Primary Sclerosing Cholangitis: Meta-Analysis of Diagnostic Performance of MR Cholangiopancreatography

Maneesh Dave, MD, MPH
B. Joseph Elmunzer, MD
Ben A. Dwamena, MD
Peter D. R. Higgins, MD, PhD, MSc

Purpose:
To determine the diagnostic accuracy of magnetic resonance cholangiopancreatography (MRCP) for detection of primary sclerosing cholangitis (PSC) in patients with biochemical cholestasis.

Materials and Methods:
Two reviewers searched MEDLINE, EMBASE, and other electronic databases to identify prospective studies in which MRCP was evaluated and compared with endoscopic retrograde cholangiopancreatography (ERCP), clinical examination, and/or histologic analysis for diagnosis of PSC in cholestasis and control cases. Main study inclusion criteria were (a) use of ERCP or percutaneous transhepatic cholangiography (PTC) as part of the reference standard for the diagnosis of PSC, (b) inclusion of patients with hepatobiliary disease other than PSC (ie, nonhealthy control subjects), (c) blinding of MRCP image readers to reference-standard results, (d) prospective study with ERCP or MRCP performed after subject recruitment into the study, and (e) inclusion of raw data (for true-positive, false-positive, true-negative, and false-negative results) that could be found or calculated from the original study data. Major exclusion criteria were duplicate article (on a primary study) that contained all or some of the original study data and inclusion of fewer than 10 patients with PSC. Methodologic quality was assessed by using the Quality Assessment of Diagnostic Accuracy Studies tool. Bivariate random-effects meta-analytic methods were used to estimate summary, sensitivity, specificity, and receiver operating characteristic (ROC) curves.

Results:
Six manuscripts with 456 subjects (with 623 independent readings)—185 with PSC—met the study inclusion criteria. The summary area under the ROC curve was 0.91. High heterogeneity (inconsistency index, 78%) was found but became moderate (inconsistency index, 36%) with the exclusion of one study in which the diagnostic threshold was set for high sensitivity. There was no evidence of publication bias ($P = .27$, bias coefficient analysis). Sensitivity and specificity of MRCP for PSC detection across all studies were 0.86 and 0.94, respectively. Positive and negative likelihood ratios with MRCP were 15.3 and 0.15, respectively. In patients with high pretest probabilities, MRCP enabled confirmation of PSC; in patients with low pretest probabilities, MRCP enabled exclusion of PSC. Worst-case-scenario (pretest probability, 50%) posttest probabilities were 94% and 13% for positive and negative MRCP results, respectively.

Conclusion:
MRCP has high sensitivity and very high specificity for diagnosis of PSC. In many cases of suspected PSC, MRCP is sufficient for diagnosis, and, thus, the risks associated with ERCP can be avoided.

Primary sclerosing cholangitis (PSC) is a chronic cholestatic liver disease characterized by diffuse inflammation and fibrosis of the medium and large bile ducts (1,2). PSC progresses to frequent biliary strictures and cholangitis that require regular endoscopic intervention and eventually to liver failure that requires liver transplantation, resulting in substantial medical costs. In a patient with biochemical evidence of cholestasis, PSC is often suspected. Anatomic evaluation of the biliary tree is necessary for the diagnosis of PSC, and endoscopic retrograde cholangiopancreatography (ERCP) has been considered the reference standard for this diagnosis (3). However, ERCP is an invasive procedure associated with complications that include pancreatitis, biliary sepsis, bleeding, perforation, and aspiration (4–6). Diagnostic ERCP is associated with a major complication rate of 1.3% and a mortality rate of 0.21% (5). In contrast, magnetic resonance cholangiopancreatography (MRCP) enables noninvasive imaging of the biliary and pancreatic trees and has been estimated to involve 11.7% lower costs compared with ERCP when sedation and supply costs are included (7).

Investigators in a number of studies have evaluated the sensitivity and specificity of MRCP for the diagnosis of biliary and pancreatic diseases. In a meta-analysis performed by Romagnuolo et al (8), MRCP had excellent sensitivity and specificity for demonstrating the presence and level of biliary obstruction; however, the authors concluded that the performance of MRCP in the diagnosis of PSC was inadequately studied. The predictive value of MRCP in patients suspected of having PSC is not well established because varying sensitivities and specificities have been reported. It is not clear whether MRCP is a sufficient test for the diagnosis of PSC in patients with clinical findings and biochemical evidence of cholestasis. A better understanding of the sensitivity, specificity, summary receiver operating characteristic (ROC) curve, and positive and negative likelihood ratios (with corresponding 95% confidence intervals [CIs]) of MRCP in the setting of suspected PSC is needed. In this study, we aimed to determine the accuracy of MRCP for the diagnosis of PSC in patients with biochemical cholestasis.

Materials and Methods

Search Strategy

This meta-analysis was performed according to guidelines for the systematic review and meta-analysis of diagnostic studies (9–13). We used the checklists described in the MOOSE (meta-analyses of observational studies) and QUOROM (Quality of Reporting of Meta-analyses) statements for reporting meta-analyses (11,12). A literature search of PubMed from its inception to September 2009; of EMBASE from 1993 through September 30, 2009; and of American College of Physicians Journal Club, DARE (Database of Abstracts of Reviews of Effects), and Google Scholar databases was conducted. The search involved the use of free-text words and MeSH (Medical Subject Headings) terms for increased sensitivity of the search strategy (13,14). The search strategy used in PubMed is given in Figure E1 (online). An information library specialist at the University of Michigan also helped us in the bibliographic search.

Abstracts from the 2002–2008 annual meetings of the American Gastroenterological Association, the American Society of Gastrointestinal Endoscopy, the American College of Gastroenterology, United European Gastroenterology Week, and Digestive Disease Week were reviewed for potentially relevant studies. We also consulted with experts in the field to identify additional published and nonpublished primary studies. No nonpublished, non–peer reviewed studies were included in the review. The search was not restricted to English-language publications. In addition, after study selection, we reviewed the bibliographies of the studies to identify any additional studies on the MRCP-based diagnosis of PSC.

Study Selection

Two independent reviewers—a physician with 2½ years experience in public health (M.D.) and a gastroenterologist with 4 years experience in advanced endoscopy (B.J.E.)—read the abstracts of all candidate articles and retrieved the full texts of the published manuscripts that could not be excluded on the basis of the title and abstract alone. The articles were independently read and checked for inclusion criteria. There was 100% agreement between these reviewers for study selection.

Study Inclusion and Exclusion Criteria

The selection criteria for inclusion in our current study were (a) use of ERCP or percutaneous transhepatic...
cholangiography (PTC) as part of the reference standard for the diagnosis of PSC, (b) inclusion of patients with hepatobiliary disease other than PSC (nonhealthy control patients), (c) blinding of MRCP image readers to the reference-standard results, (d) prospective study in which at least ERCP or MRCP was performed after recruitment of the subjects into the study, and (e) inclusion in the study of raw data (for true-positive, false-positive, true-negative, and false-negative results) that could be found or calculated from data in the original published study. Exclusion criteria were a duplicate article (on a primary study) that contained all or some of the original publication data, inclusion of fewer than 10 patients with PSC, and inclusion of healthy control subjects—specifically, individuals without cholestatic liver disease.

Data Extraction
Two independent reviewers extracted the following data from the selected studies: study characteristics (design, country of study origin, year of publication, setting or location, sample size, diagnostic criteria for PSC, time interval between MRCP examination and ERCP with or without a diagnosis of PSC, and criterion standards), population (mean age, sex, number of patients with ulcerative colitis, number of patients with Crohn disease), interventions (criteria for diagnosis, imager manufacturer, magnetic field strength for MRCP, MRCP technique, radiologist versus gastroenterologist results interpretation), and outcomes (true-positive, false-positive, true-negative, false-negative). In one study, translation of the original German-language text was required and was performed by a local German-language instructor. In a second study, translation of the French-language text by a fluent French speaker was required (15). Another variable sought was the method(s) for handling indeterminate or missing data. Discrepancies were resolved by discussion and consensus with the senior author (P.D.R.H.), a gastroenterologist with 10 years experience. The two independent reviewers (M.D., B.J.E.) also assessed the study quality by evaluating the studies for the 14 items in the Quality Assessment of Diagnostic Accuracy Studies guidelines (16), with discrepancies resolved by the senior author.

Data Synthesis and Statistical Analyses
The primary outcome in the analyses was the performance of MRCP in the diagnosis of PSC, as quantified in terms of sensitivity, specificity, diagnostic log odds ratios, and likelihood ratios, with corresponding 95% confidence intervals.

Meta-analytical model.—We applied an exact binomial rendition of the bivariate mixed-effects regression model that was developed by von Houwelingen et al (17) for treatment trial meta-analysis and modified for diagnostic test data by Reitsma et al (18). This model hierarchically accounts for within- and between-study variability. At the first (individual patient) level, the number of individuals with disease who have true-positive results and the number of individuals without disease who have true-negative results are assumed to follow independent binomial distributions (19). The second level assumes normally distributed random effects (between-study variability) for the disease groups (ie, logit-transformed sensitivity and specificity), with the degree of correlation between them being predictive of an implicit threshold effect.

Summary performance estimates.—We calculated summary sensitivity and specificity after the antilogit transformation of estimated model parameters. Corresponding positive likelihood, negative likelihood, and diagnostic odds ratios were derived as functions of these summary estimates. We also used the derived estimates of sensitivity, specificity, and respective variances to construct a summary ROC curve (20,21). The area under the ROC curve was used as an alternative global measure of test performance (22). The diagnostic odds ratio (DOR) is calculated as follows: DOR = LR+ /LR− , where LR+ is the positive likelihood ratio and LR− is the negative likelihood ratio. The diagnostic odds ratio is defined as the odds of having a positive test result among patients without the disease. This measure is a single indicator of test accuracy that comprises a combination of sensitivity and specificity information (23). The posttest probability of PSC (Ppost) was calculated from likelihood ratios by using the Bayes theorem as follows: Ppost = (LR × Ppre) / [(1 − Ppre) × (1 − LR)], where Ppre is the pretest probability—that is, the suspicion for PSC (24).

The clinical utility of a diagnostic test can be assessed on the basis of its likelihood ratios. The positive likelihood ratio was the measure of the likelihood that a positive MRCP result (PSC present) would occur in a patient with PSC, while the negative likelihood ratio was the measure of the likelihood that a negative MRCP result (PSC absent) would occur in a patient without PSC. Likelihood ratios higher than 10 and lower than 0.1 indicate that the given test can generate strong evidence to rule in or rule out a diagnosis, respectively.

The heterogeneity (ie, between-study variation) of the results between the studies was assessed graphically by using forest plots, with statistical assessments performed by using the χ² test of homogeneity and the inconsistency index. Inconsistency indexes describe the percentage of total variation across studies that is due to heterogeneity rather than to chance. An inconsistency index of 0% indicates no observed heterogeneity, and values greater than 50% may be considered to indicate substantial heterogeneity (25). A perfect test has a summary ROC curve with an area close to 1.0, and poor tests have an area under the ROC curve close to 0.5. Publication bias was assessed visually by using a scatterplot of the inverse of the square root of the effective sample size (1/ESS²) versus the diagnostic log odds ratio, which would have a symmetric funnel shape when publication bias was absent. Formal testing for publication bias was conducted by using a regression of the diagnostic log odds ratio against 1/ESS² and weighting according to the effective sample size, with P < .10 indicating significant asymmetry (26). We analyzed the data by using the MIDAS (Meta-analytical Integration of

Radiology: Volume 256: Number 2—August 2010 • radiology.rsna.org 389
Results

Literature Search

The results of the literature search are summarized in Figure 1. There was 100% agreement between the reviewers regarding selection of the included studies.

Seven studies were excluded because they did not meet the prespecified criteria (28–34), as seen in Figure 1. The selected studies involved a total of 456 subjects with 623 independent MRCP image readings. In the Berstad et al (35) and Fulcher et al (36) studies, the MRCP and ERCP images were reviewed and analyzed by two radiologists independently, without a consensus reading. Therefore, we chose our unit of analysis to be the MRCP image reading by a single radiologist and included the results from both readers in the meta-analysis. We also performed a sensitivity analysis to test the validity of this assumption by excluding the diagnoses of the most accurate readers and then excluding the diagnoses of the least accurate readers from each of these two studies to determine whether these changes would significantly alter the summary statistics. In the Weber et al study (15), in which several MRCP sequences were evaluated, we used the data from the rapid acquisition with relaxation enhancement sequence for our analysis. The study conducted by Ferrara et al (37) was unique in that it was performed with children, and the Textor et al study (38) involved the largest number of subjects. The overall prevalence of PSC among the study subjects was 41% (258 of 623 readings) (Table 1).

Assessment of Study Quality and Publication Bias

All studies (6) included in the meta-analysis fulfilled nine or more of the 14 criteria in the Quality Assessment of Diagnostic Accuracy Studies tool for methodologic quality. The results of this assessment are presented in Table 2. Results of the funnel plot asymmetry test for publication bias analyzed by means of linear regression of the log odds ratio on effective sample size were not significant (bias coefficient, −0.20; P = .27), and the slope was not significant, suggesting no major publication bias (Fig 2). The Galbraith plot to evaluate the normality assumption had no significant outliers (Fig E2 [online]).

Diagnostic Accuracy

The sensitivity and specificity of MRCP for the diagnosis of PSC were 0.86 (95% CI: 0.80, 0.90) and 0.94 (95% CI: 0.86, 0.98), respectively (Table 3). The diagnostic odds ratio was 101 (95% CI: 38, 268) (Table 3). The inconsistency index for the overall heterogeneity of the study was 78% (95% CI: 51%, 100%), which was considered to indicate significant heterogeneity.

Evaluation of Clinical Utility

The positive and negative likelihood ratios of MRCP for the diagnosis of PSC were 15.3 (95% CI: 6.2, 38.1) and 0.15 (95% CI: 0.11, 0.21) (Fig E3 [online]), respectively. We used the likelihood ratios to simulate three clinical scenarios by using different pretest probabilities of PSC (25%, indicating low clinical suspicion; 75%, indicating high clinical suspicion; and 50%, indicating worst-case scenario), and using the likelihood ratios, we calculated and plotted posttest probabilities on Fagan nomograms (Fig 5). With a pretest probability of PSC of 25% (low clinical suspicion), the posttest probability of PSC, given a negative MRCP result, was 5%, which could be considered sufficient to rule out PSC (Fig 5a). With a pretest probability of PSC of 75% (high clinical suspicion), the posttest probability of PSC, given a
### Table 1

<table>
<thead>
<tr>
<th>Study ID No. and Authors</th>
<th>Year</th>
<th>Country</th>
<th>No. of Readings</th>
<th>Patient Age (y)*</th>
<th>TP</th>
<th>FN</th>
<th>FP</th>
<th>TN</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>PPV</th>
<th>NPV</th>
<th>MR Sequence†</th>
<th>Reference Standard</th>
<th>Time Interval‡</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Angulo et al (1)</td>
<td>2000</td>
<td>United States</td>
<td>70</td>
<td>56</td>
<td>19</td>
<td>4</td>
<td>1</td>
<td>46</td>
<td>83</td>
<td>98</td>
<td>95</td>
<td>92</td>
<td>Single-shot spin echo</td>
<td>ERCP or PTC, medical history, histologic analysis, biochemical testing, and clinical follow-up</td>
<td>Within 24 h</td>
</tr>
<tr>
<td>2. Berstad et al (35)</td>
<td>2006</td>
<td>Norway</td>
<td>66</td>
<td>44</td>
<td>32</td>
<td>7</td>
<td>2</td>
<td>25</td>
<td>82</td>
<td>93</td>
<td>94</td>
<td>78</td>
<td>Thick-slab MRC and thin-section half Fourier</td>
<td>ERCP or PTC, medical history, liver biopsy, and biochemical testing</td>
<td>Within 48 h</td>
</tr>
<tr>
<td>3. Berstad et al (35)</td>
<td>2006</td>
<td>Norway</td>
<td>66</td>
<td>44</td>
<td>30</td>
<td>9</td>
<td>5</td>
<td>22</td>
<td>77</td>
<td>81</td>
<td>86</td>
<td>71</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Ferrara et al (37)</td>
<td>2002</td>
<td>United States</td>
<td>21</td>
<td>9</td>
<td>13</td>
<td>3</td>
<td>0</td>
<td>5</td>
<td>81</td>
<td>100</td>
<td>100</td>
<td>63</td>
<td>RARE and half-Fourier RARE</td>
<td>ERCP and liver biopsy</td>
<td>&lt;3 wk</td>
</tr>
<tr>
<td>5. Fulcher et al (36)</td>
<td>2000</td>
<td>United States</td>
<td>102</td>
<td>NR</td>
<td>30</td>
<td>4</td>
<td>2</td>
<td>66</td>
<td>88</td>
<td>97</td>
<td>94</td>
<td>94</td>
<td>Half-Fourier RARE</td>
<td>ERCP with or without liver biopsy or US, and CT§</td>
<td>&lt;117 mo</td>
</tr>
<tr>
<td>6. Fulcher et al (36)</td>
<td>2000</td>
<td>United States</td>
<td>102</td>
<td>NR</td>
<td>29</td>
<td>5</td>
<td>5</td>
<td>63</td>
<td>85</td>
<td>92</td>
<td>85</td>
<td>93</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Textor et al (38)</td>
<td>2002</td>
<td>Germany</td>
<td>142</td>
<td>48.6</td>
<td>29</td>
<td>4</td>
<td>1</td>
<td>108</td>
<td>88</td>
<td>99</td>
<td>97</td>
<td>96</td>
<td>T2-weighted 3D FSE with respiratory triggering</td>
<td>ERCP, medical history, and biochemical testing*</td>
<td>&lt;15 d</td>
</tr>
<tr>
<td>8. Weber et al (15)</td>
<td>2003</td>
<td>Germany</td>
<td>54</td>
<td>40</td>
<td>39</td>
<td>1</td>
<td>5</td>
<td>9</td>
<td>97</td>
<td>64</td>
<td>89</td>
<td>90</td>
<td>RARE, half-Fourier RARE, and thin-section half Fourier RARE</td>
<td>ERCP with or without liver biopsy</td>
<td>&lt;1 mo</td>
</tr>
</tbody>
</table>

Total 623 221 37 21 344

Note.—The study identification (ID) numbers correspond to study numbers on the graphs in Figures 2 and 3. In the Berstad et al (35) and Fulcher et al (36) studies, the MRCP and ERCP images were reviewed and analyzed by two radiologists independently, without a consensus reading. Sensitivity, specificity, negative predictive value (NPV), and positive predictive value (PPV) are percentages. FN = false-negative results, FP = false-positive results, TN = true-negative results, TP = true-positive results.

* Unless otherwise noted, data are mean ages. NR = not reported.
† FSE = fast spin echo, MRC = MR cholangiography, RARE = rapid acquisition with relaxation enhancement, 3D = three dimensional.
‡ Time interval between MRCP and ERCP. In the Berstad et al study (35), the time interval was within 48 hours in all but nine cases.
§ Median age.
* Control subjects did not undergo ERCP because it was already known that they had hepatobiliary diseases other than PSC. CT = computed tomography, US = ultrasonography.
* Not all control subjects underwent ERCP.
EVIDENCE-BASED PRACTICE: MR Cholangiopancreatography of Primary Sclerosing Cholangitis

Dave et al

positive MRCP result, was 98%; thus, a positive MRCP result could be considered sufficient to rule in PSC (Fig 5c).

With a pretest probability of PSC of 50% (worst-case scenario), the posttest probability of PSC, given a positive MRCP result, was 94%, and the posttest probability of PSC, given a negative MRCP result, was 13%. Thus, MRCP is a useful test in this situation. Summary curves on which posttest probabilities are plotted against all possible pretest probabilities are presented in Figure E4 (online).

Effect of Diagnostic Threshold

The sensitivities and specificities of MRCP for the diagnosis of PSC in each study are displayed on the summary

Table 3

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensitivity</td>
<td>0.86 (0.80, 0.90)</td>
</tr>
<tr>
<td>Specificity</td>
<td>0.94 (0.86, 0.98)</td>
</tr>
<tr>
<td>Positive LR</td>
<td>15.3 (6.2, 38.1)</td>
</tr>
<tr>
<td>Negative LR</td>
<td>0.15 (0.11, 0.21)</td>
</tr>
<tr>
<td>Diagnostic score</td>
<td>5 (4, 6)</td>
</tr>
<tr>
<td>DOR</td>
<td>101 (38, 268)</td>
</tr>
</tbody>
</table>

Effect of Diagnostic Threshold

The sensitivities and specificities of MRCP for the diagnosis of PSC in each study are displayed on the summary
EVIDENCE-BASED PRACTICE: MR Cholangiopancreatography of Primary Sclerosing Cholangitis

Figure 4

ROC diagram in Figure 3. The summary ROC curve and summary operating point (with a 95% confidence contour) are shown in relation to the individual studies. The area under the summary ROC curve was 0.91, consistent with high diagnostic accuracy. The Weber et al (15) investigation (study 8) can be seen as a relative outlier, in which the investigators chose a diagnostic threshold with greater sensitivity and less specificity compared with those in the other studies. This difference in diagnostic threshold is due to the investigators’ choice to value sensitivity more than the investigators in the other studies did, and this could have contributed substantially to the heterogeneity of this meta-analysis. We investigated this by performing a sensitivity analysis in which we excluded the Weber et al study.

Figure 4: Forest plots of studies involving evaluation of the sensitivity and specificity of MRCP in the diagnosis of PSC in patients with cholestasis. Summary sensitivity and specificity of MRCP for the diagnosis of PSC were 0.86 (95% CI: 0.80, 0.90) and 0.94 (95% CI: 0.86, 0.98), respectively.

Figure 5

Figure 5: Pretest probabilities (Prob) and likelihood ratios (LR). (a) With a pretest probability of PSC of 25% (low clinical suspicion), the posttest probability of PSC, given a negative MRCP result (Post-Neg Probability), is 5%, which can be considered sufficient to rule out PSC. (b) With a pretest probability of PSC of 50% (worst-case scenario), the posttest probabilities of PSC, given positive and negative MRCP results, are 94% and 13%, respectively. Thus, it is a useful test in this situation. (c) With a pretest probability of PSC of 75% (high clinical suspicion), the posttest probability of PSC, given a positive MRCP result (Post-Pos Probability), is 98%; thus, a positive MRCP result can be considered sufficient to rule in PSC.
Sensitivity Analyses
After excluding the Weber et al study (15) from the meta-analysis, the inconsistency index for heterogeneity decreased substantially, from 78% (high) to 36% (moderate), with only minimal changes in the summary statistics (Table 4).

We also evaluated the effect of having two readers each in the Berstad et al (35) and Fulcher et al (36) studies by performing two sensitivity analyses in which we used only one reading per patient, sequentially excluding either the worst or best readers from each study. These exclusions produced minimal changes in sensitivity and specificity and nonsignificant changes in likelihood ratios and diagnostic odds ratios, as seen in Table 4.

Discussion
The meta-analysis results reported herein show that MRCP has excellent accuracy in the diagnosis of PSC, with an area under the ROC curve of 0.91. This result supports the findings of the cost-effectiveness analysis performed by Meagher et al (39), which suggested that an initial MRCP examination with negative results followed by ERCP would be a cost-effective approach to diagnosing PSC if it was sufficiently sensitive and specific. The results of the cost minimization analysis performed by Talwalkar et al (7) also support the use of MRCP first, with selective use of ERCP following, as a way of minimizing costs. Thus, we believe that MRCP should gain acceptance as the initial diagnostic study for PSC because it yields accurate cholangiographic images, similar to those acquired with ERCP, without the use of ionizing radiation and with a lower risk of adverse events.

Although ERCP is considered the reference standard for the diagnosis of PSC, it can result in false-positive diagnoses if complete biliary tract distention mimics the ductal irregularities of PSC or result in false-negative diagnoses if high-grade strictures cause inadequate opacification of the intrahepatic ducts (36). In the Angulo et al study (1), intrahepatic biliary ducts proximal to high-grade strictures were not seen during ERCP in three patients with PSC, while MRCP enabled visualization of the entire biliary tree. Since the studies included in our meta-analysis often involved the use of a combined clinical, biochemical, and ERCP or PTC endpoint as the reference standard for the diagnosis of PSC, the summary characteristics for MRCP might be even better in a direct comparison of MRCP versus ERCP diagnosis of PSC. In the retrospective investigation conducted by Moff et al (28), which was excluded owing to its study design, ERCP and MRCP were directly compared for the diagnosis of PSC, and ERCP had lower sensitivity and higher specificity than did MRCP. It is noteworthy that Fulcher et al (36) found that only those patients with cirrhosis received false-positive diagnoses of PSC with MRCP; this may be an important limitation of MRCP in the diagnosis of PSC and merits further study.

The summary results of this meta-analysis had high statistical heterogeneity. The heterogeneity in a meta-analysis can have multiple sources, including differences in the test or study procedures, differences in the subject populations, differences in the study designs, or a combination of these factors (40). It is recognized that an additional source of heterogeneity in meta-analyses
diagnostic accuracy is variability in the choice of the diagnostic threshold for a positive test result (41). The Weber et al study (15), which was unlike the other investigations in that a threshold that favored sensitivity over specificity was used, was the main source of the heterogeneity in this meta-analysis, and excluding it greatly reduced the heterogeneity without substantially changing the summary estimates.

There were several limitations to our study, including the moderate sample size; however, the resulting CIs were relatively narrow. Another limitation was the inherent heterogeneity in the study designs. The optimal protocol for performing MRCP has not been defined, and differences in the MR instrument manufacturers and magnetic field strengths between different centers could have been a cause of the between-study variations. With the limited number of studies available, the significance test for publication bias was not especially powerful, and it is possible that additional negative-result studies have not been published. The accuracy of MRCP in the diagnosis of PSC may be influenced by the disease spectrum. In the Angulo et al study (1) included in our meta-analysis, PSC was prospectively diagnosed. Although disease severity was not extensively studied, among four MRCP-derived false-negative diagnoses of PSC, two cases involved very mild changes related to PSC, causing no dilatation at ERCP. This suggests that although very mild cases may be difficult to diagnose with MRCP, they may be similarly difficult to diagnose with ERCP. This may be another limitation of MRCP. In the studies included in our meta-analysis, disease severity was not described in much detail; thus, we think that a study specifically addressing this factor should be conducted in the future.

In a majority of the included studies, static ERCP images were interpreted by radiologists as opposed to the real-time interpretation of images by gastroenterologists, and this could have led to lower sensitivity and specificity for ERCP. In the Fulcher et al study (36), the control subjects did not undergo ERCP because of their previously established diagnosis of hepatobiliary disease other than PSC (ie, hepatitis C). Similarly, in the Textor et al study (38), some control subjects did not undergo ERCP; however, on the basis of their final diagnoses, 72 subjects were healthy and 39 had cholangiocarcinoma. In addition, the included studies were performed at tertiary care centers where the staff members had increased experience with PSC and MRCP. The included studies did not address whether the high level of diagnostic accuracy of MRCP for PSC is generalizable to community centers. Although we pooled multiple small studies to achieve a larger sample size, a multicenter study with real-world accuracy (in smaller centers and community hospitals) would be helpful in proving the generalizability of our meta-analysis data to centers with little MRCP experience.

Our meta-analysis results show that MRCP has high sensitivity and very high specificity for the diagnosis of PSC. The results of this meta-analysis and of previous cost-minimization and cost-effectiveness studies show that the initial diagnostic test in a patient suspected of having PSC should be MRCP and that therapeutic ERCP should follow—but only if it is needed. Confirmatory ERCP should be considered for cases in which the clinical findings are inconsistent with a diagnosis of PSC, cases involving underlying cirrhosis (given the possibility of a false-positive diagnosis with MRCP), cases in which the clinical suspicion is moderately high (pretest probability for PSC > 50%), cases with negative MRCP results, and/or cases being evaluated at centers where the technical expertise with MRCP is not well established.

Acknowledgments: We thank Andrew Smith and Amy Higgins for translating the text in some studies to English.

References

14. van der Weijden T, Uzermans CJ, Dinant GJ, van Duijn NP, de Vet R, Buntinx F. Identifying relevant diagnostic studies in MEDLINE: the diagnostic value of the erythrocyte
EVIDENCE-BASED PRACTICE: MR Cholangiopancreatography of Primary Sclerosing Cholangitis

Dave et al


