Diagnostic accuracy of cross-sectional and endoscopic imaging in ampullary tumours: systematic review

Anouk J. de Wilde1,*, Evelien J. M. de Jong2, Kurinchi S. Gurusamy3, Mohammad Abu Hilal4, Marc G. Besselink5,6, Maxime J. L. Dewulf7, Sandra M. E. Geurts2, Ulf P. Neumann1, Steven W. M. Olde Damink1, Jan-Werner Poley1, Vivianne C. G. Tjan-Heijnen4, Judith de Vos-Geelen7, Georg Wiltberger6, Marielle M. E. Coolsen1 and Stefan A. W. Bouwense1

1Department of Surgery, Maastricht University Medical Centre+, Maastricht, the Netherlands
2Department of Internal Medicine, Division of Medical Oncology, Maastricht University Medical Centre+, GROW, Maastricht University, Maastricht, the Netherlands
3Division of Surgery and Interventional Science, Royal Free Campus, University College London, London, UK
4Department of Surgery, Fondazione Poliambulanza, Brescia, Italy
5Amsterdam UMC, location University of Amsterdam, Department of Surgery, Amsterdam, the Netherlands
6Cancer Centre Amsterdam, Amsterdam, the Netherlands
7Department of Gastroenterology and Hepatology, Maastricht University Medical Centre+, Maastricht, The Netherlands
8Department of General, Visceral, and Transplantation Surgery, University Hospital of RWTH Aachen, Aachen, Germany

Correspondence to: Anouk J. de Wilde, Department of Surgery, Maastricht University Medical Centre+, P. Debyelaan 25, 6229 HX Maastricht, the Netherlands (e-mail: anouk.de.wilde@mumc.nl)

Abstract

Background: Differentiation between adenomas and carcinomas of the ampulla of Vater is crucial for therapy and prognosis. This was a systematic review of the literature on the accuracy of diagnostic modalities used to differentiate between benign and malignant ampullary tumours.

Methods: A literature search was conducted in PubMed, Embase, CINAHL, and the Cochrane Library. Studies were included if they reported diagnostic test accuracy information among benign and malignant ampullary tumours, and used pathological diagnosis as the reference standard. Risk of bias was assessed using Quality Assessment on Diagnostic Accuracy Studies (QUADAS) 2 and QUADAS-C.

Results: Ten studies comprising 397 patients were included. Frequently studied modalities were (CT; 2 studies), endoscopic ultrasonography (EUS; 3 studies), intraductal ultrasonography (IDUS; 2 studies), and endoscopic forceps biopsy (3 studies). For CT, the reported sensitivity for detecting ampullary carcinoma was 44 and 95%, and the specificity 58 and 60%. For EUS, the sensitivity ranged from 63 to 89% and the specificity between 50 and 100%. A sensitivity of 88 and 100% was reported for IDUS, with a specificity of 75 and 93%. For forceps biopsy, the sensitivity ranged from 20 to 91%, and the specificity from 75 to 86%. The overall risk of bias was scored as moderate to poor. Data were insufficient for meta-analysis.

Conclusion: To differentiate benign from malignant ampullary tumours, EUS and IDUS seem to be the best diagnostic modalities. Sufficient high-quality evidence, however, is lacking.

Introduction

Benign and malignant tumours of the ampulla of Vater (hereafter referred to as ampullary tumours) are relatively rare. For example, in the Netherlands, there were 177 patients with ampullary tumours in 2021 (0.68 per 100 000 in 2010–2016)1,2. Benign tumours have a 26–65% lifetime risk of becoming malignant3,4. To differentiate between a benign and malignant tumour, and to select appropriate treatment, clinicians rely on imaging, visual inspection of the tumour during endoscopy, and histological assessment. A clear-cut diagnostic approach to ampullary tumours, however, is lacking.

To assess local characteristics of the tumour (size, location, and depth of infiltration) and its relationship to surrounding tissues (involvement of lymph nodes and vascular structures), multiple diagnostic modalities with different advantages and disadvantages are available, such as abdominal or endoscopic (EUS) ultrasonography, endoscopic retrograde cholangiopancreatography (ERCP), MRI and magnetic resonance cholangiopancreatography, CT, PET–CT, and nuclear scintigraphy5,6. In line with the current European Society for Medical Oncology (ESMO) guideline for the assessment of pancreatic and bile duct tumours2, EUS and CT are most frequently used of these modalities. Pathological assessment might help to further differentiate between benign and malignant tumours, but sampling errors frequently occur8,9. Proper differentiation between benign and malignant ampullary tumours is particularly important in deciding which treatment is needed. For benign ampullary tumours, follow-up (with repeating imaging) or local (endoscopic or surgical) resection is sufficient, whereas oncological resection is preferred if possible for (suspected) malignant tumours5,6,10–13.

There is currently no reference standard for the diagnostic approach to ampullary tumours and no previous systematic review on this topic is available14,15. The aim of this review was to assess the accuracy of the diagnostic approach to ampullary
tumours and, more specifically, the ability to differentiate between benign and malignant tumours.

Methods

This systematic review was performed according to the Cochrane Handbook and the PRISMA guidelines, and was registered in PROSPERO database (CRD42021269453).16–19.

Search strategy

A systematic search was conducted in the PubMed, Embase, Cumulative Index to Nursing & Allied Health Literature (CINAHL), and the Cochrane Library databases to identify relevant studies assessing the accuracy of the diagnostic procedures for ampullary tumours. These studies included RCTs and (comparative) observational studies. No specific minimum volume was used in this search. The search was performed on 4 February 2022 and included the following search terms: ‘Ampulla of Vater’, ‘Neoplasms’, ‘Common Bile Duct Neoplasm’, ‘Magnetic Resonance Imaging’, ‘Magnetic resonance cholangiopancreatography’, ‘Ultrasonography’, ‘Endoscopic Ultrasound’, ‘Endoscopy, Digestive System’, ‘Endoscopic retrograde cholangiopancreatography’, ‘Tomography, X-Ray Computed’, ‘Duodenoscopy’, ‘PET/CT’, ‘Nuclear scintigraphy’, ‘Cytology’, and ‘Biopsy’. The full search is described in Appendix S1. Synonyms of these terms were also used in the search. No restrictions on language or publication date. Duplicate references were removed and all search results were uploaded into Rayyan, a web app for filtering eligible studies for a systematic review.20 If no abstract and/or full-text was available, the authors of the article were contacted by e-mail to obtain them.

Study selection

All articles were screened by two reviewers independently with respect to the prespecified inclusion and exclusion criteria based on title and abstract. Studies were included if they met the following inclusion criteria: patients had a pathologically confirmed ampullary tumour; the study assessed diagnostic accuracy of a diagnostic modality using histology as the reference standard; and, if non-ampullary tumours were included in the study, diagnostic test accuracy information was available for people with ampullary tumours. Exclusion criteria were study design such as reviews, letters, book chapters, and case reports; and study that included only malignant or only benign tumours. After abstract screening, the two reviewers independently read the full text of potentially useful articles to enable final selection.

Data extraction and data collection

Two reviewers independently screened all articles, extracted the data, and assessed the risk of bias. Disagreements were resolved by discussion. When no consensus was reached, an arbitrator resolved disagreements. Relevant data were extracted using a data extraction form. Data collected included first author, year of publication, study design (prospective or retrospective cohort study,

cross-sectional study, or RCT), total number of patients, patient characteristics (age, sex), number of patients diagnosed with an ampullary tumour (malignant versus benign), index diagnostic test, reference test, and diagnostic test accuracy information (true positive, false positive, false negative, and true negative).

Risk of bias
The Quality Assessment on Diagnostic Accuracy Studies (QUADAS) 2 tool21 and QUADAS—Comparative (QUADAS-C) tool22 were used to assess the risk of bias. The QUADAS-C is an extension of QUADAS-2 for comparative studies, in which two or more index tests were performed in the same study population. The risk of bias was assessed in four key domains including patient selection, index test(s), reference standard, and flow and timing. Concerns regarding applicability (patient selection, index test(s), and reference standard) were determined. The degree of bias and applicability were expressed as high, low, or unclear, in accordance with the guidance documents.

Statistical analysis
The statistical analysis was performed using Review Manager 5.3 (RevMan® [Computer program] Version 5.3. Copenhagen: The Nordic Cochrane Centre, The Cochrane Collaboration, 2014), for generating forest plots. The individual study estimates of sensitivity and specificity were shown as forest plots for the different index tests to examine the variation between studies. Meta-analysis was attempted using SAS® software (SAS Institute Inc., Cary, NC, USA) for calculating the summary sensitivity and specificity. Because of the sparse data, simpler hierarchical models were used for meta-analysis23. Visualization of forest plots, and model fit determined by the (-2) log likelihood values, were used to decide on the best model for undertaking meta-analysis. The forest plots of sensitivity and specificity were also inspected visually for potential sources of heterogeneity. Planned subgroup analyses or a meta-regression approach to investigate heterogeneity were not performed because of the sparse data.

Results
Selected studies and characteristics
The search yielded 2940 studies of which 1024–33 comprising 397 patients, were included for further analyses (Fig. 1). Preliminary results from three conference abstracts34–36, which were not published as peer-reviewed articles, were included in

### Table 1 Summary of characteristics of included studies

<table>
<thead>
<tr>
<th>Reference</th>
<th>Sample size</th>
<th>No. of patients with malignant or benign ampullary lesion</th>
<th>Mean age (years)</th>
<th>Index test</th>
<th>Reference standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bardales et al.24</td>
<td>22</td>
<td>Benign 15, Malignant 7</td>
<td>70</td>
<td>Brush cytology</td>
<td>Histology: 22 biopsy</td>
</tr>
<tr>
<td>Heinzow et al.25</td>
<td>72</td>
<td>Benign 32, Malignant 40</td>
<td>72</td>
<td>ETP IDUS</td>
<td>Histology of surgical resection and long-term follow-up of negative biopsies*</td>
</tr>
<tr>
<td>Ito et al.26†</td>
<td>40</td>
<td>Benign 8, Malignant 32</td>
<td>65</td>
<td>Endoscopic forceps biopsy</td>
<td>Histology: 30 surgical resections</td>
</tr>
<tr>
<td>Manta et al.27</td>
<td>24</td>
<td>Benign 5, Malignant 32</td>
<td>60</td>
<td>EUS</td>
<td>Histology: 22 surgical resections</td>
</tr>
<tr>
<td>Menzel et al.28†</td>
<td>27</td>
<td>Benign 12, Malignant 15</td>
<td>62</td>
<td>IDUS EUS</td>
<td>Histology: 27 surgical resections</td>
</tr>
<tr>
<td>Pongpornsup et al.29</td>
<td>55</td>
<td>Benign 12, Malignant 43</td>
<td>65</td>
<td>Endoscopic forceps biopsy CT</td>
<td>Histology: unclear how obtained</td>
</tr>
<tr>
<td>Rodríguez et al.30</td>
<td>31</td>
<td>Benign 14, Malignant 17</td>
<td>n.r.</td>
<td>Biopsy during ERCP</td>
<td>Histology: 31 surgical resections</td>
</tr>
<tr>
<td>Sauvanet et al.31</td>
<td>26</td>
<td>Benign 4, Malignant 22</td>
<td>57</td>
<td>SVD EUS</td>
<td>Histology: 24 surgical resections</td>
</tr>
<tr>
<td>Sperti et al.32</td>
<td>14</td>
<td>Benign 5, Malignant 9</td>
<td>65</td>
<td>Endoscopic forceps biopsy CT, [18F]FDG PET–CT</td>
<td>Histology: 14 surgical resections</td>
</tr>
<tr>
<td>Wen et al.33</td>
<td>86</td>
<td>Benign 28, Malignant 58</td>
<td>62</td>
<td>[18F]FDG PET–CT CT + MRI</td>
<td>Histology: 48 surgical resections 10 biopsies 26 long-term follow-up</td>
</tr>
<tr>
<td>Abstracts only</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chen et al.34</td>
<td>21</td>
<td>n.r.</td>
<td>n.r.</td>
<td>EUS ERCP CT</td>
<td>Surgical histology</td>
</tr>
<tr>
<td>Peng et al.35</td>
<td>102</td>
<td>Benign 76, Malignant 18</td>
<td>60</td>
<td>Ultrasonography EUS EUS + biopsy</td>
<td>Surgical or endoscopic histology</td>
</tr>
<tr>
<td>Sharaiha et al.36</td>
<td>58</td>
<td>Non-adenomatous 8, Malignant 18</td>
<td>63</td>
<td>EUS EUS + biopsy</td>
<td>Surgical or endoscopic histology</td>
</tr>
</tbody>
</table>

*The proportion of patients for each reference standard is unclear. †Prospective study. ETP, endoscopic transpapillary biopsy; IDUS, intraductal ultrasonography; EUS, endoscopic ultrasonography; n.r., not reported; ERCP, endoscopic retrograde choangio-pancreatography; SVD, side-viewing duodenoscopy; [18F]FDG, [18F]fluorodeoxyglucose.
PET–CT studies were conducted retrospectively. The studies by Ito et al. were more recent, with ERCP being performed between 1997 and 2020. The number of patients included ranged from 14 to 118. In total 260 patients (65.5%) had a malignant tumour. The most frequently studied index tests were EUS 27,28,31 and endoscopic forceps biopsy 26,28,31, which were both assessed in three studies. IDUS 25,28, CT 29,32, and PET–CT 32,33 were each examined in two studies as index tests. Brush cytology 24, endoscopic transpapillary biopsy 25, biopsy obtained by ERCP 30, side-viewing duodenoscopy 31, and a combination of CT with MRI 33 were each investigated in one study. The reference tests in the included studies consisted of pathological assessment of the resection specimen (obtained by surgical, local and/or endoscopic resection; 9 studies) or endoscopic biopsy (1 study). In two studies, some patients had long-term follow-up in the event of a negative biopsy as reference test. The studies did not mention specific symptoms per patient on which the decision was made to perform diagnostic modalities.

Quality assessment

In general, the studies had a moderate risk of bias according to the QUADAS-2 (Table S1). The studies by Manta et al. 30, Rodriguez et al. 30, and Sauvanet et al. 31 were rated as low for patient selection as these studies had inappropriate exclusions and bias was introduced owing to the selection procedure. The reference test in Heinzow et al. 25 was assessed as poor because histopathological confirmation of the final diagnosis was not available for all patients. All studies, except that of Rodriguez et al. 30, were at high risk of bias in the domain flow and timing. This related to different methods of pathology sampling used as reference tests including resection specimen or biopsy, or follow-up of negative biopsies within one cohort. Regarding applicability concerns, Bardales et al. 24 and Sauvanet et al. 31 were scored as poor in terms of patient selection. For all other domains, all studies scored well. Five studies 25,28,31,33 assessed more than one index test, for which the QUADAS-C tool was used. The risk of bias investigated using QUADAS-C could be interpreted as moderate to poor.

Diagnostic accuracy

Outcomes reported in the studies are summarized in Table 2. The sensitivity and specificity were calculated for each index test. A meta-analysis for each index test with more than two studies was attempted but, owing to the clinical and methodological heterogeneity along with poor overlap of confidence intervals, convergence was obtained only for fixed-effect meta-analysis. This was clearly inappropriate, because of the poor overlap of confidence intervals and so meta-analysis was not undertaken. Only a narrative summary is provided below.

CT

Two studies reported the sensitivity and specificity of CT. The sensitivity was 44% in Serti et al. 32 and 95% in Pongpornsup et al. 29. The reported specificity was 58% in Pongpornsup et al. 60% in Serti et al. Figure 2a shows the forest plot with corresponding confidence intervals.

PET–CT

Serti et al. 32 and Wen et al. 33 reported a sensitivity of 78% and 93% for PET–CT with a corresponding specificity of 20 and 79% (Fig. 2b).
sensitivity of 88% and a specificity of 93%. Menzel et al. reported a sensitivity of 100% and a specificity of 75% (Fig. 2d).

**Forceps biopsy**

Results for forceps biopsy were reported in three studies. The biopsies were compared with the pathology of resection specimens. The sensitivity ranged between 20 and 91%, and the specificity from 75 to 83% (Fig. 2e).

**Additional index tests**

Five different index tests were each reported only once, in five different studies. Bardales et al. reported a sensitivity and specificity of 100% for brush cytology. Endoscopic transpapillary biopsy also had a sensitivity and specificity of 100% according to Heinzow et al.

The sensitivity and specificity of biopsy during ERCP were 67 and 70% respectively. For side-viewing duodenoscopy, Sauvanet et al. reported a sensitivity of 63% and a specificity of 100%, whereas Wen et al. reported a sensitivity of 90% and specificity of 36% for CT + MRI.

The diagnostic test accuracy reported in the conference abstracts, which could not be analysed owing to missing diagnostic test accuracy information, is listed in Table S2.

**Discussion**

This systematic review of the diagnostic approach to assessment of benign and malignant ampullary tumours showed wide variation in diagnostic modalities currently being used in daily clinical practice. EUS and IDUS seem to have the best sensitivity and specificity. A meta-analysis could not be performed because of the limited amount of data, clinical and methodological heterogeneity, and poor overlap of confidence intervals between...
the studies. Therefore, no firm conclusions can be drawn on which diagnostic modality is best for assessing the nature of ampullary tumours.

Specific guidelines on how to diagnose and stage ampullary tumours are lacking and current advice is predominantly based on guidelines for patients with pancreatic cancer. Despite many similarities with pancreatic cancer, the location and nature of ampullary tumours necessitates specific data. Current pancreatic cancer guidelines are not consistent. In the ESMO guidelines, CT is recommended in all patients, and EUS with fine-needle aspiration and biopsy in case of doubt. The European Society of Gastrointestinal Endoscopy guideline only recommends endoscopic assessment with biopsy.

In daily clinical practice, the proper treatment for ampullary tumours is preferably selected based on pathological, local, and regional assessment of the tumour. CT provides information on the tumour and involvement of nearby structures, lymph node(s), and distant metastases. The reported sensitivity of CT for assessing the nature of ampullary tumours in the two included studies is highly variable. Sample sizes were small, and different methods were used for pathology sampling and image acquisition. EUS and IDUS are valuable for describing local infiltration of the tumour and provide an opportunity for biopsy. Lymph node involvement or metastases cannot be assessed. The sensitivity and specificity reported in the included studies for EUS and IDUS are higher than those for CT, and several studies have shown the advantage in tumour classification. Specific data on the diagnostic accuracy of MRI or nuclear imaging are scarce and should be the subject of future research. The value of periodic imaging follow-up as a diagnostic modality was not investigated in the studies included in the present review. Reliable follow-up was especially relevant for patients who underwent local resection or who were not good candidates for endoscopic or surgical resection.

This review had several limitations. Most studies included were retrospective, had a moderate-to-high risk of bias and were published more than 10 years ago. Diagnostic modalities evolve rapidly and newer ones might nowadays be better at discriminating ampullary tumours. No RCTs were available on this subject. Limited data and heterogeneity precluded meta-analysis.

The limited number of included studies and their quality highlights the need for continued research on this topic. Data from registries and prospective cohorts are needed to design clinical trials to further assess the best diagnostic approach. Currently, no studies have been registered at ClinicalTrials.gov or the International Clinical Trials Registry Platform. Recently, an international registry was initiated on the treatment of ampullary tumours by the International Study Group on Ampullary Cancers, which will provide more information regarding the use and accuracy of diagnostic modalities in daily clinical practice. The main challenge for future studies is to have an adequate sample size. (Inter)national collaboration should be encouraged. Histological confirmation of the diagnosis in biopsies and/or resection specimens of malignant and benign tumours is needed to assess the diagnostic accuracy of these procedures. This will allow the development of a clear algorithm, including clinical presentation, and single and combined diagnostic modalities, for choosing the best diagnostic and treatment strategy for ampullary tumours in every patient.

Funding
The authors have no funding to declare.

Acknowledgements
A.J.d.W. and E.J.M.d.J. are joint first authors.

Author contributions
Anouk De Wilde (Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Writing—original draft, Writing—review & editing), Evelien de Jong (Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Writing—original draft, Writing—review & editing), Kurinchi Gurusamy (Formal analysis, Methodology, Resources, Software, Supervision, Writing—review & editing), Sandra GE Geurts (Conceptualization, Methodology, Supervision, Writing—review & editing), UIl Neumann (Writing—review & editing), Steven W.M. Olde Damink (Writing—review & editing), Jan-Werner Poley (Writing—review & editing), Vivianne Tjan-Heijnen (Supervision, Writing—review & editing), Judith de Vos-Geelen (Conceptualization, Investigation, Methodology, Supervision, Writing—original draft, Writing—review & editing), Georg Wilibberger (Writing—review & editing), Marielle Coolsen (Conceptualization, Investigation, Methodology, Resources, Supervision, Writing—original draft, Writing—review & editing), and Stefan Bouwense (Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Supervision, Writing—original draft, Writing—review & editing).

Disclosure
J.d.V.-G. has served as a consultant for Amgen, AstraZeneca, MSD, Pierre Fabre, and Servier, and has received institutional research funding from Servier, all outside the submitted work.

S.M.E.G. reports grants from Roche, grants from Pfizer, grants from Novartis, grants from Lilly, grants from Daiichi Sankyo, grants from Gilead, and personal fees from AstraZeneca, all outside the submitted work. V.C.G.T.-H. reports grants and personal fees from Roche, grants and personal fees from Novartis, grants and personal fees from Pfizer, grants and personal fees from Lilly, personal fees from Accord Healthcare, grants from AstraZeneca, grants from Eisai, grants from Daiichi Sankyo, and grants from Gilead, all outside the submitted work. The salary and promotions of K.S.G. are linked to publications. The authors declare no other conflict of interest.

Supplementary material
Supplementary material is available at BJS online.

Data availability
The data that support the findings in this study are available from the corresponding author upon request.
References

34. Chen CH, Tseng LJ, Yang CY, Yeh YH, Mo LR. The accuracy of endoscopic ultrasound, endoscopic retrograde cholangiopancreatography, computed tomography, and transabdominal ultrasound in the detection and staging of primary ampullary tumors. Hepatogastroenterology 2001; 48: 1750–1753
35. Peng C, Lv Y. Su1387 evaluation of endoscopic ultrasonography for diagnostic and therapeutic decision making in ampullary lesions. Gastrointest Endosc 2018;87:AB334
38. Cannon ME, Carpenter SL, Elta GH, Nostrant TT, Kochman ML, Ginsberg GG et al. EUS compared with CT, magnetic resonance imaging, and angiography and the influence of biliary stenting on staging accuracy of ampullary neoplasms. Gastrointest Endosc 1999;50:27–33
A gaze in the crystal ball: Where is the role of virtual reality and artificial intelligence in colorectal surgery
Müller Beat, Basel, CH

MALIGNANT COLORECTAL DISEASE

Cytoreductive Surgery and Intrapertitoneal Chemotherapy – facts and hopes
Michel Adamina, Winterthur, CH

Metastatic Colorectal Cancer – surgical approaches and limits
Jürgen Weitz, Dresden, DE

Extended lymph node dissection for rectal cancer, is it still under debate?
Miranda Kusters, Amsterdam, NL

Organ preservation functional outcome in rectal cancer treatment – in line with patient’s needs?
(Robot – laparoscopic – open surgery?)
Hans de Wilt, Nijmegen, NL

ROBOTICS

Advances in Robotic Surgery and what we learnt so far
Parvaiz Amjad, Portsmouth, UK

Challenging the market: Robotic (assistant) Devices and how to choose wisely
(Da Vinci – Hugo Ras – Distalmotion ua)
Khan Jim, London, UK

TAMIS - Robotic Transanal Surgery, does it make it easier?
Knol Joep, Genk, BE

Live Surgery – Contonal Hospital of St.Gallen
Walter Brunner, St.Gallen, CH; Salvador Conde Morals, Sevilla, ES; Friedrich Herbst, Vienna, AUT; Amjad Parvaiz, Portsmouth, UK

Video Session

Lars Pahlmann Lecture
Markus Büchler, Lisboa, PRT

Honorary Lecture
Bill Heald, Lisboa, PRT