40, p = 0.0003) and was higher in the high-risk group stratified by 2.2x ULN ALP threshold (8.9m^{-1} vs 6.7m^{-1}, p = 0.0007) and 1.5x ULN ALP threshold (8.2m^{-1} vs 6.6m^{-1}, p = 0.0225; Figure 24). Intrahepatic SumRelSevDilat (AUC = 0.79, 95% CI 0.68–0.90) had the best discriminatory performance to detect > 2.2x ULN ALP group and numerically performed better than Anali score (AUC = 0.71, 95% CI 0.57–0.84) and ISSS (AUC = 0.72, 95% CI 0.60–0.85; Table 35). The performance of all three were weak at the 1.5x ULN threshold.

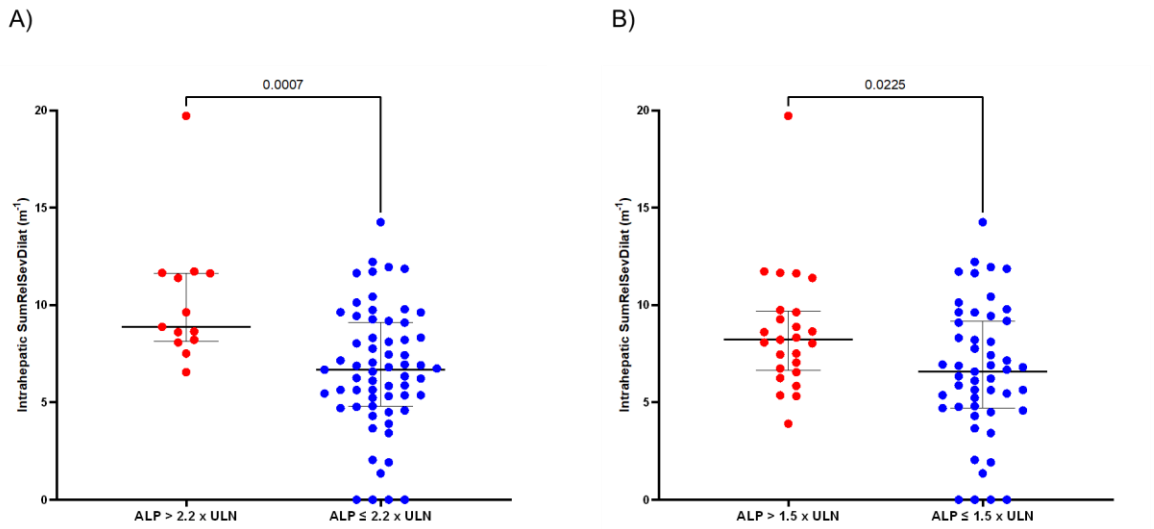


Figure 24. Intrahepatic sum of relative severity of dilatations (SumRelSevDilat) in high-risk vs low-risk groups in 76 patients with large-duct PSC stratified by (A) 2.2x; and (B) 1.5x the upper limit of normal serum ALP. Lines and error bars indicate median and interquartile range.

4.6.7.3 Liver fibrosis markers

Intrahepatic SumRelSevDilat correlated with LS (r_s = 0.46, p <0.001) and ELF (ELF: r_s = 0.30, p = 0.0085). Patients had higher intrahepatic SumRelSevDilat when stratified by LS > 9.6kPa (9.1m^{-1} vs 6.3m^{-1}, p = 0.0022) and ELF > 9.8 (8.1m^{-1} vs 6.6m^{-1}, p = 0.0189; Figure 25). Anali score discriminated ≥F3 from ≤F2 fibrosis (LS: AUC = 0.76, 95% CI 0.66–0.88; 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: 0.68, 95% CI 0.55–0.81; Table 35). The performance of ISSS was numerically lower with no significant difference between the high-risk and low-risk groups.

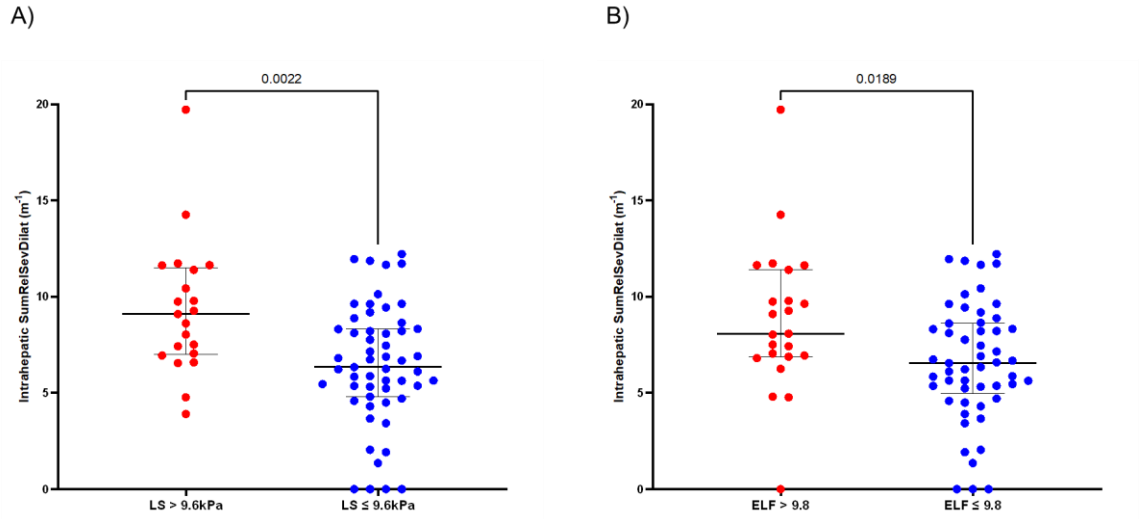


Figure 25. Intrahepatic sum of relative severity of dilatations (SumRelSevDilat) in high-risk vs low-risk groups in 76 patients with large-duct PSC stratified by (A) Transient elastography liver stiffness (LS); and (B) enhanced liver fibrosis score (ELF). Lines and error bars indicate median and interquartile range.

4.6.7.4 Qualitative MRI-MRCP scores

There was a similar correlation between intrahepatic SumRelSevDilat and either ISSS (r_s = 0.43, p = 0.0001) or Anali score (r_s = 0.42, p = 0.0001). Patients classified with ISSS >2 (8.3m^{-1} vs 5.9m^{-1}, p < 0.0001) and Anali score > 2 (8.3m^{-1} vs 6.6m^{-1}, p = 0.0299) had higher intrahepatic SumRelSevDilat (Figure 26). The performance of intrahepatic SumRelSevDilat to discriminate ISSS > 2 vs ≤ 2 (AUC = 0.76, 95% CI 0.65–0.88) was numerically better than to discriminate Anali score > 2 vs ≤ 2 (AUC = 0.65, 95% CI 0.52–0.78; Table 35).

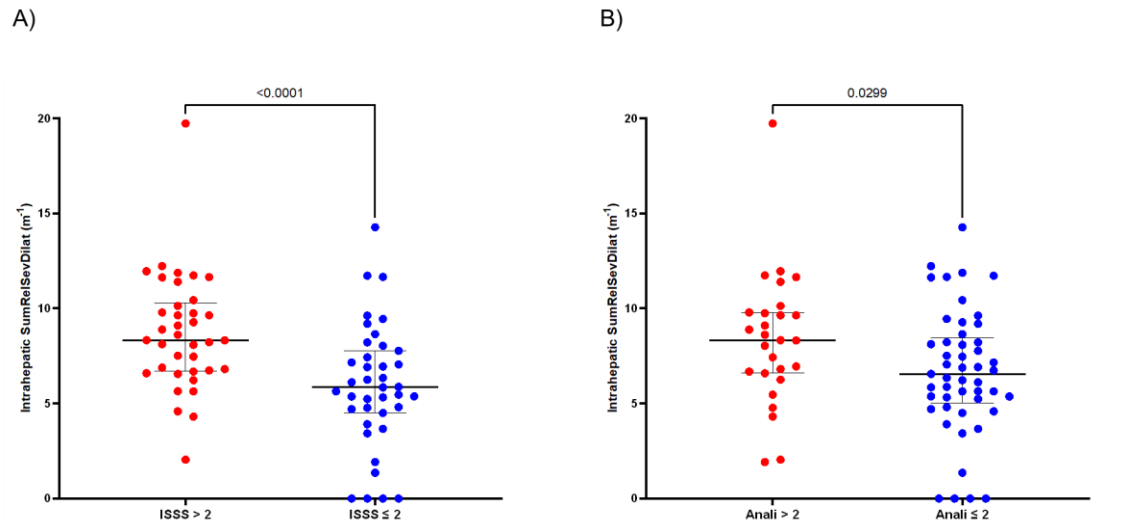


Figure 26. Intrahepatic sum of relative severity of dilatations (SumRelSevDilat) in high-risk vs low-risk groups in 76 patients with large-duct PSC stratified by (A) modified Amsterdam intrahepatic stricture severity score (ISSS); and (B) Anali score. Lines and error bars indicate median and interquartile range.

4.6.7.5 Presence of extrahepatic disease 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 (8.1m^{-1} vs 6.6m^{-1}, p = 0.0323; Figure 27). Extrahepatic stricture metrics were significantly higher in patients with both extrahepatic and intrahepatic disease (Table 33).

When a dominant stricture was present, the intrahepatic SumRelSevDilat was higher (8.3m^{-1} vs 6.3m^{-1}, p = 0.0019; Figure 27). Intrahepatic stricture metrics were also significantly higher in these patients (Table 33).

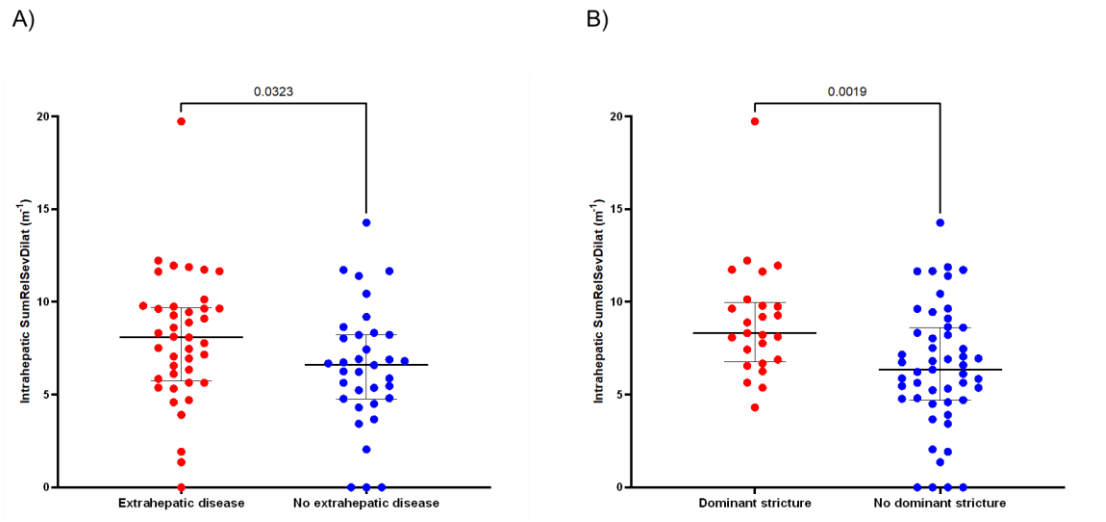


Figure 27. Intrahepatic sum of relative severity of dilatations (SumRelSevDilat) in high-risk vs low-risk groups in 76 patients with large-duct PSC stratified by the presence of (A) extrahepatic disease; and (B) dominant stricture. Lines and error bars indicate median and interquartile range.

Table 35. Performance of non-invasive markers to detect the high-risk PSC group stratified by non-invasive risk stratification tools.

		AUC	95% CI	p-value	Cut-off	Sensitivity (%)	Specificity (%)
Mayo risk score							
SumRelSevDilat (m^{-1})		0.77	0.66-0.88	<0.001	7.0	72	79
ISSS	Reader 1	0.69	0.56-0.81	0.005	-	67	76
	Reader 2	0.74	0.62-0.85	<0.001	-	72	76
Anali	Reader 1	0.70	0.58-0.81	0.004	-	70	64
score	Reader 2	0.79	0.69-0.90	<0.001	-	72	79
LS (kPa)		0.73	0.61-0.84	0.001	6.8	67	67
ELF		0.78	0.67-0.88	<0.001	9.4	74	79
xULN ALP		0.67	0.54-0.79	0.014	1.5	65	70
Amsterdam-Oxford model†							
SumRelSevDilat (m^{-1})		0.72	0.60-0.85	0.002	6.9	83	62
ISSS	Reader 1	0.64	0.50-0.78	0.057	-	70	60
	Reader 2	0.63	0.50-0.76	0.081	-	70	57
Anali	Reader 1	0.79	0.67-0.91	<0.001	-	74	81
score	Reader 2	0.80	0.68-0.91	<0.001	-	87	66
LS (kPa)		0.85	0.74-0.95	<0.001	8.8	74	85
ELF		0.79	0.65-0.92	<0.001	10.4	61	94
ALP x2.2 ULN							
SumRelSevDilat (m^{-1})		0.79	0.68-0.90	0.001	7.5	92	65
ISSS	Reader 1	0.72	0.60-0.85	0.012	-	85	59
	Reader 2	0.66	0.51-0.81	0.072	-	85	44
Anali	Reader 1	0.71	0.57-0.84	0.020	-	100	37
score	Reader 2	0.62	0.45-0.80	0.166	-	39	83
LS (kPa)		0.77	0.65-0.89	0.002	7.4	85	67
ELF		0.72	0.57-0.87	0.013	9.3	85	51
ALP x1.5 ULN							
SumRelSevDilat (m^{-1})		0.66	0.54-0.78	0.023	7.0	72	61
ISSS	Reader 1	0.61	0.48-0.74	0.110	-	64	59
	Reader 2	0.57	0.43-0.70	0.353	-	64	55
Anali	Reader 1	0.64	0.51-0.76	0.057	-	92	41
score	Reader 2	0.54	0.40-0.68	0.565	-	40	69
LS (kPa)		0.72	0.59-0.84	0.002	6.9	76	65
ELF		0.66	0.53-0.79	0.022	9.8	52	74
Liver stiffness (kPa)							
SumRelSevDilat (m^{-1})		0.73	0.60-0.85	0.002	6.4	91	61
ISSS	Reader 1	0.63	0.50-0.77	0.076	-	67	58
	Reader 2	0.62	0.49-0.75	0.109	-	67	54
Anali	Reader 1	0.77	0.66-0.88	<0.001	-	91	58
score	Reader 2	0.72	0.59-0.85	0.003	-	76	60
ULN ALP		0.68	0.54-0.81	0.018	1.6	57	78
Enhanced liver fibrosis score							
SumRelSevDilat (m^{-1})		0.68	0.55-0.81	0.013	6.8	83	57
ISSS	Reader 1	0.56	0.42-0.70	0.412	-	61	57
	Reader 2	0.55	0.41-0.69	0.476	-	83	32
Anali	Reader 1	0.71	0.58-0.83	0.004	-	61	75
score	Reader 2	0.75	0.63-0.87	0.001	-	78	72
ULN ALP		0.65	0.52-0.79	0.037	1.6	52	77

†Performance of xULN ALP not tested against AOM as the model already includes serum ALP.

4.6.7.6 Independent predictors of high-risk PSC

Univariate and multivariate analysis results are summarised in Table 36. The individual performance non-invasive markers to discriminate the high-risk vs low-risk PSC groups stratified by MRS and AOM, and the multivariate logistic regression model performance to detect the high-risk group is summarised in Figure *28*. In the multivariate logistic regression model, intrahepatic SumRelSevDilat (OR: 29.2, p = 0.038) and ELF (OR: 3.5, p = 0.042) were associated with an increased likelihood of having high-risk PSC when stratified by MRS (Table 37). The model correctly classified 83% of cases (AUC = 0.91, 95% CI 0.85–0.98) (Figure 28). LS (OR: 1.3, p = 0.033) was the only covariate associated with increased likelihood of having high-risk PSC when stratified by AOM (Table 38).

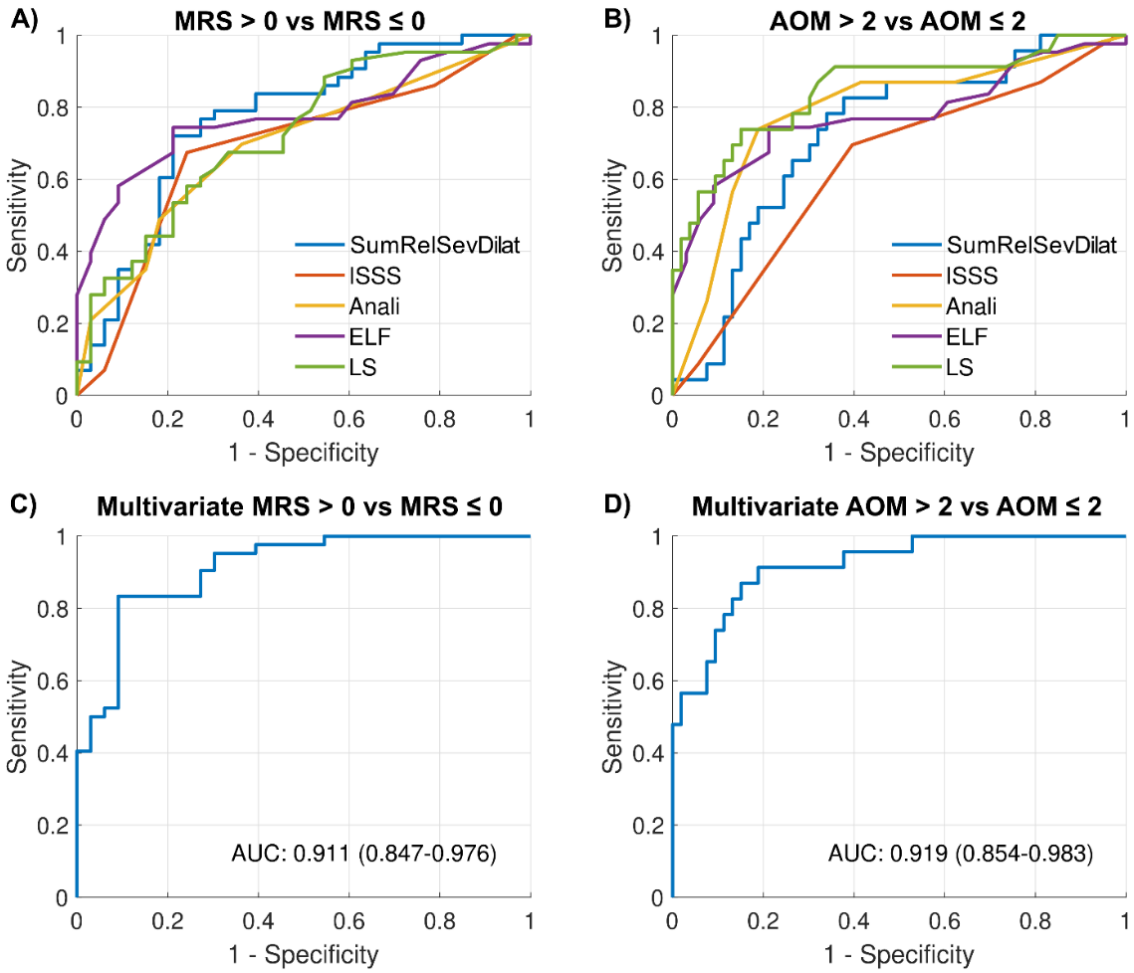


Figure 28. ROC curves summarising the individual high-risk vs low-risk discriminatory performance of non-invasive risk stratification tools and the multivariable logistic regression models for identifying the high-risk PSC group. (A) Mayo Risk Score (MRS: Intrahepatic SumRelSevDilat AUC = 0.77 (0.66–0.88), ISSS AUC = 0.69 (0.56–0.81), Anali AUC = 0.70 (0.58–0.81), ELF AUC = 0.78 (0.67–0.88), and LS AUC = 0.73 (0.61–0.84)); (B) Amsterdam-Oxford Model (AOM: Intrahepatic SumRelSevDilat AUC=0.73 (0.60–0.85), ISSS AUC = 0.64 (0.50–0.78), Anali AUC = 0.79 (0.67–0.91), ELF AUC = 0.79 (0.65–0.92), and LS AUC = 0.85 (0.74–0.95)). Multiple logistic regression model performance to detect high-risk PSC stratified by (C) MRS with independent predictors intrahepatic SumRelSevDilat (OR: 29.2, p = 0.038) 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, enhanced liver fibrosis score; ISSS, modified Amsterdam intrahepatic stricture severity score; LS, transient elastography liver stiffness.

Table 36. Univariate and multiple logistic regression analysis for the detection of high-risk PSC defined by Mayo Risk Score and Amsterdam-Oxford Model score.

Variable	Mayo Risk score		Amsterdam-Oxford Model	
	Univariate p-value	Multivariate p-value	Univariate p-value	Multivariate p-value
MRCP+ dilatation metrics				
DilatNum (m^{-1})	0.001	-	0.011	-
SumRelSevDilat (m^{-1})†	<0.001	0.035	0.005	0.330
DilatTreeProp (%)	0.001	-	0.055	-
TreeDilatScore	0.001	-	0.050	-
TreeVol (mlm^{-1})	0.001	-	0.090	-
Liver fibrosis markers				
LS (kPa)	0.012	0.944	<0.001	0.033
ELF	<0.001	0.042	0.000	0.586
Qualitative MRI-MRCP scores				
Reader 1				
ISSS	0.010	0.657	0.019	0.907
Anali score	0.004	0.439	<0.001	0.511
Dominant stricture	0.028	0.325	0.336	
*Reader 2**				
ISSS	0.001		0.072	
Anali score	<0.001		<0.001	
Dominant stricture	0.029		0.011	
Laboratory parameters				
ALP (IU/l)	0.223		‡	
GGT (IU/l)	0.039	0.802	0.243	
Total bilirubin	‡		‡	
ALT (IU/l)	0.248		0.755	
AST (IU /l)	‡		‡	
Albumin (g/l)	‡		‡	
Platelet count ($x10^9$/l)	0.690		‡	
Prothrombin time (s)	0.017	0.099	0.108	
Clinical features				
Age	‡		‡	
Gender	0.764		0.764	
PSC duration (years)	0.250		0.125	
Presence of UC	0.912		0.884	
Extrahepatic disease	0.080		0.426	

†The strongest dilatation metric variable was entered into multivariate analysis given the multicollinearity between quantitative MRCP metrics. Statistical interaction terms between covariates were not analysed.

*Reader 2 data was excluded from the multivariate analysis given the correlation with reader 1.

‡These variables are already incorporated into MRS or AOM, therefore excluded from analysis.

Abbreviations: ALP, alkaline phosphatase; ALT, alanine aminotransferase; AST, aspartate aminotransferase; DilatNum; number of candidate dilatations per metre, SumRelSevDilat; sum of relative severity of candidate dilatations per metre, DilatTreeProp; ELF, enhanced liver fibrosis score; GGT, gamma-glutamyltransferase; ISS, intrahepatic stricture severity score; LS, transient elastography liver stiffness; TreeDilatScore; tree dilatation score, TreeVol, biliary tree volume per metre, UC, ulcerative colitis; xULN, times upper limit of normal.

Table 37. Multivariate logistic regression model parameters to detect Mayo Risk Score (MRS) > 0 group.

Variable	B	OR	95% CI for OR		p-value
SumRelSevDilat	3.374	29.186	1.205	707.049	0.038
LS	0.008	1.008	0.793	1.282	0.948
ELF	1.239	3.452	1.050	11.345	0.041
GGT	0.000	1.000	0.996	1.004	0.947
PT	-1.285	0.277	0.061	1.255	0.096
ISSS†					0.400
ISSS (1)	-19.092	0.000	0.000	-	1.000
ISSS (2)	-0.421	0.656	0.031	13.968	0.787
ISSS (3)	-1.209	0.299	0.017	5.288	0.410
ISSS (4)	0.872	2.392	0.138	41.522	0.549
Anali score†					0.654
Anali (1)	0.073	1.076	0.034	34.282	0.967
Anali (2)	-1.154	0.315	0.011	8.704	0.495
Anali (3)	-1.150	0.317	0.014	7.341	0.473
Anali (4)	1.783	5.947	0.083	425.144	0.413
Anali (5)	-0.787	0.455	0.008	26.939	0.705
DS	-0.948	0.388	0.059	2.559	0.325
Intercept	0.363	1.438			0.971

†Entered as categorical variables into multivariate model. DS, dominant stricture; ELF, enhanced liver fibrosis score; GGT, gamma-glutamyltransferase; ISSS, modified Amsterdam intrahepatic stricture severity score; LS, transient elastography liver stiffness; PT, prothrombin time; SumRelSevDilat; sum of relative severity of candidate dilatations per metre.

Table 38. Multivariate logistic regression model parameters to detect Amsterdam-Oxford Model (AOM) > 2 group.

Variable	B	OR	95% CI for OR		p-value
SumRelSevDilat	1.447	4.251	0.232	77.993	0.330
LS	0.270	1.310	1.022	1.678	0.033
ELF	0.241	1.273	0.534	3.036	0.586
ISSS†					0.907
ISSS (1)	-16.915	0.000	0.000	-	1.000
ISSS (2)	0.954	2.597	0.017	407.973	0.711
ISSS (3)	0.686	1.985	0.016	247.002	0.781
ISSS (4)	1.704	5.496	0.049	616.221	0.479
Anali score†					0.511
Anali (1)	0.565	1.759	0.074	41.987	0.727
Anali (2)	-20.686	0.000	0.000		0.998
Anali (3)	-1.009	0.365	0.034	3.938	0.406
Anali (4)	1.077	2.935	0.151	57.020	0.477
Anali (5)	1.498	4.474	0.253	79.045	0.307
Intercept	-8.085	0.000			0.079

†Entered as categorical variables into multivariate model. ELF, enhanced liver fibrosis score; ISSS, modified Amsterdam intrahepatic stricture severity score; LS, transient elastography liver stiffness; SumRelSevDilat; sum of relative severity of candidate dilatations per metre.

4.6.8 Association between *MRCP+* metrics, other study variables and patient-reported outcomes in PSC

MRCP+

To recap, the SF-36 is composed of 8 domains summarised into physical and mental component scores, and PSC-PRO is made up of 2 modules, to assess for PSC symptoms and impact of symptoms (section 2.8 Figure 9).

The distribution of mean domain scores for SF-36 and PSC-PRO is shown in Figure 29 and Figure 30 respectively. In the physical component of SF-36, physical functioning mean ± standard deviation score was the highest at 87.6 ± 20.1% and general health the lowest at 49.6 ± 22.3%. In the mental component of SF-36, social functioning mean score was the highest at 83.2 ± 21.7% and energy/fatigue score the lowest at 54.5 ± 21.2%. In the PSC-PRO, the PSC symptoms score was low at 0.2 ± 0.9 with mean domain scores ranging from 1.2–1.6 out of a maximum of 5.

The intrahepatic sum of relative severity of dilatations had negative correlations with the physical functioning (r_s = -0.25, p = 0.038) and bodily pain (r_s = -0.26, p = 0.025) mean domain scores of the SF-36 (Table 39). There were no significant correlations with any of the other domains of SF-36. In the PSC-PRO, physical function (r_s = 0.26, p = 0.029) correlated with intrahepatic sum of relative severity of dilatations and activities of daily living (r_s = 0.29, p = 0.014) correlated with both intrahepatic number of candidate dilatations and sum of relative severity of dilatations. The whole tree number of candidate dilatations (r_s = 0.30, p = 0.011), sum of relative severity of dilatations (r_s = 0.29, p = 0.015), dilated tree proportion (r_s = 0.34, p = 0.004), and tree dilatation score (r_s = 0.30, p = 0.011) all showed significant positive correlations with activities of daily living domain of the PSC-PRO (Table 39). Other *MRCP+* metrics did not show significant correlation with any of the SF-36 or PSC-PRO domains.

MRCP

Amongst the qualitative MRI-MRCP features, bodily pain domain score from SF-36 showed significant correlation with the presence of dominant stricture when assessed by reader 1 (r_s = -0.28, p = 0.015) and borderline significance when assessed by reader 2 (r_s = -0.25, p = 0.044). Physical function domain score from PSC-PRO showed significant correlation with intrahepatic bile duct dilatation component of the Anali score when assessed by reader 1 (r_s = 0.36, p = 0.002) and borderline significance when assessed by reader 2 (r_s = 0.24, p = 0.049).

Serum ALP

Serum ALP was negatively correlated with mean scores of bodily pain (r_s = -0.37, p = 0.001), general health (r_s = -0.26, p = 0.026), physical component summary (r_s = -0.30, p = 0.008), energy/fatigue (r_s = -0.29, p = 0.012), and mental component summary (r_s = -0.29, p = 0.012) domains of SF-36 (Table 40). There was positive correlation with the physical function (r_s = 0.30, p = 0.009) and emotional impact (r_s = 0.38, p = 0.001) domain scores of PSC-PRO (Table 40).

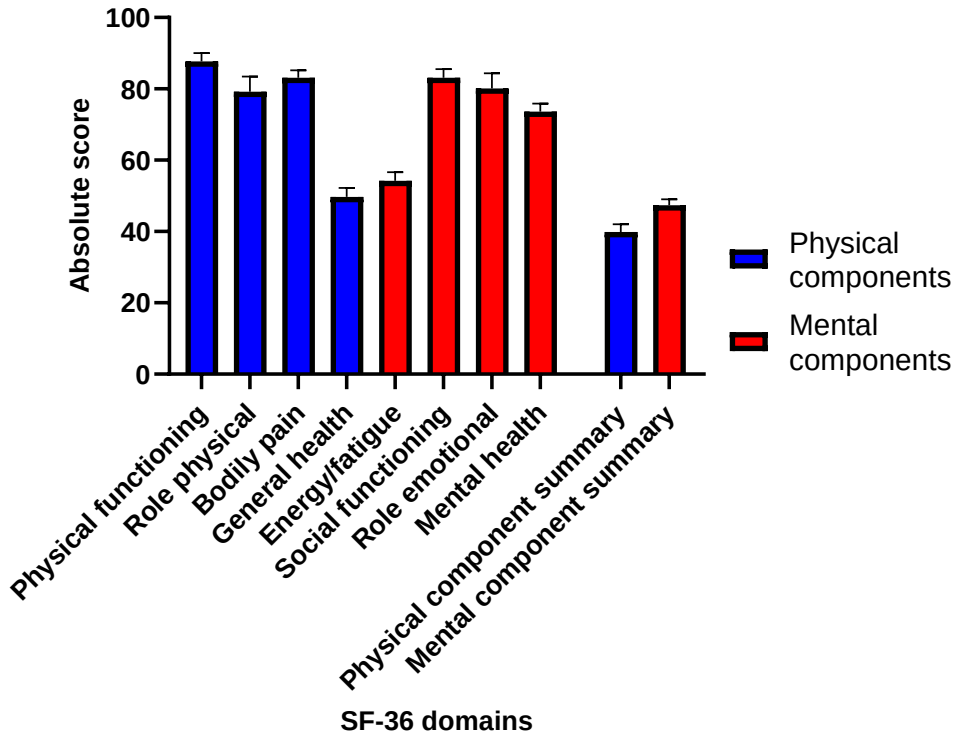


Figure 29. Summary of SF-36 mean domain and summary aggregate scores in patients with PSC. Lines and error bars indicate mean and standard error.

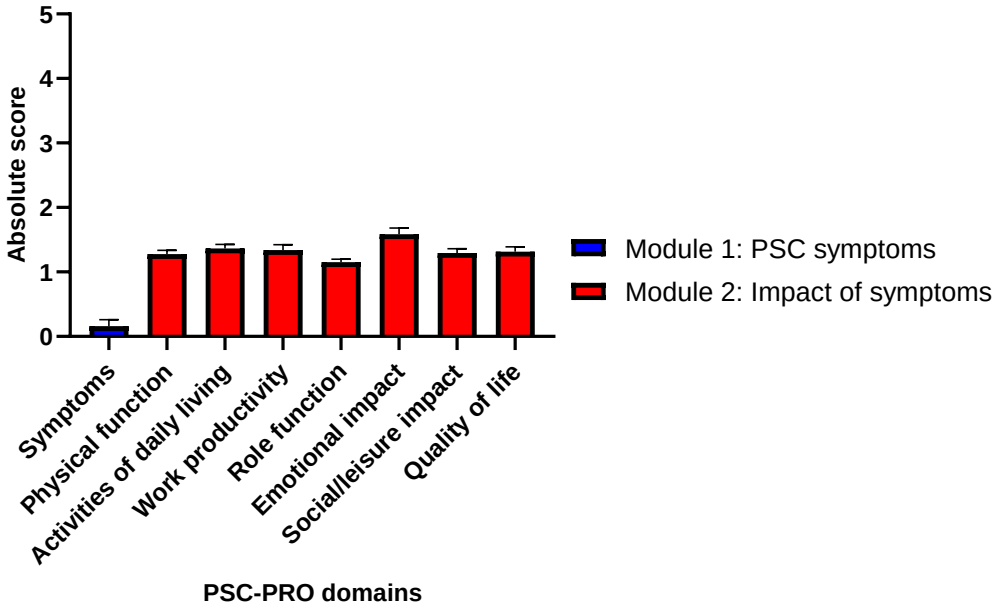


Figure 30. Summary of PSC-PRO mean domain scores in patients with PSC. Lines and error bars indicate mean and standard error.

Table 39. Correlation between SF-36 and PSC-PRO domains with the intrahepatic sum of relative severity of candidate dilatations metric (SumRelSevDilat).

	Correlation coefficient	p-value
SF-36 domain		
Physical component summary	-0.242	0.040
Physical functioning	-0.245	0.038
Role physical	-0.237	0.045
Bodily pain	-0.264	0.025
General health	-0.015	0.901
Mental component summary	-0.098	0.414
Energy/fatigue	-0.027	0.820
Social functioning	-0.145	0.225
Role emotional	-0.171	0.152
Mental health	-0.076	0.526
PSC-PRO domain		
Module 1: PSC symptoms	0.072	0.550
Module 2: Impact of symptoms		
Physical function	0.260	0.029
Activities of daily living	0.289	0.014
Work productivity	0.176	0.212
Role function	0.127	0.296
Emotional impact	0.188	0.119
Social/leisure impact	0.096	0.426
Quality of life	0.123	0.306

Table 40. Correlation between SF-36 and PSC-PRO domains and serum ALP.

	Correlation coefficient	p-value
SF-36 domains		
Physical component summary	-0.30	0.008
Bodily pain	-0.35	0.002
General health	-0.26	0.026
Mental component summary	-0.29	0.012
Energy/fatigue	-0.29	0.012
PSC-PRO domains		
Physical function	0.30	0.009
Emotional impact	0.38	0.001

4.7 Discussion

The utility and performance of a range of quantitative metrics of the biliary tree derived from a 3D image post-processing tool in a well-characterised adult population with large-duct PSC is presented here. Work was carried out to evaluate these metrics against a reference normal value, its reliability against qualitative MRI-MRCP interpretation, correlation with other non-invasive markers in PSC, and its performance to classify patients into high-risk and low-risk groups using surrogate markers of clinical endpoints as the reference. The intrahepatic tree dilatation metrics have shown promising risk-stratification ability and better inter-observer agreement compared to qualitative MRI-MRCP interpretation. In particular, the intrahepatic sum of relative severity of dilatations (SumRelSevDilat) demonstrated positive correlations with important biomarkers in PSC and was consistently higher in the high-risk group defined by published thresholds for the development of clinical endpoints. This metric was an independent predictor of high-risk PSC defined by MRS threshold and will require external validation and longitudinal follow-up for correlation with clinical outcomes.

Whilst 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[148]. *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. *MRCP+* metrics in patients with large-duct PSC were compared to matched healthy controls. An increased gallbladder volume in patients with PSC was observed, a phenomenon that has been described in PSC[149] and may assist in the development of a radiological diagnostic score for the detection of PSC[29]. We also observed higher dilatation metrics in PSC than in matched healthy volunteers, with the severity of biliary dilatations showing the highest discriminatory performance. Nevertheless, there were a small proportion of healthy volunteers who were classified as

having “candidate dilatations”. This was in keeping with a case-controlled prospective study, where the MRCP of 4/28 (14%) healthy volunteers were reported by two experienced radiologists blinded to the clinical data as ‘possible’ for PSC even after consensual assessment with irregularities of intrahepatic ducts[150].

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. One of the drawbacks of conventional MRI-MRCP is the inherent technical limitation resulting in poor depiction of anatomy and strictures in an under-distended biliary tree, making it subject to both false-positive and false-negative calls. MRCP images are often noisy, with bile ducts ranging from several millimetres to below the image resolution of the MR scanner. Data interpretation using qualitative features of disease further compounds the matter due to its susceptibility to high inter-observer variability[120, 151]. In a study among experts in PSC, the inter-observer agreement for a single time point follow-up MRI-MRCP for overall interpretation, presence of bile duct abnormalities, suspicion of dominant stricture, suspicion of cholangiocarcinoma and recommendation for ERCP was poor across all aspects, with kappa values less than 0.2[120]. Similarly, inter-observer agreement of the radiological course of morphologic bile duct changes between two serial follow-up liver MRI-MRCP examinations was also poor, with ICC values ranging between -0.45 and 0.22[151]. 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[89]. Another study found that biliary stricture severity assessed using a slightly different standardised interpretation model had poor risk stratification ability on its own compared to when combined with liver stiffness measurements using MRE[96]. In contrast, the Anali score that incorporates a biliary and two liver parenchymal components has shown promising prognostic ability in a multi-

centre study, albeit in a retrospective cohort where images were reviewed centrally by two radiologists without assessment of inter-observer agreement[134]. Liver parenchymal abnormalities, however, tend to appear in more advanced disease. In work presented here, the inter-observer variability between Anali score, modified Amsterdam stricture severity score and quantitative MRCP metrics were compared. Whilst the agreement for binary classifications such as hepatic dysmorphy and portal hypertension was good to excellent, the inter-observer agreement for 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 millimetre margins of error. The ICC for maximum duct diameter measurements from *MRCP+* decreases gradually moving proximally into the intrahepatic biliary tree from the common bile duct, suggesting a greater room for error in manual measurements of intrahepatic duct diameter profile compared to extrahepatic ducts. The agreement for ISSS and ESSS using the modified Amsterdam classification was significantly better than 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 shown for the diagnostic performance of ISSS to detect high-risk PSC group between the two readers in our study compared to Anali score. Future radiological scores should therefore consider including quantitative metrics to avoid variations in interpretations of continuous variables such as variations in bile duct width.

The severity of intrahepatic dilatations showed the best correlation with surrogate markers of clinical endpoints. The sum of relative severity of dilatations (SumRelSevDilat) was consistently higher in patients classified as high-risk by all existing risk-stratification markers. 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[91]. 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[91].

When compared with other non-invasive markers, intrahepatic SumRelSevDilat had the best performance to diagnose high-risk PSC stratified by MRS. Intrahepatic SumRelSevDilat and ELF were the only two non-invasive 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 and Anali score performed better than intrahepatic SumRelSevDilat to identify the AOM > 2 group. LS was the only independent predictor of high-risk PSC as defined by AOM in the multivariate analysis. This is not surprising as AOM includes parameters reflecting liver fibrosis severity (platelets and albumin) rather that biliary pathology. Platelet count in fact had the strongest C-statistic amongst the seven variables in the AOM[65]. Of the radiological scores, the Anali score is weighted more towards liver fibrosis abnormalities by including the presence of dysmorphy and features of portal hypertension. Intrahepatic SumRelSevDilat performed better to classify ISSS >2 than Anali score > 2, further supporting the fact that quantitative MRCP risk-stratification by biliary morphological abnormalities was better than liver parenchymal abnormalities.

Biliary strictures are thought to be the primary pathology that drives fibrosis of the liver parenchyma in PSC. Therefore, severity of strictures may expect to correlate with severity of liver fibrosis, but we found no positive correlation or significant risk-stratification ability of any of the stricture metrics in our study, except when stratified by disease distribution. This was similarly observed in a study that found no association between the severity of the biliary component of a score from standardised qualitative interpretation of MRI-MRCP and advanced stages of liver fibrosis using the Nakanuma histological classification system[97]. 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[96]. 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 was not as good for stricture metrics. An alternative explanation for the poor correlation between stricture metrics and fibrosis metrics in our and other studies is that there are mechanisms independent of biliary stricturing that drive fibrosis in PSC which is beyond the scope of this thesis. Stricture metrics may play a more significant role in quantitatively monitoring the extrahepatic biliary tree.

Lastly, it is important to consider the relationship between patient-reported outcomes and candidate biomarkers, especially as they are increasingly being included as secondary endpoints in interventional studies. To date, there is no published literature showing correlation between MRI biomarkers and patient-reported outcomes in PSC. Findings from this study suggest that there appears to be a weak but significant correlation between physical functioning, bodily pain, and activities of daily living domain scores with *MRCP+* dilatation metrics, maximum intrahepatic bile duct dilatation from the qualitative Anali score, the presence of dominant stricture on conventional MRCP and serum ALP levels. These findings could be tied in together from what would be expected clinically as patients are more likely to be symptomatic when a dominant stricture is present with consequent proximal bile duct dilatations, a surrogate marker of the severity of the downstream stricture. Serum ALP also had a significant negative correlation with some of the emotional domains of SF-36 and PSC-PRO. Adjusting for potential confounders such as symptoms from co-existing IBD were beyond the scope of this study.

One of the limitations directly related to this piece of work was that the threshold for defining a candidate stricture and dilatation in the *MRCP+* algorithm. Whilst expert consensus was sought to reach this decision by the *MRCP+* developers, it may be argued that the definition used was rather arbitrary. However, arbitrary definition of a dominant stricture on MRCP has long persisted and is still in practice. Nevertheless, as more validation studies emerge, the threshold for this definition may be adjusted if necessary

and may even be disease-specific in the future. 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 were upstream to the strictures. Third, even though the inter and intra-observer variability in PSC was assessed, a scan-rescan reproducibility study specifically in PSC was not conducted. The data supporting reproducibility come from 6 patients with PSC[140] and further validation would be needed.

In conclusion, *MRCP+* dilatation metrics, particularly related to the severity of intrahepatic biliary dilatations, are higher in patients with high-risk of disease progression in PSC as defined by several published thresholds for surrogate markers of clinical endpoints.

In the next chapter, I will be discussing the assessment of liver parenchymal abnormalities seen in PSC using Liver*MultiScan*.

5 Quantitative assessment of liver parenchyma in PSC

5.1 Abstract

Background: Median iron-corrected T1 (cT1) derived from Liver*MultiScan* correlates well with histologically assessed liver fibrosis and inflammation in homogeneous liver diseases. However, the liver parenchyma in PSC is thought to be heterogeneous due to patchy inflammation and fibrosis. The aim of this study was to use Liver*MultiScan* analysis to perform global assessment of the liver in PSC, identify novel metrics of liver parenchymal heterogeneity, and assess the performance of candidate metrics against markers of disease severity.

Methods: Whole liver segmentation methods and regional variance analysis using superpixel methods from Liver*MultiScan* were performed to derive liver cT1 metrics. Patients with PSC were classified into high-risk and low-risk groups using published thresholds of transient elastography LS, ELF, MRS, and AOM. The diagnostic performance of candidate metrics to detect the high-risk group were assessed using area under the receiver operating characteristic curve (AUC) analyses and binomial logistic regression models. The ability to distinguish PSC from other liver diseases were tested in a separate validation cohort of liver diseases and healthy controls using one-way ANOVA or Kruskal-Wallis tests. Correlations of liver cT1 metrics with quality of life assessed by SF-36 and PSC-PRO were performed.

Results: Data from 80 patients with PSC and 20 healthy controls were included in the derivation cohort. The validation cohort consisted of 192 mixed aetiology liver disease and 44 healthy controls. IQR cT1 and mean superpixel variance cT1 were identified as candidate metrics of heterogeneity. Both demonstrated strong correlations with biochemical data, fibrosis markers, prognostic risk models and qualitative MRI and MRCP features in PSC. IQR cT1 had the best performance to detect the high-risk PSC group stratified by LS (AUC = 0.842, 95% CI 0.750–0.934, sensitivity 86%, specificity 71%). Mean superpixel variance cT1 was higher in PSC compared to other

liver diseases (PSC vs. NAFLD: $p = 0.0006$, PSC vs. viral hepatitis: $p < 0.0001$, PSC vs. ArLD: $p = 0.0014$, PSC vs. autoimmune: $p = 0.0002$) and did not differentiate between the other liver diseases. Median cT1 correlated with all the domains of SF-36 with the strongest being mental health (r_s=-0.37, p=0.001) and social functioning: r_s=-0.32, p=0.005).

Conclusion: Two candidate liver cT1 metrics showed promising risk stratification ability in PSC. IQR cT1 may be a better measure of liver fibrosis in PSC than median cT1. Mean superpixel variance cT1 could be a better marker of liver parenchymal heterogeneity than median cT1 in PSC.

5.2 Introduction

The progression of the liver disease component of PSC is characterised by the accumulation of fibrosis in the liver parenchyma which can lead to liver cirrhosis, portal hypertension, and end-stage liver disease. Whilst histological staging of fibrosis has historically been the gold standard test in chronic liver diseases, a liver biopsy is rarely performed in PSC anymore. The patchy nature of the parenchymal disease has been thought to result in variability in sampling and an increased false negative results[74, 152]. A liver biopsy is usually reserved to diagnose small-duct PSC in the context of a normal MRCP scan, another cholestatic disorder, or to assess for overlap with another autoimmune liver disease. In clinical practice, liver parenchymal changes and tissue characterisation in PSC are assessed qualitatively using MRI in conjunction with biliary tree changes on MRCP. Staging of liver fibrosis in the UK is performed separately using ultrasound-based vibration-controlled transient elastography technique (FibroScan®). However, liver stiffness measurements can fluctuate in PSC depending on degree of cholestasis, inflammation, and can underestimate severity of fibrosis in the left lobe of the liver[79, 153].

PSC may affect the whole liver or evolve on a segmental or subsegmental level and correspond to focal or zonal areas of fibrosis and/or inflammation[108]. The distribution and extent of these focal or zonal areas makes the liver in PSC patchy and heterogeneous[145]. Liver parenchymal changes commonly reported on MRI include changes associated with liver cirrhosis such as atrophy of the right lobe with hypertrophy of the left and caudate lobes, nodular liver parenchyma and contour; and features indicative of portal hypertension, such as collateral blood vessels, splenomegaly, and ascites. Parenchymal changes specific to PSC can be observed on non-enhanced and contrast-enhanced MRI sequences including T1-weighted, T2-weighted, diffusion-weighted imaging (DWI) and dynamic contrast-enhanced (DCE).

Non-enhanced T1-weighted imaging may show hyperintensity of the liver parenchyma[87]. In contrast-enhanced scans, hypoenhancement corresponding to peripheral wedge-shaped areas of atrophy is observed in early phases and hyperenhancement in more delayed phases of the scan. In the arterial phase, increased heterogeneous peripheral enhancement of the liver parenchyma is seen. These are thought to correspond to areas of altered blood supply in response to parenchymal inflammation[28, 154]. In T2-weighted imaging, high intensity is demonstrated in peripheral wedge-shaped areas of parenchymal atrophy often described as confluent fibrosis[28, 155]. These areas are not due to development of liver cirrhosis but a consequence of parenchyma oedema and increased perfusion from inflammation and/or impaired lymphatic and venous drainage of the portal triads, often described as periportal oedema[155, 156]. Progression to focal fibrosis surrounding bile ducts causes ductal obliteration and ductopaenia with consequent cholestasis leading to areas of focal atrophy that appear heterogeneous to surrounding areas[157].

Several quantitative MR techniques have been proposed to assess the parenchymal changes described in PSC and quantify the accumulation of liver fibrosis. Apparent diffusion coefficient (ADC) values derived from DWI have been shown to detect early and advanced liver parenchymal fibrosis in a small study of 38 patients with PSC[28]. In another study involving 47 patients, ADC was shown to correlate with transient elastography liver stiffness measurements, discriminated between stages of liver fibrosis with reasonable accuracy, and performed better than dynamic contrast-enhanced MRI[103]. However, limitations due to variability in imaging protocol and reproducibility of ADC limits its use[158, 159].

Changes in T2 relaxation time (T2) signal intensities and relative liver enhancement (RLE) were quantified from contrast enhanced T2 weighted images. These changes have been shown to correspond to areas of hyperintensity seen in PSC[111] and can discriminate between stages of fibrosis[110]. RLE demonstrated a positive correlation with AOM, MRS

and clinical endpoints in PSC (cholangiocarcinoma, liver transplantation, and liver-related mortality)[112].

Increase in mean T1 relaxation time (T1) has been shown to correlate with liver fibrosis in chronic liver diseases[123, 124, 160] and predicted clinical outcome in one study[127]. A prospective study demonstrated that changes in liver T1 pre- and post-contrast correlated with disease severity in PSC [108]. A more recent retrospective study replicated this finding[114]. However, both these studies involved contrast administration and manual ROI-placement and interpretation by radiologists.

MR elastography (MRE) exploits the mechanical properties of the liver tissue by generating elasticity maps that can quantify parenchymal fibrosis. MRE has been shown to accurately diagnose and stage liver fibrosis in liver diseases[161], and specifically in PSC[98]. Liver fibrosis stage has prognostic significance as it predicts transplant-free survival and time to transplantation in PSC[72]. MRE has been shown to predict liver decompensation[98] and transplant-free survival[99] in patients with PSC. Although becoming more widely available in the USA, MRE is currently not widely available in the UK.

Quantitative imaging-based biomarkers that attempt to quantify the parenchymal changes seen in PSC requires administration of a gadolinium-based contrast agent. The repeated administration of a contrast agent must be weighed against concerns of accumulation of gadolinium-based contrast agents in the brain, particularly as it appears to be directly correlated to the number of administrations[162, 163]. For the quantification of liver fibrosis, MRE is a promising method that visualises the whole liver, therefore offering a more accurate quantification of liver fibrosis than transient elastography technique. However, transient elastography is cheaper and readily available by the bedside, but risks underestimating the severity of liver fibrosis in a patchy heterogeneous parenchyma of PSC when only the right lobe of the liver is examined. MRE is also operator dependent and there is often large part of the liver for which a value of liver elasticity is not attainable.

There is a need to validate an imaging method that can quantify the regional variations throughout the liver parenchyma of PSC and concurrently stage the patchy liver fibrosis without the need of repeated contrast administration. Measurement of cT1 in each voxel of the segmented liver parenchyma using Liver*MultiScan* offers the opportunity to interrogate the distribution of cT1 measurements as a surrogate of disease activity within the liver.

5.3 Hypotheses

The patchy parenchymal disease in PSC is based on visual judgements of expert radiologists and pathologists. Here, quantitative assessment of the whole liver data provides an opportunity to test this hypothesis of disease heterogeneity in the PSC liver. Using metrics that measure variability in the distribution of cT1, the hypotheses for this chapter were that these metrics will be:

- higher in PSC than normal liver
- higher in more severe PSC compared to mild disease
- correlated with surrogate markers of liver fibrosis in PSC
- unique to PSC
- associated with worse QoL

5.4 Aims

This chapter aims to use liver cT1 metrics to:

- explore the use of different measures of heterogeneity to assess regional variations in the liver parenchyma of PSC
- investigate their performance to discriminate patients with PSC from healthy controls

- assess their relationship with non-invasive markers of disease severity and progression in PSC
- assess their relationship with qualitative morphological MRI features of PSC
- determine their performance to classify patients into high-risk and low-risk groups based on surrogate risk-stratification indices
- assess whether they are independent predictors of high-risk PSC using prognostic risk models as the reference standard
- examine association with patient-reported outcomes
- validate their performance against other liver disease controls

5.5 Participants and methods

5.5.1 Participants

There are two cohorts of participants reported in this chapter. They are labelled as derivation and validation cohorts.

Derivation cohort

Patients with large-duct PSC and healthy controls were recruited to the MoBILISE study following the methods described in chapter 2.

Validation cohort

A database of chronic liver diseases and healthy volunteers from three previous Liver*MultiScan* studies conducted at OCMR were used to validate the candidate metrics identified in the MoBILISE study cohort. T1 and T2* maps acquired on patients with other liver diseases besides PSC were collected from three studies: the rapid imaging in liver disease (RIAL) study, non-invasive comprehensive liver assessment (NICOLA) study, and MRI assessment of portal hypertension and liver function (MAPLE) study. These studies were granted NHS regional ethics committee (RIAL: 11/H0504/2, NICOLA 13/SC/0243

and MAPLE 16/SC/0110) and local trust approvals. They were conducted in accordance with the Declaration of Helsinki 2013 and Good Clinical Practice Guidelines. All participants gave written informed consent for use of their data in future research.

The RIAL and NICOLA studies aimed to evaluate the utility of multiparametric MRI methods to evaluate chronic liver diseases against histology as the reference standard. Adult patients who were scheduled to have a clinically indicated liver biopsy for the assessment of fibrosis and liver disease aetiology, were invited from two UK study centres (Oxford and Reading) to have a multiparametric MRI scan between March 2011 and May 2015. Healthy volunteers were also recruited to the NICOLA study. Healthy volunteers were defined as adults between 16–80 years old with BMI less than 25kg/m^2, no history of liver disease or any components of the metabolic syndrome (type 2 diabetes, hypertension, hyperlipidaemia), alcohol consumption within the recommended limit by the UK Department of Health, and normal liver function tests.

The MAPLE study aimed to investigate the utility of multiparametric MRI methods for the assessment of portal hypertension and liver function. It predominantly invited patients with compensated and decompensated cirrhosis of any cause to have a multiparametric MRI in Oxford between June 2016 and July 2018. These patients were diagnosed and managed as having cirrhosis as part of their routine clinical follow-up. This study also included a group with acute liver injury and healthy volunteers. Healthy volunteers were defined as described in the NICOLA study.

Other liver diseases

Chronic liver diseases included for study analysis from these three studies were NAFLD, ArLD, HCV, HBV, PBC, and AIH. Scans from any other acute or chronic liver diseases or from patients with decompensated chronic liver disease at recruitment were excluded. Scans labelled as PSC were excluded as the disease subtype was unavailable. For this

study, patients with HBV and HCV were combined as one group and named viral hepatitis to avoid small group size. Patients with PBC and AIH were combined as one group.

Healthy

Healthy volunteer data from the three previous studies were pooled with healthy controls from MoBILISE study.

5.5.2 Clinical, laboratory and liver stiffness data

Clinical, biochemical, and liver stiffness measurement data for the MoBILISE study was obtained according to methods described in chapter 2. For the validation cohort, transient elastography and fasting blood samples were collected for all three studies, but ELF was only performed in the MAPLE study.

5.5.3 Risk stratification

Patients with PSC were classified into high-risk and low-risk groups according to published thresholds of risk scores reported in chapter 2. Additionally, patients were classified according to the presence of morphological features on liver MRI including hepatic dysmorphy, portal hypertension, and cirrhosis.

5.5.4 MRI acquisition

MRI acquisition for the derivation cohort followed the methods described in chapter 2 for Liver*MultiScan* analysis. For the validation cohort, participants in all three studies underwent MRI in a different scanner of same field strength (Siemens MAGNETOM TrioTim, Magnetic Field Strength 3 T) having fasted for at least four hours, using a similar standardised Liver*MultiScan* acquisition protocol but did not have pineapple juice prior to acquisition.

5.5.5 Quantitative image analysis

The variability in cT1 distribution within the liver was quantified using two methods (i) standard histogram analysis of all pixels that represents dispersion of cT1 across the whole liver, and (ii) superpixel analysis of regional pixels grouped together by local similarities in cT1.

Method 1: Standard histogram analysis

Histogram analysis using measures of dispersion in data were used to assess variability of cT1 distribution in the segmented liver parenchyma. For a parametric distribution, heterogeneity can be measured using the standard deviation (σ) or variance (σ^2) of the data from the mean. For a non-parametric distribution, the IQR of the data can be used as a measure of histogram width. IQR is given by the lower and upper quartiles, is not affected by outliers and 50% of observations lie within the interquartile range whatever the distribution data is. Two other histogram measures that could potentially be exploited here are skewness and kurtosis of the distribution. Skewness describes the degree of departure from symmetry of a distribution. A negatively skewed distribution will have a negatively pulled “tail” with skewness < 0 whereas a positively skewed distribution will have a “tail” pulled in the positive direction with skewness > 0. Kurtosis describes the degree of “peakedness” of a distribution and its deviation from a normal distribution, where kurtosis (k) = 3 is a normal distribution. A higher peak and heavier tail leads to a greater kurtosis. Figure 31 summarises these in graphical form. Therefore, it was hypothesised that the more heterogeneous the liver, the higher the IQR cT1, greater the skewness and kurtosis of the distribution of cT1.

For reporting of the MoBILISE study data, image analysis was performed as described in chapter 2. The whole liver segmentation method — as opposed to ROI-based method — was used to assess the liver parenchyma as it presents with the opportunity to interrogate the whole parenchyma in a heterogeneous disease such as PSC rather than reliance on the bias of data analysts in ROI placements. The following cT1 outcomes for the whole liver were derived from this analysis:

- Median cT1 (ms)
- IQR cT1 (ms)
- Skewness cT1
- Kurtosis cT1

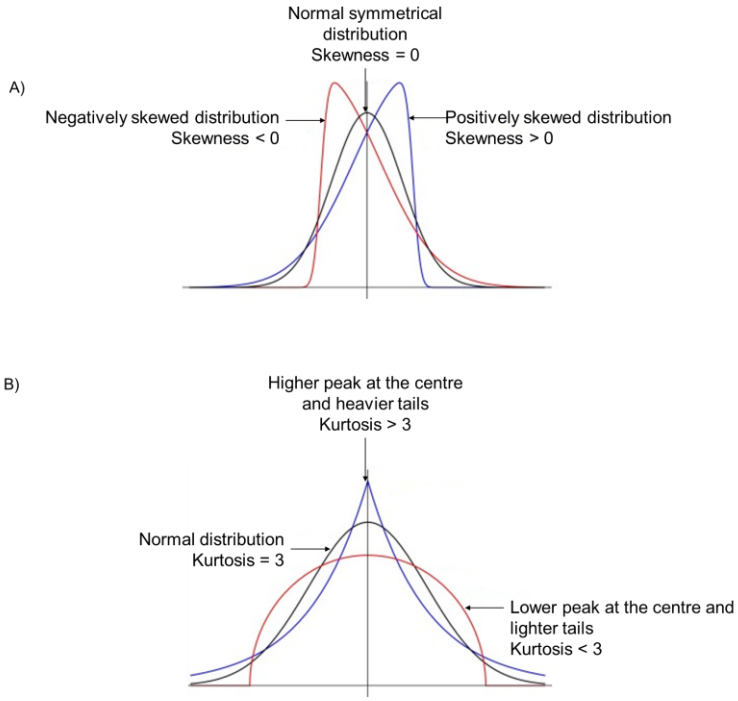


Figure 31. Schematic diagram explanation for (A) skewness and (B) kurtosis of a distribution.

Method 2: Superpixel analysis

The superpixel analysis described in section 2.4.2 chapter 2 is a form of texture analysis that is used to assess local variations in cT1 and their spatial distribution. A measure of variance in each superpixel derived from this method could provide a measure of texture within the liver parenchyma that is not affected by blood vessels and liver border transitions.

For comparison with other liver diseases, MoBILISE study scans were re-analysed as a batch together with the large dataset of scans obtained from the other three studies using the automated single-slice Liver*MultiScan* superpixel analysis technique to assess regional variance as described in chapter 2. Automated analysis of scans was rejected if the delineation of the liver by the algorithm was judged visually to be grossly misplaced outside the liver borders. Minimal displacement of the liver mask over the stomach

boundaries was edited manually and scan fed back through the automated algorithm. The following liver cT1 outcome measures were derived from this analysis:

- Mean superpixel variance cT1 (ms)
- Skewness superpixel cT1 means
- Kurtosis superpixel cT1 means

Mean superpixel variance cT1 measures the average variation of cT1 from its mean within each superpixel across the automatically segmented liver slice.

5.5.6 Qualitative image analysis

MRI scans from the MoBILISE study were interpreted independently by two specialist radiologists using a standardised interpretation model as described in chapter 2.

5.5.7 Statistical analysis

Comparison of liver cT1 metrics between high-risk and low-risk PSC groups were performed using Student's *t*-test for parametric distribution and Mann-Whitney test for non-parametric distribution. Inter-observer agreement of morphological changes on MRI between the two readers was evaluated using Cohen's kappa (κ) coefficient. κ values were classified as follows[10]: > 0.80 excellent, 0.79–0.60 substantial, 0.59–0.40 moderate, 0.39–0.20 fair, and < 0.2 poor. Receiver operating characteristic curve (ROC) analyses were used to determine the performance of liver metrics to differentiate PSC versus healthy cohorts, and high-risk versus low-risk PSC groups. Comparisons between the different liver disease groups were performed using one-way ANOVA with Tukey's multiple comparisons test for parametric distribution and Kruskal-Wallis with Dunn's multiple comparisons test for non-parametric distribution of liver cT1 metrics. Cut-off values were chosen using the Youden index and sensitivity and specificity reported. Multivariate logistic regression analyses were performed to identify the best predictive covariate of high-risk PSC stratified by AOM and MRS. Covariates with $p < 0.1$ in the univariate analysis were entered into the multivariate analysis. Variables that contributed to MRS and AOM scores were excluded. Potential multicollinearity between liver cT1

metrics were assessed using thresholds of Spearman's rank correlation coefficient > 0.80 and variance inflation factor < 1 or > 10.

5.6 Results

5.6.1 Derivation cohort: MoBILISE study

5.6.1.1 Cohort characteristics

Recruitment details are depicted in Figure 32. Eighty-one patients with large-duct PSC consented to the MoBILISE study. One patient was excluded due to claustrophobia. MRI acquisition and Liver*MultiScan* analysis were successful in the remaining 80 patients with PSC. There were no technical failures for transient elastography. The median age of patients with PSC was 42 years (IQR: 31–55) and 68% were male. The median disease duration was 8 years (IQR: 4–12) with 53% having both intrahepatic and extrahepatic disease at study recruitment and 56% on ursodeoxycholic acid therapy. Concomitant inflammatory bowel disease was present in 76% with ulcerative colitis being the predominant subtype. The clinical and biochemical parameters for patients with PSC are summarised in Table 41.

Twenty-one healthy volunteers consented to the study with successful acquisition of MRI and transient elastography liver stiffness measurements. Healthy volunteers had median age 35 years (IQR: 31–39) and 62% were male. All their blood parameters and liver stiffness measurements were within the normal range.

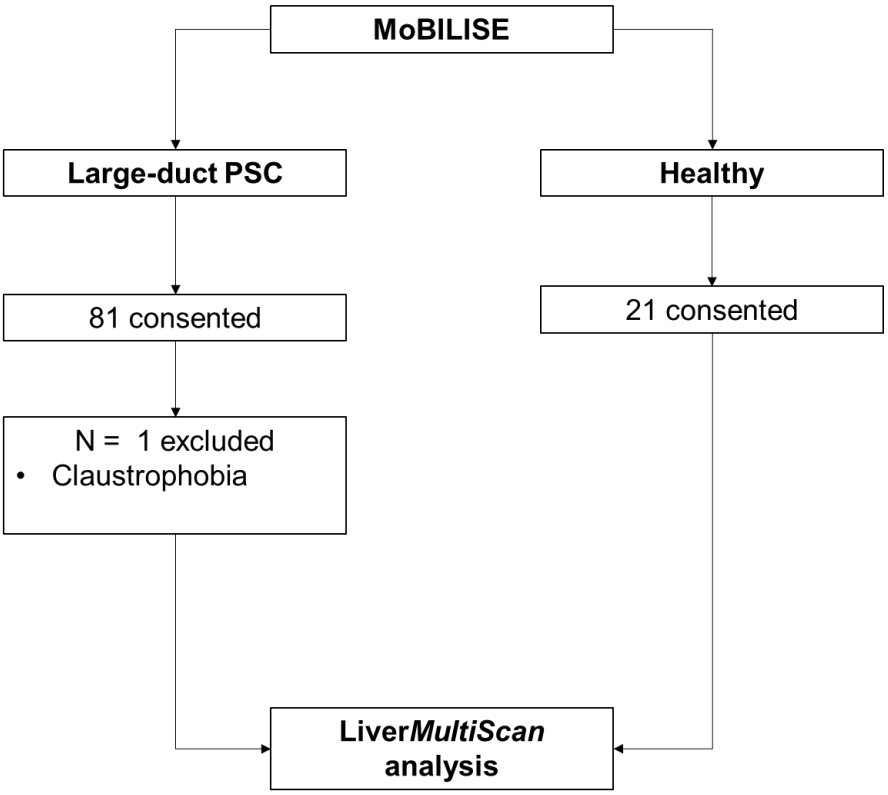


Figure 32. Study flow diagram.

Table 41. Baseline characteristics of participants included in the MoBILISE study.

	PSC (n = 80)		Healthy (n = 21)	
Demographics				
Male	54	(68)	13	(62)
Age (years)	42	(31–55)	35	(31–39)
BMI (kg/m^2)	24.5	(21.7-27.5)	22.7	(21.2–24.8)
Laboratory parameters				
Total bilirubin, µmol/L	13	(10–21)	10	(8–13)
ALT (IU/l)	42	(27–84)	20	(13–26)
AST (IU/l)	38	(27–64)	22	(19–24)
ALP (IU/l)	150	(101–234)	58	(51–70)
xULN ALP	1.2	(0.8–1.8)	0.4	(0.4–0.5)
GGT (IU/l)	122	(48–297)	17	(14–20)
Albumin (g/l)	40	(37–41)	42	(41–43)
Prothrombin time (s)	10.6	(10.3–11.0)	10.5	(10.5-11.0)
Platelet count (x 10^9/l)	253	(192–305)	256	(234–270)
Clinical data				
PSC duration (years)	8	(4–12)	-	
PSC localisation at recruitment				
Intrahepatic + extrahepatic	42	(53)	-	
Intrahepatic only	38	(47)	-	
On UDCA therapy	45	(56)	-	
Clinical cirrhosis	13	(16)	-	
IBD present	61	(76)	-	
IBD phenotype				
Ulcerative colitis	46	(58)	-	
Crohn’s	8	(10)	-	
Unspecified	7	(9)	-	
Liver fibrosis markers				
Transient elastography liver stiffness (kPa)	6.9	(5.4–10.8)	4.7	(4.1–5.3)
ELF score	9.4	(8.7–10.0)	-	
Prognostic risk models				
Mayo risk score	0.1	(-0.2–0.5)	-	
Amsterdam-Oxford model score	1.6	(1.3–2.1)	-	

Continuous variables are expressed as median (interquartile range) and nominal variables as absolute number (percentage).
ALP, alkaline phosphatase; ALT, alanine aminotransferase; AST, aspartate aminotransferase; APRI, AST to platelet ratio index; BMI, body mass index; ELF, enhanced liver fibrosis score; FIB-4, fibrosis-4 index; GGT, gamma-glutamyltransferase; IBD, inflammatory bowel disease; PSC, primary sclerosing cholangitis; UDCA, ursodeoxycholic acid; xULN, times upper limit of normal.

5.6.1.2 Comparison of quantitative metrics between PSC and healthy controls

In this exploratory analysis, all liver cT1 metrics derived from both standard and superpixel analyses were compared between patients with PSC and healthy controls. There were no correlations between the derived cT1 metrics. Metrics that measured the skewness and kurtosis of the distribution of the two cohorts as well as the number of superpixels parcellated within the liver were not able to differentiate PSC from normal, therefore were discarded. The discriminatory performance of the remaining three liver cT1 metrics were considered. Mean superpixel variance cT1 (AUC = 0.90, sensitivity 90%, specificity 80%) showed the highest discriminatory performance to distinguish between PSC and healthy compared to IQR cT1 (AUC = 0.79, sensitivity 91%, specificity 58%) and median cT1 (AUC = 0.74, sensitivity 76%, specificity 68%;Table 42).

Table 42. Diagnostic performance of Liver*MultiScan* liver cT1 metrics to distinguish between patients with PSC and healthy controls.

Liver cT1 metrics	PSC (n=80)		Healthy (n=21)		Cut-off	AUC (95% CI)	p-value	Sensitivity (%)	Specificity (%)
Standard analysis									
Median cT1 (ms)	760.0	(721.8–790.8)	713.0	(679.0–744.0)	736.5	0.74 (0.62–0.85)	0.0009	76	68
IQR cT1 (ms)	78.0	(72.0–94.0)	69.0	(63.0–74.0)	76.5	0.79 (0.70–0.88)	<0.0001	91	58
Skewness cT1	0.4	(0.0–0.8)	0.6	(0.1–0.7)	0.4	0.53 (0.41–0.65)	0.6818	67	53
Kurtosis cT1	5.2	(3.7–7.9)	4.7	(3.8–7.0)	5.0	0.52 (0.40–0.65)	0.7568	62	53
Superpixel analysis									
Total superpixel count	51.5	(45.0–58.8)	49.0	(46.5–54.0)	57.5	0.54 (0.42–0.65)	0.6156	95	26
Mean superpixel variance cT1 (ms)	59.3	(56.7–61.2)	54.3	(53.5–55.7)	56.5	0.90 (0.83–0.96)	<0.0001	90	80
Skewness superpixel cT1 means	0.8	(0.6–0.9)	0.8	(0.7–0.9)	0.8	0.57 (0.45–0.70)	0.3113	71	50
Kurtosis superpixel cT1 means	2.9	(2.5–3.3)	2.6	(2.3–3.0)	2.5	0.63 (0.51–0.76)	0.0644	48	79
Other non-invasive variables									
ALP (IU/l)	150.0	(101.3–234.8)	58.0	(50.5–70.0)	80.5	0.94 (0.88–0.98)	<0.0001	90	89
LS (kPa)	6.8	(5.3–10.0)	4.7	(4.1–5.3)	6.1	0.85 (0.77–0.92)	<0.0001	100	61
ELF	9.4	(8.7–9.9)	8.6	(8.0–8.9)	9.1	0.79 (0.70–0.89)	<0.0001	81	66

All values are presented as median (interquartile range). Abbreviations: ALP, alkaline phosphatase; LS, transient elastography liver stiffness; ELF, enhanced liver fibrosis score.

5.6.1.3 Qualitative MRI data in PSC

The morphological features assessed by qualitative interpretation of MRI and their respective inter-observer agreements are summarised in Table 43. Features related to MRCP have already been summarised in Table 19 of chapter 4. Hepatic dysmorphy (κ = 0.82) had excellent inter-observer agreement. Despite the excellent and substantial agreements for hepatic dysmorphy and portal hypertension, respectively, the inter-observer agreement for Anali score was poor (κ = 0.38), owing to the biliary component of the score as discussed in chapter 4.

Table 43. Qualitative interpretation of MRI for morphological features and their corresponding inter-observer agreements.

	Reader 1	Reader 2	Cases in agreement	Cohen's kappa (κ)
Hepatic dysmorphy			73/80 (91%)	0.819
Absent	47	48	44	
Present	33	32	29	
Portal hypertension			70/80 (88%)	0.697
Absent	54	60	52	
Present	26	20	18	
Anali score			40/80 (50%)	0.378
0	24	28	18	
1	11	11	5	
2	15	12	4	
3	9	11	4	
4	10	13	5	
5	11	5	4	
Liver cirrhosis			72/80 (90%)	0.728
Absent	57	65	57	
Present	23	15	15	
Splenomegaly			73/80 (91%)	0.790
Absent	54	59	53	
Present	26	21	20	

5.6.1.4 Association between liver cT1 metrics and non-invasive markers

The correlation between the three metrics and other variables from the study were assessed and are presented in Table 44. IQR cT1 and mean superpixel variance cT1 demonstrated stronger correlations with biochemical data, fibrosis markers, prognostic risk models, and qualitative MRI and MRCP features than median cT1. Both metrics demonstrated the strongest correlation with LS (IQR cT1 $r_s = 0.66$, $p < 0.0001$; mean superpixel variance cT1 $r_s = 0.53$, $p < 0.0001$; Figure 33).

There was positive correlation between IQR cT1 and ELF ($r_s = 0.55$, $p < 0.0001$), AOM ($r_s = 0.51$, $p < 0.0001$), MRS ($r_s = 0.45$, $p < 0.0001$), xULN ALP ($r_s = 0.26$, $p = 0.018$), modified Amsterdam ISSS (reader 1: $r_s = 0.39$; $p = 0.0002$; reader 2: $r_s = 0.43$, $p < 0.0001$) and Anali score (reader 1: $r_s = 0.53$, $p < 0.0001$; reader 2: $r_s = 0.57$, $p < 0.0001$). IQR cT1 was negatively correlated with albumin ($r_s = -0.30$, $p = 0.0007$) and platelet ($r_s = -0.27$, $p = 0.0015$). IQR cT1 showed a mixed picture with liver tests, positively correlated with bilirubin ($r_s = 0.37$, $p = 0.001$), AST ($r_s = 0.34$, $p = 0.002$), GGT ($r_s = 0.30$, $p = 0.006$), but no significant correlation with ALT ($r_s = 0.22$, $p = 0.054$).

Mean superpixel variance cT1 was positively correlated with bilirubin ($r_s = 0.41$, $p < 0.0001$), AOM ($r_s = 0.35$, $p = 0.0002$), ELF ($r_s = 0.32$, $p = 0.0004$), MRS ($r_s = 0.28$, $p = 0.012$), modified Amsterdam ISSS (reader 1: $r_s = 0.28$; $p = 0.011$; reader 2: $r_s = 0.23$, $p = 0.04$), Anali score (reader 1 and 2: $r_s = 0.34$, $p = 0.002$). and AST ($r_s = 0.25$, $p = 0.025$). This metric was also negatively correlated with albumin ($r_s = -0.26$, $p = 0.022$) and platelet ($r_s = -0.25$, $p = 0.029$).

Table 44. Correlation between quantitative liver MRI metrics and study variables.

	Median cT1 (ms)		IQR cT1 (ms)		Mean superpixel variance cT1 (ms)	
	coefficient	p-value	coefficient	p-value	coefficient	p-value
Blood markers						
Bilirubin (µmol/l)	0.118	0.297	0.369	0.001	0.414	<0.001
ALT (IU/l)	-0.265	0.017	0.216	0.054	0.095	0.400
AST (IU/l)	-0.139	0.220	0.341	0.002	0.250	0.025
ALP (IU/l)	0.032	0.776	0.264	0.018	0.036	0.752
GGT (IU/l)	-0.123	0.280	0.304	0.006	0.101	0.378
Platelets ($x10^9$/l)	0.068	0.550	-0.272	0.015	-0.245	0.029
Albumin (g/l)	-0.368	0.001	-0.299	0.007	-0.256	0.022
Prothrombin time (s)	0.061	0.588	0.006	0.957	0.146	0.197
Liver fibrosis markers						
Liver stiffness (kPa)	0.294	0.008	0.661	<0.001	0.529	<0.001
ELF	0.140	0.215	0.447	<0.001	0.320	0.004
Prognostic risk models						
MRS	0.231	0.039	0.450	<0.001	0.280	0.012
AOM	0.122	0.281	0.512	<0.001	0.345	0.002

Table 44 continued

Modified Amsterdam classification						
ISSS reader 1	0.112	0.322	0.392	<0.001	0.281	0.012
ISSS reader 2	0.111	0.326	0.434	<0.001	0.231	0.040
ESSS reader 1	0.101	0.375	0.320	0.004	0.266	0.017
ESSS reader 2	0.004	0.971	0.294	0.008	0.231	0.039
TSSS reader 1	0.102	0.366	0.411	<0.001	0.314	0.005
TSSS reader 2	0.054	0.632	0.420	<0.001	0.270	0.015
Anali classification						
IHBD dilatation reader 1	0.087	0.321	0.231	0.009	0.156	0.075
IHBD dilatation reader 2	0.160	0.069	0.313	<0.001	0.117	0.183
Dysmorphy reader 1	0.129	0.255	0.392	<0.001	0.243	0.030
Dysmorphy reader 2	0.146	0.197	0.416	<0.001	0.306	0.006
Portal hypertension reader 1	0.163	0.148	0.457	<0.001	0.362	0.001
Portal hypertension reader 2	0.165	0.143	0.424	<0.001	0.389	<0.001
Anali score reader 1	0.237	0.034	0.527	<0.001	0.341	0.002
Anali score reader 2	0.269	0.016	0.574	<0.001	0.343	0.002
Other features						
Cirrhosis reader 1	0.202	0.073	0.422	<0.001	0.387	<0.001
Cirrhosis reader 2	0.428	0.000	0.402	<0.001	0.274	0.014
Dominant stricture reader 1	0.082	0.470	0.077	0.498	0.080	0.481
Dominant stricture reader 2	0.062	0.585	0.278	0.013	0.147	0.195

ALP, alkaline phosphatase; ALT, alanine aminotransferase; AST, aspartate aminotransferase; ELF, enhanced liver fibrosis score; ESSS, modified Amsterdam extrahepatic stricture severity score; GGT, gamma-glutamyltransferase; IHBD, intrahepatic bile duct; ISSS, modified Amsterdam intrahepatic stricture severity score, LS, transient elastography liver stiffness; TSSS, modified Amsterdam total stricture severity score; xULN, times upper limit of normal.

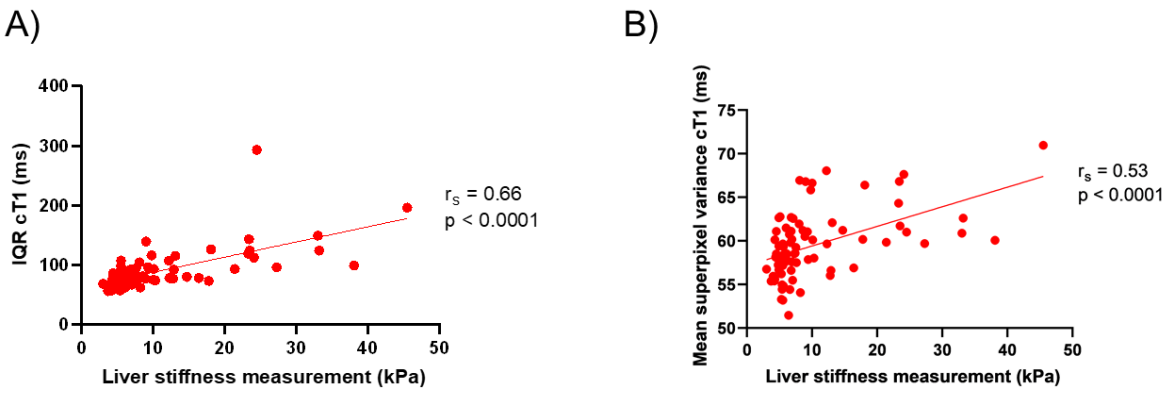


Figure 33. Correlation between (A) IQR cT1 and (B) Mean superpixel variance cT1 with transient elastography liver stiffness in 80 patients with large-duct PSC.

5.6.1.5 Risk stratification by liver fibrosis severity

Transient elastography liver stiffness (LS)

Patients with LS > 9.6kPa (n = 24) had higher median, IQR and mean superpixel variance cT1 compared to patients with LS ≤ 9.6kPa (n = 56) Table 45, Figure 34. IQR cT1 cut-off of 91ms had the best performance to detect the high-risk group stratified by LS (AUC = 0.842, 95% CI 0.750–0.934, sensitivity 86%, specificity 71%: Figure 35).

Enhanced liver fibrosis (ELF)

IQR cT1 and mean superpixel variance cT1 demonstrated significant correlation with ELF. Both were higher in the ELF > 9.8 group (n = 25) compared to ELF ≤ 9.8 group (n = 55). IQR cT1 cut-off of 88.5ms and mean superpixel variance cT1 cut-off of 58.5ms showed similar performances to detect the high-risk group (IQR cT1: AUC = 0.718, 95% CI 0.579–0.857; mean superpixel variance cT1: AUC = 0.721, 95% CI 0.597–0.846; Table 45, Figure 34). Median cT1 showed no significant correlation with ELF and poorly differentiated the two risk groups (AUC 0.641, 95% CI 0.506–0.776).

Table 45. Performances of liver cT1 metrics to discriminate between high-risk and low-risk PSC stratified by transient elastography liver stiffness (LS) and enhanced liver fibrosis score (ELF) thresholds.

	High risk		Low risk		p-value	AUC (95% CI)	Cut-off (ms)	Sensitivity (%)	Specificity (%)
	LS > 9.6kPa (n = 24)		LS ≤ 9.6kPa (n = 56)						
Median liver cT1 (ms)	788	748.8–823.0	747	713.5–783.0	0.0003	0.708 (0.583–0.833)	756.5	75	57
IQR liver cT1(ms)	103	78.5–124.0	75.5	70.0–85.6	<0.0001	0.842 (0.750–0.934)	91	71	86
Mean superpixel variance cT1 (ms)	61.1	59.7–66.3	58.2	56.0–60.4	<0.0001	0.773 (0.663–0.883)	59.5	83	68
	ELF > 9.8 (n = 25)		ELF ≤ 9.8 (n = 55)						
Median liver cT1 (ms)	775	740.5–822.0	751	715.0–789.0	0.0438	0.641 (0.506–0.776)	755.5	55	72
IQR liver cT1(ms)	96	76.0–124.0	76	71.0–86.0	0.0002	0.718 (0.579–0.857)	88.5	64	82
Mean superpixel variance cT1 (ms)	60.9	59.5–65.1	58	56.6–70.8	0.0002	0.721 (0.597–0.846)	58.5	88	60

All values are presented as median (interquartile range).

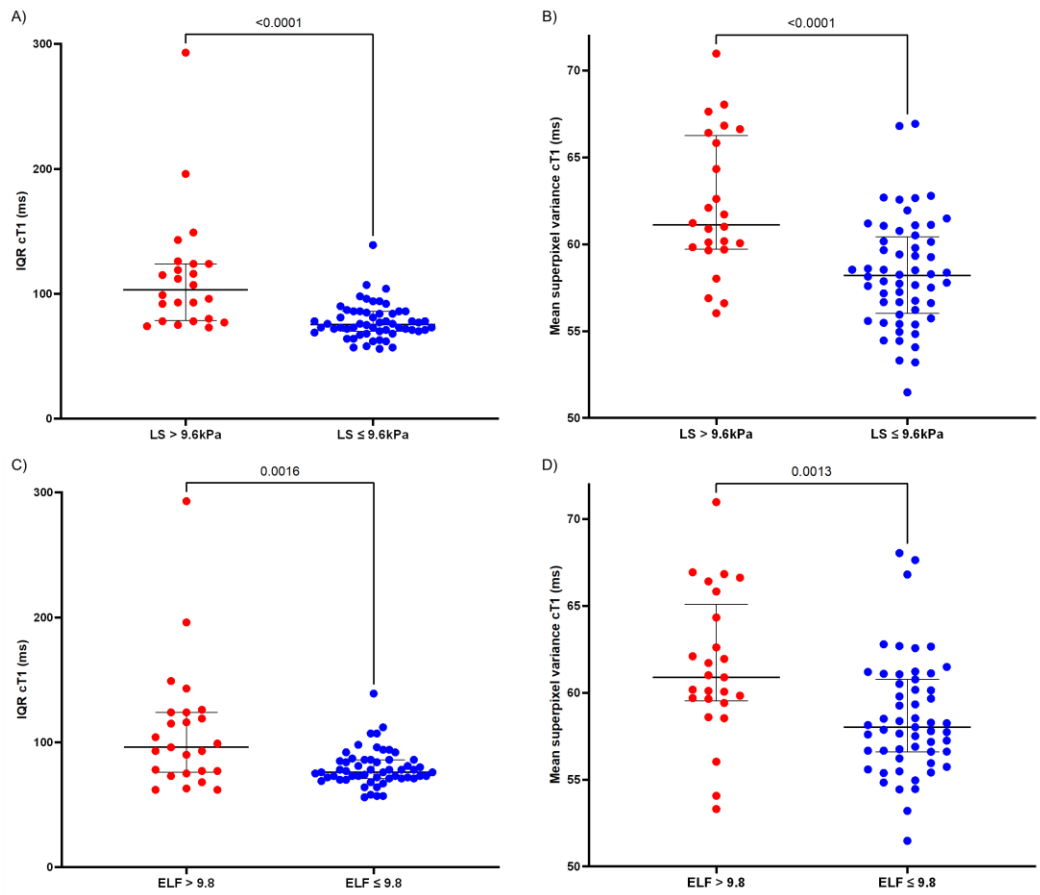


Figure 34. Comparisons of IQR cT1 and mean superpixel variance cT1 in 80 patients with PSC stratified into high-risk and low-risk groups by (A-B) transient elastography liver stiffness (LS) and (C-D) enhanced liver fibrosis score (ELF) thresholds. Lines and error bars indicate median and interquartile range.

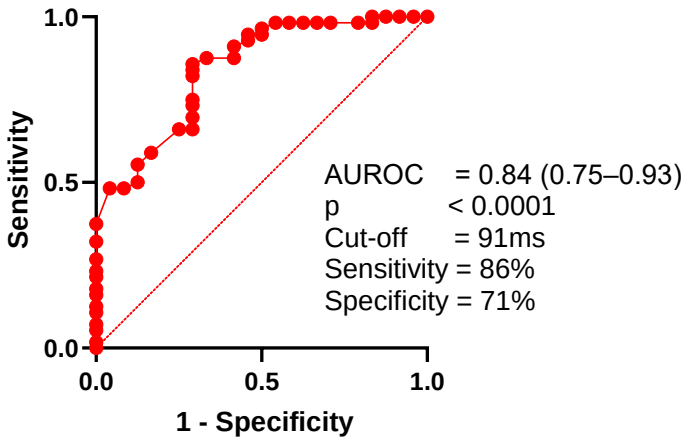


Figure 35. ROC curve of IQR cT1 for detecting advanced fibrosis (≥F3) defined by transient elastography liver stiffness measurement > 9.6kPa in 80 patients with large-duct PSC.

5.6.1.6 Risk stratification by serum alkaline phosphatase

Apart from IQR cT1 at the threshold of x2.2 upper limit normal of ALP (p = 0.0115; Figure 36), none of the other liver metrics were significantly different between the high-risk and low-risk groups stratified by x2.2 or x1.5 ULN of ALP (Table 46).

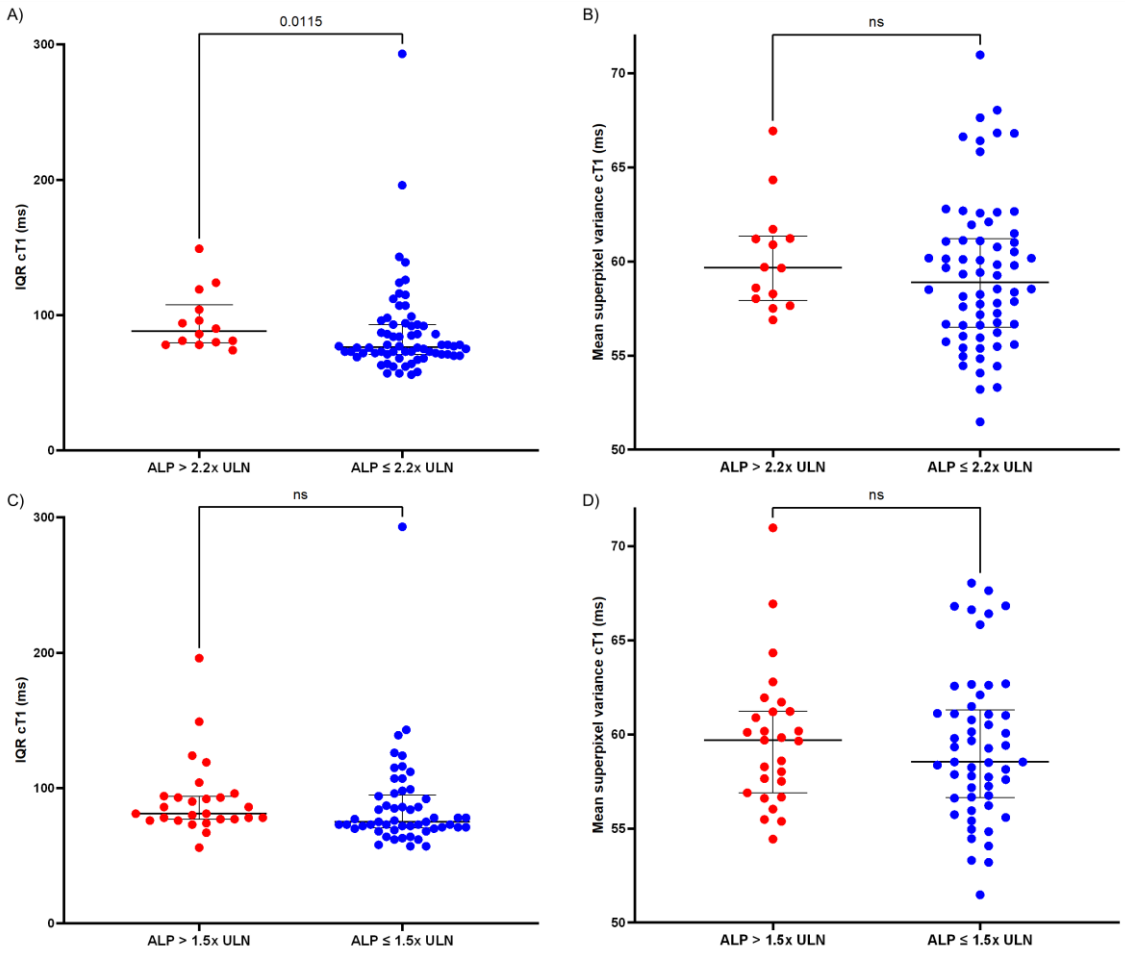


Figure 36. Comparisons of IQR cT1 and mean superpixel variance cT1 in 80 patients with PSC stratified into high-risk and low-risk groups by (A-B) 2.2x; and (C-D) 1.5x upper limit of normal serum ALP thresholds. ns, not significant. Lines and error bars indicate median and interquartile range.

Table 46. Performances of liver cT1 metrics to discriminate between high-risk and low-risk PSC stratified by upper limit of normal (ULN) serum alkaline phosphatase thresholds.

	High risk		Low risk		p-value	AUC (95% CI)	Cut-off (ms)	Sensitivity (%)	Specificity (%)
	ALP > 2.2 x ULN (n = 14)		ALP ≤ 2.2 x ULN (n = 66)						
Median liver cT1 (ms)	768	721.0–793.8	758	720.8–790.3	0.8584	-	-	-	-
IQR liver cT1(ms)	88	79.5–107.8	76.5	71.0–93.0	0.0115	0.713 (0.597–0.830)	77.5	55	93
Mean superpixel variance cT1 (ms)	59.7	57.9–61.4	58.9	56.5–61.2	0.2823	-	-	-	-
	ALP > 1.5 x ULN (n = 27)		ALP ≤ 1.5 x ULN (n = 53)						
Median liver cT1 (ms)	770	738.0–793.0	751	714.0–790.0	0.3002	-	-	-	-
IQR liver cT1(ms)	81	77.0–94.0	75	70.5–95.0	0.0542	-	-	-	-
Mean superpixel variance cT1 (ms)	59.7	56.9–61.2	58.5	56.6–61.3	0.6202	-	-	-	-

All values are presented as median (interquartile range).

5.6.1.7 Risk stratification by prognostic risk models

Amsterdam-Oxford model

Patients with AOM > 2 (n = 25) had higher median, IQR and mean superpixel variance cT1 compared to patients with AOM ≤ 2 (n = 55; Figure 37). IQR cT1 had the best performance to detect the high-risk group (AUC = 0.758, 95% CI 0.636–0.880;Table 47).

Mayo risk score

IQR cT1 was the only metric to be significantly different between the high-risk and low-risk groups with patients in the MRS > 0 group (n = 46) having higher IQR cT1 values compared to the MRS ≤ 0 group (n = 34; Figure 37). However, its performance to do so was poor (AUC = 0.685, 95% CI 0.569–0.801; Table 47).

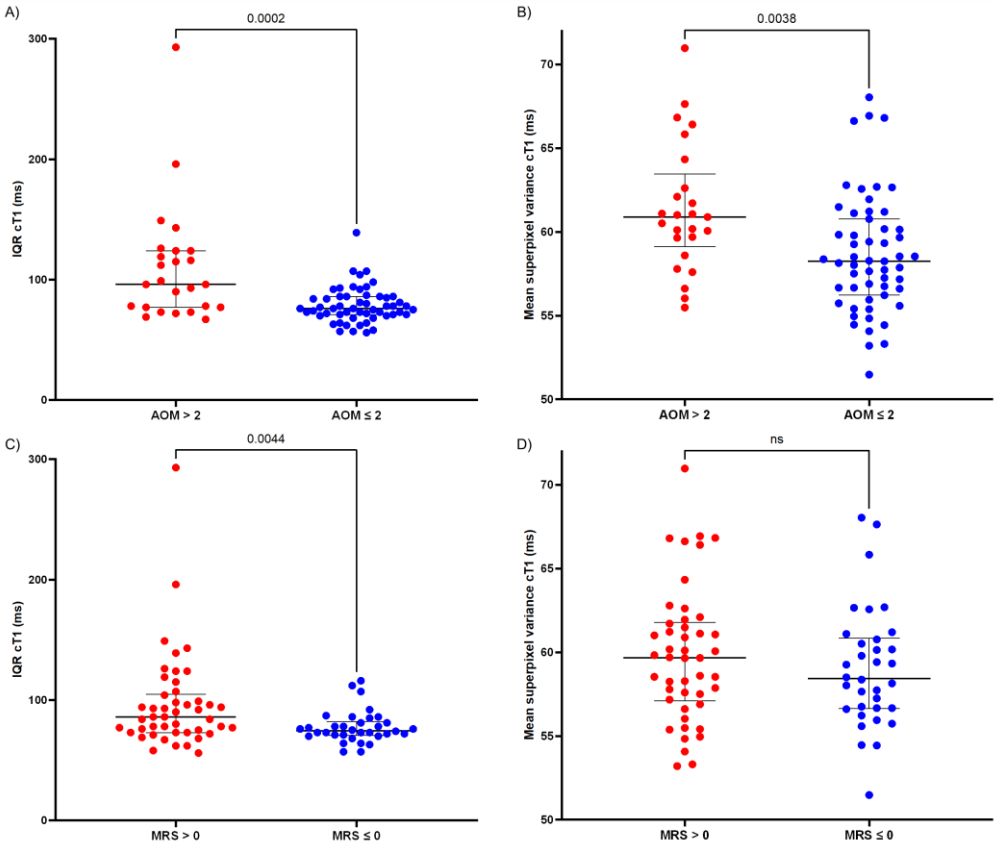


Figure 37. Comparisons of IQR cT1 and mean superpixel variance cT1 in 80 patients with PSC stratified into high-risk and low-risk groups by (A-B) Amsterdam-Oxford Model score (AOM); and (C-D) Mayo Risk Score (MRS) thresholds. ns, not significant. Lines and error bars indicate median and interquartile range.

Table 47. Performances of liver cT1 metrics to discriminate between high-risk and low-risk PSC stratified by Amsterdam-Oxford Model score (AOM) and Mayo Risk Score (MRS) thresholds.

	High risk		Low risk		p-value	AUC (95% CI)	Cut-off (ms)	Sensitivity (%)	Specificity (%)
	AOM > 2 (n=25)		AOM ≤ 2 (n=55)						
Median liver cT1 (ms)	791.0	756.5–822.0	743.0	713.0–780.0	0.0012	0.724 (0.600–0.847)	790.5	87	52
IQR liver cT1(ms)	96.0	77.0–124.0	76.0	71.0–86.0	0.0002	0.758 (0.636–0.880)	95.0	91	56
Mean superpixel variance cT1 (ms)	60.9	59.1–63.5	58.3	56.2–60.8	0.0026	0.708 (0.590–0.827)	59.5	65	76
	MRS > 0 (n=46)		MRS ≤ 0 (n=34)						
Median liver cT1 (ms)	771.0	727.8–810.5	744.5	716.0–785.3	0.0819	-	-	-	-
IQR liver cT1(ms)	86.0	73.0–104.8	74.5	70.8–82.0	0.0044	0.685 (0.569–0.801)	88.5	88	48
Mean superpixel variance cT1 (ms)	59.7	57.1–61.8	58.4	56.7–60.9	0.3700	-	-	-	-

All values are presented as median (interquartile range).

5.6.1.8 Risk stratification by morphological features on liver MRI

IQR cT1 and mean superpixel variance cT1 were significantly higher in patients with Anali score > 2, presence of hepatic dysmorphy, cirrhosis, and portal hypertension when assessed by either radiologist (Figure 38,Figure *39*). There were no differences in the median cT1 except when assessed by reader 2 for the presence of cirrhosis (Table 48).

For the classification of patients with Anali score > 2 or ≤ 2, IQR cT1 (reader 1: AUC = 0.783, 95% CI 0.678–0.888; reader 2: AUC = 0.778, 95% CI 0.671–0.885) performed better than mean superpixel variance cT1 (reader 1: AUC = 0.662, 95% CI 0.535–0.789; reader 2: AUC = 0.713, 95% CI 0.592–0.834; Table 49). The performance of IQR cT1 to detect hepatic dysmorphy (reader 1: AUC = 0.769, 95% CI 0.665–0.872; reader 2: AUC = 0.780, 95% CI 0.677–0.884), cirrhosis (reader 1: AUC = 0.777, 95% CI 0.664–0.890; reader 2: AUC = 0.753, 95% CI 0.629–0.877), and portal hypertension (reader 1: AUC = 0.802, 95% CI 0.697–0.906; reader 2: AUC = 0.752, 95% CI 0.625–0.879) was fair to good (Table 49).

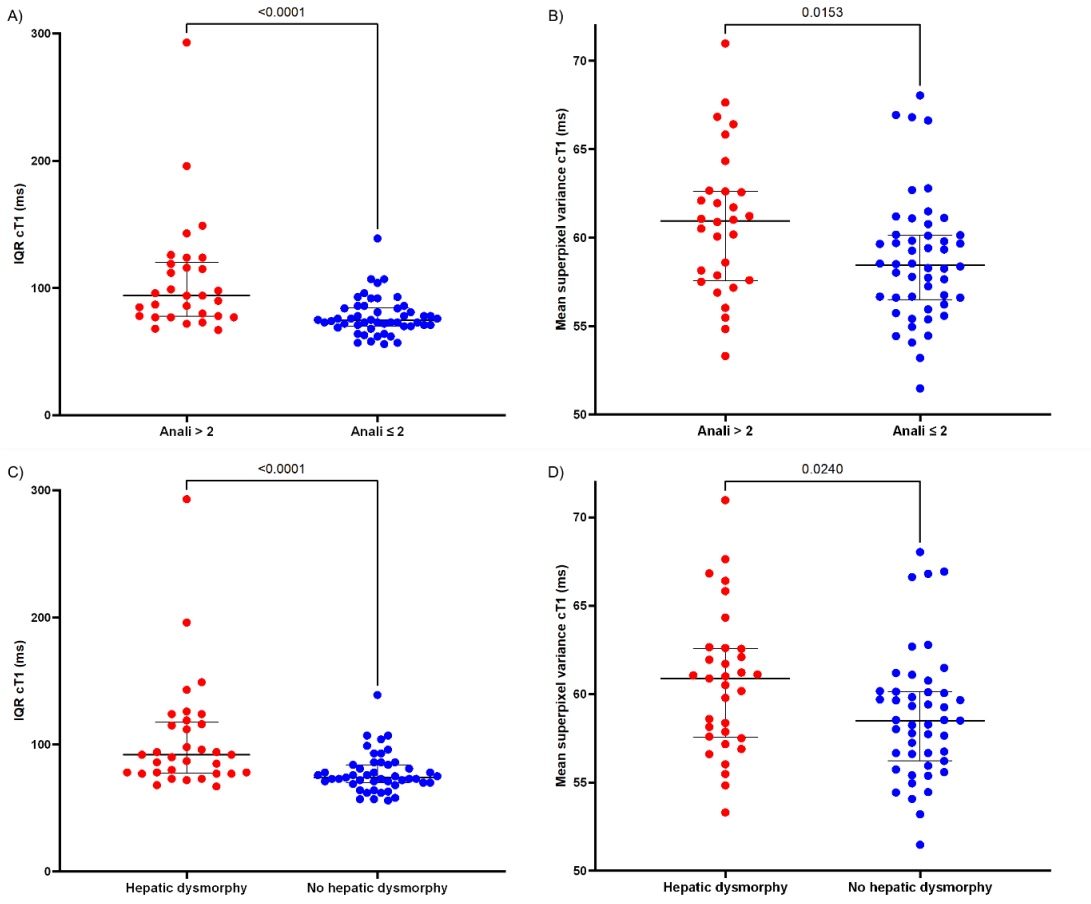


Figure 38. Comparisons of IQR cT1 and mean superpixel variance cT1 in 80 patients with PSC stratified into high-risk and low-risk groups by (A-B) Anali score threshold; and (C-D) the presence of hepatic dysmorphy on MRI. Lines and error bars indicate median and interquartile range.

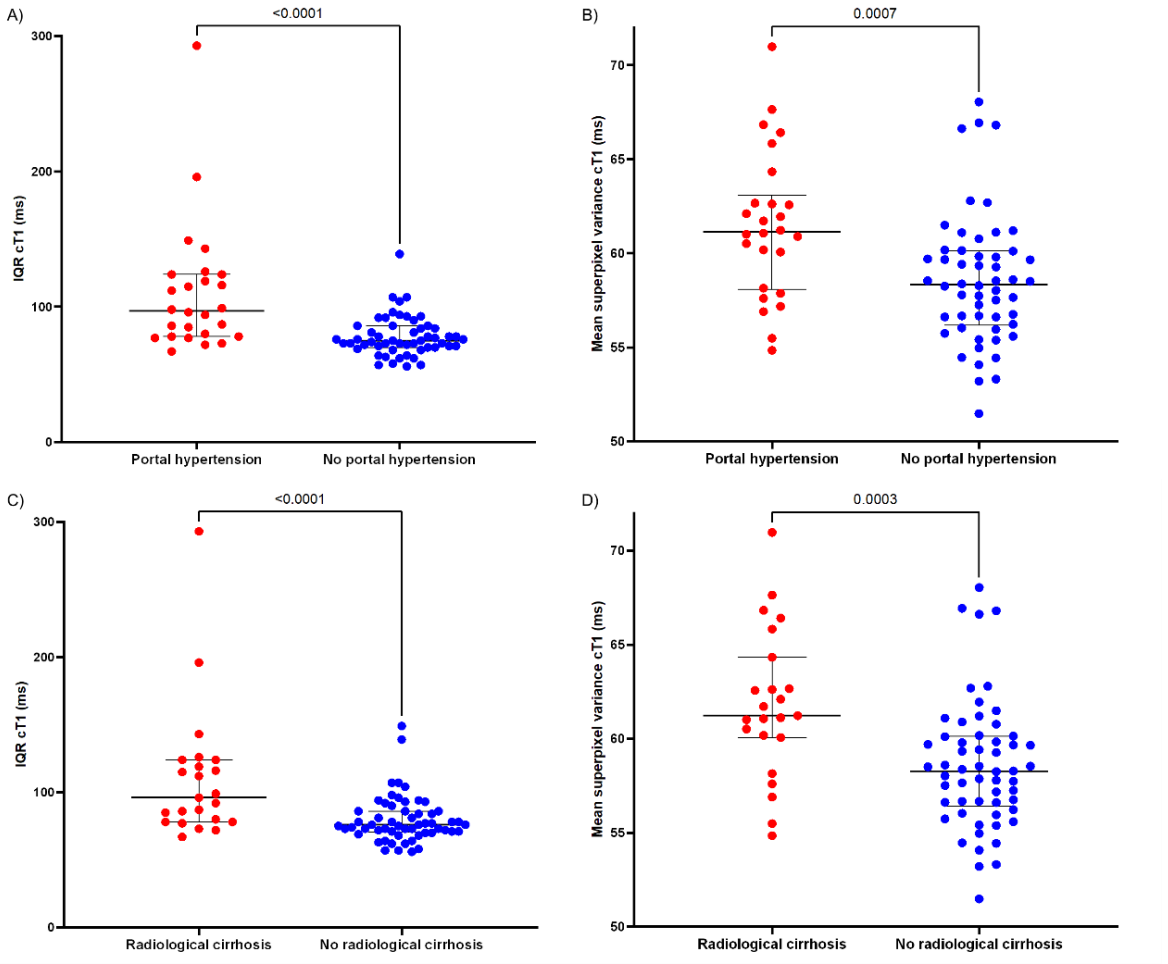


Figure 39. Comparisons of IQR cT1 and mean superpixel variance cT1 in 80 patients with PSC stratified into high-risk and low-risk groups by the presence of (A-B) portal hypertension; and (C-D) cirrhosis on MRI. Lines and error bars indicate median and interquartile range.

Table 48. Liver cT1 metrics in high-risk and low-risk PSC stratified by Anali score and morphological features on MRI.

	Reader 1					Reader 2				
	High risk		Low risk		p-value	High risk		Low risk		p-value
Anali score	> 2 (n=30)		≤ 2 (n=50)			> 2 (n=29)		≤ 2 (n=51)		
Median cT1 (ms)	775.5	736.8–817.0	752.5	713.0–784.5	0.0429	775.0	734.5–809.5	751.0	713.0–786.0	0.0540
IQR cT1 (ms)	94.0	77.8–120.3	74.5	70.0–84.5	<0.0001	94.0	77.5–117.5	75.0	70.0–84.0	<0.0001
Mean superpixel variance cT1 (ms)	61.0	57.6–62.6	58.4	56.5–60.2	0.0153	61.1	58.0–62.6	58.3	56.2–60.1	0.0014
Dysmorphy	Yes (n=33)		No (n=47)			Yes (n=32)		No (n=48)		
Median cT1 (ms)	772.0	734.5–801.0	751.0	713.0–784.0	0.0756	773.5	729.3–812.8	752.5	714.0–785.5	0.0889
IQR cT1 (ms)	92.0	77.5–117.5	74.0	70.0–84.0	<0.0001	93.0	77.3–118.3	74.5	70.0–84.0	<0.0001
Mean superpixel variance cT1 (ms)	60.9	57.6–62.6	58.5	56.2–60.1	0.0240	61.0	57.7–62.7	58.3	56.6–60.0	0.0054
Portal hypertension	Yes (n=26)		No (n=54)			Yes (n=26)		No (n=54)		
Median cT1 (ms)	775.5	729.8–817.0	754.5	714.5–786.8	0.0897	773.5	732.3–812.8	754.5	713.0–790.0	0.0876
IQR cT1 (ms)	97.0	78.0–124.0	75.0	70.0–86.0	<0.0001	97.5	77.3–122.8	76.0	71.0–86.8	0.0006
Mean superpixel variance cT1 (ms)	61.1	58.1–63.1	58.3	56.2–60.1	0.0007	61.1	60.1–65.5	58.3	56.3–60.2	0.0005
Cirrhosis	Yes (n=23)		No (n=57)			Yes (n=19)		No (n=61)		
Median cT1 (ms)	779.0	730.0–820.0	755.0	715.0–787.5	0.0570	803.0	757.0–869.0	751.0	713.0–782.0	0.0002
IQR cT1 (ms)	96.0	78.0–124.0	76.0	70.5–86.0	<0.0001	99.0	76.0–124.0	77.0	71.5–88.5	0.0069
Mean superpixel variance cT1 (ms)	61.2	60.1–64.3	58.3	56.4–60.1	0.0003	61.0	56.9–65.8	58.5	56.6–61.0	0.0416

All values are presented as median (interquartile range).

Table 49. Performances of liver cT1 metrics to discriminate between high-risk and low-risk PSC stratified by Anali score, presence of hepatic dysmorphy, portal hypertension, and cirrhosis.

	Reader	Cut-off (ms)	AUC (95% CI)		p-value	Sensitivity (%)	Specificity (%)
Anali score							
IQR cT1 (ms)	1	76.5	0.78	0.68–0.89	<0.0001	87	60
	2	84.5	0.78	0.67–0.89	<0.0001	69	77
Mean superpixel variance cT1 (ms)	1	60.2	0.66	0.54–0.79	0.0157	60	78
	2	60.2	0.71	0.59–0.83	0.0016	66	80
Median cT1 (ms)	1	789.5	0.64	0.51–0.76	0.0431	80	47
	2	755.5	0.63	0.50–0.76	0.0540	55	69
Presence of dysmorphy							
IQR cT1 (ms)	1	76.5	0.77	0.67–0.87	<0.0001	85	62
	2	76.5	0.78	0.68–0.88	<0.0001	84	60
Mean superpixel variance cT1 (ms)	1	60.2	0.65	0.54–0.77	0.0243	58	79
	2	60.2	0.68	0.56–0.81	0.0058	63	81
Median cT1 (ms)	1	789.5	0.62	0.49–0.74	0.0753	81	45
	2	789.5	0.61	0.49–0.74	0.0884	79	44
Presence of portal hypertension							
IQR cT1 (ms)	1	93.5	0.80	0.70–0.91	<0.0001	58	89
	2	95.0	0.75	0.63–0.88	0.0008	55	87
Mean superpixel variance cT1 (ms)	1	60.2	0.73	0.61–0.85	0.0009	69	80
	2	60.0	0.76	0.64–0.88	0.0006	80	72
Median cT1 (ms)	1	790.5	0.62	0.48–0.76	0.0891	83	42
	2	790.5	0.63	0.50–0.76	0.0871	80	40
Presence of cirrhosis							
IQR cT1 (ms)	1	84.5	0.78	0.66–0.89	0.0001	70	72
	2	95.0	0.70	0.55–0.86	0.0076	58	87
	Clinical†	98.5	0.80	0.67–0.93	0.0007	88	62
Mean superpixel variance cT1 (ms)	1	60.2	0.75	0.63–0.88	0.0004	74	79
	2	60.3	0.64	0.49-0.79	0.0629	58	72
	Clinical†	59.8	0.80	0.68–0.92	0.0007	67	92
Median cT1 (ms)	1	790.5	0.64	0.50–0.78	0.0571	84	48
	2	790.5	0.78	0.66–0.90	0.0003	87	63
	Clinical†	790.5	0.75	0.61–0.90	0.0042	82	62

†The diagnosis of cirrhosis was based on a combination of biochemical, radiological, and endoscopic findings.

5.6.1.9 Risk stratification by morphological features on MRCP

Modified Amsterdam stricture severity score

IQR cT1 was higher in patients with total stricture severity score > 4 and intrahepatic stricture severity score > 2 compared to total stricture severity score ≤ 4 and intrahepatic stricture severity score ≤ 2 when assessed by either radiologist (Figure 40,Table 50). There were no differences in the other metrics between the high-risk and low-risk groups. IQR cT1 was able to detect the high-risk groups with AUC ranging from 0.701–0.725 between the two radiologists (Table 51).

Presence of dominant stricture

The results for IQR cT1 was mixed (Figure 40). It was observed to be higher in patients with a dominant stricture when assessed by radiologist 2 ($p = 0.0023$), but no difference when assessed by radiologist 1 ($p = 0.0769$). All other metrics were not significantly different when a dominant stricture was deemed to be present by either radiologist.

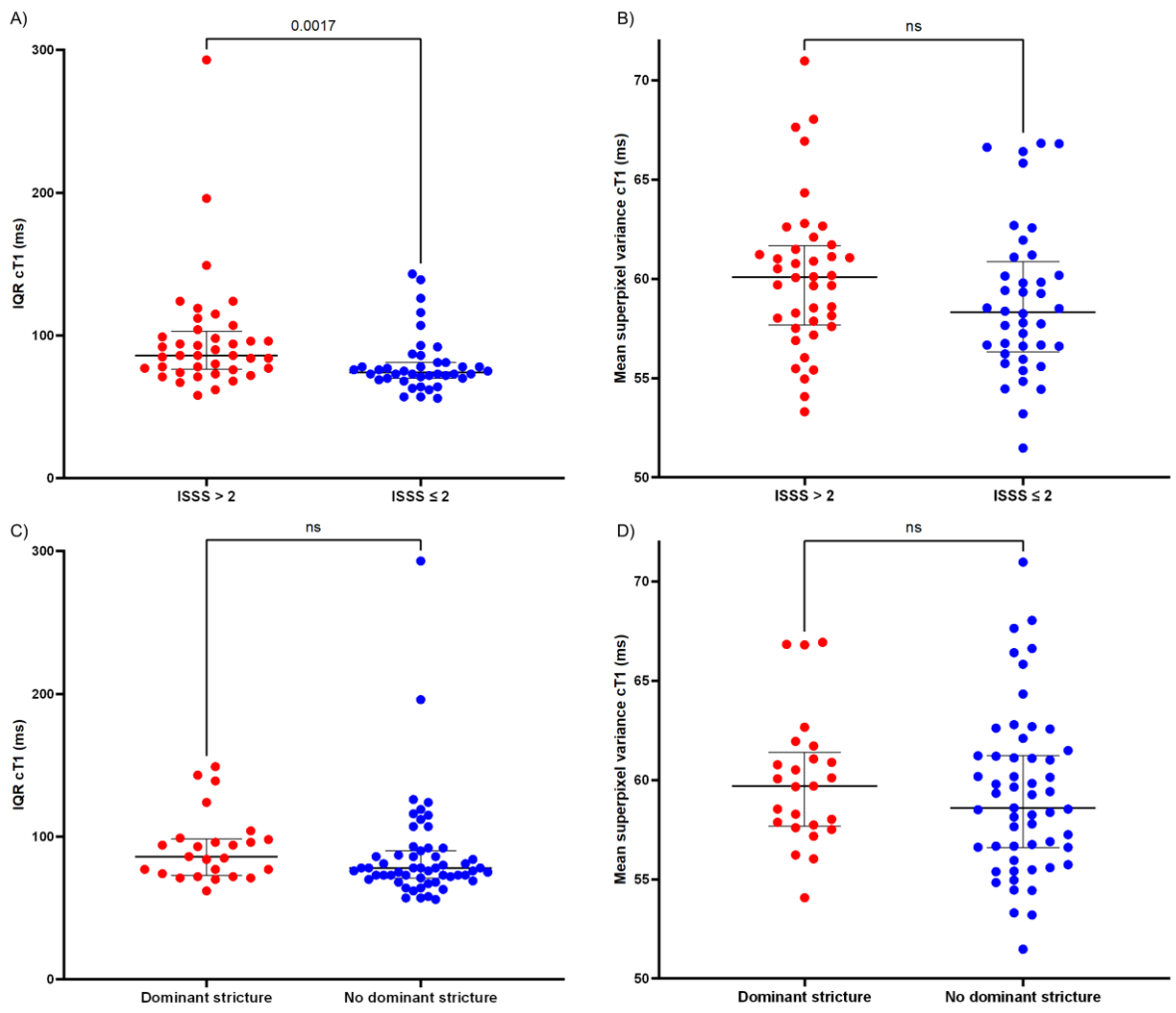


Figure 40. Comparisons of IQR cT1 and mean superpixel variance cT1 in 80 patients with PSC stratified into high-risk and low-risk groups by (A-B) modified Amsterdam intrahepatic stricture severity score (ISSS) threshold; and (C-D) the presence of dominant stricture. ns, not significant. Lines and error bars indicate median and interquartile range.

Table 50. Liver cT1 metrics in high-risk and low-risk PSC stratified by stricture severity scores and presence of dominant stricture.

	Reader 1					Reader 2				
	High risk		Low risk		p-value	High risk		Low risk		p-value
TSSS	> 4 (n = 32)		≤ 4 (n = 48)			> 4 (n = 31)		≤ 4 (n = 49)		
Median cT1 (ms)	757.0	721.8–792.5	764.0	721.5–790.0	0.9747	721.0	755.0–791.0	766.0	723.0–791.5	0.7853
IQR cT1 (ms)	89.0	77.3–110.0	75.0	69.3–85.5	0.0003	77.0	86.0–104.0	75.0	69.5–88.5	0.0019
Mean superpixel variance cT1 (ms)	60.3	57.9–61.7	58.4	56.0–60.9	0.0442	58.0	60.2–61.5	58.4	56.1–61.2	0.0538
ISSS	> 2 (n = 40)		≤ 2 (n = 40)			> 2 (n = 42)		≤ 2 (n = 38)		
Median cT1 (ms)	763.5	729.3–792.5	754.0	714.0–790.0	0.4506	763.5	727.8–795.5	754.0	716.0–790.0	0.4273
IQR cT1 (ms)	86.0	76.3–102.8	74.0	70.0–81.0	0.0017	86.0	76.8–104.8	73.0	69.8–78.8	0.0004
Mean superpixel variance cT1 (ms)	60.1	57.7–61.7	58.3	56.3–60.9	0.0742	59.9	57.6–61.6	58.4	56.5–61.1	0.1799
Dominant stricture	Yes (n = 25)		No (n = 55)			Yes (n = 30)		No (n = 50)		
Median cT1 (ms)	769.0	727.5–821.0	757.0	717.0–790.0	0.4496	765.0	729.0–813.5	757.0	716.5–790.0	0.3935
IQR cT1 (ms)	86.0	73.0–98.5	78.0	71.0–90.0	0.0769	91.5	76.3–106.8	76.0	69.8–86.0	0.0023
Mean superpixel variance cT1 (ms)	59.7	57.7–61.4	58.6	56.6–61.2	0.4476	59.9	57.8–61.8	58.4	55.9–61.2	0.1133

All values are presented as median (interquartile range). TSSS; total stricture severity score, ISSS; intrahepatic stricture severity score

Table 51. Performance of IQR cT1 to discriminate between high-risk and low-risk PSC groups stratified by modified Amsterdam severity scores and the presence of a dominant stricture.

		Cut-off (ms)	AUC	95% CI	p-value	Sensitivity (%)	Specificity (%)
TSSS							
IQR cT1 (ms)	Reader 1	76.5	0.733	0.622–0.844	0.0004	58	81
	Reader 2	76.5	0.704	0.589–0.819	0.0022	57	81
ISSS							
IQR cT1 (ms)	Reader 1	82.5	0.701	0.584–0.817	0.0020	78	65
	Reader 2	82.5	0.725	0.610–0.840	0.0005	82	64
Dominant stricture							
IQR cT1 (ms)	Reader 1	92.5	0.624	0.492–0.756	0.0768	80	48
	Reader 2	82.5	0.702	0.582–0.821	0.0026	72	67

TSSS; total stricture severity score, ISSS; intrahepatic stricture severity score

5.6.1.10 Independent predictors of high-risk PSC

Univariate and multivariate analysis results are summarised in Table 53. In the multivariate model, only LS (OR: 1.4, p = 0.038) was associated with an increased likelihood of having high-risk PSC when stratified by AOM (Table 54). The model accurately classified 89% of cases (AUC = 0.93, 95% CI:0.87–0.99; Figure 41).

When stratified by MRS, ELF (OR: 4.8, p = 0.005) was the only covariate associated with an increased likelihood of having high-risk PSC (Table 55). The model accurately classified 84% of cases (AUC = 0.91, 95% CI: 0.85–0.98; Figure 41). None of the liver cT1 metrics were independent predictors of high-risk PSC group when defined by AOM or MRS thresholds. Statistical interaction terms between covariates were not analysed.

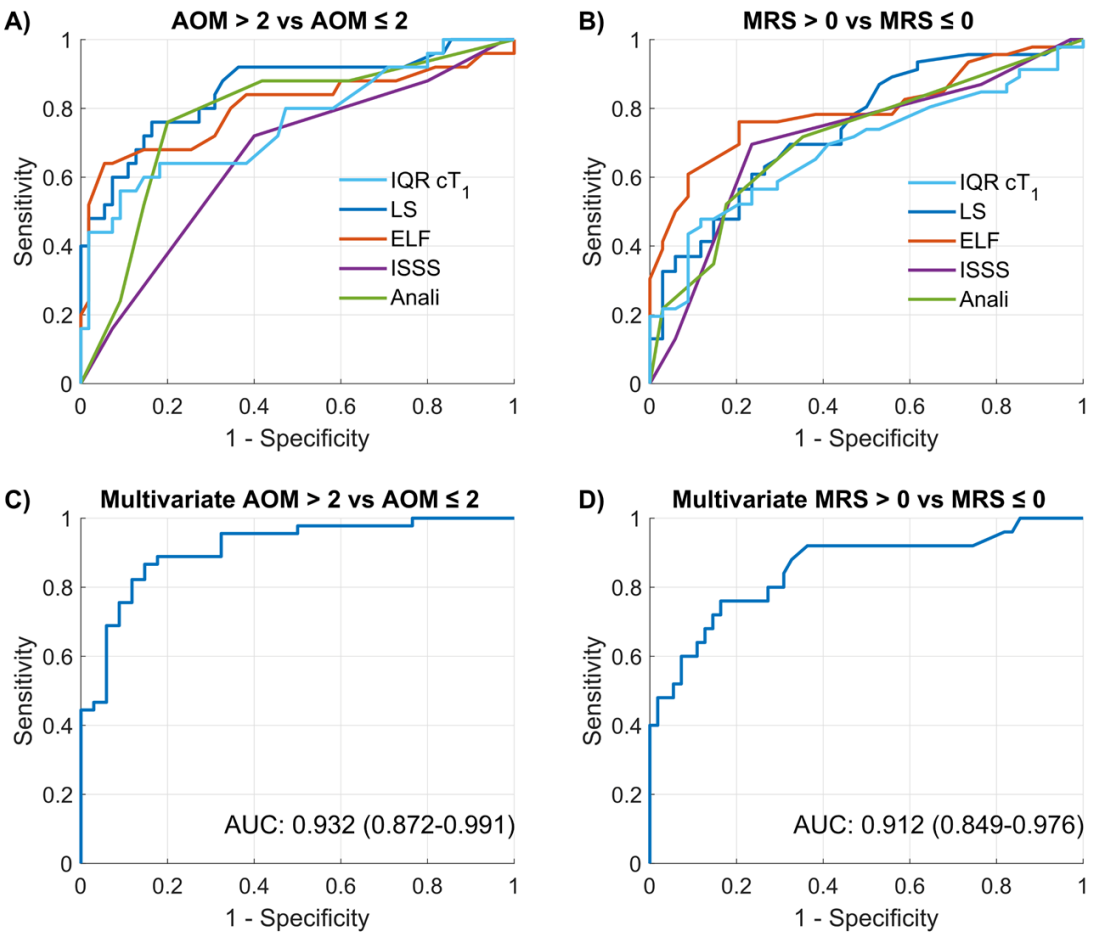


Figure 41. ROC curves summarising the individual discriminatory performance of non-invasive risk stratification tools and the multivariable logistic regression models for identifying the high-risk PSC group. (A) Amsterdam-Oxford Model (AOM: IQR cT1 AUC=0.76 (0.64–0.88), LS AUC = 0.85 (0.76–0.95), ELF AUC = 0.80 (0.68–0.93), ISSS AUC = 0.66 (0.53–0.79), Anali AUC = 0.78 (0.67–0.90)); and (B) Mayo Risk Score (MRS: IQR cT1 AUC=0.69 (0.57–0.80), LS AUC = 0.75 (0.64–0.85), ELF AUC = 0.80 (0.70–0.89), ISSS AUC = 0.71 (0.60–0.83), Anali AUC = 0.71 (0.60–0.83)). Multiple logistic regression model performance to detect high-risk PSC stratified by (C) AOM with LS (OR: 1.4, p = 0.038) as the only independent predictor; and (D) MRS with ELF (OR: 4.8, p = 0.005) as the only independent predictor. ELF, enhanced liver fibrosis score; ISSS, modified Amsterdam intrahepatic stricture severity score; LS, transient elastography liver stiffness.

Table 52. Performance of non-invasive markers to discriminate high-risk vs low-risk PSC group stratified by Amsterdam-Oxford Model and Mayo Risk Score thresholds.

		AUC	95% CI	p-value	Cut-off	Sensitivity (%)	Specificity (%)
Amsterdam-Oxford model[†]							
IQR cT1 (ms)		0.76	0.64–0.88	<0.001	95.0	56	91
LS (kPa)		0.85	0.76–0.95	<0.001	8.8	76	84
ELF		0.80	0.68–0.93	<0.001	10.5	64	95
ISSS	Reader 1	0.66	0.53–0.79	0.023	-	72	40
	Reader 2	0.65	0.52–0.78	0.034	-	72	44
Anali	Reader 1	0.78	0.67–0.90	<0.001	-	76	20
score	Reader 2	0.80	0.69–0.91	<0.001	-	88	35
Mayo risk score							
IQR cT1 (ms)		0.69	0.57–0.80	0.005	88.5	48	88
LS (kPa)		0.75	0.64–0.85	<0.001	7.2	61	77
ELF		0.80	0.70–0.89	<0.001	9.4	76	79
ISSS	Reader 1	0.71	0.60–0.83	0.001	-	70	24
	Reader 2	0.76	0.64–0.87	<0.001	-	74	24
Anali	Reader 1	0.71	0.60–0.83	0.001	-	72	35
score	Reader 2	0.81	0.71–0.90	<0.001	-	74	21
xULN ALP		0.67	0.55–0.80	0.008	1.2	65	71

[†]Performance of xULN ALP not tested against AOM as the model already includes serum ALP.

Table 53. Univariate and multiple logistic regression analysis for the detection of high-risk PSC defined by Amsterdam-Oxford Model score and Mayo Risk Score thresholds.

	Amsterdam-Oxford Model		Mayo Risk Score	
	Univariate p-value	Multivariate p-value	Univariate p-value	Multivariate p-value
Liver cT1 metrics				
Median cT1 (ms)	0.021	0.339	0.058	0.067
IQR cT1 (ms)	0.001	0.786	0.010	0.658
Mean superpixel variance cT1 (ms)	0.005	0.515	0.358	-
Liver fibrosis markers				
LS (kPa)	<0.001	**0.038**	0.008	0.644
ELF	<0.001	0.285	<0.001	**0.005**
Qualitative MRI-MRCP features				
Reader 1				
ISSS	0.027	0.497	0.002	0.316
Anali score	<0.001	0.602	0.002	0.395
Dominant stricture	0.258		0.028	0.575
*Reader 2**				
ISSS	0.031	-	<0.001	-
Anali score	<0.001	-	<0.001	-
Dominant stricture	0.006	-	0.029	-
Laboratory parameters				
ALP (IU/l)	‡		0.189	
GGT (IU/l)	0.291	-	0.045	0.287
Total bilirubin	‡		‡	
ALT (IU/l)	0.815	-	0.211	-
AST (IU /l)	‡		‡	
Albumin (g/l)	‡		‡	
Platelet count (x109/l)	‡		0.350	-
Prothrombin time (s)	0.040	0.052	0.757	-
Clinical features				
Age	‡		‡	
Gender	0.563	-	0.324	-
PSC duration (years)	0.224	-	0.254	-
Ulcerative colitis	0.884	-	0.912	-
Extrahepatic disease	0.673	-	0.083	0.656

*Reader 2 data was excluded from the multivariate analysis given the correlation with reader 1.

‡These variables are already incorporated into MRS or AOM, therefore excluded from analysis.

Abbreviations: ALP, alkaline phosphatase; ALT, alanine aminotransferase; AST, aspartate aminotransferase; ELF, enhanced liver fibrosis score; GGT, gamma-glutamyltransferase; ISSS, intrahepatic stricture severity score; LS, liver stiffness; xULN, times upper limit of normal.

Table 54. Multivariate logistic regression model parameters to detect Amsterdam-Oxford Model (AOM) > 2 group.

Covariates	B	S.E.	OR	95% CI for OR		p-value
				Lower	Upper	
Median cT1 (ms)	0.01	0.01	1.00	0.99	1.01	0.34
IQR cT1 (ms)	0.01	0.03	1.00	0.94	1.05	0.79
Mean superpixel variance cT1 (ms)	0.01	0.15	1.10	0.82	1.47	0.52
LS (kPa)	0.32	0.15	1.40	1.02	1.85	**0.04**
ELF	0.53	0.50	1.70	0.64	4.49	0.29
PT	1.72	0.89	5.57	0.98	31.64	0.05
ISSS†						0.56
ISSS (1)	-14.27	40192.97	0.00	0.00	-	1.00
ISSS (2)	2.86	4.77	17.47	0.00	199488.44	0.55
ISSS (3)	3.32	4.75	27.69	0.00	304259.41	0.48
ISSS (4)	5.41	4.84	223.94	0.02	2968793.89	0.26
Anali score†						0.50
Anali (1)	1.26	2.04	3.53	0.07	191.77	0.54
Anali (2)	-21.45	10033.07	0.00	0.00	-	1.00
Anali (3)	-1.85	1.49	0.16	0.01	2.92	0.21
Anali (4)	0.16	1.77	1.17	0.04	37.66	0.93
Anali (5)	1.68	1.97	5.36	0.11	253.00	0.39
Constant	-8.90	20.75	0.00	-	-	0.67

† Entered as categorical variables into the multivariate model

Abbreviations: ELF, enhanced liver fibrosis score; ISSS, intrahepatic stricture severity score; LS, transient elastography liver stiffness; PT, prothrombin time.

Table 55. Multivariate logistic regression model parameters to detect Mayo Risk Score (MRS) > 0 group.

Covariates	B	S.E.	OR	95% CI for OR		p-value
				Lower	Upper	
Median cT1 (ms)	0.01	0.01	1.01	1.00	1.02	0.07
IQR cT1 (ms)	0.01	0.02	1.01	0.98	1.04	0.66
LS (kPa)	-0.05	0.12	0.95	0.75	1.19	0.64
ELF	1.56	0.56	4.77	1.59	14.33	**0.01**
ISSS†						0.32
ISSS (1)	-21.19	40192.97	0.00	0.00	-	1.00
ISSS (2)	-1.36	1.45	0.26	0.02	4.39	0.35
ISSS (3)	-1.86	1.41	0.16	0.01	2.49	0.19
ISSS (4)	0.31	1.42	1.37	0.08	22.35	0.83
Anali score†						0.39
Anali (1)	-0.77	1.73	0.46	0.02	13.70	0.66
Anali (2)	-2.11	1.66	0.12	0.00	3.13	0.20
Anali (3)	-1.40	1.53	0.25	0.01	4.92	0.36
Anali (4)	0.72	2.00	2.06	0.04	104.22	0.72
Anali (5)	-2.88	1.86	0.06	0.00	2.17	0.12
Dominant stricture	-0.58	1.03	0.56	0.07	4.22	0.58
GGT	0.00	0.00	1.00	1.00	1.01	0.29
Extrahepatic disease	-0.32	0.71	0.73	0.18	2.93	0.66
Constant	-19.89	6.77	0.00	-	-	0.00

† Entered as categorical variables into the multivariate model
Abbreviations: ELF, enhanced liver fibrosis score; GGT, gamma-glutamyltransferase; ISSS, intrahepatic stricture severity score; LS, transient elastography liver stiffness.

5.6.1.11 Association between liver cT1 metrics and patient-reported outcomes in PSC

IQR cT1 and mean superpixel variance cT1 did not show any significant correlation with any domains of the SF-36 or PSC-PRO. However, median cT1 negatively correlated with all domains of the SF-36 as summarised in Table 56, supporting the link between QoL and liver fibrosis as discussed in chapter 4. Median cT1 did not correlate with any of the domains in PSC-PRO.

5.6.1.12 Association between liver fibrosis markers and patient-reported outcomes in PSC

Liver stiffness was negatively correlated with all the mean domain scores of SF-36 except general health and mental health domains (Table 57). Mean domain scores were higher in patients with LS > 9.6kPa compared to LS ≤ 9.6kPa, however, in the multiple comparisons test, only role physical was significantly lower in those with advanced fibrosis defined by transient elatography (67% vs 85%, p = 0.0106). There were no correlations between SF-36 domains and ELF.

In the PSC-PRO, LS (r_s = 0.30, p = 0.008), and ELF (r_s = 0.26, p = 0.022) correlated with the physical function domain. Activities of daily living (r_s = 0.26, p = 0.021) correlated with LS whilst role function (r_s = 0.27, p = 0.019) correlated with ELF.

Table 56. Correlation between SF-36 domains and median cT1.

SF-36 domain	Correlation coefficient	p-value
Physical component summary	-0.23	0.042
Physical functioning	-0.31	0.006
Role physical	-0.24	0.041
Bodily pain	-0.31	0.006
General health	-0.28	0.014
Mental component summary	-0.37	0.001
Energy/fatigue	-0.28	0.015
Social functioning	-0.32	0.005
Role emotional	-0.26	0.023
Mental health	-0.37	0.001

Table 57. Correlation between SF-36 and PSC-PRO domains and transient elastography liver stiffness.

	Correlation coefficient	p-value
SF-36 domains		
Physical component summary	-0.24	0.038
Physical functioning	-0.25	0.031
Role physical	-0.22	0.061
Bodily pain	-0.26	0.023
General health	-0.18	0.116
Mental component summary	-0.24	0.036
Energy/fatigue	-0.24	0.041
Social functioning	-0.27	0.021
Role emotional	-0.24	0.035
Mental health	-0.15	0.197
PSC-PRO domains		
Physical function	0.29	0.010
Activities of daily living	0.25	0.029

5.6.2 Validation cohort: RIAL, NICOLA, and MAPLE studies

5.6.2.1 Cohort characteristics

Recruitment details for this cohort of participants is summarised in Figure 42.

RIAL study

There were 99 scans available for analysis from the RIAL study and 85 were included. Reasons for exclusion were: 5 other liver diseases (sarcoidosis, cholestasis, haemosiderosis, unspecified), 4 PSC, 3 poor quality scans, and 2 classified as normal. The breakdown of liver diseases included were 32 NAFLD, 28 HCV, 7 HBV, 11 ArLD, 4 AIH, and 3 PBC.

NICOLA study

There were 97 scans available for analysis from the NICOLA study and 69 were included. Reasons for exclusion were: 13 duplicate patients (follow-up scans), 9 poor quality scans, 5 other liver diseases (alpha-1 antitrypsin, haemochromatosis, unspecified), and 1 PSC. The breakdown of included liver diseases was 28 NAFLD, 13 HCV, 1 HBV, 5 ArLD, 1 PBC, and 21 healthy.

MAPLE study

There were 153 scans available for analysis from the MAPLE study and 82 were included. Reasons for exclusion were: 51 classified as decompensated at recruitment, 12 acute liver injury, 4 other liver diseases (haemochromatosis, cryptogenic, unclassified), 2 PSC, and 2 poor quality scans. Final breakdown was 16 NAFLD, 8 HCV, 2 HBV, 23 ArLD, 4 AIH, 6 PBC, and 23 healthy.

Pooled results

In total, 236 participant scans were included in the final analysis. The baseline characteristics for each cohort is summarised in Table 58.

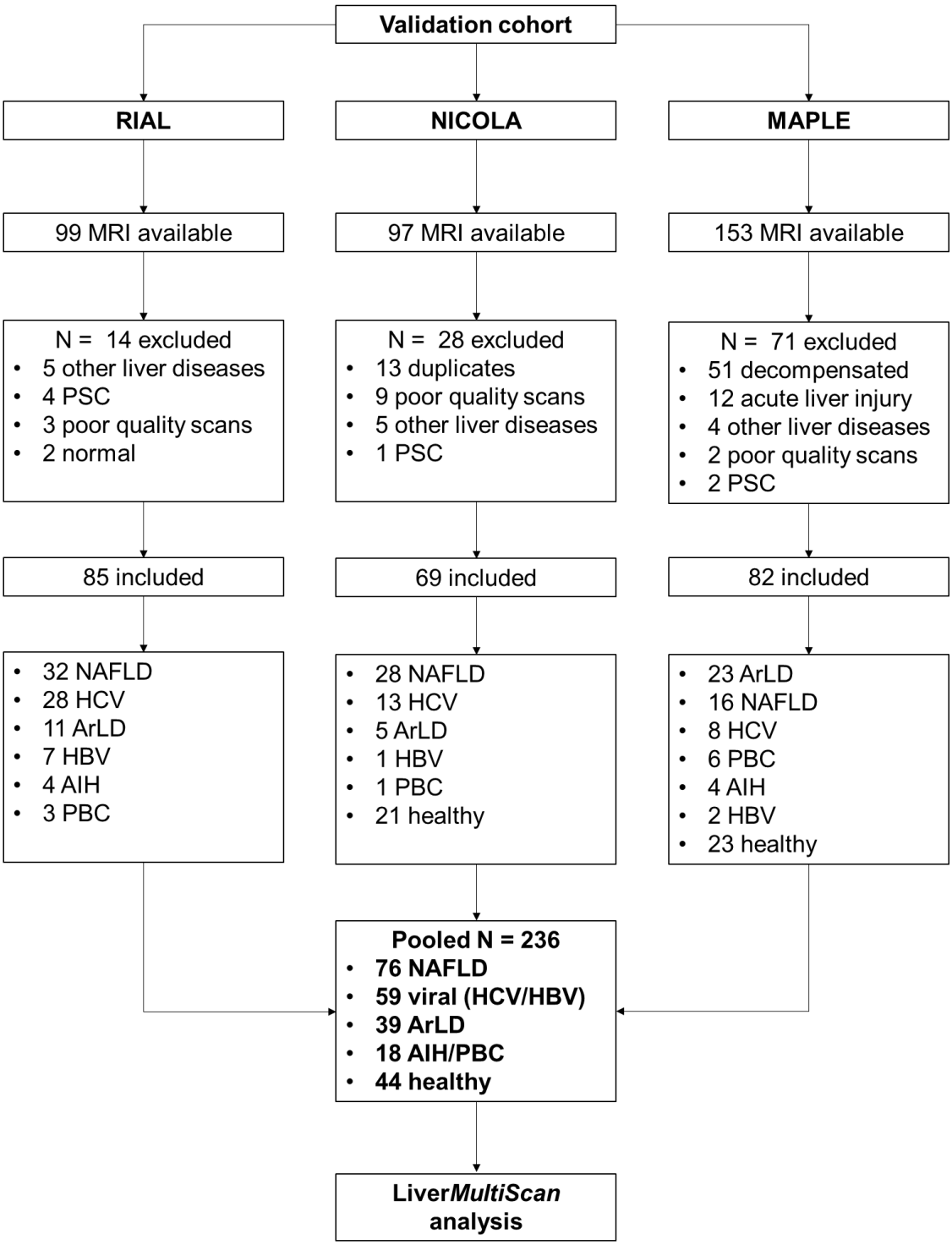


Figure 42. Breakdown of the pooled participants included in the validation cohort from the RIAL, NICOLA and MAPLE studies. AIH; autoimmune hepatitis, ArLD; alcohol-related liver disease, HBV; chronic hepatitis B, HCV; chronic hepatitis C, NAFLD; non-alcoholic fatty liver disease, PBC; primary biliary cholangitis.

Table 58. Baseline characteristics of all participants included in the final analysis.

	PSC (n = 80)	NAFLD (n = 76)	Viral (n = 59)	ArLD (n = 39)	AIH/PBC (n = 18)	Healthy (n = 65)
Age (years)	42 (31–55)	57 (47–64)	56 (48–60)	58 (49–67)	62 (51–69)	45 (33–59)
Male	54 (68)	42 (55)	47 (79)	25 (64)	4 (22)	36 (55)
Liver stiffness† (kPa)	6.9 (5.4–10.8)	11.9 (5.9–20.8)	9.5 (6.3–20.7)	20.3 (8.3–34.0)	14.7 (11.1–20.0)	4.5 (3.9–5.3)
ELF‡	9.4 (8.7–10.0)	10.3 (8.7–10.9)	10.5 (9.3–12.1)	10.4 (9.3–11.7)	10.5 (10.1–10.9)	8.6 (8.4–9.3)
xULN ALP	1.2 (0.8–1.8)	0.6 (0.5–0.8)	0.6 (0.5–0.7)	0.6 (0.6–0.8)	1.1 (0.7–2.2)	0.5 (0.4–0.5)
Bilirubin (µmol/L)	13 (10–21)	10 (7–15)	12 (9–17)	12 (9–16)	9 (8–11)	11 (9–15)
ALT (IU/l)	42 (27–84)	41 (27–64)	61 (34–121)	30 (22–44)	42 (28–70)	17 (13–22)
Prothrombin time (s)	10.6 (10.3–1.0)	13.1 (12.1–14.3)	13.3 (11.7–14.1)	12.0 (11.6–13.0)	11.7 (10.9–14.0)	11.3 (10.9–11.6)
Albumin (g/l)	40 (37–41)	44 (39–46)	42 (38–44)	39 (37–42)	38 (35–42)	42 (40–44)
Platelet count (x 10^9/l)	253 (192–305)	200 (149–265)	164 (115–205	179 (132–234)	181 (142–275)	248 (207–270)

†Transient elastography was attempted/successful/reliable in 59 NAFLD, 49 viral, 32 ArLD, 15 AIH/PBC and 63 healthy.
‡ELF test performed only in the MAPLE study on 35 NAFLD, 17 viral, 15 ArLD, 11 AIH/PBC and 23 healthy.
ALP, alkaline phosphatase; ALT, alanine aminotransferase; ELF, enhanced liver fibrosis score; xULN, times upper limit of normal.

5.6.2.2 Comparison of candidate liver cT1 metrics between liver diseases

IQR cT1 and mean superpixel variance cT1 were both higher in all liver diseases compared to healthy volunteers. There were no significant differences in IQR cT1 between PSC and the other liver diseases as depicted in Figure 43. On the contrary, mean superpixel variance cT1 was higher in PSC compared to any other liver diseases (PSC: 59.5ms; NAFLD: 56.4ms; viral hepatitis: 54.4ms, ArLD 55.9ms; AIH/PBC 54.2ms; $p < 0.0001$). In the multiple comparisons test, it was able to distinguish PSC from any other liver disease (PSC vs. NAFLD: $p = 0.0006$, PSC vs. viral hepatitis: $p < 0.0001$, PSC vs. ArLD: $p = 0.0014$, PSC vs. AIH/PBC: $p = 0.0002$) as shown in Figure 44. There were no significant differences in mean superpixel variance cT1 between any other liver diseases.

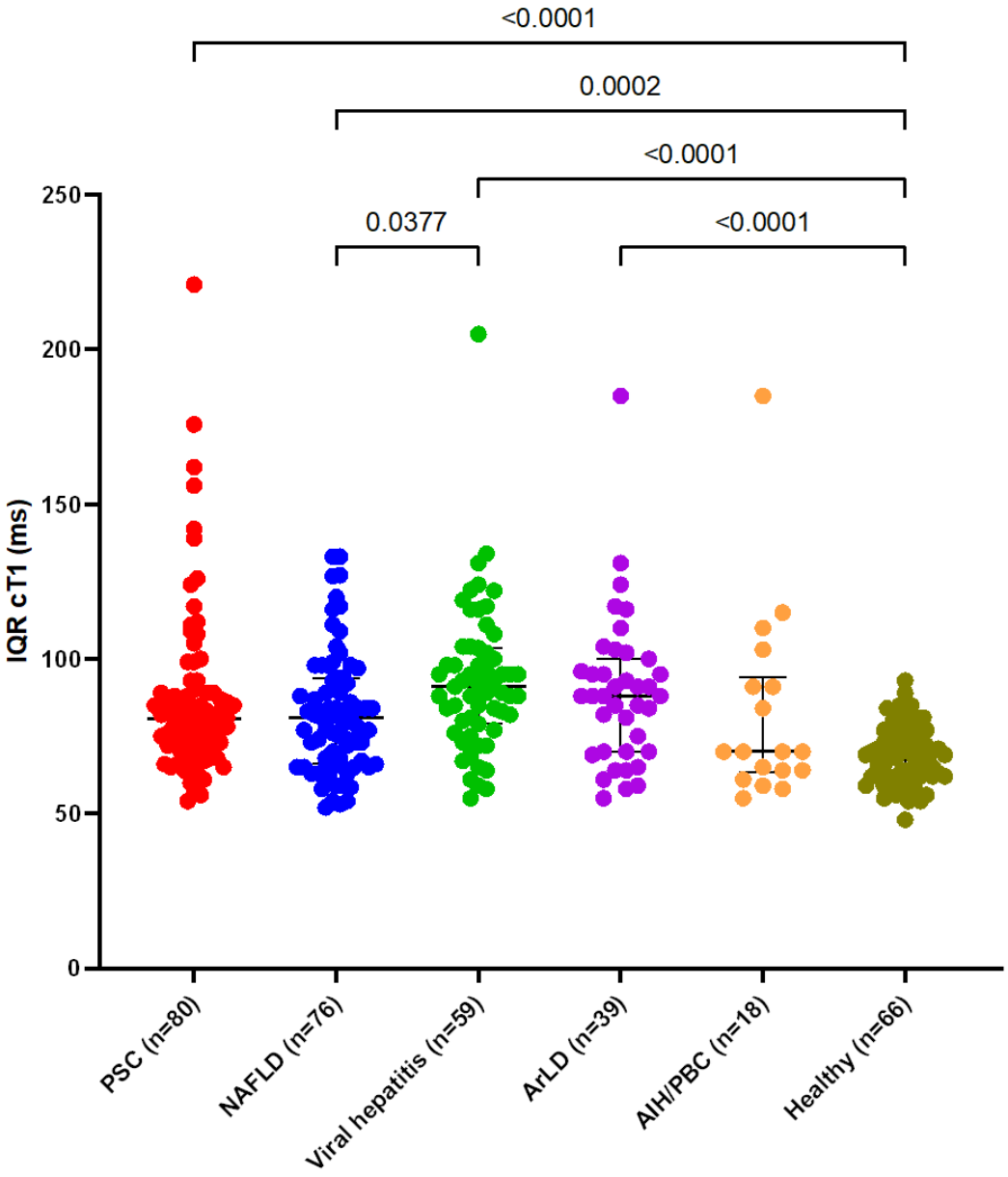


Figure 43. IQR cT1 in chronic liver diseases and healthy volunteers. ANOVA with Tukey's multiple comparisons tests failed to differentiate between liver diseases but was higher in liver diseases than healthy volunteers. PSC; primary sclerosing cholangitis, NAFLD; non-alcoholic fatty liver disease, ArLD; alcohol-related liver disease AIH/PBC; autoimmune hepatitis and primary biliary cholangitis. Only significant p values are shown. Lines and error bars indicate median and interquartile range.

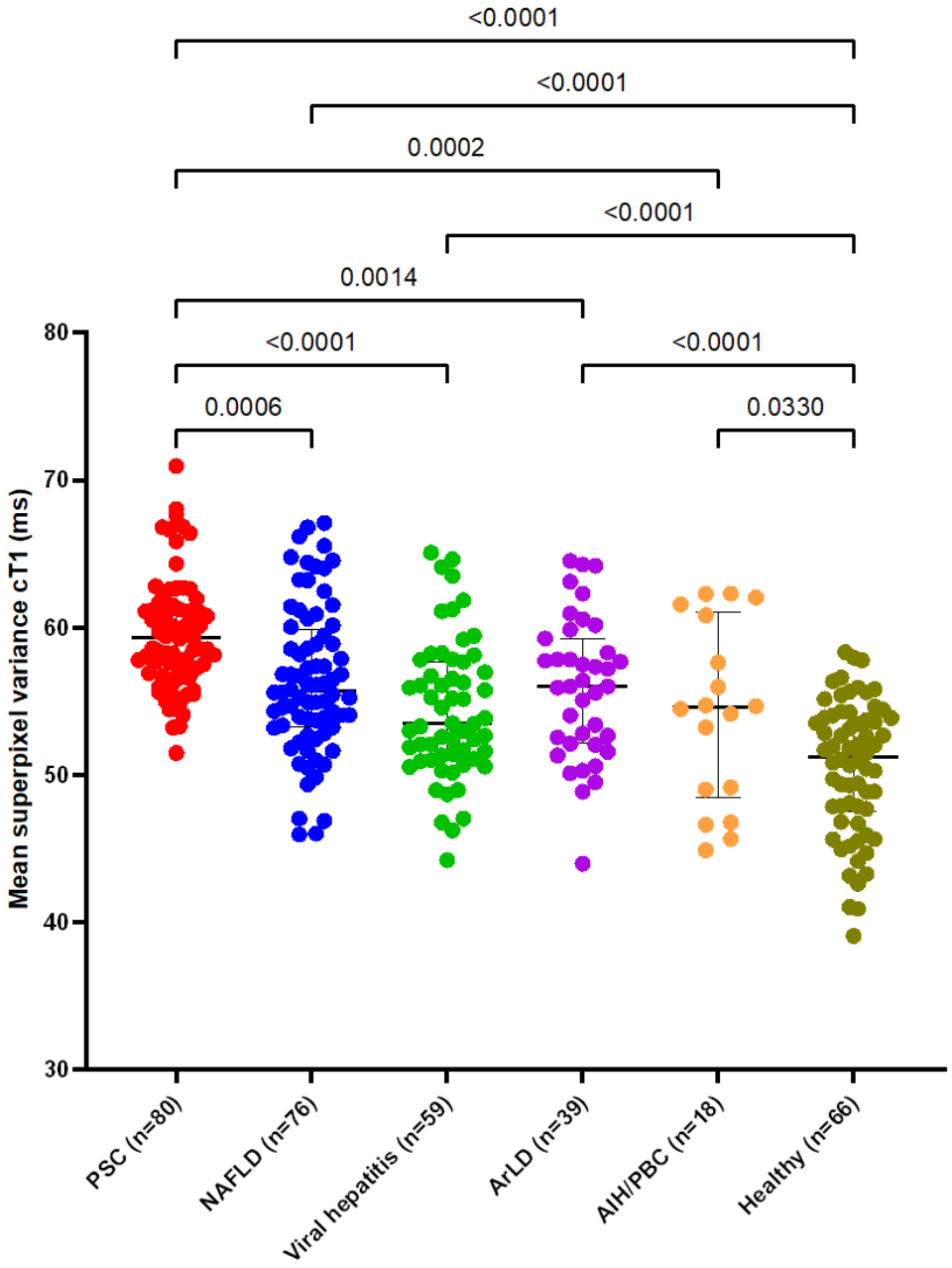


Figure 44. Mean superpixel variance cT1 in chronic liver diseases and healthy volunteers. ANOVA with Tukey's multiple comparisons tests demonstrated a significant difference between PSC and other liver diseases (PSC vs. NAFLD: p = 0.0006, PSC vs. viral hepatitis: p < 0.0001, PSC vs. ArLD: p = 0.0014, PSC vs. AIH/PBC: p = 0.0002) but no difference amongst the other liver diseases. PSC; primary sclerosing cholangitis, NAFLD; non-alcoholic fatty liver disease, ArLD; alcohol-related liver disease; AIH/PBC; autoimmune hepatitis and primary biliary cholangitis. Only significant p values are shown. Lines and error bars indicate median and interquartile range.

5.6.2.3 Liver cT1 metrics and severity of liver fibrosis

Liver cT1 metrics vs biopsy-proven fibrosis stage

In a post-hoc subgroup analysis of included participants, there were 116 paired MRI data and liver biopsy available for analysis from the RIAL and NICOLA studies. The breakdown of fibrosis stages was F0 (n = 7), F1–2 (n = 63), F3–4 (n = 27), and F5–6 (n = 19). Participants with no fibrosis (n = 7), PSC (n = 3), and other liver diseases (n = 4) were excluded, leaving 102 participant data for the final analysis. The breakdown of diseases was: 50 NAFLD, 34 viral hepatitis, 11 ArLD, and 7 PBC/AIH. Patients were grouped into F1–2 (n = 58), F3–4 (n = 26), and F5–6 (n = 18).

Although not statistically significant, fibrosis severity showed borderline positive correlation with IQR cT1 (p = 0.078) with numerically lower IQR cT1 in the F1–2 fibrosis stage (F1–2: 45.4ms; F3–4: 59.7ms; F5–6: 59.2ms, p = 0.0589; Figure 45). Mean superpixel variance cT1 did not correlate with Ishak fibrosis stages (p = 0.766) and showed no discriminatory value in fibrosis severity (F1–2: 54.6ms; F3–4: 55.5ms; F5–6: 54.2ms, p = 0.6748; Figure 45). However, as demonstrated in previous Liver*MultiScan* studies, there was a positive correlation between fibrosis stages and median cT1 (r_s = 0.49, p < 0.0001). Median cT1 was significantly higher in F5-6 compared to lower fibrosis stages (F1–2 vs F3–4: 836.9ms vs 883.8ms, p = 0.1883; F1–2 vs F5–6: 836.9ms vs 970.5ms, p < 0.0001; F3–4 vs F5–6: 883.8ms vs 970.5ms, p = 0.0228; Figure 46).

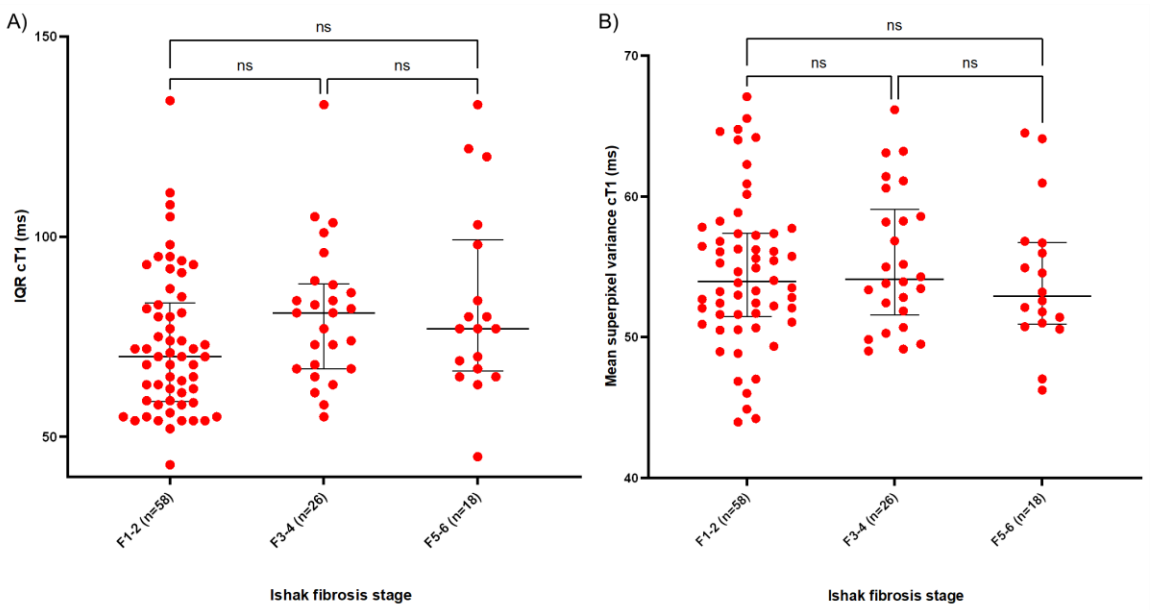


Figure 45. (A) IQR cT1 and (B) mean superpixel variance cT1 stratified by Ishak's fibrosis stages in 102 patients with mixed aetiology chronic liver diseases. ANOVA with Tukey's multiple comparisons test showed no significant differences between groups. ns, not significant. Lines and error bars indicate median and interquartile range.

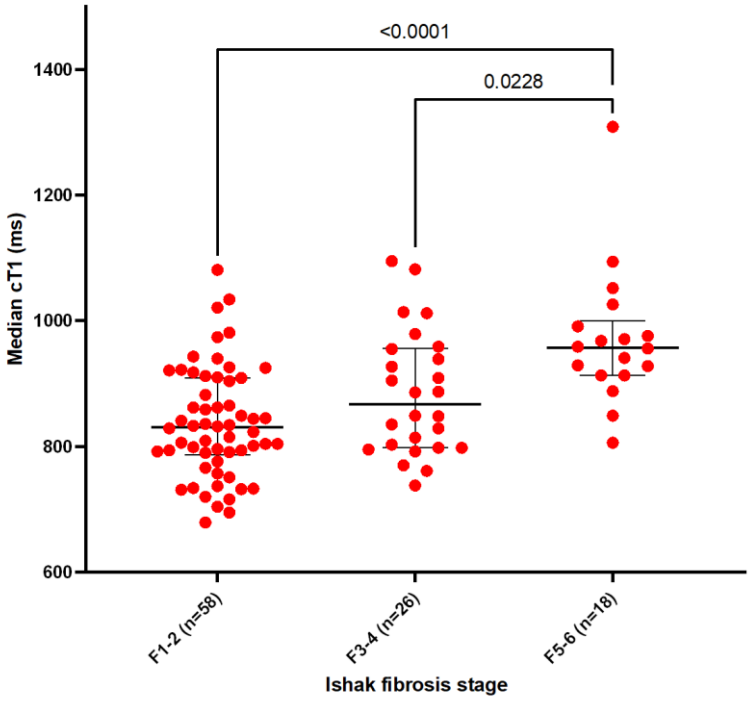


Figure 46. Median cT1 stratified by Ishak's fibrosis stages in 102 patients with mixed aetiology chronic liver diseases. ANOVA with Tukey's multiple comparisons test showed significant differences between F1–2 and F5–6 ($p < 0.0001$) groups, and F3–4 and F5–6 ($p = 0.0228$) groups. There was no difference between F1–2 and F3–4 ($p = 0.1883$) groups. Lines and error bars indicate median and interquartile range.

Liver cT1 metrics vs biopsy-proven fibrosis stage + clinical compensated cirrhosis

Given the low numbers in the F5–6 group (n = 18), data from the MAPLE studies were included to expand the F5–6 sample size. Liver biopsies were not performed in the MAPLE study but recruitment to the study involved patients with compensated (n = 67) and decompensated (n = 51) clinical cirrhosis. Given that RIAL and NICOLA studies excluded patients with decompensation at recruitment, only dataset from the compensated group were included in the F5–6 group to reduce heterogeneity between patient groups. In the compensated cirrhosis group, 8 participant data (3 other liver disease, 2 PSC, 2 poor quality MRI, and 1 diagnosis unknown) were excluded, leaving 59 for analysis. IQR cT1 and mean superpixel variance cT1 were compared between F1–2 (n = 58), F3–4 (n = 26), and F5–6 or clinical cirrhosis (n = 77). There was a significant difference in IQR cT1 between groups (F1–2: 74.8ms; F3–4: 80.6ms, F5–6: 90.3ms, $p < 0.0001$). In the multiple comparisons test, there was a significant difference between F1–2 vs F5–6 ($p < 0.0001$) and F3–4 vs F5–6 ($p = 0.0161$;Figure 47). There was no difference between F1–2 vs F3–4 ($p = 0.3294$). For comparison, median cT1 was only able to distinguish between F1–2 and F5–6 groups ($p = 0.0185$; Figure 48). There were no differences in median cT1 between F1–2 and F3–4 ($p = 0.1261$) or F3–4 and F5–6 ($p = 0.9977$) groups. Mean superpixel variance cT1 was no different between fibrosis groups (F1–2: 54.0ms; F3–4: 55.6ms; F5–6: 56.0ms, $p = 0.1041$; Figure 49).

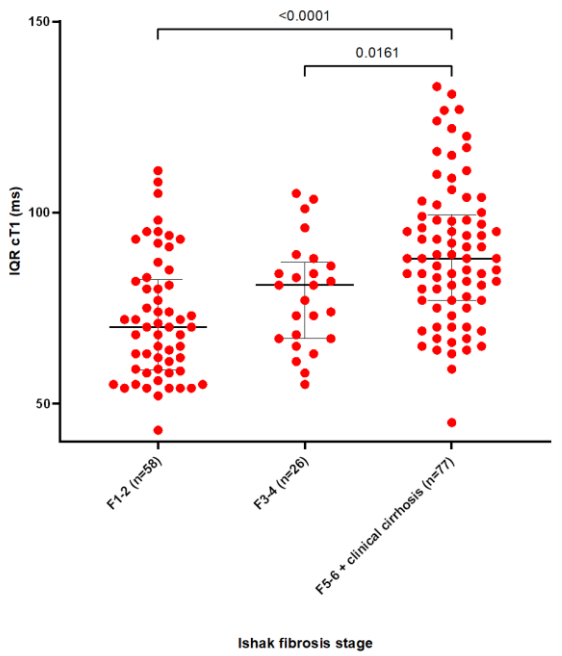


Figure 47. IQR cT1 stratified by Ishak's fibrosis stages in 161 patients with mixed aetiology chronic liver diseases. ANOVA with Tukey's multiple comparisons test showed significant differences in IQR cT1 between F1–2 and F5–6 groups ($p < 0.0001$), and between F3–4 and F5–6 groups ($p = 0.0161$). There was no difference between F1–2 and F3–4 ($p = 0.3294$) groups. Only statistically significant p-values are shown. Lines and error bars indicate median and interquartile range.

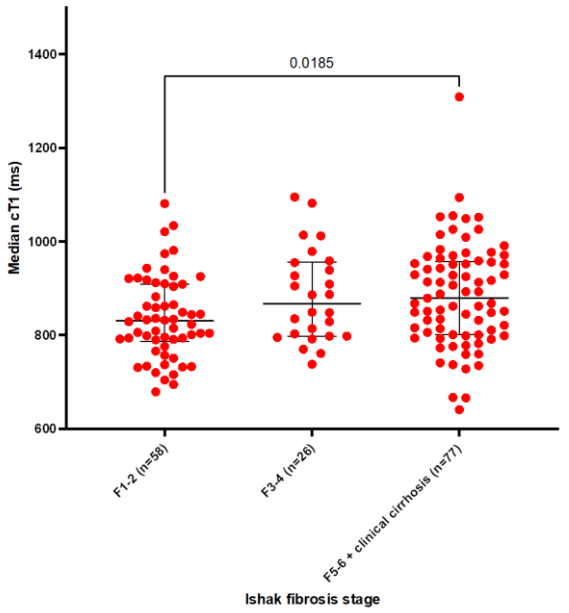


Figure 48. Median cT1 stratified by Ishak's fibrosis stages in 161 patients with mixed aetiology chronic liver diseases. ANOVA with Tukey's multiple comparisons test showed significant difference between F1–2 and F5–6 groups ($p = 0.0185$). There were no differences between F1–2 and F3–4 ($p = 0.1261$) or F3–4 and F5–6 ($p = 0.9977$) groups. Only statistically significant p-values are shown. Lines and error bars indicate median and interquartile range.

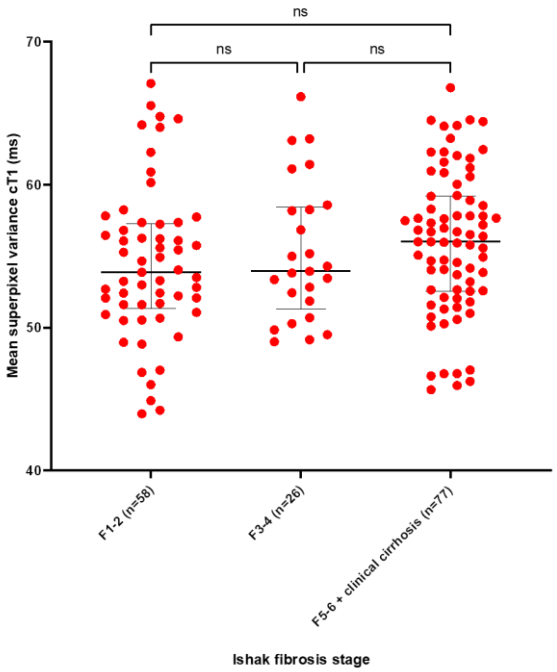


Figure 49. Mean superpixel variance cT1 stratified by Ishak's fibrosis stages in 161 patients with mixed aetiology chronic liver diseases. ANOVA with Tukey's multiple comparisons test showed no significant differences between groups. ns, not significant. Lines and error bars indicate median and interquartile range.

Liver cT1 metrics clinical compensated vs decompensated cirrhosis

In the decompensated group of MAPLE study, 6 participant data were excluded (3 other liver disease, 2 PSC, and 1 diagnosis unknown), leaving 45 for analysis. IQR cT1 and superpixel mean cT1 of standard deviations were compared between the compensated (n = 59) and decompensated (n = 45) groups. IQR cT1 (107.2ms vs 91.5ms, p = 0.0031) was higher in the decompensated group compared to the compensated group. Likewise mean superpixel variance cT1 was also higher (59.3ms vs 56.5, p = 0.0114) in the decompensated group, albeit less significantly than IQR cT1 (Figure 50).

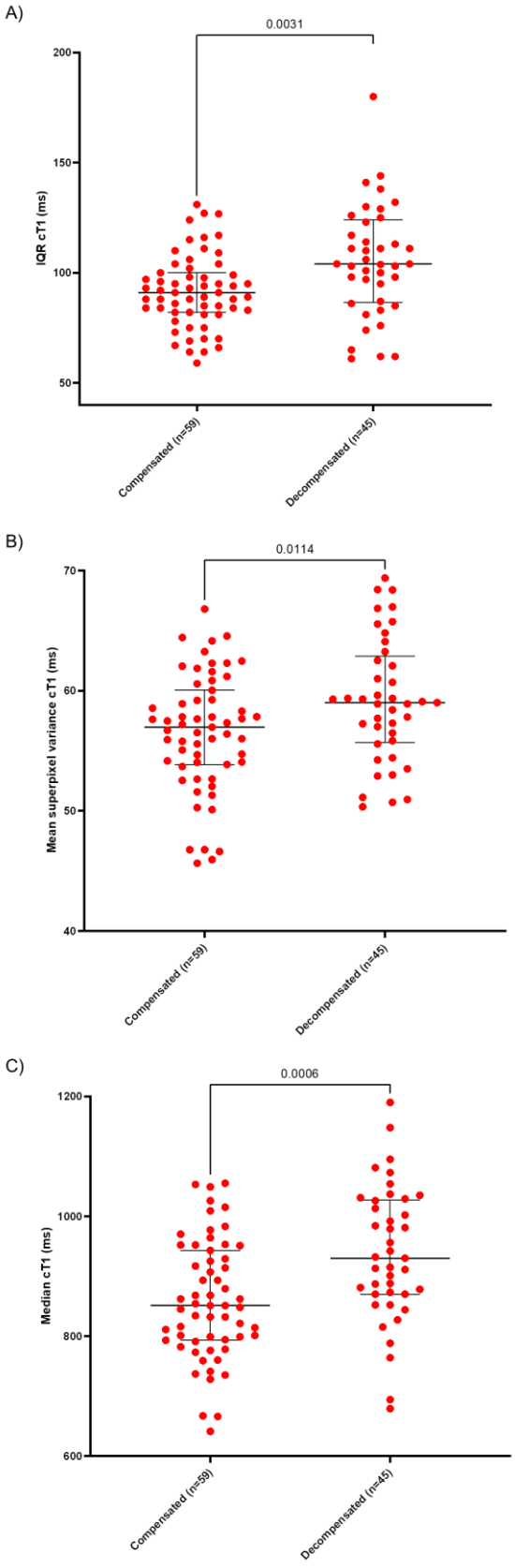


Figure 50. Difference in liver cT1 metrics between compensated and decompensated cirrhosis groups in the MAPLE study. (A) IQR cT1, (B) mean superpixel variance cT1, and (C) median cT1. Lines and error bars indicate median and interquartile range.

5.7 Discussion

This work explored the role of IQR cT1 and mean superpixel variance cT1 as surrogate markers of liver parenchymal heterogeneity in PSC and compared it to the performance of median cT1, an established surrogate marker of liver fibrosis from previous studies involving homogeneous liver diseases. IQR cT1 was a marker of disease severity in PSC, correlated with LS and ELF in PSC, as well as histologically-assessed liver fibrosis in other liver diseases. Nonetheless, it was unable to distinguish PSC from other liver diseases. Mean superpixel variance cT1, on the other hand, was also higher in more severe PSC and correlated with LS and ELF in PSC. However, it was independent of histologically assessed liver fibrosis in other liver diseases and distinguished PSC from other liver diseases. Median cT1 remained the best metric to assess severity of liver fibrosis in other liver diseases but poorly correlated with LS and did not correlate with ELF in PSC. Findings from this work suggest that IQR cT1 may not truly reflect liver parenchymal heterogeneity in PSC, but a surrogate marker that is associated with disease severity and a better marker of liver fibrosis than median cT1 in PSC. Mean superpixel variance cT1 appears to be unique to PSC, perhaps a better reflection of liver parenchymal heterogeneity in PSC than IQR cT1 compared to other liver diseases.

Previous studies that explored the use of liver cT1 found that measures of histogram data centrality such as mean and median cT1 had good correlation with histologically assessed liver fibrosis and predicted clinical outcomes[127, 164]. These studies predominantly included homogeneous liver diseases such as NAFLD, viral hepatitis and ArLD, therefore one would expect quantitative metric that measured the centrality of cT1 distribution was a fair representation of the true state of the liver parenchyma. The performance of other histogram metrics was not reported in these studies. The liver parenchyma in PSC, however, is thought to be heterogeneous in nature, particularly in the early stages of the diseases where normal liver parenchyma could be present among diseased areas. A measure of centrality of distribution of cT1 in liver segmentations could falsely

underestimate the degree of patchiness and variation in cT1 distribution, owing to the dilutional effect of averaging out the cT1. The work performed in this chapter has attempted to discover new metrics of liver cT1 that could be used to reflect the possible variability in cT1 distribution within the liver parenchyma that may be expected in PSC.

Identification of candidate metrics

The first step in assessing novel metrics is to ensure that they can sufficiently discriminate a heterogeneously diseased liver from healthy liver. Although skewness and kurtosis can be used to assess heterogeneity of a distribution, both were unable to sufficiently discriminate PSC from healthy controls, therefore were not considered further. Median cT1, a measure of centrality of distribution, and IQR cT1, a measure of dispersion of data in a distribution, were higher in PSC than healthy controls in the standard Liver*MultiScan* analysis. In the superpixel analysis, mean superpixel variance cT1, a measure of regional variance in cT1 distribution, was significantly higher in PSC than healthy. IQR cT1 and mean superpixel variance cT1 had the best performance with AUC > 0.80 to distinguish PSC from healthy compared to other metrics.

Association between candidate cT1 metrics and non-invasive markers

Median cT1 did not correlate with any markers of disease severity or progression in PSC and had poor risk-stratification ability to identify the high-risk PSC group. The skewed distribution of cT1 in PSC may have led to an averaging effect. On the contrary, IQR cT1 and mean superpixel variance cT1, had stronger positive correlations with markers of disease severity in PSC and were better classifiers of high-risk and low-risk groups.

Given that cT1 quantifies inflammation and fibrosis in the liver, metrics that measure the dispersion of cT1 can be extrapolated as a measure of the heterogeneous distribution of inflammation and/or fibrosis within the liver. IQR cT1 and mean superpixel variance cT1 showed stronger correlations with liver fibrosis markers (LS, ELF, FIB-4 and APRI) than serum markers of liver inflammation (ALT, AST). Whilst liver enzymes are not necessarily

a good indicator of liver inflammation, this may suggest that the main contributing factor to heterogeneity in the liver parenchyma in PSC is patchy accumulation of liver fibrosis rather than inflammation.

Association between candidate cT1 metrics and morphological features on MRI-MRCP

Dysmorphy of the liver is usually seen in the advanced stages of liver disease and is characterised by lobulation of the liver contours, atrophy of segments where the biliary disease predominates, and hypertrophy of relatively spared areas[28, 165]. The inflammation and cholestatic changes in PSC tend to spare the central segments of the liver including the caudate lobe[87, 102]. Therefore, caudate lobe compensatory hypertrophy, along with hypertrophy of the left lobe and atrophy of the right lobe, has been shown to be more frequent in PSC than in cirrhosis from other causes[154, 165]. These distortions to the parenchyma give rise to the appearance of heterogeneity in the liver commonly seen on cross-sectional imaging and showed excellent inter-observer agreement between the two radiologists in this study.

Both IQR cT1 and mean superpixel variance cT1 were higher in patients with hepatic dysmorphy than those without when assessed by either radiologist. Both metrics were also elevated in patients with portal hypertension changes on MRI compared to those who did not. Liver congestion from portal hypertension may have contributed towards elevated cT1 measurements. IQR cT1 showed a better discriminating performance than mean superpixel variance cT1 for both MRI features. Median cT1 was unable to discriminate between patients with and without hepatic dysmorphy or portal hypertension. The results for identifying cirrhosis were mixed between the two radiologists. This prompted comparison with clinical diagnosis of cirrhosis from review of clinical notes and imaging which confirmed that median cT1 had good diagnostic performance to identify liver cirrhosis as shown in previous studies.

IQR cT1 was higher in patients with TSSS > 4 than those ≤ 4 and showed good discriminatory performance between the two groups, but no different to distinguish between other MRCP features. None of the other liver metrics showed significant risk stratification ability, suggesting that liver heterogeneity and fibrosis may be independent to biliary stricture severity and presence of dominant strictures.

Risk stratification

In the absence of a reliable gold standard reference test and the relatively short follow-up for validation against clinical outcomes, the performance of liver cT1 metrics was assessed against published thresholds of known surrogate markers of disease severity and progression in PSC. IQR cT1 performed better than mean superpixel variance cT1 and median cT1 to classify the high-risk and low-risk groups when stratified by liver fibrosis markers (LS and ELF), AOM, Anali score, modified Amsterdam cholangiographic score, and the three morphological features on MRI — dysmorphy, cirrhosis and portal hypertension. This data suggests that indicators of severity in liver fibrosis or cirrhosis, liver function, and severity of biliary disease contribute to a higher measure in parenchymal heterogeneity. Serum ALP, a marker of cholestasis from bile duct injury, can fluctuate over the course of the disease and can vary within and among patients with PSC[60]. Whilst it has been shown to have prognostic implications in several studies, data presented in this cohort of patients with large-duct PSC suggest that it may not be an accurate reflection of liver parenchyma heterogeneity. In the multivariate analysis, neither of the candidate liver cT1 metrics were independent predictors of high-risk PSC when defined by AOM and MRS. Ultimately, clinical endpoints in PSC are required to compare and validate the performance of all non-invasive markers.

Validation of candidate liver cT1 metrics

Increasing heterogeneity in the liver parenchyma is a well-known phenomenon associated with advanced fibrosis and cirrhosis in chronic liver disease[166, 167]. Therefore, the validation work was important to investigate whether the heterogeneity quantified by IQR cT1 and mean superpixel variance cT1 in patients with PSC was simply a reflection of the severity of liver fibrosis seen in progressive chronic liver disease of any aetiology rather than a unique measure of parenchymal heterogeneity in PSC.

IQR cT1 was higher in all liver disease controls compared to healthy controls. However, IQR cT1 in PSC from the present study was not significantly different compared to other liver diseases, challenging the premise that this metric uniquely measures parenchymal heterogeneity seen in PSC. One explanation for this is that the inclusion of a higher proportion of patients with advanced fibrosis and cirrhosis in the validation cohort, as reflected by the higher medians and interquartile ranges of liver stiffness and ELF, may have skewed the analysis.

When only participants who had paired liver biopsy and MRI data from the RIAL and NICOLA studies were included, IQR cT1 was no different between the fibrosis groups, but median cT1 was higher in the higher fibrosis groups as expected. Whilst this showed some evidence that IQR cT1 metric may not be related to severity of fibrosis, the results may have been skewed by proportional bias from significantly smaller numbers in the higher fibrosis groups. To overcome this bias, participants with compensated cirrhosis from the MAPLE study was added to the F5–6 group and data re-analysed. IQR cT1 was now observed to be higher in F5–6 compared to F1–2 groups, suggesting that greater liver heterogeneity is observed in bridging fibrosis and cirrhosis compared to mild to moderate fibrosis stages. There was no difference compared to the intermediate stages of fibrosis (F3–4), which had a smaller group size. This work supports what is known clinically and histologically about the evolution of the liver parenchyma becoming more heterogenous with worsening fibrosis in chronic liver diseases. Finally, when participants

with clinical cirrhosis in the MAPLE study were analysed separately, IQR cT1 was higher in the decompensated than the compensated cirrhosis group, suggesting that IQR cT1 may also reflect the severity of end-stage liver disease just as median cT1 does. This data proposes an alternative explanation for the strong correlation observed between IQR cT1 and surrogate markers of liver fibrosis and portal hypertension, in that IQR cT1 is simply a better metric than median cT1 to assess the severity of liver fibrosis in the parenchyma of patients with PSC. This difference in choice of metric becomes less significant with increasing severity of end-stage liver disease. Nevertheless, whilst may not be useful to distinguish PSC from other liver diseases, IQR cT1 did show promising risk stratification ability in patients with PSC and needs to be validated against clinical endpoints in PSC and tested in prospective multicentre studies before it can be considered as a potential imaging biomarker.

Mean superpixel variance cT1 was not as good as IQR cT1 in risk-stratifying patients with PSC, but unlike IQR cT1, was able to differentiate PSC from every other liver disease in the validation cohort. There was no difference in this metric between the other liver diseases and it was independent of the severity of liver fibrosis in subgroup analysis involving the two studies with paired biopsies and when data from compensated cirrhosis was included in the F5–6 group. Mean superpixel variance cT1 could be uniquely quantifying the heterogeneity in cT1 distribution expected to be present in PSC but not in other liver diseases. This needs to be investigated further in prospective controlled studies.

Patient reported outcomes

SF-36 mean domain scores showed significant correlation with median cT1 and liver stiffness measurements with lower mean scores in patients with advanced fibrosis defined by liver stiffness. It can be inferred that participants with more severe liver parenchymal disease reported lower energy levels and lower rating of their general health. Future

antifibrotic drug trials in PSC should perhaps consider assessing patient-reported outcomes as one of the secondary endpoints.

Despite having similar physical and mental component themes, the domains from the PSC-PRO did not correlate with any of the MRI metrics. One explanation for this could be due to the reporting of chronicity of the components that make up each domain in the questionnaires. SF-36 asked to report outcomes for the preceding 4 weeks whilst PSC-PRO was attempting to quantify outcomes for the last 7 days only.

One of the limitations of the work carried out in this chapter was that a gold standard comparator was not available for quantifying parenchymal fibrosis and heterogeneity. Liver biopsy can underestimate fibrosis stage but correlation with histological changes may help validate findings. Assessment of heterogeneity on MRE stiffness map could also help establish the relationship between the two MR techniques. Whilst automated Liver*MultiScan* analysis provides a standardised and time-efficient analysis of a large heterogeneous dataset, the selection of a single representative liver slice may have underestimated the measure of heterogeneity in the liver. However, there was sufficient evidence to suggest good correlation between semi-automated multi-slice method and automated single-slice from the validation of methods work carried out in chapter 3. Lastly, confounders to cT1 measurements other than liver iron have not been assessed. Liver fat is a confounder to cT1 and was not adjusted for in the current algorithm. It also remains to be established more definitely whether heterogeneity measures described in this chapter was assessing inflammation rather than fibrosis.

In conclusion, work carried out in this chapter have identified two liver cT1 metrics that can potentially assess heterogeneity in the liver parenchyma of patients with PSC and showed promising risk stratification ability.

The next chapter will assess cT1 specifically around the peribiliary region in the liver parenchyma by using a combination of *MRCP+* and Liver*MultiScan* techniques.

6 Assessment of periductal abnormalities

6.1 Abstract

Background: Changes in the liver parenchyma adjacent to the bile ducts in PSC are described as periductal fibrosis on liver biopsy and peribiliary enhancement on contrast-enhanced MRI. The aim of this study was to evaluate a quantitative method to assess periductal regions macroscopically using data from two quantitative MRI scans.

Methods: Biliary data were analysed using *MRCP+* and liver data using Liver*MultiScan*. Using MRI acquisition coordinates, the 3D *MRCP+* model was co-registered with the four 2D liver cT1 maps from Liver*MultiScan*. This allowed concentric periductal cT1 to be quantified around the bile ducts. Regions of interest (ROI) 1–4 were defined as radius 1.5–3.5mm, 3.5–5.5mm, 5.5–7.5mm, 7.5–9.5mm, and ROI 5 the whole liver. The angle of intersection of bile ducts on liver were calculated and used to correct for partial volume effect on periductal cT1 measurement. Advanced fibrosis was defined as transient elastography liver stiffness (LS) > 9.6kPa.

Results: Data from 73 patients with PSC and 31 healthy controls were included. In patients with PSC, the periductal mean cT1 in ROI 1 was higher than other regions (ROI 1: 795ms; ROI 2: 780ms; ROI 3: 772ms; ROI 4: 769ms; and ROI 5: 774ms; $p < 0.0001$), even after correcting for partial volume effect of bile duct intersection. There were no differences between ROIs greater than 5.5mm radius and in healthy controls. The periductal mean cT1 in the region 1.5–3.5mm surrounding the bile ducts were higher in patients with LS > 9.6kPa compared to patients with LS ≤ 9.6kPa (833ms vs. 781ms; $p = 0.0021$). Periductal mean cT1 cut-off of 818.1ms had an AUC of 0.74 (95% CI 0.60–0.88), sensitivity 63% and specificity 83% to detect the high-risk group with advanced fibrosis.

Conclusion: Periductal mean cT1 in regions closer to the bile ducts were observed to be higher than regions further away in PSC, but not in healthy controls. Patients with advanced fibrosis showed higher periductal mean cT1 values.

6.2 Introduction

Multifocal chronic inflammation and fibrosis of bile ducts with involvement of surrounding periductal liver parenchyma is the hallmark of disease pathology in PSC. Inflammation and fibrosis of bile ducts cause concentric periductal changes involving typically the medium sized bile ducts, giving an appearance of 'onion skin' on liver histology as shown in Figure 51. Varying extent of periductal and peri-portal inflammation and fibrosis are observed causing progressive atrophy of bile duct epithelium, expansion of the peribiliary stroma, obstruction of bile duct lumen with cholestasis, and eventually leaving a fibro-obliterative scar[145]. Periportal oedema and fibrosis may be seen from obstruction of larger ducts. Biliary fibrosis originates in the portal tract and tends to expand with a portal-to-portal pattern, extending into lobular areas that eventually leads to development of biliary-type cirrhosis and end-stage liver disease[168]. As discussed in chapter 5, monitoring disease progression histologically is no longer recommended given its invasive nature, increased risk of cholangitis in PSC, susceptibility to sampling error in a patchy disease, and inter-observer variability in histological interpretation.

Changes in the adjacent liver parenchyma surrounding the bile ducts in PSC can be visualised on extracellular gadolinium-enhanced MRI and have demonstrated to correlate with survival in two studies. The first study was a small retrospective study involving 47 patients with PSC[88]. Peribiliary hyperenhancement was defined as linear hyperenhancement within the hepatic parenchyma adjacent to the bile ducts including the bile duct wall. Peribiliary hyperenhancement was classified as none, minimal (<2 mm margin of enhancement surrounding the bile duct), moderate (2-6mm margin of enhancement surrounding the bile ducts) or marked (>6mm margin of enhancement surrounding the bile ducts). Peribiliary hyperenhancement was observed in 38 patients (81%). Nine had none, 19 minimal, 13 moderate, 6 marked peribiliary hyperenhancement.

Delayed venous phase (3 minutes post-contrast) peribiliary hyperenhancement showed weak correlation with MRS.

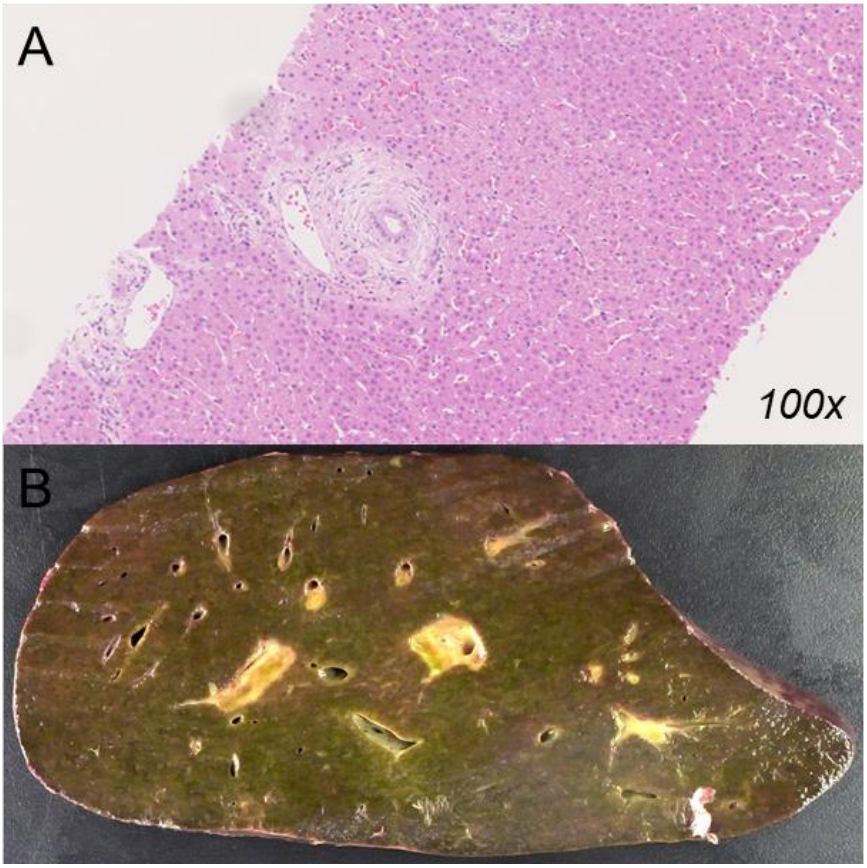


Figure 51. Histopathology of periductal fibrosis. (A) Liver biopsy sample showing periductal 'onion skin' fibrosis surrounding a bile duct at 100x magnification in PSC. (B) A slice of explanted liver with fibrosis around large bile ducts in PSC.

Image A courtesy of Dr Eve Fryer, Consultant Histopathologist at Oxford University Hospitals NHS Foundation Trust. Image B reproduced with permission from Dr Raul Gonzales and PathologyOutlines Gonzalez RS. Primary sclerosing cholangitis. PathologyOutlines.com website. https://www.pathologyoutlines.com/topic/liverpsc.html. Accessed March 11th, 2021.

The second study was also a retrospective study that investigated peribiliary hyperenhancement with the same definition but concentrated on the effect of early or arterial phase peribiliary hyperenhancement instead of delayed or venous phase in 62 patients with PSC[90]. Peribiliary hyperenhancement on any imaging phase post contrast was seen in 46 (74%) patients and specifically arterial phase hyperenhancement in 29

(47%) patients. The study showed that patients with MRS > 0 (intermediate and high-risk) had significantly increased prevalence of peribiliary hyperenhancement in arterial, portal, and venous phases compared to patients with MRS < 0. Patients with arterial peribiliary hyperenhancement had shorter survival than those who did not, however, it was not an independent predictor of survival in the multivariate model.

In another retrospective study of 55 patients with large-duct PSC, segmental ductal strictures on MRCP were associated with arterial phase hyperenhancement and increased segmental parenchymal liver stiffness measured on MRE[95]. All three studies have independently demonstrated a similar phenomenon surrounding the bile ducts in PSC, but all require contrast administration and rely on the qualitative interpretation of data, so are prone to inter/intra-observer variation. Contrast administration and accurately timed MRI acquisition is a pre-requisite for this technique. Manual measurements of the margin of signal enhancement surrounding the bile ducts will inevitably further compound the inter-observer variability in MRI interpretation. Therefore, development of a standardised quantitative MRI method that can objectively quantify these periductal changes may be able to overcome these limitations in clinical practice.

6.3 Hypotheses

Periductal inflammation and fibrosis are the hallmark of disease in PSC. Liver inflammation and fibrosis causes an increase in T1 relaxation time as previously discussed. cT1 can be measured on non-contrast MRI sequences analysed using the Liver*MultiScan* software at each liver acquisition voxel to generate a cT1 map of the liver. A 3D model of the biliary tree can be derived from *MRCP+* software for staging the cholangiopathy in PSC. Co-registration of the 3D biliary tree model to the axial liver cT1 maps may allow targeted exploration of the periductal region. The hypotheses for this chapter were that:

- patients with more active form of PSC with on-going inflammation and periductal fibrosis are likely to have higher periductal cT1 values
- patients with a greater extent of peribiliary involvement of the liver parenchyma are likely to have a more severe disease phenotype, therefore more likely to have higher periductal cT1 values compared to those with milder disease phenotype
- periductal cT1 is likely to reflect increased heterogeneity in the liver parenchyma as discussed in chapter 4 but maybe independent of bile duct dilatation severity as discussed in chapter 3

6.4 Aims

The aims of this study chapter were to:

- investigate whether a feature similar in concept to concentric periductal fibrosis can be diagnosed macroscopically on larger bile ducts using a combination of quantitative MRI scans.
- explore the relationship between quantitative periductal metrics and other non-invasive markers in PSC.
- assess whether quantitative periductal metrics can be used to risk stratify severity of disease in patients with PSC.
- examine the association between periductal metrics and QoL data.

6.5 Methods

Patient recruitment, study assessments, risk stratification, and image analysis followed the methods described in chapter 2.

6.5.1 Co-registration of axial liver cT1 maps and biliary tree model

Both liver MRI and 3D MRCP sequences were obtained on the patient consecutively in the same setting without moving the patient out from the MRI scanner. Using the MRI acquisition coordinates, *MRCP+* models were aligned with the four axial Liver*MultiScan* cT1 maps as shown in Figure 52. Only MRI and MRCP data that were acquired at the same time were included to ensure accuracy of co-registration using MRI coordinates. Patients were excluded if there was gross movement artifact on either of the sequences.

6.5.2 Periductal cT1 quantification in 3D space

The biliary tree 3D binary mask obtained from *MRCP+* analysis was first cleaned to tightly follow the *MRCP+* biliary tree model to ensure that voxels outside the biliary tree were excluded. Given the difference in resolution between 3D MRCP and T1 mapping, the corrected 3D biliary tree mask was resampled onto 2D liver cT1 slices obtained from Liver*MultiScan* analysis using standard linear interpolation. Periductal distance map was obtained by calculating 3D distance transform (Figure 53). Using the distance map, periductal cT1 was quantified over fixed radial distances surrounding the bile ducts measured from the centre line of the ducts (Figure 54). Region of interest (ROI) 1 was defined as the ring-shaped area between the circles with radius 1.5 and 3.5mm, ROI 2 radius 3.5 and 5.5mm, ROI 3 radius 5.5 and 7.5mm, and ROI 4 radius 7.5 and 9.5mm. ROI 5 was defined as the whole liver mean cT1 and was the reference measurement point. The quality criteria for inclusion of cT1 quantification was a coefficient of determination (R^2) value of > 98% within the segmented liver portion.

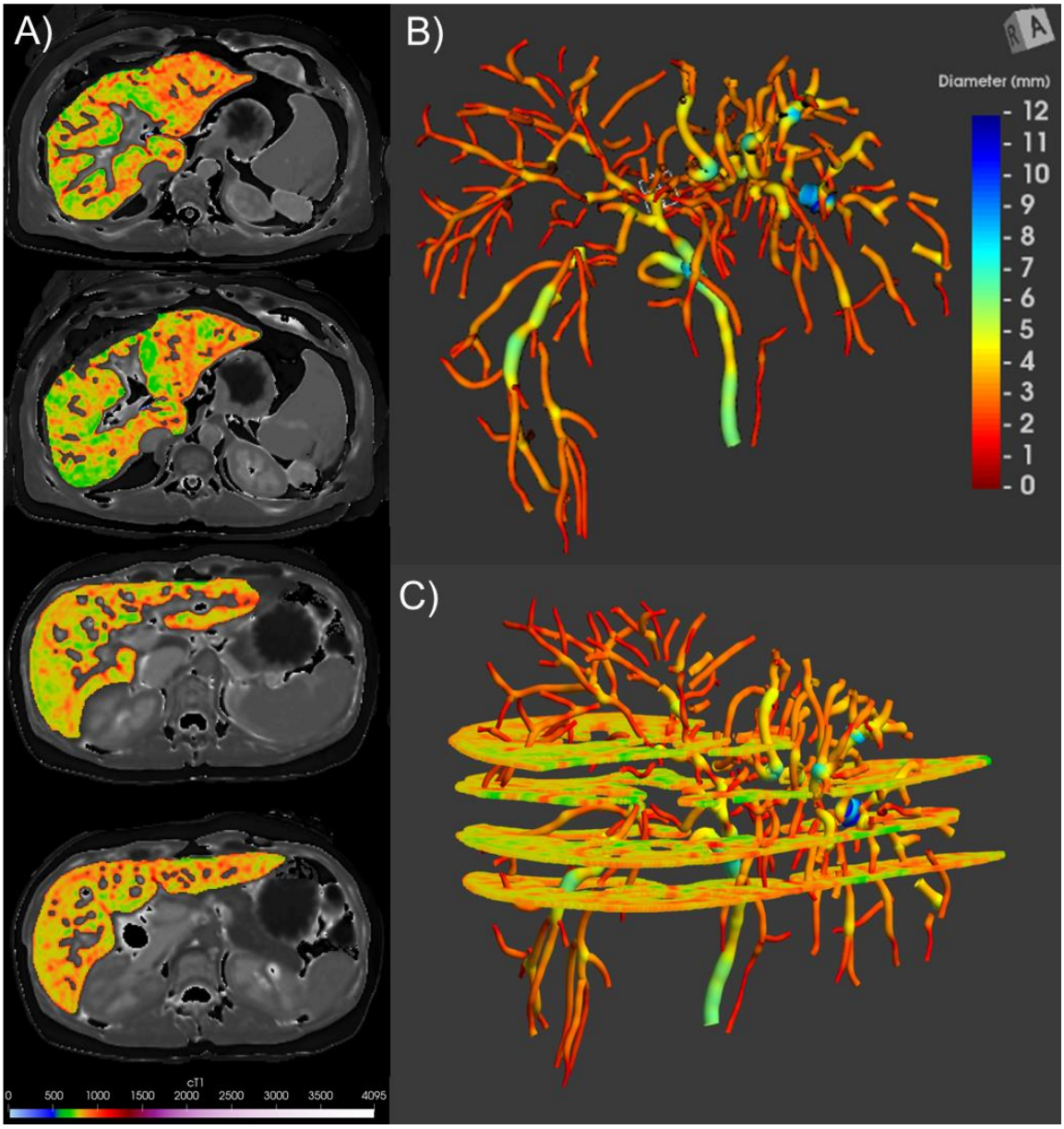


Figure 52. Co-registration of two quantitative MRI techniques for periductal cT1 quantification in a 33-year old lady with large-duct PSC. (A) Four segmented liver axial cT1 maps from Liver*MultiScan* analysis acquired at different levels with prominent blood vessels excluded and liver parenchyma colour-coded to liver cT1 values. (B) 3D biliary tree model derived from *MRCP+* analysis, colour-coded according to duct diameter size and orientation in geometric space depicted in the top right corner. (C) Co-registration of both images using MRI acquisition coordinates for the purpose of periductal cT1 quantification.

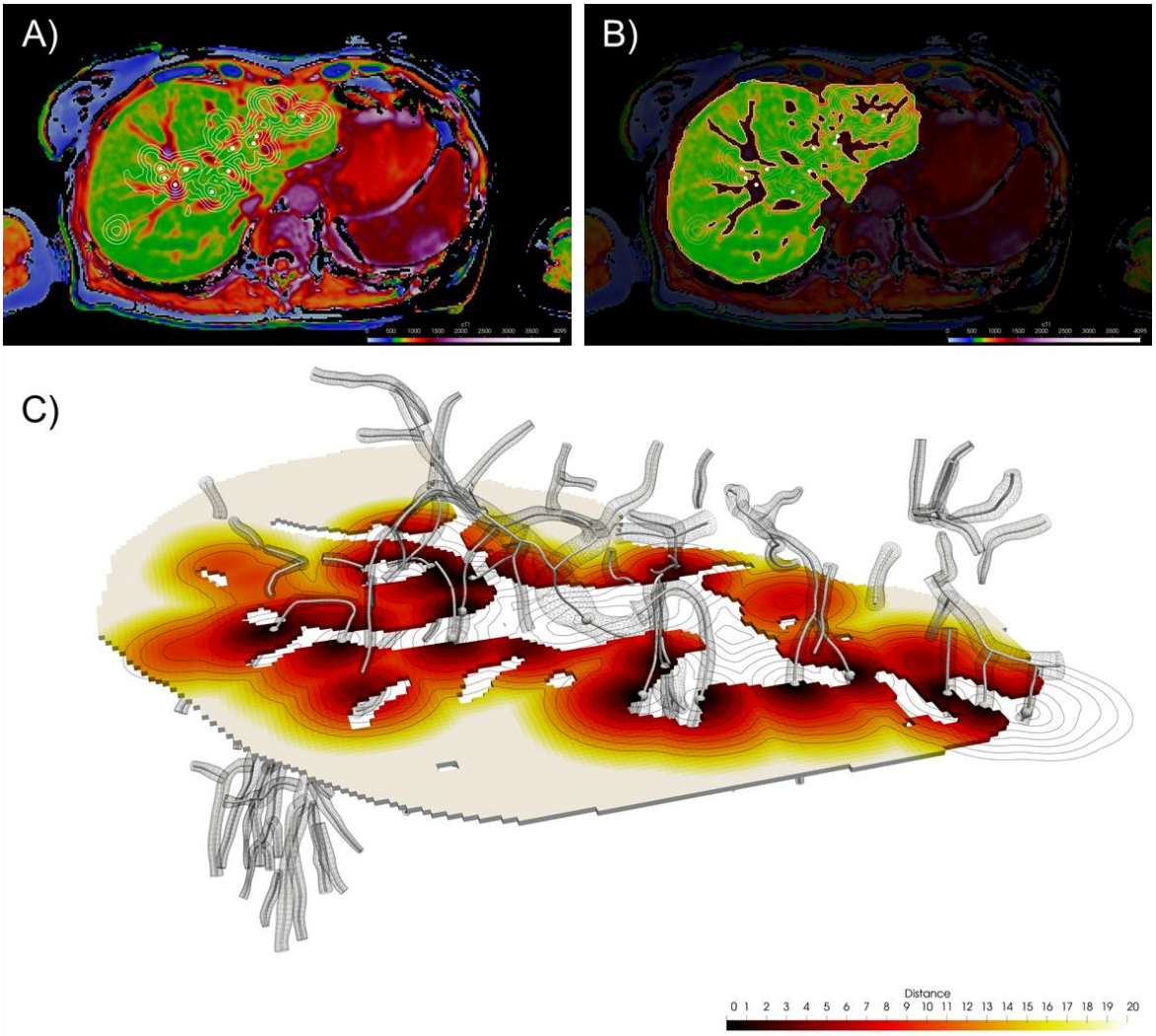


Figure 53. Periductal distance map from distance transform. 2D representation of co-registration of bile ducts (white dots) and distance map (white contour lines) A) without, and B) with an overlaying liver mask to exclude major blood vessels. White contour lines devoid of a white dot represents areas of adjacent bile duct in 3D space not represented in the 2D plane. C) 3D representation of distance transform around bile ducts. Wire frame represents bile duct centre line, small spheres represent points of bile duct intersection, and the colour map represents periductal distance ranging from 0 to 20mm.

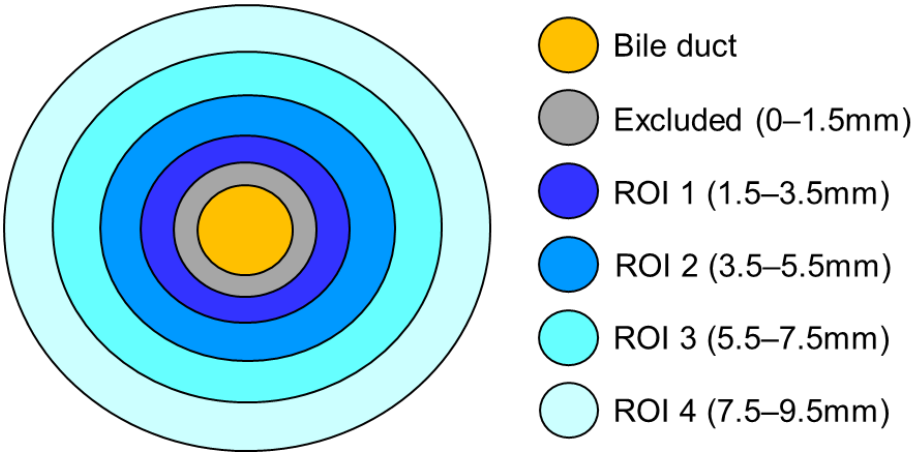


Figure 54. Schematic diagram representing the boundaries of the five regions of interest for periductal cT1 quantification. ROI; region of interest

6.5.3 Partial volume effect

In MRI, partial volume effects occur when the components contributing to signal intensity come from more than one source. In periductal cT1 quantification, partial volume effects could potentially arise at voxels of bile duct boundary, which may contain signals from within the bile duct, accompanying smaller portal blood vessel, and surrounding liver parenchyma at the level of intersection (Figure 55). Given that the MRCP acquisition voxel size was 1.1 x 1.1 x 1.1mm, periductal cT1 measurements below 1.5mm distance from the bile duct wall were excluded to minimise partial volume effect.

Oblique bile duct orientation with respect to the axial liver cT1 map can also give rise to partial volume effect. Figure 56 demonstrates this problem. Partial volume effect increases with increasing angulation from slice normal with the bile duct profile moving from cylindrical to a more ellipsoid profile. To be able to address this partial volume effect in regions where bile ducts traverse through 2D image slices at an oblique angle, angles between slice normals and bile duct segments at intersection points were calculated, where slice normals were taken to have inferior-superior direction. For the analysis, bile ducts were divided between those lesser than 30° angle, between 30–60° angle, and between 60–90° angle. Regions of interests were expanded and re-defined as ROI 1 (1.5–5.5mm), ROI 2 (5.5–9.5mm), and ROI 3 (whole liver) to avoid missing values in

certain periductal distance groups that would restrict pairwise comparisons to be performed for the cohort.

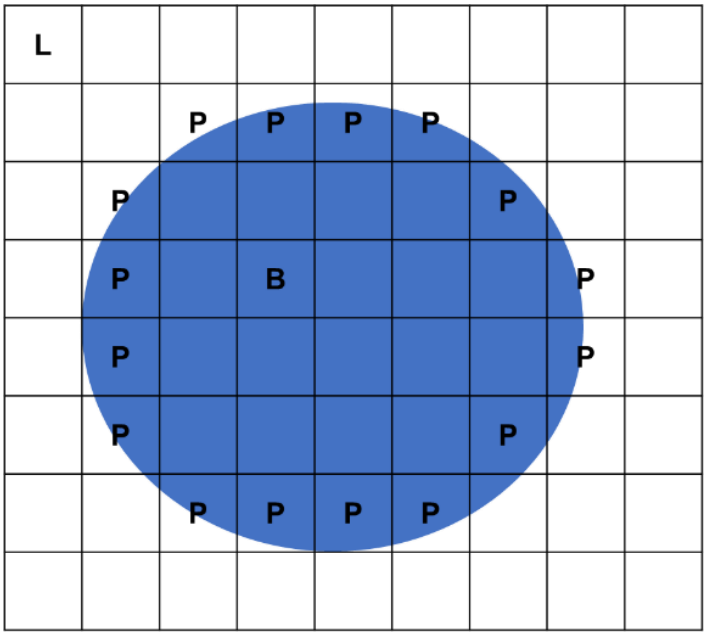


Figure 55. A schematic diagram representing a cross section of a bile duct intersecting with an axial segmented liver cT1 slice. Voxels labelled P are affected by partial volume effects and are located at the bile duct boundaries. The signal in these voxels is a weighted average of the signal originating from bile and from the liver parenchyma. Intraductal voxels (such as B) only contain signal originating from bile, while liver voxels (such as L) only contain signal originating from the liver parenchyma.

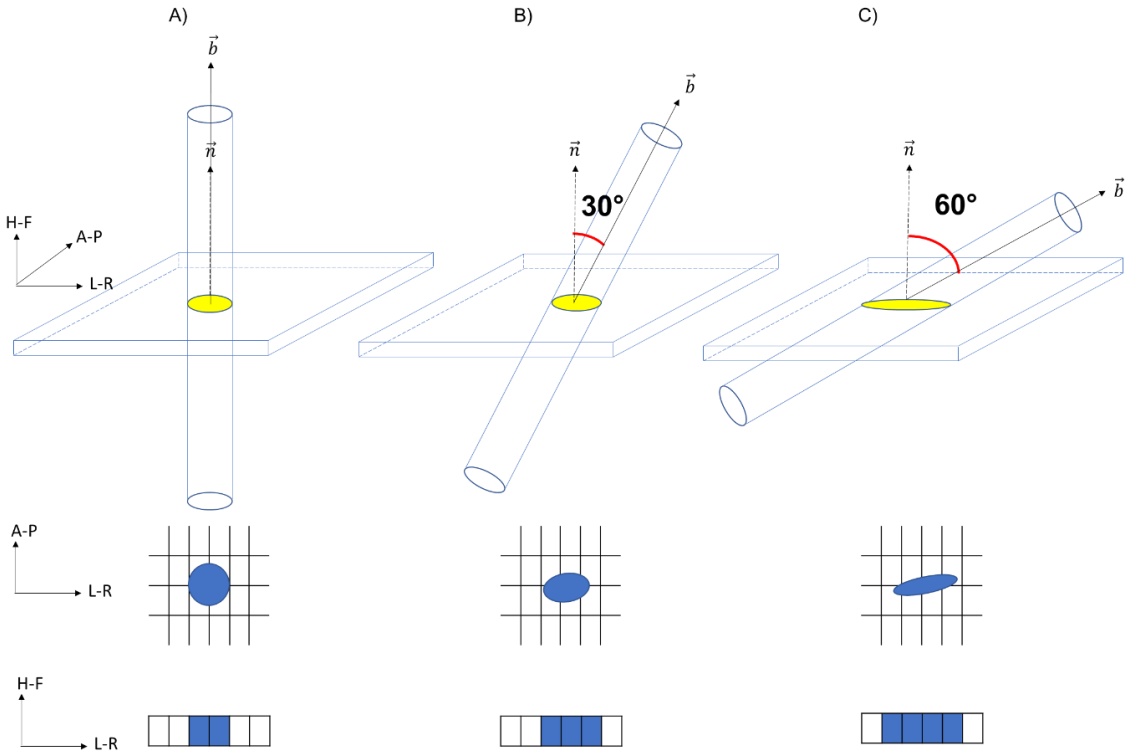


Figure 56. A schematic diagram representing partial volume effect exerted by a single bile duct intersecting an axial segmented liver cT1 slice at different angles. (A) 0°, (B) 30°, and (C) 60° to the slice normal represented by the dotted line (i.e., the angle between the normal vector of the imaging plane, $\vec{n}$ and the vector describing the orientation of the bile duct, $\vec{b}$). Their respective proportion of partial volume voxels are represented on the axial (A-P and L-R axis, middle row) and coronal (H-F and L-R axis, bottom row) views. Abbreviation: A-P, anterior-posterior; L-R, left-right; H-F, head-feet.

6.5.4 Relationship between periductal cT1 and other quantitative MRI and MRCP markers

Interquartile range (IQR) cT1 as a measure of parenchymal heterogeneity has been discussed in chapter 5. For the analysis, patients with PSC were divided equally by tertiles into low, intermediate, and high parenchymal heterogeneity by the IQR cT1 distribution. Intrahepatic sum of relative severity of dilatations has been discussed as a candidate biomarker of disease severity in chapter 4. For the analysis, patients with PSC were divided equally by tertiles into low, intermediate, and high severity of dilatations by the distribution of intrahepatic sum of relative severity. The same was performed according to the distribution of intrahepatic sum of relative severity of strictures.

6.5.5 Healthy volunteer validation cohort

T1, T2* and 3D MRCP data acquired using the same protocol as the MoBILISE study was obtained from an independent study conducted by a collaborator, Perspectum to test the accuracy and precision of *MRCP+*[146]. The repeatability and reliability (R&R) study received a favourable opinion from the South Central – Oxford C Research Ethics Committee (REC reference: 17/SC/0459). All participants gave written informed consent for use of their data in future research. The study included 20 healthy volunteers, 10 self-reported mixed aetiology liver diseases (5 NAFLD, 2 haemochromatosis, 1 HCV, 1 veno-occlusive disease, and 1 NAFLD with haemochromatosis), and 10 self-reported mixed aetiology "biliary diseases" (6 PSC subtype unknown, and 4 PBC). Given the heterogeneity in the disease group coupled with the inability to examine their medical records to verify the diagnosis, only data from the healthy volunteers (not age-sex matched) were included for analysis in this thesis. Scan series and time were checked to ensure only MRI and MRCP data acquired at the same time and on the same scanner was included in the analysis.

6.5.6 Statistical analysis

Descriptive statistics were used to summarise participant characteristics. Normality was assessed using the Shapiro-Wilk test. Results were reported using mean and standard deviation if parametric distribution or median and IQR if non-parametric distribution. Associations between continuous variables were tested using the Pearson's correlation coefficient for parametric variables and Spearman's rank correlation coefficient for non-parametric variables. Paired comparison between ROIs were performed using a one-way ANOVA for parametric distribution with Tukey's post-hoc correction for multiple comparisons, or Friedmann's test for non-parametric distribution with Dunn's post-hoc correction for multiple comparisons. Receiver operating characteristic curve (ROC) analyses were used determine the diagnostic accuracy of periductal mean cT1 to identify

the high-risk PSC group. Cut off values were selected using the Youden index with the corresponding sensitivity and specificity reported.

6.6 Results

6.6.1 Cohort characteristics

PSC

There were 73 patients included in this analysis from the 81 recruited to the study. Recruitment details are summarised in Figure 57. One patient was claustrophobic and declined MRI examination. Four were excluded for poor quality MRCP acquisition that prevented *MRCP+* analysis. Two were excluded for poor co-registration and one because MRI and MRCP were acquired at different times. The baseline characteristics of patients included in the final analysis are summarised in Table 59.

Healthy volunteers

Data from 20 age and sex-matched healthy controls were available from the MoBILISE study for the final analysis. Their baseline characteristics are summarised in chapter 4. Data from an additional 11 healthy volunteers were available from the R&R study. The remaining nine participants from the R&R study were excluded as the available liver MRI and MRCP scans were acquired at different time points and/or on different MRI manufacturer. The demographic data for the healthy volunteers in the R&R study were not available.

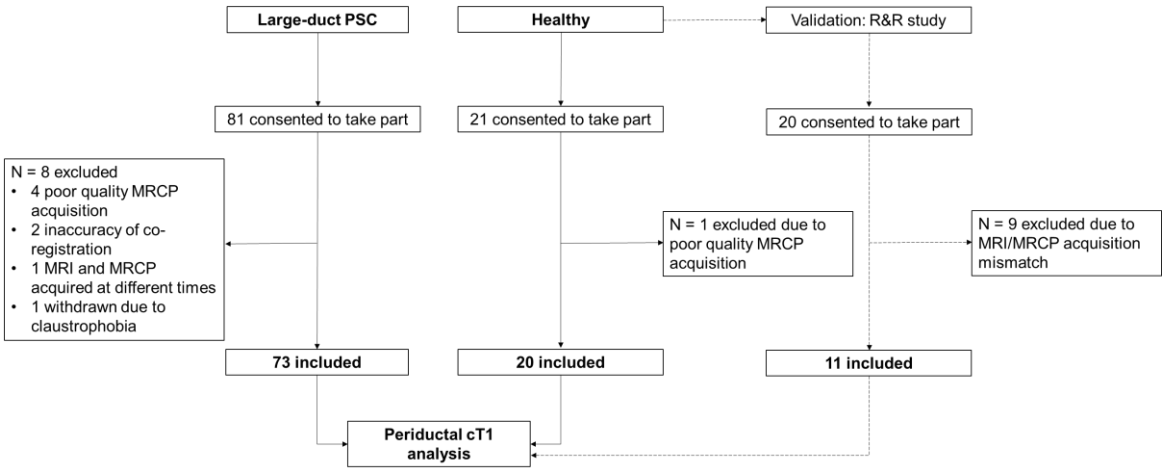


Figure 57. Study flow diagram.

Table 59. Baseline characteristics of 73 patients with PSC included in the final analysis.

	PSC (n = 73)	
Demographics		
Male	48	(66)
Age (years)	44	(31-56)
Clinical data		
PSC duration (years)	8	(4-12)
PSC disease distribution		
Intrahepatic + extrahepatic	41	(56)
Intrahepatic only	32	(44)
On UDCA therapy	41	(56)
IBD present	55	(75)
IBD phenotype		
Ulcerative colitis	41	(56)
Crohn's	8	(11)
Unspecified	6	(8)
Laboratory parameters		
Total bilirubin (µmol/l)	13	(10–20)
ALT (IU/l)	42	(27–84)
AST (IU/l)	38	(27–61)
xULN ALP	1.2	(0.8–1.8)
GGT (IU/l)	123	(50–297)
Albumin (g/l)	40	(37–41)
Prothrombin time (s)	10.6	(10.3–11.0)
Platelet count (x 10^9/l)	256	(192–305)
Liver fibrosis markers		
Transient elastography liver stiffness (kPa)	6.8	(5.3-9.5)
ELF	9.3	(8.7-9.9)
Prognostic risk models		
Mayo Risk Score	0.1	(-0.2–0.4)
Amsterdam-Oxford Model score	1.6	(1.3–2.0)

Continuous variables are expressed as median (interquartile range) and nominal variables as absolute number (percentage). Abbreviations: ALP, alkaline phosphatase; ALT, alanine aminotransferase; AST, aspartate aminotransferase; ELF, enhanced liver fibrosis score; GGT, gamma-glutamyltransferase; IBD, inflammatory bowel disease; INR, international normalised ratio; PSC, primary sclerosing cholangitis; UDCA, ursodeoxycholic acid; xULN, times upper limit of normal.

6.6.2 Periductal mean cT1 in healthy controls

The periductal mean ± standard deviation cT1 in the five ROIs were as follows: ROI 1 (1.5–3.5mm): 724.6 ± 44.0ms; ROI 2 (3.5–5.5mm): 721.9 ± 42.0ms; ROI3 (5.5-7.5mm): 719.4 ± 40.0ms; ROI 4 (7.5–9.5mm): 717.7 ± 41.0ms; and ROI 5 (whole liver): 719.5 ± 41.0ms. There were no differences in the periductal mean cT1 between any of the regions in healthy volunteers ($p = 0.3661$; Figure 58).

6.6.3 Validation of periductal cT1 in R&R study healthy volunteers

The periductal mean ± sd cT1 in the five ROIs were as follows: ROI 1 (1.5–3.5mm): 743.6 ± 80.4ms; ROI 2 (3.5–5.5mm): 740.1 ± 76.2ms; ROI 3 (5.5–7.5mm): 738.5 ± 77.6ms; ROI 4 (7.5–9.5mm): 735.6 ± 78.9ms; and ROI 5 (whole liver): 736.2 ± 79.8ms. There were no differences in the periductal mean cT1 between any of the ROIs ($p = 0.4189$; Figure 59).

6.6.4 Periductal mean cT1 in patients with PSC

The periductal mean cT1 in the five ROIs were as follows: ROI 1 (1.5–3.5mm): 794.7ms; ROI 2 (3.5–5.5mm): 779.6ms; ROI 3 (5.5–7.5mm): 771.8ms; ROI 4 (7.5–9.5mm): 769.4ms; and ROI 5 (whole liver): 774.1ms. The periductal mean cT1 in ROI 1 and 2 were higher compared to the other regions ($p < 0.0001$). There were no differences between ROIs greater than 5.5mm radius from the bile duct wall and the whole liver mean cT1 (Figure 60).

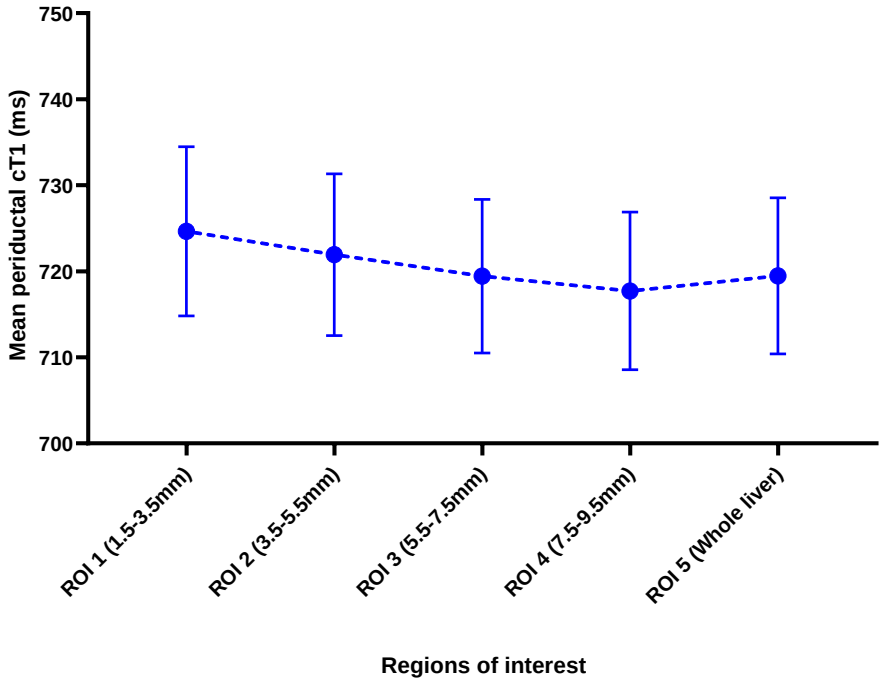


Figure 58. Mean ± SEM cT1 in periductal region of interests (ROIs) at fixed radial distances from the bile ducts in comparison to each other and the whole liver in 20 healthy controls in the MoBILISE study. There were no significant differences between any of the regions.

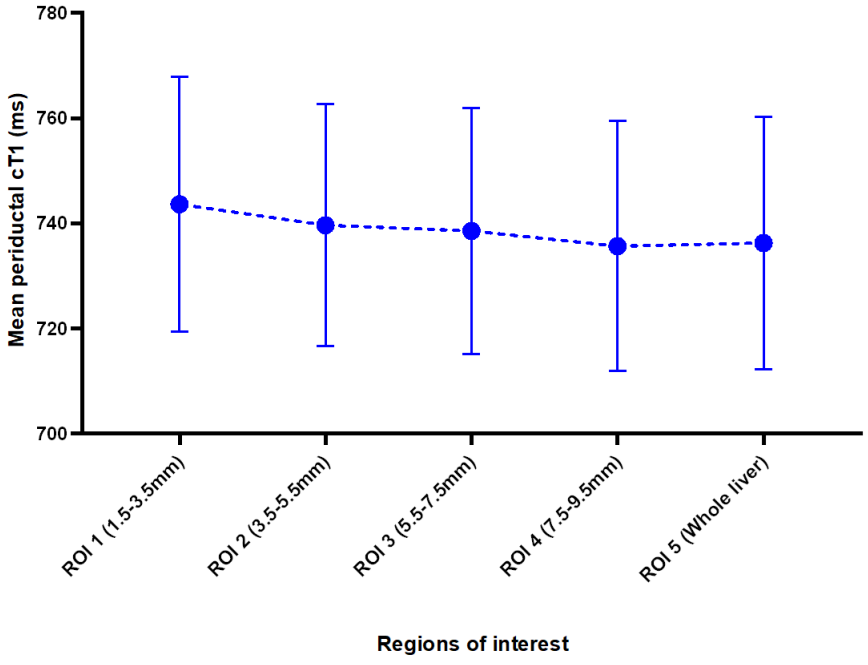


Figure 59. Mean ± SEM cT1 in periductal region of interests (ROIs) at fixed radial distances from the bile ducts in comparison to each other and the whole liver in the R&R study validation cohort of 11 healthy volunteers. There were no significant differences between any of the regions.

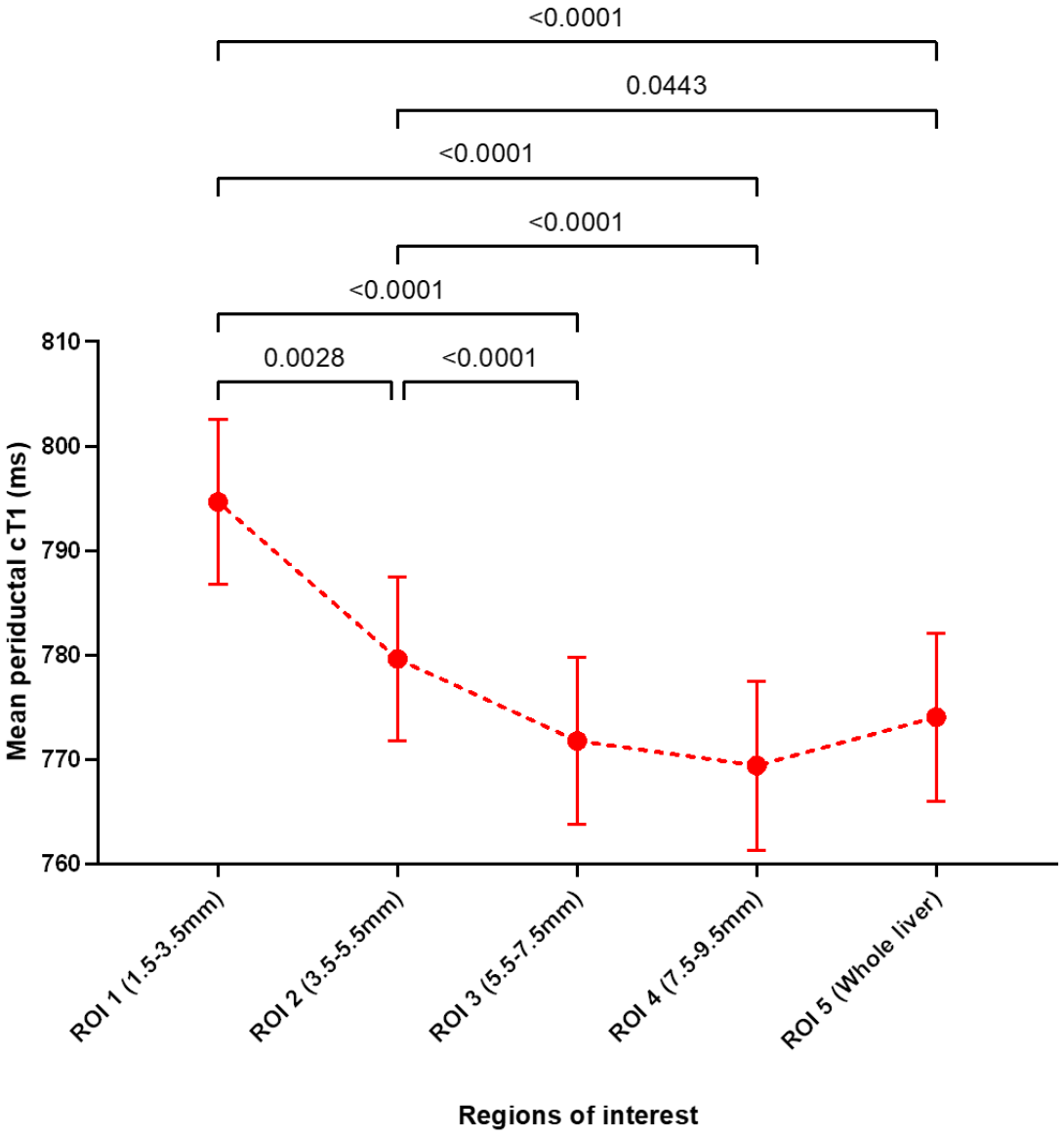


Figure 60. Mean ± SEM cT1 in periductal region of interests (ROIs) at fixed radial distances from the bile ducts in comparison to each other and the whole liver in 73 patients with PSC. Only significant pairwise comparisons are depicted.

6.6.5 Influence of oblique angle bile ducts on periductal mean cT1

In patients with PSC, the difference in mean cT1 between regions 1.5–5.5mmm (ROI 1), 5.5–9.5mm (ROI 2), and whole liver persisted ($p < 0.0001$) regardless of whether bile ducts intersected at less than 30° angle, between 30–60° angle, or between 60–90° angle to slice normal. Mean cT1 in ROI 1 showed a trend of progressive increase with increasing oblique angle intersection of bile ducts (Figure 61)

In the post-hoc multiple comparisons tests, the mean cT1 for bile ducts intersecting at less than 30° angle in ROI 1 (778.4 ± 70.7ms) were significantly higher than in ROI 2 (771.0 ± 66.7ms; $p < 0.0001$) and whole liver (773.7 ± 67.5ms; $p < 0.0050$). There was no difference between periductal regions 5.5–9.5mm and whole liver ($p > 0.9999$).

The mean cT1 for bile ducts intersecting between 30–60° angle in ROI 1 (784.5 ± 69.2ms) were also significantly higher than in ROI 2 (771.7 ± 72.1ms; $p < 0.0001$) and whole liver (774.9 ± 68.0ms; $p < 0.0001$). There was no difference however between ROI 2 and whole liver mean cT1 ($p = 0.2729$).

The mean cT1 for bile ducts intersection between 60–90° angle in ROI 1 (787.5 ± 74.3ms) were also significantly higher than in ROI 2 (770.1 ± 77.9ms; $p < 0.0001$) and whole liver (774.8 ± 69.5ms; $p < 0.0140$). There was no difference however between ROI 2 and whole liver mean cT1 ($p = 0.0618$).

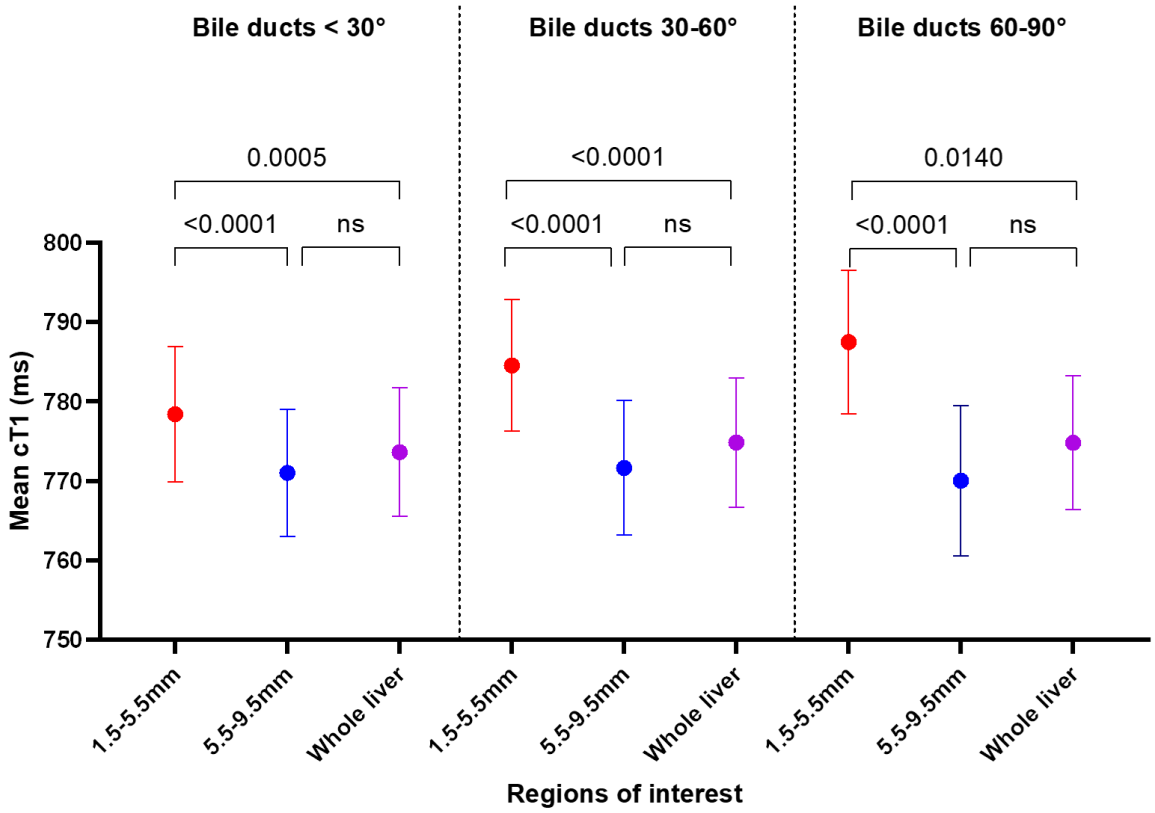


Figure 61. Mean ± SEM cT1 comparisons between ROI 1 (1.5–5.5mm), ROI 2 (5.5–9.5mm), and ROI 3 (whole liver) stratified according to angle of intersection of bile ducts with the slice normal. ns, not significant.

6.6.6 Association of periductal mean cT1 with liver fibrosis and inflammation

Liver fibrosis markers

Periductal mean cT1 1.5–3.5mm showed a positive correlation with LS (r = 0.36, p = 0.0018) (Figure 62). However, there was no correlation with ELF score (r_s = 0.13, p = 0.340). There was significant correlation with the tissue inhibitors of metalloproteinases 1 (TIMP-1; r_s = 0.30, p = 0.011) component of the ELF score.

Liver inflammation markers

Periductal mean cT1 did not show any significant correlation with any of the liver enzymes (ALP: r_s = 0.10, p =0.405; ALT: r_s = -0.23, p = 0.066; AST: r_s = -0.08, p = 0.483).

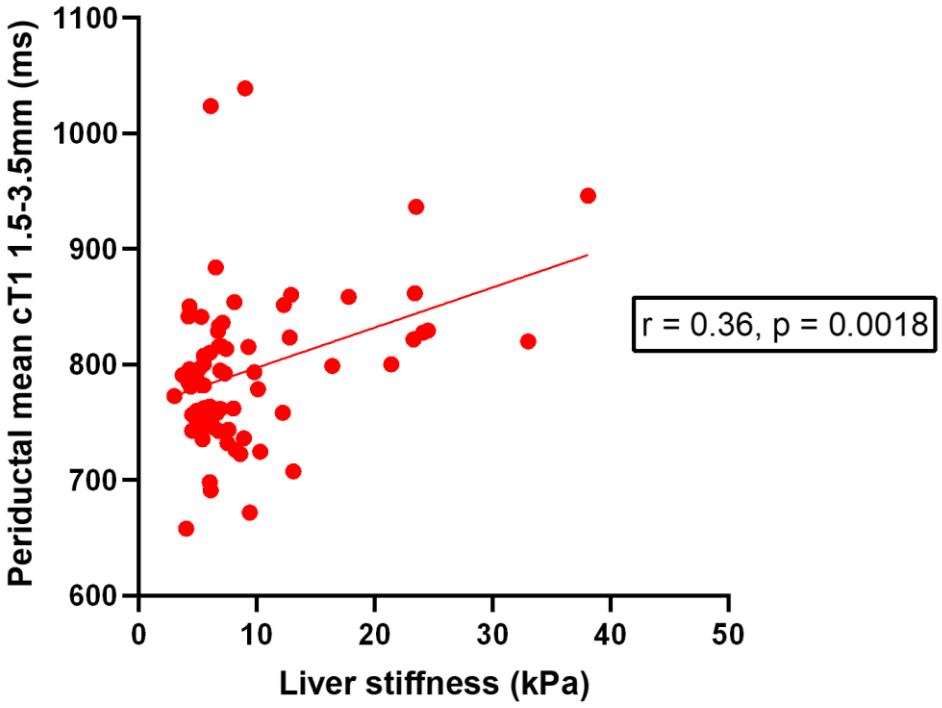


Figure 62. Correlation between transient elastography liver stiffness and periductal mean cT1 in ROI (1.5-3.5mm) in patients with PSC.

6.6.7 Risk stratification

Liver fibrosis markers

The periductal mean cT1 in the region 1.5–3.5mm surrounding the bile ducts were higher in patients with LS > 9.6kPa (n = 19; 832.7 ± 75.5ms) compared to patients with LS ≤ 9.6kPa (n = 53; 781.0 ± 58.5ms; p = 0.0021). Periductal mean cT1 cut-off of 818.1ms had an AUC of 0.74 (95% CI 0.60–0.88), sensitivity 63% and specificity 83% to detect the high-risk group with advanced fibrosis (Figure 63). Patients with ELF > 9.8 (n = 20; 824.0 ± 80.1ms) also had a higher periductal mean cT1 in the region 1.5–3.5mm compared to those with ELF ≤ 9.8 (n = 52; 783.4 ± 58.1ms; p = 0.0182). However, the diagnostic performance was poor with AUC of 0.68 (95% CI 0.53–0.83).

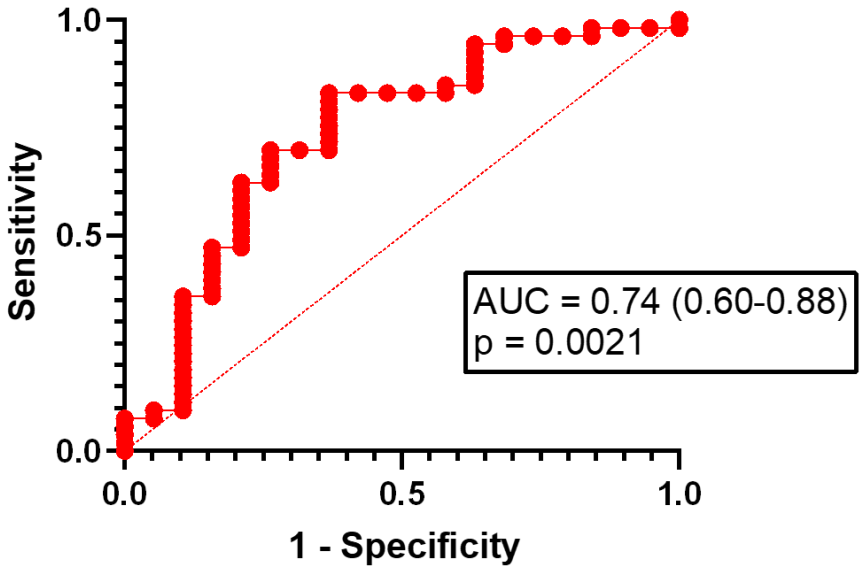


Figure 63. Receiver operator curve for the identification of patients with advanced fibrosis defined as transient elastography liver stiffness greater than 9.6kPa using periductal mean cT1 1.5–3.5mm.

Prognostic risk models

The periductal mean cT1 in region 1.5–3.5mm was higher in patients with AOM > 2 (n = 22; 826.9 ± 71.3ms) compared to those classified as AOM ≤ 2 (n = 50; 780.5 ± 58.1ms; p = 0.0031). Periductal mean cT1 cut-off of 811.7ms had an AUC of 0.72 (95% CI 0.59-0.85), 68% sensitivity and 80% specificity to detect the high-risk group (Figure 64). However, there was no difference when patients were stratified by MRS > 0 (n = 40; 806.1 ± 77.9ms) or MRS < 0 (n = 32; 780.4 ± 47.5ms; p = 0.1224).

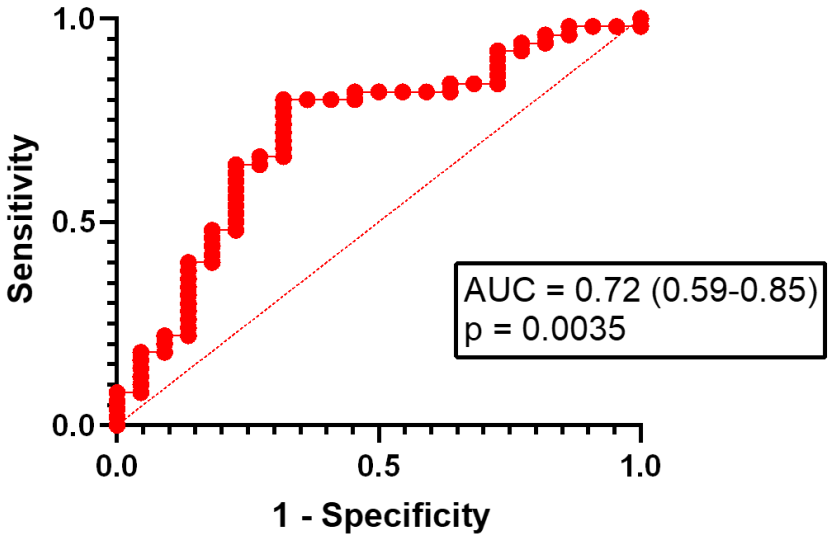


Figure 64. Receiver operator curve to detect patients with high-risk for death and/or liver transplantation stratified by Amsterdam-Oxford Model score > 2 using periductal mean cT1 1.5–3.5mm.

Cholestasis marker

There were no differences in periductal mean cT1 in the region 1.5–3.5mm surrounding bile ducts in high-risk patients with PSC stratified by 2.2x or 1.5x ULN ALP.

Qualitative MRI-MRCP scores and features

There were no differences in periductal mean cT1 in the high-risk groups classified by either Anali score (radiologist 1: p = 0.1519; radiologist 2: p = 0.0534) or modified Amsterdam intrahepatic stricture severity score (ISSS; radiologist 1: p = 0.8012; radiologist 2: p = 0.6020). There were no associations observed between periductal mean cT1 and morphological changes on MRI including hepatic dysmorphy, cirrhosis, and portal hypertension with either radiologists. There were no differences either when patients were stratified by PSC disease distribution.

6.6.8 Periductal cT1 stratified by quantitative MRI-derived marker of parenchymal heterogeneity

Whole liver IQR cT1 (r_s = 0.24, p = 0.0387) was not highly correlated with periductal cT1 1.5–3.5mm and therefore are independent of each other. IQR cT1 in the low, intermediate, and high heterogeneity groups were 778.0 ± 51.4ms, 778.9 ± 41.9ms, and 827.1 ± 88.3ms, respectively. There was no difference between the low and intermediate groups but there was a difference between low and high groups (p = 0.0257) and between intermediate and high groups (p = 0.0290).

6.6.9 Periductal cT1 stratified by quantitative MRCP-derived marker of intrahepatic duct dilatation severity

Intrahepatic sum of relative severity of dilatations showed no correlation with periductal mean cT1 1.5–3.5mm (r_s = 0.20, p = 0.0961) and there were no differences (p = 0.5857) in periductal mean cT1 between those with low (788.5 ± 75.2ms), intermediate (789.7 ± 44.3ms), and high (805.7 ± 77.6ms) severity of dilatations.

6.6.10 Association between mean periductal cT1 and patient-reported outcomes in PSC

Periductal mean cT1 1.5–3.5mm showed significant correlation with several of the domains of SF-36 as summarised in Table 60. The mental health component (r_s = 0.38, p = 0.002) and role emotional (r_s = 0.37, p = 0.010) showed the strongest correlations. There were no significant correlations with any of the PSC-PRO domains.

Table 60. Correlation between SF-36 domains and mean periductal cT1 1.5–3.5mm.

SF-36 domain	Correlation coefficient	p-value
Physical component summary	-0.20	0.098
Physical functioning	-0.26	0.034
Role physical	-0.18	0.153
Bodily pain	-0.30	0.014
General health	-0.26	0.036
Mental component summary	-0.37	0.002
Energy/fatigue	-0.20	0.105
Social functioning	-0.29	0.017
Role emotional	-0.31	0.010
Mental health	-0.38	0.002

6.7 Discussion

This work is the first to propose a method for objectively quantifying periductal changes in PSC using a combination of two quantitative MRI techniques. Periductal mean cT1 was higher in regions closer to the bile duct than further away even after adjusting for potential confounding partial volume effect from oblique angles of bile duct intersection. Periductal mean cT1 in the region 1.5–3.5mm surrounding the bile ducts showed a significant correlation with liver fibrosis. Patients with LS > 9.6kPa, ELF > 9.8, and AOM > 2 had higher periductal mean cT1 than patients with LS < 9.6kPa, ELF < 9.8, and AOM < 2, respectively. The periductal mean cT1 cut-offs were in the same region between 812–818ms.

Periductal mean cT1 was no different in regions closer to the bile ducts than further away or whole liver mean cT1 in healthy controls, both in the MoBILISE and R&R studies. This suggest that the increased periductal mean cT1 in PSC may truly be disease-driven rather than a false-positive result from chance, methodological limitations such as partial

volume effect or inaccuracy in co-registration of the two MRI techniques. Several steps were taken to minimise inaccuracy and partial volume effect including visual inspection of co-registered images, excluding voxels below 1.5mm distance from the bile duct wall, and accounting for oblique angle of bile duct intersection with cT1 slice.

Periductal liver cT1 voxels may be confounded by partial volume effect from bile in the boundaries of the bile ducts and smaller blood vessels accompanying the bile ducts. In addition to excluding cT1 voxels lesser than 1.5mm in distance from the bile ducts, this work has attempted to address the confounding effect of oblique angle of intersection of bile ducts with liver cT1 slices. Increasing the angle of deviation from slice normal bring the bile ducts proximity closer to the liver, and therefore increases the number of periductal cT1 voxels that is averaged out to contain signals from both the liver and bile. As discussed in chapter 2, freely moving protons in free water have very long T1 relaxation time — approximately 4000ms at 1.5T[169]. Bile is composed predominantly of water (97-98%)[170], therefore has the effect of increasing the cT1 in the periductal regions closer to the bile ducts. Whilst the increased periductal cT1 in the region 1.5–5.5mm surrounding the bile ducts persisted with varying oblique angles compared to regions further away, the mean value showed an increasing trend with increasing deviation of oblique angle from slice normal. This data suggests that there may in fact be partial volume effect that is contributing to the rise in cT1 when more oblique angle ducts intersect the liver.

Although liver enzymes are imperfect surrogates of liver inflammation, the trend towards a negative correlation with serum ALT and AST, and no significant correlation with serum ALP suggests that periductal mean cT1 in this cohort is likely to be quantifying periductal fibrosis rather than periductal hepatitis or cholangitis. Therefore, it is not surprising that periductal mean cT1 is able to separate the high-risk groups stratified by liver fibrosis markers and AOM. Periductal mean cT1 however is significantly different to and

independent of whole liver mean cT1, therefore it is not just a surrogate for quantification of whole liver fibrosis. In chapter 5, there is evidence to suggest that heterogeneity in the liver parenchyma is driven by distribution of fibrosis in the liver. Data presented here suggests that patients with higher IQR cT1 i.e., heterogeneity have higher periductal mean cT1. This possibly strengthens the argument that periductal fibrosis contributes to the parenchymal heterogeneity as measured by quantitative MRI. However, the incongruency with qualitative MRI features of hepatic dysmorphy and cirrhosis could refute this.

MRI-MRCP image resolution limits detection of bile ducts only up to the segmental ducts. Depiction of smaller bile ducts including the septal, interlobular, and bile ductules are not possible as their diameter is less than 400μm[171]. Therefore, periductal cT1 is quantifying periductal fibrosis in bile ducts larger than the septal ducts. There was no correlation between periductal mean cT1 and *MRCP*+-derived intrahepatic dilatation or stricture severity metrics. Similarly, conventional MRCP ISSS and maximum intrahepatic bile duct dilatation classification by both radiologists did not show any correlation with periductal mean cT1. Bile duct dilatation is a consequent of downstream stricture; therefore, it is not surprising that periductal cT1 has not shown association with either quantitative or qualitative dilatation measurements. The lack of correlation with *MRCP*+ intrahepatic stricture severity metric and qualitative MRCP ISSS could be due to the technical limitation of both to detect intrahepatic strictures in an under-distended biliary tree.

Clinical application

Periductal cT1 quantification could potentially aid in the diagnosis of early PSC. Patients with cholestatic liver enzymes with no visible features of PSC on MRI-MRCP undergo a liver biopsy to look for features of small-duct PSC on histology. Nevertheless, abnormal peribiliary enhancement involving bile ducts have been reported to be present on 'normal' MRI-MRCP of patients labelled as small-duct PSC. In a small study of patients with small-

duct PSC, seven (70%) patients had arterial phase peribiliary enhancement, and nine (90%) had delayed phase peribiliary enhancement on MRI-MRCP despite having a normal cholangiogram[172]. It has been speculated that this may indicate early stage of large-duct PSC before the appearance of visible biliary strictures on MRI-MRCP to the naked eye. In the same study, heterogeneous hepatic parenchymal signal intensity and delayed phase enhancement were observed in patients with small-duct PSC. These changes have been postulated to be related to periductal inflammation in the early stages and fibrosis in later stages, which can involve adjacent blood vessels resulting in perfusion abnormalities[111, 155]. The classification and progression of small-duct PSC to large-duct PSC remains ambiguous as to whether small-duct PSC is a separate disease condition altogether or merely represents an early or mild stage of large-duct PSC. The largest study involving 83 patients with small-duct PSC reported a progression to large-duct PSC in 23% of patients in a median of 7 years[49]. In a more recent study involving 16 patients with small-duct PSC, 5 patients (31%) progressed to large duct PSC with typical cholangiographic findings at a longer median follow-up of 12 years[173].

Another potential application of periductal cT1 quantification is early detection of cholangiocarcinoma (CCA) in PSC, particularly intrahepatic (iCCA) and perihilar (pCCA) type. CCA is often already in advanced stages when the diagnosis is made as it usually remains asymptomatic in early stages. A late diagnosis limits therapeutic options and contributes to the poor prognosis in these patients. Development of CCA in PSC is a contraindication for liver transplantation in the UK, the only proven effective life-extending intervention in PSC currently. Early diagnosis in a high-risk population such as PSC potentially offers early curative surgical intervention. Although convincing evidence for a surveillance programme in patients with PSC is yet to be established, most clinicians tend to perform MRI liver and MRCP at regular intervals. More centres are acquiring 3D MRCP sequences now making it compatible with *MRCP+* analysis. If T1 and T2* mapping sequences for Liver*MultiScan* analysis can be incorporated into clinical MRI sequences,

interval periductal cT1 measurements could offer close monitoring of the periductal region in a suspected segment to potentially differentiate cholangitis from early CCA.

Patient reported outcomes

As discussed in chapter 5, SF-36 domains were observed to be correlated with liver fibrosis severity. Given that periductal fibrosis is the main driver of PSC, it is not surprising that periductal mean cT1 closest to the bile ducts showed similar associations with some of the domains of SF-36.

One of the main limitations of this work is the absence of a 'gold standard' comparator for periductal mean cT1. In addition to specific limitations in PSC discussed above, liver biopsies sample only smaller bile ducts whilst MRCP resolution only allows delineation of the biliary tree up to segmental ducts. The absence of contrast administration prevented the study radiologists from classifying the degree of peribiliary hyperenhancement on MRI qualitatively for comparison. Liver cT1 has been demonstrated to histologically correlate with both liver inflammation and fibrosis. The increased periductal cT1 in this study was assumed to be due to periductal fibrosis based on the findings. Whilst none of the patients with PSC had recent cholangitis at study recruitment, subclinical inflammation and cholangitis may be contributing to increased periductal cT1. Although there was association with heterogeneity, the extent of periductal fibrosis may not have been truly captured from four axial liver slices in a patchy disease such as PSC. Future validation study should consider acquiring more volumetric slices to minimise partial volume effect on periductal cT1 measurements.

In conclusion, this chapter has demonstrated a novel method to quantitatively assess periductal changes in PSC using a combination of two quantitative MRI techniques. Periductal mean cT1 in regions closer to the bile ducts were observed to be higher than regions further away in PSC, but not seen in healthy volunteers. Periductal mean cT1

showed significant correlation with transient elastography liver stiffness measurements. Patients with increased liver parenchyma heterogeneity as measured by IQR cT1 showed higher periductal mean cT1 as well, demonstrating that periductal fibrosis contributes to the appearance of heterogeneity in the liver. This work demonstrates how a feature of disease that was previously only seen on histology, may be evaluated macroscopically, albeit with different magnification factor involving different size bile ducts. With further refinement in measurement technique to address potential confounders, periductal cT1 could be useful in monitoring periductal changes, especially in early PSC and CCA.

In the next chapter, I will be exploring how the candidate metrics identified from the previous chapters perform over time.

7 Temporal relationship in candidate quantitative metrics

7.1 Abstract

Background: The assessment of change in a non-invasive biomarker of disease is important both clinically to reflect disease progression and in interventional trials to assess the effectiveness of a study drug. The main aim of this study was to explore the temporal relationship in candidate MRI metrics and assess whether both time and baseline risk classification have any effect on these changes.

Methods: Follow-up biliary data were analysed using *MRCP+* and liver data using Liver*MultiScan*. Using baseline recruitment data, patients were risk-stratified into high-risk or low-risk group by MRS, AOM, LS, ELF, ALP, Anali, ISSS, presence of EHD or DS. Candidate MRI metrics were compared between visits 1 and 2 using paired tests. A two-way 2 x 2 mixed-effects ANOVA model was used to examine for interactions between candidate MRI metrics, low- or high-risk classification, and over the two time points. Diagnostic performances of baseline and change from baseline values of candidate metrics to detect the high-risk group were compared using ROC curve analyses. Differences in QoL data were examined.

Results: A total of 55/80 (69%) patients attended follow-up visit. There were no differences in mean intrahepatic SumRelSevDilat, IQR cT1, and periductal cT1 1.5–3.5mm between the two study visits. There was, however, a difference in mean superpixel variance cT1, seen in the low-risk group ($p < 0.0001$) but not in the high-risk group ($p > 0.05$). There was a significant interaction between time and risk categorisation by AOM ($p < 0.001$), LS ($p = 0.010$), ELF ($p = 0.004$), and ISSS ($p = 0.039$) on mean IQR cT1 with a large effect size for AOM ($\eta_p^2 = 0.228$) and ELF ($\eta_p^2 = 0.149$) and medium effect size for LS ($\eta_p^2 = 0.122$) and ISSS ($\eta_p^2 = 0.079$). There was no effect on SumRelSevDilat nor mean periductal cT1. The diagnostic performances of single baseline measurements were numerically better than the change between time points to detect the high-risk group. There were significant differences in bodily pain (p=0.0050), energy/fatigue

(p<0.0001) and mental summary score (p=0.0006) domains of SF-36 between the two study visits.

Conclusion: A temporal change in IQR cT1 was observed in the high-risk PSC group when stratified by baseline AOM, LS, ELF, Anali and ISSS whereas change in mean superpixel variance cT1 was seen in the low-risk PSC group. Changes in intrahepatic SumRelSevDilat and mean periductal cT1 are driven by baseline risk classification and is independent of time.

7.2 Introduction

In most patients, disease progression is slow in PSC. The median time for development of liver transplant or death and hepatopancreatobiliary malignancy was 14.5 years in a large retrospective multicentre study[37]. Therefore, prospective studies examining the prognostic performance of biomarkers in PSC are limited to development of short-term outcomes.

Changes in serum ALP levels have long been the subject of investigation in PSC. Various thresholds and relative change from baseline have been proposed to be associated with clinical outcomes in observational cohort studies[53-59]. For this reason, serum ALP has been widely used as primary or secondary endpoints in clinical trials. However, large inter- and intra-individual variations have recently been reported using a post-hoc analysis of data collected prospectively over a 2-year period in the negative phase 2 Simtuzumab trial in PSC[60]. In this study, although baseline elevated ALP level was associated with greater risk of PSC-related clinical events, changes in ALP levels did not correlate with disease progression. On the contrary, in the same trial, ELF showed relative stability over time. Both baseline values and ELF scores at different time points were associated with biopsy-proven fibrosis progression and development of cirrhosis.

The change in repeated transient elastography LS measurements over time have been investigated with mixed evidence. In a study of 142 patients, change in LS was associated with endpoints[76]. However, it is unclear whether it is an independent predictor as the changes in LS was not adjusted for baseline values. In a prospective clinical trial, when the change in LS was adjusted for baseline values, it was not associated with the primary outcome[63].

MRI-MRCP changes over a period of 4 years were systematically analysed using a standardised qualitative interpretation model to extract features that were associated with disease progression[91]. These features were later incorporated into the Anali score, which

was shown to predict clinical outcomes in a large retrospective multicentre study, however, only baseline MRCP features were examined[134]. In the prospective phase 2 simtuzumab trial, another qualitative MRI-MRCP interpretation model was developed with similar components to the Anali score (hepatic dysmorphy, portal hypertension, perihepatic lymph nodes). The baseline MRI-MRCP based score predicted progression to liver cirrhosis and development of cholangitis, but again radiological changes over two time points were not examined.

A single LS measured by MRE and changes in LS over time were both independently associated with liver decompensation in a retrospective single-centre study of 204 patients[174]. In that study, those with lower baseline LS values, the LS progression was slow compared to faster progression in those with higher baseline values.

The temporal relationship in *MRCP+* and Liver*MultiScan* metrics has never been examined in PSC.

7.3 Hypothesis

Biliary and liver parenchymal abnormalities are unlikely to have changed significantly in a year in a predominantly low to moderate risk cohort of patients. Therefore, I would not expect significant differences in candidate MRI metrics between the two time points both for the whole cohort and when stratified by baseline risk category.

7.4 Aims

The aims of this chapter were to:

1. Investigate whether there is a within-subject change in candidate MRI metrics between baseline (visit 1) and follow-up visit (visit 2). The following candidate metrics identified in chapters 4 to 6 were examined here:

- Intrahepatic sum of relative severity of candidate dilatations (SumRelSevDilat; m^{-1})
- Interquartile range of liver cT1 (IQR cT1; ms)
- Mean of liver superpixel variance cT1 (mean spx variance cT1; ms)
- Mean of periductal cT1 in ROI 1 (1.5–3.5mm) surrounding the bile ducts (mean pd cT1 ROI 1)

2. Investigate whether there is a change in other non-invasive risk stratification tools between visits 1 and 2. The following tools were examined here:

- Transient elastography LS
- ELF
- ALP

3. Assess whether there is an interaction between time and baseline risk category (high-risk or low-risk) on candidate MRI metrics and determine whether any change in metrics is the result of the interaction between risk category and time.

4. Compare the diagnostic performance of baseline measurements and temporal change (Δ) from baseline measurements of candidate MRI metrics to detect the high-risk PSC group.

5. Assess the relationship in patient reported outcomes between visits 1 and 2

7.5 Methods

7.5.1 Follow-up assessments

Follow-up study assessments mirrored baseline study assessments as described in chapter 2. MRI and MRCP data were analysed as described in chapters 4, 5, and 6 to compute the corresponding quantitative metrics.

7.5.2 Statistical analysis

Paired t-test (parametric) or Wilcoxon matched-pairs signed rank test (non-parametric) was used to assess the within-subject differences in candidate liver and biliary MRI metrics, LS, ELF, ALP, and mean domain scores of SF-36 and PSC-PRO between visits 1 and 2. A two-way 2 x 2 mixed-effects ANOVA model (Figure 65) was used to examine for potential changes in candidate MRI metrics (dependent variable) between the low-risk and high-risk groups (between-subject variable or fixed effect) over the two time points (within-subject variable or random effect). Variables with non-parametric distribution were log-transformed to ensure normality. The magnitude of effects (partial-ETA squared; η_p^2) was defined as: 0.01 (small effect), 0.06 (medium), and 0.14 (large effect)[175]. ROC analyses were performed to assess the diagnostic performance of baseline measurement (V1) and temporal change from baseline measurement (ΔV1–V2) to identify the high-risk PSC group.

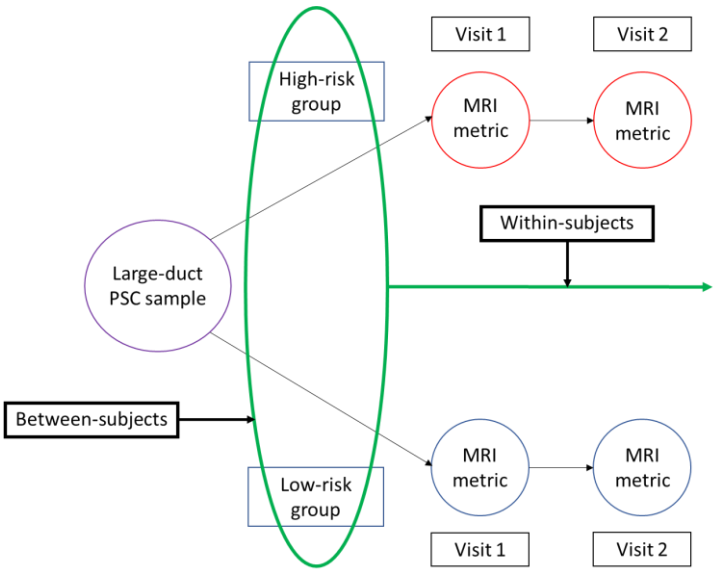


Figure 65. Summary of 2 x 2 mixed-effects ANOVA. Two independent groups are compared on how their MRI metrics taken at two study time points change over time.

7.6 Results

7.6.1 Cohort characteristics

Out of a possible 80 patients that underwent imaging at baseline, 55 (69%) attended follow-up visit. Two patients withdrew from the study, three had reached a clinical endpoint, two were not contactable, and 18 patients had their follow-up visit cancelled or postponed due to Covid-19 pandemic (Figure 66). The median (IQR) interval between baseline and follow-up visit was 417 (371–495) days. The characteristics of the 55 patients included for the temporal change analysis are summarised in Table 61.

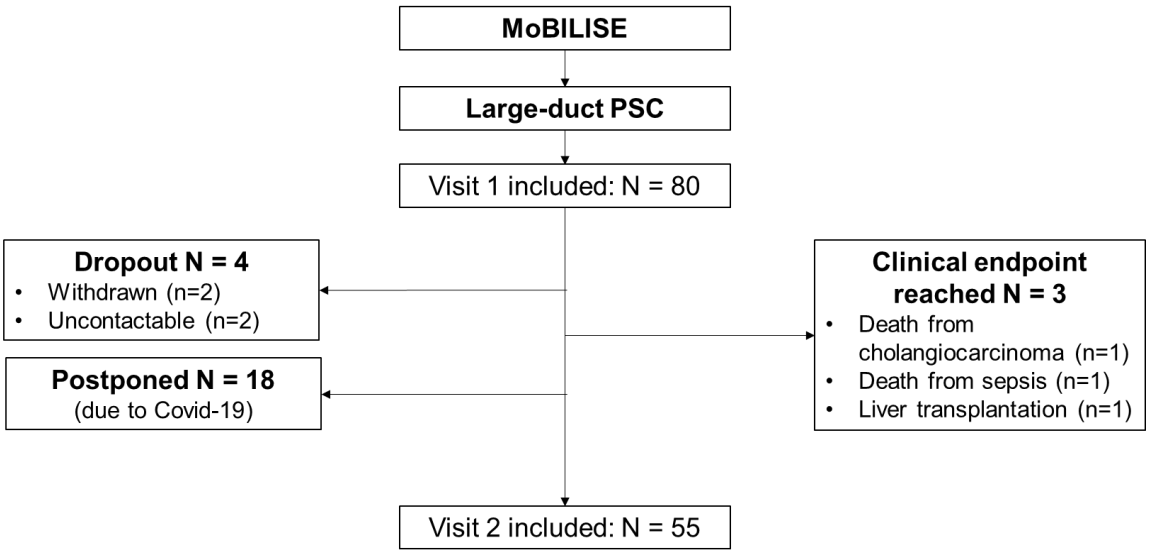


Figure 66. Study flow chart for follow-up visit.

Table 61. Characteristics of patients with large-duct PSC at baseline and follow-up visits.

	PSC (n = 55)	
	Visit 1	Visit 2
Demographics		
Male	35 (63)	35 (64)
Age (years)	43 (32–55)	45 (33–57)
Laboratory parameters		
Total bilirubin (µmol/l)	13 (11–21)	14 (10–20)
ALT (IU/l)	42 (27–84)	38 (27–68)
AST (IU/l)	38 (26–61)	35 (25–52)
ALP (IU/l)	139 (89–226)	125 (96–222)
xULN ALP	1.1 (0.7–1.7)	1.0 (0.7–1.7)
GGT (IU/l)	123 (48–308)	100 (49–269)
Albumin (g/l)	40 (38–42)	39 (37–42)
Prothrombin time (s)	10.5 (10.3–11.0)	10.7 (10.3–11.1)
Platelet count (x 10^9/l)	237 (195–293)	236 (191–285)
Clinical data		
PSC disease distribution		
Intrahepatic + extrahepatic	30 (55)	30 (55)
Intrahepatic only	25 (45)	25 (45)
On UDCA therapy	28 (51)	28 (51)
IBD present	38 (70)	38 (69)
IBD phenotype		
Ulcerative colitis	27 (49)	27 (49)
Crohn's	6 (11)	6 (11)
Unspecified	5 (9)	5 (9)
Liver fibrosis markers		
Liver stiffness (kPa)	6.9 (5.–10.0)	6.5 (5.4–11.0)
ELF score	9.4 (8.7–9.9)	9.4 (9.0–9.9)
Prognostic risk models		
Mayo risk score	0.0 (-0.2–0.4)	0.2 (-0.2–0.6)
Amsterdam-Oxford model score	1.6 (1.3–2.0)	1.5 (1.2–1.8)

Continuous variables are expressed as median (interquartile range) and nominal variables as absolute number (percentage). ALP, alkaline phosphatase; ALT, alanine aminotransferase; AST, aspartate aminotransferase; ELF, enhanced liver fibrosis score; GGT, gamma-glutamyltransferase; IBD, inflammatory bowel disease; INR, international normalised ratio; PSC, primary sclerosing cholangitis; UDCA, ursodeoxycholic acid; xULN, times upper limit of normal.

7.6.2 Within-subject temporal relationship of candidate MRI metrics and other markers

Overall, there was no difference in MRI metrics between the two study visits in all except one candidate metric. The mean superpixel variance cT1 had changed between the two time points when analysed for the whole cohort. It was higher at baseline visit compared to follow-up (59.7ms vs 55.9ms, $p < 0.0001$; Table 62). Post-hoc subgroup analysis by baseline risk categorisation revealed that in the high-risk PSC group defined by MRS > 0 (n = 28), AOM > 0 (n = 18), ULN ALP > 2.2x (n = 7), ULN ALP > 1.5x (n = 17), LS > 9.6kPa (n = 17), ELF > 9.8 (n = 16), Anali score > 2 (n = 19), ISSS > 2 (n = 24), presence of EHD (n = 31), and presence of DS (n = 18), there were no significant changes between visits 1 and 2. The reduction in the mean superpixel variance cT1 was in fact explained by a significant reduction between visits in the low-risk group as demonstrated in Figure 67, suggesting that baseline risk categorisation may have an impact on the temporal relationship of candidate metrics.

Table 62. Summary results of candidate MRI metrics and other markers at visits 1 and 2.

	Visit 1		Visit 2		p-value
Intrahepatic *MRCP*+ metrics					
SumRelSevDilat (m^{-1})	6.9	(5.0–9.2)	7.3	(5.6–9.3)	0.3347
DilatNum (m^{-1})	10.6	(8.1–13.9)	12.4	(9.3–14.9)	0.1568
DilatTreeProp (%)	7.1	(4.2–8.8)	7.7	(5.9–9.3)	0.0950
TreeDilatScore	45.2	(27.8–61.0)	47.6	(34.1–60.9)	0.1864
TreeVol (mlm^{-1})	6.3	(5.4–7.5)	6.4	(5.5–7.6)	0.1062
TreeLength (m)	1.1	(0.6–1.8)	1.2	(0.6–2.1)	0.5059
Liver cT1 metrics					
Median cT1 (ms)	748.5	(713.5–788.3)	756.5	(722.5–802.3)	0.1026
IQR cT1 (ms)	78.0	(73.0–92.8)	78.5	(71.0–91.8)	0.6760
Mean superpixel variance cT1 (ms)	59.7	(57.3–61.7)	55.9	(52.9–60.3)	<0.0001
Periductal cT1 metrics					
Mean ROI 1 (1.5–3.5mm)	783.2	(749.0–822.0)	789.1	(746.6–815.7)	0.3254
Mean ROI 2 (3.5–5.5mm)	766.9	(733.9–809.4)	776.4	(732.9–815.2)	0.6291
Mean ROI 3 (5.5–7.5mm)	758.0	(722.6–794.6)	760.7	(723.0–805.9)	0.5051
Mean ROI 4 (7.5–9.5mm)	759.3	(725.0–796.7)	759.8	(724.5–804.5)	0.4008
Other markers					
Liver stiffness, kPa	6.9	(5.2–10.2)	6.5	(5.3–11.0)	0.2217
Enhanced liver fibrosis (ELF)	9.3	(8.7–9.9)	9.4	(9.0–9.9)	0.1172
Alkaline phosphatase (ALP)	142.0	(89.0–231.0)	127.0	(101.5–225.0)	0.7605

Values are presented as median (interquartile range). DilatNum, number of candidate dilatations per metre; DilatTreeProp, proportion of tree dilated; ROI, region of interest; SumRelSevDilat, sum of relative severity of candidate dilatations per metre; TreeDilatScore, sum of tree dilatation scores; TreeLength, total length of biliary tree; TreeVol, biliary tree volume.

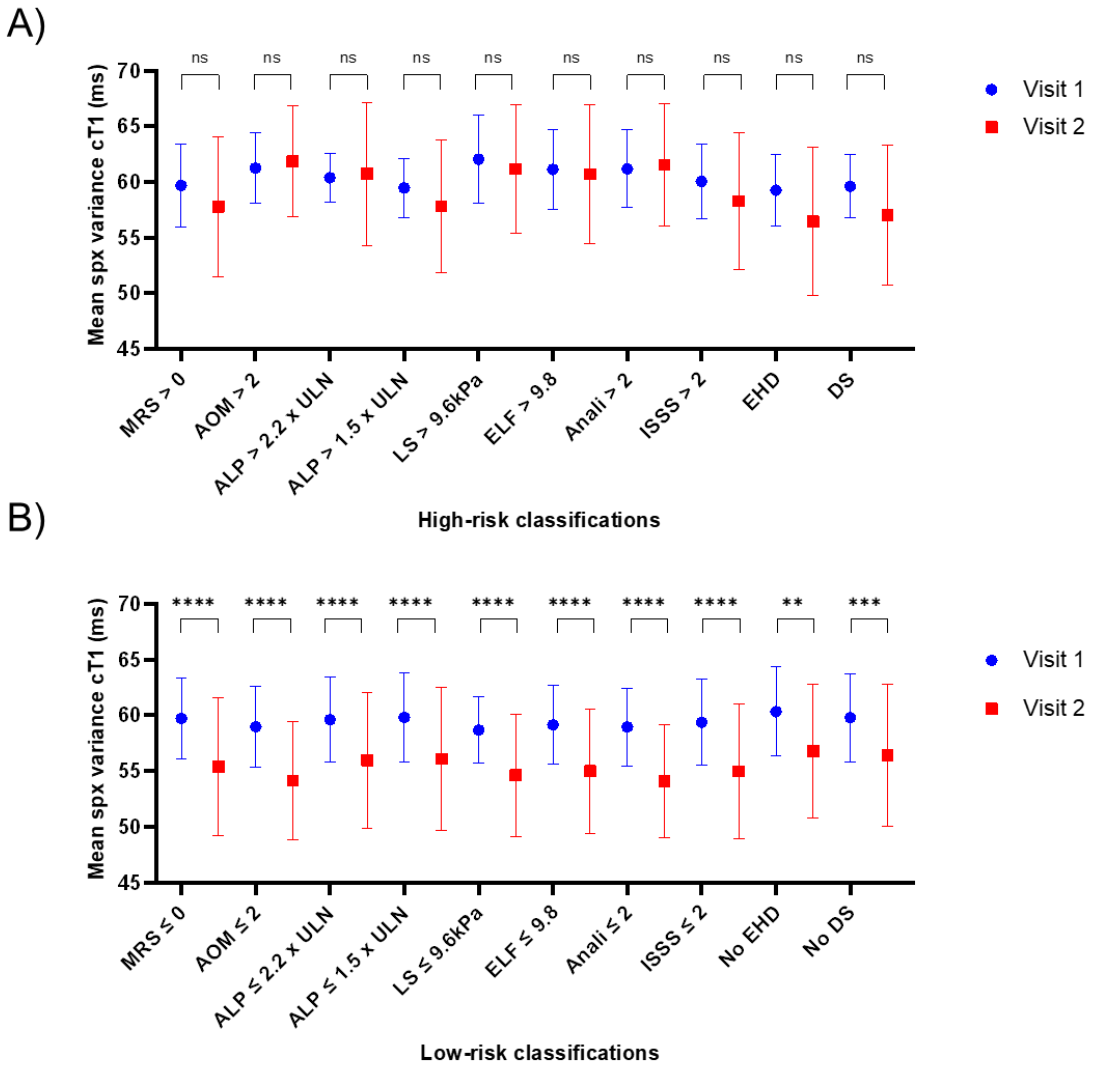


Figure 67. Differences in mean superpixel variance cT1 between study visits 1 and 2 in (A) high-risk; and (B) low-risk group stratified by non-invasive surrogate markers. MRS, Mayo Risk Score; AOM, Amsterdam-Oxford Model; ALP; alkaline phosphatase; LS; transient elastography liver stiffness; ELF, enhanced liver fibrosis score; ISSS, modified Amsterdam intrahepatic stricture severity score; EHD, presence of extrahepatic disease; DS, presence of dominant stricture. Lines and error bars indicate median and interquartile range.

7.6.3 Between-subject and within-subject effect interaction on candidate MRI metrics

The within-subject effect (time), between-subject effect (high-risk or low-risk), and interaction between the two on candidate MRI metrics were assessed. The effect of time on candidate metrics has been discussed above for the whole cohort but not examined in tandem with risk categorisation. This section will focus on the effect of the interaction between time and baseline risk classification on candidate MRI metrics examined with a 2 (time) x 2 (risk category) mixed-model ANOVA. Results of the mixed model ANOVA are presented here according to individual candidate MRI metrics. Analysis could not be performed when patients were risk stratified by 2.2x ULN ALP as it violated the Box's test of equality of covariance matrices.

7.6.3.1 Intrahepatic SumRelSevDilat

The main effect for time (within-subject effect) was not significant for any of the risk stratification methods. Therefore, there was no overall difference in the mean intrahepatic SumRelSevDilat between visits 1 and 2 within the low-risk or high-risk groups.

A significant main effect for risk categorisation (between-subject effect) was obtained for all risk stratification methods except when stratified by 1.5x ULN ALP, presence of EHD, and presence of DS. This result indicates that the mean intrahepatic SumRelSevDilat was higher in the high-risk groups stratified by MRS, AOM, LS, ELF, Anali, or ISSS at either time points as depicted in Figure 68.

Despite the general trend towards an increase in mean intrahepatic SumRelSevDilat between visits 1 and 2, the time and risk category interactions were not significant for any of the risk stratification methods as summarised in Table 63, indicating that any change in metrics observed was not dependent on time.

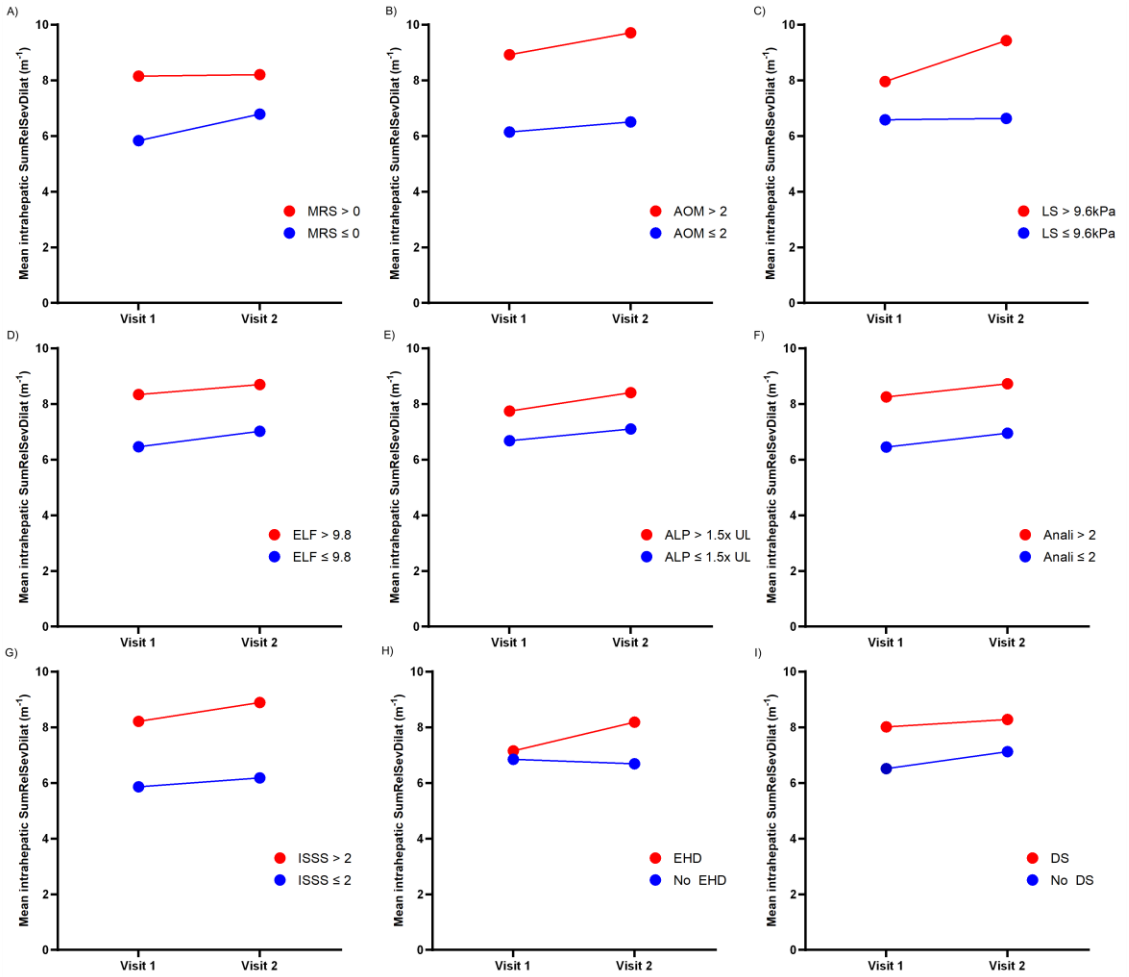


Figure 68. Differences in mean intrahepatic SumRelSevDilat between study visits 1 and 2 according to high-risk or low-risk group stratified by non-invasive surrogate markers. MRS, Mayo Risk Score; AOM, Amsterdam-Oxford Model; ALP; alkaline phosphatase; LS; transient elastography liver stiffness; ELF, enhanced liver fibrosis score; ISSS, modified Amsterdam intrahepatic stricture severity score; EHD, presence of extrahepatic disease; DS, presence of dominant stricture. Red line represents trend within the high-risk group and blue line low-risk group.

Table 63. Summary of mixed-model ANOVA results assessing the interaction of intrahepatic SumRelSevDilat with time and risk categorisation stratified by non-invasive surrogate markers of disease severity.

	Mean square	F-ratio (F)	p-value	Effect size ($\eta_p^{2)}$
MRS				
Within-subject: Time point	6.5	1.0	0.328	0.020
Time point x MRS (>0/≤0)	5.1	0.8	0.383	0.016
Between-subject: MRS (>0/≤0)	88.9	7.8	0.007	0.137
AOM				
Within-subject: Time point	6.5	1.0	0.328	0.022
Time point x AOM (>2/≤2)	1.0	0.1	0.701	0.003
Between-subject: AOM (>2/≤2)	196.5	21.3	<0.001	0.303
LS				
Within-subject: Time point	12.6	2.0	0.169	0.038
Time point x LS (>9.6/≤9.6kPa)	95.4	8.5	0.197	0.034
Between-subject: LS (>9.6/≤9.6kPa)	11.1	1.7	0.005	0.147
ELF				
Within-subject: Time point	4.4	0.7	0.423	0.013
Time point x ELF (>9.8/≤9.8)	0.2	0.0	0.864	0.001
Between-subject: ELF (>9.8/≤9.8)	66.9	5.6	0.022	0.103
1.5x ULN ALP				
Within-subject: Time point	6.4	1.0	0.333	0.019
Time point x ULN ALP (>1.5x/≤1.5x)	0.3	0.1	0.827	0.001
Between-subject: ULN ALP (>1.5x/≤1.5x)	30.8	2.4	0.124	0.048
Anali				
Within-subject: Time point	5.3	0.8	0.380	0.016
Time point x Anali (>2/≤2)	0.0	0.0	0.980	<0.001
Between-subject: Anali (>2/≤2)	70.3	6.0	0.018	0.108
ISSS				
Within-subject: Time point	6.3	0.9	0.336	0.019
Time point x ISSS (>2/≤2)	0.8	0.1	0.729	0.002
Between-subject: ISSS (>2/≤2)	163.2	16.5	<0.001	0.252
EHD				
Within-subject: Time point	4.8	0.7	0.395	0.015
Time point x EHD (Yes/No)	9.0	1.4	0.247	0.027
Between-subject: EHD (Yes/No)	20.4	1.6	0.213	0.032
DS				
Within-subject: Time point	4.3	0.6	0.425	0.013
Time point x DS (Yes/No)	0.7	0.1	0.752	0.002
Between-subject: DS (Yes/No)	39.9	3.2	0.079	0.062

MRS, Mayo Risk Score; AOM, Amsterdam-Oxford Model; ALP; alkaline phosphatase; LS; transient elastography liver stiffness; ELF, enhanced liver fibrosis score; ISSS, modified Amsterdam intrahepatic stricture severity score; EHD, presence of extrahepatic disease; DS, presence of dominant stricture.

7.6.3.2 IQR cT1

The mixed-model ANOVA results show that the main effect for time was not significant for all risk stratification methods except AOM. This indicated that there was no overall difference in the mean IQR cT1 between visits 1 and 2 within the low-risk or high-risk patient groups except when stratified by AOM.

A significant main effect for risk categorisation was obtained for all risk stratification methods except when stratified by 1.5x ULN ALP, presence of extrahepatic disease, and presence of dominant stricture. This result indicates that the mean IQR cT1 was higher in the high-risk groups stratified by MRS, AOM, LS, ELF, Anali, or ISSS at either time points.

Figure 69 shows a trend towards an increase in IQR cT1 between visit 1 and 2 in all the high-risk group but remains relatively the same or a decrease in the low-risk group between the two time points. However, results of the mixed-model ANOVA summarised in Table 64 show a significant interaction between time point and risk categorisation effect on IQR cT1 only when stratified by AOM ($p < 0.001$), LS ($p = 0.010$), ELF ($p = 0.004$), and ISSS ($p = 0.039$) with a large effect size for AOM ($\eta_p^2 = 0.228$) and ELF ($\eta_p^2 = 0.149$) and medium effect size for LS ($\eta_p^2 = 0.122$) and ISSS ($\eta_p^2 = 0.079$).

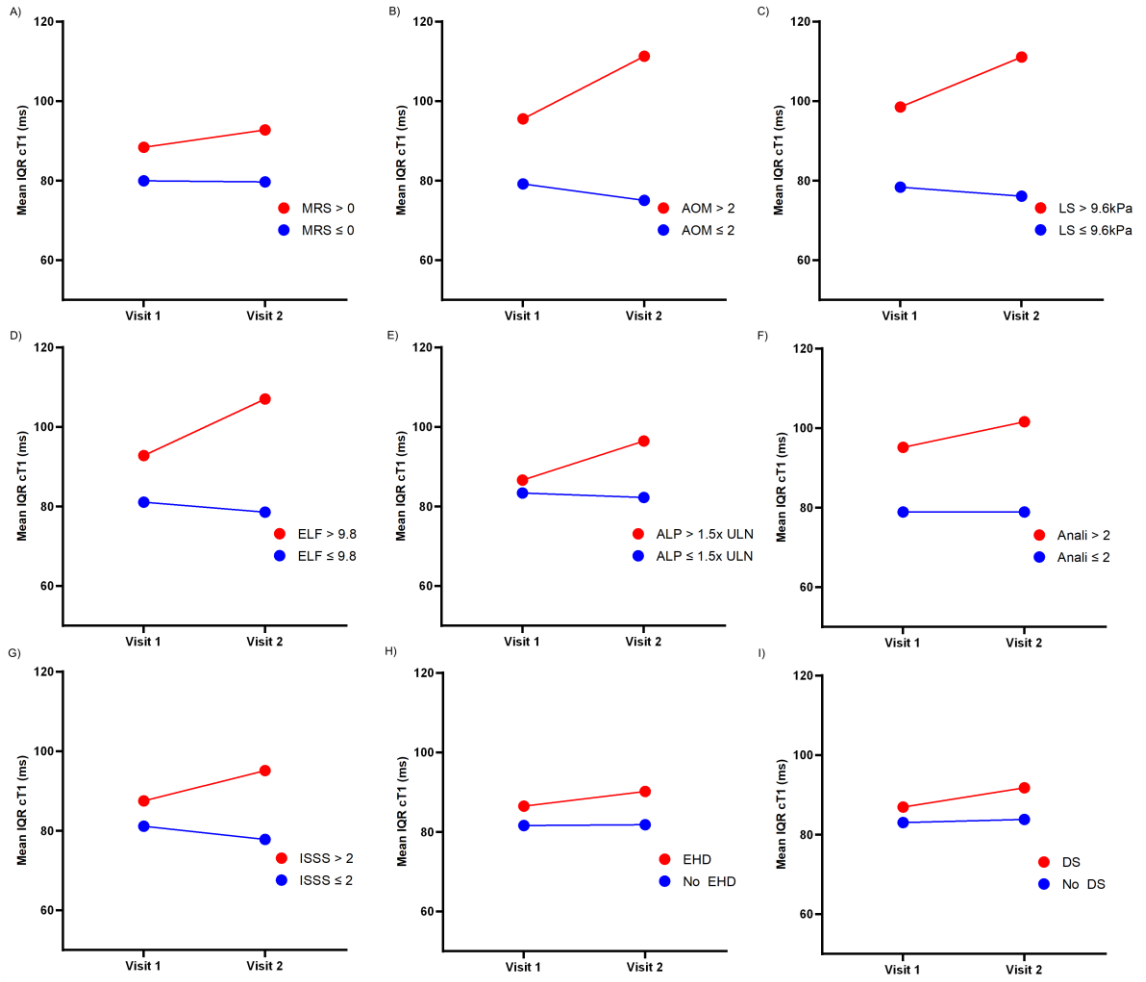


Figure 69. Differences in mean IQR cT1 between study visits 1 and 2 according to high-risk or low-risk group stratified by non-invasive surrogate markers. MRS, Mayo Risk Score; AOM, Amsterdam-Oxford Model; ALP; alkaline phosphatase; LS; transient elastography liver stiffness; ELF, enhanced liver fibrosis score; ISSS, modified Amsterdam intrahepatic stricture severity score; EHD, presence of extrahepatic disease; DS, presence of dominant stricture. Red line represents trend within the high-risk group and blue line low-risk group.

Table 64. Summary of mixed-model ANOVA results assessing the interaction of IQR cT1 with time point and risk categorisation stratified by non-invasive surrogate markers of disease severity.

	Mean square	F-ratio (F)	p-value	Effect size ($\eta_p^{2)}$
MRS				
Within-subject: Time point	112.6	0.6	0.447	0.011
Time point x MRS (>0/≤0)	144.3	0.8	0.390	0.014
Between-subject: MRS (>0/≤0)	3112.7	4.1	0.048	0.073
AOM				
Within-subject: Time point	787.7	5.2	0.026	0.091
Time point x AOM (>2/≤2)	2306.4	15.3	<0.001	0.228
Between-subject: AOM (>2/≤2)	16101.7	31.4	<0.001	0.377
LS				
Within-subject: Time point	597.2	3.5	0.067	0.063
Time point x LS (>9.6/≤9.6kPa)	1237.4	7.2	0.010	0.122
Between-subject: LS (>9.6/≤9.6kPa)	17084.7	34.6	<0.001	0.400
ELF				
Within-subject: Time point	739.9	4.5	0.039	0.079
Time point x ELF (>9.8/≤9.8)	1513.0	9.1	0.004	0.149
Between-subject: ELF (>9.8/≤9.8)	8735.7	13.4	0.001	0.204
1.5x ULN ALP				
Within-subject: Time point	426.8	2.3	0.132	0.043
Time point x ULN ALP (>1.5x/≤1.5x)	671.0	3.7	0.060	0.066
Between-subject: ULN ALP (>1.5x/≤1.5x)	1710.5	2.2	0.147	0.040
Anali				
Within-subject: Time point	247.0	1.3	0.260	0.024
Time point x Anali (>2/≤2)	251.3	1.3	0.255	0.025
Between-subject: Anali (>2/≤2)	9113.0	14.1	<0.001	0.213
ISSS				
Within-subject: Time point	122.5	0.7	0.413	0.013
Time point x ISSS (>2/≤2)	805.8	4.5	0.039	0.079
Between-subject: ISSS (>2/≤2)	3780.8	5.0	0.029	0.088
EHD				
Within-subject: Time point	100.1	0.5	0.475	0.010
Time point x EHD (Yes/No)	79.7	0.4	0.524	0.008
Between-subject: EHD (Yes/No)	1163.1	1.5	0.233	0.027
DS				
Within-subject: Time point	188.9	1.0	0.327	0.018
Time point x DS (Yes/No)	98.7	0.5	0.478	0.010
Between-subject: DS (Yes/No)	848.1	1.1	0.310	0.020

MRS, Mayo Risk Score; AOM, Amsterdam-Oxford Model; ALP; alkaline phosphatase; LS; transient elastography liver stiffness; ELF, enhanced liver fibrosis score; ISSS, modified Amsterdam intrahepatic stricture severity score; EHD, presence of extrahepatic disease; DS, presence of dominant stricture.

7.6.3.3 Mean superpixel variance cT1

The mixed-model ANOVA results show that the main effect for time was significant for all risk stratification methods confirming that there was an overall difference in the mean superpixel variance cT1 between visits 1 and 2 within the low-risk or high-risk patient group as discussed in section 7.6.2 above in the post-hoc subgroup analysis.

A significant main effect for risk categorisation was obtained for AOM, LS, ELF, and Anali. This indicates that the mean superpixel variance cT1 was higher in the high-risk groups stratified by AOM, LS, ELF, or Anali at either time points, but was not significantly different when stratified by MRS, 1.5x ULN ALP, ISSS, presence of extrahepatic disease, and presence of dominant stricture as depicted in Figure 70.

There were significant interactions between time point and risk categorisation on mean superpixel variance cT1 when stratified by AOM ($p < 0.001$, $\eta_p^2 = 0.246$), LS ($p = 0.037$, $\eta_p^2 = 0.082$), ELF ($p = 0.017$, $\eta_p^2 = 0.107$), and Anali ($p < 0.001$, $\eta_p^2 = 0.238$) as summarised in Table 65 .

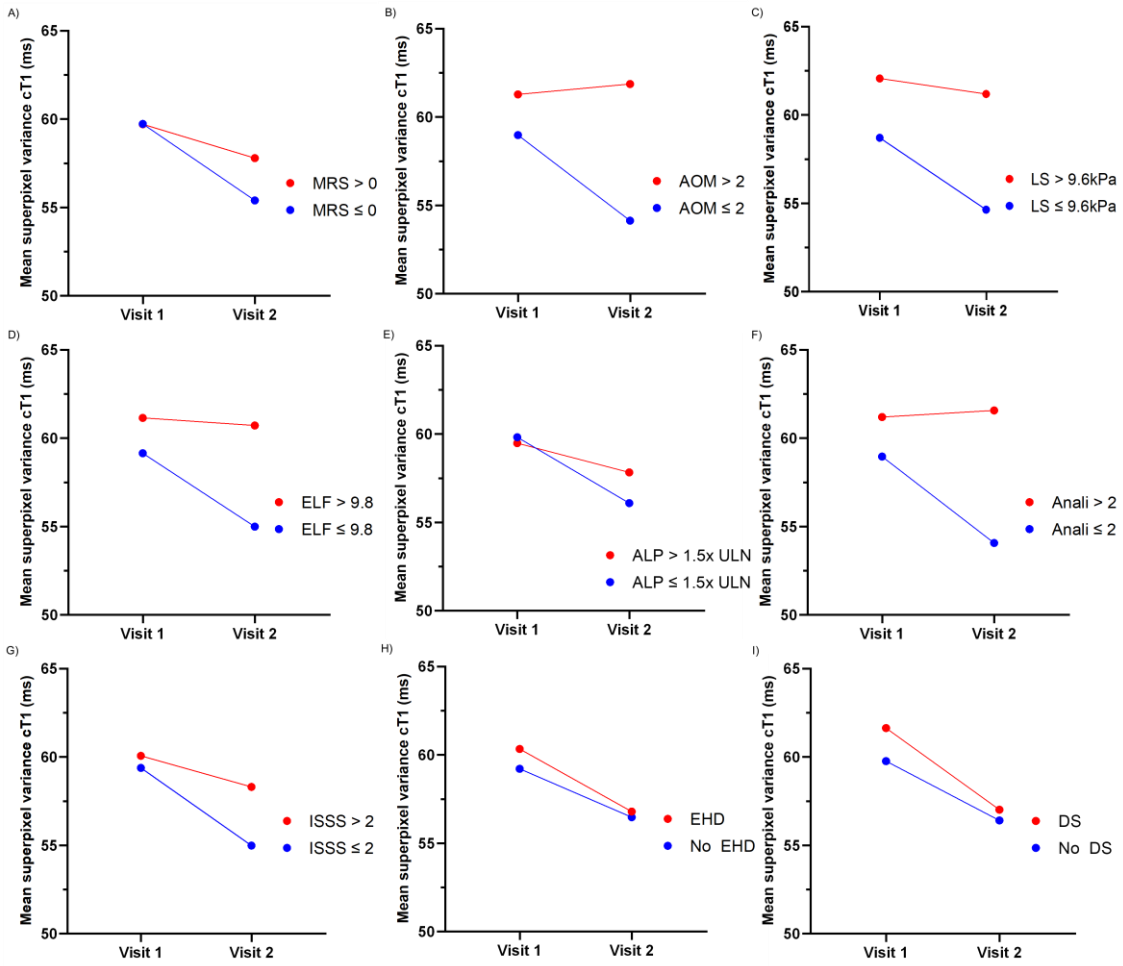


Figure 70. Differences in mean superpixel variance cT1 between study visits 1 and 2 according to high-risk or low-risk group stratified by non-invasive surrogate markers. MRS, Mayo Risk Score; AOM, Amsterdam-Oxford Model; ALP; alkaline phosphatase; LS; transient elastography liver stiffness; ELF, enhanced liver fibrosis score; ISSS, modified Amsterdam intrahepatic stricture severity score; EHD, presence of extrahepatic disease; DS, presence of dominant stricture. Red line represents trend within the high-risk group and blue line low-risk group.

Table 65. Summary of mixed-model ANOVA results assessing the interaction of mean superpixel variance cT1 with time point and risk categorisation stratified by non-invasive surrogate markers of disease severity.

	Mean square	F-ratio (F)	p-value	Effect size ($\eta_p^{2)}$
MRS				
Within-subject: Time point	258.6	20.2	<0.001	0.283
Time point x MRS (>0/≤0)	38.9	3.0	0.088	0.056
Between-subject: MRS (>0/≤0)	37.2	0.9	0.336	0.018
AOM				
Within-subject: Time point	104.6	10.2	0.002	0.167
Time point x AOM (>2/≤2)	170.2	16.6	<0.001	0.246
Between-subject: AOM (>2/≤2)	580.8	20.1	<0.001	0.283
LS				
Within-subject: Time point	136.2	10.9	0.002	0.176
Time point x LS (>9.6/≤9.6kPa)	56.9	4.6	0.037	0.082
Between-subject: LS (>9.6/≤9.6kPa)	548.7	18.6	<0.001	0.267
ELF				
Within-subject: Time point	113.3	9.3	0.004	0.155
Time point x ELF (>9.8/≤9.8)	74.4	6.1	0.017	0.107
Between-subject: ELF (>9.8/≤9.8)	320.3	9.4	0.003	0.156
1.5x ULN ALP				
Within-subject: Time point	162.0	12.4	0.001	0.195
Time point x ULN ALP (>1.5x/≤1.5x)	23.8	1.8	0.184	0.034
Between-subject: ULN ALP (>1.5x/≤1.5x)	11.0	0.3	0.603	0.005
Anali				
Within-subject: Time point	121.2	11.7	0.001	0.187
Time point x Anali (>2/≤2)	164.6	15.9	<0.001	0.238
Between-subject: Anali (>2/≤2)	562.2	19.2	<0.001	0.274
ISSS				
Within-subject: Time point	250.8	19.8	<0.001	0.279
Time point x ISSS (>2/≤2)	46.2	3.6	0.062	0.067
Between-subject: ISSS (>2/≤2)	105.8	2.8	0.102	0.052
EHD				
Within-subject: Time point	258.9	19.2	<0.001	0.273
Time point x EHD (Yes/No)	4.2	0.3	0.578	0.006
Between-subject: EHD (Yes/No)	13.7	0.3	0.560	0.007
DS				
Within-subject: Time point	211.7	15.7	<0.001	0.235
Time point x DS (Yes/No)	3.2	0.2	0.628	0.005
Between-subject: DS (Yes/No)	1.4	0.1	0.853	0.001

MRS, Mayo Risk Score; AOM, Amsterdam-Oxford Model; ALP; alkaline phosphatase; LS; transient elastography liver stiffness; ELF, enhanced liver fibrosis score; ISSS, modified Amsterdam intrahepatic stricture severity score; EHD, presence of extrahepatic disease; DS, presence of dominant stricture.

7.6.3.4 Mean periductal cT1 1.5-3.5mm

The mixed-model ANOVA results show that the main effect for time was not significant for all risk stratifications. This indicates that there was no overall difference in the mean periductal cT1 between visits 1 and 2 within the low-risk or high-risk patient groups.

A significant main effect for risk categorisation was obtained only for AOM and LS indicating that the mean periductal cT1 was higher in the AOM > 2 and LS > 9.6kPa compared to AOM ≤ 2 and LS ≤ 9.6kPa groups at both time points. Although Figure 71 shows a general trend of higher mean periductal cT1 in the high-risk groups, they were no statistically significant difference between the high-risk and low-risk groups at both time points.

Despite the general trend towards an increase in mean periductal cT1 between visits 1 and 2, the time and risk category interactions on mean periductal cT1 were not significant for any of the risk stratification methods as summarised in Table 66, indicating that any change in metrics observed was not dependent on time.

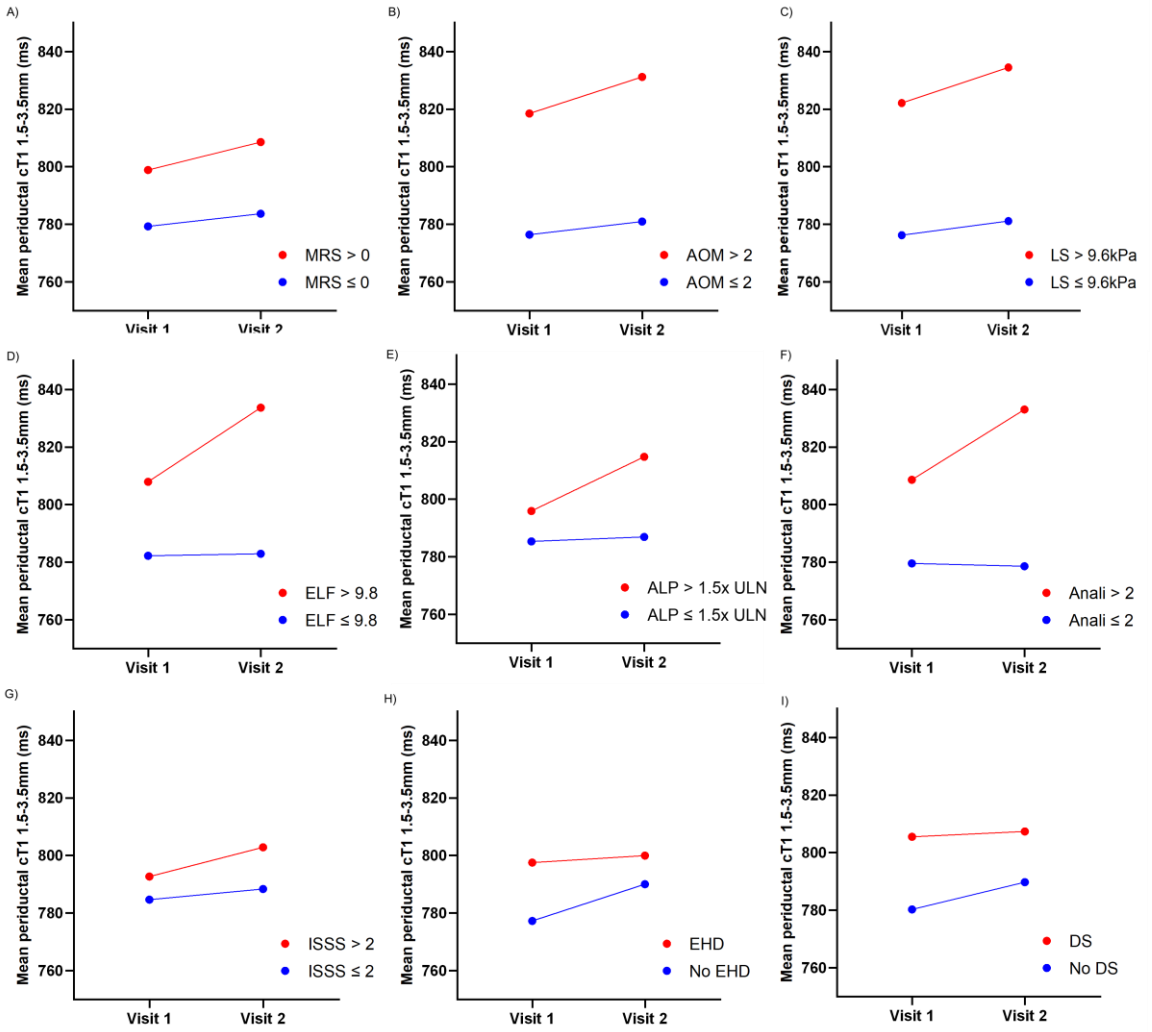


Figure 71. Differences in mean periductal cT1 1.5-3.5mm between study visits 1 and 2 according to high-risk or low-risk group stratified by non-invasive surrogate markers. MRS, Mayo Risk Score; AOM, Amsterdam-Oxford Model; ALP; alkaline phosphatase; LS; transient elastography liver stiffness; ELF, enhanced liver fibrosis score; ISSS, modified Amsterdam intrahepatic stricture severity score; EHD, presence of extrahepatic disease; DS, presence of dominant stricture. Red line represents trend within the high-risk group and blue line low-risk group.

Table 66. Summary of mixed-model ANOVA results assessing the interaction of mean periductal cT1 1.5–3.5mm with time point and risk categorisation stratified by non-invasive surrogate markers of disease severity.

	Mean square	F-ratio (F)	p-value	Effect size ($\eta_p^{2)}$
MRS				
Within-subject: Time point	1189.4	1.2	0.282	0.025
Time point x MRS (>0/≤0)	172.8	0.2	0.680	0.004
Between-subject: MRS (>0/≤0)	11840.6	1.5	0.226	0.032
AOM				
Within-subject: Time point	1482.0	1.5	0.230	0.031
Time point x AOM (>2/≤2)	334.2	0.3	0.566	0.007
Between-subject: AOM (>2/≤2)	42283.4	5.9	0.019	0.113
LS				
Within-subject: Time point	1415.1	1.4	0.241	0.030
Time point x LS (>9.6/≤9.6kPa)	262.5	0.3	0.611	0.006
Between-subject: LS (>9.6/≤9.6kPa)	46612.5	6.6	0.014	0.125
ELF				
Within-subject: Time point	3134.4	3.3	0.076	0.067
Time point x ELF (>9.8/≤9.8)	2823.3	3.0	0.091	0.061
Between-subject: ELF (>9.8/≤9.8)	26311.4	3.5	0.068	0.071
1.5x ULN ALP				
Within-subject: Time point	2137.1	2.2	0.146	0.045
Time point x ULN ALP (>1.5x/≤1.5x)	1544.7	1.6	0.215	0.033
Between-subject: ULN ALP (>1.5x/≤1.5x)	7588.6	1.0	0.334	0.020
Anali				
Within-subject: Time point	2822.0	3.0	0.089	0.061
Time point x Anali (>2/≤2)	3327.7	3.6	0.066	0.072
Between-subject: Anali (>2/≤2)	35889.7	4.9	0.032	0.096
ISSS				
Within-subject: Time point	1153.9	1.1	0.289	0.024
Time point x ISSS (>2/≤2)	253.0	0.3	0.618	0.005
Between-subject: ISSS (>2/≤2)	3028.1	0.4	0.543	0.008
EHD				
Within-subject: Time point	1356.2	1.4	0.249	0.029
Time point x EHD (Yes/No)	624.7	0.6	0.432	0.013
Between-subject: EHD (Yes/No)	5386.6	0.7	0.416	0.014
DS				
Within-subject: Time point	682.9	0.7	0.413	0.015
Time point x DS (Yes/No)	312.7	0.3	0.579	0.007
Between-subject: DS (Yes/No)	9810.3	1.2	0.271	0.026

MRS, Mayo Risk Score; AOM, Amsterdam-Oxford Model; ALP; alkaline phosphatase; LS; transient elastography liver stiffness; ELF, enhanced liver fibrosis score; ISSS, modified Amsterdam intrahepatic stricture severity score; EHD, presence of extrahepatic disease; DS, presence of dominant stricture.

7.6.4 Comparison of the diagnostic performance of candidate MRI metrics at baseline and their temporal change from baseline to detect the high-risk PSC group

The diagnostic performance of the change in candidate metric measurements between the two time points to detect the high-risk group, were compared with the diagnostic performance obtained from a single baseline measurement (Table 67Table 68). The diagnostic performances of single baseline measurements to detect the high-risk group were numerically better than the change between the two visits, except in one. The change in IQR cT1 between the two time points (AUC = 0.78, 95% CI: 0.61–0.95) had numerically a higher diagnostic performance than baseline value (AUC = 0.75, 95% CI 0.61–0.89; p = 0.0009) to detect the AOM > 2 group. There was, however, an overlap in confidence intervals and clinically not a significant difference. As for the other non-invasive tools, baseline measurements were again better than change over time (Table 69).

Table 67. Performance of intrahepatic SumRelSevDilat and IQR cT1 at baseline and their changes from baseline to discriminate high-risk vs low-risk PSC groups stratified by non-invasive surrogate markers.

	SumRelSevDilat (m^{-1})				IQR cT1 (ms)			
	Visit 1		Δ Visit 1–2		Visit 1		Δ Visit 1–2	
	AUC (95% CI)	p-value	AUC (95% CI)	p-value	AUC (95% CI)	p-value	AUC (95% CI)	p-value
MRS > 0 vs MRS ≤0	0.71 (0.57–0.86)	0.0088	0.56 (0.40–0.72)	0.4510	0.61 (0.46–0.77)	0.1508	0.58 (0.43–0.74)	0.2910
AOM >2 vs AOM ≤2	0.75 (0.62–0.88)	0.0042	0.55 (0.39–0.72)	0.5425	0.75 (0.61–0.89)	0.0033	0.78 (0.61–0.95)	0.0009
ULN ALP > 2.2x vs ≤2.2x	0.77 (0.59–0.95)	0.0231	0.55 (0.34–0.76)	0.6814	0.71 (0.53–0.89)	0.0756	0.71 (0.47–0.95)	0.0756
ULN ALP >1.5x vs ≤1.5x	0.56 (0.39–0.73)	0.4901	0.54 (0.38–0.70)	0.6699	0.63 (0.48–0.77)	0.1500	0.62 (0.45–0.78)	0.1817
LS >9.6kPa vs ≤9.6kPa	0.63 (0.46–0.80)	0.1495	0.64 (0.48–0.80)	0.1180	0.80 (0.67–0.94)	0.0005	0.70 (0.53–0.87)	0.0214
ELF >9.8 vs ≤9.8	0.67 (0.50–0.85)	0.0599	0.54 (0.36–0.72)	0.6345	0.66 (0.48–0.84)	0.0740	0.74 (0.57–0.91)	0.0073
Anali >2 vs Anali ≤2	0.66 (0.49–0.83)	0.0677	0.51 (0.33–0.68)	0.9514	0.77 (0.64–0.90)	0.0012	0.59 (0.41–0.77)	0.2914
ISSS >4 vs ≤4	0.71 (0.56–0.85)	0.0104	0.54 (0.38–0.71)	0.5978	0.64 (0.49–0.79)	0.0762	0.63 (0.47–0.78)	0.1076
EHD present vs absent	0.54 (0.38–0.70)	0.6226	0.61 (0.45–0.77)	0.1915	0.55 (0.39–0.71)	0.5309	0.59 (0.43–0.74)	0.2844
DS present vs absent	0.62 (0.46–0.78)	0.1742	0.56 (0.41–0.72)	0.4597	0.55 (0.39–0.72)	0.5387	0.53 (0.35–0.70)	0.7551

MRS, Mayo Risk Score; AOM, Amsterdam-Oxford Model; ALP; alkaline phosphatase; LS; transient elastography liver stiffness; ELF, enhanced liver fibrosis score; ISSS, modified Amsterdam intrahepatic stricture severity score; EHD, presence of extrahepatic disease; DS, presence of dominant stricture.

Table 68. Performance of mean superpixel variance cT1 and periductal cT1 1.5–3.5mm at baseline and their changes from baseline to discriminate high-risk vs low-risk PSC groups stratified by non-invasive surrogate markers.

	Mean superpixel variance cT1 (ms)				Periductal cT1 1.5–3.5mm			
	Visit 1		Δ Visit 1–2		Visit 1		Δ Visit 1–2	
	AUC (95% CI)	p-value	AUC (95% CI)	p-value	AUC (95% CI)	p-value	AUC (95% CI)	p-value
MRS > 0 vs MRS ≤0	0.52 (0.36–0.67)	0.8490	0.51 (0.35–0.68)	0.8770	0.56 (0.39–0.73)	0.4893	0.51 (0.34–0.68)	0.9096
AOM >2 vs AOM ≤2	0.68 (0.53–0.82)	0.0361	0.53 (0.34–0.71)	0.7855	0.70 (0.54–0.87)	0.0278	0.53 (0.34–0.71)	0.7855
ULN ALP > 2.2x vs ≤2.2x	0.59 (0.42–0.76)	0.4322	0.55 (0.34–0.76)	0.7083	0.52 (0.20–0.85)	0.8516	0.56 (0.36–0.76)	0.6401
ULN ALP >1.5x vs ≤1.5x	0.51 (0.35–0.67)	0.9396	0.58 (0.41–0.75)	0.3918	0.56 (0.38–0.74)	0.4975	0.58 (0.41–0.75)	0.4042
LS >9.6kPa vs ≤9.6kPa	0.72 (0.57–0.87)	0.0100	0.54 (0.36–0.72)	0.6678	0.71 (0.52–0.89)	0.0301	0.54 (0.36–0.72)	0.6847
ELF >9.8 vs ≤9.8	0.66 (0.50–0.82)	0.0710	0.67 (0.50–0.84)	0.0865	0.60 (0.39–0.81)	0.3059	0.67 (0.81–0.84)	0.0741
Anali >2 vs Anali ≤2	0.68 (0.53–0.84)	0.0290	0.64 (0.46–0.82)	0.1169	0.65 (0.48–0.82)	0.0975	0.64 (0.46–0.82)	0.1276
ISSS >2 vs ≤2	0.58 (0.42–0.73)	0.3199	0.52 (0.36–0.69)	0.7887	0.55 (0.39–0.72)	0.5362	0.52 (0.35–0.68)	0.8528
EHD present vs absent	0.57 (0.41–0.73)	0.3841	0.59 (0.42–0.75)	0.3036	0.54 (0.38–0.71)	0.6253	0.58 (0.42–0.74)	0.3443
DS present vs absent	0.50 (0.35–0.66)	0.9707	0.55 (0.35–0.74)	0.6150	0.52 (0.33–0.71)	0.8269	0.54 (0.35–0.73)	0.6150

MRS, Mayo Risk Score; AOM, Amsterdam-Oxford Model; ALP; alkaline phosphatase; LS; transient elastography liver stiffness; ELF, enhanced liver fibrosis score; ISSS, modified Amsterdam intrahepatic stricture severity score; EHD, presence of extrahepatic disease; DS, presence of dominant stricture.

Table 69. Performance of transient elastography liver stiffness (LS), enhanced liver fibrosis score (ELF), and serum alkaline phosphatase (ALP) at baseline and their changes from baseline to discriminate high-risk vs low-risk PSC groups stratified by non-invasive surrogate markers.

	LS (kPa)				ELF				ALP			
	Visit 1		Δ Visit 1–2		Visit 1		Δ Visit 1–2		Visit 1		Δ Visit 1–2	
	AUC (95% CI)	p-value	AUC (95% CI)	p-value	AUC (95% CI)	p-value	AUC (95% CI)	p-value	AUC (95% CI)	p-value	AUC (95% CI)	p-value
MRS > 0 vs ≤0	0.71 (0.27–0.84)	0.0094	0.60 (0.44–0.76)	0.2167	0.78 (0.65–0.91)	0.0005	0.54 (0.38–0.71)	0.6095	0.62 (0.47–0.77)	0.1192	0.64 (0.49–0.79)	0.0789
AOM >2 vs ≤2	0.85 (0.73–0.98)	<0.0001	0.66 (0.45–0.87)	0.0653	0.79 (0.64–0.95)	0.0006	0.57 (0.38–0.76)	0.4610	-	-	-	-
ULN ALP > 2.2x vs ≤2.2x	0.81 (0.65–0.97)	0.0086	0.54 (0.26–0.82)	0.7375	0.70 (0.45–0.94)	0.0917	0.57 (0.39–0.74)	0.6100	-	-	-	-
ULN ALP >1.5x vs ≤1.5x	0.73 (0.59–0.88)	0.0074	0.55 (0.36–0.74)	0.5653	0.69 (0.53–0.84)	0.0331	0.51 (0.34–0.69)	0.8836	-	-	-	-
LS >9.6kPa vs ≤9.6kPa	-	-	-	-	-	-	-	-	0.72 (0.57–0.87)	0.0100	0.56 (0.38–0.74)	0.5013
ELF >9.8 vs ≤9.8	-	-	-	-	-	-	-	-	0.71 (0.56–0.86)	0.0190	0.59 (0.42–0.76)	0.3199
Anali >2 vs ≤2	0.77 (0.63–0.91)	0.0015	0.65 (0.47–0.84)	0.0727	0.66 (0.49–0.83)	0.0575	0.56 (0.39–0.74)	0.4718	0.57 (0.42–0.73)	0.3986	0.56 (0.40–0.72)	0.4630
ISSS >4 vs ≤4	0.70 (0.56–0.84)	0.0127	0.56 (0.40–0.72)	0.4476	0.59 (0.44–0.75)	0.2360	0.68 (0.53–0.84)	0.0325	0.71 (0.57–0.85)	0.0090	0.56 (0.40–0.72)	0.4675
EHD present vs absent	0.58 (0.42–0.74)	0.3253	0.57 (0.41–0.72)	0.4036	0.56 (0.41–0.72)	0.4283	0.53 (0.37–0.70)	0.7001	0.55 (0.39–0.71)	0.5366	0.56 (0.40–0.71)	0.4862
DS present vs absent	0.70 (0.56–0.84)	0.0193	0.63 (0.45–0.81)	0.1397	0.55 (0.38–0.73)	0.5148	0.53 (0.35–0.74)	0.7398	0.67 (0.52–0.83)	0.0417	0.62 (0.45–0.80)	0.1497

MRS, Mayo Risk Score; AOM, Amsterdam-Oxford Model; ALP; alkaline phosphatase; LS; transient elastography liver stiffness; ELF, enhanced liver fibrosis score; ISSS, modified Amsterdam intrahepatic stricture severity score; EHD, presence of extrahepatic disease; DS, presence of dominant stricture.

7.6.5 Differences in patient reported outcomes between the two study visits

The mean domain score dropped between the two visits in three of the domains in SF-36: bodily pain (82.9% vs 73.5%, $p = 0.0050$), energy/fatigue (56.6% vs 47.2%, $p < 0.0001$), and mental summary score (71.4% vs 65.8%, $p = 0.0006$; Figure 72). There were no differences in any other domains in SF-36 or PSC-PRO between the two time points. There were no significant correlations between changes in mean score in any of the domains of SF-36 or PSC-PRO, and changes in candidate MRI metrics, ALP, LS, or ELF between the two time points.

When stratified by baseline LS risk category, patients with advanced fibrosis defined by LS > 9.6kPa had a drop in their mean domain score for physical functioning compared to a slight increase in those with LS < 9.6kPa (-9.0±16.5% vs 0.4±20.8%, $p = 0.0110$; Figure 73). In the PSC-PRO, mean domain score for emotional impact increased between the two visits in those with LS > 9.6kPa but slightly decreased in those with LS < 9.6kPa (0.6±0.7 vs -0.1±0.5, $p = 0.0097$; Figure 73). There were no significant differences in the other domains in either questionnaire.

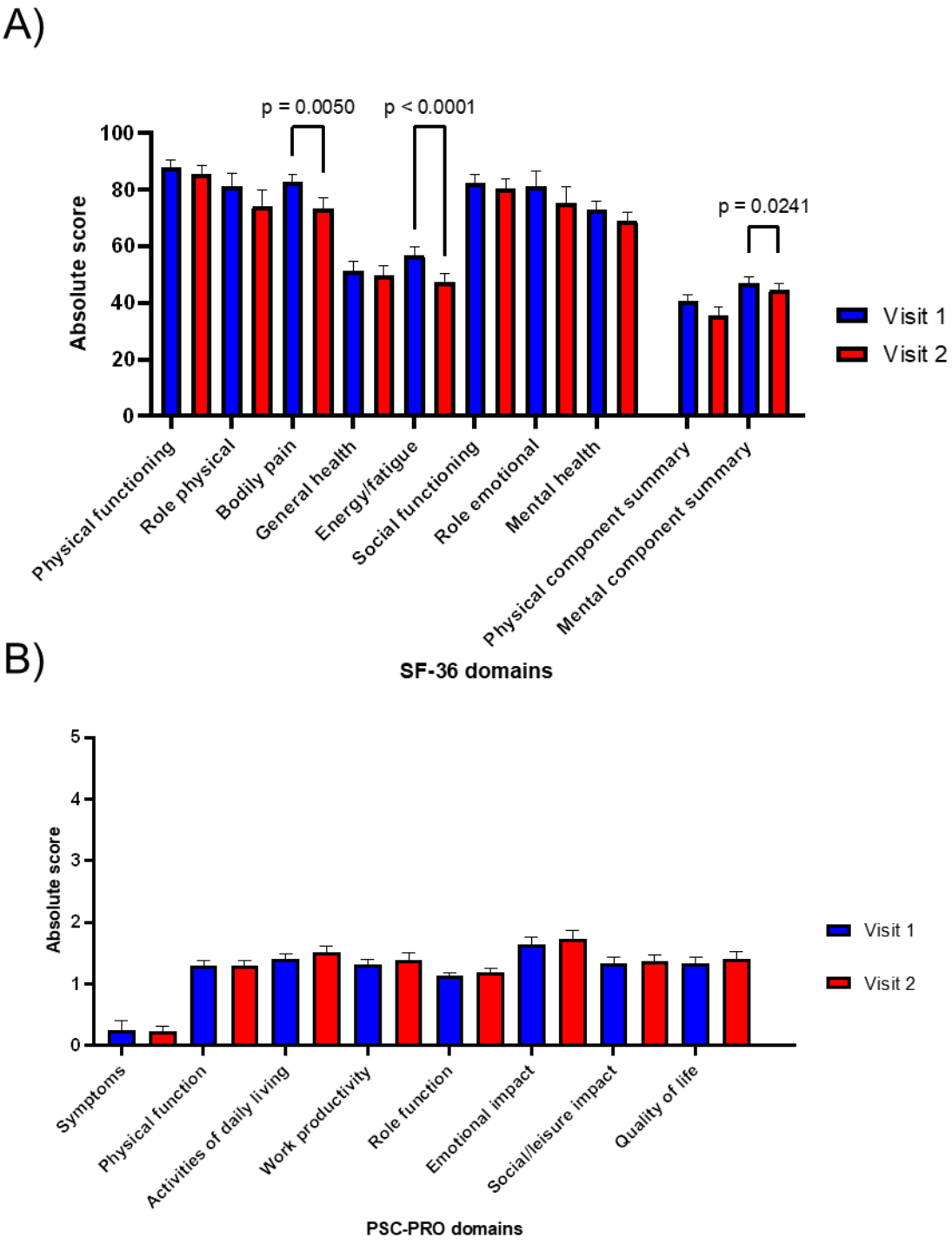


Figure 72. Differences in mean domain scores of (A) SF-36; and (B) PSC-PRO between visit 1 and 2. There were significant differences in bodily pain (p=0.0050), energy/fatigue (p<0.0001) and mental summary score (p=0.0241) domains of SF-36 between the two study visits. Lines and error bars indicate mean and standard error.

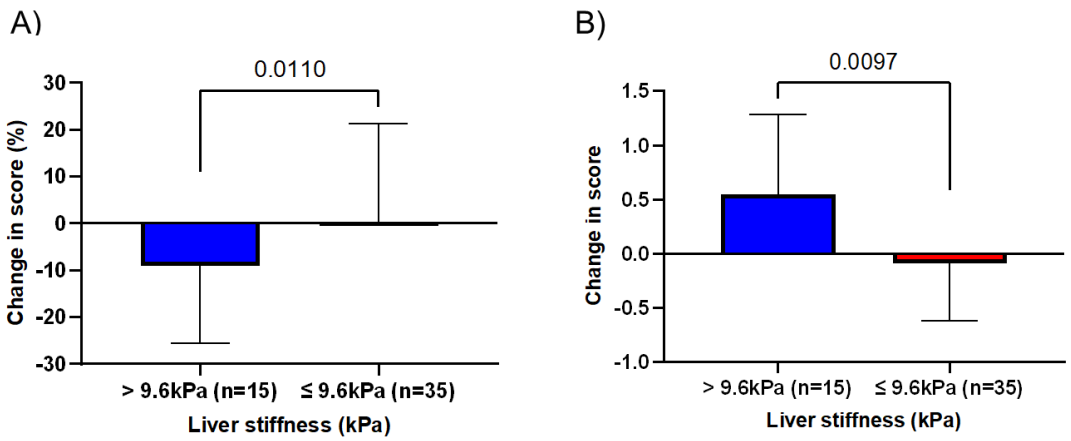


Figure 73. Changes in the mean score of (A) physical functioning domain of SF-36; and (B) emotional impact domain of PSC-PRO between the two study visits stratified by fibrosis severity. Lines and error bars indicate mean and standard deviation.

7.7 Discussion

This chapter assessed the relationship of candidate MRI metrics over two time points and its interaction with baseline risk categorisation by surrogate markers of disease severity. There were no significant changes within patients in all MRI metrics, except in mean superpixel variance cT1, which was significantly different in the low-risk PSC group but not changed in the high-risk group. The main effect of risk categorisation between patients confirmed earlier findings from baseline data that patients categorised into high-risk group by non-invasive surrogate markers of disease severity had higher values of candidate MRI metrics at both time points. This suggests that the quantitative techniques employed were stable over a year and not subjected to significant variability. Next, the interaction between the two time points and risk category on each of the four candidate metrics were assessed. The increase in mean IQR cT1 observed in the high-risk group with relative stability in the low-risk group was time dependent when stratified by AOM, LS, ELF, and ISSS. Similarly, the decrease in mean superpixel variance cT1 observed in the low-risk group with relative stability in the high-risk group was time dependent when stratified by AOM, LS, ELF, and Anali. Baseline candidate metrics had numerically better

performance than temporal change in metrics in detecting the high-risk group when stratified by surrogate markers of disease progression. This is probably because of only a short 1-year time lapse between assessments. Finally, analysis of QoL data between the two time points revealed worsening of bodily pain, levels of energy, and mental health. Patients with advanced fibrosis reported lower physical functioning and higher emotional impact of their disease compared to those without advanced fibrosis.

The most striking results from the mixed-effects ANOVA models are that changes in IQR cT1 is the result of the interaction between time and risk categorisation by AOM, LS, ELF, or ISSS. Descriptive statistics show an increase in mean IQR cT1 in the high-risk groups defined by these four risk categorisation over the two time points with measurements being relatively stable in their respective low-risk groups. These results suggest that patients who are at higher risk of liver decompensation (AOM), with advanced fibrosis (LS and ELF), or more severe intrahepatic strictures (ISSS) are likely to have a progressively higher IQR cT1 over time.

The only candidate metric observed to have significantly changed over the two time points in the whole cohort was the mean superpixel variance cT1. An overall decrease in the mean value between the two time points was unexpected as one would expect an increase in heterogeneity over time in PSC. When analysed by risk classification, there were no significant differences between the two time points in patients stratified into high-risk group at recruitment but there were significant differences in the low-risk group. The mixed-effects ANOVA models confirmed that changes in the mean superpixel variance cT1 is the result of the interaction between time and risk categorisation by AOM, LS, ELF, or Anali. There was a significant decrease in the mean superpixel variance cT1 in the low-risk group with relative stability in the high-risk group over the two time points when defined by AOM, LS, ELF or Anali. Following on from arguments made from chapter 5, this suggests that there was less heterogeneity in the liver parenchyma in the low-risk

group on the second visit. One possible explanation for this is the confounding effect of liver inflammation on cT1 measurements. Perhaps it can be speculated that patients in the low-risk group have a more inflammatory liver disease phenotype that waxes and wanes causing fluctuation in cT1 measurements whereas those in the high-risk group have developed fibrotic liver disease that was relatively stable over a course of a year. The other possibility here is that the superpixel analysis technique itself warrants further technical validation and may have limitations in repeat measurements in the same patients. Current comparison method assumes selection of the same representative slice of liver. In the future, careful alignment of both segmented livers from the two time points is required first before direct comparison of regional changes could be made. Additional time points in the future will be able to clarify this further.

Temporal change in candidate MRI metrics between two time points did not confer additional benefit to baseline values alone in detecting the high-risk group defined by surrogate markers of disease severity. This needs to be interpreted in the context of limited time points within a relatively short space of time for a slowly progressive disease. It is hoped that with additional time points and larger intervals between measurements, the progression rate in candidate metrics per year can be estimated with repeated-measures analysis using linear mixed model.

Patient-reported outcomes between the two time points numerically were lower in all the domains of SF-36 with statistically significant difference reached in three of the domains (bodily pain, energy/fatigue, mental component summary). Although fibrosis stage was no different between the two time points as would be expected, patients in the advanced fibrosis group were observed to have reported lower physical functioning and higher emotional impact of their disease compared to those in the low-risk group. These findings, however, need to be interpreted in the context of imposed restrictions during the Covid-19 pandemic which may have had a significant psycho-social impact on the patients.

In conclusion, temporal changes in intrahepatic SumRelSevDilat and mean periductal cT1 were driven by risk classification and is independent of time. Temporal changes in mean IQR cT1 were dependent on time when stratified by AOM, LS, ELF, and ISSS. Temporal changes in mean superpixel variance cT1 were dependent on time when stratified by AOM, LS, ELF, and Anali. Baseline candidate MRI metrics outperformed temporal changes from baseline values in detecting the high-risk PSC group. There was deterioration in bodily pain, energy levels, and mental health in patients between the two study visits, with worsening physical functioning and emotional impact of disease in those with advanced fibrosis.

8 General discussion and thesis conclusion

8.1 Summary of main findings

This thesis aimed to improve the non-invasive assessment of PSC using quantitative MRI techniques. It addressed some of the research priorities that have been identified by the MRI working group of the International PSC Study Group in their position statement[29]. Two standardised quantitative MRI techniques were used to assess the three important aspects of the disease. The primary pathology affecting the biliary tree causing strictures and dilatations was assessed quantitatively for the first time in PSC using *MRCP+*. The multifocal accumulation of fibrosis in the liver parenchyma as a result was assessed using cT1 from Liver*MultiScan*, a metric that has been validated in other liver diseases but not specifically in PSC. A novel method that combined both *MRCP+* and Liver*MultiScan* techniques to derive periductal cT1, allowed quantitative assessment specifically of periductular abnormalities, the hallmark of disease in PSC. The performance of candidate metrics was compared against healthy controls. Liver*MultiScan* data from previous studies provided a validation cohort of other liver disease to compare liver cT1 metrics derived from PSC. In the absence of adequate clinical endpoints from a limited follow-up period, the risk stratification performance of metrics derived from these techniques were assessed against several published surrogate markers of disease severity and progression in PSC. Finally, candidate MRI metrics within the same patients were assessed over two time points and the interaction of the effect of time and baseline risk classification on candidate MRI metrics examined concurrently.

The most promising *MRCP+* metric was a metric that quantified the sum of relative severity of candidate dilatations within the liver. This metric demonstrated significant correlation with important non-invasive biomarkers in PSC and showed promising risk-stratification ability when stratified into high-risk and low-risk groups using prognostic risk models, liver fibrosis markers, serum ALP, conventional MRI-MRCP scores and

morphological features. It also showed better inter-observer agreement than qualitative MRI-MRCP scores and morphological features. In the multivariate model, this metric was found to be an independent predictor of high-risk PSC when stratified by MRS. Stricture metrics differentiated PSC from healthy and showed significant correlation with some of the biomarkers in PSC. However, contrary to the hypothesis, the performance of candidate stricture metrics was poor overall and lacked the ability to identify the high-risk group except when stratified by disease distribution, highlighting perhaps the ongoing limitations with MRCP techniques.

One of the challenges in assessing the liver parenchyma in PSC is the patchy nature of how the disease affects the liver, particularly in the early stages. Two methods of cT1 distribution analysis carried out in this thesis have attempted to quantify this heterogeneity in the liver. Histogram analysis from the whole liver segmentation identified IQR cT1 and regional variance analysis identified mean superpixel variance cT1 as potential markers. IQR cT1 showed better risk stratification ability between high-risk and low-risk PSC but may be susceptible to the global severity of liver fibrosis in chronic liver disease rather than quantifying early patchy changes in PSC. On the contrary, mean superpixel variance cT1 appears to be uniquely elevated in PSC and was found to be independent of the severity of liver fibrosis.

A quantitative method of exploring periductal abnormalities is presented in this thesis using a co-registration technique of *MRCP*+ 3D model and cT1 map obtained from Liver*MultiScan* analysis. The mean cT1 in the surrounding regions closer to the bile ducts were higher than regions further away from the bile ducts, even after correcting for partial volume effect from co-registration technique. The same phenomenon was not seen in healthy controls. This results mimic what would be expected of fibrosis accumulation in PSC. Periductal mean cT1 showed correlation and risk stratification ability with liver fibrosis markers and AOM, but not with cholestasis markers or MRS. When stratified by

whole liver IQR cT1, patients with higher heterogeneity had higher periductal mean cT1 confirming that periductal changes contribute to the patchy heterogeneity seen in the liver.

The final aspect of this thesis examined the temporal relationship between the candidate metrics identified in the three chapters. This piece of work has been affected by the Covid-19 pandemic which hampered enthusiasm for participants to attend follow-up visits. Nevertheless, data was available from 55 participants and provided some early trends in the behaviour of candidate metrics across two time points. When examined across the whole cohort, all candidate metrics except one remained relatively stable with no significant difference across the two time points. Further analysis confirmed that the change observed in the mean superpixel variance cT1 was unique to the low-risk group. The explanation for this remains speculative at this stage and it is hoped that future data points may elucidate a less speculative reason than what is currently proposed. The effect of time and risk categorisation on candidate MRI metrics were explored, with the most notable finding was for mean IQR cT1. Compared to their respective low-risk groups, participants who are at higher risk of liver decompensation identified by AOM, or with advanced fibrosis as defined by LS and ELF, or more severe intrahepatic strictures quantified by ISSS, are likely to have a progressively higher IQR cT1 over time. Single baseline measurements alone outperformed dynamic measurements over two time points in detecting the high-risk group. However, further follow-up data points with repeated-measures analysis and correlation with clinical outcomes are needed before concrete conclusions can be drawn from this.

Lastly, this thesis captured patient-reported outcomes using SF-36 and PSC-PRO questionnaires to assess the correlation between quality of life and non-invasive markers of disease severity including quantitative MRI metrics. *MRCP*+-derived dilatation metrics showed correlation with the physical components of the SF-36. Median cT1 and LS correlated with both the physical and mental components of SF-36, with patients with advanced fibrosis defined by LS > 9.6kPa having a poorer quality of life at recruitment and

at 1-year follow-up. The lack of correlation of PSC-PRO with any of the non-invasive markers or MRI metrics perhaps reflect the more acute phase reporting of symptoms compared to SF-36. Nonetheless, there was a significant difference in reporting of emotional impact of disease from the PSC-PRO domain in patients with advanced fibrosis over the course of a year.

8.2 Implications of thesis findings

The clinical translation of these quantitative techniques will depend on the performance of these metrics in larger cohorts of patients with PSC from different centres and their validation against clinical endpoints. These metrics could have a significant impact on how we diagnose, risk stratify and monitor patients with PSC. The standardised MRI-MRCP acquisition protocol used in this thesis allows simultaneous quantitative assessment of the biliary tree, global liver parenchymal, and periductal abnormalities. Staging of cholangiopathy and liver fibrosis could be performed in the same sitting with the added benefit of providing prognostic information. MRI could negate the additional need to perform transient elastography given its ability to stage fibrosis with good diagnostic accuracy against liver stiffness measurements in this study. Moreover, this imaging protocol also allows for qualitative assessment as currently performed in clinical practice. Contrast administration could be added for characterisation of liver lesion if required. It is anticipated that these quantitative techniques are not likely to replace, rather improve current assessment by amalgamating qualitative and quantitative imaging outcomes.

One of the major potential impacts that these metrics could have is in interventional trial designs. Several reasons have hampered the development of effective medical therapies in clinical trials to date including inadequate statistical power, slow and heterogeneous disease progression, and lack of true surrogate biomarkers of disease severity. Despite being the cornerstone of diagnosis and follow-up in PSC, MRI-MRCP have not been

considered as surrogate endpoints in trials due to significant variability in imaging protocols and reporting. Therefore, trials have tended to rely on more objective measurements that can quantify change such as serum ALP, transient elastography, ELF and/or liver biopsy, despite their limitations as surrogate endpoints. With the development of a standardised imaging protocol and post-processing techniques such as *MRCP+* and Liver*MultiScan*, there is now an opportunity to include imaging biomarkers as additional surrogate endpoints in trials. If validated against clinical endpoints, it provides trialists with an alternative non-invasive, objective, and reliable modality to investigate the effect of their interventions. As an example, trials targeting fibrotic pathways specifically around bile ducts in early stages of PSC could use periductal cT1 either to determine entry criteria or assess treatment response without the need for repeated liver biopsies.

Increasingly, patients are being empowered to be actively involved in monitoring their disease. In a condition where there is no cure such as PSC, it is even more important to consider the principles of value-based medicine by involving patients in decision regarding which aspects of their disease should be captured by novel therapies or biomarkers. The patient and public involvement workshop held at the conception of this study and speaking to patients throughout the study recruitment process, revealed that patients with PSC felt that the prospect of being able to 'visualise' their disease by tracking quantitative changes in their biliary and liver metrics against changes in their biochemical parameters and symptoms would give them a sense of being able to have some degree of control on their otherwise dismal disease.

8.3 Limitations

Technical limitations particular to each chapter was discussed within the discussion section of each chapter. Limitations related to the overall design and outcome of the study are discussed here.

This study was an exploratory, proof-of-concept, single-centre study performed in a tertiary non-transplant hospital to identify candidate metrics. Therefore, the generalisability of the results presented here is limited.

Clinical outcomes are slow to develop in PSC. Published studies often relied on retrospective analysis of prospectively collected longitudinal data from routine clinical follow-up to validate their findings against clinical endpoints. Whilst *MRCP+* allows retrospective analysis of 3D MRCP scans, Liver*MultiScan* sequences needed to be acquired prospectively. Retrospective MRCP scans of the Oxford PSC cohort were obtained for *MRCP+* analysis, however, the heterogeneity in scan protocol, image quality, and incompatibility of the algorithm with 2D MRCP scans prevented this. The absence of adequate clinical endpoints from limited follow-up period meant that the predictive capability of candidate biomarkers could not be evaluated fully. Instead, risk stratification in this study relied on surrogate biomarkers that have been validated against clinical endpoints in other studies. Whilst it is reassuring that candidate imaging biomarkers from this study are comparable to the performance of other non-invasive markers, these surrogate biomarkers themselves are imperfect, and none have exceeded level 3 validation in the Fleming hierarchy of surrogate endpoints[176]. In essence, the performance of the quantitative metrics was judged based on their ability to be a surrogate of imperfect surrogate markers. Therefore, it is possible that metrics that may have otherwise be identified as clinically important could have been excluded. On the other hand, given the exploratory nature of this study, multiple metrics were tested. Some findings may be subject to chance, and therefore require validation in further studies. Longitudinal data will enable comparative assessments between all non-invasive biomarkers against clinical outcomes.

The cost-effectiveness and capital expenditure of quantitative techniques versus conventional qualitative approach to image analysis was beyond the scope of this study. Nevertheless, 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+*.

8.4 Suggested further work

The MoBILISE study is still in its follow-up phase for up to 15 years to assess the association between candidate metrics reported in this thesis with clinical outcomes. Cox regression analyses and Kaplan-Meier curves will assess their ability to predict endpoints in PSC. The value of measuring serial candidate metrics over more time points could be explored for monitoring of disease progression and prognostication. Given the heterogeneity of the disease, external validation of the candidate metrics identified in this thesis is warranted in other cohorts of patients with PSC via a multicentre study. Suggested work specific to each quantitative MRI technique is discussed below.

Given the impact of acquisition technique on *MRCP+* analysis, further work could also explore optimising MRCP techniques with significantly shorter acquisition protocols using newer methods such as single breath-hold 3D MRCP and/or 3D breath-holding MRCP with compressed sensing technique[177-180]. The approach to the visualisation and assessment of the biliary tree is considerably different between MRCP and *MRCP+*. Nonetheless, the utility of *MRCP+* in the hands of radiologists is currently unknown. It would be important to assess the complementary value of performing *MRCP+* analysis by performing a head-to-head comparison in different hepatopancreatobiliary conditions to facilitate buy-in from radiologists. Collaborations with other centres that have retrospective 3D MRCP database would allow quicker validation of *MRCP+* metrics with clinical endpoints. It could also allow development of machine-learning techniques to automate disease-specific outputs from *MRCP+*.

The main competing technique for MRI with Liver*MultiScan* is MRE, a technique that has been validated against clinical outcomes in PSC and have shown promising predictive

capability. Future studies should consider a head-to-head comparison of these two MR techniques for the assessment of liver parenchyma in patients with PSC. The advantage of performing MRI with Liver*MultiScan* analysis is that no additional equipment to the MRI scanner is required, shorter breath holds that patients are more likely to tolerate, lower technical failure rates, and allows simultaneous assessment of liver lesions.

Instead of acquiring only four slices, 3D T1 mapping technique would acquire a greater proportion of the liver and provide a better representation of the T1 distribution in the liver of PSC, albeit a significantly longer acquisition time. 3D T1 maps could be co-registered with 3D *MRCP+* model using the techniques described in this thesis to allow a more comprehensive assessment of the periductal changes in PSC.

8.5 Conclusion

This thesis has identified four candidate imaging biomarkers derived from three quantitative MRI and MRCP techniques that did not require contrast administration or additional external equipment. Quantitative metrics showed better inter-observer agreement than quantitative scores. These metrics have shown significant correlation with important biomarkers in PSC and promising ability to detect the high-risk PSC patients when stratified by published surrogate markers of clinical endpoints. Longitudinal follow-up data from this study will enable their validation against clinical outcomes. External validation of these metrics is warranted before consideration can be given for its application in clinical and trial setting.

References

1. Hirschfield GM, Karlsen TH, Lindor KD, Adams DH. Primary sclerosing cholangitis. Lancet 2013;382:1587-99.
2. Boonstra K, Weersma RK, van Erpecum KJ, Rauws EA, Spanier BW, Poen AC, et al. Population-based epidemiology, malignancy risk, and outcome of primary sclerosing cholangitis. Hepatology 2013;58:2045-55.
3. Trivedi PJ, Bowlus CL, Yimam KK, Razavi H, Estes C. Epidemiology, Natural History, and Outcomes of Primary Sclerosing Cholangitis: A Systematic Review of Population-based Studies. Clin Gastroenterol Hepatol 2021.
4. Schrumpf E, Elgjo K, Fausa O, Gjone E, Kolmannskog F, Ritland S. Sclerosing cholangitis in ulcerative colitis. Scand J Gastroenterol 1980;15:689-97.
5. Chapman RW, Arborgh BA, Rhodes JM, Summerfield JA, Dick R, Scheuer PJ, et al. Primary sclerosing cholangitis: a review of its clinical features, cholangiography, and hepatic histology. Gut 1980;21:870-7.
6. Wiesner RH, LaRusso NF. Clinicopathologic features of the syndrome of primary sclerosing cholangitis. Gastroenterology 1980;79:200-6.
7. European Association for the Study of the Liver (EASL). Clinical Practice Guidelines: management of cholestatic liver diseases. J Hepatol 2009;51:237-67.
8. Angulo P, Maor-Kendler Y, Lindor KD. Small-duct primary sclerosing cholangitis: a long-term follow-up study. Hepatology 2002;35:1494-500.
9. Kaya M, Angulo P, Lindor KD. Overlap of autoimmune hepatitis and primary sclerosing cholangitis: an evaluation of a modified scoring system. J Hepatol 2000;33:537-42.
10. Mendes FD, Jorgensen R, Keach J, Katzmann JA, Smyrk T, Donlinger J, et al. Elevated serum IgG4 concentration in patients with primary sclerosing cholangitis. Am J Gastroenterol 2006;101:2070-5.
11. Bjornsson E, Chari S, Silveira M, Gossard A, Takahashi N, Smyrk T, et al. Primary sclerosing cholangitis associated with elevated immunoglobulin G4: clinical characteristics and response to therapy. Am J Ther 2011;18:198-205.
12. Benito de Valle M, Muller T, Bjornsson E, Otten M, Volkmann M, Guckelberger O, et al. The impact of elevated serum IgG4 levels in patients with primary sclerosing cholangitis. Dig Liver Dis 2014;46:903-8.
13. Lazaridis KN, LaRusso NF. Primary Sclerosing Cholangitis. N Engl J Med 2016;375:1161-70.
14. Support P. PSC Support Patient Unmet Needs and Research Implications: pscsupport.org.uk, 2016.
15. American Society for Gastrointestinal Endoscopy (ASGE) Standards of Practice Committee, Chandrasekhara V, Khashab MA, Muthusamy VR, Acosta RD, Agrawal D, et al. Adverse events associated with ERCP. Gastrointest Endosc 2017;85:32-47.
16. Dave M, Elmunzer BJ, Dwamena BA, Higgins PD. Primary sclerosing cholangitis: meta-analysis of diagnostic performance of MR cholangiopancreatography. Radiology 2010;256:387-96.
17. Chapman R, Fevery J, Kalloo A, Nagorney DM, Boberg KM, Shneider B, et al. Diagnosis and management of primary sclerosing cholangitis. Hepatology 2010;51:660-78.
18. Talwalkar JA, Angulo P, Johnson CD, Petersen BT, Lindor KD. Cost-minimization analysis of MRC versus ERCP for the diagnosis of primary sclerosing cholangitis. Hepatology 2004;40:39-45.
19. Meagher S, Yusoff I, Kennedy W, Martel M, Adam V, Barkun A. The roles of magnetic resonance and endoscopic retrograde cholangiopancreatography (MRCP and ERCP) in the diagnosis of patients with suspected sclerosing cholangitis: a cost-effectiveness analysis. Endoscopy 2007;39:222-8.

20. Fulcher AS, Turner MA, Franklin KJ, Shiffman ML, Sterling RK, Luketic VA, et al. Primary sclerosing cholangitis: evaluation with MR cholangiography-a case-control study. Radiology 2000;215:71-80.
21. Angulo P, Pearce DH, Johnson CD, Henry JJ, LaRusso NF, Petersen BT, et al. Magnetic resonance cholangiography in patients with biliary disease: its role in primary sclerosing cholangitis. J Hepatol 2000;33:520-7.
22. Wallner BK, Schumacher KA, Weidenmaier W, Friedrich JM. Dilated biliary tract: evaluation with MR cholangiography with a T2-weighted contrast-enhanced fast sequence. Radiology 1991;181:805-8.
23. Goode EC, Clark AB, Mells GF, Srivastava B, Spiess K, Gelson WTH, et al. Factors Associated With Outcomes of Patients With Primary Sclerosing Cholangitis and Development and Validation of a Risk Scoring System. Hepatology 2019;69:2120-2135.
24. Tokala A, Khalili K, Menezes R, Hirschfield G, Jhaveri KS. Comparative MRI analysis of morphologic patterns of bile duct disease in IgG4-related systemic disease versus primary sclerosing cholangitis. AJR Am J Roentgenol 2014;202:536-43.
25. Ohara H, Okazaki K, Tsubouchi H, Inui K, Kawa S, Kamisawa T, et al. Clinical diagnostic criteria of IgG4-related sclerosing cholangitis 2012. J Hepatobiliary Pancreat Sci 2012;19:536-42.
26. Parlak E, Disibeyaz S, Odemis B, Koksal AS, Oguz D, Cicek B, et al. Demonstration of retraction of the main papilla toward the biliary system in patients with primary sclerosing cholangitis with magnetic resonance cholangiopancreatography. Dig Endosc 2012;24:384.
27. Kovac JD, Weber MA. Primary Biliary Cirrhosis and Primary Sclerosing Cholangitis: an Update on MR Imaging Findings with Recent Developments. J Gastrointestin Liver Dis 2016;25:517-524.
28. Kovac JD, Jesic R, Stanisavljevic D, Kovac B, Maksimovic R. MR imaging of primary sclerosing cholangitis: additional value of diffusion-weighted imaging and ADC measurement. Acta Radiol 2013;54:242-8.
29. Schramm C, Eaton J, Ringe KI, Venkatesh S, Yamamura J, IPSCSG MRIwg. Recommendations on the use of magnetic resonance imaging in PSC-A position statement from the International PSC Study Group. Hepatology 2017;66:1675-1688.
30. Stiehl A, Rudolph G, Kloters-Plachky P, Sauer P, Walker S. Development of dominant bile duct stenoses in patients with primary sclerosing cholangitis treated with ursodeoxycholic acid: outcome after endoscopic treatment. J Hepatol 2002;36:151-6.
31. Chapman MH, Thorburn D, Hirschfield GM, Webster GGJ, Rushbrook SM, Alexander G, et al. British Society of Gastroenterology and UK-PSC guidelines for the diagnosis and management of primary sclerosing cholangitis. Gut 2019;68:1356-1378.
32. Chapman MH, Webster GJ, 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;24:1051-8.
33. Chapman RW, Williamson KD. Are Dominant Strictures in Primary Sclerosing Cholangitis a Risk Factor for Cholangiocarcinoma? Curr Hepatol Rep 2017;16:124-129.
34. Lundberg Bave A, Bergquist A, Bottai M, Warnqvist A, von Seth E, Nordenvall C. Increased risk of cancer in patients with primary sclerosing cholangitis. Hepatol Int 2021;15:1174-1182.
35. Liang H, Manne S, Shick J, Lissoos T, Dolin P. Incidence, prevalence, and natural history of primary sclerosing cholangitis in the United Kingdom. Medicine (Baltimore) 2017;96:e7116.

36. Trivedi PJ, Crothers H, Mytton J, Bosch S, Iqbal T, Ferguson J, et al. Effects of Primary Sclerosing Cholangitis on Risks of Cancer and Death in People With Inflammatory Bowel Diseases, Based on Sex, Race, and Age. Gastroenterology 2020.
37. Weismuller TJ, Trivedi PJ, Bergquist A, Imam M, Lenzen H, Ponsioen CY, et al. Patient Age, Sex, and Inflammatory Bowel Disease Phenotype Associate With Course of Primary Sclerosing Cholangitis. Gastroenterology 2017;152:1975-1984 e8.
38. Petersen-Benz C, Stiehl A. Impact of dominant stenoses on the serum level of the tumor marker CA19-9 in patients with primary sclerosing cholangitis. Z Gastroenterol 2005;43:587-90.
39. Levy MJ, Baron TH, Clayton AC, Enders FB, Gostout CJ, Halling KC, et al. Prospective evaluation of advanced molecular markers and imaging techniques in patients with indeterminate bile duct strictures. Am J Gastroenterol 2008;103:1263-73.
40. Razumilava N, Gores GJ, Lindor KD. Cancer surveillance in patients with primary sclerosing cholangitis. Hepatology 2011;54:1842-52.
41. Ali AH, Tabibian JH, Nasser-Ghodsi N, Lennon RJ, DeLeon T, Borad MJ, et al. Surveillance for hepatobiliary cancers in patients with primary sclerosing cholangitis. Hepatology 2018;67:2338-2351.
42. Bowlus CL, Lim JK, Lindor KD. AGA Clinical Practice Update on Surveillance for Hepatobiliary Cancers in Patients With Primary Sclerosing Cholangitis: Expert Review. Clin Gastroenterol Hepatol 2019;17:2416-2422.
43. Goode EC, Rushbrook SM. A review of the medical treatment of primary sclerosing cholangitis in the 21st century. Ther Adv Chronic Dis 2016;7:68-85.
44. Liu K, Wang R, Kariyawasam V, Wells M, Strasser SI, McCaughan G, et al. Epidemiology and outcomes of primary sclerosing cholangitis with and without inflammatory bowel disease in an Australian cohort. Liver Int 2017;37:442-448.
45. Andersen IM, Fosby B, Boberg KM, Clausen OP, Jebsen P, Melum E, et al. Indications and Outcomes in Liver Transplantation in Patients With Primary Sclerosing Cholangitis in Norway. Transplant Direct 2015;1:e39.
46. 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;51:562-6.
47. Tischendorf JJ, Hecker H, Kruger M, Manns MP, Meier PN. Characterization, outcome, and prognosis in 273 patients with primary sclerosing cholangitis: A single center study. Am J Gastroenterol 2007;102:107-14.
48. Bjornsson E, Boberg KM, Cullen S, Fleming K, Clausen OP, Fausa O, et al. Patients with small duct primary sclerosing cholangitis have a favourable long term prognosis. Gut 2002;51:731-5.
49. Bjornsson E, Olsson R, Bergquist A, Lindgren S, Braden B, Chapman RW, et al. The natural history of small-duct primary sclerosing cholangitis. Gastroenterology 2008;134:975-80.
50. Marcus E, Stone P, Krooupa AM, Thorburn D, Vivat B. Quality of life in primary sclerosing cholangitis: a systematic review. Health Qual Life Outcomes 2021;19:100.
51. Ponsioen CY. Endpoints in the design of clinical trials for primary sclerosing cholangitis. Biochim Biophys Acta Mol Basis Dis 2018;1864:1410-1414.
52. 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.
53. Stanich PP, Bjornsson E, Gossard AA, Enders F, Jorgensen R, Lindor KD. Alkaline phosphatase normalization is associated with better prognosis in primary sclerosing cholangitis. Dig Liver Dis 2011;43:309-13.

54. de Vries EM, Wang J, Leeflang MM, Boonstra K, Weersma RK, Beuers UH, et al. Alkaline phosphatase at diagnosis of primary sclerosing cholangitis and 1 year later: evaluation of prognostic value. Liver Int 2016;36:1867-1875.
55. Rupp C, Rossler A, Halibasic E, Sauer P, Weiss KH, Friedrich K, et al. Reduction in alkaline phosphatase is associated with longer survival in primary sclerosing cholangitis, independent of dominant stenosis. Aliment Pharmacol Ther 2014;40:1292-301.
56. Lindstrom L, Hultcrantz R, Boberg KM, Friis-Liby I, Bergquist A. Association between reduced levels of alkaline phosphatase and survival times of patients with primary sclerosing cholangitis. Clin Gastroenterol Hepatol 2013;11:841-6.
57. Hilscher M, Enders FB, Carey EJ, Lindor KD, Tabibian JH. Alkaline phosphatase normalization is a biomarker of improved survival in primary sclerosing cholangitis. Ann Hepatol 2016;15:246-53.
58. 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 cholangiocarcinoma in primary sclerosing cholangitis. J Hepatol 2013;58:329-34.
59. Balasubramaniam K, Wiesner RH, LaRusso NF. Primary sclerosing cholangitis with normal serum alkaline phosphatase activity. Gastroenterology 1988;95:1395-8.
60. Trivedi PJ, Muir AJ, Levy C, Bowlus CL, Manns M, Lu X, et al. Inter- and Intra-individual Variation, and Limited Prognostic Utility, of Serum Alkaline Phosphatase in a Trial of Patients With Primary Sclerosing Cholangitis. Clin Gastroenterol Hepatol 2020.
61. Vesterhus M, Hov JR, Holm A, Schrumpf E, Nygard S, Godang K, et al. Enhanced liver fibrosis score predicts transplant-free survival in primary sclerosing cholangitis. Hepatology 2015;62:188-97.
62. de Vries EMG, Farkkila M, Milkiewicz P, Hov JR, Eksteen B, Thorburn D, et al. Enhanced liver fibrosis test predicts transplant-free survival in primary sclerosing cholangitis, a multi-centre study. Liver Int 2017;37:1554-1561.
63. Muir AJ, Levy C, Janssen HLA, Montano-Loza AJ, Shiffman ML, Caldwell S, et al. Simtuzumab for Primary Sclerosing Cholangitis: Phase 2 Study Results With Insights on the Natural History of the Disease. Hepatology 2019;69:684-698.
64. Saffioti F, Roccarina D, Vesterhus M, Hov JR, Rosenberg W, Pinzani M, et al. Cholangiocarcinoma is associated with a raised enhanced liver fibrosis score independent of primary sclerosing cholangitis. Eur J Clin Invest 2019;49:e13088.
65. 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 2017.
66. Farrant JM, Hayllar KM, Wilkinson ML, Karani J, Portmann BC, Westaby D, et al. Natural history and prognostic variables in primary sclerosing cholangitis. Gastroenterology 1991;100:1710-7.
67. Eaton JE, Vesterhus M, McCauley BM, Atkinson EJ, Schlicht EM, Juran BD, et al. Primary Sclerosing Cholangitis Risk Estimate Tool (PREsTo) Predicts Outcomes in PSC: A Derivation & Validation Study Using Machine Learning. Hepatology 2018.
68. 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;75:688-94.
69. 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.
70. 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.

71. Wiesner RH, Grambsch PM, Dickson ER, Ludwig J, MacCarty RL, Hunter EB, et al. Primary sclerosing cholangitis: natural history, prognostic factors and survival analysis. Hepatology 1989;10:430-6.
72. de Vries EM, Verheij J, Hubscher SG, Leeflang MM, Boonstra K, Beuers U, et al. Applicability and prognostic value of histologic scoring systems in primary sclerosing cholangitis. J Hepatol 2015;63:1212-9.
73. de Vries EM, de Krijger M, Farkkila M, Arola J, Schirmacher P, Gotthardt D, et al. Validation of the prognostic value of histologic scoring systems in primary sclerosing cholangitis: An international cohort study. Hepatology 2017;65:907-919.
74. Olsson R, Hagerstrand I, Broome U, Danielsson A, Jarnerot G, Loof L, et al. Sampling variability of percutaneous liver biopsy in primary sclerosing cholangitis. J Clin Pathol 1995;48:933-5.
75. Corpechot C, El Naggar A, Poujol-Robert A, Ziol M, Wendum D, Chazouilleres O, et al. Assessment of biliary fibrosis by transient elastography in patients with PBC and PSC. Hepatology 2006;43:1118-24.
76. 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;146:970-9; quiz e15-6.
77. Ehlken H, Wroblewski R, Corpechot C, Arrive L, Rieger T, Hartl J, et al. Validation of Transient Elastography and Comparison with Spleen Length Measurement for Staging of Fibrosis and Clinical Prognosis in Primary Sclerosing Cholangitis. PLoS One 2016;11:e0164224.
78. Ehlken H, Lohse AW, Schramm C. Transient elastography in primary sclerosing cholangitis-the value as a prognostic factor and limitations. Gastroenterology 2014;147:542-3.
79. Millonig G, Reimann FM, Friedrich S, Fonouni H, Mehrabi A, Buchler MW, et al. Extrahepatic cholestasis increases liver stiffness (FibroScan) irrespective of fibrosis. Hepatology 2008;48:1718-23.
80. Mjelle AB, Fossdal G, Gilja OH, Vesterhus M. Liver Elastography in Primary Sclerosing Cholangitis Patients Using Three Different Scanner Systems. Ultrasound Med Biol 2020;46:1854-1864.
81. Craig DA, MacCarty RL, Wiesner RH, Grambsch PM, LaRusso NF. Primary sclerosing cholangitis: value of cholangiography in determining the prognosis. AJR Am J Roentgenol 1991;157:959-64.
82. MacCarty RL, LaRusso NF, Wiesner RH, Ludwig J. Primary sclerosing cholangitis: findings on cholangiography and pancreatography. Radiology 1983;149:39-44.
83. Olsson RG, Asztely MS. Prognostic value of cholangiography in primary sclerosing cholangitis. Eur J Gastroenterol Hepatol 1995;7:251-4.
84. Majoie CB, Reeders JW, Sanders JB, Huibregtse K, Jansen PL. Primary sclerosing cholangitis: a modified classification of cholangiographic findings. AJR Am J Roentgenol 1991;157:495-7.
85. Chen LY, Goldberg HI. Sclerosing cholangitis: broad spectrum of radiographic features. Gastrointest Radiol 1984;9:39-47.
86. Ponsioen CY, Reitsma JB, Boberg KM, Aabakken L, Rauws EA, Schrumpf E. Validation of a cholangiographic prognostic model in primary sclerosing cholangitis. Endoscopy 2010;42:742-7.
87. Bader TR, Beavers KL, Semelka RC. MR imaging features of primary sclerosing cholangitis: patterns of cirrhosis in relationship to clinical severity of disease. Radiology 2003;226:675-85.
88. 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. Digestive Diseases and Sciences 2007;52:3499-3506.

89. 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.
90. 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.
91. Ruiz A, Lemoinne S, Carrat F, Corpechot C, Chazouilleres O, Arrive L. Radiologic course of primary sclerosing cholangitis: assessment by three-dimensional magnetic resonance cholangiography and predictive features of progression. Hepatology 2014;59:242-50.
92. Cazzagon N, Lemoinne S, El Mouhadi S, Trivedi PJ, Gaouar F, Kemgang A, et al. The Complementary Value of Magnetic Resonance Imaging and Vibration-Controlled Transient Elastography for Risk Stratification in Primary Sclerosing Cholangitis. Am J Gastroenterol 2019;114:1878-1885.
93. Muir AJ, Taghipour M, Hassanzadeh EH, Sahni VA, Sainani N, Ding D. A risk prediction score based on magnetic resonance cholangiopancreatography (MRCP) accurately predicts disease progression in patients with primary sclerosing cholangitis (PSC) [Abstract]. Hepatology 2017;66:81A.
94. Kitzing YX, Whitley SA, Upponi SS, Srivastava B, Alexander GJ, Lomas DJ. Association between progressive hepatic morphology changes on serial MR imaging and clinical outcome in primary sclerosing cholangitis. J Med Imaging Radiat Oncol 2017;61:636-642.
95. Bookwalter CA, Venkatesh SK, Eaton JE, Smyrk TD, Ehman RL. MR elastography in primary sclerosing cholangitis: correlating liver stiffness with bile duct strictures and parenchymal changes. Abdom Radiol (NY) 2018.
96. 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.
97. 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) 2019.
98. Eaton JE, Dzyubak B, Venkatesh SK, Smyrk TC, Gores GJ, Ehman RL, et al. Performance of magnetic resonance elastography in primary sclerosing cholangitis. J Gastroenterol Hepatol 2016;31:1184-90.
99. Jhaveri KS, Hosseini-Nik H, Sadoughi N, Janssen H, Feld JJ, Fischer S, et al. The development and validation of magnetic resonance elastography for fibrosis staging in primary sclerosing cholangitis. Eur Radiol 2018.
100. Hoodeshenas S, Welle CL, Navin PJ, Dzyubak B, Eaton JE, Ehman RL, et al. Magnetic Resonance Elastography in Primary Sclerosing Cholangitis: Interobserver Agreement for Liver Stiffness Measurement with Manual and Automated Methods. Acad Radiol 2019.
101. Taouli B, Koh DM. Diffusion-weighted MR imaging of the liver. Radiology 2010;254:47-66.
102. Kovac JD, Dakovic M, Stanisavljevic D, Alempijevic T, Jesic R, Seferovic P, et al. Diffusion-weighted MRI versus transient elastography in quantification of liver fibrosis in patients with chronic cholestatic liver diseases. Eur J Radiol 2012;81:2500-6.
103. Keller S, Sedlacik J, Schuler T, Buchert R, Avanesov M, Zenouzi R, et al. Prospective comparison of diffusion-weighted MRI and dynamic Gd-EOB-DTPA-enhanced MRI for detection and staging of hepatic fibrosis in primary sclerosing cholangitis. Eur Radiol 2018.

104. Wang QB, Zhu H, Liu HL, Zhang B. Performance of magnetic resonance elastography and diffusion-weighted imaging for the staging of hepatic fibrosis: A meta-analysis. Hepatology 2012;56:239-47.
105. Ringe KI, Hinrichs J, Merkle EM, Weismuller TJ, Wacker F, Meyer BC. Gadoxetate disodium in patients with primary sclerosing cholangitis: an analysis of hepatobiliary contrast excretion. J Magn Reson Imaging 2014;40:106-12.
106. Noren B, Forsgren MF, Dahlqvist Leinhard O, Dahlstrom N, Kihlberg J, Romu T, et al. Separation of advanced from mild hepatic fibrosis by quantification of the hepatobiliary uptake of Gd-EOB-DTPA. Eur Radiol 2013;23:174-81.
107. Nilsson H, Blomqvist L, Douglas L, Nordell A, Jacobsson H, Hagen K, et al. Dynamic gadoxetate-enhanced MRI for the assessment of total and segmental liver function and volume in primary sclerosing cholangitis. J Magn Reson Imaging 2014;39:879-86.
108. Hinrichs H, Hinrichs JB, Gutberlet M, Lenzen H, Raatschen HJ, Wacker F, et al. Functional gadoxetate disodium-enhanced MRI in patients with primary sclerosing cholangitis (PSC). Eur Radiol 2016;26:1116-24.
109. Nolz R, Asenbaum U, Schoder M, Wibmer A, Einspieler H, Prusa AM, et al. Diagnostic workup of primary sclerosing cholangitis: the benefit of adding gadoxetic acid-enhanced T1-weighted magnetic resonance cholangiography to conventional T2-weighted magnetic resonance cholangiography. Clin Radiol 2014;69:499-508.
110. Keller S, Aigner A, Zenouzi R, Kim AC, Meijer A, Weidemann SA, et al. Association of gadolinium-enhanced magnetic resonance imaging with hepatic fibrosis and inflammation in primary sclerosing cholangitis. PLoS One 2018;13:e0193929.
111. Keller S, Venkatesh SK, Avanesov M, Weinrich JM, Zenouzi R, Schramm C, et al. Gadolinium-based relative contrast enhancement in primary sclerosing cholangitis: additional benefit for clinicians? Clin Radiol 2018.
112. Schulze J, Lenzen H, Hinrichs JB, Ringe B, Manns MP, Wacker F, et al. An Imaging Biomarker for Assessing Hepatic Function in Patients with Primary Sclerosing Cholangitis. Clin Gastroenterol Hepatol 2018.
113. Elkilany A, Geisel D, Muller T, Fischer A, Denecke T. Gadoxetic acid-enhanced MRI in primary sclerosing cholangitis: added value in assessing liver function and monitoring disease progression. Abdom Radiol (NY) 2021;46:979-991.
114. Evrimler S, Swensson JK, Are VS, Tirkes T, Vuppalanchi R, Akisik F. Quantitative assessment of disease severity of primary sclerosing cholangitis with T1 mapping and extracellular volume imaging. Abdom Radiol (NY) 2020.
115. Ehlken H, Wroblewski R, Corpechot C, Arrive L, Lezius S, Hartl J, et al. Spleen size for the prediction of clinical outcome in patients with primary sclerosing cholangitis. Gut 2016;65:1230-2.
116. Khoshpouri P, Hazhirkarzar B, Ameli S, Pandey A, Ghadimi M, Rezvani Habibabadi R, et al. Quantitative spleen and liver volume changes predict survival of patients with primary sclerosing cholangitis. Clin Radiol 2019;74:734 e13-734 e20.
117. Idilman IS, Low HM, Bakhshi Z, Eaton J, Venkatesh SK. Comparison of liver stiffness measurement with MRE and liver and spleen volumetry for prediction of disease severity and hepatic decompensation in patients with primary sclerosing cholangitis. Abdom Radiol (NY) 2020.
118. Jung F, Cazzagon N, Vettorazzi E, Corpechot C, Chazouilleres O, Arrive L, et al. Rate of Spleen Length Progression Is a Marker of Outcome in Patients With Primary Sclerosing Cholangitis. Clin Gastroenterol Hepatol 2019;17:2613-2615.
119. Ringe KI, Vo Chieu VD, Wacker F, Lenzen H, Manns MP, Hundt C, et al. Fully automated detection of primary sclerosing cholangitis (PSC)-compatible bile duct changes based on 3D magnetic resonance cholangiopancreatography using machine learning. Eur Radiol 2020.

120. Zenouzi R, Liwinski T, Yamamura J, Weiler-Normann C, Sebode M, Keller S, et al. Follow-up magnetic resonance imaging/3D-magnetic resonance cholangiopancreatography in patients with primary sclerosing cholangitis: challenging for experts to interpret. Aliment Pharmacol Ther 2018.
121. Piechnik SK, Ferreira VM, Dall'Armellina E, Cochlin LE, Greiser A, Neubauer S, et al. Shortened Modified Look-Locker Inversion recovery (ShMOLLI) for clinical myocardial T1-mapping at 1.5 and 3 T within a 9 heartbeat breathhold. J Cardiovasc Magn Reson 2010;12:69.
122. Tunnicliffe EM, Banerjee R, Pavlides M, Neubauer S, Robson MD. A model for hepatic fibrosis: the competing effects of cell loss and iron on shortened modified Look-Locker inversion recovery T1 (shMOLLI-T1) in the liver. J Magn Reson Imaging 2017;45:450-462.
123. Banerjee R, Pavlides M, Tunnicliffe EM, Piechnik SK, Sarania N, Philips R, et al. Multiparametric magnetic resonance for the non-invasive diagnosis of liver disease. J Hepatol 2014;60:69-77.
124. Pavlides M, Banerjee R, Tunnicliffe EM, Kelly C, Collier J, Wang LM, et al. Multiparametric magnetic resonance imaging for the assessment of non-alcoholic fatty liver disease severity. Liver Int 2017;37:1065-1073.
125. Eddowes PJ, Newsome PN, Anstee Q, Guha IN, Sheridan DA, Tsochatzis E, et al. Staging fibrosis and excluding advanced fibrosis in patients with NAFLD: Comparison of non-invasive markers in an interim analysis from a prospective multicentre study. Hepatology 2016;64.
126. McDonald N, Eddowes PJ, Hodson J, Semple SIK, Davies NP, Kelly CJ, et al. Multiparametric magnetic resonance imaging for quantitation of liver disease: a two-centre cross-sectional observational study. Sci Rep 2018;8:9189.
127. Pavlides M, Banerjee R, Sellwood J, Kelly CJ, Robson MD, Booth JC, et al. Multiparametric magnetic resonance imaging predicts clinical outcomes in patients with chronic liver disease. J Hepatol 2016;64:308-315.
128. Jayaswal ANA, Levick C, Selvaraj EA, Dennis A, Booth JC, Collier J, et al. Prognostic value of multiparametric magnetic resonance imaging, transient elastography and blood-based fibrosis markers in patients with chronic liver disease. Liver Int 2020.
129. 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]. J Radiol 2007;88:1689-94.
130. Irving B, Hutton C, Dennis A, Vikal S, Mavar M, Kelly M, et al. Deep Quantitative Liver Segmentation and Vessel Exclusion to Assist in Liver Assessment, Cham, Springer International Publishing, 2017.
131. Irving B, Popescu IA, Bates R, Allen PD, Gomes AL, Kannan P, et al. maskSLIC: Regional Superpixel Generation with Application to Local Pathology Characterisation in Medical Images, 2017:1-7.
132. Ware JE, Jr., Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care 1992;30:473-83.
133. Younossi ZM, Afendy A, Stepanova M, Racila A, Nader F, Gomel R, et al. Development and validation of a primary sclerosing cholangitis-specific patient-reported outcomes instrument: The PSC PRO. Hepatology 2018;68:155-165.
134. 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.
135. Harrison SA, Dennis A, Fiore MM, Kelly MD, Kelly CJ, Paredes AH, et al. Utility and variability of three non-invasive liver fibrosis imaging modalities to evaluate efficacy of GR-MD-02 in subjects with NASH and bridging fibrosis during a phase-2 randomized clinical trial. PLOS ONE 2018;13:e0203054.
136. Mojtahed A, Kelly CJ, Herlihy AH, Kin S, Wilman HR, McKay A, et al. Reference range of liver corrected T1 values in a population at low risk for fatty liver disease-

a UK Biobank sub-study, with an appendix of interesting cases. Abdom Radiol (NY) 2019;44:72-84.
137. Bachtiar V, Kelly MD, Wilman HR, Jacobs J, Newbould R, Kelly CJ, et al. Repeatability and reproducibility of multiparametric magnetic resonance imaging of the liver. PLOS ONE 2019;14:e0214921.
138. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. J Chiropr Med 2016;15:155-63.
139. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;1:307-10.
140. 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.
141. Mozes FE, Tunnicliffe EM, Moolla A, Marjot T, Levick CK, Pavlides M, et al. Mapping tissue water T1 in the liver using the MOLLI T1 method in the presence of fat, iron and B0 inhomogeneity. NMR Biomed 2019;32:e4030.
142. Hellmig S, Von Schoning F, Gadow C, Katsoulis S, Hedderich J, Folsch UR, et al. Gastric emptying time of fluids and solids in healthy subjects determined by 13C breath tests: influence of age, sex and body mass index. J Gastroenterol Hepatol 2006;21:1832-8.
143. Vela D. Low hepcidin in liver fibrosis and cirrhosis; a tale of progressive disorder and a case for a new biochemical marker. Mol Med 2018;24:5.
144. Lyberopoulou A, Chachami G, Gatselis NK, Kyratzopoulou E, Saitis A, Gabeta S, et al. Low Serum Hepcidin in Patients with Autoimmune Liver Diseases. PLoS One 2015;10:e0135486.
145. Pinzani M, Luong TV. Pathogenesis of biliary fibrosis. Biochim Biophys Acta Mol Basis Dis 2018;1864:1279-1283.
146. 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 Sep;52(3):807-820.
147. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159-74.
148. 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.
149. van de Meeberg PC, Portincasa P, Wolfhagen FH, van Erpecum KJ, VanBerge-Henegouwen GP. Increased gall bladder volume in primary sclerosing cholangitis. Gut 1996;39:594-9.
150. Culver EL, Bungay HK, Betts M, Forde C, Buchel O, Manganis C, et al. Prevalence and long-term outcome of sub-clinical primary sclerosing cholangitis in patients with ulcerative colitis. Liver Int 2020;40:2744-2757.
151. 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 2020:1-8.
152. Scheuer PJ. Ludwig Symposium on biliary disorders--part II. Pathologic features and evolution of primary biliary cirrhosis and primary sclerosing cholangitis. Mayo Clin Proc 1998;73:179-83.
153. Harata M, Hashimoto S, Kawabe N, Nitta Y, Murao M, Nakano T, et al. Liver stiffness in extrahepatic cholestasis correlates positively with bilirubin and negatively with alanine aminotransferase. Hepatol Res 2011;41:423-9.
154. Ito K, Mitchell DG, Outwater EK, Blasbalg R. Primary sclerosing cholangitis: MR imaging features. AJR Am J Roentgenol 1999;172:1527-33.

155. Revelon G, Rashid A, Kawamoto S, Bluemke DA. Primary sclerosing cholangitis: MR imaging findings with pathologic correlation. AJR Am J Roentgenol 1999;173:1037-42.
156. Caldwell SH, Hespenheide EE, Harris D, de Lange EE. Imaging and clinical characteristics of focal atrophy of segments 2 and 3 in primary sclerosing cholangitis. J Gastroenterol Hepatol 2001;16:220-4.
157. Arrive L, Hodoul M, Arbache A, Slavikova-Boucher L, Menu Y, El Mouhadi S. Magnetic resonance cholangiography: Current and future perspectives. Clin Res Hepatol Gastroenterol 2015;39:659-64.
158. Tokgoz O, Unal I, Turgut GG, Yildiz S. The value of liver and spleen ADC measurements in the diagnosis and follow up of hepatic fibrosis in chronic liver disease. Acta Clin Belg 2014;69:426-32.
159. Colagrande S, Pasquinelli F, Mazzoni LN, Belli G, Virgili G. MR-diffusion weighted imaging of healthy liver parenchyma: repeatability and reproducibility of apparent diffusion coefficient measurement. J Magn Reson Imaging 2010;31:912-20.
160. Hoad CL, Palaniyappan N, Kaye P, Chernova Y, James MW, Costigan C, et al. A study of T-1 relaxation time as a measure of liver fibrosis and the influence of confounding histological factors. Nmr in Biomedicine 2015;28:706-714.
161. Singh S, Venkatesh SK, Wang Z, Miller FH, Motosugi U, Low RN, et al. Diagnostic performance of magnetic resonance elastography in staging liver fibrosis: a systematic review and meta-analysis of individual participant data. Clin Gastroenterol Hepatol 2015;13:440-451 e6.
162. Pullicino R, Radon M, Biswas S, Bhojak M, Das K. A Review of the Current Evidence on Gadolinium Deposition in the Brain. Clin Neuroradiol 2018;28:159-169.
163. Gulani V, Calamante F, Shellock FG, Kanal E, Reeder SB, International Society for Magnetic Resonance in M. Gadolinium deposition in the brain: summary of evidence and recommendations. Lancet Neurol 2017;16:564-570.
164. Jayaswal ANA, Levick C, Selvaraj EA, Dennis A, Booth JC, Collier J, et al. Prognostic value of multiparametric magnetic resonance imaging, transient elastography and blood-based fibrosis markers in patients with chronic liver disease. Liver Int 2020;40:3071-3082.
165. Khoshpouri P, Habibabadi RR, Hazhirkarzar B, Ameli S, Ghadimi M, Ghasabeh MA, et al. Imaging Features of Primary Sclerosing Cholangitis: From Diagnosis to Liver Transplant Follow-up. Radiographics 2019;39:1938-1964.
166. Aube C, Bazeries P, Lebigot J, Cartier V, Boursier J. Liver fibrosis, cirrhosis, and cirrhosis-related nodules: Imaging diagnosis and surveillance. Diagn Interv Imaging 2017;98:455-468.
167. Venkatesh SK, Wells ML, Miller FH, Jhaveri KS, Silva AC, Taouli B, et al. Magnetic resonance elastography: beyond liver fibrosis-a case-based pictorial review. Abdom Radiol (NY) 2017.
168. Colling R, Verrill C, Fryer E, Kartsonaki C, Wang LM, Chapman R, et al. Bile duct basement membrane thickening in primary sclerosing cholangitis. Histopathology 2016;68:819-24.
169. Bojorquez JZ, Bricq S, Acquitter C, Brunotte F, Walker PM, Lalande A. What are normal relaxation times of tissues at 3 T? Magn Reson Imaging 2017;35:69-80.
170. Barrett K, Barman SM, Boitano S, Brooks H. Ganong's Review of Medical Physiology. New York: McGraw-Hill Medical, 2012.
171. Nakanuma Y, Hoso M, Sanzen T, Sasaki M. Microstructure and development of the normal and pathologic biliary tract in humans, including blood supply. Microsc Res Tech 1997;38:552-70.
172. Kozaka K, Sheedy SP, Eaton JE, Venkatesh SK, Heiken JP. Magnetic resonance imaging features of small-duct primary sclerosing cholangitis. Abdom Radiol (NY) 2020;45:2388-2399.

173. Ringe KI, Bergquist A, Lenzen H, Kartalis N, Manns MP, Wacker F, et al. Clinical features and MRI progression of small duct primary sclerosing cholangitis (PSC). Eur J Radiol 2020;129:109101.
174. Eaton JE, Sen A, Hoodeshenas S, Schleck CD, Harmsen WS, Gores GJ, et al. Changes in Liver Stiffness, Measured by Magnetic Resonance Elastography, Associated With Hepatic Decompensation in Patients With Primary Sclerosing Cholangitis. Clin Gastroenterol Hepatol 2019.
175. Cohen J. Statistical power analysis for the behavioral sciences. Hillsdale, N.J.: L. Erlbaum Associates, 1988.
176. Ponsioen CY, Chapman RW, Chazouilleres O, Hirschfield GM, Karlsen TH, Lohse AW, et al. Surrogate endpoints for clinical trials in primary sclerosing cholangitis: Review and results from an International PSC Study Group consensus process. Hepatology 2016;63:1357-67.
177. Nam JG, Lee JM, Kang HJ, Lee SM, Kim E, Peeters JM, et al. GRASE Revisited: breath-hold three-dimensional (3D) magnetic resonance cholangiopancreatography using a Gradient and Spin Echo (GRASE) technique at 3T. Eur Radiol 2018;28:3721-3728.
178. Yoshida M, Nakaura T, Inoue T, Tanoue S, Takada S, Utsunomiya D, et al. Magnetic resonance cholangiopancreatography with GRASE sequence at 3.0T: does it improve image quality and acquisition time as compared with 3D TSE? Eur Radiol 2018;28:2436-2443.
179. Morimoto D, Hyodo T, Kamata K, Kadoba T, Itoh M, Fukushima H, et al. Navigator-triggered and breath-hold 3D MRCP using compressed sensing: image quality and method selection factor assessment. Abdom Radiol (NY) 2020;45:3081-3091.
180. Yoon JH, Lee SM, Kang HJ, Weiland E, Raithel E, Son Y, et al. Clinical Feasibility of 3-Dimensional Magnetic Resonance Cholangiopancreatography Using Compressed Sensing: Comparison of Image Quality and Diagnostic Performance. Invest Radiol 2017;52:612-619.