American Society for Gastrointestinal Endoscopy guideline on role of endoscopy in the diagnosis of malignancy in biliary strictures of undetermined etiology: methodology and review of evidence

Prepared by: ASGE STANDARDS OF PRACTICE COMMITTEE

Larissa L. Fujii-Lau, MD,1 Nirav C. Thosani, MD,2 Mohammad Al-Haddad, MD, MSc, FASGE,3 Jared Acoba, MD,4 Curtis J. Wray, MD, MS, FACS,5 Rodrick Zvavanjanja, MD, Msc, FRCR(UK), FSIR, DABR(VIR/DR),6 Stuart K. Amateau, MD, PhD, FASGE,7 James L. Buxbaum, MD, MS, FASGE,8 Sachin Wani, MD, FASGE,9 Audrey H. Calderwood, MD, MS, FASGE,10 Jean M. Chalhoub, MD,11 Nayantara Coelho-Prabhu, MD, FASGE,12 Madhav Desai, MD, MPH,13 Sherif E. Elhanafi, MD,14 Douglas S. Fishman, MD, FAAP, FASGE,15 Nauzer Forbes, MD, MSc, FASGE,16 Laith H. Jamil, MD, FASGE,17 Terry L. Jue, MD, FASGE,18 Divyanshoo R. Kohli, MD,19 Richard S. Kwon, MD, FASGE,20 Joanna K. Law, MD, FRCP, FASGE,21 Jeffrey K. Lee, MD, MPH,22 Jorge D. Machicado, MD, MPH,26 Neil B. Marya, MD,23 Swati Pawa, MD, FASGE,24 Wenly Ruan, MD,15 Mandeep S. Sawhney, MD, MS, FASGE,25 Sunil G. Sheth, MD, FASGE,25 Andrew Storm, MD,12 Nikhil R. Thiruvengadam, MD,26 Bashar J. Qumseya, MD, MPH, FASGE27

(ASGE Standards of Practice Committee Chair [2020-2023])

Biliary strictures of undetermined etiology pose a diagnostic challenge for endoscopists. Despite advances in technology, diagnosing malignancy in biliary strictures often requires multiple procedures. The Grading of Recommendations Assessment, Development and Evaluation (GRADE) framework was used to rigorously review and synthesize the available literature on strategies used to diagnose undetermined biliary strictures. Using a systematic review and meta-analysis of each diagnostic modality, including fluoroscopic-guided biopsy sampling, brush cytology, cholangioscopy, and EUS-guided FNA or fine-needle biopsy sampling, the American Society for Gastrointestinal Endoscopy Standards of Practice Committee provides this guideline on modalities used to diagnose biliary strictures of undetermined etiology. This document summarizes the methods used in the GRADE analysis to make recommendations, whereas the accompanying article subtitled “Summary and Recommendations” contains a concise summary of our findings and final recommendations. (Gastroint Endosc 2023;:1-19.)

This guideline document was prepared by the Standards of Practice Committee of the American Society for Gastrointestinal Endoscopy using the best available scientific evidence and considering a multitude of variables including but not limited to adverse events, patient values, and cost implications. The purpose of these guidelines is to provide the best practice recommendations that may help standardize patient care, improve patient outcomes, and reduce variability in practice. We recognize that clinical decision-making is complex. Guidelines, therefore, are not a substitute for a clinician’s judgment. Such judgements may at times seem contradictory to our guidance because of many factors that are impossible to fully consider by guideline developers. Any clinical decisions should be based on the clinician’s experience, local expertise, resource availability, and patient values and preferences. This document is not a rule and should not be construed as establishing a legal standard of care or as encouraging, advocating for, mandating, or discouraging any particular treatment. Our guidelines should not be used in support of medical complaints, legal proceedings, and/or litigation, as they were not designed for this purpose.

Biliary strictures remain a challenge for gastroenterologists and hepatologists. Such strictures pose a diagnostic dilemma because cross-sectional imaging is often nonspecific and noninvasive options are limited for the diagnosis

www.giejournal.org Volume ■, No. ■ : 2023 GASTROINTESTINAL ENDOSCOPY 1
of malignancy in these biliary strictures. Therefore, the diagnostic workup predominately lies in the hands of the advanced endoscopist and often requires multiple procedures to determine whether the stricture is benign or malignant. This may delay the diagnosis and treatment of these strictures, which may in turn worsen the overall prognosis in cases of malignancy. Furthermore, this delay in diagnosis is associated with increased patient cost, time, potential adverse events, and anxiety. Despite advances in endoscopic techniques for tissue acquisition, up to 20% of patients with suspected cholangiocarcinoma have benign disease at surgical resection. Therefore, the American Society for Gastrointestinal Endoscopy (ASGE) Standards of Practice Committee has developed guidelines for the role of endoscopy in biliary strictures of undetermined etiology. We focused on 3 important modalities, ERCP with fluoroscopic-guided biopsy sampling, cholangioscopy-guided biopsy sampling, and EUS with FNA or fine-needle biopsy sampling (FNB), to develop recommendations on the diagnostic approach to biliary strictures of undetermined etiology.

These guidelines follow the Grading of Recommendations Assessment, Development and Evaluation (GRADE) methodology. This article details guideline methodology including formulation of clinical questions, literature searches, data analyses, panel composition, evidence profiles, and other considerations like cost-effectiveness, patient preferences, and health equity. For each clinical question, this article includes outcomes of interest, pooled-effects estimates, and evidence that was considered by the panel in making final recommendations. The accompanying article subtitled “Summary and Recommendations” is published separately and provides a summary of our findings and final recommendations.

Our pediatric gastroenterologist (D.S.F.) highlighted that strictures secondary to malignancy are rare in pediatric patients. Cholangiocarcinoma in pediatric patients occurs at a rate of only .0036 per 100,000, and thus specific endoscopic sampling recommendations may not be applicable in patients under age 21 years.

### METHODS

#### Formulation of clinical questions

Our guideline addressed 3 questions using GRADE methodology (Table 1). For these questions we followed the PICO format: P, population in question; I, intervention; C, comparator; and O, outcomes of interest. For all clinical questions, potentially relevant patient-important outcomes were identified a priori and rated from “critical” to “important” through a consensus process.

For each clinical question, we included studies with any location of the biliary stricture ( hilar, extrahepatic, intrahepatic, proximal, distal). The term indeterminate biliary stricture was not used because it historically refers to patients who had negative tissue diagnosis from prior ERCP.

### Table 1. Population, intervention, comparator, outcomes questions

<table>
<thead>
<tr>
<th>Population</th>
<th>Intervention</th>
<th>Comparator</th>
<th>Outcome</th>
<th>Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biliary stricture of undetermined etiology</td>
<td>a. ERCP with fluoroscopic-guided biopsy + brush cytology</td>
<td>ERCP with brush cytology</td>
<td>1. Incremental yield</td>
<td>Critical</td>
</tr>
<tr>
<td></td>
<td>b. ERCP with fluoroscopic-guided biopsy alone</td>
<td></td>
<td>2. Sensitivity, specificity, and positive and negative predictive values</td>
<td>Critical</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3. Technical success</td>
<td>Important</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4. Specimen adequacy</td>
<td>Important</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>5. Adverse events</td>
<td>Important</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>6. Mortality</td>
<td>Important</td>
</tr>
<tr>
<td>Biliary stricture of undetermined etiology</td>
<td>ERCP without cholangioscopy guided biopsy sampling</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1. Incremental yield</td>
<td>Critical</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2. Sensitivity, specificity, and positive and negative predictive values</td>
<td>Critical</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3. Technical success</td>
<td>Important</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4. Specimen adequacy</td>
<td>Important</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>5. Adverse events</td>
<td>Important</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>6. Mortality</td>
<td>Important</td>
</tr>
<tr>
<td>Biliary stricture of undetermined etiology</td>
<td>ERCP alone with any form of tissue acquisition</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1. Incremental yield</td>
<td>Critical</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2. Sensitivity, specificity, and positive and negative predictive values</td>
<td>Critical</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3. Technical success</td>
<td>Important</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4. Specimen adequacy</td>
<td>Important</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>5. Adverse events</td>
<td>Important</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>6. Mortality</td>
<td>Important</td>
</tr>
</tbody>
</table>

*Malignant diagnosis is based on surgical or autopsy pathology, nonequivocal cytologic diagnosis, positive histology, and follow-up clinical course of at least 6 months consistent with malignant disease, whereas a benign diagnosis is based on surgical or autopsy pathology or a follow-up clinical course of at least 12 months consistent with a benign disease.*
Instead, the terms *biliary stricture of undetermined etiology* or *undetermined biliary stricture* were used so that studies including patients undergoing their initial endoscopic evaluation were incorporated in the meta-analysis. It is important to make this distinction to emphasize the importance of the potential use of multiple forms of tissue acquisition during the initial endoscopic evaluation to enhance the diagnostic approach to these strictures.

**Literature search and study selection criteria**

To inform the guideline panel, a comprehensive literature search was performed with the help of a medical librarian using Ovid MEDLINE, Embase, and Wiley Cochrane. Inclusion criteria were articles published in the English language, randomized controlled and observational studies from inception through May 28, 2021, and abstracts presented at major gastroenterology or hepatology conferences within the last 5 years. Case reports, case series with fewer than 10 patients, reviews, editorials, and animal studies were excluded. If not enough data were available to calculate our own statistical analysis for the diagnostic test characteristic (particularly sensitivity and specificity), the study was also excluded.

For each PICO question, the systematic literature search was used to identify existing systematic reviews and meta-analyses. If none were found, a full systematic review and meta-analysis was conducted using the recommendations of the Preferred Reporting Items for Systematic reviews and Meta-Analyses criteria. Citations were imported into EndNote (Thompson Reuters, Philadelphia, Penn, USA), and duplicates were removed. The EndNote library was then uploaded into Covidence (www.covidence.org) for review by 2 independent reviewers (L.L.F.-L. and M.A.). Studies were first screened by title and abstract and then by full text by 2 independent reviewers (L.L.F.-L. and M.A.), and all conflicts were resolved by consensus. When applicable, available systematic reviews and meta-analyses were updated based on literature review as described above.

**Data extraction and statistical analysis**

Data were extracted by 2 independent reviewers (L.L.F.-L. and M.A.). The primary estimate of effect was based on a priori identified outcomes of interest. After calculating the true positives, false positives, true negatives, and false negatives of each included study, pooled sensitivity, specificity, positive likelihood ratio, and negative likelihood ratio were calculated using MetaDisc V1.4 (Madrid, Spain). The summary statistic included odds ratios (ORs) for each of the other outcomes (incremental yield, technical success, specimen adequacy, and adverse events). For each PICO question, subgroup analyses were performed for the location of the biliary stricture (distal vs proximal bile duct) and whether the primary mass was in the bile duct or pancreas.

Statistical analyses were performed using RevMan V5.3 (Cochrane, London, UK) and Comprehensive Meta Analysis V3 (Biostat Inc, Englewood, NJ, USA). Pooled effects were calculated using a DerSimonian and Laird random-effects model, and studies were weighted based on size. Heterogeneity was assessed using the $I^2$ statistic, and publication bias was analyzed using funnel plots. Quality was assessed using the Cochrane risk of bias tool for randomized controlled trials and the modified Newcastle-Ottawa Scale for observational studies (Supplementary Table 1, available online at www.giejournal.org).

**Panel composition and conflict of interest management**

We assembled a virtual panel of stakeholders to review evidence and make recommendations on January 17, 2022. The panel consisted of lead authors (L.F.L., N.C.T., and M.A.), a committee member with expertise in GRADE methodology (N.F.), and content experts (J.A., oncology; C.J.W., surgical oncology; and R.Z., interventional radiology) and was chaired by the Standards of Practice Committee chair (B.J.Q.). A patient representative from the Cholangiocarcinoma Foundation was also included. All panel members were required to disclose potential financial and intellectual conflicts of interest, which were addressed according to ASGE policies set forth in the ASGE & Journal Policy for Managing Declared Conflicts of Interest found at https://www.asge.org/docs/default-source/default-document-library/coi-full-policy-for-asge-and-publications_edd_2-10-20.pdf. The primary methodologist (L.L.F.-L.) was excluded from all votes.

**Certainty in evidence, outcomes, and definitions**

The certainty in the body of evidence (also known as quality of the evidence or confidence in the estimated effects) was assessed using the GRADE framework (Table 2). Primary outcomes were the incremental yield and diagnostic accuracy (sensitivity, specificity, positive likelihood ratio, negative likelihood ratio) statistics. Other clinical outcomes were technical success, specimen adequacy, and adverse events. Although mortality was in the initial list of outcomes, no included study in any of the PICO questions had any mortality directly related to the endoscopic procedures.

A diagnosis of malignancy was based on surgical or autopsy pathology, unequivocal cytologic diagnosis of malignancy, positive histology, or follow-up course of at least 6 months consistent with malignant disease. A diagnosis of benign pathology was based on surgical or autopsy pathology or follow-up of at least 12 months consistent with benign disease.
**TABLE 2. GRADE categories of quality of evidence and corresponding meaning and interpretation and implications of the strength of GRADE recommendations on various stakeholders**

<table>
<thead>
<tr>
<th>Quality of evidence</th>
<th>Meaning</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>We are confident that the true effect lies close to that of the estimate of the effect.</td>
<td>Further research is very unlikely to change our confidence in the estimate of the effect.</td>
</tr>
<tr>
<td>Moderate</td>
<td>We are moderately confident in the estimate of the effect; the true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different.</td>
<td>Further research is likely to have an impact on our confidence in the estimate of the effect and may change the estimate.</td>
</tr>
<tr>
<td>Low</td>
<td>Our confidence in the estimate of the effect is limited; the true effect may be substantially different from the estimate of the effect.</td>
<td>Further research is very likely to have an impact on our confidence in the estimate of the effect and is likely to change the estimate.</td>
</tr>
<tr>
<td>Very low</td>
<td>We have very little confidence in the estimate of the effect; the true effect is likely to be substantially different from the estimate of the effect.</td>
<td>Any estimate of the effect is very uncertain.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Implications for</th>
<th>Strong recommendation</th>
<th>Conditional recommendation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patients</td>
<td>Most individuals in this situation would want the recommended course of action, and only a small proportion would not.</td>
<td>Most individuals in this situation would want the suggested course of action, but many would not.</td>
</tr>
<tr>
<td>Clinicians</td>
<td>Most individuals should receive the test. Formal decision aids are not likely to be needed to help individual patients make decisions consistent with their values and preferences.</td>
<td>Recognize that different choices will be appropriate for individual patients and that you must help each patient arrive at a management decision consistent with his or her values and preferences. Decision aids may be useful in helping individuals to make decisions consistent with their values and preferences.</td>
</tr>
<tr>
<td>Policymakers</td>
<td>The recommendation can be adopted as policy in most situations. Compliance with this recommendation according to the guideline could be used as a quality criterion or performance indicator.</td>
<td>Policymaking will require substantial debate and involvement of various stakeholders.</td>
</tr>
</tbody>
</table>


**RESULTS**

**Question 1:** In patients with biliary strictures of undetermined etiology, should ERCP with fluoroscopic-guided biopsy sampling be performed in addition to brush cytology versus ERCP with brush cytology alone to diagnose malignancy?

**Recommendation 1.** In patients with a biliary stricture of undetermined etiology undergoing ERCP, the ASGE suggests the addition of fluoroscopic-guided biopsies with brush cytology to brush cytology alone to diagnose malignancy.

*(Conditional recommendation/very low quality of evidence)*

We performed a systematic review and meta-analysis on patients with biliary strictures of undetermined etiology who underwent fluoroscopic-guided biopsy sampling, brush cytology, or both. An initial search yielded 2695 total studies, and an updated search yielded an additional 305 studies (Appendix 1, available online at www.giejournal.org). Fifty-two studies underwent full text review, and 21 studies (2726 patients) were included. 10-30 All 21 studies were observational studies; 20 of these were full-text publications and 1 was a meeting abstract.

**Incremental yield**

Incremental yield calculations were performed on 7 studies14,16,17,21,23,24,29 that directly compared fluoroscopic-guided biopsy sampling with brush cytology alone on the same patients and had sufficient information on how many patients had a positive biopsy sampling or brush cytology result. The standard of care was considered ERCP with brush cytology alone, so the incremental yield was expressed as the addition of fluoroscopic-guided biopsy sampling to brush cytology. To calculate the incremental yield, the total number of patients who underwent both biopsy sampling and brushings in which only the biopsy sample was positive was divided by the total number of patients who were diagnosed with malignancy. Based on the random-effects model, the addition of fluoroscopic-guided biopsy sampling to brushing resulted in a 20% (95% confidence interval [CI], 9-51; $I^2 = 54.5\%$) increase in the diagnostic yield compared with brushing alone (Fig. 1).

The miss rate of either fluoroscopic-guided biopsy sampling or brush cytology was also calculated to determine how many malignant diagnoses each modality would miss. We found that brush cytology alone missed 58% (95% CI, 46-71; $I^2 = 79.5\%$) of malignancies (Fig. 2), whereas fluoroscopic-guided biopsy sampling missed 41% (95% CI, 31-52; $I^2 = 80.3\%$) (Fig. 3).
Diagnostic test characteristics were calculated in 20 studies comparing fluoroscopic-guided biopsy sampling with brush cytology and biopsy sampling + brushing with brushing alone. The pooled diagnostic test characteristics for brushing cytology alone were sensitivity of .4 (95% CI, .37-.43), specificity of .98 (95% CI, .97-.99), positive likelihood ratio of 10.57 (95% CI, 5.56-20.12), negative likelihood ratio of .63 (95% CI, .58-.69), diagnostic OR of 18.9 (95% CI, 10.31-34.66), and area under the curve (summary receiver-operating characteristic curve [SROC]) of .615. However, for fluoroscopic-guided biopsy sampling only, the pooled diagnostic characteristics were a sensitivity of .52 (95% CI, .49-.55), specificity of .97 (95% CI, .96-.99), positive likelihood ratio of 10.25 (95% CI, 6.36-16.5), negative likelihood ratio of .51 (95% CI, .43-.59), OR of 20.96 (95% CI, 12.41-35.4), and SROC of .799.

### Figure 1.
Incremental yield of fluoroscopic-guided biopsy sampling over brush cytology. *ID-BX*, Intraductal biopsy; *CI*, confidence interval; *REML*, random effects model.

### Figure 2.
Miss rate of brush cytology in the diagnosis of malignant strictures. *ES*, estimate; *CI*, confidence interval.

### Diagnostic accuracy
Diagnostic test characteristics were calculated in 20 studies comparing fluoroscopic-guided biopsy sampling with brush cytology and biopsy sampling + brushing with brushing alone. The pooled diagnostic test characteristics for brushing cytology alone were sensitivity of .4 (95% CI, .37-.43), specificity of .98 (95% CI, .97-.99), positive likelihood ratio of 10.57 (95% CI, 5.56-20.12), negative likelihood ratio of .63 (95% CI, .58-.69), diagnostic OR of 18.9 (95% CI, 10.31-34.66), and area under the curve (summary receiver-operating characteristic curve [SROC]) of .615. However, for fluoroscopic-guided biopsy sampling only, the pooled diagnostic characteristics were a sensitivity of .52 (95% CI, .49-.55), specificity of .97 (95% CI, .96-.99), positive likelihood ratio of 10.25 (95% CI, 6.36-16.5), negative likelihood ratio of .51 (95% CI, .43-.59), OR of 20.96 (95% CI, 12.41-35.4), and SROC of .799.

The sensitivity of fluoroscopic-guided biopsy sampling (52%) alone was significantly higher than brush cytology.

---

**Table 1.**

<table>
<thead>
<tr>
<th>Study</th>
<th>ID-Bx + Brushing Detected</th>
<th>ID-Bx + Brushing Missed</th>
<th>Brushing Detected</th>
<th>Brushing Missed</th>
<th>Risk diff with 95% CI</th>
<th>Weight (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ren et al. 2018</td>
<td>34</td>
<td>8</td>
<td>16</td>
<td>26</td>
<td>0.43 [0.24, 0.62]</td>
<td>14.83</td>
</tr>
<tr>
<td>Pugliese et al. 1997</td>
<td>22</td>
<td>14</td>
<td>19</td>
<td>17</td>
<td>0.08 [-0.14, 0.31]</td>
<td>12.30</td>
</tr>
<tr>
<td>Weber et al. 2008</td>
<td>35</td>
<td>23</td>
<td>24</td>
<td>34</td>
<td>0.19 [0.01, 0.37]</td>
<td>15.57</td>
</tr>
<tr>
<td>Jailwala et al. 1999</td>
<td>57</td>
<td>47</td>
<td>31</td>
<td>73</td>
<td>0.25 [0.12, 0.38]</td>
<td>19.44</td>
</tr>
<tr>
<td>Kitajima et al. 2007</td>
<td>28</td>
<td>10</td>
<td>27</td>
<td>11</td>
<td>0.03 [-0.17, 0.23]</td>
<td>13.99</td>
</tr>
<tr>
<td>Rosch et al. 2004</td>
<td>13</td>
<td>15</td>
<td>12</td>
<td>16</td>
<td>0.04 [-0.22, 0.30]</td>
<td>10.56</td>
</tr>
<tr>
<td>Kulaksiz et al. 2011</td>
<td>28</td>
<td>6</td>
<td>17</td>
<td>17</td>
<td>0.32 [0.11, 0.53]</td>
<td>13.32</td>
</tr>
</tbody>
</table>

**Overall**

Heterogeneity: $I^2 = 0.01$, $H^2 = 54.45%$, $H^2 = 2.20$

Test of $H = 0$: Q(6) = 12.77, p = 0.05

Test of $H = 0$: Z = 3.66, p = 0.00

**Random-effects REML model**

---

**Table 2.**

<table>
<thead>
<tr>
<th>Study</th>
<th>ES (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Draganov et al. 2012</td>
<td>0.71 (0.47, 0.87)</td>
</tr>
<tr>
<td>Gerges et al. 2020</td>
<td>0.79 (0.52, 0.92)</td>
</tr>
<tr>
<td>Pugliese et al. 1997</td>
<td>0.47 (0.32, 0.63)</td>
</tr>
<tr>
<td>Weber et al. 2008</td>
<td>0.59 (0.46, 0.70)</td>
</tr>
<tr>
<td>Jailwala et al. 1999</td>
<td>0.70 (0.61, 0.78)</td>
</tr>
<tr>
<td>Kitajima et al. 2007</td>
<td>0.29 (0.17, 0.45)</td>
</tr>
<tr>
<td>Rosch et al. 2004</td>
<td>0.57 (0.38, 0.73)</td>
</tr>
<tr>
<td>Overall (I^2 = 79.5%, p = 0.000)</td>
<td>0.58 (0.46, 0.71)</td>
</tr>
</tbody>
</table>
alone (.52 vs .4, respectively; \( P < .006 \)). These results can be found in Table 3.

The pooled diagnostic test characteristics of combined fluoroscopic-guided biopsy sampling with brush cytology were a sensitivity of .66 (95% CI, .63-.69), specificity of .97 (95% CI, .95-.98), positive likelihood ratio of 11.91 (95% CI, 7.37-19.23), negative likelihood ratio of .38 (95% CI, 0.33-.43), diagnostic OR of 31.78 (95% CI, 18.59-54.35), and SROC of .7668. The sensitivity of the fluoroscopic biopsy sampling and brushing was significantly higher than brushing alone (\( P < .001 \)).10-14,16-21,23,25-29

### Technical success and specimen adequacy

Eight studies reported on technical success of brush cytology and fluoroscopic-guided biopsy sampling.10,16,17,19,21,22,25,28 Four of these studies reported 100% technical success on both forms of tissue acquisition.10,17,19,22 The remaining 4 studies showed no difference in the technical success in each group (OR, 3.27; 95% CI, .52-20.53; \( I^2 = 65\% \)) (Supplementary Fig. 1, available online at www.giejournal.org).10,16,21,25,28

Specimen adequacy was reported in 11 studies, and the adequacy of brush cytology was found to be higher than that of fluoroscopic-guided biopsy sampling (OR, 2.28;
95% CI, 1.1-4.74; $I^2 = 63\%$). This analysis was performed on an intention-to-treat basis. Of 3 studies in which the specimen adequacy favored brushing, 2 were mostly because of technical failure in obtaining the biopsy sample itself.\textsuperscript{10,20,25}

### Adverse events

Five studies reported adverse events between brush cytology and fluoroscopic-guided biopsy sampling.\textsuperscript{16,20,21,27,30} There was no difference in adverse events with either brushing or intraductal biopsy sampling (OR, 53; 95% CI, 1.14-2.05; $I^2 = 0\%$) (Fig. 4).

The 2 reported adverse events in the 503 patients within the brushing group included 1 retroperitoneal perforation treated with stent placement\textsuperscript{20} and 1 incidence of mild pancreatitis.\textsuperscript{30} Three mild and 2 severe adverse events occurred in 518 patients who underwent fluoroscopic-guided biopsy sampling. The mild adverse events were a mid-bile duct perforation treated with stent placement and 2 cases of pancreatitis.\textsuperscript{16,30} There was 1 incidence of prolonged bleeding after obtaining a biopsy sample of a proximal bile duct tumor that required hospitalization, 4 units of red blood cell transfusion, and placement of a nasobiliary tube.\textsuperscript{27} One patient with a benign stricture developed peritonitis after the ERCP with intraductal biopsy sampling and required an exploratory laparotomy with choledochotomy and suture closure of the common hepatic duct perforation.\textsuperscript{21}

### Intervention time

Brush cytology took 3.75 minutes (95% CI, 2.8-4.71) shorter to perform than fluoroscopic-guided biopsy sampling in the 2 studies that reported the mean time for each intervention.\textsuperscript{10,16}

### Subgroup analyses

Subgroup analyses did not find a difference in the sensitivities between distal and proximal strictures or primary biliary and pancreatic masses (Supplementary Table 2, available online at www.giejournal.org).

### Certainty of the evidence

The risk of bias assessment for each study can be found in Supplementary Table 3 (available online at www.giejournal.org). The certainty of evidence for all clinical outcomes for PICO question 1 were downgraded because only observational studies were included (Fig. 5). For the main outcome of incremental yield of fluoroscopic-guided biopsy sampling, no other downgrades were applied for an overall low certainty. For the other main outcome of diagnostic test characteristics, the certainty of evidence was high, except the sensitivity of fluoroscopic-guided biopsy sampling had a high $I^2$ value, lowering that to moderate. The remainder of the secondary analyses was very low and downgraded for indirectness (may not be generalizable to community centers) and imprecision (wide CIs) for technical success, risk of bias for specimen adequacy, imprecision (low number of patients) for adverse events, and imprecision (wide CIs) for intervention time.

### Other considerations

Both biopsy forceps and cytology brushes should be readily available at any endoscopy center, whether in a community setting or at tertiary referral centers. The difference in cost between biopsy forceps and brush cytology is negligible and should be similar throughout the country. In terms of cost-effectiveness, I studied the cost utility of ERCP-based techniques in the diagnosis of cholangiocarcinoma in patients with primary sclerosing cholangitis.\textsuperscript{34} When comparing ERCP with intraductal biopsy sampling and brush cytology, the authors found that ERCP with fluoroscopic-guided biopsy sampling was cost-effective based on a willingness-to-pay threshold of less than $50,000.

The patient representative preferred the modality that would more likely provide an earlier diagnosis. However, the representative was cautious about the possible severe adverse events that occurred in patients who underwent fluoroscopic-guided biopsy sampling.

### Discussion

ERCP with brush cytology is the most common modality of tissue acquisition performed in patients with biliary...
stricture because of its ease and availability.\textsuperscript{35,36} However, it is known that the sensitivity of brush cytology in the diagnosis of malignancy is low.\textsuperscript{37} Our pooled sensitivity for brush cytology was 40%, which is similar to previous meta-analyses and noted to be suboptimal.\textsuperscript{36,37} Adding fluoroscopic-guided biopsy sampling had an incremental incremental yield - brushing + biopsy vs brushing alone

<table>
<thead>
<tr>
<th>Certainty assessment</th>
<th>Ne of patients</th>
<th>Effect</th>
<th>Certainty</th>
<th>Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Incremental yield - brushing + biopsy vs brushing alone</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7 observational studies</td>
<td>not serious</td>
<td>not serious</td>
<td>not serious</td>
<td>none</td>
</tr>
<tr>
<td>Miss rate of biopsies</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10 observational studies</td>
<td>not serious</td>
<td>serious\textsuperscript{a}</td>
<td>not serious</td>
<td>not serious</td>
</tr>
<tr>
<td>Technical success</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8 observational studies</td>
<td>not serious</td>
<td>not serious\textsuperscript{a}</td>
<td>serious\textsuperscript{a}</td>
<td>serious\textsuperscript{a}</td>
</tr>
<tr>
<td>Specimen adequacy</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11 observational studies</td>
<td>serious\textsuperscript{a}</td>
<td>not serious\textsuperscript{a}</td>
<td>not serious</td>
<td>not serious</td>
</tr>
<tr>
<td>Adverse events</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7 observational studies</td>
<td>serious\textsuperscript{a}</td>
<td>not serious</td>
<td>not serious</td>
<td>serious\textsuperscript{f}</td>
</tr>
<tr>
<td>Intervention time</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 observational studies</td>
<td>not serious</td>
<td>serious\textsuperscript{a}</td>
<td>not serious</td>
<td>not serious</td>
</tr>
</tbody>
</table>

CI: confidence interval; MD: mean difference; OR: odds ratio

**Explanations**

a. high I\textsuperscript{2}  
b. not generalizable to community centers  
c. small amount of reported events  
d. only a few small studies reported on this topic  
e. more poor quality than good/fair  
f. large pooled estimate

**Figure 5.** Certainty of evidence profile for population, intervention, comparator, outcomes question 1. CI, Confidence interval; MD, mean difference; OR, odds ratio.
yield of 20% to brushings alone, resulting also in higher sensitivity than biopsy sampling alone (66% vs 52%). Therefore, our panel was in favor of routinely adding fluoroscopic-guided biopsy sampling to brush cytology in the workup of biliary strictures of undetermined etiology.

However, several concerns were raised during the panel discussion, hence the conditional recommendation. Fluoroscopic-guided biopsy sampling is most commonly performed freehand alongside the wire rather than wire-guided, which can be more time-consuming and requires extra technical skill. This can sometimes be overcome by using an over-the-wire biopsy forceps (Histoguide; Steris, Mentor, Ohio, USA) or a double-lumen cytology brush device (Cytomax II; Cook Medical, Bloomington, Ind, USA) as mini-overtubes.58 The synthesized studies were performed exclusively at tertiary care centers where intraductal biopsy sampling is more commonly performed, so the results may not be generalizable to other settings. In addition, the optimal number of biopsy samples needed for maximal accuracy is unknown, but a median of 2.9 biopsy samples (range, 2-4) was obtained in the summarized studies. Furthermore, although the overall number of adverse events was low, the only severe adverse events occurred with the biopsy forceps. The panel recognized these limitations and therefore made the routine use of intraductal biopsy sampling with brush cytology a conditional recommendation. Because the subgroup analysis did not find any difference in patients with proximal and distal bile duct strictures, this recommendation applies to any biliary structure of undetermined etiology.

**Question 2:** In patients with biliary strictures of undetermined etiology, should ERCP with cholangioscopy-guided biopsy sampling be performed versus ERCP without cholangioscopy to diagnose malignancy?

**Recommendation 2.** In patients with biliary strictures of undetermined etiology undergoing ERCP, the ASGE suggests the use of cholangioscopic-guided biopsy sampling in

a. Nondistal biliary strictures where there is a high probability of adequate drainage of the critical liver segment or

b. Previous nondiagnostic ERCP without cholangioscopy and

c. Centers with clinical expertise and easy access to the equipment.

Otherwise, the ASGE suggest ERCP with or without cholangioscopy to diagnosis malignancy.

(Conditional recommendation/very low quality of evidence)

We performed a systematic review and meta-analysis for this question. An initial search yielded 998 total studies, and the updated search yielded an additional 344 studies (Appendix 2, available online at www.giejournal.org). From these 2 searches, 31 studies underwent full-text re-

**Diagnostic accuracy**

Diagnostic test characteristics were calculated for all 13 included studies (Table 4). There was a significantly higher sensitivity for ERCP with cholangioscopy-guided biopsy sampling (.72; 95% CI, .66-.77; \( I^2 = 79.9\% \)) than without cholangioscopy (.61; 95% CI, .57-.66; \( I^2 = 71.8\% ; P = .001 \)). Furthermore, the SROC was higher for ERCP with cholangioscopy (area under the curve, .9689 for cholangioscopy vs .7495 without cholangioscopy).

**Technical success and specimen adequacy**

All studies that reported on the technical success of ERCP with and without cholangioscopy had a 100% technical success rate for both interventions.10,31,39 Specimen adequacy was mentioned in 4 studies.10,31,39,40 There was no difference in the ability to obtain adequate tissue specimens by cholangioscopy-guided biopsy sampling, fluoroscopic-guided biopsy sampling, or brush cytology (OR, 96; 95% CI, .23-4; \( I^2 = 0\% \)) (Supplementary Fig. 2, available online at www.giejournal.org).

**Adverse events**

There was no difference in the number of adverse events reported in patients who underwent ERCP with and without cholangioscopy (21/72 patients without cholangioscopy and 16/81 patients with cholangioscopy; OR, .58; 95% CI, .26-1.26; \( I^2 = 0\% \)) (Fig. 7).51,41,42 In the ERCP without cholangioscopy group, 21 of 72 patients...
had adverse events, all of which were mild (12 mild pancreatitis, 6 cholangitis, 1 cholecystitis, 1 bleeding, and 1 pulmonary disorder). In the ERCP with cholangioscopy group, 16 of 81 patients had adverse events. Of these, 3 were severe adverse events: 2 severe pancreatitis and 1 severe bleeding, although the bleeding was attributed to the sphincterotomy (which is required for cholangioscopy) itself rather than the cholangioscopic-guided biopsy. The remaining 13 patients had mild adverse events (9 cases of mild pancreatitis, 2 cases of minor bleeding related to the sphincterotomy, and 2 with cholangitis). These adverse events reported in the cholangioscopy group were not specific to the cholangioscopy technique itself and reflect the inherent risks of ERCP.

**Intervention time**

One RCT reported on the mean time required to perform each form of tissue acquisition: 31.95 (standard deviation, 3.11) minutes for performing fluoroscopic-guided biopsy sampling and brush cytology during ERCP versus 23.64 (standard deviation, 9.43) minutes for the cholangioscopy portion of the ERCP. The mean difference

---

**Figure 6.** Incremental yield of cholangioscopy. CI, Confidence interval; REML, random effects model.

**Figure 7.** Adverse events in patients who underwent ERCP with and without cholangioscopy. CI, Confidence interval.

**TABLE 4. Pooled test characteristics for ERCP with and without cholangioscopy**

<table>
<thead>
<tr>
<th>Study</th>
<th>ERCP without cholangioscopy</th>
<th>ERCP with cholangioscopy</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sensitivity</strong></td>
<td>.61 (.57-.66)</td>
<td>.72 (.66-.77)</td>
</tr>
<tr>
<td><strong>Specificity</strong></td>
<td>.93 (.91-.96)</td>
<td>.96 (.92-.98)</td>
</tr>
<tr>
<td><strong>Positive likelihood ratio</strong></td>
<td>11.31 (3.42-37.35)</td>
<td>10.61 (6.57-17.12)</td>
</tr>
<tr>
<td><strong>Negative likelihood ratio</strong></td>
<td>.48 (.39-.6)</td>
<td>.32 (.22-.46)</td>
</tr>
<tr>
<td><strong>Diagnostic odds ratio</strong></td>
<td>23.21 (9.56-56.31)</td>
<td>59.72 (27.85-128.02)</td>
</tr>
<tr>
<td><strong>Summary receiver-operating characteristic curve</strong></td>
<td>.7495</td>
<td>.9689</td>
</tr>
</tbody>
</table>

**Table 4.** Pooled test characteristics for ERCP with and without cholangioscopy.
was +14.15 minutes (95% CI, 10.33-19.97) for cholangioscopy with biopsy sampling.

**Subgroup analyses**

One study compared ERCP with and without cholangioscopy in patients with either distal or proximal bile duct stricture. In patients with distal strictures, the sensitivity of ERCP with cholangioscopic-guided biopsy sampling was only 50% as compared with 76% with fluoroscopic-guided biopsy sampling. There was no statistically significant difference in sensitivity in biopsy sampling of proximal strictures using ERCP with cholangioscopic-guided biopsy sampling (sensitivity, 67%) and fluoroscopic-guided biopsy sampling (sensitivity, 73%).

A different study showed similar sensitivities in ERCP with cholangioscopic-guided biopsy sampling and fluoroscopic-guided biopsy sampling in patients with biliary masses (sensitivities of 50% and 58.3%, respectively). Although patients with pancreatic masses had a 100% sensitivity using cholangioscopic-guided biopsy sampling in this study, only 2 patients were included in this group. Meanwhile, in the 9 patients with pancreatic masses, the sensitivity was 22.2% in those who underwent fluoroscopic-guided biopsy sampling.

**Certainty of the evidence**

Risk of bias assessment for each study can be found in Supplementary Table 4 (available online at www.giejournal.org), whereas the summary of evidence is shown in Figure 8. The certainty of evidence focusing only on the RCT for the main analysis on incremental yield was downgraded for imprecision because of a low total number of patients and large CI, making the final rating moderate. With the 4 studies combined, the overall grade was very low because of the observational study designs and large CIs. The diagnostic test characteristics of both ERCP with and without cholangioscopy were downgraded to moderate for high inconsistency. Because the other secondary analyses were predominately based on observational studies, the evidence profile was already low.

**Other considerations**

The panel considered the cost of cholangioscopy, which was deemed to be high. One study quoted the total direct cost (including procedure and recovery personnel, devices, stent placement, sterilization) of an ERCP with stent placement to be $893, whereas the total direct cost of an ERCP with Spyglass cholangioscopy (Boston Scientific Corp, Natick, Mass, USA) and stent placement was $3530. The 2022 quoted cost from the Boston Scientific representative for the Spyglass digital controller was $132,825 (although common costs may vary for individual institutions free to facilities based on contractual agreements at no extra charge), Spyglass access and delivery catheter to be $2750, and the Spyglass biopsy forceps to be $535.

One study evaluated the cost utility of ERCP-based techniques in the diagnosis of cholangiocarcinoma in primary sclerosing cholangitis patients. The use of cholangioscopy was cost-effective at willingness-to-pay thresholds of $50,000 and $100,000. In fact, cholangioscopy was the most cost-effective diagnostic strategy in this study. Another study found that the use of cholangioscopy decreased the total number of procedures required for diagnosis (31% relative reduction) and costs (~$14,125 dollars; ~5% relative variation) when compared with ERCP without cholangioscopy. The patient representative valued the overall increased diagnostic yield of cholangioscopy. The panel noted that that cholangioscopy is not widely available and may require the patient to be evaluated at a tertiary referral center with expertise.

**Discussion**

Cholangioscopy allows the endoscopist to have direct visualization of the biliary tree and target intraductal biopsy sampling. The panel noted the improvement in diagnostic yield with cholangioscopy-assisted biopsy sampling during ERCP but also noted that severe adverse events, including pancreatitis and cholangitis, only occurred in those undergoing cholangioscopy.

The panel raised several concerns that resulted in qualifying the recommendation as conditional. The main concern related to the cost and availability of the cholangioscopy system. Cholangioscopy systems have evolved tremendously, with improvements in images, device maneuverability, and devices, that may make cholangioscopy more available and easier to use. However, patients most often have to travel to centers with expertise in ERCP with cholangioscopy. Despite the high costs of ERCP with cholangioscopy, it is still considered to be a cost-effective diagnostic modality and an important tool to use when available, especially with prior nondiagnostic ERCP without cholangioscopy. Another concern the panel raised was regarding the subgroup of distal biliary strictures where the sensitivity of ERCP with cholangioscopy + biopsy sampling was lower compared with ERCP with fluoroscopic-guided biopsy sampling. The cholangioscopy system is often unstable in the pямullary location and tends to migrate out of the duct or to be torqued in such a way that visualization and biopsy sample acquisition are more difficult in the distal duct. Therefore, the panel recognized that ERCP with cholangioscopy may not be as effective in distal locations. Finally, there is a concern of inadvertently introducing infection through the use of cholangioscopy when it is unclear whether the segment proximal to the stricture is amenable for adequate drainage after cholangioscopy. Cholangioscopy requires water or saline solution to be injected into the bile duct for visualization, which can introduce bacterial contamination to proximal segments of the liver or cause bacterial translocation. Therefore, the panel wanted to emphasize the need to ensure adequate drainage of the
duct proximal to the stricture before the use of cholangioscopy. Some experts on the panel did not routinely perform cholangioscopy during the initial ERCP because of this reason. Other experts on the panel consider performing...
cholangioscopy during the initial ERCP if they believe that adequate drainage is feasible.

**Question 3:** In patients with biliary strictures of undetermined etiology, should EUS with FNA or FNB be performed versus ERCP with any form of tissue acquisition to diagnose malignancy?

**Recommendation 3.** In patients with biliary strictures of undetermined etiology undergoing ERCP, the ASGE suggests EUS in addition to ERCP for the diagnosis of malignancy in the presence of

- Prior ERCP with nondiagnostic ERCP results,
- Distal biliary stricture, or
- Presence of lymphadenopathy or metastatic disease on cross-sectional imaging.

*(Conditional recommendation/very low quality of evidence)*

We performed a systematic review and meta-analysis on patients with biliary strictures of undetermined etiology who underwent EUS and ERCP for the diagnosis of malignancy. An initial search yielded 1869 total studies, and the updated search yielded an additional 510 studies (Appendix 3, available online at www.giejournal.org). From these 2 searches, 34 studies underwent full-text review. One meta-analysis was identified and looked at the incremental yield of EUS. Twelve studies (1536 patients) were included in the remaining analyses. All studies were observational, with full-text articles.

**Incremental yield**

A meta-analysis reported on the incremental benefit of EUS in 10 studies (1162 patients). No additional studies were found in our systematic search to include in the analysis. This meta-analysis focused on the incremental benefit of EUS after a nondiagnostic ERCP with brush cytology. The authors calculated the incremental benefit of EUS by dividing the total number of patients who underwent EUS and ERCP, where only the EUS had a positive malignant diagnosis, by the total number of patients who underwent ERCP with brush cytology. The pooled incremental benefit of EUS was found to be 15% (95% CI, 9-24; $I^2 = 0$).

**Diagnostic accuracy**

The pooled diagnostic test characteristics for tissue acquisition using either EUS or ERCP were similar (Supplementary Table 5, available online at www.giejournal.org). Although the meta-analysis used to calculate the incremental yield focused only on prior negative ERCPs with brush cytology, the studies included in the pooled diagnostic test characteristics were based on 1 study with brush cytology only, 3 studies with fluoroscopic-guided biopsy sampling, 4 studies with either intraductal biopsy sampling or brushing, and 4 studies with both biopsy sampling and brushing.

In the 8 studies that assessed EUS with ERCP versus ERCP alone, there was a higher sensitivity for the combined procedures (.88; 95% CI, .85-.91; $I^2 = 86.4\%$) versus ERCP alone (.61; 95% CI, .57-.64; $I^2 = 53.6\%; P < .001) (Table 5). In addition, in these studies, the pooled SROC was also high for EUS + ERCP at .9799.

**Technical success and specimen adequacy**

No significant difference was found in the 5 studies that reported on the technical success of tissue acquisition using either ERCP or EUS (OR, .39; 95% CI, .08-1.89; $I^2 = 70\%$). Although specimen adequacy favored EUS with FNA, there was no statistical difference between the ability to acquire an adequate specimen using either EUS with FNA or ERCP (OR, .4; 95% CI, .14-1.13; $I^2 = 10\%$).

**Adverse events**

EUS-guided FNA had a statistically significant lower adverse event rate than ERCP (OR, 8.11; 95% CI, 2.95-22.29; $I^2 = 0\%$) (Table 5). In addition, in these studies, there was no statistical difference in the 5 studies that reported on the technical success of tissue acquisition using either ERCP or EUS (OR, .39; 95% CI, .08-1.89; $I^2 = 70\%$) (Fig. 9). In 3 patients, minor bleeding was reported after EUS + FNA. Forty-four adverse events were reported in the ERCP group, 1 severe pancreatitis and 43 mild events (27 mild pancreatitis, 10 cholangitis, and 6 bleeding).

**Intervention time**

One study compared the mean time to do EUS + ERCP (74 [standard deviation, 14] minutes) with historical control subjects who had an ERCP alone performed by the same endoscopists (mean time, 56 [standard deviation, 25] minutes) and found a mean difference of +23 minutes (95% CI, 14-32) with the addition of EUS-guided FNA.

**Subgroup analyses**

There was a higher sensitivity of EUS-guided FNA in distal strictures as shown in 2 studies. The pooled sensitivity of EUS-guided FNA of distal strictures was .82 (95% CI, .76-.87) versus .62 (95% CI, .55-.69) for ERCP. In these 2 studies, there was a trend but no significant difference in the sensitivity of EUS-guided FNA and ERCP in proximal strictures. The pooled sensitivity of EUS-guided FNA in distal proximal strictures was .67 (95% CI, .5-.8) versus .48 (95% CI, .32-.64) for ERCP.

In addition, EUS-guided FNA was found to have a higher sensitivity in patients with biliary strictures related to a pancreatic mass seen on cross-sectional imaging compared with ERCP. In 6 studies, the pooled sensitivity of EUS in the setting of pancreatic masses was .82 (95% CI, .78-.86) versus .46 (95% CI, .4-.51; $P < .0001$) for ERCP.

**Certainty of the evidence**

The certainty of evidence profile is summarized in Figure 10, and the risk of bias assessment is shown in Supplementary Table 6 (available online at www.giejournal.org).
The incremental yield analysis was based on 10 studies included in the meta-analysis, all of which were observational studies, precluding a low certainty of evidence. Half of the diagnostic test characteristics were downgraded to a moderate certainty of evidence because of inconsistency, whereas the others remained at a high certainty. Specimen adequacy remained a low certainty for including observational studies only, whereas technical success analysis was downgraded to very low certainty for inconsistency, indirectness (unclear if it could be applied to all endoscopy centers), and imprecision (high CIs). The adverse event analysis was downgraded to very low for imprecision because of a low overall number of events, whereas intervention time was downgraded to very low for indirectness (compared with historical control subjects) and imprecision (only 1 study).

### Other considerations

The total direct cost (including procedure and recovery personnel, devices, stent placement, sterilization) of an ERCP with stent placement in 1 study was $892.99, whereas that of EUS with FNA was $1076.25. EUS was found to be more cost-effective in patients with a biliary stricture. In this study, ERCP resulted in 9.05 quality-adjusted life-years and a cost of $34,685.11 for a cost-effectiveness ratio of $3832.33, whereas EUS resulted in an incremental increase in .13 quality-adjusted life-years and $2773.69 for an incremental cost-effectiveness ratio of $20,840.28 per quality-adjusted life-year gained. The patient representative expressed some concerns about the length of time it took to do an EUS but indicated that the higher diagnostic yield with lower adverse events of EUS outweighed this concern.

### Discussion

The incremental benefit of EUS-guided FNA in patients with nondiagnostic ERCP with brush cytology was 15% as found by the meta-analysis. Furthermore, there was a significantly higher sensitivity when EUS-guided FNA was combined with ERCP compared with ERCP alone (.88 vs .61, respectively; P < .001). This improvement in diagnostic yield using EUS-guided FNA was influenced by the presence of lymph node metastases, because it has been shown that 15% to 20% of patients with cholangiocarcinoma have lymph node metastases diagnosed by EUS after negative abdominal imaging. With the improvement in diagnostic yield using EUS and the significantly lower adverse event rate, the panel was in favor of using EUS in the diagnostic approach to biliary strictures of undetermined etiology.

There were several caveats to the conditional recommendation of EUS. The primary benefit of EUS was in

---

**Figure 9.** Adverse events in patients who underwent EUS and ERCP. CI, Confidence interval; AE, adverse event.
combination with ERCP; therefore, the panel agreed that EUS should be considered when it is available and be performed during the same session as the ERCP. In fact, some experts in the panel recommended performing EUS on any biliary strictures of undetermined etiology regardless of its location. Because the subgroup analysis showed that EUS
was particularly beneficial in patients with distal biliary strictures and pancreatic masses, the panel agreed that these indications should be emphasized. Given the ability of EUS to sample the concerning lesion at the same time and to avoid repeat procedures for a pending diagnosis, performing EUS at same time the patient is undergoing ERCP when possible was believed to be reasonable by most experts.

It must be emphasized that a significant risk of needle-tract seeding is possible during EUS with FNA or FNB of hilar cholangiocarcinoma. Heimbach et al reported that 83% of patients who had a positive transperitoneal FNA of the primary hilar mass had peritoneal metastases during operative staging before liver transplantation, whereas only 8% of patients who did not undergo transperitoneal FNA had peritoneal metastases ($P = .0097$). It was recommended that biopsy sampling of the hilar mass should not be performed in patients who are otherwise candidates for curative surgery. Therefore, if an EUS is performed in the setting of proximal or hilar strictures, the endosonographer should not perform FNA or FNB of the biliary mass itself. EUS-guided FNA or FNB may still be helpful in the diagnostic workup of hilar strictures, particularly if there are lymph nodes or metastatic lesions that can be targeted instead of the primary mass itself.

GUIDELINE UPDATE

ASGE guidelines are reviewed for updates approximately every 5 years or in the event that new data may influence a recommendation. Updates follow the same ASGE guideline development process.

DISCLOSURE

ACKNOWLEDGMENTS

We are grateful to Toni Pham from the Cholangiocarcinoma Foundation for her input as a patient advocate on this guideline panel and to Kellie Kaulback (librarian) and Robyn Rosasco (librarian) for assistance with searching for articles. We also thank Dr Tiffany Chua and Dr Ramzi Mulki on behalf of the Gastrointestinal Endoscopy Editorial Board and Dr Bret Petersen for their review of the guidelines. This guideline was funded exclusively by the American Society for Gastrointestinal Endoscopy; no outside funding was received to support the development of this guideline.

REFERENCES

Medical Center, San Francisco, California, USA (22), Division of Gastroenterology and Hepatology, UMass Chan Medical School, Worcester, Massachusetts, USA (23), Department of Gastroenterology, Wake Forest School of Medicine, Winston Salem, North Carolina, USA (24), Division of Gastroenterology, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, Massachusetts, USA (25), Division of Gastroenterology and Hepatology, Loma Linda University, Loma Linda, California, USA (26), Department of Gastroenterology, Hepatology, and Nutrition, University of Florida, Gainesville, Florida, USA (27).

Reprint requests: Bashar J. Qumseya, MD, MPH, FASGE, Department of Gastroenterology, Hepatology and Nutrition, University of Florida, PO Box 100214, 1329 SW 16th St, Ste 5251, Gainesville, FL 32610-0214.
APPENDIX

APPENDIX 1. SEARCH STRATEGY FOR POPULATION, INTERVENTION, COMPARATOR, OUTCOMES QUESTION 1

Search date: May 28, 2021
Databases searched: Ovid MEDLINE: Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Ovid MEDLINE Daily and Ovid MEDLINE 1946 to Present; Embase.com (Elsevier) (1947 to 2021 May 28; Wiley Cochrane Library [Cochrane Database of Systematic Reviews {CDSR}, Cochrane Central Register of Controlled Trials {CENTRAL}])

Limits: English, human
Excluded: letters, notes, comments, editorials, case reports; conference abstracts or congresses before 2019.

Ovid MEDLINE ALL
1 exp Bile Ducts/ use ppez 47,710
2 exp Bile Duct Neoplasms/ use ppez 18,934
3 (bile duct* or biliary or hilar or peri*hilar or klatskin).ti,ab,kw. 130,221
4 or/1-3 149,325
5 exp Constriction, Pathologic/ use ppez 31,366
6 (constriction OR stricture* OR stenosis OR obstruction OR occlusion OR blockage).ti,ab,kw. 516,699
7 or/5-6 524,413
8 4 and 7 21,148
9 cholestasis.ti,ab,kw. 15,804
10 ((Bile duct* or biliary or hilar or peri*hilar or klatskin) adj2 (carcinoma* or adenoma* or adenocarcinoma* or neoplasm* or tumor* or tumour* or cholangiocarcinoma* or malignant* or stricture* or obstruction or occlusion or stenos* or blockage)).ti,ab,kw. 18,820
11 or/8-10 42,063
12 exp Cytodiagnosis/ use ppez 312,960
13 exp Cytological Techniques/ use ppez 1,520,003
14 *Specimen Handling/ use ppez or exp Specimen Handling/mt 65,114
15 or/12-14 1,543,954
16 11 and 15 31,888
17 ((biliary or bile duct*) adj5 (brush* or scrape)).ti,ab,kw. 320
APPENDIX 2. SEARCH STRATEGY FOR POPULATION, INTERVENTION, COMPARATOR, OUTCOMES QUESTION 2

Search date: May 28, 2021

Databases searched: Ovid MEDLINE: Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Ovid MEDLINE Daily and Ovid MEDLINE 1946-Present; Embase.com (Elsevier) (1947 to 2021 May 28; Wiley Cochrane Library [Cochrane Database of Systematic Reviews {CDSR}, Cochrane Central Register of Controlled Trials {CENTRAL}])

Limits: English, Human

Exclusions: Conference abstracts pre-2014, letters, notes, comments, editorials, case reports

Ovid MEDLINE ALL
1 exp Bile Ducts/ use ppez 47,710
2 exp Bile Duct Neoplasms/ use ppez 18,934
3 (bile duct* or biliary or hilar or peri*hilar or klatskin).ti,ab,kw. 130,221
4 or/1-3 149,325
5 exp Constriction, Pathologic/ use ppez 31,366
6 (constriction or stricture* or stenosis* or obstruction or occlusion or blockage).ti,ab,kf,kw. 516,699
7 5 or 6 524,413
8 4 and 7 21,148
9 exp Cholestasis/ use ppez 34,217
10 cholestasis.ti,ab,kw. 15,804
11 ((Bile duct* or biliary or hilar or peri*hilar or klatskin) adj2 (carcinoma* or adenoma* or adenocarcinoma* or neoplasm* or tumor* or tumour* or cholangiocarcinoma* or malignanc* or stricture* or obstruction or occlusion or stenosis* or blockage)).ti,ab,kf,kw. 18,820
12 or/8-11 60,862
13 exp endoscopy, gastrointestinal/ use ppez or exp biliary tract surgical procedures/ use ppez 127,534
14 (Choledochoscope* or cholangioscope* or Cholangiopancreatoscope* or spyglass).ti,ab,kf,kw. 2188
15 13 or 14 129,205
16 12 and 15 5654
17 animals/ not (humans/ and animals/) 4,800,822
18 16 not 17 5487
19 limit 18 to English language 4129
20 (case reports or comment or editorial or letter).pt. 3,924,944
21 Case Report/ 2,180,861
22 19 not (20 or 21) 2957
23 limit 22 to dt 20190530-20211231 196

Diagnosis of malignancy in biliary strictures of undetermined etiology

Excluded: letters, notes, comments, editorials, case reports; conference abstracts or congresses before 2019.

- **INCLUSION CRITERIA**
  - English
  - Cohort, case control, randomized control studies
  - Full text
  - Abstracts presented at DDW, UEGW, ACG, AASLD, EASL conferences over the past 5 years
  - Human subjects

- **EXCLUSION CRITERIA**
  - Case report
  - Case series with n < 20 (cut off for all meta-analyses)
  - Reviews, editorials, letters to the editor
  - Insufficient data to make adequate tables

### Ovid MEDLINE ALL
No. Searches No. of results
1. exp Bile Ducts/ use ppez 47,710
2. exp Bile Duct Neoplasms/ use ppez 18,934
3. (bile duct* or biliary or hilar or peri*hilar or klatskin).-ti,ab,kf,kw. 130,221
4. or/1-3 149,325
5. exp Constriction, Pathologic/ use ppez 31,366
6. (constriction or stricture* or stenos?is or obstruction or occlusion or blockage).ti,ab,kf,kw. 516,699
7. 5 or 6 524,413
8. 4 and 7 21,148
9. exp Cholestasis/ use ppez 34,217
10. cholestasis.ti,ab,kf,kw. 15,804
11. ((Bile duct* or biliary or hilar or peri*hilar or klatskin) adj2 (carcinoma* or adenoma* or adenocarcinoma* or neoplasm* or tumor* or tumour* or cholangiocarcinoma* or malignant* or stricture* or obstruction or occlusion or stenosis? is or blockage)).ti,ab,kf,kw. 18,820
12. or/8-11 60,862
13. *Endosonography/ use ppez 7565
14. *Biopsy, Fine-Needle/ use ppez 3950
15. (eus or FNA or fine needle or (endoscop* adj2 ultrasound) or endosonograph*).ti,ab,kf,kw. 51,124
16. or/13-15 54,063
17. 12 and 16 1554
18. animals/ not (humans/ and animals/) 4,800,822
19. 17 not 18 1548
20. limit 19 to English language 1377
21. (case reports or comment or editorial or letter).pt. 3,924,944
22. Case Report/ 180,861
23. 20 not (21 or 22) 1024
24. limit 23 to dt=-20190530-20211231 179

### Embase.com (Elsevier)
No. Searches
1. “bile duct”/exp

---

**Cochrane Library (CDSR, CENTRAL – Wiley)**

#1 [mh “Bile Ducts”]
#2 [mh “Bile Duct Neoplasms”]
#3 (bile duct* or biliary or hilar or peri*hilar or klatskin)
#4 #1 or #2 or #3
#5 [mh “Constriction, Pathologic”]
#6 (constriction or stricture* or stenosis? or obstruction or occlusion or blockage)
#7 #5 or #6
#8 #4 and #7
#9 cholestasis
#10 ((bile duct* or biliary or hilar or peri*hilar or klatskin) adj2 (carcinoma* or adenoma* or adenocarcinoma* or neoplasm* or tumor* or tumour* or cholangiocarcinoma* or malignant* or stricture* or obstruction or occlusion or stenosis? or blockage))
#11 #8 or #9 or #10
#12 Choledochoscop* or cholangioscop* or Cholangiopancreatoscop* or spyglass
#13 #11 and #12

Date added to CENTRAL trials database: May 30, 2019

Results: 13

**APPENDIX 3. SEARCH STRATEGY FOR POPULATION, INTERVENTION, COMPARATOR, OUTCOMES QUESTION 3**

Search date: May 28, 2021

Databases searched: Ovid MEDLINE: Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Ovid MEDLINE Daily and Ovid MEDLINE 1946-Present; Embase.com (Elsevier) (1947 to 2021 May 28; Wiley Cochrane Library [Cochrane Database of Systematic Reviews (CDSR), Cochrane Central Register of Controlled Trials (CENTRAL)])

Limits: English, human
Diagnosis of malignancy in biliary strictures of undetermined etiology

2 'bile duct tumor'/exp
3 (bile duct* OR biliary OR hilar OR peri*hilar OR klatskin):ti,ab,kw
4 #1 OR #2 OR #3
5 ‘stenosis, occlusion and obstruction’/exp
6 (constriction OR stricture* OR stenos?s OR obstruction OR occlusion OR blockage):ti,ab,kw
7 #5 OR #6
8 #4 AND #7
9 cholestasis/exp
10 cholestasis:ti,ab,kw
11 ((‘bile duct*’ OR biliary OR hilar OR peri*hilar OR klatskin) NEAR/2 (carcinoma* OR adenoma* OR adenocarcinoma* OR neoplasm* OR tumor* OR tumour* OR cholangiocarcinoma* OR malignanc* OR stricture* OR obstruction OR occlusion OR stenos?s OR blockage)):ti,ab,kw
12 #8 OR #9 OR #10 OR #11
13 ‘endoscopic ultrasonography’/de
14 ‘fine needle aspiration biopsy’/de
15 (eus OR FNA OR fine needle OR (endoscop* NEAR/2 ultraso*) OR endosonograph*):ti,ab,kw
16 #13 OR #14 OR #15
17 #12 AND #16
18 animals/exp NOT (humans/exp AND animals/exp)
19 #17 NOT #18
20 #19 AND English:la
21 ‘case reports’:it OR comment:it OR editorial:it OR letter:it OR note:it
22 ‘Case Report’:de
23 #20 NOT (#21 OR #22)

24 #23 AND [30-05-2019]/sd
Results: 459

Cochrane Library (CDSR, CENTRAL – Wiley)
1D Search Hits
#1 [mh “Bile Ducts”]
#2 [mh “Bile Duct Neoplasms”]
#3 (bile duct* or biliary or hilar or peri*hilar or klatskin):ti,ab
#4 #1 or #2 or #3
#5 [mh “Constriction, Pathologic”]
#6 (constriction or stricture* or stenos?s or obstruction or occlusion or blockage):ti,ab
#7 #5 or #6
#8 #4 and #7
#9 cholestasis:ti,ab
#10 ((Bile duct* or biliary or hilar or peri*hilar or klatskin) NEAR/2 (carcinoma* or adenoma* or adenocarcinoma* or neoplasm* or tumor* or tumour* or cholangiocarcinoma* or malignanc* or stricture* or obstruction or occlusion or stenos?s or blockage)):ti,ab
#11 #8 or #9 or #10
#12 [mh Endosonography]
#13 [mh “Biopsy, Fine-Needle”]
#14 (eus or FNA or fine needle or (endoscop* near/2 ultra-so*) or endosonograph*):ti,ab
#15 #12 or #13 or #14
#16 #11 and #15
Date added to CENTRAL trials database: May 30, 2019 to present
Results: 59
Supplementary Figure 1. Technical success of brush cytology versus fluoroscopic-guided biopsy sampling. CI, Confidence interval.

Supplementary Figure 2. Specimen adequacy for ERCP with and without cholangioscopy. CI, Confidence interval.
**SUPPLEMENTARY TABLE 1. Newcastle-Ottawa quality assessment**

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Acceptable (star given)</th>
<th>Unacceptable (star not given)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Representativeness</td>
<td>Biliary stricture of undetermined etiology</td>
<td>Known malignant biliary stricture in all patients, only proximal or hilar strictures included, only primary sclerosing cholangitis patients included</td>
</tr>
<tr>
<td>Selection</td>
<td>Population-based or multicenter studies</td>
<td>Single-center or hospital-based studies, different technique used for tissue acquisition</td>
</tr>
<tr>
<td>Ascertainment</td>
<td>Medical records</td>
<td>Self-reported</td>
</tr>
<tr>
<td>Comparability</td>
<td>Controls for confounders: same patients that allow for direct comparisons</td>
<td>No control for confounders: consecutive patients, no baseline characteristics reported</td>
</tr>
<tr>
<td>Assessment of outcome</td>
<td>Secure records</td>
<td>Self-reported</td>
</tr>
<tr>
<td>Follow-up adequacy</td>
<td>Median follow-up 6 mo (enough time to know if malignancy is there)</td>
<td>No statement regarding missing data, medial follow-up &lt; 6 mo</td>
</tr>
</tbody>
</table>

**SUPPLEMENTARY TABLE 2. Subgroup analysis on fluoroscopic-guided biopsy sampling and brush cytology based on location of the stricture**

<table>
<thead>
<tr>
<th>Stricture location</th>
<th>Brush sample sensitivity</th>
<th>Biopsy sample sensitivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distal vs proximal strictures</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distal</td>
<td>.61 (.51-.71)</td>
<td>.64 (.54-.73)</td>
</tr>
<tr>
<td>Proximal</td>
<td>.56 (.4-.71)</td>
<td>.58 (.42-.73)</td>
</tr>
<tr>
<td>Biliary vs pancreatic mass</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Biliary mass</td>
<td>.53 (.47-.6)</td>
<td>.63 (.56-.69)</td>
</tr>
<tr>
<td>Pancreatic mass</td>
<td>.37 (.3-.44)</td>
<td>.46 (.38-.53)</td>
</tr>
</tbody>
</table>
### SUPPLEMENTARY TABLE 3. Risk of bias assessment for studies included in population, intervention, comparator, outcomes question 1

<table>
<thead>
<tr>
<th>Study</th>
<th>Selection (maximum, 4 stars)</th>
<th>Comparability (maximum, 2 stars)</th>
<th>Outcomes (maximum, 3 stars)</th>
<th>Total score (maximum, 9 stars)</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pugliese 198722</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>3</td>
<td>Poor</td>
</tr>
<tr>
<td>Ponchon 199520</td>
<td>***</td>
<td>**</td>
<td>***</td>
<td>8</td>
<td>Good</td>
</tr>
<tr>
<td>Howell 199613</td>
<td>**</td>
<td>**</td>
<td>*</td>
<td>5</td>
<td>Poor</td>
</tr>
<tr>
<td>Sugiyama 199618</td>
<td>***</td>
<td>**</td>
<td>***</td>
<td>8</td>
<td>Good</td>
</tr>
<tr>
<td>Pugliese 199721</td>
<td>***</td>
<td>**</td>
<td>***</td>
<td>8</td>
<td>Good</td>
</tr>
<tr>
<td>Schoeff 199727</td>
<td>***</td>
<td>—</td>
<td>***</td>
<td>6</td>
<td>Poor</td>
</tr>
<tr>
<td>Jailwala 200014</td>
<td>***</td>
<td>**</td>
<td>***</td>
<td>8</td>
<td>Good</td>
</tr>
<tr>
<td>Kitajima 200725</td>
<td>***</td>
<td>**</td>
<td>*</td>
<td>6</td>
<td>Poor</td>
</tr>
<tr>
<td>Weber 200826</td>
<td>*</td>
<td>**</td>
<td>***</td>
<td>6</td>
<td>Poor</td>
</tr>
<tr>
<td>Kulaksiz 201117</td>
<td>**</td>
<td>**</td>
<td>**</td>
<td>6</td>
<td>Fair</td>
</tr>
<tr>
<td>Draganov 201220</td>
<td>***</td>
<td>**</td>
<td>***</td>
<td>8</td>
<td>Good</td>
</tr>
<tr>
<td>Salomao 201528</td>
<td>**</td>
<td>—</td>
<td>*</td>
<td>3</td>
<td>Poor</td>
</tr>
<tr>
<td>Naitoh 201619</td>
<td>***</td>
<td>—</td>
<td>*</td>
<td>4</td>
<td>Poor</td>
</tr>
<tr>
<td>Sakuma 201725</td>
<td>***</td>
<td>**</td>
<td>**</td>
<td>7</td>
<td>Good</td>
</tr>
<tr>
<td>Moura 201818</td>
<td>***</td>
<td>**</td>
<td>***</td>
<td>8</td>
<td>Good</td>
</tr>
<tr>
<td>Ren 201823</td>
<td>**</td>
<td>**</td>
<td>*</td>
<td>4</td>
<td>Poor</td>
</tr>
<tr>
<td>Han 201911</td>
<td>**</td>
<td>—</td>
<td>***</td>
<td>5</td>
<td>Poor</td>
</tr>
<tr>
<td>Hartman 2020</td>
<td>*</td>
<td>**</td>
<td>*</td>
<td>4</td>
<td>Poor</td>
</tr>
<tr>
<td>Kaura 202012</td>
<td>**</td>
<td>—</td>
<td>**</td>
<td>4</td>
<td>Poor</td>
</tr>
<tr>
<td>Yang 202130</td>
<td>***</td>
<td>—</td>
<td>***</td>
<td>6</td>
<td>Poor</td>
</tr>
</tbody>
</table>

Quality assessments thresholds are as follows: Good: 3-4 stars in selection and 1-2 stars in comparability and 2-3 stars in outcomes; Fair: 2 stars in selection and 1-2 stars in comparability and 2-3 stars in outcomes; Poor: 0-1 star in selection or 0 stars in comparability or 0-1 stars in outcomes. Not available, No stars.

### SUPPLEMENTARY TABLE 4. Risk of bias assessment for studies included in population, intervention, comparator, outcomes question 2

<table>
<thead>
<tr>
<th>Study</th>
<th>Selection (maximum, 4 stars)</th>
<th>Comparability (maximum, 2 stars)</th>
<th>Outcomes (maximum, 3 stars)</th>
<th>Total score (maximum, 9 stars)</th>
<th>Quality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fukuda 200539</td>
<td>**</td>
<td>**</td>
<td>***</td>
<td>7</td>
<td>Fair</td>
</tr>
<tr>
<td>Tischendorf 200646</td>
<td>**</td>
<td>**</td>
<td>***</td>
<td>7</td>
<td>Fair</td>
</tr>
<tr>
<td>Draganov 201220</td>
<td>***</td>
<td>**</td>
<td>***</td>
<td>8</td>
<td>Good</td>
</tr>
<tr>
<td>Hartman 201225</td>
<td>***</td>
<td>—</td>
<td>***</td>
<td>6</td>
<td>Poor</td>
</tr>
<tr>
<td>Walter 201620</td>
<td>***</td>
<td>**</td>
<td>***</td>
<td>8</td>
<td>Good</td>
</tr>
<tr>
<td>Kato 201911</td>
<td>**</td>
<td>**</td>
<td>**</td>
<td>6</td>
<td>Fair</td>
</tr>
<tr>
<td>Lee 201923</td>
<td>**</td>
<td>—</td>
<td>***</td>
<td>5</td>
<td>Poor</td>
</tr>
<tr>
<td>Yan 201927</td>
<td>**</td>
<td>—</td>
<td>***</td>
<td>5</td>
<td>Poor</td>
</tr>
<tr>
<td>Kaura 202022</td>
<td>**</td>
<td>—</td>
<td>**</td>
<td>4</td>
<td>Poor</td>
</tr>
<tr>
<td>Onoyama 201914</td>
<td>***</td>
<td>**</td>
<td>***</td>
<td>8</td>
<td>Good</td>
</tr>
<tr>
<td>Han 202111</td>
<td>**</td>
<td>—</td>
<td>***</td>
<td>5</td>
<td>Poor</td>
</tr>
</tbody>
</table>

Quality assessments thresholds are as follows: Good: 3-4 stars in selection and 1-2 stars in comparability and 2-3 stars in outcomes; Fair: 2 stars in selection and 1-2 stars in comparability and 2-3 stars in outcomes; Poor: 0-1 star in selection or 0 stars in comparability or 0-1 stars in outcomes. Not available, No stars.

*Did not include abstracts (Kokoy-Mondragon).
**SUPPLEMENTARY TABLE 5. Pooled diagnostic test characteristics for tissue acquisition in ERCP and EUS**

<table>
<thead>
<tr>
<th>Test characteristic</th>
<th>ERCP</th>
<th>EUS-guided FNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensitivity</td>
<td>.7 (0.66-0.73)</td>
<td>.74 (0.71-0.77)</td>
</tr>
<tr>
<td>Specificity</td>
<td>.95 (0.91-0.97)</td>
<td>.88 (0.83-0.92)</td>
</tr>
<tr>
<td>Positive likelihood ratio</td>
<td>9.33 (5.88-14.78)</td>
<td>5.41 (3.07-9.51)</td>
</tr>
<tr>
<td>Negative likelihood ratio</td>
<td>.34 (0.24-0.47)</td>
<td>.28 (0.19-0.41)</td>
</tr>
<tr>
<td>Diagnostic odds ratio</td>
<td>58.29 (30.91-109.9)</td>
<td>22.26 (10.49-47.25)</td>
</tr>
<tr>
<td>Summary receiver-operating characteristic curve</td>
<td>.9547</td>
<td>.9128</td>
</tr>
</tbody>
</table>

**SUPPLEMENTARY TABLE 6. Risk of bias assessment for studies included in population, intervention, comparator, outcomes question 3**

<table>
<thead>
<tr>
<th>Study</th>
<th>Selection (maximum, 4 stars)</th>
<th>Comparability (maximum, 2 stars)</th>
<th>Outcomes (maximum, 3 stars)</th>
<th>Total score (maximum, 9 stars)</th>
<th>Quality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rösch 2004</td>
<td>***</td>
<td>**</td>
<td>**</td>
<td>7</td>
<td>Good</td>
</tr>
<tr>
<td>Oppong 2010</td>
<td>***</td>
<td>**</td>
<td>***</td>
<td>8</td>
<td>Good</td>
</tr>
<tr>
<td>Hijioka 2012</td>
<td>**</td>
<td>—</td>
<td>*</td>
<td>3</td>
<td>Poor</td>
</tr>
<tr>
<td>Khan 2013</td>
<td>***</td>
<td>**</td>
<td>***</td>
<td>8</td>
<td>Good</td>
</tr>
<tr>
<td>Weilert 2014</td>
<td>***</td>
<td>**</td>
<td>***</td>
<td>8</td>
<td>Good</td>
</tr>
<tr>
<td>Heinzow 2014</td>
<td>***</td>
<td>**</td>
<td>***</td>
<td>8</td>
<td>Good</td>
</tr>
<tr>
<td>Moura 2018</td>
<td>***</td>
<td>**</td>
<td>***</td>
<td>8</td>
<td>Good</td>
</tr>
<tr>
<td>Jo 2019</td>
<td>****</td>
<td>**</td>
<td>***</td>
<td>9</td>
<td>Good</td>
</tr>
<tr>
<td>Lee 2019</td>
<td>**</td>
<td>—</td>
<td>***</td>
<td>5</td>
<td>Poor</td>
</tr>
<tr>
<td>Onoyama 2019</td>
<td>**</td>
<td>**</td>
<td>**</td>
<td>6</td>
<td>Fair</td>
</tr>
<tr>
<td>Yeo 2019</td>
<td>***</td>
<td>**</td>
<td>***</td>
<td>8</td>
<td>Good</td>
</tr>
<tr>
<td>Han 2021</td>
<td>**</td>
<td>—</td>
<td>***</td>
<td>5</td>
<td>Poor</td>
</tr>
<tr>
<td>Yang 2021</td>
<td>***</td>
<td>—</td>
<td>***</td>
<td>6</td>
<td>Poor</td>
</tr>
</tbody>
</table>

Quality assessments thresholds are as follows: Good: 3-4 stars in selection and 1-2 stars in comparability and 2-3 stars in outcomes; Fair: 2 stars in selection and 1-2 stars in comparability and 2-3 stars in outcomes; Poor: 0-1 star in selection or 0 stars in comparability or 0-1 stars in outcomes.

*Not available, No stars.*