A systematic review of the staging performance of endoscopic ultrasound in gastro-oesophageal carcinoma

S Kelly, K M Harris, E Berry, J Hutton, P Roderick, J Cullingworth, L Gathercole, M A Smith

Abstract

Background—Endoscopic ultrasound (EUS) may be used for preoperative staging of gastro-oesophageal carcinoma but performance values given in the literature differ.

Aim—To identify and synthesise findings from all articles on the performance of EUS in tumour, node, metastasis (TNM) staging of gastro-oesophageal carcinoma.


Methods—Data on the staging performance of EUS were retrieved and evaluated. Summary receiver operator characteristic methodology was used for synthesis, and a summary estimate of performance, Q*, obtained. Multiple regression analysis was used to assess study validity and investigate reasons for differences in performance.

Results—Twenty seven primary articles were assessed in detail. Thirteen supplied results for staging oesophageal cancer, 13 for gastric cancer, and four for cancers at the gastro-oesophageal junction. For gastric T staging, Q*=0.89 (95% confidence interval (CI) 0.88–0.92). For gastro-oesophageal T staging, including cancers at the gastro-oesophageal junction, Q*=0.91 (95% CI 0.89–0.93). Inclusion of cases with non-traversable stenosis was found to slightly reduce staging performance. For N staging, Q*=0.79 (95% CI 0.75–0.83). In articles that compared EUS directly with incremental computed tomography, EUS performed better. None of the variables assessed in the regression analysis was significant using a Bonferroni correction. Three variables (anatomical location, traversability, and blinding) showed strong relationships for future research and validation.

Conclusions—EUS is highly effective for discrimination of stages T1 and T2 from stages T3 and T4 for primary gastro-oesophageal carcinomas. The failure rate of EUS from non-traversability of a stenotic cancer may be a limitation in some patient groups.

Keywords: endoscopic ultrasound; gastro-oesophageal cancer; TNM staging; systematic literature review; meta-analysis

Accurate staging of gastro-oesophageal cancer is essential for well informed decisions on stage dependent patient management. This is becoming increasingly important with improvements in non-surgical treatment regiments. While patients with early localised disease clearly benefit from complete surgical resection, there is increasing evidence that multimodal treatment (chemoradiotherapy) is superior to surgery alone for patients with resectable adenocarcinoma of the oesophagus.1 Accurate local cancer staging provides the information to allow such important decisions to be made so as not to deny patients potentially curative surgical resection, with or without neoadjuvant therapy, as appropriate. The development of other non-surgical techniques at both ends of the disease spectrum has also reinforced the need for accurate cancer staging. Endoscopic ultrasound (EUS) in conjunction with endoscopic mucosal resection2 is only appropriate for superficial non-invasive cancers. Metal mesh oesophageal stents can be employed to afford adequate palliation in patients with advanced unresectable oesophageal cancer.3

The decision to undertake surgical resection is complex and based on many factors in addition to accurate cancer staging, including in particular the age of the patient and the presence of significant coexisting medical problems. Accurate staging is essential in all patients to allow an informed decision to be made regarding the most appropriate method of non-surgical treatment or palliation even in those considered unfit for surgical intervention. If comparisons of the outcomes of available and future treatment protocols are to be made, comparable input data should be available from all patients. This is particularly important if the patient does not undergo primary surgical resection, with the consequent loss of pathological confirmation, as then the stage of the cancer can only be assessed from the best imaging modality or modalities.

EUS has been in use since the early 1980s but it has been slow to gain acceptance in certain countries, including the UK. There are two basic types of echoendoscope commercially available, with either radial or linear array transducer technology. In addition, miniprobes are small higher frequency probes that can be

Abbreviations used in this paper: CT, computed tomography; EUS, endoscopic ultrasound; FPR, false positive rate; SROC, summary receiver operating characteristic; TNM, tumour, node, metastasis; TPR, true positive rate.
passed down the biopsy channel of a conventional endoscope. Their high frequency ensures excellent resolution but also limited depth of penetration.

The ability of EUS to identify the component layers of the bowel wall provides the basis for cancer staging within the widely accepted tumour, node, metastasis (TNM) classification. The International Union against Cancer (UICC) TNM classification defines the extent of malignant cancer and allows easy correlation of results from more than one centre. While broadly similar there are important differences between TNM staging for oesophageal carcinoma and gastric carcinoma. The definition has recently (1997) been changed but all articles included in this review used the 1987 definition.

A systematic literature review of EUS in gastro-oesophageal carcinoma was undertaken as part of the NHS Health Technology Assessment Programme. This review addresses a subset of questions associated with staging performance. We examine the evidence on the staging performance of EUS for differentiating cancer stages T1 and T2 from stages T3 and T4, and lymph node stage N0 from N1 and N2. Furthermore, we ask if performance differs with anatomical location of the primary carcinoma and how the overall performance is affected by the occurrence of non-traversable stenoses. Finally, the staging performance of EUS is compared with that of available computed tomography (CT) techniques.

Methods

DATA SOURCES

Explicit search strategies were used to retrieve information from Medline and BIDS from 1981 to 1996 inclusive. The Cochrane Library, Embase, the British Library’s Inside and SIGLE (System for the Identification of Grey Literature) databases, and FirstSearch from the Online Computer Library Centre were searched using the keywords endoscopic ultrasound (or ultrasonography), EUS, and endosonography. Bibliographies of articles retrieved were hand searched. Authors of conference abstracts, leading manufacturers, major UK centres, and electronic mail discussion groups were contacted with a request for information on unpublished studies. An update of the systematic review was conducted in 1999 by searching Medline. This identified any new literature since completion of the review in 1996 and allowed an estimation of the progress in the field and any impact on the conclusions.

STUDY SELECTION

Review articles, abstracts, editorials or letters, case reports, and non-English language articles were excluded. A second set of exclusion criteria was applied to include only articles discussing the use of EUS for staging in a study of gastro-oesophageal carcinoma on human subjects. The final set of criteria concerned more general quality considerations. Articles were excluded if they did not provide a comparison with the gold standard reference test of pathology, were duplicate studies on the same patient group, involved 10 or fewer patients, or did not supply sufficient information to construct a 2×2 contingency table of results. Two checklists were used to extract important information for evaluation of study design and validity. The first checklist concentrated on potential threats to study validity in terms of the likely risk of bias. Twenty biases potentially arising from the design of medical imaging studies for evaluation of diagnostic performance were considered. The second checklist was used to note factors that might vary between studies. These factors included technical parameters, patient characteristics, and study design. The number of articles was expected to be small, so rather than use this validity information to exclude studies, information from the checklists was incorporated into the data synthesis.

DATA EXTRACTION

Comparison with the gold standard of pathology was required. All patients had surgical exploration and the majority underwent resection from which a precise pathological stage could be determined. In a small number of cases a resection was deemed inappropriate due to the advanced nature of the tumour. Data were extracted where results were presented for staging carcinoma, lymph nodes, or metastases according to the 1987 TNM system. Corresponding results were taken from articles that used CT on the same subjects as were investigated with EUS. To facilitate calculation of the true positive rate (TPR) and false positive rate (FPR), 2×2 contingency tables were completed. For cancer staging the contingency tables were completed for differentiation of stages T1 and T2 from stages T3 and T4. This threshold was judged to be of the most clinical significance to aid the decision between surgical and non-surgical management of patients. Similarly, for oesophageal lymph nodes the tables were completed for differentiation of stage N0 from N1 and N2.

QUANTITATIVE DATA SYNTHESIS

A range of TPR/FPR pairs was expected from the articles included, arising partly from random variations but also from small differences in the way in which the test was performed. For example, slightly differing thresholds may have been applied because of inter-operator variation or the use of different equipment. To allow a summary estimate of test performance the summary receiver operating characteristic (SROC) methodology first described by Moses and colleagues and Irwig and colleagues was applied. This technique involves logistic transformation of the TPR/FPR pairs to allow linear fitting to the data. Results are then back transformed and plotted as a SROC curve.

For comparison purposes a further statistic, Q*, the value of TPR where TPR=(1−FPR),
and its 95% confidence interval (CI) were calculated. This value was obtained from the intercept of the SROC curve and a line plotting sensitivity equals specificity. Q* is an appropriate summary statistic for EUS and represents the optimum performance for the following reasons. Due to the dichotomy chosen for cancer staging, T1 or T2 is analogous to a positive diagnosis in a conventional 2×2 table, and therefore T3 or T4 is analogous to a negative diagnosis. This implies that sensitivity is a measure of the ability of EUS to correctly stage T1/T2 and not overstage cancers as T3/T4, and conversely specificity is a measure of the ability of EUS to correctly stage T3/T4 and not understage cancers as T1/T2. Neither under-staging nor overstaging can be assumed to have more or less impact than the other: understaging cancer will result in surgical operations which are unnecessary and overstaging will result in palliative or non-surgical treatments when resection may have been possible. The most appropriate threshold is one which minimises both understaging and overstaging—a Q* value which balances sensitivity and specificity.

SROC curves were calculated for the staging of oesophageal carcinoma, gastric carcinoma, gastro-oesophageal carcinoma (including carcinoma at the cardia), and for lymph node staging of primary gastro-oesophageal carcinomas.

VALIDITY ASSESSMENT
Multiple linear regression analysis was performed to determine if any of a set of five of the factors and biases had a statistically significant effect on the fit of the linear model at the 5% level. Variables assessed were the risks of verification bias, withdrawal bias, and blinding bias, and two factors associated with spectrum bias: the anatomical location of the primary cancer and inclusion of patients with non-traversable cancers. The p value used in the analysis (p<0.01) represents the adjusted value given by the Bonferroni correction. Although potentially all factors and biases considered on our checklist could have been included in the analysis, it would have been counterproductive to include a large number because of the inverse relation, given by the Bonferroni correction, between the number and the p value required to demonstrate significance. The variables chosen included the biases representing the greatest threat to study validity. They also addressed our research questions about the effect on the results of anatomical location and traversability. If statistical significance was found in the forward stepwise approach, the linear model was used to fit separate SROC curves to each subgroup. To account for the possibility of type II error when using the conservative approach of the Bonferroni adjustment, any variables with a significance level of p<0.05 were reported and implications were discussed to encourage further research and validation. The analysis was performed for both cancer staging results and lymph node staging results.

Results
Twenty seven articles remained after the exclusion criteria had been applied. Thirteen supplied results for staging oesophageal carcinoma, and four for staging cancer at the cardia (gastro-oesophageal junction) (table 1).

<table>
<thead>
<tr>
<th>Category</th>
<th>Sensitivity range (%)</th>
<th>Specificity range (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oesophageal carcinoma staging (n=13)</td>
<td>71.4–100</td>
<td>66.7–100</td>
</tr>
<tr>
<td>Gastric carcinoma staging (n=13)</td>
<td>67.9–100</td>
<td>87.5–100</td>
</tr>
<tr>
<td>Gastro-oesophageal carcinoma staging (n=30)</td>
<td>41.7–100</td>
<td>66.7–100</td>
</tr>
<tr>
<td>Lymph node staging of gastro-oesophageal carcinomas (n=20)</td>
<td>59.5–97.2</td>
<td>40.0–100</td>
</tr>
</tbody>
</table>

The SROC curves for oesophageal carcinoma staging, gastric carcinoma staging, and gastro-oesophageal carcinoma staging are shown in fig 1, and the calculated values for Q* are also listed. In the multiple regression analysis none of the variables had a statistically significant effect. Trends were identified for the two factors related to the patient spectrum: for anatomical location p=0.02 and for inclusion of patients with non-traversable cancer p=0.04. In the latter case, the trend was a reduction in performance when such patients were included.

LYMPH NODE STAGING WITH EUS (PRIMARY GASTRO-OESOPHAGEAL CARCINOMAS)
The SROC curve for lymph node staging of primary gastro-oesophageal carcinoma is shown in fig 2. In the multiple regression analysis none of the variables had a statistically significant effect. Three articles (out of 20) that failed to correctly guard against blinding biases or to report their precautions showed a trend towards poorer performance (p=0.02).
Lymph node staging of gastro-oesophageal carcinomas (n=7) 40.0–79.3 25.0–66.7

Table 2 Ranges of sensitivity and specificity for staging, using computed tomography, reported in the included articles

<table>
<thead>
<tr>
<th>Category</th>
<th>Sensitivity range (%)</th>
<th>Specificity range (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gastro-oesophageal carcinoma staging (n=5)</td>
<td>40.0–80.0</td>
<td>14.3–97.1</td>
</tr>
<tr>
<td>Lymph node staging of gastro-oesophageal carcinomas (n=7)</td>
<td>40.0–79.3</td>
<td>25.0–66.7</td>
</tr>
</tbody>
</table>

was statistically significant. EUS performed better for staging gastric carcinoma, with a $Q^*$ value of 0.93 (0.91–0.95) compared with 0.89 (0.86–0.92) in the oesophagus. There was no statistically significant difference between lymph node staging performance for primary cancer in gastric and oesophageal regions. EUS was least effective for lymph node staging ($Q^*$=0.79 (0.75–0.83)). For comparison however, and it is emphasised that these values are not from a systematic review of the literature, alternative methods do not perform any better. In spiral CT, for example, the sensitivity for distinguishing stage N0 from N1 was found to be 24%, with specificity 100%. Methodological differences between the primary studies affected the results. Some studies counted all those individuals who underwent the examination in the total number of patients, regardless of whether or not the cancer was traversable. Others defined as the total number in the study only those for whom the cancer was traversable. The $Q^*$ for a result calculated only from traversable cancer was $Q^*$=0.92 (0.90–0.94). A more realistic value, if all patients actually undergoing the procedure were included in the 2×2 table, was $Q^*$=0.88 (0.84–0.92). The ubiquity of problems of traversability is of interest to those considering using the technique. The percentage of non-traversable cancers varies greatly from series to series and in part reflects patient selection. Published results include non-traversability rates of up to 45%. Hordijk and colleagues assessed the influence of cancer stenosis on T staging accuracy. A lower accuracy (46%) was reported with cancers that were traversable with difficulty than those that were traversable with ease or non-traversable (92% and 82%, respectively). The authors postulated that the short focal distance between the transducer and cancer when just traversable “hamper the clear visualisation of the wall layers and tumour penetration depth” as a possible explanation for the lower accuracy. Miniendoscopes capable of traversing all but the tightest of oesophageal cancers are available. Their limited depth of penetration restricts their use for TNM staging as the full extent of the cancer and adjacent lymph node groups may be beyond their field of view. This is particularly important as it has been shown that the vast majority (80% plus) of non-traversable oesophageal cancers are at least T3. For such cancers it is essential to determine invasion into adjacent structures (T4) and fully evaluate the nodal status, investigations that may be beyond the limitations of the miniendoscope. High frequency miniendoscopes allow clearer differentiation of the component layers of the bowel wall and so a more appropriate application for the miniendoscope may be in the evaluation of localised superficial carcinomas with a view to possible local endoscopic mucosal resection.

There is little published evidence of the use of miniendoscopes in gastro-oesophageal cancer. In the time since this review was conducted, evaluation of the use of miniendoscopes has been of primary interest in the literature. An update of the search of Medline, maintaining the same
inclusion criteria, found several studies addressing the question of the effectiveness of miniprobe but varied results were identified. One study reported no benefit of miniprobe in oesophageal cancer, another showed no benefit for advanced cancers due to ultrasound attenuation, a further study demonstrated benefit only in early gastric cancers, and another study found higher accuracy rates for oesophageal T staging and similar rates for N staging compared with conventional probes.

The articles comparing EUS and CT carcinoma staging accuracy included in this review all suggested the T staging performance of EUS was superior to that of incremental CT. To ensure that our comparison of CT and EUS results was valid we sought articles that performed both imaging tests on the same set of patients and compared the results to the same gold standard. However, the technological data presented regarding the CT examination were often limited or omitted and all articles included incremental rather than spiral (helical) CT. It was not possible to be certain that the CT study had been optimised for the task. Data from studies comparing the accuracy of state of the art spiral CT with that of EUS are required. It was noted that only one article discussed the complementary roles of the two modalities.

Updated literature available in Medline up to and including 1999 shows that the level of evidence for EUS versus CT has not improved. No studies were found comparing spiral CT and EUS in gastro-oesophageal cancer. A small number of studies reported limited results of incremental CT alongside EUS for oesophageal staging accuracy. These studies continue to support the hypothesis that EUS is superior to CT.

Oesophageal adenocarcinoma is increasing in incidence and adenocarcinomas of the oesophagus and stomach are now more commonly based at the cardia than in previous years. The changing natural history of gastro-oesophageal adenocarcinoma is important if the increase in incidence of carcinomas based on the gastro-oesophageal junction continues. Although only four articles were available, their results suggest that the accuracy of EUS is lower for carcinomas at the cardia.

There are several possible reasons for this, including anatomy at this site leading to a tendency to scan obliquely through the bowel wall and cancer. Such oblique scanning can give rise to artefactual misrepresentation of the true depth of penetration. No studies specifically designed to evaluate EUS in the preoperative T staging of cardia lesions have appeared in Medline since completion of the systematic review in 1996.

This systematic literature review has demonstrated the high accuracy of EUS in the staging of gastro-oesophageal carcinoma using the internationally recognised TNM classification. The performance differs for gastric and oesophageal carcinoma and appears to be lower for carcinomas at the cardia. This is a cause for concern because the proportion of carcinomas in this location is increasing. Some articles overestimate performance by including only those patients with traversable cancer. There is little published evidence comparing the performance of EUS directly with other states of the art imaging modalities, or describing the use of EUS in combination with other techniques.

Update of the literature review showed a general inertia in the direction of research in this field. Most new studies concentrated on the use of miniprobe or comparison with incremental CT. The impact of this new research on the conclusions of the review was minimal. Very little new information was found on the key issues identified within the review. There was no subsequent change in the conclusions for EUS compared with spiral CT, the value of EUS in non-traversable cases, or the likely performance of EUS staging of cardia lesions. A further update of this review is likely to be relevant to answer these key questions within the next few years.

This work was carried out with the financial support of the Secretary of State for Health under the NHS Health Technology Assessment Programme, project 94/44/03. The views and opinions expressed do not necessarily reflect those of the Secretary of State for Health. In part, this work was undertaken by the Leeds Teaching Hospitals NHS Trust who received funding from the NHS Executive; the views expressed are those of the authors and not necessarily those of the NHS Executive. The authors would like to thank the Sub-Unit for Medical Statistics (SUMS) and the Health Sciences Library, University of Leeds, for their assistance.