

Chapter 3. Results and Conclusions, Part I: Common Bile Duct Stones

This chapter reviews evidence on the following questions:

In 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 (e.g., EUS, MRCP, or CTC)? (*Part I, Section 1: Diagnostic Performance of ERCP in Detecting Common Bile Duct Stones – Comparison to Alternatives*)
- b. What are the outcomes of treatment using ERCP strategies compared to using surgical or medical management? (*Part I, Section 2: Outcomes of Treatment Using ERCP for Common Bile Duct Stones – Comparison of Strategies Using ERCP, Surgery, or Medical Management*)
- c. What is the diagnostic value of individual risk factors or predictive models for assessing the likelihood of having a common bile duct stone? (*Part I, Section 3: Diagnostic Value of Individual Risk Factors or Predictive Models for Assessing the Likelihood of Having a Common Bile Duct Stone*)

Part I, Section 1: Diagnostic Performance of ERCP In Detecting Common Bile Duct Stones—Comparison With Alternatives

Introduction

The literature review identified three techniques that could be used as alternatives for diagnostic ERCP in the diagnosis of common bile duct stones: magnetic resonance cholangiography (MRCP), endoscopic ultrasound (EUS), and computed tomography cholangiography (CTC, with and without oral or intravenous biliary contrast). This section of the review only assesses diagnostic performance, and does not consider costs, availability, or adverse effects.

All included studies enrolled patients who underwent both the diagnostic test under consideration and ERCP. However, the choice of reference standard varied between studies and needs to be taken into account when interpreting the test characteristics calculated in each study, particularly if the goal is to determine which test is superior. Although ERCP had traditionally been considered the most accurate test for diagnosis of common bile duct stones, the test can produce both false-negative and false-positive results. The studies reviewed here generally used one of three different types of reference standards.

Ideally, ERCP and the alternative diagnostic test are both compared to a perfect reference standard such as actual examination of the common bile duct, producing unbiased estimates of test characteristics for both tests. Such a reference standard would not be ethical in most circumstances. Short of that, there may be selective confirmation of positive ERCP or other

tests, producing slightly biased estimates of test characteristics that are upwardly biased. However, the relative performance of ERCP to the alternative diagnostic test can be examined.

If ERCP is used as the reference standard, then the comparator test can only be worse. In such a case, the analysis can not determine which test is superior, but only the degree of concordance between the two tests.

Finally, a few studies (Neitlich, Topazian, Smith et al., 1997; Jimenez Cuenca, del Olmo Martinez, Perez Homs et al., 2001; Sugiyama, Atomi, and Hachiya, 1998) used ERCP images *and* sphincterotomy findings as the reference standard. This does not really allow an evaluation of the comparison between ERCP and the diagnostic test of interest, because the unreported diagnostic errors of ERCP images are “corrected” by the sphincterotomy findings. The performance of diagnostic ERCP cannot be evaluated in such studies unless the interpretation of the diagnostic ERCP is reported separately.

Given that the expected difference in diagnostic performance between ERCP and the diagnostic alternatives reported here are relatively small and the number of cases with the outcome of interest is generally small, these studies may have very limited power to detect statistically significant differences in test performance. None of the studies actually calculated any statistical significance values. Thus, it is not possible to determine with confidence whether the diagnostic performance of the alternative is similar or poorer than ERCP or to accurately quantitate any difference.

Evidence Base

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). In addition to these studies reporting diagnostic performance specific to common duct stones, 2 studies on MRCP which reported only on overall detection of obstructive abnormalities (total n=121) are also presented here. Study quality assessment is outlined in Table 1.

Review of Evidence: MRCP Performance

Ten studies studying a total of 834 patients were selected which examined the performance of MRCP compared to ERCP for the diagnosis of common bile duct stones (Table 2). Nine of the studies used ERCP as the reference standard, and thus measure the concordance of the two techniques rather than the relative performance. Only one study (Sugiyama, Atomi, and Hachiya, 1998) confirmed positive tests and allowed a comparison between the two tests. All the studies were rated as good quality with the exception of Guibaud, Bret, Reinhold, et al. (1995) and Sugiyama, Atomi, and Hachiya (1998).

Seven of the 9 studies which use ERCP as a reference standard show high concordance between the two tests with both sensitivity and specificity being greater than 90 percent. Two studies showed lesser degrees of concordance (Guibaud, Bret, Reinhold, et al., [1995], sensitivity 81 percent specificity 98 percent, and Stiris, Tennoe, Aadland et al. [2000], sensitivity 88 percent and specificity 94 percent).

Table 1. Quality Assessment

Study Author, Year	Patient Enrollment	Diagnostic performance of ERCP determined without knowledge of other test results	Diagnostic Performance of other test(s) determined without knowledge of ERCP results	Summary Evaluation
<i>MRCP</i>				
Demartines, Eisner, Schnabel et al., 2000	Prospective (n=70) Uncertain enrollment of consecutive patients	Yes	Yes	Good
Guibaud, Bret, Reinhold, et al., 1995	Prospective (n=126) Some exclusions because of no ERCP confirmation	Uncertain	Yes	Fair
Holzkecht, Gauger, Sackmann et al., 1998	Prospective (n=61) 61 of 66 eligible patients enrolled, all exclusions accounted for	Yes	Yes	Good
Lomas, Bearcroft, and Gimson 1999	Prospective (n=69) Consecutive patients enrolled, all exclusions accounted for	Yes	Yes	Good
Soto, Barish, Alvarez et al., 2000	Prospective (n=49) Consecutive patients enrolled, all exclusions accounted for	Yes	Yes	Good
Stiris, Tennoe, Aadland et al., 2000	Prospective (n=50) Consecutive patients enrolled	Yes	Yes	Good
Varghese, Farrell, Courtney et al., 1999	Prospective (n=100) Consecutive patients enrolled, all exclusions accounted for	Yes	Yes	Good
Sugiyama, Atomi, and Hachiya 1998	Prospective (n=97) Nonconsecutive enrollment, but stated to be arbitrary without known selection bias	Uncertain	Yes	Fair
Varghese, Liddell, Farrell et al., 2000	Prospective (n=191) 191 of out 256 consecutive patients enrolled, all exclusions accounted for	Yes	Yes	Good

Table 1. Quality Assessment (cont'd)

Study Author, Year	Patient Enrollment	Diagnostic performance of ERCP determined without knowledge of other test results	Diagnostic Performance of other test(s) determined without knowledge of ERCP results	Summary Evaluation
<i>MRCP (cont'd)</i>				
Burtin, Palazzo, Canard et al., 1997	Prospective (n=68) Consecutive patients enrolled	Yes	Yes	Fair—unorthodox reporting of data, uncertain of data
<i>Endoscopic Ultrasound</i>				
Canto, Chak, Stellato et al., 1998	Prospective (n=64) 64 out of 70 consecutive patients enrolled, 6 refusals	Yes	Yes	Good
Dancygier and Nattermann 1994	Prospective (n=41) Unstated whether consecutive	Uncertain	Yes	Fair
Norton and Alderson 1997	Prospective (n=46) Unstated whether consecutive	Yes	Yes	Fair
Prat, Amouyal, Amouyal et al., 1996	Prospective (n=119) Consecutive patients recruited, exclusions and refusals accounted for	Yes	Yes	Good
Sugiyama and Atomi 1997	Prospective (n=142) Consecutive patients enrolled	Uncertain	Yes	Fair
Sugiyama and Atomi 1998	Prospective (n=35) Consecutive patients enrolled	Uncertain	Uncertain	Fair
Chak, Hawes, Cooper et al., 1999	Prospective (n=36) Consecutive patients enrolled	Yes	Yes	Good

Table 1. Quality Assessment (cont'd)

Study Author, Year	Patient Enrollment	Diagnostic performance of ERCP determined without knowledge of other test results	Diagnostic Performance of other test(s) determined without knowledge of ERCP results	Summary Evaluation
<i>CTC</i>				
Ishikawa, Tagami, Toyota et al., 2000	Prospective (n=45) Unstated whether enrollment truly consecutive, not full accounting of exclusions	Uncertain	Uncertain	Fair
Polkowski, Palucki, Regula et al., 1999	Prospective (n=52) Full accounting of enrolled and excluded consecutive patients	Uncertain	Yes	Fair
Soto, Velez, and Guzman 1999	Prospective (n=29) Uncertain consecutive enrollment	Yes	Uncertain	Fair
Jimenez Cuenca, del Olmo Martinez, Perez Homs et al., 2001	Prospective (n=40) 40 of 60 consecutive patients enrolled, 20 excluded due to scheduling	Yes	Yes	Good
Neitlich, Topazian, Smith et al., 1997	Prospective (n=51) 51 of 96 consecutive patients enrolled, all exclusions accounted for	Yes	Yes	Good
Soto, Alvarez, Munera et al., 2000	Prospective (n=51) 51 of 56 eligible consecutive patients enrolled, all exclusions accounted for	Yes	Yes	Good

Table 2. Studies of MRCP, choledocholithiasis outcome, ERCP used as reference standard for all studies except Sugiyama, Atomi and Hachiya (1998)

Study	N	Population	Diagnostic test	Prev (%)	Sens (%)	Spec (%)	PPV (%)	NPV (%)	Comments
Demartines, Eisner, Schnabel et al., 2000	40	Patients with suspected CBD stones referred for ERCP	MRCP	48	100	90	90	100	
Guibaud, Bret, Reinhold, et al., 1995	126	Patients with suspected CBD obstruction referred for ERCP	MRCP	25	81	98	93	94	10 patients with other methods for gold standard
Holzkecht, Gauger, Sackmann et al., 1998	61	Patients referred for ERCP	MRCP (on-site reading) MRCP (off-site independent reading)	21	92 85	96 93	86 79	98 96	
Lomas, Bearcroft, and Gimson 1999	69	Patients with suspected CBD stones or stricture referred for ERCP	MRCP	13	100	97	100	97	
Soto, Alvarez, Munera et al. 2000	51	Patients with suspected CBD stones referred for ERCP	MRCP	51	96	100	100	96	1 false-negative ERCP considered positive after stone found at sphincterotomy
Soto, Barish, Alvarez et al., 2000	49	Patients with suspected CBD stones referred for ERCP	MRCP fast Spin Echo Reviewer 1 Reviewer 2 Single Section half-Fourier RARE Reviewer 1 Reviewer 2 Multisection half-Fourier RARE Reviewer 1 Reviewer 2	49	96 92 100 92 92 96	96 100 96 96 92 92	96 100 96 96 92 92	96 93 100 92 92 96	
Stiris, Tennoe, Aadland et al., 2000	50	Patients with suspected CBD stones referred for ERCP	MRCP	68	88	94	97	81	

Table 2. Studies of MRCP, choledocholithiasis outcome, ERCP used as reference standard for all studies except Sugiyama, Atomi and Hachiya (1998) (cont'd)

Study	N	Population	Diagnostic test	Prev (%)	Sens (%)	Spec (%)	PPV (%)	NPV (%)	Comments
Varghese, Farrell, Courtney et al. 1999	100	Patients with CBD obstruction referred for ERCP	MRCP	30	93	99	97	97	12 patients with gold standard of IOC or PTC included in analyses
Varghese, Liddell, Farrell et al., 2000	191	Patients with CBD obstruction referred for ERCP	MRCP	18	91	98	91	98	5 patients with gold standard of IOC or PTC included in analyses
<i>ERCP findings confirmed</i>									
Sugiyama, Atomi, and Hachiya 1998	97	Patients with suspected CBD stones referred for ERCP	MRCP ERCP (ERCP findings confirmed)	35	91 100	100 100	100 100	95 100	Positive ERCP confirmed by sphincterotomy, negative ERCP not confirmed

Sugiyama, Atomi, and Hachiya (1998) did the only study that confirms positive ERCP tests and allows a comparison between the two tests. In that study of 97 patients, ERCP had 100 percent sensitivity, and MRCP had 91 percent sensitivity. Specificity for both tests was 100 percent. This was the only study that analyzed sensitivity by subgroups of stone diameter. Sensitivity was 100 percent for stone diameters from 11–27 mm, 89 percent for stone diameter from 6–10 mm, and 71 percent for stone diameter between 3–5 mm.

Two studies reporting on a total number of patients of 121 had a mixed category of outcomes that included common duct stones (Table 3). In the study by Adamek, Albert, Weitz et al. (1998), the abnormalities included benign and malignant strictures, cholangiocarcinoma and choledochal cyst in addition to common duct stones. MRCP had a sensitivity and specificity for detecting any abnormality of 89 percent and 92 percent, whereas ERCP had a sensitivity of 91 percent and 92 percent.

In the study by Holzkecht, Gauger, Sackmann et al. (1998), the abnormalities detected included common bile duct dilatation and stenosis, in addition to common duct stones. Only the concordance with ERCP was evaluated. According to an image interpretation performed on-site, the sensitivity was 91 percent and the specificity was 80 percent. An off-site interpretation showed similar results.

In conclusion, most of the evidence on MRCP allows only conclusions as to whether MRCP and ERCP are concordant, rather than which test is superior. Most studies show fairly good concordance, with sensitivities and specificities both higher than 90 percent. Evidence limited to one study may indicate that ERCP is slightly better than MRCP.

Review of Evidence: Endoscopic Ultrasound Performance

There are 9 studies (total n=601) reporting on the capability of endoscopic ultrasound to diagnose common duct stones compared to ERCP (Table 4).. In all the studies except 1 (Sugiyama and Atomi, 1998), positive tests of either method were confirmed with sphincterotomy, allowing for inferences regarding comparative performance. The study by Prat, Amouyal, Amouyal et al. (1996) stands out in this regard by subjecting all patients to sphincterotomy and endoscopic exploration, and thus is the only study in this whole section examining common bile duct stones with a truly independent reference standard. Chak, Hawes, Cooper et al. (1999) and Canto, Chak, Stellato et al. (1998) were also rated as “good” quality studies.

Given the small differences in performance noted in most of the studies, none of the studies is likely to detect statistically significant differences in test performance. In three of the studies, the sensitivity of EUS was higher than ERCP (Prat, Amouyal, Amouyal et al., 1996, Norton and Alderson 1997; Burtin, Palazzo, Canard et al., 1997). In three studies, the sensitivity of ERCP was higher than EUS (Canto, Chak, Stellato et al., 1998; Dancygier and Nattermann 1994, Sugiyama and Atomi, 1997) and in the two other studies the sensitivities were within 1 percent (Polkowski, Palucki, Regula et al., 1999; Chak, Hawes, Cooper et al., 1999). The specificities were very close in all studies except Chak, Hawes, Cooper et al. (EUS 100 percent, ERCP 87 percent).

Table 3. Studies of MRCP, mixed outcome including CBD stones, stratified by reference standard

Study	N	Population	Diagnostic test	outcome	Prev (%)	Sens (%)	Spec (%)	PPV (%)	NPV (%)	Comments
<i>ERCP findings confirmed</i>										
Adamek, Albert, Weitz et al., 1998	60	Referrals for ERCP with suspected CBD obstruction	MRCP ERCP	Any abnormality	78	89 91	92 92	98 98	71 75	Uncertain method of ascertaining reference standard
<i>ERCP used as reference standard</i>										
Holzknicht, Gauger, Sackmann et al., 1998	61	Patients referred for ERCP	MRCP (on-site reading) MRCP (off-site reading)	Any abnormality	75	91 94	80 80	93 94	75 80	

Table 4. Studies comparing ERCP to endoscopic ultrasonography, ERCP findings confirmed except for one study (Sugiyama and Atomi, 1998)

Study	N	Population	Diagnostic test	Prevalence (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Comments
Prat, Amouyal, Amouyal et al., 1996	119	High suspicion of CBD stones, sphincterotomy candidates	EUS ERCP	66	94 90	98 100	99 100	89 84	Sphincterotomy and endoscopic exploration on all patients. Numbers differ from published report due to rounding errors in published report
Burtin, Palazzo, Canard et al., 1997	68	Patients with suspected CBD obstruction referred for ERCP	EUS ERCP	50	97 91	97 97	97 97	97 92	Unorthodox presentation of data in report, test characteristics calculated from text descriptions, technical failures counted as neg tests
Canto, Chak, Stellato et al., 1998	64	Patients with suspected CBD stones referred for ERCP	EUS ERCP	31	84 95	98 98	94 no report	93 no report	Actual numbers not reported, all values quoted from study. Positive ERCP confirmed with stone extraction, negatives with 12 mo clinical follow up
Norton and Alderson 1997	46	Patients with suspected CBD stones referred for ERCP	EUS ERCP	52	88 79	96 92	95 90	89 83	Positive ERCP and EUS confirmed by sphincterotomy, no confirmation of negative ERCP and EUS
Dancygier and Nattermann 1994	41	Patients with obstructive jaundice, referred for ERCP	EUS ERCP	39	94 100	100 100	100 100	96 100	Positive ERCP confirmed by sphincterotomy, no apparent confirmation of negative ERCP
Polkowski, Palucki, Regula et al., 1999	50	Patients referred for ERCP for suspected CBD stones	EUS ERCP	68	91 91	100 100	100 100	84 84	Positive ERCP confirmed by sphincterotomy, selective confirmation of negative ERCP
Sugiyama and Atomi 1997	142	Patients referred for ERCP for suspected CBD stones	EUS ERCP	36	96 100	100 100	100 100	98 100	Positive ERCP confirmed by sphincterotomy, no apparent confirmation of negative ERCP

Table 4. Studies comparing ERCP to endoscopic ultrasonography, ERCP findings confirmed except for one study (Sugiyama and Atomi, 1998) (cont'd)

Study	N	Population	Diagnostic test	Prevalence (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Comments
Chak, Hawes, Cooper et al., 1999	36	Patients with suspected acute biliary pancreatitis	EUS ERCP	33	91 92	100 87	100 79	95 94	Positives for either test confirmed with sphincterotomy, negatives not confirmed
<i>ERCP + sphincterotomy as ref standard</i>									
Sugiyama and Atomi 1998	35	Patients with suspected acute biliary pancreatitis	EUS	43	100	100	100	100	ERCP reference standard, but positive ERCP confirmed with stone removal

Although most of the studies are small, within the limits of the evidence available, it appears that EUS is similar to ERCP in the detection of common bile duct stones.

Review of Evidence: CTC Performance

Seven sets of findings report the diagnostic characteristics of CTC compared to ERCP for the diagnosis of common bile duct stones (Table 5). The studies varied considerably in the reference standard used. Three studies used ERCP as a reference standard, 2 studies used an independent reference standard, and 2 studies used ERCP and sphincterotomy findings as a reference standard. Three variations of CTC were used—no biliary contrast (3 studies, total n=142), intravenous biliary contrast (2 studies, total n=95) and oral contrast (2 studies, total n=80). This results in a body of literature in which, at most, 2 studies share the same CT technique and reference standard. The studies by Jimenez Cuenca, del Olmo Martinez, Perez Homs et al. (2001), Neitlich, Topazian, Smith et al. (1997), and Soto, Alvarez, Munera et al. (2000) were rated as “good” quality.

Three sets of findings from 2 studies, all from the same principal author (Soto, Velez, Guzman et al., 1999 and Soto, Alvarez, Munera et al., 2000), used ERCP images as the reference standard. Soto, Alvarez, Munera et al. (2000, n=51), which used no biliary contrast, showed poor concordance with ERCP (sensitivity 65 percent and 84 percent specificity). The other two sets of findings (Soto, Velez, Guzman et al., 1999, n=29 and Soto, Alvarez, Munera et al., 2000, n=51), found higher concordance with ERCP when using oral biliary contrast (sensitivities and specificities both greater than 90 percent).

Two studies (Ishikawa, Tagami, Toyota et al., 2000, n=45 and Polkowski, Palucki, Regula et al., 1999, n=50) examined CTC with IV biliary contrast, and both studies used methods where ERCP findings were confirmed. In both studies ERCP was more sensitive and specific than CTC (Ishikawa, Tagami, Toyota et al., 2000, ERCP 100 percent sensitivity, 100 percent specificity, CTC 71 percent sensitivity, 95 percent specificity; Polkowski, Palucki, Regula et al., 1999, ERCP 91 percent sensitivity, 100 percent specificity, CTC 85 percent sensitivity, 88 percent specificity).

Finally, the two studies that use ERCP sphincterotomy results as the reference standard (Jimenez Cuenca, del Olmo Martinez, Perez Homs et al., 2001, n=40 and Neitlich, Topazian, Smith et al., 1997, n=51) showed sensitivities of 80 percent and 88 percent, respectively, and specificities of 100 percent and 97 percent. A direct comparison to ERCP cannot be done with these data, but these sensitivities are lower than generally has been shown for ERCP.

In conclusion, most studies show a fair concordance with ERCP diagnosis of common bile duct stones, but in studies which allow a determination of which test is superior ERCP seems to have better sensitivity and specificity. However, no estimate of the magnitude of this superiority can be made from this evidence.

Table 5. Studies comparing CTC to ERCP, stratified by reference standard and presence and by type of contrast

Study	N	Population	Diagnostic test	Prev (%)	Sens (%)	Spec (%)	PPV (%)	NPV (%)	Comments
<i>ERCP used as reference standard (No biliary contrast)</i>									
Soto, Alvarez, Munera et al., 2000	51	Patients referred for ERCP for suspected CBD stones	CTC	51	65	84	81	70	
<i>ERCP used as reference standard (Oral biliary contrast)</i>									
Soto, Alvarez, Munera et al., 2000	51	Patients referred for ERCP for suspected CBD stones	CTC with oral biliary contrast	51	92	92	92	92	
Soto, Velez, Guzman et al. 1999	29	Patients referred for ERCP for suspected CBD stones	CTC with oral biliary contrast Observer 1 Observer 2	48	93 86	100 100	100 100	94 88	
<i>ERCP findings confirmed (independent reference standard)</i>									
<i>IV biliary contrast</i>									
Ishikawa, Tagami, Toyota et al., 2000	45	Laparoscopic patients undergoing routine preoperative ERCP	CTC with IV biliary contrast ERCP	16	71 100	95 100	71 100	95 100	Positive ERCP apparently confirmed during cholecystectomy, negative ERCP unlikely to be confirmed
Polkowski, Palucki, Regula et al., 1999	50	Patients referred for ERCP for suspected CBD stones	CTC with IV biliary contrast ERCP	68	85 91	88 100	94 100	74 84	Positive ERCP confirmed by sphincterotomy, selective confirmation of negative ERCP

Table 5. Studies comparing CTC to ERCP, stratified by reference standard and presence and by type of contrast

Study	N	Population	Diagnostic test	Prev (%)	Sens (%)	Spec (%)	PPV (%)	NPV (%)	Comments
<i>No biliary contrast, ERCP + sphincterotomy findings used as reference standard</i>									
Jimenez Cuenca, del Olmo Martinez, Perez Homs et al., 2001	40	Patients referred for ERCP for suspected CBD stones	CTC	50	80	100	100	83	ERCP reference standard based on image and/or sphincterotomy findings, not only images
Neitlich, Topazian, Smith et al., 1997	51	Patients referred for ERCP for suspected CBD stones	CTC	33	88	97	94	94	ERCP reference standard based on image and/or sphincterotomy findings, not only images

Conclusion

The evidence about the relative performance of EUS compared to ERCP is the strongest, because most of the studies used reference standards which allowed inferences regarding comparative performance. With some studies showing EUS is better, and other studies showing ERCP is better, and no remarkable outlying results, the weight of the evidence suggest that EUS is similar to ERCP in detecting common bile duct stones.

MRCP has a concordance with ERCP that results in sensitivities and specificities greater than 90 percent in most studies when using ERCP as a reference standard. Along with evidence limited to one study regarding comparative performance of MRCP and ERCP, MRCP may be slightly worse than ERCP in detecting common bile duct stones.

CTC also has reasonable concordance with ERCP, but the range of sensitivities and specificities is lower, with sensitivities dipping down to the 80 percent level in some studies. Again with evidence limited to only 2 small studies on the relative performance of CTC to ERCP, it appears that CTC is not as good as ERCP in detecting common bile duct stones.

Although some tests may not perform quite as well as ERCP, the role of these tests in the management of patients with suspected common bile duct stones cannot be determined strictly by an examination of their test characteristics. The costs and risks of the tests, and the costs and risks of actions based on their results, along with the pretest probability of stone needs to be taken into account to determine the optimal strategy that most efficiently treats patients with suspected common duct stones.

Part I, Section 2: Outcomes of Treatment Using ERCP for Common Bile Duct Stones—Comparison of Strategies Using ERCP, Surgery, or Medical Management

Introduction

ERCP can both provide diagnosis and treatment of common bile duct stones in one session in a less-invasive manner than an open surgical procedure. Commonly performed in conjunction with cholecystectomy, it could be performed before or after or, rarely, during surgery. However, there are risks from the procedure and it may not be successful at removing the common bile duct stones. Common bile duct exploration was the traditional surgical treatment to remove stones. This used to be performed with an open surgical incision. Then laparoscopic cholecystectomy became a common operation, and in order to avoid an open incision, ERCP was used in the diagnosis and removal of common duct stones. Recently, laparoscopic methods of exploring the common bile duct and removing stones have evolved, making for even more varied potential treatment options.

In order to appropriately evaluate ERCP treatment strategies, studies must properly account for the patients throughout the diagnostic and treatment process, including additional procedures needed for failed initial procedures. Alternatively, studies can assess outcomes through identical stages of the diagnostic or treatment process. Complication rates in and of themselves may not be fair measures of outcomes between treatment strategies if the baseline morbidity of procedures (e.g., open common bile duct exploration versus ERCP common duct stone extraction) are very different. Ideally, a measure of morbidity that could fairly assess both the quantity of procedures and total morbidity endured during each procedure would be a fair comparison between treatment strategies.

Evidence Base

For the purposes of this evidence review, the literature remaining after selection criteria were applied was very thin and spread out over many different research questions. Generally, there was only one or at most, two, studies on a specific comparison of interest. Study quality assessment is outlined in Table 6.

Review of Evidence: ERCP with Laparoscopic Cholecystectomy to Remove Common Bile Duct Stones

Three randomized controlled trials enrolling a total of 289 patients compared alternative strategies for removal of common bile duct stones in patients undergoing laparoscopic cholecystectomy (Tables 7–9). Although all 3 trials were judged to be of good quality, the evidence is limited because there is only a single study addressing each comparison of interest. Each trial reported on a different comparison, with respect to both the procedures compared and the patient population selected.

Table 6. Quality Assessment

Study Author, Year	Comparable Initial Groups?	Comparable Groups Maintained?	Comparable Performance of Intervention?	Comparable Measurement of Outcomes?	Appropriate Analysis	Summary Evaluation
Cuschieri, Lezoché, Morino et al., 1999	RCT (n=300) Good comparability — computerized randomization — comparable characteristics	31 patients not treated according to random allocation, reported separately	Adequate for comparison	Adequate outcome measures used.	Those treated to assigned treatment reported as principal findings. Patients not treated by assigned treatment reported separately.	good
Rhodes, Sussman, Cohen et al., 1998	RCT (n=80) Uncertain comparability — randomization technique unknown — limited data on comparability	All patients retained for analysis	Adequate for comparison	Outcomes were not assessed blindly Uncertain how morbidity rates determined	All retained patients analyzed	Good
Chang, Lo, Stabile et al., 2000	RCT (n=59) Good comparability — sealed envelope randomization — comparable characteristics	All patients retained for analysis	Adequate for comparison	Outcomes were not assessed blindly Definition of morbidity not provided	All retained patients analyzed	Good

Table 6. Quality Assessment (cont'd)

Study Author, Year	Comparable Initial Groups?	Comparable Groups Maintained?	Comparable Performance of Intervention?	Comparable Measurement of Outcomes?	Appropriate Analysis	Summary Evaluation
Targarona, Ayuso, Bordas et al., 1996	RCT (n=98) Good comparability — stratified randomization with sealed envelopes — patient characteristics comparable	2 out of 100 patients excluded because of incorrect randomization	Adequate for comparison	Outcomes were not assessed blindly Short-term morbidity rates do not capture difference in invasiveness between treatments	All patients retained for short-term outcomes analysis 89/93 surviving patients retained for long term outcomes analysis	Good
Trias, Targarona, Ros et al., 1997	Prospective study with historical control group (n=110) Good comparability Patient characteristics comparable	All patients prospectively identified as eligible enrolled	Surgical arm may include endoscopic sphincterotomy, more intensive treatment	Outcomes were not assessed blindly Short-term morbidity rates do not capture difference in invasiveness between treatments	All patients retained for short-term outcomes analysis 99/105 surviving patients retained for long term outcomes analysis	Fair
Hammarstrom, Holmin, Stridbeck et al., 1995	RCT (n=80) Good comparability — random numbers — patient characteristics comparable	All potential patients accounted for, few refusals	Adequate for comparison	Outcomes not systematically defined or enumerated	Adequate follow up	Poor, most results could not be tabulated

Table 6. Quality Assessment (cont'd)

Study Author, Year	Comparable Initial Groups?	Comparable Groups Maintained?	Comparable Performance of Intervention?	Comparable Measurement of Outcomes?	Appropriate Analysis	Summary Evaluation
Lai, Mok, Tan et al., 1992	RCT (n=82) Good comparability — randomized by consecutive envelopes — patient characteristics comparable	82 of 96 patients with severe acute cholangitis enrolled	Adequate for comparison	Outcomes were not assessed blindly Complication rates do not capture difference in invasiveness between treatments	All patients retained for analysis	Good
Leese, Neoptolemos, Baker et al., 1986	Retrospective observational study (n=82) Not very comparable Patients undergoing ERCP older, greater numbers of risk factors	Not applicable- retrospective study	Adequate for comparison	Outcomes were not assessed blindly	Analysis does not take into account difference in risk factors	Poor

Table 6. Quality Assessment (cont'd)

Study Author, Year	Comparable Initial Groups?	Comparable Groups Maintained?	Comparable Performance of Intervention?	Comparable Measurement of Outcomes?	Appropriate Analysis	Summary Evaluation
Adamek, Maier, Jakobs et al., 1996	Retrospective observational study (n=145) Fair comparability Patients comparable on all measured characteristics	Not applicable-retrospective study	Adequate for comparison	Outcomes were not assessed blindly	Simple unadjusted comparisons	Fair/poor
Neuhaus, Zillinger, Born et al., 1998	RCT (n=60) Good comparability — randomization technique unknown — patients comparable on all measured characteristics	All patients retained for analysis	Adequate for comparison	Outcomes were not assessed blindly	All patients retained for analysis	Good
Bergman, Rauws, Fockens et al., 1997	RCT (n=202) Good comparability — blinded computer-generated randomization — patients comparable on all measured characteristics	16 out of 218 excluded after randomization because of ineligibility	Adequate for comparison	Outcomes were not assessed blindly	All patients retained for analysis	Good

Table 6. Quality Assessment (cont'd)

Study Author, Year	Comparable Initial Groups?	Comparable Groups Maintained?	Comparable Performance of Intervention?	Comparable Measurement of Outcomes?	Appropriate Analysis	Summary Evaluation
Ochi, Mukawa, Kiyosawa et al., 1999	RCT (n=110) Good comparability — randomization not described — patients comparable on all measured characteristics	All patients retained for analysis	Adequate for comparison	Outcomes were not assessed blindly	All patients retained for short-term outcome analysis 105/110 patients retained for long-term outcome analysis	Good
Mavrogiannis, Liatsos, Romanos et al., 1999	RCT (n=153) Good comparability — randomization by sealed envelopes — Baseline characteristics similar for age, gender, presence of GB and gallstones	No cross-overs, drop outs reported.	Adequate for comparison.	Adequate outcome measures used. Outcomes were not assessed blindly.	Intention to treat analysis used.	Good
Chopra, Peters, O'Toole et al., 1996	RCT (n=86) Good comparability — Randomization by sealed envelopes — patients comparable on all measured characteristics	All patients retained for analysis	Adequate for comparison	Outcomes not blindly assessed Adequate for comparison	All patients analyzed for short term outcomes, 82/86 followed for long term outcomes	good

Table 7. Preoperative versus Postoperative ERCP in Cholecystectomy: Randomized Trials

Study	N	Population and Interventions	Outcomes	P	Adverse effects, complications	P	Resource utilization	P
Chang, Lo, Stabile et al., 2000	59	59 patients with mild to moderate gallstone pancreatitis, undergoing cholecystectomy after acute pancreatitis Mandatory preoperative ERCP (n=30) vs. selective postoperative ERCP (n=29) based on IOC findings	Stone Removal, successful ERCP/ERCP with stones: Preop ERCP: 12/12, 100% Postop ERCP: 7/7 , 100%		Morbidity rates (not defined) Preop ERCP: 10% Postop ERCP: 10%	n.s.	Hospital stay: mean, median days Preop ERCP: 11.7, 9.5 Post op ERCP: 9.0 , 8 ICU days: mean, median Preop ERCP: 1.7, 1 Post op ERCP: 1.9 ,1 Total Costs: Preop ERCP: \$10,210 Postop ERCP: \$8,586	.04 n.s. .049

Table 8. Preoperative ERCP versus Intraoperative cholangiogram and laparoscopic common bile duct exploration in patients undergoing laparoscopic cholecystectomy in patients with suspected common bile duct stones, randomized trials

Study	N	Population and Interventions	Outcomes	P	Adverse effects, complications	P	Resource utilization	P
Cuschieri, Lezoche, Morino et al., 1999	269	Patients with suspected CBD stones needing cholecystectomy Preoperative ERCP (n=136) versus IOC and laparoscopic CBD exploration (n=133) as initial strategies for removing stones	Stone clearance: Preop ERCP: 84% IOC, LCBDE: 84%	n.s.	Conversion to open cholecystectomy: Preop ERCP: 6% IOC, LCBDE: 13% Overall morbidity: Preop ERCP: 12.8% IOC, LCBDE: 15.8% Mortality: Preop ERCP: 1.5% IOC, LCBDE: 0.75%	.08 n.s. n.s.	Hospital stay, mean days: Preop ERCP: 9 IOC, LCBDE: 6	<.05

Table 9. Postoperative ERCP versus laparoscopic exploration of common bile duct in patients with common duct stones found on intraoperative cholangiography, randomized trials

Study	N	Population and Interventions	Outcomes	P	Adverse effects, complications	P	Resource utilization	P
Rhodes, Sussman, Cohen et al., 1998	80	80 patients with CBD stones found on cholangiography during cholecystectomy	Initial clearance of CBD stones: LCBDE: 75% Postop ERCP: 75%	n.s.	Overall Morbidity: LCBDE: 18% Postop ERCP: 15%	n.s.	Hospital stay, median days: LCBDE: 1 Postop ERCP: 3.5	<.01
		Laparoscopic CBD exploration (LCBDE) (n=40) versus postoperative ERCP (n=40)	Final clearance of CBD stones: LCBDE: 100% Postop ERCP: 93%	n.s.				

Overall, both arms in each of these 3 studies reported similar rates of stone clearance and morbidity, although morbidity was not well defined in two of these trials (Chang, Lo, Stabile et al., 2000; Rhodes, Sussman, Cohen et al., 1998). Thus, the main outcome of interest is relative resource utilization for each pair of alternative strategies for stone removal.

Mandatory Preoperative ERCP versus Selective Postoperative ERCP

Chang, Lo, Stabile et al. (2000) randomized 59 patients undergoing cholecystectomy during recovery from acute gallstone pancreatitis. Selective postoperative ERCP was based on findings from intraoperative cholangiogram. Resource utilization was lower in the selective postoperative ERCP group as measured by mean total hospital stay (9.0 vs. 11.7 days, $p=0.04$), and total costs (\$8,586 vs. \$10,210, $p=0.049$).

Preoperative ERCP versus intraoperative cholangiogram and laparoscopic common bile duct exploration (LCBDE)

Cuschieri, Lezoche, Morino et al. (1999) randomized 300 patients undergoing laparoscopic cholecystectomy who had suspected common bile duct stones. In one treatment arm, preoperative ERCP was performed, and sphincterotomy and stone removal was attempted if stones were detected. In the other treatment arm, LCBDE was performed if stones were detected on intraoperative cholangiogram. Mean hospital stay was reduced in the LCBDE treatment group (6 versus 9 days, $p<0.05$).

LCBDE versus Postoperative ERCP

Rhodes, Sussman, Cohen et al. (1998) randomized 80 patients with common bile duct stones found on intraoperative cholangiography during laparoscopic cholecystectomy. The hospital stay was reduced in the LCBDE group (median days, 1 vs. 3.5, $p<0.01$).

Summary

There is insufficient evidence determine whether there is an optimal strategy for common bile duct stone removal in patients undergoing cholecystectomy. The available evidence suggests that resource utilization is lower when:

- (1) selective postoperative ERCP is performed, as compared to routine ERCP prior to cholecystectomy; and
- (2) when laparoscopic common bile duct exploration is performed during laparoscopic cholecystectomy, as compared to adjunctive pre- or postoperative ERCP.

However, since success and complications of ERCP and laparoscopic cholecystectomy with LCBDE may be operator dependent, findings may not be generalizable across clinical settings. The availability of expertise in LCBDE may be limited at present.

Review of Evidence: ERCP Sphincterotomy alone versus Definitive Surgery for suspected common duct stones

Patients at High Surgical Risk

One randomized, controlled trial (Targarona, Ayuso, Bordas et al., 1996) and an observational study derived from the Targarona trial (Trias, Targarona, Ros et al., 1997) addressed whether removal of common duct stones with endoscopic sphincterotomy alone has lower morbidity and mortality than approaches which also remove the gall bladder during initial treatment (Table 10 and Table 11). The population of interest is patients at high surgical risk if subjected to cholecystectomy. For patients at high surgical risk, there may be advantages to a nonsurgical approach for removing common duct stones during acute symptomatic episodes. However, there may be differences in long term outcome if the gall bladder is not removed. Study quality was judged to be “Good” for the Targarona, Ayuso, Bordas et al. (1996) trial, and “Fair” for the Trias, Targarona, Ros et al. (1997) study.

The Targarona and Trias studies included high-risk surgical candidates based on age, cardiac risk, and pulmonary disease. The technique used in the Targarona, Ayuso, Bordas et al. (1996) study may not be representative of current surgical practice as the investigators performed open cholecystectomy for the definitive surgery arm; only the observational study by Trias, Targarona, Ros et al. (1997) used laparoscopic cholecystectomy.

Targarona, Ayuso, Bordas et al. (1996; n=98) found that both groups had similar short-term treatment failure, mortality, and morbidity, but initial postoperative length of stay favored endoscopic sphincterotomy alone (5 versus 11 days, $p<0.001$). However, over the longer term, the cholecystectomy patients had fewer biliary complications (6 percent versus 21 percent, $p=0.04$) and fewer readmissions (4 percent versus 23 percent, $p<0.01$). Eventually, 15 percent of patients in the sphincterotomy group underwent cholecystectomy.

Trias and colleagues performed laparoscopic cholecystectomy with preoperative ERCP as needed in 60 high-risk patients, and compared outcomes to the endoscopic sphincterotomy arm of the Targarona, Ayuso, Bordas et al. (1996) trial. Short-term and long-term results were similar to the Targarona trial, but initial hospital length of stay no longer favored the endoscopic sphincterotomy group when compared to laparoscopic, rather than open, cholecystectomy.

Patients Not at High Surgical Risk

One randomized controlled trial by Hammarstrom, Holmin, Stridbeck et al. (1995) enrolled 80 patients with intact gallbladders diagnosed with common bile duct stones on ERCP (Table 12). Patients either received sphincterotomy alone or open cholecystectomy and common bile duct exploration. Patients were followed for 5 years.

The study does not coherently define and compare outcomes between treatment groups for the most part; rather, various post-procedure events are unsystematically enumerated, making it difficult to tabulate any overall sense of outcomes. Total hospital stay (short term and follow up

stays) was compared between the groups and was not statistically significantly different (median stay, 13 days sphincterotomy, 16 days surgery, $p=ns$). Of patients who received sphincterotomy, 13 were subsequently treated with cholecystectomy, 4 urgently because of acute cholecystitis. The authors also noted that the death rate from non-biliary related causes was higher in the endoscopic sphincterotomy group (30 percent vs. 10 percent, $p=0.02$). The authors conclude that the two alternatives are equally effective in the long term, but that due to the difference in heart disease mortality surgery might be the better option.

Summary

The very limited available evidence shows that definitive treatment prevents long term recurrence of biliary symptoms, hospitalization, and need for further treatment. In high-risk patients as defined in these studies, definitive treatment can be performed with acceptable short term morbidity and equivalent mortality as sphincterotomy alone. Not all patients develop recurrent problems, so the choice of definitive treatment versus sphincterotomy alone involves the weighing of short term morbidity of treatment, be it sphincterotomy alone, open or laparoscopic surgery, against the probability of recurrent biliary symptoms.

Table 10. Endoscopic sphincterotomy alone versus open cholecystectomy in high risk surgical patients as primary treatment for common bile duct stones, randomized trials

Study	N	Population and Interventions	Outcomes	P	Adverse effects, complications	P	Resource utilization	P
Targarona, Ayuso, Bordas et al., 1996	98	Surgical high risk patients presenting with symptoms consistent with CBD stones Endoscopic sphincterotomy only (n=50) versus open cholecystectomy and CBD exploration if necessary (n=48)	Initial failure of treatment: ES: 12% Surgery: 6% Immediate mortality: ES: 6% Surgery: 4%	0.3 .5	Immediate morbidity: ES: 16% Surgery: 23% LONG TERM Biliary complications: ES (n=46): 21% Surgery(n=43): 6% Readmissions: ES: 23% Surgery: 4% Cholecystectomy: ES: 15% Surgery: 0% Need for sphincterotomy: ES: 2% Surgery: 4%	0.4 .04 .01 .01 .9	Post-treatment length of stay, mean days: ES: 5 Surgery: 11	.001

Table 11. Endoscopic sphincterotomy alone versus laparoscopic cholecystectomy (with or without preoperative ERCP) in high risk surgical patients as primary treatment for common bile duct stones, observational studies

Study	N	Population and Interventions	Outcomes	P	Adverse effects, complications	P	Resource utilization	P
Trias, Targarona, Ros et al., 1997	110	Surgical high risk patients presenting with symptoms consistent with CBD stones Endoscopic sphincterotomy only (n=50) versus laparoscopic cholecystectomy and with preoperative ERCP if necessary (n=60)	Initial failure of treatment: ES: 12% Surgery: 11% Immediate mortality: ES: 6% Surgery: 3%	n.s. 0.5	Immediate morbidity: ES: 16% Surgery: 18% LONG TERM Biliary complications: ES (n=46): 21% Surgery(n=53): 4% P Readmissions: ES: 23% Surgery: 2% P Need for reoperation: ES: 15% Surgery: 2%	n.s. <.04 <.01 <.01	Post-treatment length of stay, mean days: ES: 5 Surgery: 4.4	n.s.

Table 12. Endoscopic sphincterotomy alone versus open cholecystectomy and CBD exploration in non-high risk surgical patients as primary treatment for common bile duct stones, randomized trials

Study	N	Population and Interventions	Outcomes	P	Adverse effects, complications	P	Resource utilization	P
Hammarstrom, Holman, Stridbeck et al., 1995	80	<p>Patients presenting with CBD stones on ERCP with intact gall bladder</p> <p>Endoscopic sphincterotomy only (n=39) versus open cholecystectomy and CBD exploration if necessary (n=41)</p>	Biliary outcomes not coherently tabulated		<p>Biliary complications not coherently tabulated</p> <p>Deaths from non-biliary related disease ES: 30% Surgery: 10%</p> <p>13 patients in ES group required cholecystectomy on follow up</p>	0.02	<p>Total hospitalization days, median ES: 13 Surgery: 16</p>	NS

Review of Evidence: ERCP versus surgery for patients with acute cholangitis

Two studies compared of ERCP treatment to open surgery for patients with acute cholangitis due to common bile duct stones (Table 13 and Table 14). Lai, Mok, Tan et al. (1992) randomized 82 patients diagnosed with common bile duct stones by ERCP to endoscopic nasobiliary drainage or open common bile duct exploration. This study is from Hong Kong, where oriental cholangiohepatitis is a common cause of common duct stones, and may not generalize to populations with a different spectrum of disease. Leese, Neoptolemos, Baker et al. (1986) conducted a retrospective review of 43 patients treated with endoscopic sphincterotomy to 28 contemporaneous patients undergoing surgical decompression for relief of cholangitis.

The Leese, Neoptolemos, Baker et al. (1986) study was judged to be of poor quality due to imbalance of patient characteristics between groups.

Acute severe cholangitis is a condition of very high mortality, thus the important outcome is to reduce the acute mortality rate. Both studies show that short-term mortality from acute cholangitis is lower in the ERCP-treated group compared to open surgery. Lai, Mok, Tan et al. (1992) reported lower hospital mortality (10 percent versus 32 percent, $p<0.05$) in the group treated with endoscopic nasobiliary drainage. Despite prognostic factors favoring the open surgery group, Leese, Neoptolemos, Baker et al. (1986) found that mortality at 30 days was lower in the endoscopic sphincterotomy group (5 percent versus 21 percent, $p<0.02$).

Review of Evidence: Endoscopic lithotripsy vs. extracorporeal shock wave lithotripsy (ESWL) in stones not removable with standard endoscopic techniques

Two studies compared endoscopic lithotripsy techniques to extracorporeal shock wave lithotripsy (ESWL) in removing common bile duct stones that cannot be removed with standard endoscopic techniques (which includes mechanical lithotripsy) (Neuhaus, Zillinger, Born et al., 1998 and Adamek, Maier, Jakobs et al., 1996; Table 15 and Table 16). In these studies, successful removal of stones is the important outcome.

Neuhaus, Zillinger, Born et al. (1998) randomized 60 patients to ESWL or intracorporeal laser lithotripsy. Adamek, Maier, Jakobs et al. (1996) performed an observational comparison between ESWL ($n=79$) and intracorporeal electrohydraulic lithotripsy ($n=46$).

Neuhaus, Zillinger, Born et al. (1998), found that intracorporeal laser lithotripsy was more successful than ESWL in clearing the bile duct of stones (97 percent versus 73 percent, $p<0.05$). Adamek, Maier, Jakobs et al. (1996) found no significant difference between ESWL and electrohydraulic lithotripsy.

Table 13. Endoscopic drainage for treatment of acute cholangitis due to common bile duct stones, randomized trials

Study	N	Population and Interventions	Outcomes	P	Adverse effects, complications	P	Resource utilization	P
Lai, Mok, Tan et al., 1992	82	82 patients with acute severe cholangitis due to CBD stones diagnosed with diagnostic ERCP Nasobiliary drainage placed by ERCP (n=41) versus open CBD exploration (n=41)	Hospital mortality rate: ERCP: 10% Surgery: 32%	<.03	Overall complication rate: ERCP: 34% Surgery: 66%	>.05		

Table 14. Sphincterotomy for treatment of acute cholangitis due to common bile duct stones, observational studies

Study	N	Population and Interventions	Outcomes	P	Adverse effects, complications	P	Resource utilization	P
Leese, Neoptolemos, Baker et al., 1986	71	Retrospective review of patients with acute cholangitis due to CBD stones Early sphincterotomy (n=43) versus early surgery (n=28)	30 day mortality ERCP: 5% Surgery: 21%	<.02	Total % of patients with complications: ERCP: 28% Surgery: 57%	N/A	Hospital stay, median days: ERCP: 20 Surgery 23	n.s.

Patients receiving ERCP had greater baseline medical risk factors than patients having surgery (2 vs. 1, $P<.05$)

Table 15. Intracorporeal vs. extracorporeal lithotripsy for common bile duct stones, randomized trials

Study	N	Population and Interventions	Outcomes	P	Adverse effects, complications	P	Resource utilization	P
Neuhaus, Zillinger, Born et al. 1998	60	<p>Patients with stones not removable with ERCP techniques due to impacted stones or inaccessible bile duct. 33 patients with endoscope access, 27 patients with percutaneous access</p> <p>Extracorporeal shock wave lithotripsy (ESWL) (n=30) versus intracorporeal laser lithotripsy (ILL) (n=30)</p>	<p>Bile duct clearance:</p> <p>ESWL: 73%</p> <p>ILL: 97%</p>	<.05	Not formally enumerated, appeared to be mild		<p>Treatment sessions needed, mean:</p> <p>ESWL: 3.0</p> <p>ILL: 1.2</p> <p>Duration of treatment, mean days:</p> <p>ESWL: 3.9</p> <p>ILL: 0.9</p>	<p><.001</p> <p><.001</p>

Table 16. Intracorporeal vs. extracorporeal lithotripsy for common bile duct stones, observational studies

Study	N	Population and Interventions	Outcomes	P	Adverse effects, complications	P	Resource utilization	P
Adamek, Maier, Jakobs et al., 1996	125	<p>Patients with stones not removeable with ERCP techniques due to large stone size, impaction, biliary stricture, inaccessible bile duct</p> <p>Extracorporeal shock wave lithotripsy (ESWL) (n=79) versus intracorporeal electrohydraulic lithotripsy (EHL) (n=46)</p>	<p>Fragmentation of stones: ESWL: 97% EHL: 93%</p> <p>Bile duct clearance: ESWL: 79% EHL: 74%</p>	<p>n.s.</p> <p>n.s.</p>	Not formally compared between treatments		<p>Treatment sessions needed, mean: ESWL: 2.0 EHL: 1.1</p> <p>Hospital stay, mean days: ESWL: 13 EHL: 11</p>	<p>N/A</p> <p>N/A</p>

Characteristics of patients, stone size, number of stones, stone location not statistically significantly different between treatment groups.

Review of Evidence: Endoscopic balloon dilation versus endoscopic sphincterotomy

Two randomized controlled trials (Bergman, Rauws, Fockens et al., 1997 and Ochi, Mukawa, Kiyosawa et al., 1999) compared endoscopic balloon dilation to endoscopic sphincterotomy for removal of common bile duct stones in a total of 312 patients (Table 17). Study quality was judged as “Good” for both trials.

Concern about possible long term effects of sphincterotomy on biliary function, plus concern about hemorrhage induced by sphincterotomy have led to consideration of dilation of the biliary sphincter as an alternative method to remove common bile duct stones. Dilation would potentially preserve the function of the biliary sphincter. However, concern has been raised that pancreatitis may occur more often as a complication after balloon dilation.

However, neither study assesses long term outcomes, so the only outcomes that can be assessed are success in removing common bile duct stones and early complications. Both studies found that although balloon dilation ultimately produces equivalent stone removal rates (Bergman, Rauws, Fockens et al., 1997, balloon 89 percent success, sphincterotomy 91 percent success; Ochi, Mukawa, Kiyosawa et al., 1999, balloon 93 percent success, sphincterotomy 98 percent). Some patients in the balloon treatment arm must either cross over or be subject to additional procedures such as mechanical lithotripsy to compensate for the lower initial success rate. Early complications and follow-up complications were not statistically significantly different in the Bergman, Rauws, Fockens et al. (1997) study. In the Ochi, Mukawa, Kiyosawa et al. (1999) study, early complications were not statistically different. Late complications were reported (balloon 4 percent, sphincterotomy 15 percent), but statistical significance tests were not reported.

DiSario, Freeman, Bjorkman et al., (1998) also completed a randomized controlled trial comparing balloon dilation to sphincterotomy, but this trial had only been reported in abstract form in 1998. The results of this study are summarized here because it is commonly cited in reviews and the findings on post-procedure pancreatitis are striking. In this randomized controlled trial of 240 patients, stone clearance was achieved in 99 percent of patients. However, morbidity occurred in 15 percent of balloon dilation patients and 4 percent of sphincterotomy patients ($p=0.014$) Most of the morbidity in the dilation group was due to moderate or severe pancreatitis which occurred in 4 patients and resulted in 2 deaths.

Review of Evidence: Needle-knife fistulotomy versus needle-knife precut papillotomy for the treatment of common bile duct stones in patients with difficult cannulations

Mavrogiannis, Liatsos, Romanos et al. (1999) performed a randomized, controlled trial ($n=153$) comparing two precutting techniques for cannulating the common bile duct when difficulty is encountered when trying to cannulate the common bile duct. (Table 18). Needle-knife fistulotomy (NKF) has been proposed as a safer method of precutting than traditional needle-

Table 17. Endoscopic balloon dilation versus endoscopic sphincterotomy for removal of bile duct stones, randomized trials

Study	N	Population and Interventions	Outcomes	P	Adverse effects, complications	P	Resource utilization	P
Bergman, Rauws, Fockens et al., 1997	202	Patients referred for ERCP for removal of CBD stones, stones visualized Balloon dilation and stone removal versus sphincterotomy and stone removal	Stone removal in one session: Balloon: 89% Sphincterotomy: 91% *9 patients in Balloon group required sphincterotomy to remove stones	n.s.	Early complications: Balloon: 17% Sphincterotomy: 24% Follow-up complications: Balloon: 18% Sphincterotomy: 23%	n.s. n.s.		
Ochi, Mukawa, Kiyosawa et al., 1999	110	Patients referred for ERCP for removal of CBD stones, stones visualized, < 15 mm and less than 10 stones Balloon dilation and stone removal versus sphincterotomy and stone removal	Stone removal, final: Balloon: 93% Sphincterotomy: 98% Stone removal after initial procedure (before lithotripsy): Balloon: 78% Sphincterotomy: 94%	.36 .02	Early complications: Balloon: 2% Sphincterotomy: 6% Late complications: Balloon: 4% Sphincterotomy: 15%	n.s. n/a		

Table 18. Needle-knife fistulotomy versus needle-knife precut papillotomy for the treatment of common bile duct stones

Study	N	Population and Interventions	Outcomes	P	Adverse effects, complications	P	Resource utilization	P
Mavrogiannis, Liatsos, Romanos et al., 1999	153	Consecutive patients who required treatment of suspected choledocholithiasis who had difficulty achieving selective CBD cannulation were randomized to either needle-knife fistulotomy (NKF, n=74) or needle-knife precut papillotomy (NKPP, n=79). All patients had biochemical cholestasis and one or more of the following: biliary pain, bile duct cannulation, and gallbladder stones.	Cannulation success rates (overall): NKF=90.5% NKPP=88.6% Successful stone extraction without lithotripsy NKF (40/48) = 83% NKPP (45/46) =98% Overall stone extraction NKF =100% NKPP =100%	n.s. n.s.	<u>Comp (%)</u> : <u>NKF</u> <u>NKPP</u> Bleeding 6.75 5.06 Perforation 2.7 2.53 Cholangitis 1.35 0 Pancreatitis 0 7.59 Total 10.81 15.18 Hyperamylasemia 2.7 17.72 Death 0 1.26	n.s. n.s. n.s. .05 n.s. .01 n.s.		

knife precut papillotomy (NKPP), with the potential disadvantage of a smaller opening into the bile duct which may prevent successful stone removal.

Overall success in cannulating the common bile duct (after second attempts) was equivalent between the two techniques (NKF 91 percent, NKPP 89 percent, $p=n.s.$) Stone removal without use of lithotripsy was greater for NKPP than for NKF (98 percent versus 83 percent), but final stone removal rates were 100 percent for both groups. Overall complications were not statistically significantly different (NKF 11 percent, NKPP 15 percent, $p=n.s.$), but NKPP had a greater pancreatitis rate (7.6 percent versus 0 percent, $p<0.05$) and a higher rate of hyperamylasemia (17.7 percent versus 2.7 percent, $p<0.01$). Both methods appear to be similar in the management of patients with common bile duct stones.

Review of Evidence: Endoscopic biliary endoprosthesis versus endoscopic sphincterotomy and stone extraction for common bile duct stones in high risk patients

One randomized study (Chopra, Peters, O'Toole, et al., 1996) compared biliary endoprosthesis placement to conventional endoscopic sphincterotomy and stone extraction for patients with common duct stones who were at high risk because of old age or serious debilitating disease. It was theorized that placement of the endoprosthesis might successfully prevent biliary complications with lower short term morbidity than endoscopic sphincterotomy.

Early complications arising within 72 hours after the procedure were 3/43 in the endoprosthesis group and 7/43 in the endoscopic sphincterotomy group ($p=0.18$). Among the 82 patients followed long term for a median of 16 to 20 months, 9 patients in the endoprosthesis group had 11 episodes of cholangitis, and 6 patients in the endoscopic sphincterotomy group developed cholangitis. Overall, a higher proportion of the sphincterotomy group (86 percent) remained free of biliary complications at 20 months than the endoprosthesis group (64%, $p=0.03$). Thus although endoprosthesis placement is as effective and safe as sphincterotomy over the short term, complications and cholangitis are higher over the long term.

Conclusion

Overall, a very thin literature spread out over many different comparisons of interest prevents strong conclusions about any specific treatment comparison. Keeping in mind this thin literature base, the available evidence suggests that:

- Laparoscopic common bile duct exploration may be better than ERCP strategies to manage cholecystectomy patients with the least resource use.
- Definitive surgery prevents long term complications at acceptable short-term morbidity when compared to sphincterotomy alone in high-risk surgical patients.
- Endoscopic treatment of acute cholangitis reduces short-term mortality when compared to emergency surgery.

- Limited evidence suggests that intracorporeal and extracorporeal lithotripsy methods show similar outcomes in removing large common bile duct stones.
- Limited evidence suggests similar stone removal rates and short-term complications when comparing balloon dilation and sphincterotomy.
- Limited evidence suggests similar stone removal rates and complications when comparing needle-knife fistulotomy to needle-knife precut papillotomy.
- Limited evidence suggests that endoscopic sphincterotomy and duct stone clearance is more effective than biliary endoprosthesis placement for prevention of long term complications in patients considered to be high surgical risks.

Part I, Section 3: Diagnostic Value of Individual Risk Factors or Predictive Models for Assessing the Likelihood of Having a Common Bile Duct Stone

Introduction

In trying to determine optimum diagnostic and treatment strategies, many investigators have analyzed individual risk factors and combinations of risk factors that may predict the presence or absence of common bile duct stones. With information about the probability of a common bile duct stone, it may be possible to design a diagnostic and treatment strategy that minimizes patient morbidity and/or minimizes medical resource utilization.

The data reviewed here cannot be directly translated into optimum diagnostic and treatment strategies because there are many possible strategies, given the variety of methods possible to diagnose common bile duct stones (ERCP, MRCP, endoscopic ultrasound, intraoperative cholangiogram) and treat them (preoperative ERCP, laparoscopic common bile duct exploration, postoperative ERCP, expectant management).

However, a few simple principles surface. From the perspective of the individual patient, the probability of a common duct stone is the key factor in determining which approach may be best. If the preoperative probability of a common bile duct stone is high enough, ERCP tends to become efficient and effective because both diagnosis and therapy can be carried out in a single procedure in one setting. If the preoperative probability of a common duct stone is low enough, then it may be possible to avoid any diagnostic procedure to diagnose common duct stones and rely on expectant postoperative management with ERCP to manage any stones that were missed. In the middle range of probability, use of diagnostic tests such as EUS, MRCP, or intraoperative cholangiogram may be efficient methods to treat patients.

All the risk factors or decision rules evaluated in this section have potentially variable cutoff thresholds, so that sensitivity or specificity can be manipulated with the expected trade-offs to produce a particular positive or negative predictive value. However, at a particular cutoff point that produces the desired predictive value, a superior risk factor or decision rule will have higher sensitivities and specificities than other decision rules, and thus better performance in discriminating between those patients who do and do not have stones.

For example, suppose that a probability of stone of 60 percent or greater makes preoperative ERCP the optimum strategy for that particular patient. For example, risk factor A at a particular cutoff produces a positive predictive value of 60 percent, and risk factor B at a particular cutoff point also produces a positive predictive value of 60 percent in the same population. However, risk factor A only identifies 40 percent of the patients with stones at that cutoff (40 percent sensitive), and risk factor B identifies 80 percent of the patients with stones at that cutoff (80 percent sensitivity). Thus, using risk factor B, 80 percent of the patients with stones can be managed by a strategy which requires a 60 percent probability of stone to be optimal.

In sum, then, given that the particular cutoff threshold can be varied to meet desired criteria, then the exact sensitivity and specificity calculated in any single study is not important. The critical factor differentiating any of these risk factors or decision rules is the capability to have both the highest sensitivity and specificity, or in the parlance of diagnostic decision-making, the best receiver-operator characteristic (ROC). Then the cutoff point can be defined that produces the sensitivities and specificities that result in the desired positive predictive value. The studies reviewed here did not in general calculate ROC curves. A risk factor or decision rule with both high sensitivity and specificity would have the best ROC.

Evidence Base

A total of 13 studies with a total of 7,409 patients contributed to the findings reported here. Most studies reported on several of the individual risk factors, some reported on individual risk factors and a multivariate risk prediction model.

Review of Evidence: Univariate Risk Factors for Common Bile Duct Stones

The single risk factors commonly examined in studies included clinical jaundice or elevated bilirubin, liver function tests, and ultrasound findings of a dilated common bile duct. Studies varied in the definitions and cutoff thresholds for the various tests

Five studies (total n=2,661) reported on clinical jaundice as a risk factor (Table 19). Positive predictive values ranged from 29 percent to 86 percent, sensitivity from 24 percent to 56 percent, and specificity from 87 percent to 99 percent. Clinical jaundice does not have an exact threshold cutoff value, nor is the reliability of measurement certain. In general, though, sensitivities are low, specificities are higher, and in the situation of a low prevalence condition such as common bile duct stones, the high specificity drives the predictive values to be high.

Six studies (total n=2369) reported on bilirubin levels. At varying cutoff levels, positive predictive values ranged from 42 percent to 95 percent, sensitivity from 31 percent to 56 percent, and specificity from 48 percent to 99 percent. In general, sensitivities were low, specificities higher, and the resulting positive predictive values are reasonably high.

Eight studies (total n=3,551) reported on various liver function tests (Table 20). Some studies examined more than 1 cutoff level. There was a broad range of predictive values, sensitivities and specificities for all the different liver function tests examined. In general, the trade off between sensitivity and specificity can be noted in all the studies. The studies with cutoff values that produce high specificity tend to have low sensitivity, but this type of cutoff produces the highest positive predictive values.

Ten studies (total n=4,321) reported on the finding of a dilated common bile duct seen on ultrasound (Table 21). The threshold for a dilated duct varied from 5 to 10 mm, and was undefined in a few studies. Predictive values ranged from 28 percent to 91 percent, sensitivities from 28 percent to 94 percent, and specificities from 72 percent to 98 percent. Studies with high sensitivity tend to have low specificity, and vice versa.

Table 19. Jaundice or elevated bilirubin as a risk factor for CBD stone

Study	Population	% prevalence of stone in population	n	Rule tested	Predictive Value	Sensitivity	Specificity	Comments
Alponat, Kum, Rajnakova et al., 1997	Patients with risk factors for CBD stones having ERCP	32	192	jaundice	67	56	87	
Barkun, Barkun, Fried et al., 1994	Patients undergoing lap cholecystectomy who had ERCP	48	139	bilirubin>1.8	57	48	48	
Bergamaschi, Tuech, Braconier et al., 1999	Patients undergoing lap cholecystectomy	15	990	jaundice	76	24	99	
Hauer-Jensen, Karesen, Nygaard et al., 1985	Patients undergoing cholecystectomy	12	319	jaundice bilirubin>1.5	29 42	26 45	91 91	
Kim, Kim, Lee et al., 1997a	Patients undergoing lap cholecystectomy	17	561	jaundice bilirubin >2	52 53	36 41	93 92	
Koo and Traverso 1996	Patients undergoing lap cholecystectomy	12	420	bilirubin>1.2	47	31	93	
Menezes, Marson, Debeaux et al. 2000	Patients undergoing lap cholecystectomy	33	233	bilirubin>nl bilirubin>2xnl	95 92	48 31	98 99	
Santucci, Natalini, Sarpi et al., 1996	Patients undergoing lap cholecystectomy	9	697	bilirubin>3	83	56	82	
Trondsen, Edwin, Reiertsen et al., 1995	Patients undergoing lap cholecystectomy	38	599	jaundice	86	46	95	

Table 20. Elevated liver function tests as a risk factor for CBD stone

Study	Population	% prevalence of stone in population	n	Rule tested	Predictive Value	Sensitivity	Specificity	Comments
Alponat, Kum, Rajnakova et al., 1997	Patients with risk factors for CBD stones having ERCP	32	192	Any LFT>2xnl AST > 2xnl ALT > 2xnl Alk phos >2xnl GGT > 2xnl LDH > 2xnl	37 41 40 43 35 38	84 89 87 84 87 68	33 40 38 46 22 46	Numbers for any LFT do not make sense, cannot be less sensitive
Barkun, Barkun, Fried et al., 1994	Patients undergoing lap cholecystectomy who had ERCP	48	139	AST>120 Alk phos>300	49 53	81 79	25 35	
Bergamaschi, Tuech, Braconier et al., 1999	Patients undergoing lap cholecystectomy	15	990	Alk phos >400 <u>and</u> GGT>200	87	58	99	
Hauer-Jensen, Karesen, Nygaard et al., 1985	Patients undergoing cholecystectomy	12	319	Alk phos>250	37	58	87	
Kim, Kim, Lee et al., 1997a	Patients undergoing lap cholecystectomy	17	561	SGOT>50 SGPT>50 Alk phos>160	43 39 50	65 67 75	82 79 85	
Koo and Traverso 1996	Patients undergoing lap cholecystectomy	12	420	SGOT>44 Alk phos>140	48 48	40 31	94 93	
Menezes, Marson, Debeaux et al. 2000	Patients undergoing lap cholecystectomy	33	233	SGOT>nl SGOT>2xnl Alkphos>nl Alkphos>2xnl	88 93 77 97	47 35 66 44	97 99 90 99	
Santucci, Natalini, Sarpi et al., 1996	Patients undergoing lap cholecystectomy	9	697	ALT> 40 AST> 40 GGT>150 Alk phos>300	88 76 75 94	94 78 80 72	79 78 76 90	Cutoffs established by ROC analysis, maximize sensitivity and specificity

Table 21. Dilated CBD as a risk factor for CBD stone

Study	Population	% prevalence of stone in population	n	Rule tested	Predictive Value	Sensitivity	Specificity	Comments
Alponat, Kum, Rajnakova et al., 1997	Patients with risk factors for CBD stones having ERCP	32	192	Dilated CBD with stone on ultrasound Dilated CBD without stone on ultrasound	72 36	42 31	92 74	
Barkun, Barkun, Fried et al., 1994	Patients undergoing lap cholecystectomy who had ERCP	48	139	Dilated CBD, subjective	64	53	73	
Bergamaschi, Tuech, Braconier et al., 1999	Patients undergoing lap cholecystectomy	15	990	CBD > 8mm	75	28	98	
Hauer-Jensen, Karesen, Nygaard et al., 1985	Patients undergoing cholecystectomy	12	319	CBD >10 mm	34	63	92	
Kim, Kim, Lee et al., 1997a	Patients undergoing lap cholecystectomy	17	561	CBD > 10 mm	61	94	88	
Koo and Traverso 1996	Patients undergoing lap cholecystectomy	12	420	CBD> 5mm + 1 mm per decade over age 50	28	22	92	
Menezes, Marson, Debeaux et al. 2000	Patients undergoing lap cholecystectomy	33	233	CBD dilated (not defined)	91	51	97	
Santucci, Natalini, Sarpi et al., 1996	Patients undergoing lap cholecystectomy	9	697	CBD> 8 mm	74	59	72	

Table 21. Dilated CBD as a risk factor for CBD stone (cont'd)

Study	Population	% prevalence of stone in population	n	Rule tested	Predictive Value	Sensitivity	Specificity	Comments
Trondsen, Edwin, Reiertsen et al., 1998	Patients undergoing lap cholecystectomy	15	171	CBD > 6 mm	35	64	79	
Trondsen, Edwin, Reiertsen et al., 1995	Patients undergoing lap cholecystectomy	38	599	CBD dilated (not defined)	85	31	96	

In sum, although all the previously mentioned single risk factors for common duct stones have significant associations with the presence of stones, none of them have outstanding ROC characteristics. The presence of any of these factors certainly increases the probability of the presence of a common bile duct stone, possibly high enough to change clinical decision-making. However, changing the cutoff value to increase the positive predictive value (by increasing the specificity) usually results in poor sensitivity.

Review of Evidence: Multivariable Predictors for Common Bile Duct Stones

Four studies (total n=1,461) examined the use of multiple risk factors for prediction of the presence of common bile duct stones (Table 22). Many studies that simply used the criterion of “any one risk factor” as a prediction rule were not included in this evidence review, as such a criterion has been used for many years to select patients for ERCP and has a known poor specificity and low positive predictive value.

The four studies varied in the analytic technique used to develop the prediction rule. Hawasli, Lloyd, Pozios et al. (1993) did not use any quantitative technique but defined combinations of risk factors to classify patients at high risk of stones. Menezes, Marson, Debeaux et al. (2000) developed a logistic model based on age, sex, jaundice, presence of cholangitis, liver function tests, and ultrasound examination of the common bile duct. Trondsen, Edwin, Reiertsen et al. (1995) used a discriminant analysis technique based on age, bilirubin, alanine aminotransferase, and gamma glutamyltransferase. In Trondsen, Edwin, Reiertsen et al. (1998), a new rule was not developed, but the previously developed discriminant analysis rule was prospectively validated in a new population of patients.

Thus, except for Trondsen, Edwin, Reiertsen et al. (1998), the findings of the three other studies should be viewed as optimistic estimates of stone prediction, since the performance of the rules was only evaluated on the set of patients used to develop the rule.

All the studies produced decision rules in which both the sensitivity and specificity were greater than 80 percent. However, these findings should be viewed cautiously, since there has been no independent validation. The prospective validation study by Trondsen, Edwin, Reiertsen et al. (1998) is a particularly strong finding, since the rule was derived from an independent population—the sensitivity was 94 percent and the specificity was 88 percent in an independent set of patients. The discriminant function cutoff could be varied to increase sensitivity at the expense of specificity or vice-versa, but since both are high the actual discriminative capability of the rule compared to individual risk factors was far superior.

In conclusion, multivariable modeling of risk factors for prediction of common duct stones shows promise as a method of triage for determining appropriate treatments, given that they appear to have superior discriminatory power. These prediction models have yet to be integrated into clinical decision models to determine optimal cutoffs.

Table 22. Decision rules for prediction of stones

Study	population	% prevalence of stone in population	n	Rule tested	Predictive value	Sensitivity	Specificity	Comments
Hawasli, Lloyd, Pozios et al., 1993	Patients undergoing lap cholecystectomy	4	459	High suspicion combination	75	83	99	
Menezes, Marson, Debeaux et al. 2000	Patients undergoing lap cholecystectomy	15	211	Score \geq 2 Score \geq 3 Based on logistic regress	56 67	86 82	66 80	
Trondsen, Edwin, Reiertsen et al., 1995	Patients undergoing lap cholecystectomy	38	599	Discriminant function	91	95	94	Rule applied to same data used to develop function
Trondsen, Edwin, Reiertsen et al., 1998	Patients undergoing lap cholecystectomy	17	192	Discriminant function	60	94	88	Same 2 by 2 data as Trondsen, Edwin, Reiertsen et al., 1995, above

Review of Evidence: Absence of Any Risk Factor as A Predictor of Common Bile Duct Stone Absence

Seven studies (total n=599) examined the prediction of absence of common duct stones (Table 23). Usually, the absence of any of the known risk factors (all the individual factors reviewed previously) was used as the indicator. Trondsen, Edwin, Reiertsen et al. (1995) and Trondsen, Edwin, Reiertsen et al. (1998) reviewed previously, are also included here because the discriminant function used to predict stones can also be used to predict the absence of stone.

If the prevalence of stone is low enough in some patients, then some clinicians might avoid use of any diagnostic test to diagnose common duct stones. Such a case would be very compelling if the probability of stone is in the same range or lower as it is in the case of a negative ERCP examination. Although ERCP is selectively performed on patients with higher risk of common duct stones, if physicians are willing to believe a negative ERCP, they should be willing to believe a prediction rule if the probabilities of stones are equally low.

The seven studies reported a probability of common duct stones in those predicted not to have stones between a range of 0.25 percent to 7 percent. In all studies, a reasonable sensitivity for stone-free patients was shown, from 60 percent to 98 percent, and reasonable specificity, 60 percent to 96 percent. Thus, the decision rules all can identify more than half of the patients that do not have stones.

The strongest finding is Trondsen, Edwin, Reiertsen et al. (1998), in which the same discriminant function which identifies stones can rule out stones with both high sensitivity (88 percent) and specificity (94 percent). This study is also a validation study of an independently developed discriminant function, which further increases its validity.

These probabilities of stones compare quite favorably to the probabilities of stones in patients having a negative ERCP. If the probability is calculated, using the equation “1-NPV” and some of the reported NPVs of the ERCP studies in the section of this report comparing ERCP to EUS, a range of stone probabilities is calculated from 0 percent to 17 percent.

In conclusion, the absence of any risk factors for stones (or a discriminant function indicating absence of stone) is a very strong predictor of the absence of stones, producing probabilities of stones that are in the same range as a negative ERCP exam in a patient with risk factors for stones.

Conclusions

The probability of a common duct stone is the key factor to determining diagnostic and treatment strategies. When preoperative probability of a common bile duct stone is high enough, ERCP may be preferred because diagnosis and therapy can be carried out in a single procedure. If the preoperative probability of a common duct stone is low enough, then expectant management may be preferred in order to avoid unnecessary procedures. In the middle range of probability, use of diagnostic tests such as EUS, MRCP, or intraoperative cholangiogram may be used to further discriminate patients with high or low probability of common bile duct stones.

Table 23. Rules ruling out stones, absence of stone is the outcome

Study	population	% prevalence of stones in population	n	Rule tested	Prevalence of stone in those ruled out by rule (1 – PPV)	Sensitivity--% of stone-free patients detected by rule	Specificity--% of patients with stones ruled out by rule	Comments
Carroll, Phillips, Rosenthal et al., 1996	Patients undergoing lap cholecystectomy	15	100	Normal LFTs, CBD, past history	4	61	87	
Hawasli, Lloyd, and Cacucci 2000	Patients undergoing lap cholecystectomy	5	2834	Normal LFTs, CBD, past history	0.25	89	96	Hawasli, Lloyd, Pozios et al. 1993 results of this same question included in these data
Khaira, Ridings, and Gompertz 1999	Patients undergoing lap cholecystectomy	5	154	Normal LFTs, CBD, past history	1	60	88	
Koo and Traverso 1996	Patients undergoing lap cholecystectomy	12	420	Normal LFTs, US, past history	7	78	60	
Santucci, Natalini, Sarpi et al., 1996	Patients undergoing lap cholecystectomy	9	697	Normal LFTs, US, past history	1.4	98	86	Clinical followup to detect stones in patients with no indications
Trondsen, Edwin, Reiertsen et al., 1998	Patients undergoing lap cholecystectomy	17	192	Discriminant function value negative	1.4	88	94	Rule applied to validation set of patients
Trondsen, Edwin, Reiertsen et al., 1995	Patients undergoing lap cholecystectomy	38	599	Discriminant function value negative	3	94	95	Rule applied to same data used to develop function

Thirteen studies with a total patient population of 7,409 patients that reported multiple findings of sensitivities and specificities of a single or combination of risk factors to predict the presence of common bile duct stones were reviewed.

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.

Four studies tested prediction rules based on combinations of risk factors for the presence of stones. All the studies produced decision rules in which both the sensitivity and specificity were greater than 80 percent. These findings must be viewed cautiously, since only one study was a validation of an independently developed prediction rule. Presently, multivariable modeling of risk factors for prediction of common duct stones is a promising approach.

The absence of any risk factors for stones (or a discriminant function indicating absence of stone) is a very strong predictor of the absence of stones, producing 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).

Results and Conclusions, Part II: Pancreaticobiliary Malignancy

This chapter reviews evidence on the following questions:

In patients with known or suspected pancreaticobiliary malignancy,

- a. What is the diagnostic performance of ERCP tissue sampling techniques, in establishing a tissue biopsy diagnosis of pancreaticobiliary malignancy in comparison to each other or alternative nonsurgical tissue sampling techniques (e.g., endoscopic ultrasound-guided fine-needle aspiration (FNA) or percutaneous FNA)? (*Section 1: Diagnostic Performance of Nonsurgical Tissue Sampling Techniques in Pancreaticobiliary Malignancy – Comparison of Strategies Using ERCP, EUS, or Percutaneous Approach*)
 - 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)? (*Section 2: Diagnostic Performance of ERCP in Pancreaticobiliary Malignant Obstruction – Comparison To Alternatives*)
 - c. What are the outcomes of treatment using ERCP strategies to treat malignant pancreaticobiliary obstruction compared to using surgical or interventional radiology treatment? (*Section 3: Outcomes of Treatment Using ERCP for Palliation of Pancreaticobiliary Malignancy – Comparison of Strategies Using ERCP, Surgery, or Interventional Radiology; A. Comparison of ERCP stent versus Surgical Bypass; B. Comparison of Metal vs. Plastic stents During ERCP; C. Additional Comparisons of ERCP Strategies*)
- (*Section 4: Outcomes of Treatment Using Preoperative ERCP Drainage for Relief of Malignant Obstructive Jaundice*)

Part II, Section 1: Diagnostic Performance of Nonsurgical Tissue Sampling Techniques in Pancreaticobiliary Malignancy—Comparison of Strategies Using ERCP, EUS, or Percutaneous Approach

Introduction

When a malignant cause is suspected for biliary obstruction, preoperative tissue confirmation of malignancy may be helpful in guiding management decisions. Nonsurgical tissue sampling methods include endoscopic and percutaneous approaches. Cytologic assessment can be performed on endoscopically acquired specimens such as aspirated biliary or pancreatic fluid, wire brushing specimens, or fine-needle aspiration (FNA) specimens. FNA specimens can be obtained during ERCP, EUS, or through a percutaneous approach using imaging guidance. Endoscopic tissue biopsy can be performed during ERCP with a forceps device.

The goal of tissue sampling techniques is to provide sufficient cellular material to make an accurate pathologic diagnosis. Theoretically, increasing the numbers of samples and/or the types of samples might yield more cellular tissue for assessment and might improve diagnostic accuracy, but the extent to which combinations of different sampling techniques increase the diagnostic accuracy is still being investigated (Lee and Leung 1998).

It is outside the scope of this systematic review to determine whether biliary versus pancreatic location of sampling is related to differences in diagnostic performance of sampling techniques. A recent review summarized the diagnostic sensitivity of brush cytology for detection of pancreatic cancer (Lee and Leung 1998). In a total sample of 362 patients who had pancreatic cancer, brush cytology samples diagnosed 55% of cases with a range among studies of 0–85%. When the subset of 190 brush cytology samples taken from the pancreatic duct was analyzed separately, 66% of pancreatic cancers were detected. The few studies using blinded readings reported a lower range of sensitivity (0–40%).

Cytology findings may be interpreted as definite malignancy or may be reported according to the degree of atypia. The sensitivity and specificity of cytology will be dependent on where the criterion is set for calling the test positive. Using a strict criterion where only definite malignancy is counted as positive will achieve the highest specificity, but the associated sensitivity will usually be the lowest. Likewise, considering any degree of atypia as a positive test will increase the test's sensitivity, but the specificity will generally be reduced.

This systematic review selected studies comparing the diagnostic performance of at least 2 of the available nonsurgical tissue sampling techniques in patients with pancreaticobiliary malignancy. Comparative studies including at least one ERCP tissue sampling technique compared to an alternative technique were the primary focus defined prospectively in the systematic review protocol. None of the studies identified with this set of selection criteria included any comparison of ERCP tissue techniques and EUS sampling techniques. Upon discussion of this result with the Technical Advisory Group, a supplementary request was made to review single arm studies reporting the diagnostic performance of endoscopic ultrasound (EUS) fine-needle aspiration (FNA). Studies included in this secondary analysis were not selected using a formalized systematic review, but were identified by manually searching for recent reports on EUS-FNA and carefully reviewing prior articles referenced in these studies to identify additional studies.

Evidence Base

Twelve studies comparing at least two tissue sampling techniques were identified in this systematic review. Quality ratings are displayed in Table 24. Five of these studies were rated as “Good” quality, signifying the use of blinded interpretation of test results. Only three studies include over 100 patients, and six studies include less than 50 subjects.

There is considerable variation in reported estimates of sensitivity for each tissue sampling technique, and comparison of results for the same technique across studies may be limited due to differences in populations with regard to distribution of tumor types as well as differences in tissue sampling technique and interpretation methods. To minimize this problem, this analysis

Table 24. Quality Assessment

Study Author, Year	Patient Enrollment	Diagnostic performance of ERCP determined without knowledge of other test results	Diagnostic Performance of other test(s) determined without knowledge of ERCP results	Summary Evaluation
Jaiwala, Fogel, Sherman et al., 2000	(n=133 pts) Prospective Study Enrollment of subjects stated to be selected and nonconsecutive and reasons for exclusion were stated.	No	No	Fair
Kurzwinski, Deery, Dooley et al., 1993	(n=46 pts) Prospective study of 37 of 46 consecutive pts w/ biliary tract stricture had ERCP and 9 had PTC cytology. Reasons for exclusions provided.	No	No	Fair
de Peralta-Venturina, Wong, Purslow et al., 1996	(n=74 pts; 104 spec) Retrospective review of all eligible cytology specimens during 1990 to mid 1994 in pts with verified diagnosis.	Yes	Yes	Good
Foutch et al. 1991	(n=30 pts; 78 specimens) Prospective study 30 consecutive patients with bile duct stricture	Yes	Yes	Good
Mansfield et al. 1997	(n=43 pts; 54 procedures) Prospective study All pts with biliary stricture suspicious for malignancy	Yes	Yes	Good

Table 24. Quality Assessment (cont'd)

Study Author, Year	Patient Enrollment	Diagnostic performance of ERCP determined without knowledge of other test results	Diagnostic Performance of other test(s) determined without knowledge of ERCP results	Summary Evaluation
Sugiyama, Atomi, Wada et al., 1996	(n= 43 pts) Prospective study 52 Consecutive pts with stricture (n=48) or filling defect (n=4) Papillary lesions excluded. Analysis includes 43 pts with all 3 techniques	No	No	Fair
Howell, Beveridge, Bosco et al., 1992	?Prospective 31 consecutive patients with malignant appearing strictures	No	No	Fair
Ferrari, Lichtenstein, Slivka et al., 1994	(n=74) Retrospective study of all pts who had ERCP with brush cytology of biliary or pancreatic duct stricture	No	No	Fair
Ponchon, Gagnon, Berger et al., 1995	(n=193) Prospective study Enrolled subjects meeting entry criteria. Complete explanation of enrollment process provided.	Yes	Yes	Good
Schoefl, Haefner, Wrba et al., 1997	119 consecutive pts (133 samples) ?retrospective	No	No	Fair

Table 24. Quality Assessment (cont'd)

Study Author, Year	Patient Enrollment	Diagnostic performance of ERCP determined without knowledge of other test results	Diagnostic Performance of other test(s) determined without knowledge of ERCP results	Summary Evaluation
Pugliese, Antonelli, Vincenti et al., 1997	(n=52) Prospective enrollment of consecutive biliary strictures at ERCP Excluded strictures associated with bile duct stones, perampullary tumors, or postop stricture	Yes	Yes	Good
Gmelin and Weiss 1981	(n=32) 32 proven malignant or benign tumors in papillary region out of 36 consecutive cases.	Uncertain	Uncertain	Fair

will focus primarily on within-study comparisons of the relative sensitivity of alternative sampling techniques. However, this problem is not completely avoided because the selected comparative studies frequently reported diagnostic performance for individual sampling techniques being compared on a different number of patients and thus slight differences in the population characteristics may be present.

Given that the expected difference in diagnostic performance between tissue sampling techniques and the diagnostic alternatives reported here are frequently relatively small and the number of cases with the outcome of interest is generally small, these studies may have limited power to detect statistically significant differences in test performance. Only 4 of 12 studies (Jaiwala, Fogel, Sherman et al., 2000; Sugiyama, Atomi, Wada et al., 1996; Ponchon, Gagnon, Berger et al., 1995; Kurzawinski, Deery, Dooley et al., 1993) actually reported any statistical comparisons, and all of these only reported chi square comparisons of sensitivity.

The specificity estimates for cytology techniques reported in these studies were generally close to 100%, though Jaiwala, Fogel, Sherman et al. (2000; n=133) found that specificity fell to 90% when any atypia was considered equivalent to malignancy.

The nonsurgical tissue sampling techniques being evaluated in these studies are measured against a reference standard incorporating the best available information from surgical findings, surgical or nonsurgical pathology, autopsy, imaging follow-up, and clinical follow-up.

Review of Evidence: Diagnostic Performance

Bile Aspiration Cytology Compared to Brush Cytology

Five studies (total n=approximately 178), including 3 with “Good” quality, (Kurzawinski, Deery, Dooley et al., 1993; de Peralta-Venturina, Wong, Purslow et al., 1996; Foutch et al. 1991; Mansfield et al. 1997; Sugiyama, Atomi, Wada et al., 1996) provided comparisons between bile cytology and brush cytology for biliary strictures (Table 25 and Table 26). In each comparison, brush cytology provided higher sensitivity than bile aspirate cytology, although only one study reported a statistical assessment. The absolute increase in sensitivity ranged from 16 to 50%. Reported range of bile cytology sensitivity was 6–50% and that for brush cytology was 33–100%.

Two studies reported comparative data for tissue sampling using an ERC approach versus a percutaneous transhepatic cholangiographic (PTC) approach. de Peralta-Venturina, Wong, Purslow et al. (1996) noted lower sensitivity with PTC compared with ERC, 43 versus 100%. Kurzawinski, Deery, Dooley et al. (1993) observed similar sensitivity for brush cytology techniques using either approach and possibly lower sensitivity for bile aspirates with PTC.

In sum, the available studies are relatively small and most are limited by lack of statistical analysis but do provide suggestive evidence that brush cytology is more sensitive than bile aspiration cytology.

Table 25. Comparisons of Bile Cytology and Brush Cytology

Study	N Pts	N Spec	Diagnostic test	Prevalence (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Adequate Specimens (%)	Quality Rating and Comments
Kurzawinski, Deery, Dooley et al., 1993	37	37	ERCP-Bile cytology	81	33 ^a	100	100	26		Fair p< 0.05 a vs. b p< 0.01 c vs. d
	31	31	ERCP-Brush cytology	77	71 ^b	100	100	50		
	9 15	9 15	PTC-Bile cytology PTC-Brush cytology	? ?	0 ^c 67 ^d	n.r. n.r.				
de Peralta- Venturina, Wong, Purslow et al., 1996	74	13 61	Bile cytology Brush cytology ¹⁰	? ?	50 100	100 95	100 95	40 100	69 98	Good Stratified results for bile vs. brushing not reported by ERCP vs. PTC technique
		55 19	ERCP PTC	? ?	100 43	95 100	96 100	100 57	98 79	

Table 26. Comparisons of Bile Cytology, Brush Cytology, and Other Technique

Study	N Pt	N Sp	Diagnostic test	Prevalence (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Adequate Specimens (%)	Quality Rating and Comments
Foutch et al. 1991	30	31 31 16	Bile cytology Brush cytology ¹ Stent cytology	58 58 69	6 33 36	100 100 100	100 100 100	43 52 42		Good
Mansfield et al. 1997	43	54 54 19 19 54	Bile cytology Brush cytology ² Soehendra stent retriever screw head Stent Combined	96 96 ? ? ? ?	12 42 25 37 54	100 100 ? ? 100	100 100 ? ? 100	4 6 ? ? 8	44 96 70 84	Good Clearly malignant or suspicious cytology = (+)
Sugiyama, Atomi, Wada et al., 1996 ³	43 43 43	43 43 43	Bile cytology Brush cytology ⁴ Forceps biopsy	72 72 72	32 ^a 48 ^b 81 ^c	100 100 100	100 100 100	36 43 67	100 88 87	Fair p<0.01, a vs c; p<0.05, b vs. c; p = n.r., a vs b

¹ Milrose Lab, 230 cm, 2.5-mm diameter

² Combocath, Microvasive, Boston Scientific

³ Specifically excluded patients with papillary tumor.

⁴ BC-23Q cytology brush (outer diameter, 1.8 mm, Olympus, Tokyo, Japan)

Brush Cytology Compared to FNA Cytology

Three studies (total n=approximately 193), all rated “Fair” (Jaiwala, Fogel, Sherman et al., 2000; Howell, Beveridge, Bosco et al., 1992; Ferrari, Lichtenstein, Slivka et al., 1994) compare brush cytology with FNA cytology (Table 27 and Table 28). The first two studies use ERCP to obtain both the FNA specimen and the brush cytology specimens while Ferrari, Lichtenstein, Slivka et al. (1994) compares ERCP brush cytology with percutaneous CT-guided FNA. The largest study, (Jaiwala, Fogel, Sherman et al., 2000, n=133) reports similar sensitivity for FNA and for brush cytology and the combination of both techniques increased overall sensitivity by about 9%. This difference was not statistically significant in 2 of 3 comparisons and was found significant ($p<0.05$) only when high-grade atypia was considered equivalent to malignancy.

The study by Howell, Beveridge, Bosco et al. (1992, n=31) notes a higher sensitivity for FNA than for brush cytology (62% vs. 8%) but the combination of both techniques only yielded a slight increase to 65% sensitivity. Ferrari, Lichtenstein, Slivka et al. (1994, n=29 with FNA and 70 for brush cytology) found percutaneous CT-guided FNA to be more sensitive than brush cytology (91% versus 56%) but the large difference in sample sizes makes direct comparison limited. Furthermore, the small size and lack of statistical analysis of these two studies limits the interpretation of these findings.

Among these studies, the findings of Jaiwala, Fogel, Sherman et al. (2000) provide the more reliable information and suggest that brush cytology and ERCP-FNA may be similar in sensitivity. When used together, the available evidence does not demonstrate a statistically significant increase in sensitivity.

Forceps Biopsy Sampling Compared to Brush Cytology

Six studies (total n=approximately 437), including the 3 largest studies and 3 “Good” quality studies, compared forceps biopsy sampling to brush cytology (Tables 25–28). Gmelin and Weiss (1981) exclusively studied papillary tumors and found an increase in sensitivity of about 30% using forceps biopsy over brush cytology (86% versus 55%), but statistical analysis was not reported. Sugiyama, Atomi, Wada et al. (1996) specifically excluded papillary tumors and also found a large increase in sensitivity with forceps biopsy, 81% versus 48%, $p<0.05$. The remaining studies (Jaiwala, Fogel, Sherman et al., 2000; Ponchon, Gagnon, Berger et al., 1995; Schoefl, Haefner, Wrba et al., 1997; Pugliese, Antonelli, Vincenti et al., 1997) included a mixture of pancreaticobiliary malignancies. These studies reported generally similar sensitivity with forceps biopsy compared with brush cytology, though one study (Jaiwala, Fogel, Sherman et al., 2000) noted statistically significant increases for forceps biopsy over brush cytology when atypia was not interpreted as malignancy).

In addition, each of these studies reports that the combination of forceps biopsy and brush cytology increases the sensitivity in detecting malignancy between 5-20%. Jaiwala, Fogel, Sherman et al. (2000) and Ponchon, Gagnon, Berger et al. (1995) both reported the increase in sensitivity for the combination of forceps biopsy plus brush cytology compared to forceps biopsy alone to be statistically significant ($p<0.05$).

Table 27. Comparisons of Brush Cytology and Biopsy Technique

Study	N Pt	N Sp	Diagnostic test	Prevalence (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Adequate Specimens (%)	Quality Rating and Comments
Howell, Beveridge, Bosco et al., 1992	31		Brush cytology FNA – ERCP Combined	84 84 84	8 62 65	100 100 100	100 100 100	17 33 36		Fair
Ferrari, Lichtenstein, Slivka et al., 1994	70 51 19 29		Brush cytology – Overall – Biliary – Pancreatic FNA – percutaneous	76 ?	56 54 64 91	100 100 100 75	100 100 100 95	51 45 67 60	93	Fair
Ponchon, Gagnon, Berger et al., 1995	233	193 118 105	Brush cytology Forceps biopsy ⁵ Combination	66 69 70	35 ^a 43 ^b 63 ^c	97 97 97	96 97 98	66 69 70	90 57	Good p= n.s. for a vs b p<0.001 for a vs c p<0.05 for b vs. c
Schoefl, Haefner, Wrba et al., 1997	59 106 48	65 119 51	Brush cytology ⁶ Forceps biopsy ⁷ Combination	? 	47 65 70	100 100 100	100 100 100	62 69 71		Fair
Pugliese, Antonelli, Vincenti et al., 1997	52	52	Brush cytology ⁸ Forceps biopsy ⁹ Combination	69 69 69	53 53 61	100 100 100	100 100 100	48 48 53		Good Uncertain cytology was considered negative.
Gmelin and Weiss 1981	32	32 26 26	Papillary tumors Brush cytology Forceps biopsy	85 81	18 71 55 86	100 100 100 100	100 100 100 100	18 45 29 63		Fair Suspicious cells considered negative Suspicious cells considered positive

⁵ Either Biomed 31010 (Paris, France: 175 cm length, 2mm diameter, round and fenestrated jaw with 2mm diameter, flexible tip, no needle) or Olympus prototype (Scop Medecine; 180cm length, 2.2mm diameter, round and fenestrated jaw with 2mm diameter, teflon sheath, no needle)

⁶ Endo-Flex 42 22E-A

⁷ Olympus FB-19N for about 60% and FB26N for about 30% and FB-39Q for about 10%

⁸ Olympus mod. BC-19Q or Wilson-Cook Medical Inc., Winston-Salem, NC, Mod. GBC-200-3-3.5

⁹ Olympus FB-19K or FB-39Q

Table 28. Comparison of Brush Cytology, FNA cytology, and Forceps biopsy in biliary strictures

Study	N Pts	N Spec	Diagnostic test	Prevalence (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Adequate Specimens (%)	Quality Rating and Comments
Jaiwala, Fogel, Sherman et al., 2000	133	133	Brush cytology ¹⁰	78	48 ^a	90	94	33	n.r.	Fair Any atypia on cytology was considered equivalent to cancer.
			FNA cytology ¹¹		38 ^b	97	98	30	n.r.	
			Forceps biopsy ^{12 or 13}		54 ^c	76	89	31	n.r.	
			Brush + FNA		57 ^d	86	94	36	n.r.	
			Brush + Biopsy		71 ^e	69	89	40	n.r.	
			Biopsy + FNA		64 ^f	72	89	36	n.r.	P<0.05 for: a vs. e, f, g; b vs. c, d, e, f, g; c vs. e, f, g; d vs. e, g; f vs. g
			Brush+Biopsy+FNA		77 ^g	66	89	44	n.r.	
			Brush cytology		30 ^a	100	100	28		
			FNA cytology		30 ^b	100	100	28		
			Forceps biopsy		43 ^c	90	94	31		
			Brush + FNA		39 ^d	100	100	32		Only high-grade atypia considered equivalent to cancer. P<0.05 for: a vs. c, d, e, f, g; b vs. c, d, e, f, g; c vs. e, f, g; d vs. e, f, g
			Brush + Biopsy		55 ^e	90	95	36		
			Biopsy + FNA		53 ^f	90	95	35		
			Brush+Biopsy+FNA		62 ^g	90	96	39		
			Brush cytology		26 ^a	100	100	27		
			FNA cytology		25 ^b	100	100	27		All atypia on cytology considered negative. P<0.05 for: a vs. c, e, f, g; b vs. c, e, f, g; c vs. e, d, f; d vs. e, f, g.
			Forceps biopsy		37 ^c	100	100	31		
			Brush + FNA		34 ^d	100	100	30		
			Brush + Biopsy		48 ^e	100	100	35		
			Biopsy + FNA		46 ^f	100	100	34		
			Brush+Biopsy+FNA		52 ^g	100	100	37		

¹⁰ Geenan brush system (Wilson-Cook Medical, Inc. Winston-Salem, N.C.)

¹¹ Howell needle system (Wilson-Cook)

¹² Malleable forceps (Olympus America, Inc., Melville, N.Y.)

¹³ Standard colonoscopic pinch forceps (Ballard Medical Products, Draper, Utah)

In sum, the available evidence suggests that forceps biopsy provides similar, or higher, sensitivity compared to brush cytology, and both tests used in combination may slightly increase sensitivity over that achieved with either technique alone.

Combination of Three Sampling Techniques

Jaiwala, Fogel, Sherman et al. (2000; n=133) also reports on the combination of brush cytology, FNA cytology, and forceps biopsy (Table 28). This study reports increases in overall sensitivity for detecting pancreaticobiliary malignancy as more sampling techniques are added together. The size of incremental the gains in sensitivity and statistically significance associated with adding the third sampling technique vary depending on the criteria used to interpret positive results on cytology. The largest gains are observed when forceps biopsy is being added as the third procedure (approximately 18–23% higher sensitivity, $p<0.05$), but smaller gains are still noted when one of the cytology techniques is added as the third procedure (approximately 4–13%).

Comparison of ERCP-FNA with EUS-FNA

In the absence of comparative studies directly comparing EUS-FNA and ERCP-FNA, an indirect comparison of single arm studies was attempted. Ten articles were identified, including one large multicenter report (Wiersema, Vilmann, Giovannini et al., 1997), three reports from Indiana University (Gress, Gottlieb, Sherman et al., 2001; Gress, Hawes, Savides et al., 1997; Wiersema, Kochman, Cramer et al., 1994), one report from Massachusetts General Hospital (Brandwein, Farrell, Centano et al., 2001), two reports from University of South Carolina (Williams, Sahai, Aabakken et al., 1999; Bhutani, Hawes, Baron et al., 1997), two reports from University of California (Chang, Nguyen, Erickson et al., 1997; Chang, Katz, Durbin et al., 1994), and one report from University of Pennsylvania (Bentz, Kochman, Faigel et al., 1998) (Table 29). Overlap of patient populations and data from separate reports from the same institution is difficult to assess due to limitations in reported detail. An attempt was made to minimize duplicate reporting of subjects. Earlier reports of studies from the same institution that were later published with more subjects have omitted from Table 29. However, some duplication of results likely remains between the multicenter report and separate reports from contributing institutions. The two reports by Gress et al. (Gress, Gottlieb, Sherman et al., 2001 and Gress, Hawes, Savides et al., 1997) address differently selected, but probably overlapping patient groups; however, both are included as they address slightly different questions.

All of these studies reported results separately for diagnosis of pancreatic mass. Additional results on lymph node evaluation and intestinal lesions were not relevant to this review. Despite uncertainties over the exact number of subjects included among the reports detailed in Table 29, the available studies include at least 400 subjects with pancreatic mass and report a range of sensitivity in detecting pancreatic malignancy of 60-94% with a specificity of 100%. Brandwein, Farrell, Centano et al. (2001; n=93) reported results separately for cystic versus solid pancreatic masses and found slightly lower sensitivity for cystic lesions, 50% versus 60%.

Table 29. Supplemental Analysis: Single Arm Studies Reporting Diagnostic Operating Characteristics of EUS-FNA in Pancreatic Mass

Study	N Enr	N Res	Diagnostic test Population setting	Prev (%)	Sens (%)	Spec (%)	PPV (%)	NPV (%)	Adequate Specimens (%)	Comments
Wiersema, Vilmann, Giovannini et al., 1997 Multicenter – Including Indiana University and University of California	124	124	EUS-FNA Subgroup with pancreatic mass	74	89	100	100	76	97	Prospective 4 inadequate specimens excluded. Results in article are unclear regarding 5 cases of suspicious or atypical cytology.
Gress, Gottlieb, Sherman et al., 2001 ¹⁴ Indiana University	102	94	EUS-FNA Suspected pancreatic ca after negative CT-FNA or ERCP cytology	64	88	100	100	92		Prospective 8 inconclusive or nondiagnostic results excluded
Gress, Hawes, Savides et al., 1997 ¹⁴ Indiana University	121	121	EUS-FNA Pancreatic mass	42	80	100	100	88		Prospective
Brandwein, Farrell, Centano et al., 2001 Massachusetts General Hospital	96	93	EUS-FNA Suspected pancreatic ca underwent surgery	85 23 58	60 50 60	100 100 100	100 100 100	29 60 60		Retrospective Solid lesions (n=43) Cystic Lesions (n=26) Dilated duct (n=24)
Williams, Sahai, Aabakken et al., 1999 University of South Carolina	144	144	EUS-FNA All EUS-FNA referrals to single center	85	72 73 70	100 100 100	100 100 100	38 34 45		Retrospective All pancreatic masses Pancreatic mass \geq 3 cm Pancreatic mass $<$ 3 cm
Bentz, Kochman, Faigel et al., 1998 University of Pennsylvania	45	38	EUS-FNA Pancreatic mass	82	94	100	100	78	84	Prospective

¹⁴ Both studies by Gress et al. are reported from the same institution, but patient selection criteria differ with the 2001 report choosing only the subset with persistently high clinical suspicion of pancreatic cancer following otherwise negative workup. The earlier study provides more generally selected patients.

Table 29. Supplemental Analysis: Single Arm Studies Reporting Diagnostic Operating Characteristics of EUS-FNA in Pancreatic Mass (cont'd)

Study	N Enr	N Res	Diagnostic test Population setting	Prev (%)	Sens (%)	Spec (%)	PPV (%)	NPV (%)	Adequate Specimens (%)	Comments
Chang, Nguyen, Erickson et al., 1997 University of California	44 pts 47 les	44	EUS-FNA Pancreatic mass	70	92	100	100	75	95	Retrospective

The sensitivity estimates for ERCP-FNA derived from the two studies identified in the systematic review (Jaiwala, Fogel, Sherman et al., 2000, n=133; Howell, Beveridge, Bosco et al. (1992, n=31) were obtained in subjects with a mixture of pancreaticobiliary malignancy and included subjects with pancreatic cancer, ampullary tumors, cholangiocarcinoma, and metastases. While the reported range of sensitivity of 25-62% for ERCP-FNA appears to be lower than that reported for EUS-FNA, direct comparisons do not seem appropriate due to differences in the case mix of tumors between studies. Further limitations secondary to relatively small numbers of subjects in ERCP-FNA studies and potential differences in cytology techniques and interpretations between studies preclude direct comparison of these estimated ranges of sensitivity.

Summary

There is a modest body of evidence directly comparing the diagnostic performance of nonsurgical tissue sampling techniques for the evaluation of suspected pancreaticobiliary malignancy. 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 versus bile aspiration cytology, similar for FNA cytology versus brush cytology, and similar or higher for forceps biopsy versus brush cytology. Using combinations of two or more sampling techniques may increase the 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 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% with a specificity of 100%. Two studies of ERCP-FNA including 164 subjects with various pancreatobiliary tumors reported of sensitivities ranging from 25% to 62%. While sensitivity in 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.

Part II, Section 2: Diagnostic Performance of ERCP In Pancreaticobiliary Malignant Obstruction—Comparison To Alternatives

Introduction

The evaluation of suspected malignant obstructive jaundice includes imaging evaluation to determine if there is an anatomic narrowing or stricture of the biliary or pancreatic ducts. If a stricture is identified, the appearance and location of the stricture are characterized to determine the likelihood of malignancy and to guide subsequent treatment decisions.

Images of the pancreaticobiliary system can be obtained using a variety of techniques. Direct cholangiopancreatography performed via an ERCP approach is the subject of this systematic review, and the primary diagnostic alternatives to ERCP are magnetic resonance cholangiopancreatography (MRCP), endoscopic ultrasonography (EUS), computed tomography cholangiography (CTC), and percutaneous transhepatic cholangiography (PTC). Both ERCP and PTC are minimally invasive procedures involving injection of contrast directly into the biliary tree. EUS involves endoscopy, but does not directly invade the biliary system. MRCP and CTC are both noninvasive procedures, though oral or intravenous biliary contrast agents may be used to enhance CTC while MRCP does not require the administration of a contrast agent to visualize the biliary tree.

This systematic review selected studies that directly compared the diagnostic performance of ERCP with at least one of the primary alternative diagnostic tests. Given that the expected difference in diagnostic performance between tissue sampling techniques and the diagnostic alternatives reported here are relatively small and the number of cases with the outcome of interest is generally small, these studies may have very limited power to detect statistically significant differences in test performance.

Evidence Base

ERCP vs. MRCP

Eight studies (total n=538) were identified that compared ERCP with MRCP and that used current MRCP technique. Five studies utilized an independent reference standard consisting of best available information derived from surgery, biopsy, imaging, and clinical follow-up to establish the final diagnosis, thus providing comparative data for ERCP and MRCP. The remaining three studies considered ERCP to be the reference standard against which MRCP was measured, yielding concordance of findings of MRCP with ERCP. Four studies were rated “Good” quality, signifying use of blinded interpretation of tests (Table 30). Four of these studies included over 100 subjects and the smallest study contained 46 subjects.

Table 30. Quality Assessment

Study Author, Year	Patient Enrollment	Diagnostic performance of ERCP determined without knowledge of other test results	Diagnostic Performance of other test(s) determined without knowledge of ERCP results	Summary Evaluation
MRCP Studies				
Varghese, Farrell, Courtney et al., 1999	Prospective (n=100) Complete explanation provided of 113 consecutive enrolled and 13 excluded subjects	Yes	Yes	Good
Adamek, Albert, Weitz et al., 1998	Prospective (n=60) 60 of 86 pts w/ suspected biliary obstruction Reasons for exclusions fully explained	Yes	Yes	Good
Arslan, Geitung, Viktil et al., 2000	Retrospective (n=135) 135 of 153 consecutive patients had diagnostic MRCP and ERCP Results reported in 78 patients with diagnostic quality MRCP and ERCP among of 85 patients with obstruction	Uncertain	Uncertain	Fair
Lee, Lee, Kim et al., 1997	? Retrospective (n=46) Complete explanation of 71 consecutive eligible patients and 25 exclusions	Yes	No	Fair MRCP results seem to factor into the reference standard determination
Holzkecht, Gauger, Sackmann et al., 1998	Prospective (n=61) Complete explanation provided of 66 consecutive enrolled patients and 5 excluded subjects	Yes	Yes	Good
Lomas, Bearcroft, and Gimson 1999	Prospective (n=69) Complete explanation provided of 76 enrolled and 7 excluded subjects	Yes	Uncertain	Fair

Table 30. Quality Assessment (cont'd)

Study Author, Year	Patient Enrollment	Diagnostic performance of ERCP determined without knowledge of other test results	Diagnostic Performance of other test(s) determined without knowledge of ERCP results	Summary Evaluation
MRCP Studies (cont'd)				
Adamek, Albert, Breer et al., 2000	Prospective (n=124) 124 of 141 pts w/ suspected pancreatic malignancy Reasons for exclusion fully explained	Yes	Yes	Good
Guibaud, Bret, Reinhold et al., 1995	Prospective (n=126) Some exclusions because of no ERCP confirmation	Uncertain	Yes	Fair
EUS Studies				
Kaneko, Nakao, Inoue et al., 2001	Prospective (n=27) Consecutive patients with no reported exclusions	No	No	Fair
Glasbrenner, Schwarz, Pauls et al., 2000	Prospective (n=95) Consecutive patients referred for surgical resection of pancreatic mass	Yes	Yes	Good
Rosch, Schusdziarra, Born et al., 2000	Retrospective (n=184) Full explanation of 18 exclusions provided but selection based on having all 3 diagnostic tests creates a potential bias.	Yes	Yes	Fair
Cellier, Cuillerier, Palazzo et al., 1998	Retrospective (n=47) Consecutive patients with partial explanations for 17 excluded patients.	Uncertain	Yes	Fair
Burtin. Palazzo, Canard et al., 1997	Prospective (n=68) Consecutive patients enrolled	Yes	Yes	Fair —unorthodox reporting of data, uncertain of data
Dancygier and Nattermann 1994	Prospective (n=41) Unstated whether consecutive	Uncertain	Yes	Fair
Snady, Cooperman, Siegel et al., 1992	Retrospective (n=60) Methods not well described other than pts were “diagnostically problematic”	No	No	Fair

ERCP vs. EUS

Seven studies (total n=466) were identified that compared ERCP with EUS. Six of these employed an independent reference standard consisting of best available information derived from surgery, biopsy, imaging, and clinical follow-up to establish the final diagnosis, and therefore reported data for both EUS and ERCP. Only one study was rated “Good” (Glasbrenner, Schwarz, Pauls et al., 2000, n=90–91) (Table 30). Three studies addressed populations with obstructive jaundice, two studies addressed populations with suspected pancreatic cancer, and two studies addressed patients with either known or suspected intraductal papillary mucinous tumors of the pancreas.

Review of Evidence: Diagnostic Performance

Presence of Malignant Stricture/Lesion

ERCP vs. MRCP. Five studies including a total of 379 patients reported on diagnostic performance of MRCP in identifying and characterizing a malignant stricture (Table 31). In the two studies where ERCP was the reference standard (Guibaud, Bret, Reinhold et al., 1995; n=126; Lomas, Bearcroft, and Gimson 1999, n=69; both rated “Fair”), MRCP showed 86% and 92% sensitivity and 98 and 100% specificity. These data suggest good concordance between MRCP and ERCP results.

The three studies comparing MRCP and ERCP with an independent reference standard report slight differences in estimates of sensitivity and specificity, but none of these differences is statistically significant. The one study rated “Good” quality (Adamek, Albert, Weitz et al., 1998, n=60), reported slightly lower sensitivity (81% vs. 93%) and higher specificity (100% vs. 94%) for MRCP compared with ERCP, but both tests were considered equivalent. The largest study (Arslan, Geitung, Viktil et al., 2000, n=78) found similar sensitivity (86% vs. 89%) and reports lower specificity (82% vs. 94%) for MRCP, but 95% confidence intervals overlap significantly. Finally, Lee et al. (1998; n=46) reports higher sensitivity (81% vs. 71%) and similar specificity (92% vs. 92%) for MRCP, but overall accuracy was not statistically different.

ERCP vs. EUS. Three studies, all rated “Fair” quality and including a total of 129 patients with obstructive jaundice, reported on the diagnostic performance of EUS in identifying the presence of a malignant lesion/stricture (Table 32). One study (Burtin, Palazzo, Canard et al., 1997, n=34) reported similar diagnostic performance for ERCP and EUS, with both tests achieving 89% sensitivity and similar specificity (96% for EUS and 92% for ERCP). Dancygier and Nattermann (1994, n=41) reported complete concordance between EUS and ERCP. One study (Snady, Cooperman, Siegel et al., 1992, n=54–60) compared EUS with the combination of ERCP plus CT and reports both higher sensitivity and specificity for EUS, 85% vs. 75% sensitivity, and 80% vs. 65% specificity, respectively, but these differences were not statistically significant.

In summary, individual studies were relatively small and did not identify significant differences in diagnostic performance between ERCP and either MRCP or EUS. These data permit

Table 31. Comparison of MRCP and ERCP

Study	N Pt	N Res	Diag test	Outcome	Prev (%)	Sens (%)	Spec (%)	PPV (%)	NPV (%)	Adeq Studies (%)	Comments
Independent Reference Standard¹⁵											
Adamek, Albert, Weitz et al., 1998	86	60	MRCP ERCP	Presence of malignant stricture	45 45	81 93	100 94	100 93	87 94	97 79	Good, prospective p=n.r., but “equivalent”
Arslan, Geitung, Viktil et al., 2000	153	78	MRCP ERCP	Presence of malignant stricture		86 (74-94) 89 (77-96)	82 (67-93) 94 (82-99)			98.7 90	Fair, retrospective Kappa = 0.82
Lee, Lee, Kim et al., 1997 ¹⁶	71	46	MRCP ERCP	Presence of malignant stricture	46 46	81 71	92 92	89 88	85 79	98 n.r.	Fair, ?retrospective McNemar p>0.05
Adamek, Albert, Breer et al., 2000	141	124	MRCP ERCP	Presence of pancreatic cancer	30 30	84 70	97 94	91 84	93 88	n.r. n.r.	Good, prospective McNemar p=0.059
Varghese, Farrell, Courtney et al., 1999 ¹⁷	113	100 98	MRCP ERCP	Presence of stricture	28 28	100 100	100 100	100 100	100 100	97 89	Good, prospective No statistical analysis
	113	100 98	MRCP ERCP	Level of stricture	28 28	100 100	100 100	100 100	100 100	97 89	

¹⁵ Independent reference standards relied on best available information from surgery, biopsy, cytology, imaging, and clinical follow-up.

¹⁶ Reference standard also took into consideration MRCP and ERCP results as well as surgery

¹⁷ MRCP provided additional information over ERCP regarding cause of stricture in one case of 1.5 cm periampullary adenocarcinoma

Table 31. Comparison of MRCP and ERCP (cont'd)

Study	N Pt	N Res	Diag test	Outcome	Prev (%)	Sens (%)	Spec (%)	PPV (%)	NPV (%)	Adeq Studies (%)	Comments
ERCP Reference Standard											
Guibaud, Bret, Reinhold et al., 1995	126	126	MRCP	Presence of malignant stricture	11	86 (67-100)	98 (96-100)	86	97	99	Fair, prospective
Lomas, Bearcroft, and Gimson 1999	76	69	MRCP	Presence of malignant stricture	17	92	100	100	98	97	Fair, prospective Kappa = 0.88
	76	69		Presence of stricture	29	100	98 (94-100)	95 (85-100)	100	97	
	76	69		Level of stricture	n.r.	100	100	100	100		
Holzknrecht, Gauger, Sackmann et al., 1998	66	61	MRCP ¹⁸	Presence of stricture	59	89	84	89	84		Good, prospective No statistical analysis

¹⁸ This study performed MRCP using only “snapshot” techniques (RARE and half-Fourier RARE) in the coronal and angles sagittal planes. It is unclear whether axial images were routinely obtained.

Table 32. Comparison of EUS and ERCP

Study	N Pt	N Res	Diag test	Outcome	Prev (%)	Sens (%)	Spec (%)	PPV (%)	NPV (%)	Adeq Stud (%)	Comments
<i>Population with obstructive jaundice</i>											
Independent Reference Standard											
Burtin. Palazzo, Canard et al., 1997	34	34	EUS ERCP	Presence of malignant lesion	36 36	89 89	96 92	89 80	96 96	97 97	Fair, prospective data not clearly reported p=n.s., diagnostic accuracy
Snady, Cooperman, Siegel et al., 1992	60	60 54	EUS ERCP+CT	Presence of malignant lesion	67 67	85 75	80 65	89 81	73 57		Fair, retrospective p=n.s.
ERCP Reference Standard											
Dancygier and Nattermann 1994	41	41	EUS	Presence of malignant lesion	100	100	100	100	100		Fair, prospective No statistical analysis
	41	41	EUS	Level of stricture	100	100	100	100	100		
<i>Population with suspected pancreatic disease</i>											
Independent Reference Standard											
Glasbrenner, Schwarz, Pauls et al., 2000	95	90 91 90	EUS ERCP Combo	Presence of pancreatic cancer	54 53 53	78 81 92	93 88 86	93 89 88	78 80 90		Good, prospective p=n.s. for all comparisons
Rosch, Schusdziarra, Born et al., 2000	184	184 184	EUS ERCP Clinical	Presence of pancreatic cancer vs. chronic pancreatitis	42	86 81 81	87 85 85				Fair, retrospective p=n.s. p=n.s.
	184	184 184	EUS ERCP Clinical	Presence of pancreatic cancer vs. inflammatory tumor	42	86 81 81	72 61 72				
<i>Population with IPMT</i>											
Independent Reference Standard¹⁹											
Kaneko, Nakao, Inoue et al., 2001	27	27 27	EUS ERP	Presence of mural nodules ²⁰	81 81	59 50	100 100	100 100	36 31		Fair, prospective p=n.s.
Cellier, Cuillerier, Palazzo et al., 1998	47	21 29	EUS ERCP	Presence of invasive tumor ²¹	43 31	78 55	75 90	70 71	82 82		Fair, retrospective No statistical analysis

¹⁹ Reference standard consists of surgical specimen histology and/or pancreatography

²⁰ Population of patients with suspected intraductal papillary mucinous tumors of the pancreas

²¹ population of patients with histologically proven diagnosis of intraductal papillary mucinous tumors of the pancreas

preliminary conclusions that MRCP and EUS provide similar diagnostic assessment as ERCP for detection of malignant pancreaticobiliary obstruction.

Diagnosis of Pancreatic Cancer

MRCP vs. ERCP. Diagnostic performance for demonstrating pancreatic cancer in 37 of 124 was reported by Adamek, Albert, Breer et al. (2000; Table 31). This study compares MRCP and ERCP and reported slightly higher sensitivity (84% vs. 70%) and similar specificity (97% vs. 94%) for MRCP and ERCP, respectively, but these differences did not reach statistical significance (McNemar $p=0.059$). This study was rated “Good” for quality.

EUS vs. ERCP. Diagnostic performance for pancreatic cancer was reported in two studies specifically addressing populations with suspected pancreatic disease (Table 32). Rosch, Schusdziarra, Born et al. (2000) retrospectively evaluated 184 patients who had ERCP, EUS, and CT and compared the diagnostic performance of clinical assessment with the various imaging tests. This study finds similar performance for clinical assessment, ERCP, or EUS in distinguishing pancreatic cancer from chronic pancreatitis and in distinguishing pancreatic cancer from inflammatory tumor. Interpretation of Rosch, Schusdziarra, Born et al. (2000) is somewhat limited by the retrospective selection of patients on the basis of having all three imaging tests, which might bias the study toward cases where findings were inconclusive. Glasbrenner, Schwarz, Pauls et al. (2000; $n=95$) noted ERCP and EUS to have similar sensitivity (81% vs. 78%, respectively) and specificity (88% vs. 93%, respectively), and the combination of the two tests yielded 92% sensitivity and 86% specificity, but these differences were not statistically significant.

Summary. In summary, there is little evidence directly comparing ERCP with either MRCP or EUS in diagnosing pancreatic cancer. The available evidence does not demonstrate statistically significant differences between ERCP and either MRCP or EUS.

Presence of Stricture

ERCP vs. MRCP. Three studies reported diagnostic performance in demonstrating the presence of stricture (either benign or malignant) (Table 31). One of the two studies rated as “Good” independently verified results and found 100% sensitivity and 100% specificity for both MRCP and ERCP (Varghese, Farrell, Courtney et al., 1999, $n=98-100$). The other (Holzknecht, Gauger, Sackmann et al., 1998, $n=61$) used ERCP as reference standard and reported 89% sensitivity and 85% specificity for MRCP relative to ERCP, though this study utilized only projection (“snapshot”) MRCP techniques without additional multislice techniques which may limit its comparability. One additional study (Lomas, Bearcroft, and Gimson 1999, $n=69$) rated as “Fair” quality because of uncertainties with regard to complete blinding of interpretation, noted 100% concordance for MRCP with ERCP.

ERCP vs. EUS. No studies reported this specific analysis.

Summary. In summary, the evidence specifically evaluating MRCP in relation to ERCP for detecting strictures is sparse and suggests similar results for MRCP and ERCP in identifying the

presence of a stricture. However, these studies do not report full statistical analysis. The relative performance of EUS and ERCP in this setting has not been reported.

Level of Stricture

ERCP vs. MRCP. One study comparing ERCP and MRCP (Varghese, Farrell, Courtney et al., 1999, n=98-100, “Good”) specifically reported 100% sensitivity and specificity for both MRCP and ERCP in defining the level of the stricture (Table 31). Lomas, Bearcroft, and Gimson (1999, n=69, “Fair”) also reported complete concordance for MRCP with ERCP in defining the level of malignant strictures.

ERCP vs. EUS. Only one study comparing ERCP and EUS (Dancygier and Nattermann 1994, n=41, “Fair”) specifically reported sensitivity and specificity in defining the level of the stricture (Table 32). This study reports 100% sensitivity and specificity for both ERCP and EUS.

Summary. In summary, there is little evidence specifically reporting the diagnostic accuracy of MRCP or EUS relative to ERCP in defining the level of stricture, but the available studies suggest that all three tests provide highly accurate localization of pancreaticobiliary stricture.

Evaluation of Suspected Intraductal Papillary Mucinous Tumors (IPMT) of the Pancreas

ERCP vs. MRCP. No studies reported this specific analysis

ERCP vs. EUS. Two studies evaluated EUS in comparison with endoscopic retrograde pancreatography (ERP) in patients with either known or suspected IPMT of the pancreas (Table 32). Kaneko, Nakao, Inoue et al. (2001; n=27, “Fair”) found that EUS and ERP were similarly sensitive (59% vs. 50%, respectively) in detecting mural nodules while both tests were 100% specific for this finding. Cellier, Cuillerier, Palazzo et al. (1998; n=47, “Fair”) compared ERCP and EUS in defining the presence of invasive tumor and reported EUS to be more sensitive (78% vs. 55%) and less specific (75% vs. 90%), but no statistical analysis was reported.

These two small studies, reporting estimates of diagnostic performance relating to different diagnostic endpoints, suggest that EUS may provide a similar information to ERCP in patients with known or suspected intraductal papillary mucinous tumors of the pancreas, but confirmation of these findings would be helpful.

Conclusions

The body of evidence directly comparing ERCP with either MRCP or EUS is modest in size and of varying methodological quality. The evidence comparing ERCP with MRCP is slightly stronger than that comparing ERCP with EUS both in terms of number of subjects and study quality. The available studies do not demonstrate statistically significant differences in diagnostic performance for ERCP versus MRCP or for ERCP versus EUS for characterizing malignant strictures. In sum, the available studies suggest that either MRCP or EUS provides similar diagnostic performance as ERCP in detecting pancreaticobiliary malignant obstruction.

Part II, Section 3: Outcomes of Treatment Using ERCP and Endoscopic Sphincterotomy and Endoscopic Stent for Palliation of Pancreaticobiliary Malignancy—Comparison of Strategies Using ERCP, Surgery, or Interventional Radiology

Introduction

Biliary obstruction is a frequent presenting feature of pancreaticobiliary malignancy. Unfortunately, patients with pancreaticobiliary malignancy are usually incurable at the time of diagnosis (Conio, Demarquay, De Luca et al., 2001; England and Martin 1996). Whether surgical resection for attempted cure is feasible or not, management of biliary obstruction is desirable to palliate the morbidity of jaundice. Endoscopic stent drainage has been proposed as an alternative to biliary-enteric bypass surgery to palliate malignant biliary obstruction. In addition, alternative approaches to biliary stenting have been compared with particular interest to determining optimal stent material, design, and placement strategies.

Part II, Section 3A. Comparison of ERCP Stent Versus Surgical Bypass

Body of Evidence

Five studies compared results of surgical bypass with endoscopic stent drainage for palliation of malignant obstructive jaundice. Quality assessments are described in Table 33. Results of these studies are detailed in the “Evidence Tables” section and summarized in Tables 34–37. Three randomized, controlled trials were identified comparing surgical biliary bypass with endoscopic biliary stent placement. Two of these (Smith, Dowsett, Russell et al., 1994, n=204; Andersen, Sorensen, Kruse et al., 1989, n=50) were rated as “Good” quality, and Shepherd, Royal, Ross et al. (1988, n=52) was rated as “Fair”). Two retrospective comparisons (Raikar, Melin, Ress et al., 1996, n=66; Leung, Emery, Cotton et al., 1983, n=98) were both rated as “Poor.”

Review of Evidence: Treatment Outcomes

All studies reported that there was no significant difference in overall patient survival between the ERCP and the surgery groups (Table 35). Two randomized controlled trials reported both treatments to have high rates for relief of jaundice but no statistically significant difference. A third study reported on quality of life, as measured by mean percentage of survival time with normal activity or limited activity with no aid; there were no significant differences.

Review of Evidence: Adverse Outcomes

There were no significant differences in perioperative mortality (Table 36). The randomized controlled trial by Smith, Dowsett, Russell et al. (1994) was designed to show a 5–20% decrease in 30-day mortality at 95% power with 115 patients entered into each arm. Accrual was stopped at 204 patients when interim analysis indicated that additional accrual would not change the outcome. While this trial did not show a statistically significant difference in perioperative (30-

Table 33. Quality Assessment

Study Author, Year	Comparable Initial Groups?	Comparable Groups Maintained?	Comparable Performance of Intervention?	Comparable Measurement of Outcomes?	Appropriate Analysis	Summary Evaluation
Smith, Dowsett, Russell et al., 1994	RCT (n=204) Good comparability – Randomization by computer minimization on age, bilirubin, albumin, urea, and Hb conc. – Patient characteristics not significantly different	<u>Surgery:</u> (n=103) 2 excluded due to benign disease 7 did not get surgery (2 technical failures, 1 elected crossover, 3 deteriorated clinically and got stents, 1 deteriorated and got no further rx) <u>Stent:</u> (n=101) 1 excluded due to benign disease 5 did not get stents (1 elected crossover, 3 technical failures got surgery, 1 technical failure got no further rx)	Adequate for comparison	Adequate outcome measures used. Outcomes were not assessed blindly.	Intention-to-treat analysis used	Good
Andersen, Sorensen, Kruse et al., 1989	RCT (n=50) Good comparability – Sealed envelopes – Patient characteristics not significantly different	<u>Surgery:</u> n=25 6 did not undergo surgery (2 wanted crossed over, 1 found inoperable at surgery, 2 psychological compromise, 1 surgeon not available) <u>Endoprosthesis:</u> n=25 None	Adequate for comparison	Adequate outcome measures used. Outcomes were not assessed blindly.	Intention-to-treat analysis used Results also analyzed by treatment received and findings were consistent.	Good

Table 33. Quality Assessment (cont'd)

Study Author, Year	Comparable Initial Groups?	Comparable Groups Maintained?	Comparable Performance of Intervention?	Comparable Measurement of Outcomes?	Appropriate Analysis	Summary Evaluation
Shepherd, Royal, Ross et al., 1988	RCT (n=52) Fair comparability – Randomization method not specified – Patient characteristics mostly comparable	<u>Surgical</u> : n=27 4 total: 2 withdrawn (1 died pre-op and 1 had attempted curative surgery). 2 technical failures crossed over to endoprosthesis. <u>Endoprosthesis</u> : n=25 6 total: 1 had benign biopsies but later found to have cancer at surgery; 4 failed and crossed-over to surgery; 1 failed both stent and surgery	Adequate for comparison	Adequate outcome measures used. Outcomes were not assessed blindly.	Does not clearly state method of analysis	Fair

Table 33. Quality Assessment (cont'd)

Study Author, Year	Comparable Initial Groups?	Comparable Groups Maintained?	Comparable Performance of Intervention?	Comparable Measurement of Outcomes?	Appropriate Analysis	Summary Evaluation
Raikar, Melin, Ress et al., 1996	Retrospective series (n=66) Fair to Poor comparability Baseline patient characteristics show no SSD but differences in performance status distribution noted with ERCP subjects having relatively higher percentages of good and poor PS while surgery had relatively higher midrange PS.	All subjects included in analysis	Adequate for comparison	Adequate outcome measures used. Outcomes were not assessed blindly.	Univariate analysis does not account for important confounders	Poor
Leung, Emergy, Cotton et al., 1983	Retrospective series (n=98) Poor comparability Baseline patient characteristics show differences in age and lesion location.	All subjects included in analysis	Adequate for comparison	Adequate outcome measures used. Outcomes were not assessed blindly.	Univariate analysis does not account for important confounders	Poor

Table 34. Overview of studies and reported outcomes

Study	Population	Procedure	N ERCP Surg (treated)	Outcome Measures Reported									Study Quality
				Total Hospital Days	Initial Hospital Days	Readmissions	Need for Add'l Procedure	Survival	Jaundice Relief	Quality of Life	Perioperative Mortality	Perioperative Morbidity	
Randomized Controlled Trials													
Smith, Dowsett, Russell et al., 1994	Malignant distal CBD obstruction and jaundice Mean age 70	10 Fr stents ²² vs. Bypass Surgery	101 (100) 103 (101)	X			X	X		X	X	X	Good
Andersen, Sorensen, Kruse et al., 1989	Malignant distal CBD obstruction and jaundice Age>60y	7-10 Fr stents vs. Bypass Surgery	25 (19) 25 (30)	X			X	X		X	X	X	Good
Shepherd, Royal, Ross et al., 1988	Malignant distal CBD obstruction Mean age 73	10 Fr stents vs. Bypass Surgery	27 (23) 25	X	X	X	X	X	X		X	X	Fair

²² 19 of 101 stent patients required combined ERCP and percutaneous transhepatic approach to place stent

Table 34. Overview of studies and reported outcomes (cont'd)

Study	Population	Procedure	N ERCP Surg (treated)	Outcome Measures Reported									Study Quality
				Total Hospital Days	Initial Hospital Days	Readmissions	Need for Add'l Procedure	Survival	Jaundice Relief	Quality of Life	Perioperative Mortality	Perioperative Morbidity	
Retrospective Studies													
Raikar, Melin, Ress et al., 1996	Unresectable pancreatic carcinoma	10-12 Fr stents vs. Bypass Surgery	34			X	X	X			X	X	Poor
			32										
Leung, Emergy, Cotton et al., 1983	Malignant obstructive jaundice (CBD location not specific)	8-10 Fr stents vs. Bypass Surgery	64			X	X	X			X		Poor
			34										

Table 35. Treatment Outcomes

Study	Study arm N Enrolled/ (treated or results)	Survival (median) (*mean) (**Life Table Analysis)	P	Relief of Jaundice	p	Quality of Life	p
Randomized Controlled Trials							
Smith, Dowsett, Russell et al., 1994	ERCP ²³ 101 (100)	21 weeks	ns	97%	ns		
	Surgery 103 (101)	26 weeks		98%			
Andersen, Sorensen, Kruse et al., 1989	ERCP 25 (19)	**84 days (3-498) ²⁴	ns			57% survival time mean normal activity or limited, no aid	ns
	Surgery 25 (30)	**100 days (10-642)				51% survival time mean normal activity or limited, no aid	
Shepherd, Royal, Ross et al., 1988	ERCP 27 (23)	**152 days (39-411)	ns	91%	nr		
	Surgery 25	**125 days (52-354)		92%			

²³ Stent placement was attempted first with ERCP approach. In 19 patients a combined transhepatic-endoscopic approach was required when initial ERCP failed.

²⁴ No significant difference when analyzed by treatment received.

Table 35. Treatment Outcomes (cont'd)

Study	Study arm N Enrolled/ (treated or results)	Survival (median) (*mean) (**Life Table Analysis)	P	Relief of Jaundice	p	Quality of Life	p
Retrospective Studies							
Raikar, Melin, Ress et al., 1996	ERCP 34	*9.7 months (10d-35)	0.13				
	Surgery 32	*7.3 month (7d-29)					
Leung, Emery, Cotton et al., 1983	ERCP 64	6 mos. approximate	Ns				
	Surgery 34	6 mos. approximate					

Table 36. Adverse Outcomes

Study	Study arm N Enrolled/ (treated or results)	Perioperative Mortality	P	Perioperative Complications	p
Randomized Controlled Trials					
Andersen, Sorensen, Kruse et al., 1989	ERCP 25 (19)	5 (20%)	Nr	36% (total severe infection)	Ns
	Surgery 25 (30)	6 (24%)		20% (total severe infection)	
	Shepherd, Royal, Ross et al., 1988	ERCP 27 (23)		2 (9)%	
Surgery 25	5 (20%)	14 procedure-related complication events			
Smith, Dowsett, Russell et al., 1994	ERCP ²⁵ 101 (100)	8% ²⁶	Ns	11% major complications	0.02
	Surgery 103 (101)2 (n)	15%		29% major complications	
Retrospective Studies					
Leung, Emergy, Cotton et al., 1983	ERCP 64	1 (3%)	Nr	21%	ns
	Surgery 34	1 (4%)		33%	
Raikar, Melin, Ress et al., 1996	ERCP 34	10 (16%)	Nr		
	Surgery 32	3 (9%)			

²⁵ Stent placement was attempted first with ERCP approach. In 19 patients a combined transhepatic-endoscopic approach was required when initial ERCP failed.

²⁶ Procedure related mortality was significantly higher in the surgery group (14% vs. 3% , p=0.006). Also of note, 3 deaths in the surgical group were in patients who did not undergo surgery.

Table 37. Resource Utilization Outcomes

Study	Study arm N Enrolled/ (Treated or Results)	Total Hospital Days median ²⁷ (range)	p	Initial Hospital Days (median) (*mean)	p	Readmission to Hospital N (%)	p	Need for Additional Procedure	p
Randomized Controlled Trials									
Smith, Dowsett, Russell et al., 1994	ERCP ²⁸ 101 (100)	19 (4-59)	ns					Recurrent obstructive jaundice requiring stent replacement in 36 (36%)	ns
	Surgery 103 (101)	26 (8-85)						Late gastric outlet obstruction requiring gastric bypass in 10 (10%) Recurrent obstructive jaundice in 2 (2%). One required stent. Late gastric outlet obstruction requiring gastric bypass in 5 (5%)	
Andersen, Sorensen, Kruse et al., 1989	ERCP 25 (19)	26 (3-210)	ns ²⁹					1 (4%) early failure requiring surgical bypass.	nr
	Surgery 25 (30)	27 (10-202)						3 (12%) early failure requiring stent placement.	
Shepherd, Royal, Ross et al.,1988	ERCP 27 (23)	8 ³⁰ (2-30)	<0.01	5 (2-16)	<0.002	10 (43%)	nr	Gastric outlet obstruction developed in 2 (9%)	nr
	Surgery 25	13 (8-49)		13 (8-49)		3 (12%)		Gastric outlet obstruction developed in 1 (4%)	

²⁷ Results generally reported as median. Results reported as mean are demarcated by an asterisk (*)

²⁸ Stent placement was attempted first with ERCP approach. In 19 patients a combined transhepatic-endoscopic approach was required when initial ERCP failed.

²⁹ Comparison of hospital stay was not statistically significant when analyzed by treatment received.

³⁰ Calculated only in patients who were alive 30 days post-op.

Table 37. Resource Utilization Outcomes (cont'd)

Study	Study arm N Enrolled/ (Treated or Results)	Total Hospital Days median ³¹ (range)	p	Initial Hospital Days (median) (*mean)	p	Readmission to Hospital N (%)	p	Need for Additional Procedure	p
Retrospective Studies									
Raikar, Melin, Ress et al., 1996	ERCP 34	\$17,738	.05	7*	<0.001	12 (35%)	nr	Average of 1.7 stent replacements per patient	nr
	Surgery 32	\$25,101		14*		8 (25%)		One patient developed gastric outlet obstruction requiring surgical gastric bypass. Two patients required stent placement for recurrent jaundice. No report of surgical patients developing gastric outlet obstruction.	nr
Leung, Emergy, Cotton et al., 1983	ERCP 64			14* (4-30)	Nr	8 (13%) ³²	nr	Recurrent jaundice developed in 3 (5%)	nr
	Surgery 34			30* (14-79)		3 (9%)		Gastric outlet obstruction developed in 2 (3%) Recurrent jaundice developed in 1 (3%) Gastric outlet obstruction developed in 2 (6%)	nr

³¹ Results generally reported as median. Results reported as mean are demarcated by an asterisk (*)

³² Local complications included cholangitis, recurrent jaundice, duodenal obstruction, or chest wall metastasis

day) mortality, intent-to-treat analysis showed significantly greater procedure-related mortality in the surgery arm (14% vs. 3%, $p=0.006$). Smith, Dowsett, Russell et al., (1994) also found that major complications were significantly greater in the surgery group than in the ERCP group (29% vs. 11%, $p=0.02$). Andersen, Sorensen, Kruse et al. (1989) reported severe infections in 36% of ERCP patients compared to 20% of surgical patients, but the difference was not statistically significant. Shepherd, Royal, Ross et al. (1988) found twice the rate of complications in the surgical group, but again this was not statistically significant.

Review of Evidence: Resource Utilization

The two randomized controlled trials rated as good quality found no significant difference in total days of hospitalization, including the largest of trials in this group of studies (Smith, Dowsett, Russell et al., 1994, $n=203$) (Table 37). Three studies report on initial hospitalization; including 1 randomized controlled trial (Shepherd, Royal, Ross et al., 1988, $n=52$). All show fewer days of initial hospitalization with ERCP, and 2 report that the difference is statistically significant. Readmissions were more common with ERCP, but tests of statistical significance were not reported. The randomized controlled trial by Shepherd, Royal, Ross et al. (1988) reports significantly fewer initial and total hospitalization days with ERCP, despite a readmission rate twice that of surgery. However, this randomized controlled trial was judged of lesser quality (“fair”), largely due to lack of clarity in the method of analysis.

Stent replacement was reported in the Smith, Dowsett, Russell et al., (1994) study as necessary in 37% of patients, all but 1 case due to recurrence of obstructive jaundice. Raikar, Melin, Ress et al. (1996) reported an average of 1.7 stent replacements per patient.

Summary

The most robust evidence is provided in the randomized controlled trial by Smith, Dowsett, Russell et al. (1994). 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% vs. 29%, $p=0.02$), presumably reflecting the more invasive nature of surgical versus endoscopic treatment. Stent replacement was required in 37% of ERCP patients.

Part II, Section 3B. Comparison of Metal vs. Plastic Stents During ERCP

Evidence Base

Three studies were identified comparing endoscopically placed metal or plastic stents for palliation of biliary obstruction due to malignancy. Quality ratings are described in Table 38. Results are detailed in the “Evidence Tables” chapter and summarized in Tables 39–42. Two randomized, controlled trials (total n=206) were identified. Davids, Groen, Rauws et al. (1992, n=105, “Fair” quality) compared metal versus plastic stents. Prat, Chapat, Ducot et al. (1998, n=101, “Fair” quality) randomized patients into 3 arms (either metal stents, plastic stents with exchange as needed for stent dysfunction, or plastic stents with routine exchange every 3 months). In addition, Schmassmann, Von Gunten, Knuchel et al. (1996, n=165, “Poor” quality) retrospectively compared results with metal versus plastic stents.

Review of Evidence: Treatment Outcomes

Metal stents showed statistically significantly longer patency rates compared with plastic stents in all three studies (Table 40). Two of the studies reported that median duration of patency with metal stents was twice as long as plastic stents (9.1–10 months versus 4–4.2 months, $p<0.006$), but one of the randomized trials showed a smaller benefit for metal stents (4.8 months versus 3.2 months, $p<0.05$).

The two randomized studies reported no significant difference in overall survival for patients treated with metal or plastic stents, with median survival ranging from 4.5–5.8 months. In contrast, the retrospective study found slightly longer median survival in the metal stent group (6.5 months versus 4 months, $p<0.05$), but related this observation to increased mortality in 18% of subjects (predominantly plastic stent group) who did not receive treatment for stent dysfunction.

All studies reported both treatments to have high rates for relief of jaundice with no statistically significant differences reported.

Review of Evidence: Adverse Outcomes

Two studies (Prat, Chapat, Ducot et al., 1998; Schmassmann, Von Gunten, Knuchel et al., 1996) reported no significant difference in perioperative mortality (Table 41). The randomized, controlled trial by Davids, Groen, Rauws et al. (1992) noted a higher perioperative mortality rate in the metal stent group (14% vs. 4%, $p=0.047$), but the causes of death in 6 of 7 cases were completely unrelated to biliary pathology. No significant differences were noted in complications in the two randomized studies and the retrospective study did not specifically report complications other than perioperative mortality.

Table 38. Study Quality Assessment

Study Author, Year	Comparable Initial Groups?	Comparable Groups Maintained?	Comparable Performance of Intervention?	Comparable Measurement of Outcomes?	Appropriate Analysis	Summary Evaluation
Dauids, Groen, Rauws et al., 1992	<p>RCT (n=105)</p> <p>Good comparability</p> <ul style="list-style-type: none"> - Randomization by computer generated random number - patient characteristics well-balanced 	<p>115 initially randomized and 105 included in analysis</p> <p>10 patients excluded. 5 due to prior history of malignancy in past 10 years and 5 due to selection for surgical therapy.</p> <p>None lost to follow-up</p>	Adequate for comparison.	<p>Adequate outcome measures used.</p> <p>Outcomes were not assessed blindly.</p>	Method of analysis not clearly stated.	Fair
Prat, Chapat, Ducot et al., 1998	<p>RCT (n=101)</p> <p>Good comparability</p> <ul style="list-style-type: none"> - Randomization by blocks of six and stratified for gender and investigation center - patient characteristics well-balanced 	<p>4 of 105 excluded</p> <p>Three for failed endoprosthesis insertion and one for not complying with required quarterly stent changes for group 2</p> <p>Four lost to follow-up (3 moved away and 1 no follow-up information)</p>	Adequate for comparison.	<p>Adequate outcome measures used.</p> <p>Outcomes were not assessed blindly.</p>	Method of analysis not clearly stated	Fair

Table 38. Study Quality Assessment (cont'd)

Study Author, Year	Comparable Initial Groups?	Comparable Groups Maintained?	Comparable Performance of Intervention?	Comparable Measurement of Outcomes?	Appropriate Analysis	Summary Evaluation
Schmassmann, Von Gunten, Knuchel et al., 1996	Retrospective study (n=165) Fair comparability Baseline patient characteristics similar for age, gender, bilirubin, type of tumor and stage, location of stricture, or associated procedures	All subjects included in analysis	Adequate for comparison 87% of metal stent and 100% of plastic stent patients had sphincterotomy	Adequate outcome measures used. Outcomes were not assessed blindly.	Univariate analysis does not account for confounders	Poor

Table 39. Overview of studies and reported outcomes

Study	Population	Procedure	N (treated)	Outcome Measures Reported									STUDY QUALITY	
				Metal	Plastic	Total Hospital Days	Initial Hospital Days	Cost Utilization	Need for Add'l Procedure	Survival	Jaundice Relief	Stent Patency		Periop Mortality
Randomized Controlled Trials														
Davids, Groen, Rauws et al., 1992	Patients with irresectable distal bile-duct malignancy Pancreatic ca = 93 Papillary ca = 12	Metal stent ³³	49				X	X	X	X	X	X	Fair	
		Straight 10 Fr polyethylene stent ³⁴	56											
Prat, Chapat, Ducot et al., 1998	Patients with malignant CBD strictures Not involving hilum Pancreatic ca = 65 Cholangioca = 21 Ampullary ca = 3 Metastatic = 12	Metal stent	34	X		X	X	X	X		X	X	Fair	
		Polyethylene 11.5 Fr stent ³⁵ w/ routine exchange	33											
		Polyethylene 11.5 Fr stent w/ as needed exchange	34											
Retrospective Studies														
Schmassmann, Von Gunten, Knuchel et al., 1996	Consecutive patients with unresectable malignant biliary obstruction	Metal stent	95				X	X	X	X	X		Poor	
		Straight 12 Fr or 10 Fr polyethylene stent ³⁶	70											

³³ Metal stents were of the Wallstent type (Schneider, Switzerland (Davids et al.; Schmassmann et al.)) or (Schneider-Howmedical, Lyons, France (Prat et al.)).

³⁴ Polyethylene stents were made by PBN Medicals (Stenlose, Denmark)

³⁵ Polyethylene stents were made by Wilson-Cook (Winston-Salem, N.C.)

³⁶ Polyethylene stents 12 Fr were made by Olympus (Volketswil, Switzerland) and 10 Fr Huibregtse (Cook, Nottwil, Switzerland)

Table 40. Treatment Outcomes

Study	Study arm N Enrolled/ (treated or results)	Survival (median)	P	Relief of Jaundice N (%)	p	First Stent Patency (median)	p
Randomized Controlled Trials							
Davids, Groen, Rauws et al., 1992	Metal 49	5.8 months ³⁷	0.45	47/49 (96%)	n.r.	9.1 months	0.006
	Plastic 56	4.9 months		53/56 (95%)		4.2 months	
Prat, Chapat, Ducot et al., 1998	Metal 34	4.5 months	n.s.	48h Decrease in bilirubin: 41%	n.s.	4.8 months	<0.05
	Plastic-routine 33	5.6 months		34.3%		Not reported separately	
	Plastic-as needed 34	4.8 months		35.4%		3.2 months	
Retrospective Studies							
Schmassmann, Von Gunten, Knuchel et al. 1996	Metal 95	6.5 months ³⁸	<0.05	95%	n.s.	10 months ³⁹	<0.001
	Plastic 70	4 months		88%		4 months	

³⁷ Data were converted to months from reported days by dividing by 30.

³⁸ When 29 subjects (8 metal stent, 21 plastic stent) who died related to untreated stent dysfunction were excluded from the analysis, the remaining 136 subjects had similar survival between the two groups.

³⁹ Subgroup analysis did not show any significant difference between different locations (common bile duct vs. hilar or intrahepatic stricture) but numbers were small in the hilar and intrahepatic subgroups.

Table 41. Adverse Outcomes

Study	Study arm N Enrolled/ (treated or results)	Perioperative Mortality	P	Complications	p
Randomized Controlled Trials					
Davids, Groen, Rauws et al., 1992	Metal 49	7 (14%) ⁴⁰	0.047	6 (12%) ⁴¹	n.r.
	Plastic 56	2 (4%) ⁴²		6 (11%)	
Prat, Chapat, Ducot et al., 1998	Metal 34	Overall rate was 3.9%		Overall rate was 11.9%	
	Plastic-routine 33	No significant difference between groups		No significant difference between groups	
	Plastic-as needed 34				
Retrospective Studies					
Schmassmann, Von Gunten, Knuchel et al. 1996	Metal 95	2%	n.s.		
	Plastic 70	3%			

⁴⁰ Causes of death were sepsis after recurrent cholangitis (1); cardiac failure (2); cachexia (4).

⁴¹ Complications in Davids et al. were measured in 7 days after procedure.

⁴² Causes of death were cachexia (2).

Table 42. Resource Utilization Outcomes

Study	Study arm N Enrolled/ (Treated or Results)	Total Hospital Days median (range)	p	Resource Utilization Costs	p	Need for Additional Procedure	p
Randomized Controlled Trials							
Davids, Groen, Rauws et al., 1992	Metal 49					1.3 per person	n.r.
	Plastic 56					1.8 per person	
Prat, Chapat, Ducot et al., 1998	Metal 34	5.5 ± 1.4*	*0.01 others n.s.	Mean costs (95% CI) \$4643 (4207-5079)	n.r.	1.2 ± 0.4 per patient	0.01 ANOV A
	Plastic-routine 33	10.6 ± 1.7*		\$6770 (5394-8146)		2.5 ± 1.9 per patient	
	Plastic-as needed 34	7.4 ± 1.5		\$5547 (4082-7013)		1.7 ± 1.3 per patient	
Retrospective Studies							
Schmassmann, Von Gunten, Knuchel et al., 1996	Metal 95					1.2 per patient	<0.005
	Plastic 70					1.58 per patient	

Review of Evidence: Resource Utilization Outcomes

All studies examined the relative utilization of ERCP procedures and found patients receiving metal stents to require the fewest ERCP procedures (Table 42). Patients receiving metal stents required 1.2–1.3 ERCP procedures on average and those receiving plastic stents and undergoing stent exchange only when needed required 1.58–1.8 ERCP procedures. The study by Prat, Chapat, Ducot et al. (1998) examined the strategy of routine plastic stent exchange every 3 months which necessitated an average of 2.5 ERCP procedures per patient. The differences in ERCP utilization between metal and plastic stents were reported to be statistically significant in two studies and a statistical comparison was not reported in the third study.

Prat, Chapat, Ducot et al. (1998) also examined utilization of total hospital days and found the metal stent group averaged 5.5 days while the plastic stent groups required 7.4 to 10.6 days on average, depending on whether “as needed” or routine stent exchange was used, respectively. The difference between metal stents and routinely exchanged plastic stents was statistically significant (5.5 ± 1.4 versus 10.6 ± 1.7 , $p=0.01$) while the differences between metal stents and plastic stents exchanged as needed were not statistically significant.

Prat, Chapat, Ducot et al. (1998) also reported lower average total costs for the metal stent group than costs associated with either of the plastic stent strategies, but statistical analysis was not reported for these results.

Summary

Three studies including a total of 371 subjects provide consistent evidence that metal stents remain patent longer than plastic stents. 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.

Part II, Section 3C. Additional Comparisons of ERCP Strategies

Evidence Base

The ERCP literature systematically reviewed for this report also included nine studies comparing various alternative ERCP treatment techniques. The comparisons reported in these studies were sufficiently dissimilar from the studies reviewed in preceding sections on palliative treatments of pancreaticobiliary malignancy that they are briefly summarized separately in this section. The quality assessments of these studies are detailed in Table 43 and the results of these studies are in Tables 44–46.

Review of Evidence: Stent Material and Design

Four studies, including two randomized controlled trials (one quality rated as “Good” and one as “Fair”) and two nonrandomized studies (both rated “Poor” quality) compared different features of endoscopically placed stents for palliation of pancreaticobiliary malignancy (Tables 44–46.).

van Berkel, Boland, Redekop et al. (1998, n=84, “Fair”) randomized patients to receive stents made of Teflon™ versus stents made of polyethylene and found no significant differences in efficacy or complications (Table 44). Median stent patency duration was 83 days for Teflon™ stents and 80 days for polyethylene stents (p=0.93).

Pedersen (1993, n=89, “Poor”) and Speer, Cotton, MacRae et al. (1988, n=79, “Poor”) both compared outcomes using different caliber stents, but neither of these studies uses a randomized, controlled design (Table 45). Speer, Cotton, MacRae et al. (1988) found significantly longer median stent patency for 10Fr stents compared with 8Fr stents (32 weeks vs. 12 weeks, p<0.001). Complications reported included a lower rate of cholangitis with 10 Fr stents (5% vs. 34%, p<0.05), and similar rates of local perforation and stent migration. However, the 8Fr stents had pigtail-shaped ends compared with straight-shaped 10Fr catheters, a potential confounding factor in interpreting this study. Pedersen (1993) did not reveal a statistically significant difference in stent patency comparing 10Fr and 7 Fr, and did not show significant differences in total complication rates. However, this study also suffered from baseline differences in age, with younger patients receiving 7 Fr stents, increasing concerns over interpretation of findings.

Sung, Chung, Tsui et al. (1994, n=70, “Good”) randomized patients to receive 10Fr stents with or without sideholes (Table 46). No statistically significant differences were noted in stent patency and reported complications appeared similar, although statistical analysis was not reported.

None of these studies provides a sufficient basis for a conclusion regarding the relative efficacy the stent features being compared.

Table 43. Quality Assessment

Study Author, Year Record Number	Comparable Initial Groups?	Comparable Groups Maintained?	Comparable Performance of Intervention?	Comparable Measurement of Outcomes?	Appropriate Analysis	Summary Evaluation
van Berkel, Boland, Redekop et al., 1998	<p>RCT (n=84)</p> <p>Good comparability</p> <ul style="list-style-type: none"> - Randomization by computer generated numbers in sealed envelopes - Patient characteristics similar 	<p>97 consecutive patients enrolled.</p> <p>13 excluded for protocol violations (11 had surgical resection, 1 had PTH drainage, 1 refused treatment). Details about which treatment arm patients were assigned to were not provided.</p> <p>None lost to follow-up.</p>	Adequate for comparison.	<p>Adequate outcome measures used.</p> <p>Outcomes were not assessed blindly.</p>	Method of analysis not stated but all 84 included in analysis.	Fair
Pedersen 1993	<p>Prospective study (n=89)</p> <p>Fair comparability</p> <p>Differences in age noted with younger 7Fr group. No SSD in stenosis location, gender, or type of cancer.</p>	All subjects included in analysis	<p>Adequate for comparison.</p> <p>Adjunctive sphincterotomy was performed equally in 7Fr and 10Fr groups.</p>	<p>Adequate outcome measures used.</p> <p>Outcomes were not assessed blindly.</p>	Univariate analysis does not account for important confounders	Poor
Speer, Cotton, MacRae et al., 1988	<p>Retrospective study (n=79)</p> <p>Fair comparability</p> <p>Baseline patient characteristics similar for age and site of obstruction.</p>	All subjects included in analysis	<p>Limitations for comparison</p> <p>8 Fr stents had pigtails whereas 10Fr stents were straight</p>	<p>Adequate outcome measures used.</p> <p>Outcomes were not assessed blindly.</p>	Univariate analysis does not account for important confounders	Poor

Table 43. Quality Assessment (cont'd)

Study Author, Year Record Number	Comparable Initial Groups?	Comparable Groups Maintained?	Comparable Performance of Intervention?	Comparable Measurement of Outcomes?	Appropriate Analysis	Summary Evaluation
Sung, Chung, Tsui et al., 1994	RCT (n=70) Good comparability - Sealed envelopes - Patient characteristics show no SSD	<u>SH</u> : (n=35) <u>NSH</u> : (n=35) 3 subjects dropped out before 4 week f/u and were excluded from analysis	Adequate for comparison	Adequate outcome measures used. Patient and follow-up physician were blinded to type of stent placed.	Method of analysis not reported but no crossover reported.	Good
Speer, Cotton, Russell et al., 1987	RCT (n=75) Good comparability - Computer generated random numbers and stratified by referring center - Patient characteristics similar for age, ASA ⁴³ grade, duration of jaundice, bilirubin, albumin, creatinine, and Hb, but ERCP group had more proximal obstructions, more unrelated medical problems, and more elevated WBC. No statistical results reported.	<u>ERCP</u> : (n=39) No dropouts 4 failures <u>Percutaneous</u> : (n=36) No dropouts 8 failures	Percutaneous stents were initially 6Fr and exchanged 2-3 days later to 12 Fr while endoscopic stents were 10 Fr in size	Adequate outcome measures used. Outcomes were not assessed blindly.	Intention-to-treat analysis used. Results were also analyzed taking into account relevant confounders that were not balanced.	Good

⁴³ American Society of Anesthesiology's performance status classification

Table 43. Quality Assessment (cont'd)

Study Author, Year Record Number	Comparable Initial Groups?	Comparable Groups Maintained?	Comparable Performance of Intervention?	Comparable Measurement of Outcomes?	Appropriate Analysis	Summary Evaluation
Pedersen, Lassen, De Muckadell et al., 1998	RCT (n=34) Good comparability - Randomization by computer generated numbers and sealed numbered envelopes - Baseline characteristics similar for age, type of cancer, and no SSD for gender	<u>Stent above SO (n=22)</u> 22 randomized - 5 technical failures crossed over. Final n=17. No other dropouts. <u>Stent across SO (n=19)</u> 19 randomized - 2 withdrawn for curative surgery. Final n=17. No other dropouts.	Adequate for comparison.	Adequate outcome measures used. Outcomes were not assessed blindly.	Method of analysis primarily based on treatment received. Results for one outcome reported using intention-to-treat.	Fair
DePalma, Galloro, Iovino et al., 2001	RCT (n=157) Good comparability - Randomization by sealed opaque envelopes - Baseline characteristics similar	<u>Unilateral stent (n=79)</u> No dropouts <u>Bilateral stent (n=78)</u> No dropouts	Adequate for comparison.	Adequate outcome measures used. Outcomes were not assessed blindly.	Intention to treat used.	Good
Chang, Kortan, and Haber 1998	Retrospective study (n=141) Baseline patient characteristics were comparable for age, gender, and tumor type	All subjects included in analysis	Adequate for comparison.	Adequate outcome measures used. Outcomes were not assessed blindly.	Analysis made some attempts to stratify results by Bismuth type, but did not fully consider possible confounders.	Fair

Table 43. Quality Assessment (cont'd)

Study Author, Year	Comparable Initial Groups?	Comparable Groups Maintained?	Comparable Performance of Intervention?	Comparable Measurement of Outcomes?	Appropriate Analysis	Summary Evaluation
Deviere, Baize, de Toeuf et al., 1988	Retrospective study (n=70) Baseline patient characteristics were not reported other than stricture type	All subjects included in analysis	Adequate for comparison.	Adequate outcome measures used. Outcomes were not assessed blindly.	Analysis made some attempts to stratify results by Bismuth type, but did not fully consider possible confounders.	Poor

Table 44. Comparison of Plastic versus Teflon™ stents

Study	N	Population and Interventions	Outcomes	Adverse Events	Comments												
Randomized Controlled Trials																	
van Berkel, Boland, Redekop et al., 1998	84	<p>Patients with distal malignant biliary stricture. No previous drainage procedure.</p> <p>Pancreas ca = 76 Papilla ca = 1 Bile duct ca = 5 Metastasis = 2</p> <p>42 Teflon™ stents 42 polyethylene stents (Amsterdam-type) All stents 10Fr and 9cm</p> <p>Baseline characteristics comparable.</p>	<p>Median survival (days) Teflon™ 165 Poly 140 p=0.6</p> <p>Successful biliary drainage Teflon™ 90% Poly 92%</p> <p>Median stent patency (days) Teflon™ 83 Poly 80 p=0.93</p> <p>No significant differences found in: Mean weight gain for 26 removed stents</p>	<p><u>Perioperative mortality</u> Teflon™ 14% Poly 14%</p> <p>Early procedure-related complications Teflon™ 4 (10%) Poly 4 (10%)</p> <p>Late complications</p> <table><thead><tr><th></th><th>Stent dysfunc</th><th>Repeat ERCP</th><th># ERCP</th></tr></thead><tbody><tr><td>Teflon™</td><td>28</td><td>24</td><td>79</td></tr><tr><td>Poly</td><td>29</td><td>25</td><td>75</td></tr></tbody></table>		Stent dysfunc	Repeat ERCP	# ERCP	Teflon™	28	24	79	Poly	29	25	75	<p>Univariate analysis of factors associated with reduced stent patency was reported.</p> <p>Previous failure of cannulation (p=0.03) Previous CBD contrast injection without papillotomy (p=0.004) Previous papillotomy (p=0.08)</p> <p>Gender, age>75, jaundice> 14 days, bilirubin > 300 µmol/L not significant factors.</p>
	Stent dysfunc	Repeat ERCP	# ERCP														
Teflon™	28	24	79														
Poly	29	25	75														

Table 45. Comparison of different caliber stents

Study	N	Population and Interventions	Outcomes	Adverse Events	Comments
Prospective observational studies					
Pedersen 1993	89	<p>Pts with malignant biliary strictures</p> <p>31 Single 7 Fr (S7) 45 Single 10 Fr (S10) 13 Double 7Fr (D7)</p> <p>85% of all patients also had sphincterotomy, evenly distributed between 7 and 10 Fr.</p> <p>7 Fr stent chosen when no large bore ERCP scope available.</p> <p>Baseline patient characteristics were different for age (7Fr group younger than 10Fr group). No SSD in stenosis location, gender, or type of cancer.</p>	<p>Median Stent Patency (days) Median, 25%-75% range</p> <p>S7 67 (20-336) S10 144 (39-237) D7 110 (62-145) Total 110 (33-237) P=0.11, comparing 7Fr vs. 10Fr</p>	<p><u>Mortality (2-week)</u></p> <p>S7 (n=31) 4 (13%) S10 (n=45) 4 (9%) D7 (n=13) 2 (15%) p=0.84</p> <p>Total Early Complications</p> <p>S7 (n=31) 13% S10 (n=45) 22.1% D7 (n=13) 23.1% p=n.s.</p> <p><u>Fever</u></p> <p>S7 (n=31) 9.7% S10 (n=45) 17.7% D7 (n=13) 23.1% p=n.r.</p> <p><u>Bleeding</u></p> <p>S7 (n=31) 6.5% S10 (n=45) 4.4% D7 (n=13) 0% p=n.r.</p> <p><u>Perforation</u></p> <p>S7 (n=31) 3.2% S10 (n=45) 0% D7 (n=13) 0% p=n.r.</p>	

Table 45. Comparison of different caliber stents (cont'd)

Study	N	Population and Interventions	Outcomes	Adverse Events	Comments
Retrospective studies					
Speer, Cotton, MacRae et al., 1988	79	<p>All patients receiving stent palliation for malignant obstructive jaundice</p> <p>28 8Fr <u>pigtail</u> stents 51 10Fr <u>straight</u> stents</p> <p>Baseline patient characteristics similar for age and site of obstruction.</p>	<p>Median Stent Patency (weeks)</p> <p>8 Fr 12 10 Fr 32 p<0.001</p> <p>Patency advantage of 10Fr stents primarily in first month.</p>	<p><u>Early complications (2 week)</u></p> <p><u>Cholangitis</u> 8 Fr (n=28) 13 (34%) 10 Fr (n=51) 3 (5%) p<0.01 (text)</p> <p>Local perforation 8 Fr (n=28) 2 (5%) 10 Fr (n=51) 4 (5%) p=n.s.</p> <p>Stent migration 8 Fr (n=28) 3 (8%) 10 Fr (n=51) 2 (3%) p=n.s.</p> <p><u>Late complications</u></p> <p>Need for stent replacement 8 Fr 12 (43%) 10 Fr 13 (25%) p=n.r.</p>	

Table 46. Comparison of stents with or without sideholes

Study	N	Population and Interventions	Outcomes	Adverse Events	Comments
Randomized Controlled Trials					
Sung, Chung, Tsui et al., 1994	70	<p>Most pts (93%) had malignant obstruction</p> <p>SH= side-hole stent (n=35) NSH = no side-hole (n=35)</p> <p>10Fr stents</p> <p>Patient characteristics show no SSD for age, gender, diagnosis, location of stent, prior stent</p>	<p>Biochemical improvement at 4 weeks SH (n=35) 95% NSH (n=32) 78% p>0.1</p> <p>All stent patency (weeks), median (range) SH (n=35) 7.8 (2.6-28) NSH (n=32) 7.9 (0.6-28) p>0.1</p> <p>Initial stent patency (weeks), median (range) SH (n=35) 9.5 (6.3-28) NSH (n=32) 8.0 (0.6-28) p>0.1</p> <p>Second stent patency (weeks), median (range) SH (n=35) 6.6 (2.6-19.9) NSH (n=32) 5.6 (0.9-23.3) p>0.1</p>	<p><u>Mortality</u> SH (n=35) 8 (23%) NSH (n=32) 8 (25%) p=n.r.</p> <p><u>Fever</u> SH (n=35) 82% NSH (n=32) 83% p=n.r.</p>	

Review of Evidence: Comparisons of Stent Placement

Five studies including three RCT (two quality rated as “Good” and one as “Fair”) and two retrospective studies (one “Fair” and one “Poor” quality) looked at issues of stent placement (Tables 47–49).

Speer, Cotton, Russell et al. (1987, n=75, “Good”) randomized patients to undergo percutaneous transhepatic placement of 12 Fr stents or endoscopic placement of 10 Fr stents (Table 47). This trial was terminated early when a prespecified statistical criterion was reached, specifically increased perioperative mortality was observed in subjects randomized to percutaneous stent insertion, 33% vs. 15%, $p=0.016$. Early complications also favored endoscopic over percutaneous placement (19% vs. 67%, $p=n.r.$). Patient survival and stent patency results did not demonstrate statistically significant differences.

Pedersen, Lassen, De Muckadell et al. (1998, n=34, “Fair”) randomized patients to have 10Fr stents placed with the inferior tip above the sphincter of Oddi or across the sphincter of Oddi (Table 48). Stents placed across the sphincter of Oddi were less likely to become dislocated (12% vs. 53%, $p=0.026$). Otherwise, no statistically significant differences were observed between the two groups with regard to patient survival, stent patency, procedure-related mortality, or complications.

Three studies compared results of unilateral versus bilateral stent placement in patients with biliary obstruction secondary to hilar malignancy (Table 49). DePalma, Galloro, Iovino et al. (2001, n=157, “Good”) provides the best evidence derived from a randomized controlled trial. This study finds no statistically significant differences in overall patient survival, perioperative mortality, procedure-related mortality, or late complications between those randomized to receive a unilateral versus bilateral stent. Moreover, the significant results reported favored unilateral stent placement over bilateral stents. Those randomized to receive bilateral stents had significantly lower rates of successful drainage (73% versus 81%, $p=0.049$), significantly more early complications (26.9% versus 18.9%, $p=0.026$), and significantly higher rates of cholangitis (16.6% versus 8.8%, $p=0.013$).

The two earlier retrospective studies, Chang, Kortan, and Haber (1998, n=141, “Fair”) and Deviere, Baize, de Toeuf et al. (1988, n=70, “Poor”) both examined patients who all had hilar malignancy and compared outcomes for those receiving unilateral or bilateral stents. Chang, Kortan, and Haber (1998) further considered subgroups who had different combinations of having received unilateral versus bilateral diagnostic biliary opacification and unilateral versus bilateral stent drainage. Deviere, Baize, de Toeuf et al. (1988) restricted analysis only to deceased patients. The results of these studies are complex with primary findings reported to be longer median patient survival in patients receiving bilateral drainage procedures, and higher perioperative mortality and increased rate of acute cholangitis among the subgroup which had unilateral stent placement in Deviere, Baize, de Toeuf et al. (1988) and the subgroup with unilateral drainage but bilateral diagnostic opacification performed in Chang, Kortan, and Haber (1998). However, the reported analyses do not fully account for various possible confounding influences and in light of findings of the randomized controlled trial, these retrospective findings are likely related to unmeasured differences in the groups being compared.

Table 47. Comparison of Percutaneous versus Endoscopic Stent Insertion

Study	N	Population and Interventions	Outcomes	Adverse Events	Comments												
Randomized Controlled Trials																	
Speer, Cotton, Russell et al., 1987	75	<p>Malignant biliary obstruction, unresectable</p> <p>Stents: 39 ERCP 10 Fr 36 Percutaneous 12 Fr</p> <p>Patient characteristics similar for age, ASA⁴⁴ grade, duration of jaundice, bilirubin, albumin, creatinine, and Hb, but ERCP group had more proximal obstructions, more unrelated medical problems, and more elevated WBC. No statistical results reported.</p>	<p>Survival (days), median (range)</p> <table><tr><td></td><td>Hilar</td><td>Low bile duct</td><td>Total</td></tr><tr><td>ERCP</td><td>65 (8-623)</td><td>160 (14-598)</td><td>119 (9-623)</td></tr><tr><td>PTH</td><td>24 (2-351)</td><td>94 (4-391)</td><td>88 (2-391)</td></tr></table> <p>p=0.35</p> <p>Stent patency (days) No significant difference in median time to blockage, p=0.16</p> <p>Failed Insertion ERCP (n=37) 4 PTH (n=33) 8</p> <p>Successful Insertion but No Drainage ERCP (n=37) 3 PTH (n=33) 5</p> <p>Relief of Jaundice ERCP (n=37) 30 (81%) PTH (n=33) 20 (61%) p=0.017</p> <p>Initial Hospitalization (days) (for those surviving at least 30 days) ERCP 11 (2-49) PTH 17 (3-24) p=0.4</p>		Hilar	Low bile duct	Total	ERCP	65 (8-623)	160 (14-598)	119 (9-623)	PTH	24 (2-351)	94 (4-391)	88 (2-391)	<p><u>Early complications</u> ERCP (n=37) 7 (19%) PTH (n=33) 22 (67%)</p> <p><u>Perioperative Mortality</u> ERCP 6 (15%) PTH 12 (33%) p=0.016 And Cox regression analysis confirmed that ERCP had significantly lower 30-day mortality (p=0.008).</p> <p>Cox proportional hazards model was performed. Predictors of 30-day mortality were ASA grade of 3 or more (p=0.002), randomization to PTH (p=0.008), WBC > 10 x10⁹ cells/l (p=0.018), hilar obstruction (p=0.01), and age 69-76 y (p=0.016). Predictors of decreased overall survival were WBC > 10 x10⁹ cells/l (p=0.01) and hilar obstruction (p=0.05)</p>	This trial was originally planned to enroll 200 patients. After the 1 st of 3 planned interim data analyses, the trial was halted based on prospectively defined statistical criteria.
	Hilar	Low bile duct	Total														
ERCP	65 (8-623)	160 (14-598)	119 (9-623)														
PTH	24 (2-351)	94 (4-391)	88 (2-391)														

⁴⁴ American Society of Anesthesiology's performance status classification

Table 48. Comparison of stent placement above versus across sphincter of Oddi

Study	N	Population and Interventions	Outcomes	Adverse Events	Comments									
Randomized Controlled Trial														
Pedersen, Lassen, De Muckadell et al., 1998	34	<p>Pts with unresectable CBD biliary obstruction</p> <p>17 placed above SO 17 placed across SO</p> <p>10 Fr straight stents</p> <p>Baseline characteristics Similar for age, type of cancer, and no SSD for gender</p>	<p><u>Patient survival (days)</u> Median (25%-75% range) Above SO (n=17) 144 (82-347) Across SO (n=17) 46 (35-155) p=n.s.</p> <p>Median stent patency (days) Median (25%-75% range) Above SO (n=17) 110 (61-320) Across SO (n=17) 126 (89-175) p=n.s.</p> <p>Intent-to-treat analysis: <u>Median stent patency (days)</u> Above SO (n=17) 99 (53-320) Across SO (n=17) 126 (89-175) p=n.s.</p> <p><u>Stent Function</u></p> <table><thead><tr><th></th><th># w/ Stent Dysfunction</th><th>Time to dysfunction</th></tr></thead><tbody><tr><td>Above SO</td><td>10</td><td>82 (31-185)</td></tr><tr><td>Across SO</td><td>5</td><td>89 (13-150)</td></tr></tbody></table> <p>p=n.s.</p>		# w/ Stent Dysfunction	Time to dysfunction	Above SO	10	82 (31-185)	Across SO	5	89 (13-150)	<p><u>Mortality (2 weeks)</u> Above SO (n=17) 2 (12%) Across SO (n=17) 1 (12%) p=n.s.</p> <p><u>Early complications (1 week)</u> Above SO (n=17) 2 (12%) Across SO (n=17) 4 (24%) p=n.s.</p> <p><u>Dislocation of stent</u> Above SO (n=17) 9 (53%) Across SO (n=17) 2 (12%) p=0.026</p>	
	# w/ Stent Dysfunction	Time to dysfunction												
Above SO	10	82 (31-185)												
Across SO	5	89 (13-150)												

Table 49. Comparison of unilateral versus bilateral drainage in hilar malignancy

Study	N	Population and Interventions	Outcomes	Adverse Events	Comments
Randomized Controlled Trials					
DePalma, Galloro, Iovino et al., 2001	157	<p>Pts w/ hilar obstruction due to cholangio-carcinoma, gallbladder cancer, or lymph node metastasis</p> <p>Type I (n=49) Type II (n=56) Type III (n=52)</p> <p>Randomized to unilateral (group A) or bilateral (Group B) stents</p>	<p><u>Median Survival (days)</u></p> <p>A 140 (21-612) B 142 (24-498) p=0.48</p> <p><u>Technical Success Drainage Success</u></p> <p>A 88.6 % 81% B 76.9 % 73% p= 0.041 0.049</p>	<p><u>Perioperative Mortality</u></p> <p>A 11.3% B 14.1% p=0.638</p> <p><u>Procedure-related Mortality</u></p> <p>A 2.5% B 3.8% p=0.681</p> <p><u>Early complications</u></p> <p>A 18.9% B 26.9% p=0.026</p> <p><u>Cholangitis</u></p> <p>A 8.8% B 16.6% p=0.013</p> <p><u>Late complications</u></p> <p>A 39.7% B 39.1% p=0.735</p>	

Table 49. Comparison of unilateral versus bilateral drainage in hilar malignancy (cont'd)

Study	N	Population and Interventions	Outcomes	Adverse Events	Comments
Retrospective Studies					
Chang, Kortan, and Haber 1998	141	<p>Pts w/ bifurcation tumors Bismuth Type: Type I (n=43) Type II (n=58) Type III (n=40)</p> <p>Types II and III were divided into 3 groups: N=32 A= one lobe of liver opacified with contrast and 1 side drained N=29 B = both lobes liver opacified and both drained N=37 C = both lobes liver opacified and one drained</p> <p>Single stents (n=104) 11 – 7 Fr; 40 – 10 Fr 53 – 11.5 Fr 3 – metal stents Double ERCP stents (n=15) 21 – 7 Fr; 7 – 10 Fr 2 – 11.5 Fr</p> <p>18 technical failures drained percutaneously Among those with double drains, 15 ERCP only, 3 PTH only, and 11 ERCP and PTH</p>	<p>Median survival (days)</p> <p>I 160 A 145 B 225 C 46 p<0.001</p> <p>Comparing single drains (groups A + C) versus double drains (group B), double drains had significantly better survival p<0.0001</p>	<p><u>Perioperative Mortality</u></p> <p>I 2 (5%) A 0 B 1 (3%) C 11 (30%) p<0.01</p> <p><u>Early complications</u></p> <p>Acute cholangitis I 2 (5%) A 2 (6%) B 0 C 12 (32%) p<0.01</p> <p>Stent migration I 1 (2%) A 0 B 0 C 1 (3%) p=n.s.</p> <p>Pancreatitis I 0 A 0 B 1 (3%) C 1 (3%) p=n.s.</p> <p><u>Total early complications</u></p> <p>I 3 (7%) A 2 (6%) B 1 (3%) C 14 (38%) p=n.s.</p> <p><u>Late complications</u></p> <p>Need for stent replacement I 19 (44%) A 16 (50%) B 12 (41%) C 2 (5%) p=n.r.</p>	This is a study comparing unilateral versus bilateral drainage of bifurcation tumors

Table 49. Comparison of unilateral versus bilateral drainage in hilar malignancy (cont'd)

Study	N	Population and Interventions	Outcomes	Adverse Events	Comments
Retrospective Studies (cont'd)					
Deviere, Baize, de Toeuf et al., 1988	70	Deceased pts with hilar tumors and biliary obstruction Type I stricture (n=20) 1 stent (Gr I-1) Type II or III (n=50) 24 w/ 1 stent (Gr II/III-1) 24 w/ 2 stent (Gr II/III-2) 2 w/ failed (Gr II/III-0)	Mean Survival (days) Median ⁴⁵ Gr I-1 156 (6-570) 156 Gr II/III-1 119 ^a (2-760) 162 Gr II/III-2 176 ^a (4-660) 198 Gr II/III-0 16 (6-26) ^a = p<0.01	<u>Perioperative Mortality</u> Gr I-1 0% Gr II/III-1 29% Gr II/III-2 8% Gr II/III-0 100%	

⁴⁵ Median survival after exclusion of patients who died within 30 days

Summary

Several additional comparative studies addressing variations in stent design and stent placement were identified in this systematic review. Since each research comparison has only one or no randomized controlled trial available, the results of these studies support only preliminary conclusions regarding the relative efficacy of these alternative approaches to stent palliation of pancreaticobiliary malignancy.

Part II, Section 4: Outcomes of Treatment Using Preoperative ERCP Drainage for Relief of Malignant Obstructive Jaundice

Introduction

Biliary obstruction results in a variety of biochemical and physiological disturbances such as elevated bilirubin and other liver function tests, as well as impaired hepatic and renal function with associated coagulation problems. In patients who are scheduled for potentially curative surgery, it has been postulated that using a course of preoperative biliary drainage to alleviate biliary obstruction may result in reduced surgical morbidity and mortality.

Evidence Base

Six studies addressed preoperative stenting compared to no stenting prior to surgery for malignant obstruction. Quality assessments are described in Table 50. Results are displayed in detail in the “Evidence Tables” chapter and summarized in Tables 51 and 52. The four nonrandomized series (Sewnath, Birjmohun, Rauws et al., 2001, n=290; Karsten, Allema, Reinders et al., 1996, n=241; ten Hoopen-Neumann, Gerhards, van Gulik et al., 1998, n=52; Heslin, Brooks, Hochwald et al., 1998, n=74) were judged to be of poor quality, largely due to lack of between-group comparability of patients or performance of intervention; and the randomized controlled trial by Lygidakis, van der Heyde, Lubbers et al. (1987, n=38) suffered from inappropriate use of statistical tests. Accompanying letters to the editor suggest that the conclusions as stated in the Lygidakis, van der Heyde, Lubbers et al. (1987) paper are not substantiated by the reported data. The randomized controlled trial by Lai, Mok, Fan et al. (1994, n=87) was judged to be of “Fair” quality, but is limited by insufficient sample size, which is the reason the trial was terminated by the investigators after initial analysis. Outcomes reported in these studies are largely limited to laboratory values and perioperative mortality and morbidity and postoperative hospital stay.

Review of Evidence: Treatment Outcomes

One randomized trial (Lygidakis, van der Heyde, Lubbers et al., 1987) and two nonrandomized comparisons reported on hospital days (Table 52). Lygidakis, van der Heyde, Lubbers et al. (1987) reported that preoperative ERCP group had higher initial hospital days (7 vs. 3.7) and lower total hospital days (23 vs. 26.7) than the no stent group, respectively. Tests of statistical significance were not reported. Heslin, Brooks, Hochwald et al. (1998, n=74) found patients receiving preoperative stents had slightly longer postoperative hospital stay (median of 11 versus 10 days, $p=0.04$) but Sewnath, Birjmohun, Rauws et al. (2001, n=290) reported slightly shorter postoperative stays in the stented groups that did not reach statistical significance (median of 13-15 days versus 16 days, $p=0.09$).

Lai, Mok, Fan et al. (1994) reported on technical success of preoperative stenting, which was 87%.

Table 50. Quality Assessment

Study Author, Year	Comparable Initial Groups?	Comparable Groups Maintained?	Comparable Performance of Intervention?	Comparable Measurement of Outcomes?	Appropriate Analysis	Summary Evaluation
Randomized Controlled Trials						
Lygidakis, van der Heyde, Lubbers et al., 1987	RCT (n=38) Patient characteristics similar. Method of randomization not specified	All subjects included in analysis	Adequate for comparison	Adequate outcome measures used. Outcomes were not assessed blindly.	All subjects enrolled were included in analysis. Inappropriate statistical tests used ⁴⁶	Poor
Lai, Mok, Fan et al., 1994	RCT (n=87) Fair comparability – Randomization: Consecutive numbered envelopes – Patient characteristics showed no SSD but early surgery w/o stent group tended to be higher risk with more medical problems	<u>Preop Stent:</u> (n=43) 6 technical failures crossed over 2 refused surgery after successful stent placement. <u>No Stent:</u> (n=44) No changes reported.	Adequate for comparison	Adequate outcome measures used. Outcomes were not assessed blindly.	Intention-to-treat analysis used in most comparisons. This trial was terminated because interim analysis showed that planned sample size was inadequate.	Fair

⁴⁶ Soreide O and Eide GE, Letter to the Editor: Preoperative Biliary Drainage. Acta Chir Scand 156:251-252 1990.

Table 50. Quality Assessment (cont'd)

Study Author, Year	Comparable Initial Groups?	Comparable Groups Maintained?	Comparable Performance of Intervention?	Comparable Measurement of Outcomes?	Appropriate Analysis	Summary Evaluation
Prospective Studies						
<p>Sewnath, Birjmohun, Rauws et al., 2001</p> <p>Same series as Karsten, Allema, Reinders et al., 1996, but subjects accrued June 1992 – Dec 2000</p>	<p>Prospective series (n=290)</p> <p>Excluded 21 patients who had external biliary drainage</p> <p>Fair comparability of baseline patient characteristics</p> <p>Patients without preop drainage were usually not jaundiced</p>	All subjects included in analysis	Adequate for comparison	<p>Adequate outcome measures used.</p> <p>Outcomes were not assessed blindly.</p>	<p>Analysis did compare preop drainage and no drainage for primary outcomes.</p> <p>Additional analysis by subgroups based on degree of preop jaundice</p>	Poor

Table 50. Quality Assessment (cont'd)

Study Author, Year	Comparable Initial Groups?	Comparable Groups Maintained?	Comparable Performance of Intervention?	Comparable Measurement of Outcomes?	Appropriate Analysis	Summary Evaluation
Retrospective Studies						
Karsten, Allema, Reinders et al., 1996 Subjects accrued Oct 1983 – June 1992	Retrospective series (n=241) Patients without preop drainage were usually not jaundiced; patients with jaundice assigned to ERCP Fair comparability of other baseline patient characteristics	All subjects included in analysis except for bile culture results obtained only in 195/241 (81%).	Adequate for comparison ERCP group received stent only if papillotomy alone was insufficient	Adequate outcome measures used. Outcomes were not assessed blindly.	Comparison of pre-op ERCP vs. immediate surgery outcomes lacking for most outcomes	Poor

Table 50. Quality Assessment (cont'd)

Study Author, Year	Comparable Initial Groups?	Comparable Groups Maintained?	Comparable Performance of Intervention?	Comparable Measurement of Outcomes?	Appropriate Analysis	Summary Evaluation
Retrospective Studies (cont'd)						
Heslin, Brooks, Hochwald et al., 1998	Retrospective series (n=74) Patients undergoing pancreaticoduodenectomy Slight imbalances in baseline patient characteristics such as gender and presence of positive nodes	All subjects included in analysis	Adequate for comparison	Adequate outcome measures used. Complications were assessed by an independent physician.	Analysis considered important outcomes. Secondary multivariable analysis did consider potential confounding factors. However, multivariable model may include too many candidate variables making it susceptible to overfitting.	Poor
ten Hoopen-Neumann, Gerhards, van Gulik et al., 1998	Retrospective series (n=52) Fair comparability Baseline patient characteristics showed no SSD for age, gender, tumor classification, type of surgery	All subjects included in analysis	No stent group included ERCP technical failures Post-operative radiation therapy performed in 37% of stent patients vs. 27% of immediate surgery patients.	Adequate outcome measures used. Outcomes were not assessed blindly.	Analysis did qualitatively identify possible confounding factors such as radiation therapy.	Poor

Table 51. Overview of studies and outcomes reported

Study	Population	Procedure	N Stent No Stent	Outcome Measures Reported						STUDY QUALITY
				Hospital Days	Laboratory Values	Technical Success	Perioperative Mortality	Perioperative Complications	Implantation Metastases	
Randomized Controlled Trials										
Lygidakis, van der Heyde, Lubbers et al., 1987	Patient with resectable pancreatic head carcinoma	preop ERCP placed stent	19	X	X		X	X		Poor
		vs. no pre-op stent	19							
Lai, Mok, Fan et al., 1994	Malignant obstructive jaundice	preop ERCP placed stent	43		X	X	X	X		Fair
		vs. no pre-op stent	44							
Prospective Studies										
Sewnath, Birjmohun, Rauws et al., 2001 Same series as Karsten, Allema, Reinders et al., 1996, but subjects accrued June 1992 – Dec 2000	Patients with presumed resectable tumor in pancreatic head region	232 had preop drainage - 192 stent+papillotomy - 27 papillotomy alone - 13 required percutaneous combined drainage procedure	232	X	X		X	X		Poor
		58 with no drainage were - 25 had dx ERCP only - 24 not jaundiced - 9 failed drainage and got immediate surgery	58							

Table 51. Overview of studies and outcomes reported (cont'd)

Study	Population	Procedure	N Stent No Stent	Outcome Measures Reported						STUDY QUALITY
				Hospital Days	Laboratory Values	Technical Success	Perioperative Mortality	Perioperative Complications	Implantation Metastases	
Retrospective Studies										
Karsten, Allema, Reinders et al., 1996 Subjects accrued Oct 1983 – June 1992	Patients with presumed resectable tumor in pancreatic head region	184 had preop drainage - 149 stent + papillotomy when papillotomy alone not sufficient - 25 papillotomy alone - 10 external drainage when ERCP stent not possible 57 with no drainage were not jaundiced (n=33) or had immediate operation planned (n=24)	149 57		X			X		Poor
Heslin, Brooks, Hochwald et al., 1998	Patients undergoing pancreaticoduodenectomy	39 had preop drainage 35 had no drainage preop	39 35	X	X		X	X		Poor
ten Hoopen-Neumann, Gerhards, van Gulik et al., 1998	Patients with Klatskin tumor with planned resection	41 of 52 had preop stent Main reasons for no stent were technical failure or lack of proximal congestion of bile	41 11		X				X	Poor

Table 52. Treatment Outcomes and Adverse Outcomes

Study	Study arm N	Hospital Days	p	Laboratory Values	p	Technical Success	p	Periop Mortality	p	Periop Complications	p	Implantation Metastases	p
Randomized Controlled Trials													
Lygidakis, van der Heyde, Lubbers et al., 1987	ERCP 19	Preop: 7 Total: 23 (Days for group/n)	nr	Significant reduction in Serum bilirubin, alkaline phosphatase, AST/SGOT, ALT/SGPT after stent	<.002			0 (0%)		3 (16%)	47		
	No stent 19	Preop: 3.7 Total: 26.7 (Days for group/n)		Significant increase in white blood cell count after stent Hct, creatinine, albumin, and clotting parameters unchanged	<.001			2 (11%) (1 sepsis; 1 aneurysm)		14 (74%) ⁴⁸			
				No significant change in laboratory values between baseline and preoperative testing									

⁴⁷ Inappropriate statistical tests reported raising concerns over appropriateness of conclusions reported.

⁴⁸ This study has a high baseline rate of cholangitis in the no stent group, which may contribute to the higher rate of complications in this group. Perioperative blood loss (800+/-100 vs/ 1800+/-200 mL.) and operative time (5+/- 2 vs. 7+/-2 h) were greater in the no stent group. Tests of statistical significance were not reported for these outcomes.

Table 52. Treatment Outcomes and Adverse Outcomes (cont'd)

Study	Study arm N	Hospital Days	p	Laboratory Values	p	Technical Success	p	Periop Mortality	p	Periop Complications	p	Implantation Metastases	p
Randomized Controlled Trials (cont'd)													
Lai, Mok, Fan et al., 1994	Stent 43			Serum bilirubin, alkaline phosphatase, ALT/SGPT but not AST/SGOT significantly lower than no stent group	<0.05	86%		6 (14%)	ns	Post- op:	16 (39)%	ns	
										Total ⁴⁹	23 (56%)		
	No Stent 44			Hb, Hct, BUN, creatinine, albumin no different. WBC not reported.				6 (14%)		Post- op	18 (41%)		
										Total	18 (41%)		

⁴⁹ In addition, 7 of the 23 patients had complications from both procedures (preoperative stenting and surgery.)

Table 52. Treatment Outcomes and Adverse Outcomes (cont'd)

Study	Study arm N	Hospital Days	p	Laboratory Values	p	Tech nical Succ ess	p	Periop erative Mortal ity	p	Perioperative Complications	p	Implan tation Metast eses	p
Prospective Studies													
Sewnath, Birjmohun, Rauws et al., 2001 Same series as Karsten, Allema, Reinders et al., 1996, but subjects accrued June 1992 – Dec 2000	Pre-op Drain (n=232)		0.09	Median decrease in bilirubin				1.3%	n.r.	50%	0.69		
	177 relieved of jaundice	13 (6-167)		82%*									
	32 with moderate jaundice	15 (12-39)		57%									
	23 with severe jaundice	15 (10-70)		37%* * p<0.01									
	No drainage 58	16 (8-222)		None reported				0%		55%			

Table 52. Treatment Outcomes and Adverse Outcomes (cont'd)

Study	Study arm N	Hospital Days	p	Laboratory Values	p	Technical Success	p	Perioperative Mortality	p	Perioperative Complications	p	Implan tation Metast eses	p
Retrospective Studies													
Karsten, Allema, Reinders et al., 1996 Subjects accrued Oct 1983 – June 1992	Pre-op Drain (n=184)			Median decrease in bilirubin	nr					<u>Infectious Complication</u> ⁵⁰	nr		
	149 stent+papillotomy			82%						Stent 49/149 (33%)			
	25 papillotomy alone			74%						Papillotomy 11/25 (44%)			
	10 external drainage			50%						External drain 6/10 (60%)			
	No drainage			None reported						No drainage 18/57 (32%)			
	57												

⁵⁰ The relationship between use of pre-operative drainage and postoperative complications was not significant when analyzed by preoperative bilirubin level.

Table 52. Treatment Outcomes and Adverse Outcomes (cont'd)

Study	Study arm N	Hospital Days	p	Laboratory Values	p	Tech nical Succ ess	p	Periop erative Mortal ity	p	Perioperative Complications	p	Implan tation Metast eses	p
Retrospective Studies (cont'd)													
Heslin, Brooks, Hochwald et al., 1998	Stent 39	11	0.04	Serum bilirubin, AST/SGOT significantly lower than no stent group. Albumin and alkaline phosphatase trended lower. BUN, creatinine, albumin, WBC no different.				2.6%	0.34	23 (59%)	0.04		
	No stent 35	10						0		12 (34%)			

Table 52. Treatment Outcomes and Adverse Outcomes (cont'd)

Study	Study arm N	Hospital Days	p	Laboratory Values	p	Tec hnic al Suc cess	p	Periop erative Mortal ity	p	Perioperative Complications	p	Implan tation Metast eses	p
Retrospective Studies (cont'd)													
ten Hoopen- Neumann, Gerhards, van Gulik et al., 1998	Stent 41			Bilirubin, mean (range) 117 (12-511)	0.008							8/41 (20%) ⁵¹	0.18
	No stent 11			235 (14-412)								0	

⁵¹ At 1 year, 4 of 8 patients with implantation metastases did not receive any postoperative radiation therapy. Overall, 37% of stented patients and 27% of non-stented patients did not receive radiotherapy (p=not reported)

Comparison of changes in laboratory values before and after placement of a preoperative stent consistently showed a reduction in serum bilirubin and liver function tests. One study showed a significant increase in white blood cell count in the preoperative stent group after stenting. These changes were significantly different from the pattern of laboratory values seen in the “no stent” group that went immediately to surgery. No significant changes were noted in hemoglobin, hematocrit, creatinine, blood urea nitrogen, albumin or coagulation profiles.

Review of Evidence: Adverse Outcomes

The available data shows no apparent differences in perioperative mortality (Table 52). Lygidakis, van der Heyde, Lubbers et al. (1987) reported no deaths in the stent group and 2 (11%) in the “no stent” group; and Lai, Mok, Fan et al. (1994) reported 14% mortality for both groups. However, the sample sizes ($n=34$ and $n=87$, respectively) in these randomized controlled trials are likely too small to make a meaningful comparison. A larger but nonrandomized comparative study (Sewnath, Birjmohun, Rauws et al., 2001, $n=290$) and a smaller retrospective comparison (Heslin, Brooks, Hochwald et al., 1998, $n=74$) also reported no statistically significant differences in mortality.

Only Lai, Mok, Fan et al. (1994) reported on total complications, including complications from preoperative endoscopic stenting plus those from surgery. Total complications were greater in the preoperative stent group (56% vs. 41%), but results were not statistically significant. Of patients in the preoperative stent group who had complications, 30% had complications from both preoperative endoscopic stenting and from surgery. Sewnath, Birjmohun, Rauws et al. (2001) reported no significant difference in postoperative complications (50% for stented versus 55% without stent, $p=0.69$) but also reported that 6% of those receiving preoperative stenting experienced a stent-related complication. Lygidakis, van der Heyde, Lubbers et al. (1987), Karsten, Allema, Reinders et al. (1996), and Heslin, Brooks, Hochwald et al. (1998) reported only postoperative complications. The nonrandomized comparison by Heslin, Brooks, Hochwald et al. (1998) reported higher complications in the stent group (59% versus 34%, $p=0.04$), and the study by Karsten, Allema, Reinders et al. (1996) reported the same rate of infective complications (39%) in no drainage group as in the preoperative ERCP papillotomy plus stent group.

The retrospective series by ten Hoopen-Neumann, Gerhards, van Gulik et al. (1998) reports that implantation metastases (i.e., metastases presumed to be attributable to an invasive procedure) occurred in 20% of patients with preoperative stent and none in patient without stent, but the difference was not statistically significant. Moreover, this study did not control for whether patients received postoperative radiation therapy.

Summary

The evidence available is limited by poor methodological quality and fails to demonstrate that preoperative stenting improves health outcomes. Five of the six studies were judged to be of poor quality and the sixth, a randomized controlled trial judged to be of fair quality, is limited by insufficient sample size. 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. The other studies did not report on total complications, and thus fail to account for the morbidity associated with undergoing two procedures rather than one. 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.

Results and Conclusions, Part III: Pancreatitis

This chapter reviews evidence on the following questions:

In 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)? (*Section 1: Diagnostic Performance of ERCP in Detecting Underlying Causes or Complications of Pancreatitis Amenable to Treatment – Comparison to Alternatives*)
- b. What are the outcomes of treatment using ERCP strategies compared to using surgical or medical therapy? (*Section 2: Outcomes of Treatment Using ERCP for Pancreatitis – Comparison of Strategies Using ERCP, Surgery, or Medical Management*)

Part III, Section 1: Diagnostic Performance of ERCP in Detecting Underlying Causes or Complications of Pancreatitis Amenable to Treatment—Comparison to Alternatives

Introduction

In this section, evidence was sought to find studies that compared the diagnostic performance of ERCP and another diagnostic modality to diagnose treatable causes or complications of pancreatitis. Studies that demonstrate the utility of a single diagnostic modality in detecting treatable conditions did not meet selection criteria; only studies comparing ERCP with an alternative method were included. Studies whose aim was to diagnose or characterize chronic pancreatitis itself by two diagnostic modalities also did not meet selection criteria. Common duct stones can cause pancreatitis, but these studies were included in the review of studies evaluating diagnosis of common duct stones (*see “ERCP Evidence Report Results and Conclusions, Part I: Common Bile Duct Stones”*).

Evidence Base

Only 3 studies were found that met selection criteria. Study quality is outlined in Table 53.

Review of Evidence

Duvnjak, Rotkvic, Vucelic et al. (1991, n=43, “Fair to Poor”; Table 54) compared ERCP to percutaneous cystopancreatography with measurement of pseudocyst amylase concentration to detect whether the pseudocyst communicates with the pancreatic duct. Knowledge of such a communication would help determine appropriate treatment for the pseudocyst. Although

Table 53. Quality Assessment

Study Author, Year	Patient Enrollment	Diagnostic performance of ERCP determined without knowledge of other test results	Diagnostic Performance of other test(s) determined without knowledge of ERCP results	Summary Evaluation
Duvnjak, Rotkvic, Vucelic et al., 1991	Prospective (n=43) States that patients were “randomly” selected, but otherwise not stated	Uncertain	Percutaneous pancreatography- Uncertain Amylase concentration- uncertain if 64 WU cutoff determined prospectively or post-hoc	Fair to poor
Bret, Reinhold, Taourel et al., 1996	Prospective (n=108) Most patients prospectively recruited, uncertain number with referral bias	Yes	Yes	Good
Takehara, Ichijo, Tooyama et al., 1994	Prospective (n=39) Not stated whether consecutive	Yes	Yes	Fair, small sample size

Table 54. Percutaneous pseudocystogram or percutaneous amylase measurement versus ERCP to diagnose communication between pseudocyst and pancreatic duct

Study	N	Population	Diagnostic test	Prevalence (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Comments
Duvnjak, Rotkvic, Vucelic et al., 1991	43	Patients with persistent pseudocysts >25 cm area on cross-section image	Percutaneous cystogram Amylase > 64 WU	51% communication	59 100	100 90	100 92	70 100	ERCP was the reference standard

cystopancreatography alone has poor sensitivity compared to ERCP, measurement of the amylase concentration showed that amylase concentration greater than 64 WU had a sensitivity of 100 percent and a specificity of 90 percent compared to ERCP. It is not stated whether the 64 WU cutoff was prospectively defined. These results require further prospective validation.

Bret, Reinhold, Taourel et al. (1996, n=108, “Good”; Table 55) compared ERCP to MRCP for the diagnosis of pancreas divisum. Out of 108 undergoing both ERCP and MRCP, pancreas divisum was demonstrated by both techniques in 6 patients with complete concordance. The clinical significance of this finding is uncertain, as it is not reported or known whether the demonstration of the pancreas divisum alone determined the etiology or treatment of the clinical problem.

Takehara, Ichijo, Tooyama et al. (1994, n=39, “Fair”; Table 56) compared ERCP to MRCP to examine morphology of the pancreatic ducts in 39 patients with chronic pancreatitis. Ductal narrowing is potentially treatable with surgery or endoscopy, although evidence supporting effectiveness is lacking. In the area of the pancreas with the highest prevalence of stenosis, MRCP had only fair sensitivity, 57 percent, and fair specificity, 73 percent. The prevalence of lesions in other parts of the pancreas is too low to make any conclusions comparing MRCP to ERCP.

Conclusion

In sum, there is an inadequate literature base to compare ERCP and other diagnostic modalities for the identification of treatable complications of pancreatitis.

Table 55. MRCP versus ERCP to diagnose pancreas divisum

Study	N	Population	Diagnostic test	Prevalence (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Comments
Bret, Reinhold, Taourel et al., 1996	108	Patients referred for ERCP for pancreatic disease	MRCP	6	100	100	100	100	ERCP was the reference standard

Table 56. MRCP versus ERCP to diagnose pancreatic duct stenoses and filling defects in patients with pancreatitis

Study	N	Population	Outcome studied	Prevalence (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Comments
Takehara, Ichijo, Tooyama et al., 1994	39	Patients with chronic pancreatitis	Stenosis head:	18	100	81	36	100	ERCP reference standard for all comparisons. 2 sets of data presented in paper, each observer compared with ERCP, only 1 set abstracted
			Stenosis body:	31	57	73	31	89	
			Stenosis Tail:	6	50	91	25	97	
			Filling defect head:	5	100	100	100	100	
			Filling defect body:	6	100	100	100	100	
			Filling defect Tail:	5	50	94	33	97	

Part III, Section 2: Outcomes of Treatment Using ERCP for Pancreatitis—Comparison of Strategies Using ERCP, Surgery, or Medical Management

Introduction

This chapter reviews the evidence on ERCP for the treatment of pancreatitis. Pancreatitis encompasses a number of distinct entities with differing etiologies, clinical expression, and treatment options. Each will be addressed separately to the extent allowed by the available literature. Also, there are a number of different endoscopic techniques employed for varying clinical situations. For the purposes of this chapter, “ERCP” will refer to the spectrum of interventional endoscopic techniques that are employed in the treatment of pancreatitis.

Evidence Base

Pancreatitis was classified as “acute,” “acute recurring,” and “chronic,” and evidence was sought to address a total of 9 separate indications within these classifications (Table 57). However, evidence meeting study selection criteria for this systematic review was available for only 4 of 9 indications of interest. These are: acute biliary pancreatitis; pancreas divisum; idiopathic recurrent pancreatitis, and pancreatic pseudocyst. Table 58 shows the quality and type of available evidence on pancreatitis together with the number of studies that met our inclusion criteria for each indication.. A more detailed account of the reason(s) for each of the excluded studies can be found in Table 59.

For acute pancreatitis, comparative studies are included that evaluate ERCP in the treatment of acute biliary pancreatitis. For acute recurrent pancreatitis (ARP) and chronic pancreatitis, there is a notable lack of comparative and/or prospective studies. To address the paucity of evidence on the indications, study selection criteria were relaxed to include retrospective, single arm studies that met a minimum threshold for reporting outcome measurements. Chronic pain, one of the most important outcome measures in chronic pancreatitis, is a subjective outcome that is prone to bias, especially when assessed in the absence of a comparison group. Therefore, retrospective single arm studies of acute relapsing and chronic pancreatitis were restricted to those that reported quantifiable pre and post measurements of pain and/or other similar outcomes such as analgesic use or hospitalization rates.

Review of Evidence: Acute Pancreatitis

Three randomized controlled trials compared early ERCP to delayed or selective ERCP. One associational study of a Veterans Administration database compared ERCP to surgery (Aiyer, Burdick, Sonnenberg et al., 1999).

Early ERCP Vs. Delayed or Selective ERCP for Acute Biliary Pancreatitis

There are three randomized controlled trials included in this review that compare early ERCP vs. delayed or selective ERCP for acute biliary pancreatitis. Two of these three trials were rated as “Good” (Fan, Lai, Mok et al., 1993; Folsch, Nitsche, Ludtke et al., 1997) by the quality

Table 57. ERCP in the treatment of pancreatitis: Overview of the literature by indication and study type

		Comparative studies			Single arm studies		
Indication	Status	RCT	Prospective non-randomized	Retrospective	Prospective	Retrospective	Total
Acute Pancreatitis							
Acute biliary pancreatitis	Reviewed	3	--	2	1	2	8
	Included	3	--	1	--	--	4
Acute non-biliary pancreatitis	Reviewed	--	--	--	--	--	--
	Included	--	--	--	--	--	--
Acute recurrent pancreatitis							
Pancreas divisum	Reviewed	1	--	--	--	7	8
	Included	1	--	--	--	2	3
Sphincter of Oddi dysfunction	Reviewed	--	--	--	--	--	--
	Included	--	--	--	--	--	--
Idiopathic ARP	Reviewed	1	1	--	1	1	4
	Included	1	0	--	--	--	1
Chronic pancreatitis							
Drainage of pseudocyst	Reviewed	--	--	1	1	3	5
	Included	--	--	1	1	1	3
Pancreatic duct stones (ERCP plus ESWL)	Reviewed	--	--	--	--	9	9
	Included	--	--	--	--	--	--
Pancreatic duct stricture (ERCP plus stenting)	Reviewed	--	--	--	--	11	11
	Included	--	--	--	--	--	--
Other chronic pancreatitis	Reviewed	--	--	--	--	6	6
	Included	--	--	--	--	--	--
Total	Reviewed	5	1	3	3	39	51
	Included	5	1	2	1	3	11

Table 58. Quality Assessment

Study, Year	Comparable Initial Groups?	Comparable Groups Maintained?	Comparable Performance of Intervention?	Comparable Measurement of Outcomes?	Appropriate Analysis	Summary Evaluation
Randomized controlled trials						
Neoptolemos, Carr-Locke, London et al., 1988	No <ul style="list-style-type: none"> Randomization process not well described Some baseline group differences present 	No	Yes	Yes	Yes Intent-to-treat analysis not performed, but exclusions <10% overall and ratio less than 2:1 between arms	FAIR Does not meet all quality indicators, but does not contain any fatal flaws
Fan, Lai, Mok et al., 1993	Yes (?) <ul style="list-style-type: none"> Randomization process not well-described groups appear balanced 	Yes	Yes Adequate for comparison	Yes	Yes Intent-to-treat analysis not performed, but exclusions <10% overall and ratio less than 2:1 between arms	GOOD Meets all quality indicators
Folsch, Nitsche, Ludtke et al., 1997	Yes	Yes	Yes	Yes	Yes	GOOD Meets all quality indicators
Lans, Geenen, Johanson et al., 1992	Yes (?) <ul style="list-style-type: none"> Randomization by 'card selection', ? adequate Small numbers make prone to selection bias Comparability of groups not demonstrated 	Yes (?) No dropouts	Yes	No <ul style="list-style-type: none"> Pt reported outcomes, no blinding to treatment No blinded outcome assessment 	Yes	FAIR Does not meet all quality indicators, but does not contain any fatal flaws

Table 58. Quality Assessment (cont'd)

Study, Year	Comparable Initial Groups?	Comparable Groups Maintained?	Comparable Performance of Intervention?	Comparable Measurement of Outcomes?	Appropriate Analysis	Summary Evaluation
Randomized controlled trials (cont'd)						
Jacob, Geenen, Catalano et al., 2001	Yes (?) <ul style="list-style-type: none"> Randomization process not described Small numbers make prone to selection bias Comparability of groups not demonstrated 	Yes (?) No dropouts	Yes	No <ul style="list-style-type: none"> Pt reported outcomes, no blinding to treatment No blinded outcome assessment 	Yes	FAIR Does not meet all quality indicators, but does not contain any fatal flaws
Non-randomized, retrospective comparative studies						
Aiyer, Burdick, Sonnenberg et al., 1999	No <ul style="list-style-type: none"> Database study, no randomized treatment assignment Highly prone to selection bias Comparability of groups not demonstrated 	No	No Cannot control for unequal intensity of treatment	Yes	Yes	POOR Lack of comparability of groups is a fatal flaw
Froeschle, Meyer-Pannwitt, Brueckner et al., 1993	No <ul style="list-style-type: none"> No randomized treatment assignment Highly prone to selection bias Comparability of groups not demonstrated Located 76% of treated patients 	No	No Cannot control for unequal intensity of treatment	Yes	No Statistical analysis not described or reported	POOR Lack of comparability of groups is a fatal flaw

Table 59. excluded articles

Study/yr.	Study description	Reason for exclusion
Acute pancreatitis		
Rosseland and Solhaug 1984	Retrospective comparative clinical series Compared early ERCP with delayed ERCP (historical controls) in acute biliary pancreatitis	No objective pre and post measurements
Uomo, Galloro, Rabitti et al., 1991	Prospective clinical series 50 patients with acute biliary pancreatitis treated with early ERCP	No comparison group
al Karawi, el Shiekh Mohamed, al Shahri et al. 1993 1062	Retrospective clinical series 35 patients with acute biliary pancreatitis treated with ERCP and EX at one institution	No comparison group
Chronic pancreatitis (not otherwise specified)		
Ell, Rabenstein, Schneider 1998	Retrospective clinical series 118 patients with chronic pancreatitis treated with guidewire versus needle-knife pancreatic sphincterotomy	Only short term complications reported Techniques not randomized, needle knife used if guidewire failed
Kim, Myung, Kim et al., 1998	Clinical trial 60 patients with chronic pancreatitis, treated with dual sphincterotomy vs. pancreatic sphincterotomy only	Only short term complications reported Only outcomes on small (n<25) subgroups reported
Kozarek and Terrance 1994	Retrospective clinical series 56 patients with chronic pancreatitis who were treated with ERCP and pancreatic duct sphincterotomy.	NR study question Primarily evaluated complications of stenting
Treacy and Worthley 1996	Retrospective (?) clinical series 9 patients with chronic pancreatitis treated with stents over a 3yr period at one institution	<25 patients
Guelrud, Mujica, Jaen et al., 1994	Retrospective clinical series 51 children and adolescents with acute recurrent pancreatitis over an 8-year period at one institution. 18 patients treated endoscopically	No objective pre and post measurements <25 patients (therapeutic)
Festen, Severijnen, vd Staak et al., 1991	Case reports of two children with chronic relapsing pancreatitis evaluated and treated with ERCP	<25 patients
Fuji, Amano, Ohmura et al., 1989	Retrospective clinical series 21 patients with chronic pancreatitis from one institution, treated with ERCP and endoscopic sphincterotomy	No objective pre and post measurements <25 patients
Bornman, Marks, Girdwood et al., 1980	Retrospective clinical series 52 patients with calcific pancreatitis who underwent ERCP	NR study question Evaluated the association of obstruction and pain in this population

Table 59. excluded articles (cont'd)

Study/yr.	Study description	Reason for exclusion
Stent treatment in chronic pancreatitis with stricture		
Grimm, Meyer, Nam et al., 1989	Retrospective clinical series 70 patients with obstructive chronic pancreatitis treated with ERCP with or without ESWL	No objective pre and post measurements
Ashby and Lo 1995	Retrospective, clinical series 21 patients with chronic pancreatitis and stricture, treated with ERCP and stent at one institution	<25 patients
Binmoeller, Jue, Seifert et al., 1995	Retrospective, clinical series 93 patients with chronic pancreatitis and stricture, treated with endoscopic stent at one institution over a 9-year period	No objective pre and post measurements
Smits, Badiga, Rauws et al., 1995	Retrospective clinical series. 51 patients with chronic pancreatitis and stricture of pancreatic duct, treated with ERCP over an 11-year period at one institution	No objective pre and post measurements
Cremer, Deviere, Delhaye et al., 1991	Retrospective clinical series. 76 patients with severe chronic pancreatitis and stricture, treated with endoscopic stent at one institution over a 4-year period.	No objective pre and post measurements
Kozarek, Patterson, Ball et al., 1989	Retrospective clinical series. 17 patients with chronic pancreatitis treated endoscopically with either stents or drains	Mixture of stents and drains for different indications
McCarthy, Geenen, and Hogan 1988	Retrospective clinical series. 35 patients with benign pancreatic disease and suspected obstruction treated with endoscopic stent	No objective pre and post measurements Mixed population (CP, pancreas divisum, unexplained pain)
Ponchon, Gagnon, Berger et al., 1995	Retrospective clinical series 23 patients with chronic pancreatitis, pain and MPD stricture treated with ERCP stenting	No objective pre and post measurements <25 patients
Smith and Sherman 1996	Retrospective clinical series 61 patients treated with pancreatic stenting at one institution	NR study question Primarily evaluated complications of stenting
Sherman, Hawes, Savides, et al., 1996	Retrospective clinical series 61 patients with stent treatment who had long term follow-up after stent removal	NR study question Primarily evaluated complications of stenting
Vitale, Reed, Nguyen, et al., 2000	Retrospective clinical series 25 patients with chronic pancreatitis and CBD stricture, treated with ERCP stent	No objective pre and post measurements

Table 59. Excluded articles (cont'd)

Study/yr.	Study description	Reason for exclusion
Endoscopic treatment of pancreatic pseudocysts		
Kolars, Allen, Ansel, et al., 1989	Retrospective clinical series 51 patients with pseudocyst, treated either with surgery alone, ERCP alone, or ERCP followed by surgery	No relevant outcome data No objective pre and post measurements
Ahearne, Baillie, Cotton, et al., 1992	Retrospective clinical series 102 patients with pseudocysts, treated according to algorithm at one institution. Most patients (69/102) received surgical drainage	NR study question Did not evaluate outcomes of ERCP treatment
Endoscopic treatment of pancreatic duct stones		
Smits, Rauws, Tytgat, et al. 1996	Retrospective clinical series. 53 patients with chronic pancreatitis and pancreatic stones treated with ERCP from one institution over a 9-year period	No objective pre and post measurements
Dumonceau, Deviere, Le Moine, et al., 1996	Retrospective clinical series 70 patients with chronic pancreatitis and pancreatic stones, treated with ERCP at one institution over a 15-year period	No objective pre and post measurements
Kozarek, Ball, Patterson, et al., 1992	Retrospective clinical series. 12 patients with chronic pancreatitis and pancreatic duct stones treated with ERCP at one institution	No objective pre and post measurements <25 patients
Sherman, Lehman, Hawes, et al., 1991	Retrospective clinical series. 32 patients with chronic pancreatitis and pancreatic stones treated with ERCP at two institutions	No objective pre and post measurements
Ponsky and Duppler 1987	Case report Description of technique and response to therapy by patient	<25 patients No objective pre and post measurements
ERCP plus lithotripsy for pancreatic stones		
Ohara and Oshino 1996	Retrospective clinical series 32 patients with chronic pancreatitis and pancreatic duct stones, treated with ERCP and lithotripsy at one institution over a 4-year period	No objective pre and post measurements
Schreiber, Gurakuqi, Pristautz, et al., 1996	Retrospective clinical series. 10 patients with pancreatic stones and chronic pancreatitis treated with ERCP and lithotripsy over a 2-year period from a single institution	No objective pre and post measurements <25 patients
Schneider and May 1994	Retrospective clinical series 50 patients with chronic pancreatitis and pancreatic stones treated with ERCP and lithotripsy at one institution	No objective pre and post measurements
Delhay, Vandermeeren, Baize, et al., 1992	Retrospective clinical series 123 patients referred for chronic pancreatitis who were treated with ERCP and lithotripsy at one institution over a 2-year period	No objective pre and post measurements

Table 59. excluded articles (cont'd)

Study/yr.	Study description	Reason for exclusion
Pancreas divisum		
Satterfield, McCarthy, Geenen, et al., 1988	Retrospective clinical series 82 patients with pancreas divisum seen at 2 institutions over a 4-year period Descriptive analysis of multiple subgroups	Outcomes not reported for all patients Reported outcome data on only 10/33 patients with pancreatitis
Chevillotte, Sahel, Pietri, et al., 1984 (French with English abstract)	Retrospective clinical series Descriptive analysis of 63 cases of pancreas divisum, from a series of 2800 ERCP procedures over a 6-year period at one institution	No objective pre and post measurements
Warshaw, Richter, and Schapiro, 1983	Retrospective clinical series 40 patients with pancreas divisum and recurrent pancreatitis or refractory pain, treated endoscopically over an 8-year period at one institution	No objective pre and post measurements
Keith, Shapero, and Sabil, 1982	Retrospective case series 5 patients with chronic or recurrent acute pancreatitis and pancreas divisum treated with ERCP and sphincterotomy, from 480 patients seen with pancreatitis at one institution over a 5 year period.	No objective pre and post measurements
Other studies		
Guelrud, Morera, Rodriguez, et al., 1999	Retrospective clinical series 128 children with pancreatobiliary disease who underwent ERCP at one institution over a 14-year period	NR study question (evaluated prevalence of sphincter of Oddi dysfunction in children with recurrent pancreatitis) Mixed population of patients with pancreatobiliary pathology
Hammarstrom, Stridbeck, and Ihse, 1997	Retrospective clinical series 28 patients who received ERCP treatment for benign pancreatic disease, from 319 patients who underwent ERCP at one institution for suspected pancreatic disease over a 13-year period	Mixed population of patients with benign pancreatic disease No objective pre and post measurements
He, Zheng, Zhang, et al., 2000	Retrospective clinical series 56 patients with congenital choledochal cysts, 39 evaluated and treated with ERCP	No objective pre and post measurements
Kozarek and Traverso 1996	Review and expert opinion	No primary data
Mori, Nagakawa, Ohta, et al., 1991	Retrospective clinical series 48 patients with anomalous union of pancreatic ducts, identified over an 11-year period at one institution	NR study question Evaluated prevalence of pancreatitis in patients with anomalous union of the ductal system
Malfertheiner and Buchler 1991	Review	No primary data

Table 59. excluded articles (cont'd)

Study/yr.	Study description	Reason for exclusion
Other studies (cont'd)		
Venu, Geenen, Hogan, et al., 1989	Retrospective clinical series 116 patients with idiopathic recurrent pancreatitis referred for ERCP at one institution	NR study question (yield study) Evaluated diagnostic yield of ERCP in this population
Ammann, Akovbiantz, Larglader, et al., 1984	Prospective cohort study 163 patients with chronic pancreatitis at two hospitals over a 19-year period.	NR study question Evaluated natural history of chronic pancreatitis
Himal 1999	Retrospective clinical series 55 patients with mild biliary pancreatitis. Evaluated ERCP preoperatively prior to cholecystectomy	NR study question
Testoni, Caporuscio, Bagnolo, et al., 2000	Prospective (?) clinical series 40 patients with idiopathic recurrent pancreatitis. Evaluated yield of ERCP for etiology and follow-up after treatment. Microlithiasis (n=11), sphincter of Oddi dysfunction (n=14), pancreas divisum (n=3), no etiology (n=12)	<25 patients for any one category

assessment, the third was rated as “Fair” (Neoptolemos, Carr-Locke, London et al., 1988). Among the three randomized controlled trials, there are differences in the patient eligibility criteria, severity of pancreatitis and application of ERCP intervention that are important to interpretation of the results (Table 60, Table 61). With respect to patient population: Neoptolemos, Carr-Locke, London et al. (1988, n=121) is restricted to patients with acute biliary pancreatitis; Fan, Lai, Mok et al. (1993, n=195) includes patients with non-biliary pancreatitis; and Folsch, Nitsche, Ludtke et al. (1997, n=238) excluded patients with signs of obstructive jaundice, and the remaining population largely represented patients with mild pancreatitis. Thus, the likelihood that pancreatitis was associated with ongoing biliary obstruction was highest in the Neoptolemos, Carr-Locke, London et al. (1988) study; lower in the Fan, Lai, Mok et al. (1993) study because patients with nonbiliary causes of pancreatitis were included; and lowest in the Folsch, Nitsche, Ludtke et al. (1997) study, which excluded patients with obvious obstruction.

In all three studies, patients were classified with mild or severe pancreatitis based on commonly used scales. These scales use readily available clinical information to predict prognosis in acute pancreatitis, but are not specifically meant to select patients for ERCP or to identify patients with biliary obstruction. Given the sophistication of contemporary imaging techniques, such classification systems may be of less clinical significance in predicting which patients are likely to benefit from ERCP treatment.

In these studies, ERCP was performed in 20–28 percent of patients in the delayed or selective groups. This represents a substantial minority of patients in the control group that actually underwent ERCP; but is a much lower percentage compared to the early ERCP groups, where almost all patients had the procedure.

Treatment Outcomes. No study reported statistically significant differences in mortality between groups (Table 62). Neoptolemos, Carr-Locke, London et al. (1988) and Fan, Lai, Mok et al. (1993) found numerically greater mortality in the delayed or selective ERCP group, but only for patients with severe pancreatitis. Consistent with these data, in a study population with milder disease, Folsch, Nitsche, Ludtke et al. (1997) found numerically greater mortality in the early ERCP group. This trial was terminated prematurely as the question of interest was whether early ERCP might lead to reduced mortality in the study population.

The lack of benefit for early ERCP in Folsch, Nitsche, Ludtke et al. (1997) is seen in conjunction with the exclusion of patients with ongoing biliary obstruction. This implies that the potential mortality benefit of ERCP is limited to patients with obstruction. Additionally, the overall magnitude of benefit among these studies appears to be related to the likelihood of biliary obstruction in the population. Neoptolemos, Carr-Locke, London et al. (1988), which reports the greatest benefit, also has the highest likelihood of obstruction in their population, while the study with the least benefit, Folsch, Nitsche, Ludtke et al. (1997), has a population with the lowest likelihood of obstruction. The population in the Fan, Lai, Mok et al. (1993) study had a higher likelihood of obstruction compared to Folsch, Nitsche, Ludtke et al. (1997). Neoptolemos, Carr-Locke, London et al. (1988), reported a degree of benefit intermediate between those studies.

For total complications, Neoptolemos, Carr-Locke, London et al. (1988) reported a statistically significant reduction for the early ERCP group. Fan, Lai, Mok et al. (1993) and Folsch, Nitsche, Ludtke et al. (1997) reported no significant difference in total complication rates. However, Fan,

Lai, Mok et al. (1993) observed half as many total complications with early ERCP (22 of 41 patients vs. 44 of 40) among the subgroup of patients with severe pancreatitis, but did not report statistical significance. In a subgroup analysis of patients with severe pancreatitis and documented common bile duct stone, Fan, Lai, Mok et al. (1993) reported a significantly lower rate of total complications for early ERCP group (3/19 vs. 10/16, $p=0.005$). In a study population presenting mainly with mild pancreatitis, Folsch, Nitsche, Ludtke et al. (1997) reported a significantly greater respiratory failure (15/126 vs. 5/112, $p=0.03$) with early ERCP.

In summary, the interpretation of this group of studies is that early ERCP reduces complications in patient populations with acute pancreatitis and biliary obstruction. In studies that report benefit for patients with severe pancreatitis, but not mild pancreatitis, this finding likely represents the correlation of biliary obstruction with more severe disease. In patients with low likelihood of biliary obstruction, a clinical approach that includes delayed or selective ERCP may result in lower complications, and permits many patients to avoid the procedure.

Previous meta-analysis. Sharma and Howden (1999), pooled four randomized controlled trials of early vs. delayed or selective ERCP for acute biliary pancreatitis, three of which are the studies discussed here. The fourth randomized controlled trial, Nowak, Nowakowska-Dulawa, Marek et al. (1995), has been published only in abstract form. This meta-analysis is flawed because it combines studies that have different patient populations and interventions. Also, these studies report subgroup analyses suggesting that aggregate outcomes may be misleading when applied to subsets of patients that are stratified on the severity of pancreatitis or the likelihood of biliary obstruction.

The authors computed summary estimates for total mortality and complications, and reported the relative risk reduction associated with the early ERCP strategy. For overall mortality, the combined relative risk reduction associated with early ERCP was 42.9 percent. For total complications, there was a 34.6 percent relative risk reduction associated with early ERCP. These summary results are driven largely by the results of Neoptolemos, Carr-Locke, London et al. (1988) and Nowak, Nowakowska-Dulawa, Marek et al. (1995), neither of which allowed selective early ERCP in the control group for clinical indications. The authors did not perform sensitivity analyses or stratified analysis of the data.

The authors concluded that all patients with acute biliary pancreatitis should undergo early ERCP. Given the differences in the methodology of these studies and the lack of rigor in the meta-analysis, this conclusion is not supported by a critical analysis of the data.

ERCP vs. Surgery for Acute Pancreatitis

There was a single study that met the inclusion criteria for this comparison (Table 63, Table 64). This study (Aiyer, Burdick, Sonnenberg et al., 1999) was a retrospective comparison of outcomes for patients with biliary pancreatitis that were treated initially either by ERCP or surgery, using the United States Veterans Administration computerized database. Investigators identified all hospitalizations in the VA database that had simultaneous diagnoses of pancreatitis and cholelithiasis. Outcomes for 650 patients treated initially with ERCP were compared with 1,425 patients treated initially with surgery.

This study was assigned a quality rating of “Poor” by quality assessment. The major methodologic limitation of this study is that the two groups being compared are likely to differ substantially on a variety of clinical factors. Limited information contained in the database on severity of illness indicated that the patients in ERCP group were older and had higher baseline Charlsson score as compared to patients initially treated with surgery. Also, a higher percentage of patients in the ERCP group had cholangitis, choledocholithiasis, and pancreatic cysts.

Outcomes for the two groups were generally similar or favorable towards ERCP, despite the fact that the ERCP group appeared to be more severely ill. Mortality was 4 percent for the surgery group and 2 percent for the ERCP group ($p=0.08$), while the rate of total complications was identical for the two groups at 2 percent.

Conclusions

Early ERCP Vs. Delayed or Selective ERCP for Acute Biliary Pancreatitis

Evidence from three randomized controlled trials 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 complications.

ERCP vs. Surgery for Acute Pancreatitis

A single retrospective study suggests that outcomes from ERCP are at least as good as those from surgery. This study reported comparable outcomes for the two groups despite evidence for a higher severity of illness in ERCP group. However, this is a retrospective database study and confidence in the conclusions is limited by a number of methodologic factors, especially the potential for imbalances among the groups that are compared. Also, given the limited clinical information available, this study cannot ascertain the best strategy to employ given particular patient characteristics and/or clinical presentation.

Review of Evidence: Acute Recurrent Pancreatitis

Four studies, two randomized controlled trials and two single-arm retrospective series, met the inclusion criteria for this category. The main outcomes reported in these studies were pain, episodes of recurrent pancreatitis and/or hospitalization (Table 65).

Acute, Recurrent Pancreatitis Associated with Pancreas Divisum

Three studies, one randomized controlled trial (Lans, Geenen, Johanson et al., 1992) and two retrospective single-arm studies (Lehman, Sherman, Nisi et al., 1993; Kozarek, Ball, Patterson et al., 1995), reporting on a total of 110 patients, evaluated ERCP treatment for acute, recurrent pancreatitis associated with pancreas divisum. Lans, Geenen, Johanson et al. (1992) was a randomized controlled trial in 19 patients with pancreas divisum and recurrent acute pancreatitis. All patients received diagnostic ERCP, and patients who were amenable to stenting were randomized to stent or no stent. Patients were followed for a mean of approximately 30 months for the outcomes of recurrent pancreatitis, emergency room visits/hospitalizations, and clinical improvement. The quality of this study was rated “Fair.” Confidence in the results of this study

is limited by its small size, lack of blinding, and lack of comparison with alternatives. Quality ratings were not applied to the two retrospective single studies, which are prone to confounding by the placebo effect, natural history of the disease, and a potentially large number of clinical factors.

The small randomized controlled trial by Lans, Geenen, Johanson et al. (1992, n=19) and the two retrospective single-arm studies (n=91) reported that ERCP treatment with stent or sphincterotomy decreased recurrent episodes of pancreatitis, and reduced pain as measured on visual analog scales. None of these studies met the threshold study selection criteria initially set for this systematic review. Although the body of evidence is sparse and largely uncontrolled, the observation that hospitalizations and emergency room visits were significantly reduced is consistent for both the single randomized controlled trial and the less rigorous single arm studies.

Idiopathic Acute, Recurrent Pancreatitis

A single, small, randomized controlled trial (Jacob, Geenen, Catalano et al., 2001, n=34) in patients with idiopathic acute, recurrent pancreatitis reported that ERCP plus stenting reduces episodes of recurrent acute pancreatitis as compared to diagnostic ERCP alone. However, the percent of patients with persistent pain was no less in the ERCP plus stent group as compared to the diagnostic ERCP group. Thus, this trial provides evidence that ERCP treatment reduces subsequent episodes of pancreatitis in idiopathic recurrent acute pancreatitis, similar to the results seen in patients with pancreas divisum. However, this single small, unblinded trial is insufficient to determine whether ERCP treatment reduces pain in patients who present with idiopathic acute recurrent pancreatitis.

Review of Evidence: Chronic Pancreatitis

The three studies (n=187) included in this review evaluate ERCP drainage of pancreatic pseudocysts (Table 66). There are a number of different endoscopic approaches for drainage of pseudocysts. The available studies generally report aggregate outcomes and are not adequately robust to compare outcomes among different approaches to drainage. Thus, this review will not attempt to differentiate among variations of endoscopic drainage. Only one of these studies is prospective (Barthet, Sahel, Bodiou-Bertei et al., 1995), and none provides robust information on prospective, long-term outcomes from these procedures.

One of the three studies met the threshold study selection criteria initially set for this systematic review (Froeschle, Meyer-Pannwitt, Brueckner et al., 1993). Results of this retrospective comparative study initial suggest that ERCP drainage results in a similar rate of pain relief as compared with surgery, with equivalent or lower mortality. Two additional single arm series that met the relaxed selection criteria suggest that regression of pseudocysts occurs in a majority of cases following ERCP drainage, in the range of 70–86 percent (Libera, Siqueira, Morais et al., 2000; Barthet, Sahel, Bodiou-Bertei et al., 1995). Pain relief after ERCP drainage was reported in the comparative study and in one case series, with approximately half of patients reporting complete pain relief following the procedure. The uncontrolled trial by Libera, Siqueira, Morais et al. (2000) also reported a significant improvement in pain scores following ERCP drainage.

Using a 0–3 pain scale, the mean pain score was reduced from 2.48 pre-treatment to 0.28 post-treatment ($p<0.001$).

Conclusions

For treatment of acute pancreatitis, 3 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 complications. In addition, one retrospective associational study of a Veterans Administration database of patient 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 in order to address this question. 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.

Table 60. Comparison of population and intervention in RCTs of ERCP for acute biliary pancreatitis

	Patient population	Early ERCP	Delayed/selective ERCP	Severity Pancreatitis	
				mild	severe
Neoptolemos, Carr-Locke, London et al., 1988	<ul style="list-style-type: none"> Patients hospitalized with acute biliary pancreatitis No other cause for pancreatitis 	ERCP \pm ES within 72 hours of admission for all patients	No patient received ERCP within first five days. Selective ERCP performed in 23% of control patients after day five for clinical indications (not specified).	56%	44%
Fan, Lai, Mok et al., 1993	<ul style="list-style-type: none"> Patients hospitalized with acute pancreatitis (all causes) No prior work-up for biliary stones Pancreatitis not induced by ERCP 	ERCP \pm ES within 24 hours of admission for all patients	Selective ERCP performed in 28% of control patients for rising fever, leukocytosis or tachycardia; increasing jaundice or bilirubin; shock	58%	42%
Folsch, Nitsche, Ludtke et al., 1997	<ul style="list-style-type: none"> Patients hospitalized with acute pancreatitis No signs of obstructive jaundice No other potential causes of pancreatitis 	ERCP \pm ES within 72 hours of onset of symptoms in all patients	Selective ERCP performed in 20% of control patients for signs of obstructive jaundice	78%	22%

Table 61. Early ERCP for treatment of acute biliary pancreatitis – study characteristics

Study	Population	Study design	Interventions(s)	Outcomes	Comments
Early ERCP vs. delayed/selective ERCP					
Neoptolemos, Carr-Locke, London et al., 1988	131 pts with suspected acute biliary pancreatitis, drawn from 223 consecutive pts admitted with acute pancreatitis <u>Exclusions:</u> 1) age less than 18yrs, 2) chronic alcoholism or acute alcohol intake, 3) pregnancy, and 4) identifiable secondary cause for pancreatitis.	Single center RCT Patients randomized to immediate ERCP or conventional management. Patients followed until discharged from hospital. All ERCP procedures performed by one “highly skilled” endoscopist.	<u>Immediate ERCP</u> – ERCP +/- ES within 72hrs of hospitalization. <u>Control</u> – Conventional management for first five days. Patients in conventional management group offered ERCP + ES after 5 days if clinically indicated.	Mortality Local complications (pseudocysts, ascites, duodenal obstruction) Systemic complications (respiratory failure, cardiovascular failure, stroke, DIC, renal failure)	No patients in control group got ERCP until at least day 5.
Fan, Lai, Mok et al., 1993	195 pts with acute biliary pancreatitis, selected from 206 consecutive patients with acute pancreatitis <u>Exclusions:</u> 1) prior workup for biliary stones 2) iatrogenic pancreatitis	Single center RCT Patients randomized to immediate ERCP or selective ERCP. Patients followed until discharge from hospital.	<u>Immediate ERCP</u> – ERCP +/- ES within 24hrs of hospitalization. <u>Control</u> – Selective ERCP for: rising fever, leukocytosis, or tachycardia; increasing jaundice or bilirubin; shock. All control patients had elective ERCP after acute attack resolved if selective ERCP not performed.	Mortality Local complications (pseudocysts, abscess, phlegmon, bleeding) Systemic complications (respiratory failure, cardiovascular failure, sepsis, DIC, renal failure, GI bleeding)	ERCP performed selectively in 27/98 (28%) control patients. Study included patients with etiologies for pancreatitis other than biliary stones. 64% of patients in study had documented biliary stones.
Folsch, Nitsche, Ludtke et al., 1997	238 adult patients with suspected acute biliary pancreatitis, selected from 339 consecutive patients <u>Exclusions:</u> 1) Indications for early ERCP (bilirubin >5, temp >39°), 2) age <18yrs, 3) pregnancy, 4) inability to perform ERCP within 72hrs of onset of symptoms.	Multi-center RCT, 22 clinical centers Patients randomized to immediate ERCP or selective ERCP. Patients followed for three months	<u>Immediate ERCP</u> – ERCP +/- ES within 72hrs of onset of symptoms. <u>Control</u> – Conventional management. ERCP performed for persistent biliary colic, temp >39°, or increased bilirubin. After 3 weeks, ERCP could be performed in any patient if indicated.	Mortality Local complications (pseudocysts, ascites, duodenal obstruction) Systemic complications (respiratory failure, cardiovascular failure, stroke, DIC, renal failure)	ERCP performed selectively in 22/112 (20%) of patients. Study terminated early due to inability to show a benefit in the early ERCP group.

Table 62. Early ERCP for treatment of acute biliary pancreatitis – outcomes

Study/yr.	Severity	Mortality		P value	Complications								
		Early ¹	D/S ²		Overall		P value	Systemic		P value	Local		P value
Early ERCP vs. delayed/selective ERCP													
Neoptolemos, Carr-Locke, London et al., 1988	Overall (n=121)	1.7% (1/59)	8.1% (5/62)	0.23	17% (10/59)	34% (17/62)	0.03	7% (4/59)	19% (12/62)	0.08	12% (7/59)	24% (15/62)	0.08
	Mild (n=68)	0% (0/34)	0% (0/34)	NS	12% (4/34)	12% (4/34)	NS	2.9% (1/34)	0% (0/34)	NR	12% (4/34)	12% (4/34)	NS
	Severe (n=53)	4% (1/25)	18% (5/28)	NR	24% (6/25)	61% (17/28)	<0.01	12% (3/25)	43% (12/28)	NR	12% (3/25)	39% (11/28)	NR
Fan, Lai, Mok et al., 1993	Overall (n=195)	5.2% (5/97)	9.2% (9/98)	0.40	18% (17/97)	29% (28/98)	NR	10% (10/97)	14% (14/98)	NS	10% (10/97)	12% (12/98)	NS
	Mild (n=114)	0% (0/56)	0% (0/58)	NS	8 total/ 56 pts	6 total/ 58 pts		1 total/ 56 pts	5 total/ 58 pts		7 total/ total/ 56 pts	1 58	
	Severe (n=81)	12% (5/41)	23% (9/40)	NR	22 total/ 41 pts	44 total 40 pts		16 total/ 41 pts	33 total/ 40 pts		6 total/ total/ 41 pts	11 40	
Folsch, Nitsche, Ludtke et al., 1997	Overall (n=238)	11% (14/126)	6.3% (7/112)	0.10	46% (58/126)	51% (57/112)	NS	91 total/ 126 pts	89 total/ 112 pts		25% (31/126)	25% (28/112)	
	Mild (n=160)												
	Severe (n=46)												

¹ Early ERCP group

² Delayed and/or selective ERCP group

Table 63. ERCP vs. surgery for treatment of acute biliary pancreatitis – study characteristics

Study	Population	Study design	Interventions(s)	Outcomes	Comments
ERCP vs. surgery					
Aiyer, Burdick, Sonnenberg et al., 1999	2075 pts with acute biliary pancreatitis from VA system, 650 treated with endoscopy and 1425 treated with surgery.	Retrospective analysis of VA database, comparing outcomes and complications of endoscopy versus surgery	<u>ERCP</u> – Received ERCP as initial intervention during hospitalization for acute biliary pancreatitis <u>Surgery</u> – Had cholecystectomy and/or other biliary/pancreatic surgery as initial intervention during hospitalization for acute biliary pancreatitis	Mortality Local complications (pseudocysts) Systemic complications (respiratory failure, sepsis, GI bleed, DIC, renal failure, hypocalcemia) Complications from therapy (hemorrhage, laceration/puncture of viscus organ)	

Table 64. ERCP vs. surgery for treatment of acute biliary pancreatitis – outcomes

Study/yr.	Populations/Severity	Mortality	P value	Complications (overall)	P value
ERCP vs. surgery					
Aiyer, Burdick, Sonnenberg et al., 1999	<u>ERCP:</u> (n=650) average SOI by Charlsson score 0.9	2% (15/650)	0.08	2% (14/650)	0.94
	<u>Surgery:</u> (n=1425) average SOI by Charlsson score 0.8	4% (56/1425)		2% (33/1425)	

*32 patients had undefined severity level

Table 65. ERCP for treatment of acute recurrent pancreatitis

Study	Population	Study design	Interventions(s)	Outcomes	Comments																																
Acute recurrent pancreatitis associated with pancreas divisum																																					
Lans, Geenen, Johanson et al., 1992	19 patients with pancreas divisum and recurrent acute pancreatitis at one institution over a 5yr period <u>Exclusions:</u> other potential causes of pancreatitis; prior pancreatic resection or sphincterotomy	Randomized controlled trial ERCP alone vs. ERCP plus stent. F/U every 4 mos. in both groups Mean F/U 28.6 mos. for stent group, 31.5 mos. for controls	Stent placement in dorsal pancreatic duct. Stent replaced every 4 mos. in stent group. Stents removed after one year	1) Number of hospitalizations ER visits Stent (n=10) 0 Control (n=9) 7 p<0.05 2) Number of episodes acute pancreatitis Stent (n=10) 1 Control (n=9) 7 p<0.05 3) Number of pts with subjective improvement on visual analogue scale Stent (n=10) 9 Control (n=9) 1 p<0.05																																	
Kozarek, Ball, Patterson et al., 1995	39 pts with pancreas divisum and chronic pancreatitis (CP) (n=19), acute relapsing pancreatitis (ARP) (n=15), or chronic abdominal pain (CAP) (n=5)	Retrospective (?) single arm case series	ERCP treatment determined at time of treatment: Stent 13 pts Sphincterotomy 4 pts Stent + Sphinct 22 pts	1) Pain (0-10 scale) <table><tr><td></td><td><u>Pre</u></td><td><u>Post</u></td><td><u>p value*</u></td></tr><tr><td>CP</td><td>9.4</td><td>4.8</td><td><0.001</td></tr><tr><td>Pain</td><td>8.3</td><td>7.3</td><td></td></tr><tr><td>ARP</td><td>NR</td><td>NR</td><td></td></tr></table> * pre vs. post 2) number of episodes pancreatitis/year <table><tr><td></td><td><u>Pre</u></td><td><u>Post</u></td><td><u>p value*</u></td></tr><tr><td>CP</td><td>2.0</td><td>1.6</td><td>0.025</td></tr><tr><td>Pain</td><td>NR</td><td>NR</td><td></td></tr><tr><td>ARP</td><td>2.1</td><td>0.3</td><td>0.016</td></tr></table> * pre vs. post		<u>Pre</u>	<u>Post</u>	<u>p value*</u>	CP	9.4	4.8	<0.001	Pain	8.3	7.3		ARP	NR	NR			<u>Pre</u>	<u>Post</u>	<u>p value*</u>	CP	2.0	1.6	0.025	Pain	NR	NR		ARP	2.1	0.3	0.016	
	<u>Pre</u>	<u>Post</u>	<u>p value*</u>																																		
CP	9.4	4.8	<0.001																																		
Pain	8.3	7.3																																			
ARP	NR	NR																																			
	<u>Pre</u>	<u>Post</u>	<u>p value*</u>																																		
CP	2.0	1.6	0.025																																		
Pain	NR	NR																																			
ARP	2.1	0.3	0.016																																		

Table 65. ERCP for treatment of acute recurrent pancreatitis (cont'd)

Study	Population	Study design	Interventions(s)	Outcomes	Comments																																
Acute recurrent pancreatitis associated with pancreas divisum (cont'd)																																					
Lehman, Sherman, Nisi et al., 1993	52 previously untreated pts with pancreas divisum and chronic pancreatitis (CP) (n=11), acute recurrent pancreatitis (ARP) (n=17), or disabling pancreatic pain (Pain) (n=24)	Retrospective (?) single arm case series	ERCP plus sphincterotomy of minor papilla	<div>1) Pain (0-10 scale)</div> <table><thead><tr><th></th><th>Pre</th><th>Post</th><th>p value*</th></tr></thead><tbody><tr><td>CP</td><td>9.5 ± 0.3</td><td>6.6 ± 1.3</td><td>NS</td></tr><tr><td>Pain</td><td>8.4 ± 0.2</td><td>6.6 ± 0.8</td><td>0.02</td></tr><tr><td>ARP</td><td>9.1 ± 0.3</td><td>2.1 ± 0.8**</td><td><0.001</td></tr></tbody></table> <div>* pre vs. post ** significantly greater change in symptom score as compared to CP (p=0.007) and pain (p<0.001)</div> <div>2) number of hospital days/month</div> <table><thead><tr><th></th><th>Pre</th><th>Post</th><th>p value*</th></tr></thead><tbody><tr><td>CP</td><td>1.7 ± 0.3</td><td>1.5 ± 0.5</td><td>NS</td></tr><tr><td>Pain</td><td>1.4 ± 0.4</td><td>1.0 ± 0.2</td><td>NS</td></tr><tr><td>ARP</td><td>1.6 ± 0.4</td><td>0.1 ± 0.1**</td><td><0.001</td></tr></tbody></table> <div>* pre vs. post ** significantly greater change in hospital days as compared to CP (p<0.05) and pain (p=0.003)</div>		Pre	Post	p value*	CP	9.5 ± 0.3	6.6 ± 1.3	NS	Pain	8.4 ± 0.2	6.6 ± 0.8	0.02	ARP	9.1 ± 0.3	2.1 ± 0.8**	<0.001		Pre	Post	p value*	CP	1.7 ± 0.3	1.5 ± 0.5	NS	Pain	1.4 ± 0.4	1.0 ± 0.2	NS	ARP	1.6 ± 0.4	0.1 ± 0.1**	<0.001	
	Pre	Post	p value*																																		
CP	9.5 ± 0.3	6.6 ± 1.3	NS																																		
Pain	8.4 ± 0.2	6.6 ± 0.8	0.02																																		
ARP	9.1 ± 0.3	2.1 ± 0.8**	<0.001																																		
	Pre	Post	p value*																																		
CP	1.7 ± 0.3	1.5 ± 0.5	NS																																		
Pain	1.4 ± 0.4	1.0 ± 0.2	NS																																		
ARP	1.6 ± 0.4	0.1 ± 0.1**	<0.001																																		

Table 65. ERCP for treatment of acute recurrent pancreatitis (cont'd)

Study	Population	Study design	Interventions(s)	Outcomes	Comments																		
Idiopathic acute recurrent pancreatitis																							
Jacob, Geenen, Catalano et al., 2001	34 patients with idiopathic acute recurrent pancreatitis randomized to ERCP alone or ERCP plus stenting of pancreatic duct	Prospective, randomized, non-blinded clinical trial	<u>ERCP alone:</u> diagnostic ERCP and pancreatogram at baseline and every 3 mos. for 9 mos. Mean follow-up 35 mos. <u>ERCP plus stent:</u> ERCP plus stenting of pancreatic duct, stent changed every 3 mos. for 9 mos.. Mean follow-up 33 mos.	Recurrent episodes of pancreatitis: <table><tr><td></td><td></td><td><u>P value</u></td></tr><tr><td>ERCP alone</td><td>53% (8/15)</td><td></td></tr><tr><td>ERCP plus stent</td><td>11% (2/19)</td><td><0.02</td></tr></table> Persistence of pain*: <table><tr><td></td><td></td><td><u>P value</u></td></tr><tr><td>ERCP alone</td><td>40% (6/15)</td><td></td></tr><tr><td>ERCP plus stent</td><td>32% (6/19)</td><td>NS</td></tr></table> *Presence of pancreatic type pain of at least moderate intensity (4 or greater on 0-10 scale) post-treatment			<u>P value</u>	ERCP alone	53% (8/15)		ERCP plus stent	11% (2/19)	<0.02			<u>P value</u>	ERCP alone	40% (6/15)		ERCP plus stent	32% (6/19)	NS	
		<u>P value</u>																					
ERCP alone	53% (8/15)																						
ERCP plus stent	11% (2/19)	<0.02																					
		<u>P value</u>																					
ERCP alone	40% (6/15)																						
ERCP plus stent	32% (6/19)	NS																					

Table 66. ERCP for treatment of chronic pancreatitis

Study	Population	Study design	Interventions(s)	Outcomes	Comments						
Endoscopic drainage of pseudocysts											
Libera, Siqueira, Morais et al., 2000	30 pts referred for drainage of pseudocysts. <u>Inclusion:</u> 1) Pseudocyst >4cm for at least 6 weeks with persistent abdominal pain, 2) progressive increase in size, 3) complications from pseudocyst	Retrospective (?) single arm case series	ERCP drainage performed in one of four ways: 1) transpapillary 2) cyst-gastrostomy 3) cyst-duodenoscopy 4) combined procedure Drainage performed with or without stent, as clinically indicated Treatments were repeated, or alternate drainage attempted, if clinically indicated.	1) Abdominal pain (0-3 scale): <table><tr><td><u>Pre</u></td><td><u>Post</u></td><td><u>p value</u></td></tr><tr><td>2.48 ± 0.51</td><td>0.28 ± 0.64</td><td><0.001</td></tr></table> Complete pain relief in 17/30 pts (57%) 2) Regression of pseudocyst on CT: 21/30 (70%) pts had regression. 21/25 (84%) pts with successful procedure had regression 3) Complications: 6 complications among 37 procedures (16.2%) 2 stent migration 1 duodenal perforation 1 bleeding 1 pancreatitis 1 pneumoperitoneum	<u>Pre</u>	<u>Post</u>	<u>p value</u>	2.48 ± 0.51	0.28 ± 0.64	<0.001	
<u>Pre</u>	<u>Post</u>	<u>p value</u>									
2.48 ± 0.51	0.28 ± 0.64	<0.001									
Barthet, Sahel, Bodiou-Bertei et al., 1995	30 pts with pancreatic pseudocyst amenable to drainage by ERCP. <u>Exclusions:</u> none	Prospective single arm clinical series	Transpapillary ERCP performed in all cases. Serial US and/or CT at 4 mo. intervals. F/U ERCP performed if cyst no longer present on imaging	Early resolution of pseudocyst: 26/30 (87%) Recurrence of pseudocyst: 3/26 (12%) Complications: 4/30 (13%)	7/30 patients needed surgical intervention, 3 for failure of pseudocyst to resolve and 4 for recurrence						

Table 66. ERCP for treatment of chronic pancreatitis (cont'd)

Study	Population	Study design	Interventions(s)	Outcomes	Comments																																
Endoscopic drainage of pseudocysts (cont'd)																																					
Froeschle, Meyer-Pannwitt, Brueckner et al., 1993	127 pts treated for pancreatic pseudocysts from one hospital. 35% treated surgically, 29% endoscopically, 6% percutaneously	Retrospective comparative analysis of outcomes and complications among the three approaches used	Surgery (n=44) Endoscopy (n=37) Percutaneous (n=7) Combined procedure (n=26) No procedure (n=13) F/U performed a mean of 33 mos. after intervention 30/127 (23.6%) lost to F/U.	<div>1) Mortality<table><thead><tr><th></th><th>Post-op</th><th>F/U</th><th>p value</th></tr></thead><tbody><tr><td>Surgery</td><td>6.8%</td><td>13.6%</td><td>NR</td></tr><tr><td>Endoscopy</td><td>0</td><td>2.7%</td><td>NR</td></tr><tr><td>Combined</td><td>0</td><td>15.4%</td><td>NR</td></tr></tbody></table></div> <div>2) Percent of patients free of pain at F/U<table><thead><tr><th></th><th></th><th></th><th>p value</th></tr></thead><tbody><tr><td>Surgery</td><td>50%</td><td>(16/32)</td><td>NR</td></tr><tr><td>Endoscopy</td><td>52%</td><td>(16/31)</td><td>NR</td></tr><tr><td>Combined</td><td>54%</td><td>(10/18)</td><td>NR</td></tr></tbody></table></div>		Post-op	F/U	p value	Surgery	6.8%	13.6%	NR	Endoscopy	0	2.7%	NR	Combined	0	15.4%	NR				p value	Surgery	50%	(16/32)	NR	Endoscopy	52%	(16/31)	NR	Combined	54%	(10/18)	NR	
	Post-op	F/U	p value																																		
Surgery	6.8%	13.6%	NR																																		
Endoscopy	0	2.7%	NR																																		
Combined	0	15.4%	NR																																		
			p value																																		
Surgery	50%	(16/32)	NR																																		
Endoscopy	52%	(16/31)	NR																																		
Combined	54%	(10/18)	NR																																		

Results and Conclusions, Part IV: Abdominal Pain Of Possible Pancreaticobiliary Origin

This chapter reviews evidence on the following questions:

In 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)? (*Section 1: Diagnostic Performance of ERCP Manometry in Evaluation of Abdominal Pain of Possible Pancreaticobiliary Origin—Comparison To Alternatives*)
- b. What are the outcomes of treatment using ERCP strategies compared to using surgical or medical therapy? (*Section 2: Outcomes of Treatment Using ERCP for Abdominal Pain of Possible Pancreaticobiliary Origin*)

Part IV, Section 1: Diagnostic Performance of ERCP Manometry In Evaluation of Abdominal Pain of Possible Pancreaticobiliary Origin—Comparison With Alternatives

Evidence Base

Three studies comparing biliary scintigraphy with ERCP with or without manometry for the diagnosis of sphincter of Oddi dysfunction met the inclusion criteria for this chapter. There were a total of 136 patients enrolled in these studies, 54 of whom had sphincter of Oddi dysfunction. Quality assessment of these studies is available in Table 67. The study characteristics and diagnostic performance of biliary scintigraphy in these studies are summarized in Table 68.

Review of Evidence

There are notable differences in the study objectives, populations, diagnostic criteria for biliary scintigraphy, and reference standards that limit the ability to synthesize results from these studies. The earliest study (Kloiber, AuCoin, Hershfield et al., 1988) evaluated the ability of biliary scintigraphy to diagnose obstruction of the biliary tree postcholecystectomy. In this study, not all patients with obstruction had sphincter of Oddi dysfunction. Sostre, Kalloo, Spiegler et al. (1992) compared a number of different biliary scintigraphy diagnostic criteria for sphincter of Oddi dysfunction in a consecutive sample of postcholecystectomy patients, with the intent of determining the optimal criterion for diagnosing sphincter of Oddi dysfunction. The most recent study, Peng, Lai, Tsay et al. (1994), attempted to define the performance characteristics of biliary scintigraphy in a group of patients with suspected sphincter of Oddi

Table 67. Quality Assessment

Study Author, Year	Patient Enrollment	Diagnostic performance of ERCP determined without knowledge of other test results	Diagnostic Performance of other test(s) determined without knowledge of ERCP results	Summary Evaluation
Peng, Lai, Tsay et al., 1994	Retrospective study Partial description provided of method of enrollment of 60 patients.	No	No	Fair
Sostre, Kalloo, Spiegler et al., 1992	Prospective study 26 consecutive patients	Yes	Yes	Good
Kloiber, AuCoin, Hershfield et al., 1988	Retrospective study (?) Partial description provided of method of enrollment of 50 consecutive patients	No	No	Fair

Table 68. Study Details

Study	Pt population N enrolled	N evaluable	Diagnostic Test criterion	Prev (%)	Sens (%)	Spec (%)	PPV (%)	NPV (%)	Adeq Studies (%)	Comments
ERCP + Manometry Reference Standard										
Peng, Lai, Tsay et al., 1994	34 pts with: <ul style="list-style-type: none">PostcholecystectomyRUQ symptomsNormal LFT'sNo other pathology on UGI, US, ERCP	26	Quantitative scintigraphy Time activity curve	62	69	80	85	62	n.r.	
			Common bile duct dynamics	62	69	90	92	64	n.r.	
	26 control pts: <ul style="list-style-type: none">PostcholecystectomyAsymptomaticNormal LFT's									
Sostre, Kalloo, Spiegler et al., 1992	26 consecutive postcholecystectomy patients, some with biliary pain, some with non-biliary pain and some with no symptoms	26	Quantitative scintigraphy						n.r.	This study administered CCK routinely to all patients before scintigraphy. 12/26 pts thought to have SOD
			Liver peak	46	83	79	77	85		
			Biliary visualization	46	50	100	100	70		
			Biliary prominence	46	100	79	80	100		
			Bowel visualization	46	92	71	73	91		
			CBD emptying	46	100	93	92	100		
			CBD-to-Liver ratio	46	100	86	86	100		
			Final scintigraphic score	46	100	100	100	100		
ERCP Reference Standard										
Kloiber, AuCoin, Hershfield et al., 1988	50 consecutive pts with <ul style="list-style-type: none">PostcholecystectomyRUQ pain	50	Quantitative scintigraphy Time to peak bile duct activity	18	93	64	n.r.	n.r.	n.r.	Scintigraphy was used to assess presence of obstruction in post- choly syndrome. 9/50 pts thought to have SOD

dysfunction and a control group of asymptomatic postcholecystectomy patients. Other differences in the study populations, diagnostic criteria, and reference standards for biliary scintigraphy are summarized in Table 68.

The reported performance characteristics varied among these studies. The sensitivity of biliary scintigraphy for diagnosing sphincter of Oddi dysfunction ranged from 50–100 percent. The specificity ranged from 64–100 percent. The positive predictive value ranged from 73–100 percent and the negative predictive value ranged from 62–100 percent. Confidence intervals were not reported around the point estimates for these values in any of the studies. While it is likely that differences in study methodology and populations are related to the variability in reported outcomes, it cannot be determined which variables are associated with variability in outcomes.

Conclusions

The 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.

Part IV, Section 2: Outcomes Of Treatment Using ERCP For Abdominal Pain of Possible Pancreaticobiliary Origin

Introduction

Patients with abdominal pain showing a typical biliary or pancreatic pattern who have undergone diagnostic evaluation excluding a pancreaticobiliary anatomic or structural cause for the pain may have what is termed “sphincter of Oddi dysfunction.” This diagnostic category of functional abdominal pain encompasses both sphincter of Oddi stenosis and sphincter of Oddi dyskinesia. In sphincter of Oddi stenosis, there is persistent narrowing in the region of the sphincter of Oddi with abnormal pancreaticobiliary manometry findings of elevated basal pressure and abnormality of phasic contraction patterns. In sphincter of Oddi dyskinesia, there is intermittent functional obstruction in the sphincter of Oddi, and, like sphincter of Oddi stenosis, basal sphincter of Oddi pressures may be elevated at manometry, but in sphincter of Oddi dyskinesia abnormal manometry pressures may be temporarily reversible following administration of a smooth muscle relaxant (Tzovaras and Rowlands, 1998).

Classification systems for biliary type pain have been proposed with one frequently cited system derived by Hogan and Geenen (1998). In this system, patients are classified into Types I, II, and III, depending on the number of features present. Type I biliary patients have all features present including: typical biliary type pain, elevated alanine transaminase (ALT) and aspartate transaminase (AST) on two separate occasions, dilated common bile duct on ultrasound or ERCP, and delayed biliary drainage. Type II biliary patients have biliary type pain and only one or two of the additional features required for Type I. Finally, Type III patients have biliary type pain but none of the accompanying features. The prevalence of sphincter of Oddi dysfunction is generally highest for Type I biliary patients and decreases among Type II and Type III biliary patients. Additional modifications of this classification system have been made reflecting the limited role of delayed biliary drainage as a criterion (personal communication, Elta G.).

Pancreatic type sphincter of Oddi dysfunction has been classified into three types by Sherman, Troiano, Hawes, et al., (1991). In this system, Type I patients demonstrate recurrent pancreatitis and/or typical pancreatic-type pain, elevated amylase and/or lipase, dilated pancreatic duct, and prolonged drainage of pancreatic duct. Type II pancreatic type patients have typical pancreatic-type pain and one or two of the additional features listed for Type I patients. Type III pancreatic type patients have typical pancreatic type pain but none of the accompanying features.

Evidence Base

This systematic review selected studies reporting results of endoscopic treatment with sphincterotomy in patients with abdominal pain of suspected pancreaticobiliary origin (e.g., suspected sphincter of Oddi dysfunction). Studies comparing outcomes of ERCP sphincterotomy with alternative treatment strategies were included.

There were 7 studies that met the selection criteria for this question. Quality ratings are described in Table 69 and results of these studies are detailed in Tables 70 and 71. Two of these studies were prospective randomized, controlled trials (Geenen, Hogan, Dodds et al., 1989; Toouli, Robert-Thomson, Kellow et al., 2000) and met the study selection criteria as originally defined. Because of the paucity of evidence found using the original selection criteria, criteria were relaxed to include single arm studies that reported quantifiable pre- and post-outcome measures, or that compared outcomes among relevant clinical subgroups. Four studies were identified that met these modified selection criteria. One was a prospective single-arm study that evaluated consecutive patients treated with endoscopic sphincterotomy and used quantifiable pre- and post-outcome measures. Three additional articles were retrospective single-arm studies in which outcomes were compared among different clinical subgroups of patients. These studies evaluated the relative success of treatment in relation to specific clinical factors.

Finally, an eighth study, a randomized controlled trial (Jamidar, Sherman, and Hawes, 1992) was only available in abstract form and has not been submitted for publication (personal communication, Sherman S, August 2001). This abstract was not included in the review of evidence.

Review of Evidence: Randomized Controlled Trials

There were 2 double-blind randomized, controlled trials reporting on a total of 126 patients, comparing endoscopic sphincterotomy with a sham procedure (Table 70). Both of the published randomized, controlled trials were rated as “Good” by quality assessment. Strengths of these randomized, controlled trials include double blinding, the use of a sham procedure in the control group, and independent blinded assessment of outcomes. For both studies, the primary outcome was improvement in abdominal pain. Geenen, Hogan, Dodds et al. (1989) compared outcomes between groups at 1 year and Toouli, Robert-Thomson, Kellow et al. (2000) compared outcomes at 2 years. Geenen, Hogan, Dodds, et al. (1989) also reports the number of patients in each group who have persistent objective abnormalities (increased liver enzymes, dilatation of common bile duct, delayed contrast drainage) following treatment.

In the Geenen, Hogan, Dodds, et al. (1989) study, there was a significantly greater improvement in pain scores for the overall endoscopic sphincterotomy group as compared to control (65 percent vs. 30 percent with good/fair improvement, $p < 0.01$). In Toouli, Robert-Thomson, Kellow et al. (2000), more patients in the endoscopic sphincterotomy group had improvement in pain scores than in the sham endoscopic sphincterotomy group (62 percent vs. 43 percent), however, statistical significance was not reported for the overall group comparison.

Both studies evaluated subgroups of patients with and without an elevated sphincter of Oddi pressure, defined as greater than 40mmHg. In patients with an elevated pressure, both studies report a statistically significant benefit for the endoscopic sphincterotomy group. Geenen, Hogan, Dodds, et al. (1989) reported that 91 percent (10/11) patients in the endoscopic sphincterotomy group had good or fair improvement in pain scores, compared with 25 percent (3/12) in the sham group. Similarly, Toouli, Robert-Thomson, Kellow et al. (2000) reported that 85 percent of patients in the endoscopic sphincterotomy group with elevated pressure had

Table 69. Quality Assessment in studies comparing endoscopic treatment in patients with abdominal pain of suspected pancreaticobiliary origin

Study Author, Year	Comparable Initial Groups?	Comparable Groups Maintained?	Comparable Performance of Intervention?	Comparable Measurement of Outcomes?	Appropriate Analysis	Summary Evaluation
Geenen, Hogan, Dodds, et al., 1989	RCT (n=47) Unknown comparability - Randomization by sealed opaque envelopes - patient characteristics not reported	All subjects included in one-year outcome analysis Four-year follow-up only in 40 of 47. All 7 had normal SO pressure (5 ES; 2 sham). Four lost to f/u and 3 dropped out.	Adequate for comparison.	Double-blinded assessment for 1-year outcomes. Outcome measurement instruments for pain not well described.	Method of first-year outcomes analysis not stated but equivalent to intention-to-treat because all subjects enrolled were included in analysis. Four-year analysis equivalent to treatment received because sham cross-overs were analyzed with ES group.	Good
Toouli, Robert-Thomson, Kellow et al., 2000	RCT (n=81) Comparability - randomized by draw of cards - patient characteristics not reported	One lost to follow-up and 1 dropout due to pancreatitis x 2.	Adequate for comparison.	Double-blinded assessment for two-year outcomes. Outcome measurement instruments for pain not well described.	Does not clearly state method of analysis	Good

Table 70. Randomized Controlled Trials

Study	N	Study Group	Improved Pain Scores	P	Mean Symptom Score	P	Objective Abnormalities ¹	P	Complications	P
Geenen, Hogan, Dodds, et al., 1989 ² Group II Biliary patients	23	<u>Overall:</u> ES	<u>One-Year:</u> Good/fair improvement	<0.01			Baseline 1-year	n.r.	1 Hemorrhage 1 Perforation 2 Pancreatitis	
	24	Sham	15/23 (65%) 7/17 (30%)				37 6 49 30			
		<u>SOM >40 mmHg³</u> ES	10/11 (91%)	<0.005	Baseline 10	1-year 1.8	21	n.r.		
	11	Sham	3/12 (25%)		10	6.7	30 22			
	12	<u>SOM <40 mmHg³</u> ES	5/12 (42%)	n.r.	10	5.7	16	n.r.		
	12	Sham	4/12 (33%)		10	6.3	19 8			
	30	<u>Overall:</u> ES ³	<u>Four-Year:</u> Good/fair improvement	n.r.						
	10	Sham	21/30 (70%) 4/10 (40%)							
	18	<u>SOM >40 mmHg</u> ES	17/18 (94%)	<0.005						
	5	Sham	2/5 (40%)							

¹ Summary score of presence of abnormal liver function tests, enlarged common bile duct (>12 mm), delayed drainage of contrast/bile (>45 minutes).

² Common bile duct dilatation (≥ 12 mm), abnormal liver function tests, or delayed drainage of contrast/bile (>45 minutes) were not statistically significant predictors of treatment response after ES; however, sample size was small limiting statistical power to detect a difference.

³ At 1-year, 17 sham subjects were considered treatment failures and were offered cross-over treatment with ES. 7 of 9 sham subjects w/ SO pressure > 40 mm Hg crossed over to ES. After 3 years follow-up, 7 of 7 (100%) were virtually symptom free. Five of 8 sham subjects w/ SO pressure <40 mmHG crossed over to ES. After 3 years follow-up, 2 of 5 (40%) showed Good or Fair improvement in pain scores.

Table 70. Randomized Controlled Trials (cont'd)

Study	N	Study Group	Improved Pain Scores	P	Mean Symptom Score	P	Objective Abnormalities ⁴	P	Complications	P
Toouli, Robert-Thomson, Kellow et al., 2000(n=79)	13	<u>SOM >40mmHg</u>	<u>2-year</u>						7 Mild pancreatitis	
	13	ES	11 (85%)	0.041					1 Perforation	
	13	Sham	5 (38%)							
	11	<u>SO Dyskinesia</u>								
	10	ES	4 (36%)	0.67						
	10	Sham	5 (50%)							
	13	<u>Normal SOM</u>								
	13	ES	8 (62%)	0.473						
	19	Sham	8 (42%)							

⁴ Summary score of presence of abnormal liver function tests, enlarged common bile duct (>12 mm), delayed drainage of contrast/bile (>45 minutes).

Table 71. Single-arm studies of results of endoscopic sphincterotomy for abdominal pain of suspected pancreaticobiliary origin

Study	N1	N2	Study Group	Improved Pain Scores	P	Objective Abnormalities ⁵	P	Complications	P
Brand, Wiese, Thonke, et al., 2001		29	29 consecutive patients with: abd pain of suspected pancreatobiliary origin. Elevated liver enzymes No other pathology on diagnostic ERCP	<u>Pre-treatment:</u> median pain score 8 (0-10) <u>Post-treatment:</u> 26/28 (93%) pts pain-free at 12wks (1 pt lost to f/u)	n.r.	Normalization of liver enzymes post-treatment: 22/29 (76%)		procedure induced pancreatitis in 1/29 pts (3%)	
Wehrmann, Wiemer, Lembcke, et al., 1996	108	33	33 of 108 consecutive pts w/ unexplained abdominal pain referred for workup 35 type II SOD - 20 got ES 29 type III SOD - 13 got ES ES performed only in those with SO pressure > 40mmHg	Mean pain score (0-10) <u>Pre-treatment</u> Type II: 7.2+/-1.4 Type III: 6.8+/-1.3 <u>Post-treatment</u> 4-6 weeks Type II: 2.3+/-2.6 Type III: 3.7+/-2.6 <u>Post-treatment</u> Median f/u 2.5 y Type II: 2.5+/-2.8 Type III: 5.1+/-2.0 Type II SOD 12/20 (60%) improved Type III SOD 1/13 (8%) improved	n.s. <0.01 <0.01	Bile duct dilatation (>9mm) Type II SOD Pre ES = 5 pts Post ES = 2 pts Type III SOD No significant changes	n.s.	Pancreatitis 15% No perforation	

⁵ Summary score of presence of abnormal liver function tests, enlarged common bile duct (>12 mm), delayed drainage of contrast/bile (>45 minutes).

Table 71. Single-arm studies of results of endoscopic sphincterotomy for abdominal pain of suspected pancreaticobiliary origin (cont'd)

Study	N1	N2	Study Group	Improved Pain Scores	P	Objective Abnormalities ⁶	P	Complications	P
Botoman, Kozarek, Novell, et al., 1994 ⁷		19 16	SO Pressure ≥ 40 mm Hg Type II Type III	Mean f/u 3.1 y 13/19 (68%) 9/16 (56%)	n.s.				
Choudhry, Ruffolo, Jamidar, et al., 1993		35	SO Pressure >40 mmHg	1 Month 43% pain-free 34% good 0% fair 23% no response During follow-up 56% of responders stayed well 44% relapsed					
		1 18 16	SO Pressure ≥ 40 mmHg Type I Type II Type III	0% 38% 56%	>0.05				

⁶ Summary score of presence of abnormal liver function tests, enlarged common bile duct (>12 mm), delayed drainage of contrast/bile (>45 minutes).

⁷ Common bile duct dilatation (≥ 12 mm) and presence of cholecystectomy were not statistically significant predictors of treatment response after ES; however, sample size was small limiting statistical power to detect a difference.

Table 71. Single-arm studies of results of endoscopic sphincterotomy for abdominal pain of suspected pancreaticobiliary origin (cont'd)

Study	N1	N2	Study Group	Improved Pain Scores	P	Objective Abnormalities ⁸	P	Complications	P
Thatcher, Sivak, Tedesco, et al., 1987 ⁹	34	31	Group 1 ¹⁰	Pain-free at 3-months n=N2 27/31 (87%)	n.r.			N=N1 4 perforations 2 pancreatitis 2 hemorrhage	
	17	15	Group 2	10/15 (67%)					
			Group 1	Pain free at 12-months 25/31 (81%)	n.r.				
			Group 2	7/15 (47%)					
			Group 1	Pain free at Last evaluation Mean f/u=12.5 m 24/31 (77%)	0.05				
			Group 2	Mean f/u=20.3 m 7/15 (47%)					

⁸ Summary score of presence of abnormal liver function tests, enlarged common bile duct (>12 mm), delayed drainage of contrast/bile (>45 minutes).

⁹ Statistically significant associations were noted between satisfactory response to ES and dilated CBD (p=0.02), delayed drainage of contrast (p=0.04), and combination of both of these (p=0.01). No significant association was seen for abnormal manometry or abnormal biochemical parameters.

¹⁰ Group 1 (roughly similar to Type II) had “a dilated bile duct and a clinical history compatible with sphincter dysfunction. These patients had evidence of bile duct obstruction which was defined as either a dilated common bile duct (CBD) at ERCP or CT scan (greater than 12 mm in diameter) and/or delayed drainage of contrast material (greater than 45 min in the absence of a gallbladder).” Group 2 (roughly similar to Type III) “did not have CBD dilation or delayed contrast drainage at ERCP. The sphincter of Oddi dysfunction was based on a typical history combined with abnormal sphincter of Oddi manometry.”

improvement in pain, as compared with 38 percent in the sham group ($p < 0.04$). In patients without an elevated sphincter of Oddi pressure, both studies reported that the improvement in pain scores was not statistically significant for the endoscopic sphincterotomy group as compared to the sham group.

Geenen, Hogan, Dodds et al. (1989) reported the number of patients with objective abnormalities post treatment. At 1 year, objective abnormalities were found in 16 percent of patients in the endoscopic sphincterotomy group and 61 percent of patients in the sham group. Statistical tests were not reported for this comparison. This study also allowed crossover from sham to endoscopic sphincterotomy after one year and continued to follow patients for up to four years. After four years, the improvement in pain scores was maintained for the endoscopic sphincterotomy group. The patients who crossed over from sham to endoscopic sphincterotomy had similar outcomes as the initial endoscopic sphincterotomy group.

Review of Evidence: Nonrandomized Controlled Trials

Five nonrandomized studies reported outcomes of endoscopic sphincterotomy in patients with abdominal pain of suspected pancreaticobiliary origin (Table 71). Brand, Wiese, Thonke, et al. (2001) was a prospective single-arm study that reported quantifiable pre and post values for pain. This study treated 29 consecutive patients with biliary-type pain, increased liver enzymes, and no evidence of other pancreatobiliary pathology with ERCP and endoscopic sphincterotomy. At 12 weeks post-treatment, 26 of the remaining 28 patients available for follow-up were pain-free, and all 26 patients remained pain-free after a median follow-up of 19 months. Wehrmann, Wiemer, Lembcke, et al. (1996) prospectively compared the results after endoscopic sphincterotomy in 20 patients with Type II SOD and 13 patients with Type III SOD. After a median of 2.5 years follow-up, 60 percent of the Type II SOD patients and only 8 percent of the Type III SOD patients maintained symptomatic relief.

The 3 retrospective single-arm studies compare outcomes among subgroups of patients who underwent ERCP and endoscopic sphincterotomy (Botoman, Kozarek, Novell, et al., 1994; Choudhry, Ruffolo, Jamidar, et al., 1993; Thatcher, Sivak, Tedesco, et al., 1987). In particular, these studies explore the relationship between improvement in pain following endoscopic sphincterotomy, baseline sphincter of Oddi pressure, and/or the presence of a dilated common bile duct. Because of the retrospective, uncontrolled nature of these studies, they do not provide strong data on the absolute improvement seen following treatment with endoscopic sphincterotomy. However, comparison of outcomes among clinical subgroups in these studies may provide useful information regarding the relative success of this treatment in different patient groups.

Among all patients treated with endoscopic sphincterotomy, these studies report good/fair improvement in over 60 percent. The presence of baseline sphincter of Oddi pressure greater than 40 mm Hg, a dilated common bile duct and/or delayed common bile duct emptying appear to be associated with slightly higher success rates after endoscopic sphincterotomy. However, confidence in this conclusion is limited by the small numbers of patients in the subgroup analyses, and the lack of tests of statistical significance in some cases.

Conclusions

The randomized controlled trials by Geenen, Hogan, Dodds et al. (1989) and Toouli, Robert-Thomson, Kellow et al. (2000) provide strong and consistent evidence that endoscopic sphincterotomy provides effective relief of pain in patients with pancreaticobiliary pain, sphincter of Oddi dysfunction, and elevated basal sphincter of Oddi pressure on manometry (greater than 40 mm Hg). The results of the nonrandomized studies 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 for the endoscopic sphincterotomy group.

ERCP Evidence Review Results and Conclusions, Part V: Patient, Procedure or Operator Determinants of ERCP Complications

This chapter reviews evidence on the following questions:

What patient, procedure, or provider factors are determinants of adverse events of ERCP?

(Section 1: Multivariable Analyses)

(Section 2: Randomized, Controlled Comparison Trials)

Part V, Section 1: Multivariable Analyses

Body of Evidence

Thirteen studies reported on multivariable logistic regression analyses of factors associated with complications of ERCP (Table 72; see also “Evidence Tables” chapter). The four largest studies each included more than 1,800 patients, and the total number of complications observed in these studies ranged from 98 to 229 (Loperfido, Angelini, Benedetti, et al., 1998; Freeman, DiSario, Nelson, et al., 2001; Freeman, Nelson, Sherman, et al., 1996; Masci, Toti, Mariani, et al., 2001). The remaining 9 studies ranged from 100 to 535 patients, and the number of complications observed ranged from 10–34. Seven studies reported on therapeutic ERCP, 5 studies combined therapeutic and diagnostic ERCP, and one study reported on diagnostic ERCP.

Total complications were analyzed in seven studies. The specific complications most commonly analyzed separately were pancreatitis (7 studies) and hemorrhage (4 studies). The number of cases of pancreatitis observed ranged from 17 to 131; and cases of hemorrhage ranged from 10 to 48. Other complications analyzed separately in these studies include cholangitis, septicemia, and retroperitoneal perforation, with number of cases observed ranging from 10 to 34.

This systematic review addresses the relationship of patient, procedure, and operator factors to complications. The 13 included studies assessed numerous factors suspected to be related to the likelihood of complications. The various measures used in the literature were classified into categories. There are 12 categories for patient factors, 13 for procedure factors; and 4 categories for operator factors. Independent variables reported to be statistically significant risk factors for complications are listed for each study along with an estimate of the magnitude of the effect when available (i.e., odds ratio and confidence interval). Independent variables that were considered in the study but not found to be significantly associated with complications are denoted by an “X” under the appropriate category for that factor.

Table 72. Overview Table

Study	N Pts	Pop	Patient Factors	Procedure Factors	Operator Factors	Outcomes Analyzed
Fair Quality						
Masci, Toti, Mariani, et al., 2001	2444	M	X	X		Total complications (121) Pancreatitis (44) Hemorrhage (30)
Freeman, DiSario, Nelson, et al., 2001	1963	M	X	X	X	Pancreatitis (131)
Freeman, Nelson, Sherman, et al., 1996	2347	T (ES)	X	X	X	Total complications (229) Pancreatitis (127) Hemorrhage (48)
Fair Minus Quality						
Rabenstein, Schneider, Bulling, et al., 2000	438	T (ES)	X	X	X	Total complications (33) Pancreatitis (19)
Loperfido, Angelini, Benedetti, et al., 1998	1827	T ¹	X	X	X	Total complications (98) Pancreatitis (29) Hemorrhage (21) Cholangitis (21) Retroperitoneal perforation (12)
Mehta, Pavone, Barkun, et al., 1998	535	M	X	X		Pancreatitis (34)
Neoptolemos, Shaw, and Carr-Locke, 1989	190	T (ES)	X			Total complications (32)
Motte, Deviere, Dumonceau, et al., 1991	105	T (ST)	X	X		Septicemia (34)
Tzovaras, Shukla, Kow, et al., 2000	372	M	X	X		Total complications (21)
Lai, Lo, Choi, et al., 1989	323	D	X			Acute cholangitis (21)
Boender, Nix, de Ridder, et al., 1994	242	T (ES)	X	X		Total complications (34)
Nelson and Freeman, 1994	189	T (ES)	X	X		Hemorrhage (10)
Maldonado, Brady, Mamel, et al., 1999	100	M ²	X	X		Pancreatitis (17)

¹ Loperfido included a broad population of both diagnostic and therapeutic ERCP. However, multivariate analysis of risk factors was reported only for therapeutic subpopulation.

² Maldonado was restricted to a specific population with suspected sphincter of Oddi dysfunction who were undergoing sphincter of Oddi manometry

Study Quality

The number of events observed is the primary determinant of the power of a study to detect a significant association between a factor and an outcome of interest. When multivariable analysis is performed, the number of events also constrains the number of potential relationships that can be appropriately tested. A commonly accepted benchmark is a minimum of 10 outcome events per independent variable tested. A larger number of variables relative to events can lead to unstable results, spurious findings of significance, and unreliable estimates of the magnitude of the association. Extremely wide confidence intervals are a hallmark of such “overfitted” models. Another problem is that when multiple variables are incorporated in a model, some may be highly correlated. As a result, some independently significant factors can be obscured. Concato, Feinstein, and Holford (1993) offer an overview of the methodologic deficiencies that are common in multivariable analyses published in the medical literature.

Overall, the multivariable analyses included in this systematic review demonstrated overfitting, i.e., testing an excessive number of factors relative to the number of complications observed. Consequently, this literature is exploratory in nature. Candidate variables included in the analyses are often likely to be closely related to each other (potentially leading to collinearity) resulting in potentially spurious results from multivariable analysis including all variables. Instances where multiple factors identified to be highly associated with complications on univariate analysis disappear entirely from the multivariable models raises concern over the stability of the findings. Reported magnitudes of association are not reliable, significant independent variables may have been overlooked, and some significant associations may be misleading. Moreover, the existing studies do not use common, standardized definitions for the complications and factors of interest. Thus, caution should be used in drawing inferences for clinical practice from these studies.

This body of literature was overall rated as “Fair” (Table 73). The associations found in these analyses are hypothesis generating, but not predictive. The three studies with notably larger numbers of cases of complications (121–229 vs. 10–98) were designated as “Fair” quality for purposes of this review (Freeman, DiSario, Nelson, et al., 2001; Freeman, Nelson, Sherman, et al., 1996; Masci, Toti, Mariani, et al., 2001) while the remaining 10 studies were rated “Fair Minus.” The results of the three “Fair” studies are slightly more robust, despite some degree of overfitting. The study by Loperfido, Angelini, Benedetti, et al. (1998) had 98 cases, but was classified as “Fair Minus” because confidence intervals were not reported and problems with missing data were noted.

This review focuses on factors that were found to be significant either in the more robust studies or in several studies. Also, factors are noted that were found to be not significant in all analyses. Rarely was a factor found to be significant in all studies in which it was analyzed; which is not surprising given the characteristics of the available studies. Extremely wide confidence intervals also are noted, which may suggest a spurious association.

Table 73. Quality Assessment

Study	N	No. candidate variables	Total complications	Pancreatitis	Hemorrhage	Cholangitis	Retroperitoneal perforation	Septicemia	Ratio of group size/# variables	Degree of Overfitting	Statistical reporting	Internal validity	Overall Quality Rating
Masci, Toti, Mariani, et al., 2001	2444	16	121	44	30	--	--	--	7.6 – 1.9	Mild to Severe	S	No	Fair
Freeman, DiSario, Nelson, et al., 2001	1963	32	--	131	--	--	--	--	4.1	Moderate	S	No	Fair
Freeman, Nelson, Sherman, et al., 1996	2347	22	229	127	48	--	--	--	10.4 - 2.2	Satisfactory to Severe	S	No	Fair
Rabenstein, Schneider, Bulling, et al., 2000	438	26	33	19	--	--	--	--	1.3 - 0.7	Severe	S	No	Fair Minus
Loperfido, Angelini, Benedetti, et al., 1998	1827	13	98	29	21	21	12	--	7.5 - 0.9	Mild to Severe	U	No	Fair Minus
Mehta, Pavone, Barkun, et al., 1998	535	9	--	34	--	--	--	--	3.7	Severe	U	No	Fair Minus
Neoptolemos, Shaw, and Carr-Locke, 1989	190	19	32	--	--	--	--	--	1.7	Severe	U	No	Fair Minus
Motte, Deviere, Dumonceau, et al., 1991	105	13	--	--	--	--	--	34	2.6	Severe	U	No	Fair Minus
Tzovaras, Shukla, Kow, et al., 2000	372	16	21	--	--	--	--	--	1.3	Severe	S	No	Fair Minus
Lai, Lo, Choi, et al., 1989	323	9	--	--	--	21	--	--	2.3	Severe	S	No	Fair Minus

Table 73. Quality Assessment (cont'd)

Study	N	No. candidate variables	Total complications	Pancreatitis	Hemorrhage	Cholangitis	Retroperitoneal perforation	Septicemia	Ratio of group size/# variables	Degree of Overfitting	Statistical reporting	Internal validity	Overall Quality Rating
Boender, Nix, de Ridder, et al., 1994	242	9	34	--	--	--	--	--	3.7	Severe	S	No	Fair Minus
Nelson and Freeman, 1994	189	7	--	--	10	--	--	--	0.14	Severe	S	No	Fair Minus
Maldonado, Brady, Mamel, et al., 1999	100	9	--	17	--	--	--	--	1.9	Severe	U	No	Fair Minus

Explanation of categorization:

Degree of Overfitting assessed using the ratio of number of endpoints over number of candidate variables: Satisfactory, ratio ≥ 10 ; Mild, ratio – 7 to 10; Moderate, ratio 4-7; Severe, ratio <4 .

Statistical reporting: S=satisfactory, reported both magnitude of effect estimates as well as associated confidence intervals or p-value for statistically significant findings; U = unsatisfactory, did not report both magnitude of effect estimate and statistical significance information for statistically significant findings.

Internal validity: Yes = the study used procedures (e.g., test-validation split samples or bootstrapping) to guard against overfitting the model and spurious results; No = the study did not utilize such procedures

Quality Rating:

Good = use of procedures to guard against overfitting the model and spurious results, degree of overfitting not severe for at least one analysis, and satisfactory statistical reporting

Fair = degree of overfitting not severe for at least one analysis, satisfactory statistical reporting, but no use of procedures to guard against overfitting the model and spurious results.

Fair Minus = Severe degree of overfitting

Review of Evidence: Patient Factors

All 13 studies reported on patient factors associated with complications. These various factors were classified into 12 categories: age, gender, common bile duct size/diameter, cholangitis, anatomic variation, coagulopathy, laboratory values, comorbidities, indication for ERCP procedure, previous gastrectomy, history of jaundice, and history of allergy to contrast media.

Total Complications

Seven studies reported on total complications (Table 74). Two factors were found to be significant in a study rated as “Fair” and in one additional study. These were age equal to or less than 60 years (Masci, Toti, Mariani, et al., 2001; Rabenstein, Schneider, Bulling, et al., 2000) and suspected sphincter of Oddi dysfunction (Freeman, Nelson, Sherman, et al., 1996; Tzovaras, Shukla, Kow, et al., 2000).

Jaundice of malignancy was significant in the study by Tzovaras, Shukla, Kow, et al. (2000) and elevated serum bilirubin in Neoptolemos, Shaw, and Carr-Locke (1989). Factors found to be significant in a single study rated as “Fair Minus” were: pancreas divisum, coagulopathy, pancreatic obstruction (Rabenstein, Schneider, Bulling, et al., 2000), and juxtapapillary diverticulum (Boender, Nix, de Ridder, et al., 1994). However, confidence intervals were extremely wide for pancreas divisum (1.56–36.6) and coagulopathy (1.95–48.1).

The following factors were analyzed, but were not found to be significant for total complications in any study: gender (6 studies); common bile duct size/diameter (4 studies); cholangitis (2 studies); previous gastrectomy (3 studies);

Pancreatitis

Seven studies reported on patient factors associated with pancreatitis (Table 75). Younger age was significant in four studies, two rated as “Fair” quality. Each of the four studies used a different age cut-off: 70 years in Loperfido, Angelini, Benedetti, et al. (1998); 60 years in Masci, Toti, Mariani, et al. (2001); 59 years in Mehta, Pavone, Barkun, et al., (1998); and 30 years vs. 70 years in Freeman, Nelson, Sherman, et al. (1996). Suspected sphincter of Oddi dysfunction was significant in two studies, both rated “Fair” (Freeman, Nelson, Sherman, et al., 1996; Freeman, DiSario, Nelson, et al., 2001). Note that the two studies by Freeman and co-workers included different patient populations.

Factors found to be significant in a single study rated “Fair” (Freeman, DiSario, Nelson, et al., 2001) were: normal bilirubin, female gender, absence of chronic pancreatitis, and history of post-ERCP pancreatitis.

Factors found to be significant in a single study rated as “Fair Minus” were: absence of a common bile duct stone at ERCP (Mehta, Pavone, Barkun, et al., 1998); and pancreas divisum, but with an extremely wide (1.91-34.79) confidence interval (Rabenstein, Schneider, Bulling, et

Table 74. Relationship between Patient Factors and Total Complications³

Study	N Pts Cx	Age	Gender	CBD Size\ Diameter	Cholangitis	Anatomic variation/ features ⁴	Coagulopathy ⁵	Laboratory values	Other ⁶ Comorbidities	Indication for ERCP proc ⁷	Previous Gastrectomy	Hx Jaundice	Hx Contrast Allergy
Fair Quality													
Masci, Toti, Mariani, et al., 2001	2444 121	Age ≤60 years OR=1.53 (1.06-2.2)	X	X		X Stone size Papilla features GB stones				X			
Freeman, Nelson, Sherman, et al., 1996	2347 229	X	X	X	X	X	X		Cirrhosis OR=2.93 (1.48- 5.90)	Susp. SOD OR=2.9 (1.70-4.94) All pts had ES	X		
Fair Minus Quality													
Rabenstein, Schneider, Bulling, et al., 2000	438 33	Age ≤60 years OR=2.9 (1.33-6.21)	X			Pancreas divisum OR=7.6 (1.56- 36.6)	Coagulopathy OR=9.7 (1.95- 48.10)		X	Pancreatic obstruction OR=0.07 (0.01-0.59) All pts had ES	X		

³ Independent variables reported to be statistically significant risk factors for complications are listed for each study along with an estimate of the magnitude of the effect when available (i.e., odds ratio and confidence interval). Independent variables that were considered in the study but not found to be significantly associated with complications are denoted by an "X" under the appropriate category for that factor

⁴ Summary of pancreas divisum, juxtapapillary diverticulum, acinarization

⁵ Summary of related factors – anticoagulation, coagulopathy, PT time, ASA/NSAID use, bleeding

⁶ "Comorbidities" includes reports of cirrhosis diabetes, anemia, hemodialysis etc.

⁷ Summary of related factors - Pancreatitis or Obstruction, sphincter of Oddi dysfunction, indication of than bile duct stone

Table 74. Relationship between Patient Factors and Total Complications (cont'd)

Study	N Pts Cx	Age	Gender	CBD Size\Diameter	Cholangitis	Anatomic variation/ features ⁸	Coagulopathy ⁹	Laboratory values	Other ¹⁰ Comorbidities	Indication for ERCP proc ¹¹	Previous Gastrectomy	Hx Jaundice	Hx Contrast Allergy
Fair Minus Quality													
Loperfido, Angelini, Benedetti, et al., 1998	1827 98	X	X	X		X					X	X	
Neoptolemos, Shaw, and Carr-Locke, 1989	190 32	X	X		X		X	elevated bilirubin elevated serum albumin	X	X All pts had ES			
Tzovaras, Shukla, Kow, et al., 2000	372 21	X	X							Suspected SOD OR=8.57 (2.59- 28.43); Malignant jaundice OR=4.76 (1.46-15.58)			

⁸ Summary of pancreas divisum, juxtapapillary diverticulum, acinarization

⁹ Summary of related factors – anticoagulation, coagulopathy, PT time, ASA/NSAID use, bleeding

¹⁰ “Comorbidities” includes reports of cirrhosis diabetes, anemia, hemodialysis etc.

¹¹ Summary of related factors - Pancreatitis or Obstruction, sphincter of Oddi dysfunction, indication of than bile duct stone

Table 74. Relationship between Patient Factors and Total Complications (cont'd)

Study	N Pts Cx	Age	Gender	CBD Size\Diameter	Cholangitis	Anatomic variation/ features ¹²	Coagulopathy ¹³	Laboratory values	Other ¹⁴ Comorbidities	Indication for ERCP proc ¹⁵	Previous Gastrectomy	Hx Jaundice	Hx Contrast Allergy
Fair Minus Quality													
Boender, Nix, de Ridder, et al., 1994	242 34	X		X		JPD Outside OR=3.1 (p=.072) Lower rim OR=4.3 (p=.015) Inside OR=9.4 (p=.002) Presence of GB NS				All pts had ES			

¹² Summary of pancreas divisum, juxtapapillary diverticulum, acinarization

¹³ Summary of related factors – anticoagulation, coagulopathy, PT time, ASA/NSAID use, bleeding

¹⁴ “Comorbidities” includes reports of cirrhosis diabetes, anemia, hemodialysis etc.

¹⁵ Summary of related factors - Pancreatitis or Obstruction, sphincter of Oddi dysfunction, indication of than bile duct stone

Table 75. Relationship between Patient Factors and Pancreatitis

Study	N Pts Cx	Age	Gender	CBD Size\Diameter	Cholangitis	Anatomic variation/ features ¹	Coagulopathy ²	Laboratory values	Other ³ Comorbidities	Indication for ERCP proc ⁴	Previous Gastrectomy	Hx Jaundice	Hx Contrast Allergy
Fair Quality													
Masci, Toti, Mariani, et al., 2001	2444 44	Age ≤60y OR=2.11 (1.16-3.8)	X	X		X				X			
Freeman, DiSario, Nelson, et al., 2001	1963 131	X	Female OR=2.51 (1.49- 4.24)	X		X		Normal bilirubin OR=1.89 (1.22- 2.93)	Absence of CP OR=1.87 (1.00-3.48) Hx post- ERCP pancreatitis OR=5.35 (2.97-9.66)	Susp. SOD OR=2.6 (1.59- 4.26)			
Freeman, Nelson, Sherman, et al., 1996	2347 127	Age 30 vs. Age 70y OR=2.14 (1.41- 3.25)	X	X	X	X	X		X	Susp. SOD OR=5.01 (2.73- 9.22)	X		
Fair Minus Quality													
Rabenstein, Schneider, Bulling, et al., 2000	438 19	X	X			Pancreas divisium OR=8.2 (1.91- 34.79)	X		X	X	X		

Table 75. Relationship between Patient Factors and Pancreatitis (cont'd)

Study	N Pts Cx	Age	Gender	CBD Size\Diameter	Cholangitis	Anatomic variation/ features ¹	Coagulopathy ²	Laboratory values	Other ³ Comorbidities	Indication for ERCP proc ⁴	Previous Gastrectomy	Hx Jaundice	Hx Contrast Allergy
Fair Minus Quality													
Loperfido, Angelini, Benedetti, et al., 1998	1827 29	Age <70 OR=1.11 n.r.	X	Nondilated duct OR=2.85 n.r.		X						X	
Mehta, Pavone, Barkun, et al., 1998	535 34	Age <59 years (p=0.04)	X	X		Absence of a CBD stone at ERCP (p=0.004)		X	X History of pancrea- titis	X Pre-lap choly			
Maldonado, Brady, Mamel, et al., 1999	100 17	X	X							X			

al., 2000). Loperfido, Angelini, Benedetti, et al. (1998) found non-dilated duct to be significant, but did not report the confidence interval.

Previous gastrectomy was analyzed in two studies, but was not significant.

Hemorrhage

Four studies reported on patient factors associated with hemorrhage (Table 76). Coagulopathy was significant in a study rated as “Fair” (Freeman, Nelson, Sherman, et al., 1996), prothrombin time and hemodialysis (Nelson and Freeman, 1994) were significant in one additional study. Factors found to be significant in a single study rated as “Fair” were: cholangitis (Freeman, Nelson, Sherman, et al., 1996), and obstructed papilla of Vater orifice (Masci, Toti, Mariani, et al., 2001).

Factors that were not significant in any analysis were: age (3 studies), gender (3 studies); common bile duct size/diameter (4 studies); indications for ERCP (3 studies); previous gastrectomy (2 studies); and history of jaundice (1 study).

Cholangitis

Two studies, both rated as “Fair Minus” quality, reported on patient factors associated with cholangitis (Table 77). Loperfido, Angelini, Benedetti, et al. (1998) reported that jaundice had a significant association with cholangitis. Lai, Lo, Choi, et al. (1989) reported significant associations for fever greater than 37.5 degrees Celsius within prior 72 hours; malignant obstruction; and serum AST of 70 IU or less.

The study by Loperfido, Angelini, Benedetti, et al. (1998) also included age, gender, common bile duct size and diameter, anatomic features, and previous gastrectomy in the analysis, but none were significant.

Septicemia and Retroperitoneal Perforation

Septicemia (Table 78) and retroperitoneal perforation (Table 79) were each addressed in a single study of “Fair Minus” quality.

Motte, Deviere, Dumonceau, et al. (1991) reported that prior cholangitis and elevated white blood count were significant factors for septicemia, but did not report p-values. Age, gender, anatomic variation, other comorbidities, and history of jaundice were not significant in this analysis.

Loperfido, Angelini, Benedetti, et al. (1998) reported that previous gastrectomy was a significant factor for retroperitoneal perforation, but did not report confidence intervals. Age, gender, common bile duct size/diameter; anatomic variation, and history of jaundice were not significant in this analysis.

Table 76. Relationship between Patient Factors and Hemorrhage

Study	N Pts Cx	Age	Gender	CBD Size\Diameter	Cholangitis	Anatomic variation/ features ¹	Coagulopathy ²	Laboratory values	Other ³ Comorbidities	Indication for ERCP proc ⁴	Previous Gastrectomy	Hx Jaundice	Hx Contrast Allergy
Fair Quality													
Masci, Toti, Mariani, et al., 2001	2444 30	X	X	X		Obstructed orifice of papilla of Vater OR=2.57 (1.69-6.17)				X			
Freeman, Nelson, Sherman, et al., 1996	2347 48	X	X	X	OR=2.59 (1.38-4.86)	X	OR=3.32 (1.54-7.18)		X	X	X		
Fair Minus Quality													
Loperfido, Angelini, Benedetti, et al., 1998	1827 21	X	X	X		X					X	X	
Nelson and Freeman, 1994	189 10			X			Prothrombin time 2x > control OR=12.1 (1.8-90.9)		Hemodial ysis OR=16.4 (2.9- 93.1)	X All pts had ES			

Table 77. Relationship between Patient Factors and Cholangitis

Study	N Pts Cx	Age	Gender	CBD Size\Diameter	Cholangitis	Anatomic variation/ features ¹	Coagulopathy ²	Laboratory values	Other ³ Comorbidities	Indication for ERCP proc ⁴	Previous Gastrectomy	Hx Jaundice	Hx Contrast Allergy
Fair Minus Quality													
Loperfido, Angelini, Benedetti, et al., 1998	1827 21	X	X	X		X					X	OR=4.14	
Lai, Lo, Choi, et al., 1989	323 21							Subgroup analysis excluding 43 febrile patients Serum AST ≤70IU (discriminant coefficient= 2.09, p<0.04)	Fever (>37.5° C) within 72 hours prior to examination (discriminant coefficient= 2.73, p<0.0001)	Pathologic nature of the obstructive lesion, malignant vs. benign (discriminant coefficient= 1.75, p<0.002)			

Table 78. Relationship between Patient Factors and Septicemia

Study	N Pts Cx	Age	Gender	CBD Size\Diameter	Cholangitis	Anatomic variation/ features ¹	Coagulopathy ²	Laboratory values	Other ³ Comorbidities	Indication for ERCP proc ⁴	Previous Gastrectomy	Hx Jaundice	Hx Contrast Allergy
Fair Minus Quality													
Motte, Deviere, Dumonceau, et al., 1991	105 34	X	X		Prior Cholangitis (F=7.1)	X		WBC count (F=6.6) Alk Phos n.s.	X			X	

Table 79. Relationship between Patient Factors and Retroperitoneal Perforation

Study	N Pts Cx	Age	Gender	CBD Size\Diameter	Cholangitis	Anatomic variation/ features ¹	Coagulopathy ²	Laboratory values	Other ³ Comorbidities	Indication for ERCP proc ⁴	Previous Gastrectomy	Hx Jaundice	Hx Contrast Allergy
Fair Minus Quality													
Loperfido, Angelini, Benedetti, et al., 1998	1827 12	X	X	X		X					OR=11.7 n.r.	X	

Relationship of Total and Specific Complications

Pancreatitis and hemorrhage together comprise the majority of total complications in the three studies that report all 3 outcomes (Masci, Toti, Mariani, et al., 2001; Freeman, Nelson, Sherman, et al., 1996; Loperfido, Angelini, Benedetti, et al., 1998). Pancreatitis was 36 percent, 55 percent, and 30 percent, respectively in these studies; and hemorrhage was 25 percent, 21 percent and 21 percent.

In the study by Masci, Toti, Mariani, et al. (2001), younger age was a significant factor for both pancreatitis and total complications. There was no other overlap between risk factors for total complications and pancreatitis or hemorrhage.

In Freeman, Nelson, Sherman, et al. (1996), suspected sphincter of Oddi dysfunction was a significant factor for both pancreatitis and total complications. There was no other overlap between total complications and pancreatitis or hemorrhage. In contrast to Masci, Toti, Mariani, et al. (2001), younger age was significant only for pancreatitis, not for total complications.

Loperfido, Angelini, Benedetti, et al. (1998) found no significant relationships between patient factors and overall complications.

The inconsistencies noted here might suggest that analysis of patient factors related to specific complications may be more informative than total complications. Analysis of total complications may not be sufficiently sensitive. This suggests that large studies with adequate numbers of cases of the specific complications of interest will be more useful in identifying patient-related factors that might be used to improve clinical outcomes.

Review of Evidence: Procedure Factors

Eleven studies reported on patient factors associated with complications. The various measures were classified into 13 categories: papillotomy/endoscopic sphincterotomy; pre-cut endoscopic sphincterotomy; biliary drainage; failed procedure; length of endoscopic sphincterotomy; bleeding during endoscopic sphincterotomy; combination with other procedures; difficulty of cannulation; pancreatic opacification; post-procedure care; intramural injection; sphincter of Oddi manometry; emergency procedure.

Total Complications

Six studies reported on procedure factors associated with total complications (Table 80). Precut endoscopic sphincterotomy was significant in all four studies that tested for this association; including two studies rated as “Fair” (Masci, Toti, Mariani, et al., 2001; Freeman, Nelson, Sherman, et al., 1996). Freeman, Nelson, Sherman, et al. (1996) also found two additional significant factors, combined percutaneous-endoscopic procedures and difficulty in cannulation. Masci, Toti, Mariani, et al. (2001) found that failed stone removal, another indicator of a difficult procedure, was a significant factor for total complications.

Table 80. Relationship between Procedure Factors and Total Complications

Study	N Pts Cx	Standard Papillotomy/ ES	Precut ES	Biliary drainage	Failed Procedure	ES length	Bleeding during ES	Combined with other procedures	Difficulty of cannulation	Pancreatic opacification	Postprocedure Care	Sphincter Manometry	Emergency procedure
Fair Quality													
Masci, Toti, Mariani, et al., 2001	2444 121		OR=1.70 (1.10-2.68)	X	No stone removal OR=2.52 (1.44- 4.53)					X			
Freeman, Nelson, Sherman, et al., 1996	2347 229	All pts had ES	OR=3.61 (1.78-7.34)		X		X	Comb. percut.- endo. proc. OR=3.40 (1.04-11.13)	OR=3.05 (1.83- 5.08)				X
Fair Minus Quality													
Rabenstein, Schneider, Bulling, et al., 2000	438 33	All pts had ES			X								X
Loperfido, Angelini, Benedetti, et al., 1998	1827 98		OR=1.73							X			X
Tzovaras, Shukla, Kow, et al., 2000	372 21				Previous failed ERCP OR=4.66 (1-21.80)			Need for PTC OR=10.3 (2.30-45.83)				X	X
Boender, Nix, de Ridder, et al., 1994	242 34	All pts had ES	OR=4.9 p=0.001	X	Failed biliary drainage OR=34.8 p=0.007	X							

Failed biliary drainage was significant in the study by Boender, Nix, de Ridder, et al. (1994). Tzovaras, Shukla, Kow, et al. (2000) reported two significant factors: previous failed ERCP (CI=1–21.8) and need for percutaneous procedure (CI=2.3–45.8); but confidence intervals were extremely wide for both factors.

Factors not significant were: emergency procedure (4 studies); pancreatic opacification (2 studies); and bleeding during endoscopic sphincterotomy (1 study).

Pancreatitis

Seven studies reported on procedure factors associated with pancreatitis (Table 81). Precut endoscopic sphincterotomy was significant in two studies rated as “Fair” (Masci, Toti, Mariani, et al., 2001; Freeman, Nelson, Sherman, et al., 1996); as was difficulty in cannulation and multiple pancreatic contrast injections (Freeman, Nelson, Sherman, et al., 1996 and Freeman, DiSario, Nelson, et al., 2001). Multiple pancreatic contrast injections was also a significant risk factor in Loperfido, Angelini, Benedetti, et al. (1998); and in Mehta, Pavone, Barkun, et al. (1998) for the subgroup of patients that did not undergo endoscopic sphincterotomy.

Masci, Toti, Mariani, et al. (2001) also reported that failed stone removal was a significant factor; and Freeman, DiSario, Nelson, et al. (2001) found that pancreatic sphincterotomy and balloon biliary sphincter dilatation were also significant factors.

Maldonado, Brady, Mamel, et al. (1999) identified performing a complete ERCP procedure in addition to sphincter of Oddi manometry as a significant risk factor for pancreatitis among patients who all underwent sphincter of Oddi manometry.

Factors not significant were: emergency procedure (3 studies); biliary drainage (1 study); and bleeding during endoscopic sphincterotomy (1 study).

Hemorrhage

Four studies reported on procedure factors associated with hemorrhage (Table 82). Bleeding during endoscopic sphincterotomy was significant in two studies, one of which was rated as “Fair” (Freeman, Nelson, Sherman, et al., 1996; Nelson and Freeman, 1994). Precut endoscopic sphincterotomy (Masci, Toti, Mariani, et al., 2001) and anticoagulation less than 3 days after procedure (Freeman, Nelson, Sherman, et al., 1996) were significant in a single study rated “Fair.”

Factors not significant were: pancreatic opacification (3 studies) emergency procedure (2 studies); combined with other procedures (2 studies); biliary drainage (1 study); failed procedure (1 study); endoscopic sphincterotomy length (1 study); and difficulty of cannulation (1 study).

Cholangitis, Septicemia and Retroperitoneal Perforation

Cholangitis (Table 83), septicemia (Table 84) and retroperitoneal perforation (Table 85) were each addressed in a single study of “Fair Minus” quality.

Table 81. Relationship between Procedure Factors and Pancreatitis

Study	N Pts Cx	Standard Papillotomy/ ES	Precut ES	Biliary drainage	Failed Procedure	ES length	Bleeding during ES	Combined with other procedures	Difficulty of cannulation	Pancreatic opacification	Postprocedure Care	Sphincter Manometry	Emergency procedure
Fair Quality													
Masci, Toti, Mariani, et al., 2001	2444 44		OR=2.8 (1.38-5.84)	X	No stone removal OR=3.35 (1.33- 9.1)					X			
Freeman, DiSario, Nelson, et al., 2001	1963 131	Pancreatic ES OR=3.07 (1.64- 5.75)	X		X			Biliary Balloon Sphincter Dilation OR=4.51 (1.51- 13.46)	Moderate to Difficult OR=3.41 (2.13-5.47)	>1 pancreatic contrast injection OR=2.72 (1.43-5.17)		X	
Freeman, Nelson, Sherman, et al., 1996	2347 127	All pts had ES	OR=4.34 (1.73-10.88)		X		X	X	OR=2.4 (1.07-5.36)	OR=1.35 (1.04-1.75)			X
Fair Minus Quality													
Rabenstein, Schneider, Bulling, et al., 2000	438 19	All pts had ES			X								X
Loperfido, Angelini, Benedetti, et al., 1998	1827 29		X							OR=2.84 n.r.			X

Table 81. Relationship between Procedure Factors and Pancreatitis (cont'd)

Study	N Pts Cx	Standard Papillotomy/ ES	Precut ES	Biliary drainage	Failed Procedure	ES length	Bleeding during ES	Combined with other procedures	Difficulty of cannulation	Pancreatic opacification	Postprocedure Care	Sphincter Manometry	Emergency procedure
Fair Minus Quality													
Mehta, Pavone, Barkun, et al., 1998	535 34	X						X		Subgroup with ES n.s. Subgroup without ES p=0.05			
Maldonado, Brady, Mamel, et al., 1999	100 17	X ES no added risk						X	Length of procedure			X ERCP was risk factor but not SOM	

Table 82. Relationship between Procedure Factors and Hemorrhage

Study	N Pts Cx	Standard Papillotomy/ ES	Precut ES	Biliary drainage	Failed Procedure	ES length	Bleeding during ES	Combined with other procedures	Difficulty of cannulation	Pancreatic opacification	Postprocedure Care	Sphincter Manometry	Emergency procedure
Fair Quality													
Masci, Toti, Mariani, et al., 2001	2444 30		OR=2.45 (1.6- 5.39)	X	X					X			
Freeman, Nelson, Sherman, et al., 1996	2347 48	All pts had ES	X		X		OR=1.74 (1.15- 2.65)	X	X	X	Anticoag <3d after procedure OR=5.11 (1.57- 16.68)		X
Fair Minus Quality													
Loperfido, Angelini, Benedetti, et al., 1998	1827 21		X							X			X
Nelson and Freeman, 1994	189 10	All pts had ES				X	OR=13.7 (2.2- 87.3)						

Table 83. Relationship between Procedure Factors and Cholangitis

Study	N Pts Cx	Standard Papillotomy/ ES	Precut ES	Biliary drainage	Failed Procedure	ES length	Bleeding during ES	Combined with other procedures	Difficulty of cannulation	Pancreatic opacification	Postprocedure Care	Sphincter Manometry	Emergency procedure
Fair Minus Quality													
Loperfido, Angelini, Benedetti, et al., 1998	1827 21		X							X			X

Table 84. Relationship between Procedure Factors and Septicemia

Study	N Pts Cx	Standard Papillotomy/ ES	Precut ES	Biliary drainage	Failed Procedure	ES length	Bleeding during ES	Combined with other procedures	Difficulty of cannulation	Pancreatic opacification	Postprocedure Care	Sphincter Manometry	Emergency procedure
Fair Minus Quality													
Motte, Devriere, Dumonceau, et al., 1991	105 34			Incomplete Drainage (F=319.2)				X					

Table 85. Relationship between Procedure Factors and Retroperitoneal Perforation

Study	N Pts Cx	Standard Papilotomy/ ES	Precut ES	Biliary drainage	Failed Procedure	ES length	Bleeding during ES	Combined with other procedures	Difficulty of cannulation	Pancreatic opacification	Intramural Injection	Sphincter Manometry	Emergency procedure
Fair Minus Quality													
Loperfido, Angelini, Benedetti, et al., 1998	1827 12		OR=7.19 n.r.							X	OR=6.86		X

Loperfido, Angelini, Benedetti, et al. (1998) analyzed precut endoscopic sphincterotomy, pancreatic opacification; and emergency procedure; but none of these factors were significant for cholangitis.

Motte, Deviere, Dumonceau, et al. (1991) reported that incomplete biliary drainage was a significant factor for septicemia, but did not report p-values. Combination with another procedure was not significant in this analysis.

Loperfido, Angelini, Benedetti, et al. (1998) reported that precut endoscopic sphincterotomy and intramural injection were significant factors for retroperitoneal perforation, but did not report confidence intervals. Pancreatic opacification and emergency procedure were not significant in this analysis.

Relationship of Total and Specific Complications

Pancreatitis and hemorrhage together comprise the majority of total complications in the three studies that report all three outcomes (Masci, Toti, Mariani, et al., 2001; Freeman, Nelson, Sherman, et al., 1996; Loperfido, Angelini, Benedetti, et al., 1998).

Masci, Toti, Mariani, et al. (2001) found the precut endoscopic sphincterotomy was a significant factor for total complications, pancreatitis and hemorrhage. Failed stone removal was a significant factor for total complications and pancreatitis, but not for hemorrhage. There was no other overlap between total complications and pancreatitis or hemorrhage.

Freeman, Nelson, Sherman, et al. (1996) found that precut endoscopic sphincterotomy and difficulty in cannulation were significant factors for total complications and pancreatitis. There was no other overlap between total complications and pancreatitis or hemorrhage.

Loperfido, Angelini, Benedetti, et al. (1998) found no overlap between total complications and pancreatitis or hemorrhage.

This suggests that procedure factors may be more generalizable across total and specific complications than is the case with patient factors.

Review of Evidence: Operator Factors

Operator factors were analyzed in four studies (Freeman, DiSario, Nelson, et al., 2001; Freeman, Nelson, Sherman, et al., 1996; Loperfido, Angelini, Benedetti, et al., 1998; Rabenstein, Schneider, Bulling, et al., 2000); two of which were rated as “Fair” quality (Table 86). Case volume was analyzed in all four studies; participation of a trainee in three studies; university affiliated center in one study and center size in one study. Only case volume was a significant factor for complications in any of these analyses. Importantly, cut-off points for classification as a low-volume operator varied significantly across studies. Freeman, Nelson, Sherman, et al. (1996) used a cut-off of centers with 1 or fewer procedures per endoscopist per week; Loperfido, Angelini, Benedetti, et al. (1998) defined lower volume centers as those with fewer than 200 procedures per year.

Table 86. Relationship between Operator Factors and Total Complications

Study	N Pts Cx	Case volume	Participation of a trainee	University affiliated center	Center size
Fair Quality					
Freeman, Nelson, Sherman, et al., 1996	2347 229	X ¹⁶	X	X	
Fair Minus Quality					
Rabenstein, Schneider, Bulling, et al., 2000	438 33	X	X		
Loperfido, Angelini, Benedetti, et al., 1998	1827 98	Centers which performed <200 ERCPs per year OR=2.93			X

¹⁶ Case volume was not independently significant in the primary multivariate analysis of total complications conducted by Freeman 1996, probably because of the close relationship with intraoperative technique. In a multivariable model that was based solely on data available prior to the procedure, lower case volume (average <1 case/week per endoscopist vs > 1 case) was independently associated with higher complications (OR 1.43, CI=1.07-1.89).

Table 87. Relationship between Operator Factors and Hemorrhage

Study	N Pts Cx	Case volume	Participation of a trainee	University affiliated center	Center size
Fair Quality					
Freeman, Nelson, Sherman, et al., 1996	2347 48	Endoscopist volume ≤1/week OR=2.17 (1.12-4.17)	X	X	
Fair Minus Quality					
Loperfido, Angelini, Benedetti, et al., 1998	1827 21	Centers which performed <200 ERCPs per year OR=2.98			X

Case volume was not independently significant in the primary multivariate analysis of total complications conducted by Freeman, Nelson, Sherman, et al. (1996), probably because of the close relationship with intraoperative technique. In a multivariable model that was based solely on data available prior to the procedure, lower case volume (average less than 1 case/week per endoscopist vs more than one 1 case) was independently associated with higher complications (OR 1.43, CI=1.07–1.89). This suggests that endoscopist skill in avoiding specific procedural technique is the basis for the association between case volume and complications.

Lower volume of ERCP procedures was associated with hemorrhage in two studies (Freeman, Nelson, Sherman, et al., 1996 and Loperfido, Angelini, Benedetti, et al., 1998) (Table 87). Rabenstein, Schneider, Bulling, et al. (2000) was the only study to find a significant association between lower case volume and pancreatitis (Table 88). The cut off used was fewer than 40 endoscopic sphincterotomies per endoscopist per year. Loperfido, Angelini, Benedetti, et al., (1998) also explored the relationship between case volume and cholangitis or retroperitoneal perforation (Tables 89 and 90) and reported an odds ratio of 4.22 for cholangitis and no association with retroperitoneal perforation.

Conclusion

- Thirteen studies reported on multivariable logistic regression analyses of factors associated with complications of ERCP. The four largest studies each included more than 1,800 patients, and the total number of complications observed in these studies ranged from 98 to 229. Overall, the methodologic quality of the available analyses is limited by overfitting, i.e., testing an excessive number of factors relative to the number of complications observed. Consequently, this literature is exploratory in nature. Reported magnitudes of association are not reliable, significant independent variables may have been overlooked, and some significant associations may be misleading. Moreover, the existing studies do not use common, standardized definitions for the complications and factors of interest. Thus, caution should be used in drawing inferences for clinical practice from these studies.
- Patient, procedure and operator factors were identified that were found to be significantly associated with complications in several of the more robust studies. Younger age (using various cut-offs, but generally 60 years or less) was significantly associated with total complications and with pancreatitis; as was suspected sphincter of Oddi dysfunction. Precut endoscopic sphincterotomy was the procedure-related factor most commonly associated with total complications or pancreatitis; a significant association with difficulty in cannulation was also reported, but less frequently. Multiple pancreatic contrast injections was associated with pancreatitis. For hemorrhage, the clearest association was patient factors related to coagulopathy. Case volume was the only operator-related factor found to be significantly associated with complications. These studies used various cut-offs to define lower volume centers: 1 or fewer procedures per endoscopist per week; fewer than 40 endoscopic sphincterotomies per endoscopist per year; and fewer than 200 procedures per year.

Table 88. Relationship between Operator Factors and Pancreatitis

Study	N Pts Cx	Case volume	Participation of a trainee	University affiliated center	Center size
Fair Quality					
Freeman, DiSario, Nelson, et al., 2001	1963 131	X	X		
Freeman, Nelson, Sherman, et al., 1996	2347 127	X	X	X	
Fair Minus Quality					
Rabenstein, Schneider, Bulling, et al., 2000	438 19	Endoscopist ES case load <40/year OR=3.8 (1.44-10.00)	X		
Loperfido, Angelini, Benedetti, et al., 1998	1827 29	X			X

Table 89. Relationship between Operator Factors and Cholangitis

Study	N Pts Cx	Case volume	Participation of a trainee	University affiliated center	Center size
Fair Minus Quality					
Loperfido, Angelini, Benedetti, et al., 1998	1827 21	Centers which performed <200 ERCPs per year OR=4.22			X

Table 90. Relationship between Operator Factors and Retroperitoneal Perforation

Study	N Pts Cx	Case volume	Participation of a trainee	University affiliated center	Center size
Fair Minus Quality					
Loperfido, Angelini, Benedetti, et al., 1998	1827 12	X			X

Part V, Section 2: Randomized, Controlled Comparison Trials

Introduction

This section summarizes the available randomized, controlled trials that compare technical variations in performing the ERCP procedure and compare associated complication rates. Quality ratings for these studies are available in Table 91. In addition, some of these studies provide comparative information on technical success of the procedure. Based on discussion with this project's Technical Advisory Group, studies evaluating the use of pharmacologic agents or different contrast agents in preventing ERCP-induced pancreatitis were specifically excluded from this systematic review as the volume of this literature could not be incorporated within the scope of this project.

Review of Evidence

Sphincterotome versus Standard Catheter to Achieve Selective Common Bile Duct Cannulation

Two randomized controlled trials (total n=147) compared standard catheterization versus techniques using sphincterotomes to achieve higher success rates in selectively cannulating the common bile duct (Table 92). Cortas, Mehta, Abraham, et al. (1999) randomized 47 patients to standard catheter versus either a standard or wire-guided sphincterotome, and was rated a "Good" quality study. Fifteen attempts were made to cannulate the common bile duct with the randomly assigned catheter, after which patients crossed over. In the initial attempt, the sphincterotome was more successful than the standard catheter in achieving cannulation (97 percent vs. 67 percent, $p=0.009$). After cross overs, the techniques were equivalent (standard catheter 94 percent sphincterotome 97 percent, $p=n.s.$), but successful cannulation was achieved in the sphincterotome group with fewer attempts (12.4 vs. 2.8, $p<0.001$) and in less time (13.5 vs. 3.1 minutes, $p<0.001$). Pancreatitis occurred in 5.6 percent of standard catheter group, and 10.3 percent of the sphincterotome group, but numbers are too small to assess statistical significance.

Schwacha, Allgaier, Deibert, et al. (2000) randomized 100 patients to standard catheter versus sphincterotome and was rated "Fair." If the randomly assigned technique was unsuccessful patients underwent attempts with a tapered cannula, crossing over to the other treatment arm, and then needle knife sphincterotomy. In the initial attempts, the sphincterotome was more successful than the standard catheter (84 percent vs. 62 percent, $p=0.023$). Eventually, cannulation was equally successful in both groups (91 percent for both). Complications were not statistically different between the two groups.

Based on limited evidence, techniques using a sphincterotome appear to have greater success in selective cannulation of the common bile duct than standard catheter, but no definite conclusion can be made regarding the effect of this variation on complications.

Table 91. Quality Assessment

Study Author, Year	Comparable Initial Groups?	Comparable Groups Maintained?	Comparable Performance of Intervention?	Comparable Measurement of Outcomes?	Appropriate Analysis	Summary Evaluation
Randomized Controlled Trials						
Schwacha, Allgaier, Deibert, et al., 2000	RCT (n=100) Good comparability - Randomization not described - Patient characteristics similar	<u>Standard catheter (n=50):</u> 19 crossed over to GS Guidewire <u>Sphincterotome (n=50):</u> 8 crossed over to SC	Adequate for comparison.	Adequate outcome measures used. Outcomes were not assessed blindly.	Method of analysis not clearly stated to be intention to treat Complications reported only in those with primary success	Fair
Cortas, Mehta, Abraham, et al., 1999	RCT (n=47) Good comparability - Randomization method not fully described - Patient characteristics not reported	<u>Standard catheter (n=18)</u> 6 crossed over <u>Sphincterotome (n=29)</u>	Adequate for comparison.	Adequate outcome measures used. Outcomes were not assessed blindly.	Intention to treat analysis was used.	Good
Elta, Barnett, Wille, et al., 1998	RCT (n=170) Good comparability - Randomization by even or odd calendar date - Patient characteristics similar for age, gender, reason for ES	<u>Pure cut (n=86)</u> 8 crossed over to BC <u>Blended current (n=84)</u> No crossover reported	Adequate for comparison.	Adequate outcome measures used. Outcomes reported to be assessed blindly.	Method of analysis not clearly stated to be intention to treat	Fair

Table 91. Quality Assessment (cont'd)

Study Author, Year	Comparable Initial Groups?	Comparable Groups Maintained?	Comparable Performance of Intervention?	Comparable Measurement of Outcomes?	Appropriate Analysis	Summary Evaluation
Randomized Controlled Trials						
Kohler, Maier, Benz et al., 1998	RCT (n=100) Good comparability – Randomization method not fully described – Patient characteristics similar for age, gender, and indication for sphincterotomy	<u>Conventional Current (n=50)</u> No dropouts or exclusion <u>Controlled Current (n=50)</u> No dropouts or exclusion	Adequate for comparison.	Adequate outcome measures used. Outcomes were not assessed blindly.	Method of analysis not clearly stated but equivalent to intent to treat	Good
Siegel, Veerappan, and Tucker, 1994	RCT (n=100) Fair comparability – Randomization method not fully described – Baseline characteristics similar for biliary diagnosis and reason for ES	<u>Monopolar (n=50)</u> 3 crossed over to BP <u>Bipolar (n=50)</u> 5 crossed over to MP	Adequate for comparison	Adequate outcome measures used. Complication outcomes were reportedly assessed blindly.	Method of analysis not clearly reported.	Fair

Table 91. Quality Assessment (cont'd)

Study Author, Year	Comparable Initial Groups?	Comparable Groups Maintained?	Comparable Performance of Intervention?	Comparable Measurement of Outcomes?	Appropriate Analysis	Summary Evaluation
Randomized Controlled Trials						
Kim, Lee, Lee, et al., 1997	RCT (n=45) Fair comparability <ul style="list-style-type: none"> – Randomization technique not specified – Baseline characteristics similar for age, gender, type of Billroth II anastomosis 	No crossovers or exclusions from analysis reported	Adequate for comparison	Adequate outcome measures used. Outcomes were not assessed blindly.	Method of analysis not stated.	Fair
Bergman, Rauws, Fockens, et al., 1997	RCT (n=202) Good comparability <ul style="list-style-type: none"> – blinded computer-generated randomization – patients comparable on all measured characteristics 	16 out of 218 excluded after randomization because of ineligibility	Adequate for comparison	Adequate outcome measures used. Outcomes were not assessed blindly.	All patients retained for analysis	Good

Table 91. Quality Assessment (cont'd)

Study Author, Year	Comparable Initial Groups?	Comparable Groups Maintained?	Comparable Performance of Intervention?	Comparable Measurement of Outcomes?	Appropriate Analysis	Summary Evaluation
Randomized Controlled Trials						
Tarnasky, Palesch, Cunningham et al., 1998	RCT (n=80) Fair comparability – Randomization method not reported – Baseline characteristics were similar except for two areas: biliary cannulation more difficult in No stent group (p=0.03) and longer mean time to repeat pancreatic access in the No stent group (p=0.04)	<u>Stent (n=41)</u> <u>No Stent (n=39)</u> No crossovers or loss to follow-up reported	Adequate for comparison.	Adequate outcome measures used. Outcomes were not assessed blindly.	Analysis not stated to be intention to treat but equivalent because all subjects included in analysis. Analysis did include multivariate adjustment to account for baseline differences.	Good

Table 91. Quality Assessment (cont'd)

Study Author, Year	Comparable Initial Groups?	Comparable Groups Maintained?	Comparable Performance of Intervention?	Comparable Measurement of Outcomes?	Appropriate Analysis	Summary Evaluation
Randomized Controlled Trials						
Smithline, Silverman, Rogers, et al., 1993	RCT (n=98) Fair comparability <ul style="list-style-type: none"> – Randomization method not reported – Patient characteristics similar for age, gender, clinical history of pancreatitis, suspected SOD, abnormal SOM 	<u>Stent (n=48)</u> 5 technical failures excluded 8 who required pre-cut were assigned out of sequence to stent placement <u>No Stent (n=50)</u> No dropouts or exclusions. No crossovers reported.	Adequate for comparison	Adequate outcome measures used. Outcomes were not assessed blindly	Method of analysis not stated.	Fair
Ochi, Mukawa, Kiyosawa, et al., 1999	RCT (n=110) Good comparability <ul style="list-style-type: none"> – randomization not described – patients comparable on all measured characteristics 	All patients retained for analysis	Adequate for comparison	Outcomes were not assessed blindly	All patients retained for short-term outcome analysis 105/110 patients retained for long-term outcome analysis	Good

Table 92. Randomized Clinical Trials Comparing Different ERCP Methods

Article	N	Population and Interventions	Complications/Outcomes																																																									
Schwacha, Allgaier, Deibert, et al., 2000 Research Issue: Techniques to achieve selective CBD cannulation Standard catheter vs. sphincterotome	100	<p>100 consecutive patients randomized to a group undergoing CBD and PD cannulation using and SC with a metallic tip or a GS without guidewire.</p> <p>Exclusion criteria: ERCP within 1 week before randomization Emergency ERCP Previous therapeutic ERCP Previous surgery of the upper GI tract</p> <table><tr><td>Indications*:</td><td>SC</td><td>GS</td></tr><tr><td>Choledocholithiasis</td><td>9</td><td>13</td></tr><tr><td>Pancreato-biliary</td><td></td><td></td></tr><tr><td> Malignancy</td><td>11</td><td>9</td></tr><tr><td> Acute pancreatitis</td><td>6</td><td>4</td></tr><tr><td> Chronic pancreatitis</td><td>5</td><td>3</td></tr><tr><td> Cholestasis of unknown origin</td><td>13</td><td>13</td></tr><tr><td> PSC</td><td>2</td><td>3</td></tr><tr><td> Cholangitis</td><td>0</td><td>2</td></tr><tr><td> Tumor of papilla</td><td>1</td><td>1</td></tr><tr><td> Others</td><td>3</td><td>2</td></tr></table> <p>* No statistical difference between groups</p>	Indications*:	SC	GS	Choledocholithiasis	9	13	Pancreato-biliary			Malignancy	11	9	Acute pancreatitis	6	4	Chronic pancreatitis	5	3	Cholestasis of unknown origin	13	13	PSC	2	3	Cholangitis	0	2	Tumor of papilla	1	1	Others	3	2	<p>Initial Success rates (4 to 5 attempts with assigned technique)</p> <p>Standard catheter (SC) =62% Guidewire sphincterotome (GS)=84% P=0.023</p> <p>Final Success rates (crossovers, needle-knife attempted on failures) Standard catheter (SC)=91% Guidewire sphincterotome (GS)=91%</p> <table><tr><td>Complications (%)**</td><td>SC</td><td>GS</td><td></td></tr><tr><td>None</td><td>65</td><td>69</td><td>n.s.</td></tr><tr><td>Clinical pancreatitis</td><td>10</td><td>5</td><td>n.s.</td></tr><tr><td>Biochemical pancreatitis</td><td>10</td><td>12</td><td>n.s.</td></tr><tr><td>Intramural injection</td><td>3</td><td>5</td><td>n.s.</td></tr><tr><td>Other, not relevant</td><td>12</td><td>9</td><td>n.s.</td></tr></table> <p>** Among patients for whom ERCP was primarily successful (SC n=31; GS n=42)</p>	Complications (%)**	SC	GS		None	65	69	n.s.	Clinical pancreatitis	10	5	n.s.	Biochemical pancreatitis	10	12	n.s.	Intramural injection	3	5	n.s.	Other, not relevant	12	9	n.s.
Indications*:	SC	GS																																																										
Choledocholithiasis	9	13																																																										
Pancreato-biliary																																																												
Malignancy	11	9																																																										
Acute pancreatitis	6	4																																																										
Chronic pancreatitis	5	3																																																										
Cholestasis of unknown origin	13	13																																																										
PSC	2	3																																																										
Cholangitis	0	2																																																										
Tumor of papilla	1	1																																																										
Others	3	2																																																										
Complications (%)**	SC	GS																																																										
None	65	69	n.s.																																																									
Clinical pancreatitis	10	5	n.s.																																																									
Biochemical pancreatitis	10	12	n.s.																																																									
Intramural injection	3	5	n.s.																																																									
Other, not relevant	12	9	n.s.																																																									

Table 92. Randomized Clinical Trials Comparing Different ERCP Methods (cont'd)

Article	N	Population and Interventions	Complications/Outcomes
<p>Cortas, Mehta, Abraham, et al., 1999</p> <p>Research Issue: Techniques to achieve selective CBD cannulation</p> <p>Standard catheter vs. sphincterotome</p>	47	<p>Consecutive patients undergoing ERCP with the intent to selectively cannulate the CBD. Patients randomized to cannulation of the CBD with either a standard catheter (n=18) or a sphincterome (standard or guidewire) (n=29). There were 6 crossovers from SC to SS after initial attempt (15 tries)</p> <p>Exclusion criteria: Patients who had undergone a previous therapeutic ERCP, selective cannulation was not sought as first intention, or a gastroduodenal anatomic anomaly was present.</p> <p>Indication (N): Suspected CBD stones=41 Pancreatico-biliary malignancies=4 Bile leak=2</p>	<p>Initial CBD cannulation success (% , 95% CI): Standard catheter=67% (41-87) Sphincterotome=97% (82-100) p=0.009</p> <p>After crossovers, Final selective CBD cannulation (% , 95% CI): Standard catheter=94% (73-99) Sphincterotome=97% (82-100) P= n.s.</p> <p>Complications:</p> <p>Pancreatitis (% , CI):* SC=5.6 (0.1-27) SS/WS=10.3 (2.2-27.4)</p> <p>*Numbers too small to assess statistical significance</p>

Table 92. Randomized Clinical Trials Comparing Different ERCP Methods (cont'd)

Article	N	Population and Interventions	Complications/Outcomes		
Elta, Barnett, Wille, et al., 1998 Research Issue: Techniques of ES Pure cute vs. blended current	170	170 consecutive patients undergoing biliary endoscopic sphincterotomy between November 1994 and June 1995 were randomized to either blended or pure cut current. Patients undergoing sphincterotomy on even calendar dates received blended current, whereas patients receiving sphincterotomy on odd calendar dates received pure cut*	Complications (N):	Pure	Blended
			Mild pancreatitis*	3	7
			Moderate pancreatitis*	0	2
			Severe pancreatitis*	0	1
			Bleeding	1	1
			Cholangitis	0	1
			Total	4	12
			*Patients with SOD (n=36) actually had a higher rate of pancreatitis (17% vs. 28%), but not significantly different due to low numbers. Difference in the proportion of patients who developed pancreatitis (including SOD patients) was statistically significant (p<0.05). When SOD patients were excluded, the difference in the rate of pancreatitis was still statistically different (p=0.018).		

Table 92. Randomized Clinical Trials Comparing Different ERCP Methods (cont'd)

Article	N	Population and Interventions	Complications/Outcomes																														
Siegel, Veerappan, and Tucker, 1994 Research Issue: Techniques of ES Monopolar vs. Bipolar device using blended current for both	100	Consecutive patients requiring ERCP and sphincterotomy at one institution were randomly assigned to either standard monopolar electrocautery current (n=50) or the bipolar system (n=50).* <table><tr><td><u>Indication:</u></td><td><u>Monopolar</u></td><td><u>Bipolar</u></td></tr><tr><td>CBD stones</td><td>21</td><td>23</td></tr><tr><td>Pancreatitis</td><td>7</td><td>6</td></tr><tr><td>Pancreatic CA</td><td>7</td><td>6</td></tr><tr><td>SOD</td><td>11</td><td>6</td></tr><tr><td>CBD stricture</td><td>3</td><td>7</td></tr><tr><td>Ampullary CA</td><td>1</td><td>0</td></tr><tr><td>Biliary fistula</td><td>0</td><td>2</td></tr><tr><td>Total</td><td>50</td><td>50</td></tr></table> <p>*5 patients assigned to the bipolar group were switched to monopolar group due to difficulties in the insertion of the sphincterome. 3 patients assigned to the monopolar group were crossed over to the bipolar group. The first 50 patients in each group in whom sphincterotomy was performed were included in the study.</p>	<u>Indication:</u>	<u>Monopolar</u>	<u>Bipolar</u>	CBD stones	21	23	Pancreatitis	7	6	Pancreatic CA	7	6	SOD	11	6	CBD stricture	3	7	Ampullary CA	1	0	Biliary fistula	0	2	Total	50	50	Complications (N):	MP	BP	
<u>Indication:</u>	<u>Monopolar</u>	<u>Bipolar</u>																															
CBD stones	21	23																															
Pancreatitis	7	6																															
Pancreatic CA	7	6																															
SOD	11	6																															
CBD stricture	3	7																															
Ampullary CA	1	0																															
Biliary fistula	0	2																															
Total	50	50																															
			Pancreatitis	6	0	p<0.047																											
			Bleeding	1	0	n.s.																											
			Cholangitis	4	3	n.s.																											
			Perforation	0	0	n.s.																											
			Death	1	0	n.s.																											

Table 92. Randomized Clinical Trials Comparing Different ERCP Methods (cont'd)

Article	N	Population and Interventions	Complications/Outcomes												
Kim, Lee, Lee, et al., 1997 Research Issue: Techniques to achieve ERCP and ES in Billroth II patients Forward vs. Side viewing scope	45	<p>Patients s/p Billroth II gastrectomy who required ERCP with sphincterotomy.</p> <p>Patients were randomized to either a forward-viewing (FV) endoscope (n=23) or a side-viewing (SV) endoscope (n=22).</p> <p>Exclusion criteria: Cases of Roux-en Y surgery</p>	<p>Successful cannulation of the papulla*(%): FV= 20 of 23 (87%) SV= 15 of 22 (68%) p= n.s.</p> <p>Successful endoscopic sphincterotomy (%): FV= 10 of 12 (83%) SV= 8 of 10 (80%) p= n.s.</p> <p>Complications advancing endoscope (%): FV=0 of 23 (0%) SV= 4 of 22 (18%) p<0.05</p> <p>* Among the causes of failure to cannulate the papulla, jejunal perforation occurred in 0 patients in the FV group and 4 patients in the SV group.</p> <p>Complications of endoscopic needle-knife sphincterotomy</p> <table> <tr> <th></th><th>FV n=12</th><th>SV n=10</th><th></th></tr> <tr> <td>Pancreatitis</td><td>1</td><td>2</td><td>n.s.</td></tr> <tr> <td>Retroperitoneal perforation</td><td>0</td><td>1</td><td>n.s.</td></tr> </table>		FV n=12	SV n=10		Pancreatitis	1	2	n.s.	Retroperitoneal perforation	0	1	n.s.
	FV n=12	SV n=10													
Pancreatitis	1	2	n.s.												
Retroperitoneal perforation	0	1	n.s.												

Table 92. Randomized Clinical Trials Comparing Different ERCP Methods (cont'd)

Article	N	Population and Interventions	Complications/Outcomes
Bergman, Rauws, Fockens, et al., 1997 Research Issue: Techniques to remove CBD stone Balloon dilation vs. ES	202	Consecutive patients referred for ERCP because of symptoms of CBD stones. Patients meeting inclusion and exclusion criteria were randomized to either endoscopic sphincterotomy (n=101) or endoscopic balloon dilation (n=101). Eligibility criteria: Over age 18 years BDS visualized at ERCP Deep cannulation of the BD achieved without sphincterotomy Exclusion criteria: Signs of acute cholangitis Acute pancreatitis Acute cholecystitis History of previous sphincterotomy Choledochoduodenal fistula Hemostatic disorders Intrahepatic stone disease Hemolytic anemia Concomitant pancreatic or biliary malignant disorders Coexisting bile leakage or choledochoduodenal fistula Previous participation in this study Life expectancy of less than 1 month	Complete stone removal in one endoscopic session (%): EBD=89 EST=91 n.s. <u>Early Complications (N):</u> <u>EBD</u> <u>EST</u> Pancreatitis 7 7 Fever 4 5 Bleeding 0 4 Perforation 2 1 Pain in right upper abdomen 0 4 Slow resolution of jaundice 2 1 Bile leakage 1 1 Cardiopulmonary 1 1 Total 17 24 n.s. (continued next page)

Table 92. Randomized Clinical Trials Comparing Different ERCP Methods (cont'd)

Article	N	Population and Interventions	Complications/Outcomes			
Bergman, Rauws, Fockens, et al., 1997 (cont'd) Research Issue: Techniques to remove CBD stone Balloon dilation vs. ES	202	(see previous page)	<u>Complications during follow-up (N):</u>			
			Recurrence of symptoms	14		
			Stones on repeat ERCP	8	7	
			No stones on repeat ERCP			
			ERC	6	5	
			No repeat ERCP done	0	2	
			Acute cholecystitis*	1	7	
			Symptomatic cholecystolithiasis	2	1	
			Liver abscess	0	1	
			Abnormal liver function at follow-up	1	0	
			Total	18	23	n.s.
			* Statistically significantly lower in the EBD group			
			Logistic regression analysis of treatment allocation, stone size, stone number, gender, perampullary diverticulum, and Billroth II gastrectomy on successful stone removal identified stone size (p=0.0008), and stone number (p=0.0216) as the only significant predictors of this outcome. Further subgroup analyses were undertaken (not reported in this table).			

Table 92. Randomized Clinical Trials Comparing Different ERCP Methods (cont'd)

Article	N	Population and Interventions	Complications/Outcomes
Ochi, Mukawa, Kiyosawa, et al., 1999	110	Patients with bile duct stones up to 15 mm in diameter and less than 10 in number as indicated by ERCP were randomly treated with either endoscopic papillary dilation (n=55) or endoscopic sphincterotomy (n=55).	Successful bile duct clearance (%): EPD=92.7 EST=98.1 n.s. Successful bile duct clearance achieved in the initial procedure (%): EPD=78.4 EST=94.4 p=0.02
Research Issue: Techniques to remove CBD stone		Exclusion criteria: Recurrent stones following previous procedures Intrahepatic stone disease Acute cholangitis Cholecystitis Pancreatitis Pancreatic or biliary malignant disorders	Early complications (total)(%) (EPD n=51, EST n=54): EPD=2.0 EST=5.6 n.s.
Balloon dilation vs. ES			Specific complications (N) EPD EST Progression of jaundice 1 0 Perforation 0 2
			Late complications (total/eligible for follow-up)(N): EPD=2/51 EST=8/54 n.s.
			Specific complications (N) EPD EST Recurrence of BDS 2 3 n.s. Acute cholangitis 2 2 n.s. Acute cholecystitis 1/30 5/27 n.s. Acute cholecystitis in patients with gallbladder stones in situ 1/22 5/17 p<0.03

Table 92. Randomized Clinical Trials Comparing Different ERCP Methods (cont'd)

Article	N	Population and Interventions	Complications/Outcomes												
Tarnasky, Palesch, Cunningham et al., 1998 Research Issue: Pancreatic stenting to reduce pancreatitis after ES	80	<p>Consecutive adult patients scheduled for ERCP with SOD manometry, for evaluation of unexplained pancreatobiliary pain or pancreatitis, were randomized to either pancreatic duct stents (n=41) or no stents (n=39).</p> <p>Exclusions: Pancreatic SOM results normal SOM failure or not attempted Severe chronic pancreatitis Pancreas divisum Prior gastric surgery PSH No sphincterotomy Both biliary and pancreatic sphincterotomy Precut sphincterotomy required to achieve biliary access Preference of physician or patient not to participate Failure to gain repeat pancreatic access after biliary sphincterotomy</p> <table><tr><td><u>Indications (%)</u>:</td><td><u>Stent</u></td><td><u>No Stent</u></td></tr><tr><td>Pancreatobiliary pain (gallbladder out)</td><td>51</td><td>72</td></tr><tr><td>Pancreatobiliary pain (gallbladder in)</td><td>20</td><td>5</td></tr><tr><td>Prior acute pancreatitis</td><td>29</td><td>23</td></tr></table>	<u>Indications (%)</u> :	<u>Stent</u>	<u>No Stent</u>	Pancreatobiliary pain (gallbladder out)	51	72	Pancreatobiliary pain (gallbladder in)	20	5	Prior acute pancreatitis	29	23	<p><u>Complications:</u></p> <p>Incidence of post-ERCP pancreatitis (%): Stent=2 No Stent=26 p=0.003</p> <p>RR of post-ERCP pancreatitis after biliary sphincterotomy in the no stent group=10.5, 95% CI=1.4-78.3</p> <p>Logistic regression analysis controlling for differences in baseline data (difficulty of biliary cannulation and time to repeat pancreatic access) resulted in an AOR=14.4, 95% CI=1.7-125.0 for the risk of post-ERCP pancreatitis among patients in the no stent group.</p>
<u>Indications (%)</u> :	<u>Stent</u>	<u>No Stent</u>													
Pancreatobiliary pain (gallbladder out)	51	72													
Pancreatobiliary pain (gallbladder in)	20	5													
Prior acute pancreatitis	29	23													

Table 92. Randomized Clinical Trials Comparing Different ERCP Methods (cont'd)

Article	N	Population and Interventions	Complications/Outcomes
Smithline, Silverman, Rogers, et al., 1993 Research Issue: Pancreatic stenting to reduce pancreatitis after ES	98	High risk patients (those with SOD or CBD <10 mm and patients requiring pre-cut biliary ES) were randomized to receive a main pancreatic duct stent or no stent following biliary sphincterotomy. Exclusions: Patients with pancreatic divisum, pancreatobiliary tumors, or those undergoing pancreatic septotomy	<u>Complications:</u> Incidence of pancreatitis (%): MPD Stent=14 No Stent=18 n.s. * Severity of pancreatitis (%): Mild MPD Stent=13 No Stent=12 n.s. Moderate MPD Stent=0 No Stent=6 n.s. Severe MPD Stent=0 No Stent=6 n.s. Other suspected risk factors for pancreatitis were examined including acinarization, precut ES, and history of pancreatitis. None of these risk factors were found to be independent risk factors of pancreatitis in high-risk patients. * Pancreatitis developed in 2 of 5 patients in whom stent placement failed

Variations in Electric Current Used in Sphincterotomy to Reduce Post-ERCP Complications

Three randomized clinical trials (all rated “Fair” quality) compared variations of the electric current used in performing sphincterotomy as methods to reduce post-procedure complications such as hemorrhage or pancreatitis.

Elta, Barnett, Wille, et al. (1998) randomized 170 patients to either blended or pure cut current when undergoing sphincterotomy. Blended current combines intermittent high voltage pulses with continuous low voltage current, whereas pure cut current is simply continuous low voltage current. Total complications were significantly lower in the pure cut group (5 percent vs. 14 percent, $p<0.05$).

Kohler, Maier, Benz et al. (1998) randomized 100 patients to either conventional high-frequency blended current or a newly developed high-frequency system with automatically controlled cutting mode (Endocut). Mild bleeding during sphincterotomy was significantly reduced (4 percent compared to 26 percent, $p=0.002$), but no significant difference was observed in moderate/severe bleeding or mild pancreatitis, which both occurred very infrequently.

Siegel Veerappan, and Tucker (1994) randomized 100 patients to receive either a bipolar or monopolar electric current device when undergoing sphincterotomy. Pancreatitis occurred in 6 patients receiving monopolar electrocautery and 1 patients receiving bipolar electrocautery ($p<0.05$). Other complications were very uncommon and numbers were too small to make conclusions about statistical significance.

Forward-Viewing Endoscope versus Side-Viewing Endoscope to Achieve Successful Cannulation and Sphincterotomy in Patients with Billroth II Gastrectomy

Kim, Lee, Lee, et al. (1997) randomized 45 patients with Billroth II gastrectomy who required ERCP and sphincterotomy to have the procedure done with either a forward-viewing (FV) endoscope or side-viewing (SV) duodenoscope. Successful cannulation occurred in 87 percent of FV group and 68 percent of SV group ($p=n.s.$) Successful sphincterotomy was not statistically different (FV 83 percent, SV 80 percent). Jejunal perforation occurred in 4 patients using the SV duodenoscope and 0 patients using the FV endoscope ($p<0.05$). Use of the FV endoscope may cause fewer perforations than the SV duodenoscope.

Pancreatic Stenting to Reducing Pancreatitis after Sphincterotomy

Two small randomized controlled trials examined whether placing pancreatic stents after sphincterotomy reduces the incidence of post-ERCP pancreatitis among certain patients considered to be at high risk for such a complication.

Smithline, Silverman, Rogers, et al. (1993) randomized 98 patients using an alternate assignment scheme and was rated Fair quality. The patients included those with abnormal SOD manometry, clinical suspicion of SOD, a common bile duct ≤ 10 mm or patients requiring a pre-cut sphincterotomy. Some patients requiring a pre-cut sphincterotomy were assigned a stent out of

the randomization scheme. The results are analyzed only among those who received intended treatment, as patients with failed stent placement (5 patients) are analyzed separately. The no-stent group had an 18 percent rate of pancreatitis, the stent group had a 14 percent rate of pancreatitis ($p=n.s.$) If appropriately analyzed by intent-to-treat, the pancreatitis rates would be even more similar.

Tarnasky, Palesch, Cunningham et al. (1998) randomized 80 patients to receive stents or no stent and was rated “Good” quality. The selection criteria appear to be more selective than the study by Smithline, Silverman, Rogers, et al. (1993), as only patients with confirmed abnormal sphincter of Oddi manometry and pancreatic sphincter hypertension were included. The incidence of post-ERCP pancreatitis in the stent group was 2 percent, and in the no stent group was 26 percent ($p=0.003$). After correction for some baseline differences between study groups, the risk of post-ERCP pancreatitis was still highly associated with lack of stent placement (odds ratio 14.4, $p=0.002$).

An important distinction between the two studies is the selection criteria. Smithline, Silverman, Rogers, et al. (1993) included several types of patients that are thought to be at risk of post-ERCP pancreatitis, Tarnasky, Palesch, Cunningham et al. (1998) included only patients with both confirmed abnormal sphincter of Oddi manometry and pancreatic sphincter hypertension. About three-fourths of the patients in the Smithline, Silverman, Rogers, et al. (1993) study had abnormal sphincter of Oddi manometry, and among those, pancreatic sphincter pressure was not assessed. Thus the results may not be inconsistent, even though the same intervention is assessed using identical outcome measures.

In conclusion, evidence limited to only one trial shows some evidence of efficacy of pancreatic stent placement in preventing post-ERCP pancreatitis, but only among patients with confirmed sphincter of Oddi manometry and concurrent pancreatic sphincter hypertension.