Magnetic resonance cholangiopancreatography: a meta-analysis of test performance in suspected biliary disease
Romagnuolo J, Bardou M, Rahme E, Joseph L, Reinhold C, Barkun A N

CRD summary
This review assessed the diagnostic accuracy of magnetic resonance cholangiopancreatography (MRCP) in suspected biliary obstruction. The authors concluded that MRCP has high sensitivity and specificity for the level and presence of biliary stones, but is less accurate for detecting stones or malignancy. The authors’ conclusions are likely to be reliable.

Authors’ objectives
To assess the diagnostic accuracy of magnetic resonance cholangiopancreatography (MRCP) in patients with suspected biliary obstruction.

Searching
MEDLINE was searched from January 1987 to March 2003 for studies published in English or French. The reference lists in identified studies and reviews were screened. Two international experts in the field were contacted for details of any additional unpublished studies.

Study selection
Study designs of evaluations included in the review
Comparative studies were eligible for inclusion. Studies were excluded if they were: case series of patients with a particular diagnosis; studies in which only patients with a positive MRCP result had the reference standard test applied; or studies in which only patients testing positive on the reference standard test had an MRCP.

Only studies with more than 5 patients were included in the assessment of stone detection.

Specific interventions included in the review
Studies of MRCP were eligible for inclusion.

Reference standard test against which the new test was compared
Studies that compared MRCP with a single or composite acceptable reference standard test were eligible for inclusion. The paper listed the acceptable reference standard tests. The reference standards used in the included studies were intra-operative cholangiography followed by endoscopic retrograde cholangiopancreatography (ERCP) or ductal exploration, intra-operative cholangiography alone, or intravenous cholangiography plus ERCP plus surgery.

Participants included in the review
Studies of patients with suspected biliary obstruction were eligible for inclusion.

Outcomes assessed in the review
Studies that assessed the presence, level or cause of biliary obstruction, and presented sufficient data to allow the construction of 2x2 contingency tables, were eligible for inclusion. The review assessed the presence of common bile duct stones and the differentiation of benign from malignant obstruction.

How were decisions on the relevance of primary studies made?
More than one reviewer selected the studies, but no further details were reported in the paper.

Assessment of study quality
Studies were assessed for consecutive enrolment, blinding, the use of a single (versus a composite) reference standard, and the use of the reference standard for all patients. Two reviewers independently assessed validity and resolved any
disagreements through discussion with a third reviewer.

**Data extraction**
The authors did not state how the data were extracted for the review, or how many reviewers performed the data extraction.

Studies by the same author were examined for overlapping samples. The extracted data included the following potential confounders: imaging end point of the study, clinical context (suspicion of stone, malignancy, or either, or a wide variety of pancreaticobiliary disease), MRCP technology era (before or after December 1966), frequency of direct visualisation of common bile duct by the reference standard (greater or less than 90%) and prevalence of disease. Data were extracted from each study to construct 2x2 tables for each specified outcome. For studies reporting results from more than one radiologist, the overall results reported for the study were extracted. The sensitivity, specificity, positive and negative predictive values were calculated.

**Methods of synthesis**

How were the studies combined?
The authors assumed that the studies would be heterogeneous and combined the studies in a meta-analysis, taking account of specified potential confounding factors. A logistic regression was used to explore the influence on diagnostic accuracy of the year of publication, quality score, proportion of patients having a direct reference standard, and clinical context. The regression model was weighted by study size and included the potential confounding factors using the following reference categories: quality score 1 or 2; imaging end point of the presence of obstruction; clinical context of suspicion of common bile duct; MRCP technology era after December 1997; and frequency of direct visualisation of common bile duct by the reference standard of 90% or less. A random-effects model was used to pool the studies and calculate the overall sensitivity and specificity adjusted for potential confounding variables. Summary receiver operating characteristic (ROC) curves were created, taking account of specific study characteristics (i.e. potential confounders plus sample size less or greater than 50). The area under the ROC curve and the 95% confidence interval (CI) were calculated. Positive and negative likelihood ratios (LRs) were used to calculate the post-test odds using a wide variety of pre-test probabilities of disease. Full details of the calculations were reported in the paper. Funnel plots were used to assess publication bias.

How were differences between studies investigated?
Heterogeneity was not tested, but differences between the studies were adjusted for in the analysis (see How Were the Studies Combined). Odds ratios and 95% CIs were calculated from the logistic regression model for each confounding variable, including each validity criterion. The meta-analysis was repeated using separate models for each imaging end point.

**Results of the review**

Sixty-seven diagnostic accuracy studies (n=4,711) were included.

Forty studies (61%) were blinded, 33 (59%) used consecutive recruitment, 61 (92%) used some type of reference standard in all patients, while 20 (30%) used the same reference standard test in all patients. Study quality was also reported separately for studies used in the assessment of obstruction, bile duct stones and malignancy.

All imaging end points: MRCP had a high adjusted sensitivity of 95% (95% CI: 75, 99) and specificity of 94% (95% CI: 86, 99).

MRCP had a high sensitivity (97%, 95% CI: 91, 99) and specificity (98%, 95% CI: 91, 99) for the presence of obstruction (30 studies, 1,954 patients); the LR of a positive test result was 49 (95% CI: 25, 62). Similarly high values were also found for the level of obstruction: sensitivity 98% (95% CI: 94, 99), specificity 98% (95% CI: 94, 100), and LR of a positive test result 49 (95% CI: 25, 135). The sensitivity and specificity for the diagnosis of stones were 92% (95% CI: 80, 97) and 97% (95% CI: 90, 99) respectively, (46 studies, 3,592 patients), and the LR of a positive test result was 29 (95% CI: 23, 49). Whilst the sensitivity and specificity for differentiating benign from malignant biliary obstruction were 88% (95% CI: 70, 96) and 95% (95% CI: 82, 99), respectively (22 studies, 12,943 patients), the LR of
A positive test result was 16 (95% CI: 10, 30).

Area under the ROC curve: the area was 0.99 (95% CI: 0.93, 1.0) for the presence of obstruction, 0.99 (95% CI: 0.90, 1.0) for the level of obstruction, 0.98 (95% CI: 0.97, 0.99) for bile duct stones, and 0.97 (95% CI: 0.85, 1.0) for malignancy.

The diagnostic accuracy was higher for larger studies, for studies that did not use consecutive enrolment, and for studies that did not use the 'gold' standard for all patients. The results were reported in the paper.

The funnel plots did not show any evidence for publication bias.

**Authors' conclusions**
MRCP has high sensitivity and specificity for the level and presence of biliary stones, but is less accurate for detecting stones or differentiating malignant and benign causes of obstruction.

**CRD commentary**
The review question was clear in terms of the study design, intervention and outcomes. Only one database was searched and this might have resulted in the omission of other relevant studies. By including publications in two languages and making some attempts to identify unpublished studies, the authors attempted to reduce language and publication bias. Full details of the methods used to select the studies were not reported. Methods were used to minimise bias in the data extraction and validity assessment processes. Validity was assessed using specified established criteria.

The authors stated that details of the individual studies were available on the Annals of Internal Medicine website. See Web Address at end of abstract. Statistical heterogeneity was not assessed, but the meta-analyses were adjusted for potential confounders. In addition, the influence of potential confounders of the results was explored. The evidence presented appears to support the authors’ conclusions.

**Implications of the review for practice and research**
Practice: The authors did not state any implications for practice.

Research: The authors stated that further research is required to assess the role and cost-effectiveness of MRCP in selected samples of patients.

**Bibliographic details**

**PubMedID**
14530225

**Original Paper URL**
http://www.annals.org/cgi/content/full/139/7/547

**Other publications of related interest**
This additional published commentary may also be of interest. Nietsch HH, Kowdley KV. Review: Magnetic resonance cholangiopancreatography is accurate in diagnosing biliary disease. ACP J Club 2004;141:25.

**Indexing Status**
Subject indexing assigned by NLM
MeSH
Bile Duct Neoplasms /diagnosis; Cholangiography /methods; Cholelithiasis /diagnosis; Cholestasis /diagnosis /etiology; Confounding Factors (Epidemiology); Diagnosis, Differential; Humans; Magnetic Resonance Imaging; Pancreatic Ducts /pathology; ROC Curve; Sensitivity and Specificity

AccessionNumber
12003008662

Date bibliographic record published
31/05/2005

Date abstract record published
31/05/2005

Record Status
This is a critical abstract of a systematic review that meets the criteria for inclusion on DARE. Each critical abstract contains a brief summary of the review methods, results and conclusions followed by a detailed critical assessment on the reliability of the review and the conclusions drawn.