Esophageal cancer: CT, endoscopic US, and FDG PET for assessment of response to neoadjuvant therapy. Systematic review


CRD summary
The review compared the diagnostic accuracy of three imaging modalities at assessing response to neoadjuvant therapy in oesophageal cancer. The authors concluded that computed tomography has poor accuracy, while endoscopic ultrasound and the more widely feasible fluorine-18-fluorodeoxyglucose positron emission tomography have equivalent, good accuracy. Sparse data and a lack of direct comparisons mean that the conclusions should be viewed with caution.

Authors' objectives
To compare the diagnostic accuracy of computed tomography (CT), endoscopic ultrasound (EUS) and fluorine-18 fluorodeoxyglucose (FDG) positron emission tomography (PET) for the assessment of response to neoadjuvant therapy in oesophageal cancer.

Searching
MEDLINE (1980 to 2004), EMBASE (1980 to 2004) and the Cochrane Database of Systematic Reviews (1993 to 2004) were searched for relevant articles; the search terms were reported. A manual search with cross-reference of eligible articles was used to identify additional studies. Conference abstracts were excluded.

Study selection
Study designs of evaluations included in the review
No inclusion criteria were specified and the study designs were not reported. The studies were required to have a minimum of 10 patients.

Specific interventions included in the review
Studies evaluating CT, EUS or FDG-PET for the assessment of response to neoadjuvant therapy (before and after therapy) were eligible for inclusion. The included studies used chemotherapy with and without radiotherapy as neoadjuvant treatment, and most of the included studies assessed response after the completion of neoadjuvant treatment. Studies evaluating CT used thick section thickness (8 to 10 mm). Studies evaluating FDG uptake used a wide range of cut-off values (from 30 to 80% of the standardised uptake values).

Reference standard test against which the new test was compared
The included studies were required to use a valid reference standard; this was defined as pathologic findings.

Participants included in the review
The study participants were required to have histologically proven cancer of the oesophagus. The majority of participants had adenocarcinoma or squamous cell carcinoma, and disease stage varied between I and IV.

Outcomes assessed in the review
No inclusion criteria for the outcome measures were specified. Where data were available, the prevalence of responders was reported for individual studies while sensitivity and specificity values were presented in forest plots.

How were decisions on the relevance of primary studies made?
Two authors independently assessed articles for inclusion. The final decision was based on the full article and any disagreements were resolved by consensus.

Assessment of study quality
The methodological quality of the included studies was assessed using criteria based on those recommended by the Cochrane Methods Working Group on Systematic Reviews of Screening and Diagnostic Tests.

Criteria relating to internal validity were: use of a valid reference test; definition of the index test; assessment of index test results without knowledge of reference standard results; assessment of reference standard results without knowledge of reference index test results; avoidance of verification bias; index test results interpreted without knowledge of all clinical information; prospective study.

Criteria relating to external validity were: consecutive series of patients used; adequate reporting of spectrum of disease, demographic information, inclusion and exclusion criteria and index test methods. Items classified as unclear were given a negative score and quality scores were assessed as a percentage of the maximum, with subtotals for internal and external validity.

Two reviewers independently assessed methodological quality.

**Data extraction**
The authors did not state how the data were extracted for the review, or how many reviewers performed the data extraction. For all included studies, an attempt was made to extract data to populate a 2x2 contingency table and to calculate the sensitivity and specificity.

**Methods of synthesis**
How were the studies combined?
The studies were described in a narrative synthesis. In addition, for studies that reported sufficient data to populate 2x2 contingency tables, a summary receiver operating characteristic (ROC) curve was generated for each imaging modality. Summary ROC curves were fitted using the Moses and Littenburg method, which captures variation in diagnostic accuracy caused by differences in diagnostic threshold between studies. The Q point (maximum joint sensitivity and specificity), derived from the summary ROC curves, was used to assess differences in accuracy between the imaging modalities.

How were differences between studies investigated?
The authors discussed potential source of heterogeneity in the text, but did not report a formal assessment of heterogeneity.

**Results of the review**
A total of 24 studies were included in the review. Four studies (n=146) assessed CT, 13 studies (n=441) assessed EUS and 7 studies (n=189) assessed FDG-PET. Three studies of CT, four of EUS and four of FDG-PET reported sufficient data to enable the calculation of sensitivity and specificity (n=318 included in meta-analyses).

The included studies showed limited methodological quality. The quality scores ranged from 15 to 100%: from 46 to 92% for CT studies, from 15 to 85% for EUS studies and from 54 to 100% for FDG-PET studies.

Full quality assessment data were reported in the article.

The sensitivity of CT (3 studies, n=108) ranged from 33 to 55% and the specificity ranged from 50 to 71%. The sensitivity of EUS (4 studies, n=120) ranged from 50 to 100% and the specificity ranged from 36 to 100%. The sensitivity of FDG-PET (4 studies, n=116) ranged from 71 to 100% and the specificity ranged from 55 to 100%.

The Q points for CT, EUS and FDG-PET were 54% (95% confidence interval, CI: 31, 77), 86% (95% CI: 80, 93) and 85% (95% CI: 77, 93), respectively. Based on the Q point estimates, the overall accuracy of CT was significantly lower than that of EUS (P<0.003) and FDG-PET (P<0.006); EUS and FDG-PET had similar overall accuracy.

Evaluations were not feasible in one FDG-PET patient and seven (6% in 2 studies) EUS patients, and were suboptimal in a further 17 (14% in 1 study) EUS patients.
An update study published in 2006 (see Other Publications of Related Interest) identified two additional studies. These studies indicated a relationship between tumour response and decreased FDG uptake, but did not substantively alter the conclusions of the current review.

**Authors’ conclusions**
CT has poor accuracy for assessment of response to neoadjuvant therapy in oesophageal cancer. EUS and FDG-PET have equivalent good accuracy, but FDG-PET is more widely feasible.

**CRD commentary**
The review addressed a clearly stated research question, with appropriately defined inclusion criteria. An adequate literature search was reported, although the absence of searches for grey literature and the explicit exclusion of conference abstracts might have resulted in the omission of relevant, unpublished data. Appropriate measures were taken to avoid the introduction of error and bias during the study selection and quality assessment processes, although it was unclear whether the same measures were applied to the data extraction. Appropriate criteria were used to assess the methodological quality of the included studies, and the results of the quality assessment were reported in full. The results of the quality assessment were used to generate quality scores, a process which is not generally recommended: it cannot produce reliable indicators of overall quality and it results in a potential loss of information about individual components. The small number of available studies precluded meaningful incorporation of the quality assessment in the analysis.

The meta-analytic methods used were broadly appropriate, with the caveat that the summary ROC curves presented were based on very small numbers of data points. The authors’ conclusions follow generally from the data presented, though reporting of sensitivity and specificity values from the individual included studies would have aided interpretation. It should also be noted that the conclusion about the relative accuracy of the three imaging modalities are based upon indirect comparisons (no study assessed more than one imaging modality).

**Implications of the review for practice and research**
Practice: The authors stated that CT is inaccurate for the assessment of neoadjuvant response in oesophageal cancer and, therefore, is not recommended. EUS can be used to identify pathologic response, but is not always feasible during or shortly after chemoradiation. FDG-PET seems a promising, noninvasive tool for assessing response.

Research: The authors stated that adequately powered studies, which focus on the prediction of tumour response early in therapy and which include direct comparisons between imaging modalities, are needed. There is also a need for research to determine the optimal cut-off values for FDG-PET to discriminate between patients who respond and those who do not.

**Bibliographic details**

**PubMedID**
16118165

**DOI**
10.1148/radiol.2363041042

**Original Paper URL**
http://radiology.rsna.org/cgi/content/full/236/3/841

**Other publications of related interest**

Indexing Status
Subject indexing assigned by NLM

MeSH
Endosonography; Esophageal Neoplasms /diagnosis /therapy; Fluorodeoxyglucose F18; Humans; Neoadjuvant Therapy; Positron-Emission Tomography; ROC Curve; Radiopharmaceuticals; Sensitivity and Specificity; Tomography, Emission-Computed

AccessionNumber
12005001152

Date bibliographic record published
28/02/2007

Date abstract record published
28/02/2007

Record Status
This is a critical abstract of a systematic review that meets the criteria for inclusion on DARE. Each critical abstract contains a brief summary of the review methods, results and conclusions followed by a detailed critical assessment on the reliability of the review and the conclusions drawn.