DR ROMAN ZENOUZI (Orcid ID : 0000-0003-0136-0924)

Article type : Original Scientific Paper

Corresponding author mail id: r.zenouzi@uke.de

Follow-up magnetic resonance imaging/3D-magnetic resonance cholangiopancreatography in patients with primary sclerosing cholangitis: challenging for experts to interpret

Short title: Follow-up MRI/MRCP in PSC

R. Zenouzi[1], T. Liwinski[1], J. Yamamura[2], C. Weiler-Normann[1], M. Sebode[1], S. Keller[2], A. W. Lohse[1], C. Schramm[1]

[1] 1st Department of Medicine, University Medical Centre Hamburg-Eppendorf, Martinistraße 52, 20246 Hamburg, Germany.

[2] Department of Diagnostic and Interventional Radiology and Nuclear Medicine, University Medical Centre Hamburg-Eppendorf, Martinistraße 52, 20246 Hamburg, Germany.

Acknowledgments:

R. Zenouzi is the guarantor of the article, performed the research, collected and analysed the data, designed the research study and wrote the paper. T. Liwinski significantly contributed to the analysis of the data. J. Yamamura contributed to the design of the study. T. Weiler-Normann contributed to the design of the study and critically revised the manuscript. M. Sebode contributed to the design of the study and

critically revised the manuscript. S. Keller technically supported the study. A.W. Lohse contributed to the design of the study and critically revised the manuscript. C. Schramm designed and supervised the research study and critically revised the manuscript.

Participants of the survey (suggested for banner-co-authorship):

L. Aabakken[1], L. Arrivé[2], C.L. Bowlus[3], H. Bungay[4], H.R. van Buuren[5], V. Cardinale[6], E.J. Carey[7], O. Chazouillères[8], A. Cheung[9], E.L. Culver[10], J.F. Dufour[11], J.M.Dumonceau[12], J.E. Eaton[13], P.J. Eddowes[14], M. Färkkilä[15], A. Floreani[16], I. Franceschet[16], S.D. Hohenester[17], G. Kemmerich[18], M. Krawczyk[19], H. Lenzen[20], C. Levy[21], H.U. Marschall[22], M. Marzioni[23], R. Motta[24], L. Muratori[25], S.P. Pereira[26], J.W. Poley[27], J. Rimola[28], K.I. Ringe[29], S. Rushbrook[30], E. Schrumpf[31], B.W. Simpson[30], J.C. Spina[32], B. Terziroli Beretta-Piccoli[33], M. Trauner[34], A. Tringali[35], S.K. Venkatesh[36], M. Vesterhus[37,38], A. Villamil[39], T.J. Weismüller[40], H. Ytting[41], V. Zimmer[19]

[1] Rikshospitalet University Hospital, Hospital and Faculty of Medicine, University of Oslo, Oslo, Norway. [2] Department of Radiology, Hôpital Saint-Antoine, UPMC Sorbonne Universités, Paris, France. [3] University of California Davis School of Medicine, Sacramento, CA, USA. [4] Oxford University Hospitals NHS Foundation Trust, Oxford, UK. [5] Department of Gastroenterology and Hepatology, Erasmus University Medical Centre Rotterdam, Rotterdam, The Netherlands. [6] Department of Medico-Surgical Sciences and Biotechnologies, Sapienza University of Rome, Rome, Italy. [7] Division of Gastroenterology & Hepatology, Mayo Clinic in Arizona, Phoenix, AZ, USA. [8] Department of Hepatology, AP-HP, Hôpital Saint Antoine, Paris, France. [9] General Internal Medicine, University Health Network, Toronto General Hospital, Toronto, Canada. [10] Royal Free Hospital, London and John Radcliffe Hospital, Oxford.

Wellcome Trust Research Fellow, Oxford University. Oxford, UK. [11] University Clinic of Visceral Surgery and Medicine, Inselspital Bern, Bern, Switzerland. [12] Gedyt Endoscopy Center, Buenos Aires, Argentina. [13] Department of Medicine, Mayo Clinic College of Medicine, Mayo Clinic, Rochester, MN, USA. [14] NIHR Birmingham Liver Biomedical Research Unit and Centre for Liver Research, University of Birmingham, Birmingham, UK. [15] Clinic of Gastroenterology, University of Helsinki and Helsinki University Hospital, Helsinki, Finland. [16] Department of Surgery, Oncology and Gastroenterology, University of Padua, Padua, Italy. [17] Department of Medicine II, University of Munich (LMU), Munich, Germany. [18] Department for Abdominal Radiology, Rikshospitalet University Hospital, University of Oslo, Oslo, Norway. [19] Department of Medicine II, Saarland University Medical Center, Saarland University, Homburg, Germany. [20] Department of Gastroenterology, Hepatology and Endocrinology, Hannover Medical School, Hannover, Germany. [21] Division of Hepatology, Miller School of Medicine, University of Miami, Miami, FL, USA. [22] Department of Molecular and Clinical Medicine, Sahlgrenska Academy, Institute of Medicine, University of Gothenburg, Gothenburg, Sweden. [23] Department of Gastroenterology and Hepatology, Università Politecnica delle Marche, Ancona, Italy. [24] Department of Medicine, University of Padua, Padua, Italy. [25] Department of Digestive System Disease and Internal Medicine, Saint Orsola-Malpighi Hospital, Bologna, Italy. [26] Department of Gastroenterology, Royal Liverpool and Broadgreen University Hospitals NHS Trust, Liverpool, UK. [27] Department of Gastroenterology & Hepatology, Endoscopy Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands. [28] Department of Radiology, Hospital Clinic Barcelona, Barcelona, Spain. [29] Department of Diagnostic and Interventional Radiology, Hannover Medical School, Hannover, Germany. [30] Norwich Medical School, Faculty of Medicine and Health Sciences, University of East Anglia, Norwich, UK. [31] Section of

Gastroenterology, Department of Transplantation Medicine, Division of Cancer Medicine, Surgery and Transplantation, Oslo University Hospital, Rikshospitalet, Oslo, Norway. [32] Department of Radiology, Hospital Italiano of Buenos Aires, Buenos Aires, Argentina. [33] Epatocentro TicinoLuganoSwitzerland, Switzerland. [34] Division of Gastroenterology and Hepatology, Department of Internal Medicine III, Medical University of Vienna, Vienna, Austria. [35] Digestive Endoscopy Unit, Catholic University, Rome, Italy. [36] Department of Radiology, Mayo Clinic College of Medicine, Mayo Clinic, MN, USA. [37] Norwegian PSC Research Center, Department of Transplantation Medicine, Division of Cancer Medicine, Surgery, Inflammatory Diseases and Transplantation, Oslo University Hospital, Oslo, Norway. [38] National Centre for Ultrasound in Gastroenterology, Haukeland University Hospital, Bergen, Norway. [39] Hepatology and Liver Transplantation Unit, Department for Gastroenterology, Hospital Italiano of Buenos Aires, Buenos Aires, Argentina. [40] Department of Internal Medicine 1, University of Bonn, Bonn, Germany. [41] Department of Hepatology, Rigshospitalet, University of Copenhagen, Copenhagen, Denmark.

This study was supported by the "Deutsche Forschungsgemeinschaft" (DFG SFB841 and KFO306), the YAEL Foundation and the Helmut and Hannelore Greve Foundation.

Summary:

Background: In patients with primary sclerosing cholangitis follow-up magnetic resonance imaging with magnetic resonance cholangiopancreatography is performed by many centres, particularly for the early detection of biliary malignancies and strictures. Clinically meaningful MRI-based definitions of PSC-related complications are, however, lacking.

Aims: To investigate how primary sclerosing cholangitis experts interpret follow-up magnetic resonance imaging/magnetic resonance cholangiopancreatography with a focus on conclusions that may impact clinical decision-making in PSC.

Methods: Within the International Primary Sclerosing Cholangitis Study Group, an online-survey on 16 real-life primary sclerosing cholangitis cases including clinical and biochemical information as well as a T2-weighted liver magnetic resonance imaging/3D-magnetic resonance cholangiopancreatography was conducted. The interpretation of images and subsequent recommendations were assessed using a multiple-choice questionnaire. An inter-rater reliability calculation (Fleiss` kappa) was performed and factors potentially affecting the interpretation of magnetic resonance images were analysed using generalized linear mixed effects models.

Results: 44 members/associates of the International Primary Sclerosing Cholangitis Study Group (median experience in the care of primary sclerosing cholangitis patients: 14 years) completed the survey. The magnetic resonance imaging interpretation significantly varied among the participants. The lowest agreement was found with respect to the indication to perform subsequent endoscopic retrograde cholangiopancreatography (κ=0.12, 95%CI 0.11-0.14). Elevated total bilirubin was the variable with the strongest effect on the rate of suspected dominant strictures, cholangiocarcinoma or endoscopic retrograde cholangiopancreatography recommendations. Liver cirrhosis did not prevent participants from recommending endoscopic retrograde cholangiopancreatography. Overall, the survey participants´ recommendations contrasted the real-life management and outcome.

Conclusions: In primary sclerosing cholangitis, the interpretation of follow-up magnetic resonance imaging/3D-magnetic resonance cholangiopancreatography significantly varies even among experts and seems to be primarily affected by bilirubin levels.

Generally accepted magnetic resonance imaging-based definitions of primary sclerosing cholangitis-related complications are urgently needed.

Key words: autoimmune liver disease, biliary tract, cholangiocarcinoma, primary sclerosing cholangitis,

Introduction

Primary sclerosing cholangitis (PSC) is a chronic cholestatic liver disease characterized by inflammation and progressive fibrosis of the intra- and/or extrahepatic bile ducts. PSC affects predominantly younger men and is frequently associated with a distinct form of colitis.[1] To date there is no effective therapy for PSC, and the disease progresses to biliary cirrhosis within 10-20 years in most patients.[2] PSC is one of the leading indications for liver transplantation in Northern Europe and the United States.[3,4] One of the major clinical challenges in PSC is the early detection of disease related complications, most importantly biliary malignancies and functionally relevant biliary strictures. Cholangiocarcinoma (CCA) develops in approximately 15-20% of patients and is regarded as the leading cause of death in PSC.[5-7] Most commonly, CCA presents as a stricture of the bile duct and less often as a mass lesion.[6] Dominant bile duct strictures develop in more than 50% of PSC patients over time and in itself seem to be a risk factor for the development of CCA.[6] Screening patients for novel or worsening biliary strictures, therefore, appears reasonable.[8] Magnetic resonance imaging (MRI) with magnetic resonance cholangiopancreatography (MRCP) has developed into the standard imaging modality to diagnose PSC and appears to be useful for the detection of bile duct and parenchymal changes associated with disease

progression as well as the diagnosis of PSC associated CCA.[8,9] Albeit the clinical perceptions of biliary strictures may vary across practitioners and institutions,[10] it is recommended that novel or worsening bile duct strictures should be further assessed using endoscopic retrograde cholangiopancreatography (ERCP) with tissue sampling in order to exclude CCA.[9] Therefore, most large volume centres perform follow-up MRI/MRCP in their PSC patients.[6,7,11-15]

Retrospective data recently published by Ali et al. support the notion that surveillance of hepatobiliary cancer may significantly improve outcome in PSC.[16] Prospective data on CCA surveillance and its impact on the detection of early stage cancer and on patient survival are, however, lacking. As a consequence, recommendations on the use of MRI for CCA surveillance vary between current guidelines.[3,4,17,18] Recently, the International PSC Study Group (IPSCSG) recommended that the use of MRI for surveillance of CCA in PSC should be an individualized decision.[8]

In clinical practice, even high quality MRCP pictures sometimes are difficult to interpret. In addition, imaging standards for MRI/MRCP vary across different centres and there is no accepted definition of dominant or clinically relevant strictures based on MRI. This indicates that the interpretation of MRI/MRCP in PSC may differ across clinicians and centres.

We hypothesised that in PSC the interpretation of follow-up MRI/MRCP and subsequent treatment recommendations, particularly with respect to the indication to perform ERCP, may significantly vary even across PSC-experienced clinicians. We therefore conducted an online survey within the IPSCSG and associated experts on 16 real-life cases of patients with established PSC diagnosis. On the basis of selected, relevant clinical, biochemical and MRI/3D-MRCP data survey participants were asked about their interpretation of the follow-up MRI/3D-MRCP and their recommendations with respect to subsequent ERCP performance. In addition, recommendations were

compared to the real-life management and outcome at the University Medical Centre Hamburg-Eppendorf.

Materials and Methods

Study design

The study protocol conformed to the ethical guidelines of the 1975 Declaration of Helsinki and was approved by the local ethics committee (PV4081-L). To evaluate the potential differences in the interpretation of follow-up MRI/3D-MRCP in PSC among experienced clinicians, an online-survey was conducted using the internet-platform *surveymonkey.com*. The survey started with a short users manual and information on the survey´s purpose. Next, survey participants were asked to provide personal information and information with respect to their centres´ clinical practice. These questions included information on the survey participants´ specialization (a), the years of clinical experience with PSC patients (b), the yearly number of patients treated in the survey participants´ centre (c) and the policy with respect to follow-up MRI/MRCP in the survey participants´ centres (d). Except for the years of experience with PSC patients, a multiple-choice questionnaire was used. Possible answers are provided in table 1.

Next, 16 anonymised real-life cases of patients with established large-duct PSC from the specialized outpatient clinic of the 1st Department of Medicine (YAEL-Centre for Autoimmune Liver Diseases), University Medical Centre Hamburg-Eppendorf, Germany were presented in a randomized order. Each case consisted of general patient characteristics (a) as well as relevant clinical (b) and biochemical information (c) closest to the time of imaging, which were considered to be used in clinical practice

to assess stage of disease and PSC related complications. Patient characteristics (a) included the patients´ age and gender, disease duration, the presence of liver cirrhosis, the presence of inflammatory bowel disease, information on previous ERCPs and on ursodeoxycholic acid and immunosuppressive treatment. Clinical Information (b) included the presence of fever, pruritus, jaundice, right upper quadrant abdominal pain and weight loss. Biochemical information (c) included the white blood cell and platelet counts and aspartate aminotransferase, alkaline phosphatase, total bilirubin, C-reactive protein and carbohydrate antigen 19-9 levels. Gender specific reference values were provided. In addition, each case included a T2-weighted liver MRI (axial) and a T2-weighted 3D-MRCP provided as video material. All MRI/3D-MRCPs had been performed on follow-up in patients with known PSC with a 1.5T scanner (slice thickness: 4 mm) between 2013 and 2015 at the Department for Diagnostic and Interventional Radiology and Nuclear Medicine, University Medical Centre Hamburg-Eppendorf, Germany. Study participants were able to watch the video files as often as required.

To identify factors, which could potentially impact the MRI-interpretation and ERCP-recommendation, a wide spectrum of PSC patients was chosen for the above mentioned cases. Within these 16 PSC cases 8 patients had liver cirrhosis, of which 4 patients were symptomatic (e.g. fever), of which again 2 patients showed additional biochemical signs of cholangitis. Likewise, within the remaining 8 non-cirrhotic PSC cases, 4 patients were symptomatic, of which 2 patients showed additional biochemical signs of cholangitis. Two patients had CCA (case 15 and 16). An example of a case presentation can be found in Figure S1.

Each of these 16 PSC cases was followed by a multiple-choice questionnaire. This questionnaire assessed the survey participant´s overall MRI/3D-MRCP interpretation (a), the presence of bile duct changes (b), the presence of dominant strictures (c), the

presence of CCA (d), whether ERCP should be performed (e) and arguments for/against ERCP recommendation (f). Possible answers are provided in Table 1.

Survey participants were recruited within the IPSCSG and associates and were primarily specialized in gastroenterology (including 5 participants primarily specialized in endoscopy), hepatology or radiology. A participation in the survey was possible between March and June 2016 in an anonymous fashion. Since completing the survey was associated with a significant amount of time (60-90 minutes), survey participants were invited to email their personal data after finishing the survey in order to be acknowledged by banner-co-authorship.

Statistical Analysis

Only results from participants, who completed the survey, went into the final analysis. Survey participants´ characteristics were described using the median value accompanied by the first (Q1) and the third quartile (Q3). To assess the agreement with respect to the MRI/3D-MRCP interpretation within the cohort of survey participants an inter-rater reliability calculation was performed using the Fleiss´ kappa value. To assess the systematic influence of case and respondent related variables on the respondent ratings we additionally applied generalized linear mixed effects models (GLMM) controlling for the individual raters and cases as random factors. We aimed to assess the fixed effects influencing the respondents´ votes on: the presence of dominant strictures, the presence of CCA and the indication for ERCP. As fixed effects the following independent variables were considered for each of the above listed dependent variables: the presence of liver cirrhosis, symptoms, signs of cholangitis (defined by elevated white blood cell count or C-reactive protein plus symptoms), bilirubin level (normal or elevated; upper limit of normal: 1.1 mg/dl) and the specialization of the respective respondent. The latter was split in dichotomous dummy

variables: gastroenterologist, hepatologist and radiologist. First, we explored unconditional models by assessing the association of each independent variable with the respective outcomes by univariate GLMM analyses. Variables that retained a *P*-value <.05 were intended to be included in multivariate GLMM. The potential interference between predictor variables was tested post-hoc by Chi-squared test or Fisher's exact test (where appropriate) to guarantee the model assumption of independent predictors. In case of a significant relationship, the independent variable displaying a higher *P*-value and higher Akaike information criterion (AIC) was dropped. This model building approach has been described previously.[19,20] For the GLMM Laplace approximation a logit link was used as provided in the "glmer function" of the "lme4" R package.[21] All statistical analyses were performed with the R statistical programming language version 3.4.1 (R Foundation for Statistical Computing, Vienna, Austria).

As all provided data in the survey were based on real-life cases follow-up data were available, enabling a comparison of the survey participants´ recommendations with respect to ERCP performance with the real-life management and outcome. In this context, ERCP was rated to be beneficial, if one or more of the following features were present: technically successful bile duct dilatation, positive microbiological report from bile fluid sampling, decreasing biochemical cholestatic markers after ERCP, improvement of symptoms, diagnosis of CCA by brush cytology.

Results

Survey participants

120 members and associates of the IPSCSG were invited to participate in the survey. Out of 89 respondents, 44 participants (19 hepatologists, 16 gastroenterologists, 9 radiologists) completed the survey and went into the final analysis. Within this cohort the median clinical experience in the care of PSC patients was 14 (7-20) years. The level of expertise was also reflected by the annual number of PSC patients treated at the survey participants´ centres: 14 (32%) participants quoted that < 50, 22 (50%) participants quoted that 51-150 and 8 (18%) participants quoted that > 150 PSC patients were treated in their centre per year. Despite the lack of clear recommendations in the current guidelines, follow-up MRI/MRCP for PSC patients was widely accepted among the survey participants: 23 (52%) participants quoted to perform follow-up MRI/MRCP annually at their centre, 8 (18%) participants each quoted that MRI/MRCP was performed every other year or irregularly, respectively. Only 5 (11%) participants stated that MRI/MRCP was not performed for PSC surveillance in their centre.

Inter-rater reliability of agreement

To assess the level of agreement with respect to the MRI/3D-MRCP interpretation and subsequent recommendations among the study participants an inter-rater reliability calculation was performed using the Fleiss´ kappa value. This approach provided a low level of agreement within the cohort.

In detail, an agreement of over 75% of the survey participants was achieved with respect to the overall MRI/3D-MRCP interpretation (Figure 1a) in only 2/16 cases (κ=0.13, 95%CI 0.12-0.15) and with respect to the interpretation of presence and localization of biliary changes (Figure 1b) in 10/16 cases (κ=0.19, 95%CI 0.18-0.21). With regard to the presence and localization of dominant strictures (Figure 1c) an agreement of over 75% of the survey participants was found in not even a single case

(K=0.16, 95%CI 0.15-0.17). In 4/16 cases over 75% of the survey participants came to an agreement with respect to CCA suspicion (Figure 1d) (K=0.13, 95%CI 0.11-0.15). The lowest kappa value was found with regard to the recommendation whether subsequently ERCP should be performed (Fig. 1e), which was consistent among more than 75% of the participants in only 4/16 cases (K=0.12, 95%CI 0.11-0.14).

Assessment of variables associated with recommendation for/against ERCP

To assess the rationale behind the survey participants´ decision to recommend for or against subsequent ERCP, a multiple choice questionnaire was used. As multiple answers were possible, the total number of votes below exceeds the number of survey participants (n=44) multiplied with the number of cases (n=16).

If ERCP was recommended (n=398), this was predominantly in order to “obtain brush cytology/histology” (n=349) followed by dilatation of strictures, including “balloon/bougie dilatation” (n=264) and “stent placement” (n=175). Less frequent arguments for ERCP recommendation were “to obtain bile for microbial culture” (n=104) and “for diagnostic purpose only” (n=17).

If ERCP was not recommended (n=306), the survey participants most frequently quoted that the “risk of ERCP outweighs the potential benefit” (n=242). Less frequently selected arguments were “too many strictures” (n=46), a “too advanced disease” (n=22) or that “ERCP is generally not appropriate for PSC treatment” (n=22) (Figure 2).

Influence of case-/rater-related factors on MRI/3D-MRCP interpretation and subsequent ERCP recommendation

The main outcome of follow-up MRI in PSC should be to identify novel or worsening dominant strictures, to suspect CCA which should then prompt ERCP and tissue

sampling or to decide upon endoscopic dilation treatment. In order to identify potential factors influencing the survey participants´ decisions generalized linear mixed effects models (GLMM) using a univariate GLMM as a first step were applied.

1. Suspicion of dominant strictures

The univariate analysis revealed no associations between symptoms, signs of cholangitis or the rater´s specialization and the frequency of suspecting dominant strictures. Presence of liver cirrhosis or elevated total bilirubin levels, however, were associated with suspecting dominant strictures. Due to the strong relationship between elevation of total bilirubin and liver cirrhosis (Chi-squared test, *P* <.01), it seems reasonable that total bilirubin as the factor with the strongest effect was responsible for that association (Odds ratio 3.7, 95% CI 1.6-8.6, *P* <.01*)*. Nevertheless, reassessing the impact of the factor liver cirrhosis while additionally controlling for total bilirubin as a random effect revealed that liver cirrhosis still was associated with a significantly higher rate of suspected dominant strictures (Odds ratio 2.8 95% CI 1.1-7.6, *P* =.02) (Table 3).

2. Suspicion of CCA

One of the main reasons for performing follow-up MRI scans in patients with PSC is the detection of early stage CCA, thus offering potential curative treatment. Two of the 16 cases (case 15 and 16) actually suffered from CCA. While in case 15 the presented history and MRI scans raised suspicion of CCA in 71% of the survey participants, in case 16 only 52% quoted that CCA was suspected.

Analyzing factors that may trigger the suspicion of CCA, significantly higher rates were only found in patients with elevated levels of total bilirubin (Odds ratio 5.2, 95% CI 2.5-

11.4, *P*<.01). Symptoms, signs of cholangitis, liver cirrhosis or the raters´ specialization did not seem to significantly impact on the frequency of CCA suspicion (Table 3).

3. ERCP recommendation

There are several accepted indications for ERCP in patients with PSC, among them signs of bacterial cholangitis with strictures or otherwise symptomatic disease with strictures amenable to treatment. In this survey, ERCP was more frequently recommended only in patients with signs of cholangitis or elevated total bilirubin. Due to the strong relationship between elevated total bilirubin and signs of cholangitis (Chi-squared test, *P*<.01), only total bilirubin, as the factor with the strongest effect, was considered significant (Odds ratio 6.4, 95% CI 2.9-15.1, *P*<.01). Controlling for total bilirubin as a random effect no longer showed a significant influence of the factor cholangitis on the rate of ERCP recommendations (Table 3).

There is a paucity of data on the role of ERCP in PSC with established cirrhosis and the recent ESGE (European Society of Gastrointestinal Endoscopy)/EASL (European Association for the Study of the Liver) guideline cautions on the use of ERCP in this population.[22] Interestingly, in this study presence of cirrhosis did not prevent raters to recommend ERCP after reviewing the MRI scans.

Comparison of disease course with survey recommendations

As all provided data in the survey were based on real-life PSC cases follow-up data were available, enabling a comparison of the real-life management and outcome with the survey participants´ recommendations.

Overall, 10/16 patients underwent ERCP after follow-up MRI/3D-MRCP was performed. In 7/10 patients ERCP was recognized to be beneficial (technically successful bile duct dilatation, positive microbiological report from bile fluid sampling,

decreasing biochemical cholestatic markers after ERCP, improvement of symptoms and/or diagnosis of CCA by brush cytology), while in 3/10 patients there was no benefit evident after ERCP. The real-life contrasted the survey participants recommendations: while the majority of the respondents (>50%) recommended ERCP in only 4 out of the 7 cases with clinical benefit, the majority of the respondents recommended ERCP in all of the 3 patients, who did not seem to have a benefit from ERCP. Likewise, ERCP was recommended by the majority of the survey participants in 3 out of 6 patients, who were – despite not undergoing ERCP in real-life – not affected by biliary complications during follow-up of 30 months (95% CI 20-33 months).

Discussion

MRI/MRCP has been established as the imaging modality of choice in patients with suspected PSC. In addition, MRI/MRCP is used for surveillance purposes by many centers,[3,6,7,11-13,15,17] although prospective data demonstrating benefit on the early detection of CCA or patient survival is lacking.[8] Imaging standards and protocols vary across institutions and MRI definitions for PSC related conditions such as dominant strictures and data on the interpretation of obtained images are lacking. Using an online questionnaire on 16 real-life PSC cases we therefore investigated how the interpretation of follow-up MRI/3D-MRCP differs among a group of PSC experts and which factors influence these experts in their decision making, particularly with respect to ERCP recommendation, which is the major decision to be taken after reviewing the MR images.

The most important finding of our study is the notable lack of agreement between the survey participants with respect to the interpretation of the given MRI/3D-MRCP data.

Despite a high expertise among the survey participants and relevant clinical information provided within each of the 16 cases, the inter-rater reliability calculation revealed kappa values not higher than 0.2 with respect to all case-related questions. Albeit statistical limitations of an inter-rater reliability calculation must be considered, in none of the case related queries therefore a level of consensus higher than what in general is claimed as a "slight agreement" was reached.[23] With respect to the recommendation to perform subsequent ERCP, a comparatively low kappa value of 0.12 was found among the survey participants. Taking into account potential complications associated with ERCP, this lack of agreement on the indication for performing ERCP is worrisome and highlights, that the interpretation of follow-up MRI/MRCP is highly challenging even for PSC experienced clinicians. Since in many centres the diagnostic report is being made by radiologists who may not be able to review the patient´s respective clinical information, these results probably underscore the situation in real life.

While in our study no association was found between the survey participants´ specialization and their MRI/3D-MRCP interpretation, total bilirubin turned out to be the variable with the strongest effect on the rate of suspected dominant strictures, suspected CCAs or recommendations for subsequent ERCP. Even though our study was not designed to draw firm conclusions on causality, this information should be considered when defining what a clinically relevant bile duct stricture could be using MRI/MRCP. These data also suggest that in particular biochemical information supports clinicians in their decision-making regarding ERCP performance in PSC.

Interestingly, next to total bilirubin the information on the presence of liver cirrhosis was associated with an increased rate of suspected dominant strictures and tended to be associated with a higher rate of ERCP recommendations. Whether these associations can be attributed to the clinical factor cirrhosis alone, to potentially more advanced bile

duct changes in cirrhotic patients or elevated bilirubin levels cannot be answered by the present study. Nevertheless, the data outline that in terms of ERCP associated complications liver cirrhosis in PSC does not prevent experienced clinicians from performing ERCP. Along this line, in the present study only a minority of survey participants gave advanced disease as an argument against ERCP (Figure 2). This contrasts the recent practice guideline of ESGE/EASLE on the use of ERCP in PSC, which cautions against the use of ERCP in PSC cirrhosis.[9]

In our study, the survey participants´ recommendations with respect to ERCP performance contrasted the real-life clinical course. Particularly, the overwhelming recommendation to perform ERCP in patients, who in real life did not seem to benefit from ERCP or were not affected by biliary complications during follow-up despite not undergoing ERCP underline the difficulty in interpretation of obtained images. These findings should, however, be interpreted with great caution, since a reasonable definition of a “beneficial” ERCP was not possible due to our study design.

In our study, the survey participants´ recommendations with respect to ERCP performance contrasted the real-life management and outcome. Particularly, the overwhelming recommendation to perform ERCP in patients, who in real-life did not seem to benefit from ERCP or were not affected by biliary complications during follow-up despite not undergoing ERCP, underline the difficulty in interpretation of obtained images. These findings should, however, be interpreted with great caution, since an evidence based definition of a “beneficial” ERCP cannot be applied.

In the recent ESGE/EASL practice guideline on endoscopy in PSC, ERCP is recommended in patients with established PSC with relevant or aggravating symptoms, new or progressive dominant strictures or with a rapid increase of cholestatic enzyme levels.[9] However, the quality of evidence underlying these recommendations is low and a MRI-based definition of dominant strictures is lacking.

Originally, the term “dominant stricture” derives from ERCP studies and defines strictures in the common bile duct of less than 1.5 mm and in the left or right or common hepatic duct of less than 1 mm.[24] Nevertheless, from our experience the term “dominant stricture” is widely used by clinicians for MRI reporting as well, and is indeed part of liver transplantation listing criteria in some countries, independent from the diagnostic technique applied.[25] Considering this and the lack of agreement among the survey participants in the present study, we suggest that the term dominant stricture should not be used in MRI reporting. A novel definition could be based on functional significance of a stricture. Currently, a new definition of relevant strictures based on MRI is being developed within the MRI working group of the IPSCSG.[8]

This is the first study to evaluate the interpretation of follow-up MRI in patients with PSC amongst a considerable number of PSC experts. However, as a survey this study was explorative by nature and has several obvious limitations. First, patient characteristics as well as the clinical and biochemical data provided within each of the 16 cases were preselected upon clinical experience and limited to the items that were thought to be most relevant. Second, despite the use of high-resolution videos, the imaging quality of the provided MRI/3D-MRCP was different from what is common standard at least on a radiological workstation. Third, to ensure a smooth feasibility of the survey and reveal potential specialization-related differences among the survey participants, we chose a non-interdisciplinary approach for the assessment of the cases. Ideally, in clinical practice MRI scans in PSC should be reviewed by multidisciplinary teams.[8] Last but not least, the survey took at least an hour to be finished and due to the randomization of cases could not be interrupted. This may have caused issues with concentration and/or motivation in some of the participants.

To summarize, our survey demonstrates that in PSC the interpretation of follow-up MRI/3D-MRCP significantly varies even among highly experienced experts. Presence of liver cirrhosis does not seem to prevent clinicians from recommending ERCP. Total serum bilirubin level was the most relevant single factor to aid clinicians in the assessment of follow-up MRI/3D-MRCP and their decision to subsequently recommend ERCP. Taking into consideration the potential of follow-up MRI/3D-MRCP for the early detection of disease related complications, there is an urgent need for the development and validation of MRI-based clinically meaningful definitions of PSC related complications.

References

1. Karlsen TH, Folseraas T, Thorburn D, et al. Primary sclerosing cholangitis - a comprehensive review. J Hepatol 2017.

2. Boonstra K, Weersma RK, van Erpecum KJ, et al. Population-based epidemiology, malignancy risk, and outcome of primary sclerosing cholangitis. Hepatology 2013;58:2045-2055.

3. European Association for the Study of the L. EASL Clinical Practice Guidelines: management of cholestatic liver diseases. J Hepatol 2009;51:237-267.

4. Lindor KD, Kowdley KV, Harrison ME, et al. ACG Clinical Guideline: Primary Sclerosing Cholangitis. Am J Gastroenterol 2015;110:646-659; quiz 660.

5. Boonstra K, Beuers U, Ponsioen CY. Epidemiology of primary sclerosing cholangitis and primary biliary cirrhosis: a systematic review. J Hepatol 2012;56:1181-1188.

6. Ehlken H, Zenouzi R, Schramm C. Risk of cholangiocarcinoma in patients with primary sclerosing cholangitis: diagnosis and surveillance. Curr Opin Gastroenterol 2017;33:78-84.

7. Razumilava N, Gores GJ, Lindor KD. Cancer surveillance in patients with primary sclerosing cholangitis. Hepatology 2011;54:1842-1852.

8. Schramm C, Eaton J, Ringe KI, et al. Recommendations on the use of MRI in PSC-A position statement from the International PSC study group. Hepatology 2017.

9. European Society of Gastrointestinal E, European Association for the Study of the Liver. Electronic address eee, European Association for the Study of the L. Role of endoscopy in primary sclerosing cholangitis: European Society of Gastrointestinal Endoscopy (ESGE) and European Association for the Study of the Liver (EASL) Clinical Guideline. J Hepatol 2017;66:1265-1281.

10. Tabibian JH, Hilscher M, Carey EJ, et al. Dominant Stricutres in Primary Sclerosing Cholangitis: A Multi-Center Survey of Clinical Perceptions and Practices. 2016;83:AB615.

11. Eaton JE, Gossard AA, Talwalkar JA. Recall processes for biliary cytology in primary sclerosing cholangitis. Curr Opin Gastroenterol 2014;30:287-294.

12. Eaton JE, Talwalkar JA, Lazaridis KN, et al. Pathogenesis of primary sclerosing cholangitis and advances in diagnosis and management. Gastroenterology 2013;145:521-536.

13. Horsley-Silva JL, Rodriguez EA, Franco DL, et al. An update on cancer risk and surveillance in primary sclerosing cholangitis. Liver Int 2017;37:1103-1109.

14. Karlsen TH, Vesterhus M, Boberg KM. Review article: controversies in the management of primary biliary cirrhosis and primary sclerosing cholangitis. Aliment Pharmacol Ther 2014;39:282-301.

15. Rizvi S, Eaton JE, Gores GJ. Primary Sclerosing Cholangitis as a Premalignant Biliary Tract Disease: Surveillance and Management. Clin Gastroenterol Hepatol 2015;13:2152-2165.

16. Ali AH, Tabibian JH, Nasser-Ghodsi N, et al. Surveillance for Hepatobiliary Cancers in Patients with Primary Sclerosing Cholangitis. Hepatology 2017.

17. Chapman R, Fevery J, Kalloo A, et al. Diagnosis and management of primary sclerosing cholangitis. Hepatology 2010;51:660-678.

18. Khan SA, Davidson BR, Goldin RD, et al. Guidelines for the diagnosis and treatment of cholangiocarcinoma: an update. Gut 2012;61:1657-1669.

19. Duff P, Goldenberg S, Deering K, et al. Barriers to Viral Suppression Among Female Sex Workers: Role of Structural and Intimate Partner Dynamics. J Acquir Immune Defic Syndr 2016;73:83-90.

20. Papp Z, Clark RG, Parmley EJ, et al. The ecology of avian influenza viruses in wild dabbling ducks (Anas spp.) in Canada. PLoS One 2017;12:e0176297.

21. Bates D, Maechler M, Bolker BM, et al. Fitting Linear Mixed-Effects Models Using lme4. J Stat Softw 2015;67.

22. Aabakken L, Karlsen TH, Albert J, et al. Role of endoscopy in primary sclerosing cholangitis: European Society of Gastrointestinal Endoscopy (ESGE) and European Association for the Study of the Liver (EASL) Clinical Guideline. Endoscopy 2017;49:588-608.

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

24. Stiehl A, Rudolph G, Kloters-Plachky P, et al. Development of dominant bile duct stenoses in patients with primary sclerosing cholangitis treated with ursodeoxycholic acid: outcome after endoscopic treatment. J Hepatol 2002;36:151-156.

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

TABLE 1 Study questionnaire

	Possible answers
1. General questionnaire	
(a) Survey participants´ specialization	Hepatology (including Endoscopy) Gastroenterology Radiology

(b) Survey participants´ experience with PSC patients	Time in years
(c) Yearly number of PSC patients treated at the survey participants´ centre	<50 51-150 >150
(d) Performance of follow-up MRI/MRCP at the survey participants´ centre	Yearly Every other year Irregularly Not at all
2. Case-related questionnaire	
(a) Overall MRI/3D-MRCP interpretation	Typical for PSC Compatible with PSC Atypical for PSC
(b) Presence of bile duct changes	Intra- and extrahepatic Intrahepatic only Extrahepatic only No bile duct changes
(c) Suspicion of dominant strictures	Intra- and extrahepatic Intrahepatic only Extrahepatic only No dominant strictures
(d) Suspicion of CCA	Yes No
(e) ERCP recommendation	Yes No
(f) If ERCP was recommended, for what reason (multiple answers possible)	To perform balloon/bougie dilation To place a stent To obtain brush cytology/histology To obtain bile for microbial culture For diagnostic purpose only
(f) If ERCP was not recommended, for what reason (multiple answers possible)	Disease was too advanced Too many strictures Risk of ERCP outweighs potential benefit ERCP is generally not appropriate for PSC treatment

Survey participants were asked for their personal/centre´s experience and practice policy in PSC (1. general questionnaire) and at the end of each of the provided 16 real-life PSC cases (2. case related questionnaire). Except for the years of experience with PSC patients (1b) a multiple-choice questionnaire was used. Note, that in 2f multiple answers were possible. CCA, cholangiocarcinoma; ERCP endoscopic retrograde

cholangiopancreatography; MRCP, magnetic resonance cholangiopancreatography; MRI, magnetic resonance imaging; PSC, primary sclerosing cholangitis.

TABLE 2 Survey participants´ characteristics

Feature	**Number of survey participants (%)**
Total number of survey participants	44 (100)
Primary specialization Hepatology Gastroenterology Radiology	 19 (43) 16 (36) 9 (21)

Annual number of PSC patients treated at the survey participants centres < 50 patients 51-150 patients > 150 patients	 14 (32) 22 (50) 8 (18)
Performance of follow-up MRI/MRCP at the survey participants centres Yearly Every other year Irregularly Not at all	 23 (52) 8 (18) 8 (18) 5 (11)

44 participants with a median clinical experience in the care of patients with primary sclerosing cholangitis (PSC) of 14 (7-20) years completed the survey. MRCP, magnetic resonance cholangiopancreatography; MRI, magnetic resonance imaging.

TABLE 3. Influence of case-/rater-related factors on follow-up MRI/3D-MRCP interpretation/ERCP recommendation

Increased rate of	Evaluated patient-/rater-related factors	Odds ratio (95% CI)	*P-value*
suspected dominant stricture(s)	Liver cirrhosis present	**2.8 (1.1-7.6)**	**.02†**
	Symptoms present	1.1 (0.4-3.3)	.79
	Cholangitis present	2.5 (0.8-7.9)	.10
	Total bilirubin elevation present	**3.7 (1.6-8.6)**	**<.01**
	Rater´s specialization: hepatology	0.6 (0.3-1.0)	.06
	Rater´s specialization: gastroenterology	1.2 (0.6-2.3)	.55
	Rater´s specialization: radiology	1.8 (0.8-4.1)	.12

suspected cholangio-carcinoma	Liver cirrhosis present	1.8 (0.6-5.5)	.25
	Symptoms present	0.6 (0.2-1.8)	.31
	Cholangitis present	1.5 (0.4-5.6)	.51
	Total bilirubin elevation present	**5.2 (2.5-11.4)**	**<.01**
	Rater´s specialization: hepatology	1.1 (0.5-2.1)	.87
	Rater´s specialization: gastroenterology	1.1 (0.6-2.3)	.71
	Rater´s specialization: radiology	0.8 (0.3-1.8)	.52
recommendation to perform subsequent ERCP	Liver cirrhosis present	2.9 (0.9-9.4)	.06
	Symptoms present	1.7 (0.5-6.1)	.38
	Cholangitis present	4.2 (1.2-15.7)	.02‡
	Total bilirubin elevation present	**6.4 (2.9-15.1)**	**<.01**
	Rater´s specialization: hepatology	1.1 (0.5-2.5)	.74
	Rater´s specialization: gastroenterology	1.4 (0.6-3.2)	.41
	Rater´s specialization: radiology	0.5 (0.2-1.3)	.16

Univariate generalized linear mixed effects models (GLMM) were used to analyze potential factors associated with a higher rate of suspected dominant strictures, cholangiocarcinoma or ERCP recommendations. In case of significant results in more than one evaluated factor and an interference between these factors (evaluated with a post-hoc test), only the factor with the strongest level of significance was considered meaningful. † Nevertheless, reassessing the impact of the factor liver cirrhosis on the rate of suspected dominant strictures while additionally controlling for total bilirubin as a random effect revealed that liver cirrhosis still was associated with a significantly higher rate of suspected dominant strictures. ‡ Contrary, reassessing the impact of the factor cholangitis on the rate of ERCP recommendations while additionally controlling for total bilirubin levels revealed that cholangitis was no longer significantly associated with a higher rate of ERCP recommendations. For details, please see the methods section. ERCP, endoscopic retrograde cholangiopancreatography; MRCP, magnetic resonance cholangiopancreatography; MRI, magnetic resonance imaging.

FIGURE 1. Analysis of agreement in the interpretation of follow-up MRI/3D-MRCP among the survey participants. Graphical presentation of the agreement within a group of 44 PSC-experienced clinicians in the interpretation of 16 follow-up MRI/3D-MRCP of patients with established PSC. The case related queries involved the overall

MRI/3D-MRCP interpretation (A), the presence of biliary changes (B), whether dominant strictures (C) or cholangiocarcinoma (D) were suspected and the resulting indication to perform endoscopic retrograde cholangiopancreatography (ERCP) (E). Each bar represents the relative number of survey participants, who chose one of the possible answers presented on the right hand side of the graphs. The inter-rater reliability of agreement was calculated using the Fleiss´ kappa value (K). MRCP, magnetic resonance cholangiopancreatography; MRI, magnetic resonance imaging; PSC, primary sclerosing cholangitis.

FIGURE 2. Assessment of variables associated with recommendation for/against ERCP. Absolute number of arguments for/against endoscopic retrograde cholangiopancreatography (ERCP) chosen by 44 PSC-experienced survey participants on the basis of follow-up MRI/3D-MRCP of 16 patients with established PSC. Note, that multiple answers were possible. MRCP, magnetic resonance cholangiopancreatography; MRI, magnetic resonance imaging; PSC, primary sclerosing cholangitis.

FIGURE S1. Example of a case presentation in the survey.