Overview

Diseases of the pancreas and biliary tree are common in the United States. An estimated 6 per 100,000 people are afflicted with common bile duct stones, representing only a small fraction of those with gallstones. There are approximately 57,400 newly diagnosed cases of malignancy of the pancreas, gallbladder, or extrahepatic biliary tract each year, and the prognosis is usually poor. Pancreatitis can occur in an acute, acute recurrent, or chronic pattern, with common etiologic factors including alcohol consumption and choledocholithiasis.

This report is the product of a systematic literature review of the evidence on the diagnostic and therapeutic effectiveness of endoscopic retrograde pancreatography (ERCP) focusing on four clinical conditions: common bile duct stones, pancreaticobiliary malignancy, pancreatitis, and abdominal pain of possible pancreaticobiliary origin. In addition, the evidence describing patient, procedure, or operator determinants of complications of ERCP is systematically reviewed. The evidence on the prediction of common bile duct stones is reviewed as well.

Reporting the Evidence

The clinical topic areas addressed in this evidence report were developed by the planning committee for the National Institutes of Health State-of-the-Science Conference (January 2002) on Endoscopic Retrograde Cholangiopancreatography. For each major topic, there are several key questions that address the most pertinent diagnostic and therapeutic issues.

Topic 1. Patients with known or suspected common bile duct stones

a. What is the diagnostic performance of ERCP in detecting common bile duct stones in comparison to alternatives? Alternatives include endoscopic ultrasound (EUS), magnetic resonance cholangiopancreatography (MRCP), or computed tomography cholangiography (CTC).

b. What are the outcomes of treatment using ERCP strategies compared to using surgical or medical management?

c. What is the diagnostic value of specific risk factors or predictive models for assessing the likelihood of having a common bile duct stone?

Topic 2. Patients with known or suspected pancreaticobiliary malignancy

a. What is the comparative diagnostic performance of ERCP tissue sampling techniques in establishing a tissue biopsy diagnosis of pancreaticobiliary malignancy, and how do these techniques compare to alternative nonsurgical tissue sampling techniques (e.g., endoscopic ultrasound-guided fine-needle aspiration [FNA] or percutaneous FNA)?

b. What is the diagnostic performance of ERCP in diagnosing the presence of malignant pancreaticobiliary obstruction in comparison to other imaging alternatives (e.g., EUS or MRCP)?
c. What are the outcomes of treatment using ERCP strategies to treat malignant pancreaticobiliary obstruction compared to using surgical or interventional radiology treatment?

**Topic 3. Patients with pancreatitis**

a. What is the diagnostic performance of ERCP in detecting underlying causes or complications of pancreatitis that are amenable to treatment in comparison to alternatives (e.g., EUS or MRCP)?

b. What are the outcomes of treatment using ERCP strategies compared to using surgical or medical therapy?

**Topic 4. Patients with abdominal pain of possible pancreaticobiliary origin**

a. What is the diagnostic performance of ERCP with sphincter of Oddi manometry in identifying a pancreaticobiliary origin of pain in comparison to alternatives (e.g., biliary scintigraphy, EUS, or MRCP)?

b. What are the outcomes of treatment using ERCP strategies compared to using surgical or medical therapy?

**Topic 5. What patient, procedure, or operator factors are determinants of complications of ERCP?**

**Methodology**

The protocol for this review was designed prospectively to define study objectives, search strategy, patient populations of interest, study selection criteria, outcomes of interest, data elements to be abstracted and methods for abstraction, and methods for study quality assessment.

One reviewer performed primary data abstraction of all data elements into the evidence tables, and a second reviewer checked accuracy of the evidence tables. Disagreements were resolved between the two reviewers, or if necessary, in consultation with the Evidence-based Practice Center Director or members of the Technical Advisory Group.

**Search Strategy for the Identification of Articles**

The National Library of Medicine (NLM) staff conducted a comprehensive literature search for journal articles on ERCP from the PubMed®/MEDLINE®, BIOSIS, EM BASE, and SciSearch® databases with a publication date from 1980 through August 13, 2001. Articles which had been indexed to the NLM Medical Subject Heading (MeSH®) “cholangiopancreatography, endoscopic retrograde” as well as those containing the following list of ERCP synonyms and textword combinations were retrieved:

- Endoscopic retrograde pancreatocholangiography
- Endoscopic retrograde pancreato-choangiography
- ERCP
- ERCPs
- Endoscopic retrograde cholangiography
- ERC and endoscopy
- ERC and cholangiography
- Endoscopic cholangiography
- Endoscopic retrograde pancreatography
- ERP and endoscopy
- ERP and pancreatography
- Endoscopic pancreatography
- Endoscopic cholangiopancreatography
- Endoscopic cholangio-pancreatography
- ECP and endoscopy
- ECP and cholangiography
- Endoscopic pancreatocholangiography
- Endoscopic pancreato-choangiography
- EPC and endoscopy
- EPC and pancreatography

The “?” is a truncation symbol used to permit retrieval for variant word endings, as cholangiopancreatography, cholangiopancreatographic, etc.

Excluded from the search results were articles that:

- Were written in a foreign language.
- Did not have abstracts as a part of the online record in any of the databases searched.
- Did not include human subjects.
- Contained reports of only a single case.

The literature search for Topic 1c on prediction of common bile duct stones and for additional studies selected by the secondary selection criteria for Topics 3 and 4 used a streamlined search process to identify key articles addressing the clinical issue of interest. Reference lists from these articles were reviewed, focused MEDLINE searches were performed, and related articles were identified.

The Technical Advisory Group and peer reviewers for this project were asked to inform the project team of any studies relevant to the key questions addressed in this evidence report that were not retrieved by either of the search strategies.

**Search Results**

The online searches of the PubMed, EM BASE, BIOSIS, and SciSearch databases in conjunction with additional citations identified through manual searching yielded a total of 5,698 titles and abstracts for review. Based on review of abstracts, 789 articles were selected for review in full text.
Approximately 117 of these articles were excluded as review articles. Primary and secondary selection criteria were applied to articles identified as potential clinical trial reports. This process yielded a total of 149 included studies for the review of evidence.

**Study Selection Criteria**

**Primary Selection Criteria**

The selection criteria for all topics in this report were:

1. Full-length report in peer-reviewed medical journals.
2. Published in English.
3. Reported outcomes relevant to this systematic review.
4. Where there were multiple reports of a single study, only the report judged to be most recent and complete, based on number of included patients and length of followup, was included. If additional relevant outcomes were included in the duplicate reports, these data were abstracted and added to the data from the primary report with citation to the supplementary articles.
5. Prospective in design, or if retrospective, enrolled consecutive patients or used appropriate sampling methods (e.g., case-control sampling method).

In order to keep readers informed of ongoing studies, studies published only in abstract form since 1999 and judged to be important are noted in this systematic review; but data were not abstracted into the evidence tables.

Studies of diagnostic performance met the following additional selection criteria:

1. Compared ERCP and at least one of the relevant diagnostic alternatives or compared two ERCP alternatives.
2. Subjected at least 90 percent of participants to both ERCP and the relevant diagnostic alternative.
3. Addressed a relevant patient population.
4. Included at least 25 subjects.
5. Reported sufficient information to be able to calculate 2x2 contingency tables of diagnostic performance.

Studies of therapeutic outcomes met the following additional selection criteria:

1. Compared ERCP strategies with at least one of the relevant therapeutic alternatives.
2. Addressed a relevant patient population.
3. Included at least 25 subjects in each treatment group being analyzed separately.
4. Reported on at least one relevant outcome measure.
5. Were a contemporaneous comparison studies. If not contemporaneous, the populations and treatment settings were comparable.

Studies of predictors of ERCP complications met the following additional selection criteria:

1. Included a multivariable analysis of the relationship between patient, procedure, or operator factors and ERCP complications.
2. Enrolled at least 100 patients if a cohort study, or at least 25 cases if a case-control study.
3. Addressed potential confounding variables in either the selection of subjects or analysis.

Studies on the prediction of common bile duct stones met the following additional selection criteria:

1. Reported the association of either (a) specific risk factors of interest and the presence of a common bile duct stone (specific risk factors of interest were jaundice, liver function test results, and ultrasound finding of a dilated common bile duct), or (b) a prediction rule or model predicting likelihood of having a common bile duct stone and the presence of a common bile duct stone.
2. Enrolled at least 100 patients.
3. Reported sufficient information to be able to calculate 2x2 contingency tables of diagnostic performance in the prediction of presence or absence of a common bile duct stone.

**Secondary Selection Criteria**

There was a paucity of literature that met the primary selection criteria for questions on ERCP treatment of chronic pancreatitis (Topic 3b) and ERCP treatment of chronic abdominal pain of possible pancreaticobiliary origin (Topic 4b). In order to examine these questions, the original study selection criteria were relaxed for these topics to include:

1. Randomized controlled trials or otherwise concurrently controlled studies of an ERCP intervention compared to a relevant therapeutic alternative, regardless of sample size for pancreatitis.
2. Single arm pre-post-intervention studies which selected a well-defined population with a predictable natural history ascertained by baseline evaluation over 3 months. These studies must also have used an appropriate well-designed outcome measure over at least 6 months of followup.
Outcomes of Interest

For diagnostic performance studies, the outcomes of interest were test performance characteristics (i.e., sensitivity, specificity) in diagnosing clinically relevant findings.

For therapeutic outcome studies, the primary outcomes of interest include:

1. Measures of technical success (e.g., removal of stone, relief of obstruction, cyst drainage, need for repeat procedure or placement of stent).
2. Measures of clinical success (e.g., survival, quality of life, performance scores, relief of jaundice, relief of infection, symptom scores, or pain scores).
3. Resource utilization (e.g., hospitalization, perioperative care, return to work, intensity of post-procedure care).
4. Procedure-related morbidity (e.g., stent-related problems, cholangitis, sepsis, sedation-related outcomes, bleeding, perforation, pancreatitis, long-term effects of sphincterotomy, mortality).

For studies of factors predicting ERCP complications, the primary outcomes of interest were measures of relative risk or predictive value associated with patient, procedure, or operator factors.

Study Quality Assessment

The approach to assessing the quality of evidence used domains commonly recognized as important in the literature on study quality. Quality criteria were developed for each of the three types of studies included in this systematic review: studies of therapeutic effectiveness; studies of diagnostic performance; and multivariable regression analysis. For many topics addressed in this evidence review, studies meeting the most rigorous standards of quality do not exist. Thus, the main purpose of quality assessment in this systematic review is to discriminate between the better and lesser quality studies in the available evidence base.

For studies of therapeutic efficacy, the approach to quality assessment was adapted from that of the U.S. Public Health Preventive Services Task Force. Study quality domains of interest were: initial assembly of comparable groups (includes adequacy of randomization and controls for confounders); maintenance of comparable groups (includes attrition, crossovers, adherence, contamination); comparable performance of interventions; comparable measurements (unbiased, reliable, and valid); and appropriate analysis of outcomes (includes intent-to-treat analysis). A study was rated as “Good” if it clearly met all quality parameters. A study was rated “Fair” if it reasonably met these parameters and had no fatal flaw. A study was rated “Poor” if it was fatally flawed on one or more parameters (e.g., if comparable groups were not assembled or maintained or outcome measures were invalid or not applied equally among groups).

For studies of diagnostic performance, criteria for assessing study quality were developed using key references in the field of study quality assessment. The selection criteria used for this systematic review eliminated poor quality studies from inclusion. Study quality domains of interest to discriminate between good and fair quality studies were: enrollment of representative subjects (includes appropriate spectrum of patients, unbiased enrollment, complete enrollment of eligible patients, accounting for all eligible subjects); ERCP interpreted independently of diagnostic alternative; and diagnostic alternative interpreted independently from ERCP. As relevant, issues of suitability and interpretation of reference standards are addressed qualitatively in the discussion of each question.

For multivariable logistic regression analysis studies, the quality domains of interest were the degree of over-fitting present in the multivariable models, the nature of statistical reporting, and the use of procedures to establish internal validity. Degree of over-fitting was assessed using the ratio of the number of endpoints divided by the number of candidate variables in the model and was classified as satisfactory (ratio ≥10) to severe (ratio <4).

Findings

Topic 1. Patients with known or suspected common bile duct stones

Diagnostic performance of ERCP compared to alternatives:

• The search and selection process yielded 10 studies on MRCP (total n=834), 9 studies on EUS (total n=601), and 6 studies with 7 sets of findings on CTC (total n=266), but reference standards were not consistent among studies.
• Individual studies were relatively small and unlikely to have adequate power to detect a statistically significant difference; and no studies reported tests of statistical significance. Thus, it is not possible to determine with confidence whether the diagnostic performance is similar or poorer than ERCP or to accurately quantify any difference.
• The evidence comparing EUS to ERCP employs a reference standard that permits inferences regarding comparative performance. The evidence suggests that EUS is similar to ERCP in detecting common bile duct stones.
• MRCP has a degree of concordance with ERCP that results in sensitivities and specificities greater than 90 percent in most studies. Concordance of CTC with
ERCP appears to be lower, with sensitivities as low as 80 percent in some studies.

- The role of alternative tests in the management of patients with suspected common bile duct stones cannot be determined strictly by diagnostic performance. The costs and risks of the tests, and the costs and risks of actions based on test results, along with the pretest probability of stones must all be considered to determine the optimal management strategy.

ERCP treatment strategies compared to surgical or medical management:

- In order to evaluate ERCP treatment strategies, studies must account for patients through the diagnostic and treatment process, including additional procedures needed when initial treatment fails, and total morbidity of the alternative strategies. Overall, the literature is very thin and spread out over many different comparisons of interest, preventing strong conclusions about any specific comparison of treatment strategies.

- The limited evidence available suggests that: laparoscopic common bile duct exploration may be better than ERCP strategies to manage cholecystectomy patients with the least resource use; definitive surgery with cholecystectomy prevents long term complications at acceptable short-term morbidity when compared to sphincterotomy alone in high-risk surgical patients with suspected common bile duct stones; and endoscopic treatment of acute cholangitis reduces short-term mortality when compared to emergency surgery.

- Limited evidence suggests that the following techniques have similar stone removal rates and short-term complications: intracorporeal and extracorporeal lithotripsy methods for removing large common bile duct stones; balloon dilation and sphincterotomy; and needle-knife fistulotomy and needle-knife precut papillotomy.

Diagnostic value of specific risk factors or predictive models for assessing the likelihood of having a common bile duct stone:

- The probability of a common duct stone is one important factor in determining diagnostic and treatment strategies. When preoperative probability is high, ERCP may be preferred. When probability is low, expectant management is preferred. Additional diagnostic tests may be used to discriminate among patients in the middle range of probability. The exact probability cutoffs depend on the risks and benefits of the diagnostic and treatment alternatives. The risk factor or prediction model with the best receiver-operating characteristics (ROC) would make the best decision rule if the cutoff threshold were set correctly.

- Thirteen studies (total n=7,409) reported multiple findings of sensitivities and specificities of a single or combination of risk factors to predict the presence of common bile duct stones. The single risk factors most commonly assessed were: clinical jaundice or elevated bilirubin, liver function tests, and ultrasound findings of a dilated common bile duct. All have significant associations with the presence of common duct stones, but none have both high sensitivity and specificity. Of the four studies testing prediction rules based on combinations of risk factors, only one study was a validation of an independently developed prediction rule. Multivariable prediction rules appear to have superior ROCs compared to individual risk factors.

- The absence of any risk factors for stones (or a discriminant function indicating absence of stones) is a very strong predictor of the absence of stones. Absence of any risk factor produces probabilities of stones that are in the same range as a negative ERCP exam in a patient with risk factors for stones (0 percent to 17 percent).

**Topic 2. Patients with known or suspected pancreaticobiliary malignancy**

Diagnostic performance of ERCP tissue sampling techniques in establishing a tissue biopsy diagnosis of pancreaticobiliary malignancy in comparison to each other and compared to alternative nonsurgical tissue sampling techniques:

- Twelve studies comparing at least two tissue sampling techniques were identified in this systematic review. The available studies are limited by small size and do not consistently compare techniques in the same group of patients. Most studies do not report statistical tests, so it is not possible to determine with confidence whether reported differences in sensitivity are significantly different. While available evidence is suggestive, larger studies are needed to draw conclusions on relative performance of tissue sampling techniques.

- The available evidence suggests that sensitivity for detecting malignancy is similar or higher for brush cytology vs. bile aspiration cytology, similar for fine-needle aspiration (FNA) cytology vs. brush cytology, and similar or higher for forceps biopsy vs. brush cytology. Using combinations of two or more sampling techniques may increase overall sensitivity. No comparative studies evaluated whether incremental
improvement could also be achieved by repeated sampling using the same technique.

- In the absence of comparative studies of endoscopic ultrasound (EUS)-FNA and ERCP-FNA, indirect comparison of single-arm studies was attempted. Results from 10 studies including at least 400 subjects with pancreatic mass suggest a range of sensitivity in detecting pancreatic malignancy of 60-94 percent with a specificity of 100 percent. Two studies of ERCP-FNA including 164 subjects with various pancreaticobiliary tumors reported sensitivities ranging from 25 percent to 62 percent. While sensitivity reported in these studies appears to be lower than that for EUS-FNA, such a comparison is not valid due to differences in study populations, cytology techniques, and study settings.

Diagnostic performance of ERCP compared to alternatives in detecting malignant pancreaticobiliary obstruction:

- The available evidence directly comparing ERCP with either MRCP or EUS is modest in size and of varying methodologic quality. The evidence comparing ERCP with MRCP is some what stronger than that comparing ERCP with EUS.

- Individual studies do not demonstrate statistically significant differences in diagnostic performance for ERCP vs. MRCP or for ERCP vs. EUS for characterizing malignant strictures. In sum, the available studies suggest that both MRCP and EUS provide similar diagnostic performance as ERCP in detecting pancreaticobiliary malignant obstruction.

Treatment outcomes using ERCP strategies to treat malignant pancreaticobiliary obstruction compared to using surgical or interventional radiology treatment:

- Five studies compared endoscopic stent drainage with surgical bypass for palliation of malignant obstructive jaundice, and a randomized controlled trial of 204 patients provided the most robust evidence. There were no significant differences in overall survival, relief of jaundice, technical success, total hospitalization days, or perioperative mortality. Major complications were more frequent in the surgery group (11 percent vs. 29 percent, p=0.02); and stent replacement was required in 37 percent of patients treated with ERCP stents.

- Two randomized controlled trials (total n=206) and one nonrandomized trial (n=165) compared metal to plastic stents placed by ERCP for palliation of biliary obstruction due to malignancy. Both types of stents offer initial relief of jaundice and the available evidence does not conclusively show any difference in perioperative adverse events. Overall patient survival is not significantly different when stent occlusions are treated with stent exchange as needed. Total resource utilization including need for repeat ERCP, total hospital days, and costs was reported to be lower with metal stents compared with plastic stents.

- Six studies (total n=782), addressed preoperative stenting compared to no stenting prior to surgery for malignant pancreaticobiliary obstruction. The available evidence is of poor methodologic quality and fails to demonstrate that preoperative stenting improves health outcomes. Few studies report overall complications including both those related to the preoperative stent and the surgery, and these suggest that when complications of preoperative endoscopic stenting are considered along with the perioperative complications of surgery, preoperative stenting is associated with more complications. Preoperative stenting does appear to significantly improve elevated bilirubin and liver function tests, but the available evidence does not suggest that surgical outcomes are improved as a result.

**Topic 3. Patients with pancreatitis**

Diagnostic performance of ERCP compared to alternatives to detect underlying causes or complications of pancreatitis that are amenable to treatment:

- Three studies (total n=190) were found which met selection criteria. Each study addresses a different potential cause or complication of pancreatitis amenable to treatment. The available evidence is insufficient to compare ERCP and other diagnostic modalities for the identification of treatable causes or complications of pancreatitis.

Treatment outcomes of ERCP strategies compared to surgical or medical therapy:

- For treatment of acute pancreatitis, three randomized controlled trials (total n=554) compared early ERCP to delayed or selective ERCP. The available evidence suggests that early ERCP reduces complications in patient populations with acute pancreatitis and signs and symptoms suggesting biliary obstruction. In patients with low likelihood of biliary obstruction, delayed or selective ERCP permits many patients to avoid the procedure, and may result in lower complication rates. In addition, one retrospective associational study of a Veterans Administration database of patients with acute pancreatitis (n=2,075) suggests that outcomes of ERCP treatment are similar to those of surgery.

- For ERCP treatment in patients with acute recurrent or chronic pancreatitis, study selection criteria were relaxed as described above. Although the available evidence is
sparse and largely uncontrolled, it suggests that ERCP treatment reduces emergency room visits and hospitalization in patients with pancreas divisum and acute recurrent pancreatitis. Evidence on ERCP drainage of pseudocysts is also sparse and poorly controlled, but suggests that pain relief with ERCP is similar to results of surgery.

**Topic 4. Patients with abdominal pain of possible pancreaticobiliary origin**

Diagnostic performance of ERCP with sphincter of Oddi manometry compared with alternatives to identify a pancreaticobiliary origin of pain:

- The available evidence is not sufficient to permit conclusions on the diagnostic performance of biliary scintigraphy for sphincter of Oddi dysfunction. The body of evidence consists of three studies that included only 54 patients with sphincter of Oddi dysfunction; results of these studies cannot be synthesized due to differences in populations and methodology. There was substantial variability in the reported performance characteristics of biliary scintigraphy.

Treatment outcomes of ERCP strategies compared to surgical or medical therapy:

- Two randomized controlled trials (total n=128) show that endoscopic sphincterotomy relieves pain in patients with pancreaticobiliary pain, sphincter of Oddi dysfunction, and elevated basal sphincter of Oddi pressure on manometry (greater than 40mm Hg). The results of five single arm studies (total n=183) corroborate these data and suggest that patients with a dilated common bile duct and/or delayed contrast emptying may also benefit from endoscopic sphincterotomy.

- There is insufficient evidence to determine whether endoscopic sphincterotomy improves outcomes in patients with normal manometry findings. For this group, the small studies included in this review do not report significant improvements in pain with endoscopic sphincterotomy.

**Future Research**

Recommendations for future research include the following:

- Rigorous studies are required in order to reliably quantify the relative performance of diagnostic ERCP compared to alternatives. Existing studies do not consistently use common reference standards and frequently do not report tests of statistical significance. Thus, assumptions about equivalence or difference among alternative diagnostic technologies are not supported by robust empirical evidence.

- Comparative studies of alternative diagnostic and treatment strategies are urgently needed. It is imperative to use a comprehensive approach to outcomes assessment, taking into account the total burden of morbidity and resource utilization.

- Evidence on treatment of chronic pancreatitis and relapsing or recurrent pancreatitis is sparse. Rigorously designed controlled trials are needed to assess the outcomes of treatment for this debilitating condition.
Risk factors for complications of diagnostic and therapeutic ERCP have been explored using multivariable model analysis. Such analyses generate hypotheses for reducing complications, but cannot demonstrate cause and effect. Thus, interventions intended to reduce complications should incorporate prospectively defined studies to evaluate the results.

Availability of Full Report

The full evidence report from which this summary was derived was prepared for AHRQ by the Technology Evaluation Center, an Evidence-based Practice Center, under contract number 290-97-001-5. It is expected to be available in early 2002. At that time, printed copies may be obtained free of charge from the AHRQ Publications Clearinghouse by calling 800-358-9295. Requestors should ask for Evidence Report/Technology Assessment No. 50, Endoscopic Retrograde Cholangiopancreatography. Internet users will be able to access the report online through AHRQ’s Web site at: www.ahrq.gov.