Positron Emission Tomography in Staging of Esophageal Cancer

Westreenen, Henderik Leendert van

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2005

Citation for published version (APA):
SYSTEMATIC REVIEW OF THE STAGING PERFORMANCE OF 18F-FLUORODEOXYGLUCOSE POSITRON EMISSION TOMOGRAPHY IN ESOPHAGEAL CANCER

HL van Westreenen¹
M Westerterp²
PMM Bossuyt³
J Pruim⁴
GW Sloof⁵
JJB van Lanschot²
H Groen⁶
JThM Plukker¹

Departments of Surgery¹, Nuclear Medicine/PET center⁴,
Office for Medical Technology Assessment⁶
University Hospital Groningen, The Netherlands
Departments of Surgery², Clinical Epidemiology and Biostatistics³,
Nuclear Medicine⁵
Academic Medical Center, Amsterdam, The Netherlands

Journal of Clinical Oncology 2004;22:3805-12
CHAPTER 2

ABSTRACT

Introduction: Despite the increasing number of publications concerning $^{18}$F-fluorodeoxyglucose positron emission tomography (FDG-PET) for staging of esophageal cancer and the increasing availability of this novel diagnostic modality, its exact role in preoperative staging of these tumors is still unknown. The aim of this study was to systematically review the literature regarding the diagnostic performance of FDG-PET in preoperative staging of patients with esophageal cancer and to calculate summary estimates of its sensitivity and specificity.

Materials and Methods: The databases of PubMed, Embase and Cochrane were searched for relevant studies. Two reviewers independently assessed the methodological quality of each study. A meta-analysis of the reported sensitivity and specificity of each study was performed.

Results: Twelve studies met the inclusion criteria. The studies had several design deficiencies. Pooled sensitivity and specificity for the detection of locoregional metastases were 0.51 (95% CI, 0.34-0.69) and 0.84 (95% CI, 0.76-0.91) respectively. For distant metastases, pooled sensitivity and specificity were 0.67 (95% CI, 0.58-0.76) and 0.97 (95% CI, 0.90-1.0), respectively.

Conclusion: FDG-PET showed moderate sensitivity and specificity for the detection of locoregional metastases, and reasonable sensitivity and specificity in detection of distant lymphatic and hematogenous metastases.
INTRODUCTION

The incidence of esophageal carcinoma (EC) has been rising steadily over the 1980s and 1990s. This has been attributed most to the increasing frequency of adenocarcinoma particularly in a preexisting Barrett’s esophagus. Surgical resection is currently the best curative treatment in patients without distant metastases and locally advanced tumor growth. However, esophagectomy is associated with a substantial morbidity and mortality and may have a negative impact on quality of life over a period of several months. Therefore, conventional imaging techniques are employed to select only patients with resectable disease for esophagectomy. Despite these efforts, metastatic spread is encountered during operation in up to 60% of patients. As a result, accurate preoperative staging is essential to select those patients who will benefit from surgery and to avoid unnecessary operations in patients with distant metastases.

Currently, the most common conventional modalities for staging of EC are endoscopic ultrasonography (EUS) with or without fine needle aspiration (FNA), computed tomography (CT) of the chest, and abdomen and ultrasonography of the neck and abdomen. Occasionally, barium swallow, bronchoscopy, bone scintigraphy and diagnostic laparoscopy with or without laparoscopic ultrasonography are employed. EUS is highly effective to distinguish stages T1 and T2 from stages T3 and T4. The accuracy of EUS has increased in combination with FNA to assess nodal involvement, especially of lymph nodes in the celiac trunk region. CT plays an important role in detecting distant metastases and in assessing the extent of invasion of surrounding structures by the primary tumor. The main limitations of CT are its insensitivity to the identification of irresectability (T4) and its inability to identify metastatic disease in normal-sized lymph nodes.

Positron emission tomography (PET) is a rapidly developing noninvasive method for staging of various types of cancer. The increased glucose metabolism of malignant cells is the rationale behind the common use of FDG as a radiotracer in oncological PET studies. The use of whole body FDG-PET as a staging method in EC was first described by Yasuda et al. This case report was followed by several studies investigating the accuracy, sensitivity and specificity of FDG-PET in staging esophageal carcinoma. These studies have demonstrated that FDG-PET is a promising noninvasive method of detecting both distant nodal and hematogenous metastases, and might thus prevent futile esophagectomies.

Despite the increasing number of publications concerning FDG-PET in staging of esophageal cancer, its exact role is still unknown. The aim of this study was to systematically review the available literature regarding the diagnostic performance of FDG-PET in preoperative staging of patients with EC, which might contribute to the development of guidelines for the effective use of PET.

MATERIALS AND METHODS

The systematic review and meta-analysis were conducted according to recently presented guidelines for diagnostic reviews. After conducting a comprehensive literature search, the methodological quality of all retrieved reports was assessed in terms of the potential for bias and lack of generalizibility of the identified studies.
A computer-aided search of the databases PubMed/Medline, Embase and Cochrane was conducted in June 2003. We used the search terms ‘positron emission tomography’ and ‘esophageal cancer’ without any language restrictions. All searches were performed using text word or medical subject heading (MeSH). We looked for clinical studies evaluating the diagnostic accuracy of FDG-PET in patients with histologically proven cancer of the esophagus or gastroesophageal junction and studies using pathology or surgery as reference standard.

We augmented our computerized literature search by manually reviewing the reference lists of identified studies and relevant reviews. Unpublished data and conference proceedings were not included.

Criteria for exclusion were insufficient information to construct 2x2 contingency tables, and duplicate studies on the same patients. Two reviewers (HLvW, MW) independently selected studies for possible inclusion in the review by checking titles and abstracts. The final decision regarding inclusion was based on the full article. Disagreement was resolved in a consensus meeting.

Both reviewers independently assessed the methodological quality of the selected studies. The criteria list recommended by the Cochrane Methods Working Group on Systematic Review of Screening and Diagnostic Tests was used. Some items on the list were modified for this specific review. The complete criteria list used is shown in Table 1. Internal validity criteria (IV) were scored as ‘positive’ (adequate methods), ‘negative’ (inadequate methods, potential bias), or ‘unclear’ if insufficient information had been provided on a specific item. External validity criteria (EV) were assessed to evaluate generalizibility. Standard performance of FDG-PET was scored ‘positive’ when the type of PET camera, the dose of FDG, the time between injection and scanning and the method of reconstruction were described. The criteria for external validity scored positive if sufficient information was provided to judge generalizibility of findings. After the consensus meeting we decided to score unclear scores as negative. Agreement between both reviewers was quantified by Cohen’s kappa. Quality scores were expressed as a percentage of the maximum score. Subtotals were calculated for internal (maximum 6) and external validity (maximum 6) separately.

Data on sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) of FDG-PET in the detection of both locoregional and distant metastases were calculated from the original numbers given in the publications to avoid rounding-off effects. For some articles, which did not present their data according to the TNM classification, the reviewers restaged patients according to the TNM classification if data were presented in a sufficiently detailed manner.

Numbers of patients with locoregional metastases and distant metastases were placed in a 2x2 table independently by the two reviewers. If data were available for only a subset of patients, those data were included.

Meta-analysis was performed using a bivariate random effects approach to pool the sensitivity and specificity for locoregional lymph nodes and distant metastases. This model assumes a bivariate normal distribution for the logit-transformed sensitivity and specificity values across studies, allowing for additional heterogeneity between studies due to differences in study characteristics. With this model estimates of the mean logit-transformed sensitivity and specificity were obtained. PPV and NPV were not subjected to
Summary estimates of sensitivity and specificity with 95% confidence intervals (CI) were calculated after anti-logarithm transformation of these logit estimates. Statistical analyses were executed with the statistical software package SAS version 8.02 (SAS institute Inc., Cary, NC, USA).

RESULTS

Literature search
A total of 119 studies about initial staging of esophageal cancer with FDG-PET were identified. After reviewing the title and the abstract 98 studies had to be excluded. These studies included reviews, case reports, studies reporting on the use of FDG-PET for response evaluation to neoadjuvant therapy or studies comprising other carcinomas besides EC. Of the remaining 21 studies, 9 were excluded after reviewing the full article. Four of these 9 studies were excluded because of reporting data on the same patients. The other five studies were excluded because of insufficient information to construct a 2x2 table, and because the data of some of these studies were based on number of identified lesions and not on number of identified patients.

Twelve studies met the inclusion criteria. The characteristics of the included studies are shown in Table 2. The total number of patients in a study ranged from 18 to 81 (median, 33 patients). Reported age ranged from 22 to 90 years, and the proportion of male patients ranged from 83% to 100%. Most studies comprised both squamous cell carcinoma and adenocarcinoma. Reference tests consisted of histopathology of resected specimens or

Table 1. Criteria List Used to Assess the Methodological Quality of the Studies

<table>
<thead>
<tr>
<th>Criteria of Validity</th>
<th>Positive score</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Internal Validity (IV)</strong></td>
<td></td>
</tr>
<tr>
<td>1. Valid reference test</td>
<td>Histology, cytology, surgery</td>
</tr>
<tr>
<td>2. Blind measurement of FDG-PET without knowledge of reference test</td>
<td>Mentioned in publication</td>
</tr>
<tr>
<td>3. Blind measurement of reference test without knowledge of FDG-PET</td>
<td>Mentioned in publication</td>
</tr>
<tr>
<td>4. Avoidance of verification bias</td>
<td>Assessment by reference test independent of FDG-PET results</td>
</tr>
<tr>
<td>5. FDG-PET interpreted independently of all clinical information</td>
<td>Mentioned in publication</td>
</tr>
<tr>
<td>6. Prospective study</td>
<td>Mentioned in publication</td>
</tr>
<tr>
<td><strong>External Validity (EV)</strong></td>
<td></td>
</tr>
<tr>
<td>1. Spectrum of disease</td>
<td>All stages of disease</td>
</tr>
<tr>
<td>2. Demographic information</td>
<td>Age and gender information given</td>
</tr>
<tr>
<td>3. Inclusion criteria</td>
<td>Mentioned in publication</td>
</tr>
<tr>
<td>4. Exclusion criteria</td>
<td>Mentioned in publication</td>
</tr>
<tr>
<td>5. Avoidance of selection bias</td>
<td>Consecutive series of patients</td>
</tr>
<tr>
<td>6. Standard execution of FDG-PET</td>
<td>Type of camera, dose FDG, time interval, reconstruction</td>
</tr>
</tbody>
</table>

FDG: 18F-fluorodeoxyglucose; PET: positron emission tomography

this analysis because these values depend on prevalence, which is rarely constant across studies included in a systematic review. Summary estimates of sensitivity and specificity with 95% confidence intervals (CI) were calculated after anti-logarithm transformation of these logit estimates.

Statistical analyses were executed with the statistical software package SAS version 8.02 (SAS institute Inc., Cary, NC, USA).
CHAPTER 2

biopsies obtained during surgical procedures or radiographic follow up.

Methodological quality assessment

Methodological quality was assessed by 12 items for each of the 12 selected studies. There was disagreement in 40 of 144 scores with a Cohen’s kappa of 0.70. Main disagreement was in the questions IV3 and IV5. Disagreements were caused by reading errors and differences in interpretation. The scores for internal and external validity of the 12 selected studies are presented in Table 3. All studies had a valid reference test, but most studies (92%) did not describe whether the reference test was interpreted without knowledge of the FDG-PET findings. In 9 of the 12 studies (75%) verification bias was avoided because patients were

<table>
<thead>
<tr>
<th>Study</th>
<th>Year</th>
<th>No. of Patients</th>
<th>Sex (male/female)</th>
<th>Age</th>
<th>Histology (adenocarcinoma/squamous/carcinoma)</th>
<th>Preoperative Work-Up</th>
<th>No. of Excluded Patients</th>
<th>Reference Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Block et al</td>
<td>1997</td>
<td>58</td>
<td>42/16</td>
<td>Range, 44–84</td>
<td>34/22/2 CT, Chest X-ray, barium swallow, endoscopy</td>
<td>1 PA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kole et al</td>
<td>1998</td>
<td>26</td>
<td>22/24</td>
<td>Range, 41–76</td>
<td>21/4/1 CT, EUS</td>
<td>6 PA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rankin et al</td>
<td>1998</td>
<td>19</td>
<td>-</td>
<td>-</td>
<td>13/- CT</td>
<td>6 PA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kobori et al</td>
<td>1999</td>
<td>33</td>
<td>28/5</td>
<td>Range, 50–81</td>
<td>- endoscopy CT, EUS, bone scintigraphy, bronchoscopy, US of neck</td>
<td>13 PA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Choi et al</td>
<td>2000</td>
<td>48</td>
<td>45/3</td>
<td>Range, 46–77</td>
<td>-48/- CT, EUS, bone scintigraphy, bronchoscopy, US of neck</td>
<td>13 PA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Flamen et al</td>
<td>2000</td>
<td>74</td>
<td>-</td>
<td>-</td>
<td>53/21/- US of neck</td>
<td>1 PA/FU</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Meltzer et al</td>
<td>2000</td>
<td>47</td>
<td>39/8</td>
<td>Mean ± SD, 63±10</td>
<td>37/10/- CT, EUS</td>
<td>20 PA/surgery</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jager et al</td>
<td>2001</td>
<td>18</td>
<td>15/3</td>
<td>Range, 22–75</td>
<td>11/4/3 CT</td>
<td>20 PA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Junginger et al</td>
<td>2002</td>
<td>30</td>
<td>25/5</td>
<td>Median, 63</td>
<td>16/14/- CT</td>
<td>20 PA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kato et al</td>
<td>2002</td>
<td>32</td>
<td>29/3</td>
<td>Range, 42–76</td>
<td>-52/- CT, EUS, bone scintigraphy</td>
<td>20 PA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wren et al</td>
<td>2002</td>
<td>24</td>
<td>24/1</td>
<td>Mean ± SD, 66±11</td>
<td>15/7/2 CT</td>
<td>30 PA/FU</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yoon et al</td>
<td>2003</td>
<td>81</td>
<td>78/3</td>
<td>Range, 31–90</td>
<td>-81/- CT</td>
<td>55 PA</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

CT: computed tomography; BS: barium swallow; PA: pathology; EUS: endoscopic ultrasound; US: ultrasound (external); FU: follow-up (radiographic); SD: standard deviation

Table 2. Characteristics of the 12 Included Studies

Table 3. Quality Assessment of the 12 Diagnostic Studies Included in the Present Review

IV1-IV6: six criteria for internal validity [IV; see Table 1]; EV1-EV6: six criteria for external validity [EV; see Table 1]
selected for assessment by the reference test independently of the FDG-PET results (IV4). Eight studies were prospective (67%) and in six studies (50%) the patients entered the study consecutively. In a majority of the selected studies (67%), all stages of disease were included. Only in a minority of studies were the inclusion (33%) and exclusion (25%) criteria described. The total score for the combined internal and external validity, expressed as a fraction of the maximum score, ranged from 33% to 83% with a median of 63%. Ten of the 12 studies had a total score above 50%.

Analysis
For all studies, a 2x2 table was constructed regarding the detection of locoregional lymph node metastases (N stage) and distant hematogenous and/or distant lymph node metastases (M stage). Supraclavicular, celiac, retroperitoneal and nonadjacent lymph nodes as defined in the study of Block et al were classified as distant lymph nodes (M1). The celiac nodes reported by Kole et al were considered N1 disease because of insufficient details. However, the supraclavicular nodal involvement was sufficiently described and considered as distant metastases. Rankin et al provided detailed information about periesophageal and left gastric lymph nodes only. Therefore, this study was entered in the analysis only concerning the N stage of disease. For both Kobori et al and Kato et al, the involvement of cervical and abdominal lymph nodes was classified as M1 disease, all other nodes were considered N1 disease. Common hepatic, celiac and para-aortic lymph nodes were classified M1 disease in the study presented by Choi et al. The remaining nodes were staged as N1 disease. N and M stage were presented according to the revised TNM classification in the studies of Junginger et al and Wren et al. Meltzer et al did not describe in detail the N and M stage. Therefore, the nodal staging was scored as N disease and distant metastases as M disease. The description in the studies of Flamen et al, Jager et al and Yoon et al enabled restaging of patients to the revised classification. The studies of Kobori et al, Choi et al, Kato et al and Yoon et al only described locoregional and distant lymph nodes. Distant hematogenous metastases had not been detected in their series.

The data of each study and the results of the statistical pooling are shown in Tables 4 and 5 for N stage and M stage, respectively. The total number of patients included for analysis concerning N stage was 421 and the ranges of sensitivity, specificity, PPV and NPV were 0.08 to 0.92, 0.67 to 1.00, 0.70 to 1.00, and 0.64 to 1.00, respectively. The overall pooled sensitivity was 0.51 (95% CI, 0.34 to 0.69), and pooled specificity 0.84 (95% CI, 0.76 to 0.91) for staging locoregional lymph node involvement. The median prevalence of locoregional lymph node metastasis was 0.55, with PPV and NPV of 0.60 and 0.46, respectively. Sensitivity and specificity for distant metastases (M stage) were determined in 452 patients in 11 studies. Sensitivity, specificity, PPV and NPV ranged from 0.33 to 1.00, 0.90 to 1.00, 0.60 to 1.00, and from 0.24 to 0.88, respectively. The overall pooled sensitivity and specificity for M stage were 0.67 (95% CI, 0.58 to 0.76), and 0.97 (95% CI, 0.90 to 1.0), respectively. The median prevalence of distant lymph node and organ metastasis was 0.36, with PPV and NPV of 0.92 and 0.83, respectively.
DISCUSSION

This systematic review and meta-analysis included 12 studies concerning the value of FDG-PET in the staging performance for both locoregional and distant metastases in patients with newly diagnosed cancer of the esophagus or gastroesophageal junction. The included studies had a moderate methodological quality score, with a median of 63% for combined internal and external validity. Pooled sensitivity and specificity for the detection of locoregional metastases were 0.51 and 0.84, respectively. For the detection of distant metastases, pooled sensitivity and specificity were 0.67 and 0.97, respectively.

### Table 4. Parameters of Diagnostic Accuracy of FDG-PET for the Detection of Locoregional Lymph Node Metastases (N stage)

<table>
<thead>
<tr>
<th>Study</th>
<th>Year</th>
<th>Sensitivity</th>
<th>95% CI</th>
<th>Specificity</th>
<th>95% CI</th>
<th>Positive Predictive Value</th>
<th>95% CI</th>
<th>Negative Predictive Value</th>
<th>95% CI</th>
<th>Prevalence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Block et al</td>
<td>1997</td>
<td>0.58</td>
<td>(0.39-0.78)</td>
<td>0.87</td>
<td>(0.73-1.01)</td>
<td>0.82</td>
<td>(0.64-1.01)</td>
<td>0.67</td>
<td>(0.44-0.89)</td>
<td>0.51</td>
</tr>
<tr>
<td>Kole et al</td>
<td>1998</td>
<td>0.92</td>
<td>(0.78-1.01)</td>
<td>0.88</td>
<td>(0.65-1.10)</td>
<td>0.92</td>
<td>(0.78-1.07)</td>
<td>0.88</td>
<td>(0.70-1.06)</td>
<td>0.62</td>
</tr>
<tr>
<td>Rankin et al</td>
<td>1999</td>
<td>0.89</td>
<td>(0.62-1.06)</td>
<td>0.67</td>
<td>(0.39-1.04)</td>
<td>0.71</td>
<td>(0.38-1.05)</td>
<td>0.33</td>
<td>(-0.02-0.68)</td>
<td>0.68</td>
</tr>
<tr>
<td>Kobori et al</td>
<td>2000</td>
<td>0.87</td>
<td>(0.10-0.58)</td>
<td>0.88</td>
<td>(0.71-1.04)</td>
<td>0.75</td>
<td>(0.45-1.05)</td>
<td>0.56</td>
<td>(0.22-0.90)</td>
<td>0.52</td>
</tr>
<tr>
<td>Choi et al</td>
<td>2000</td>
<td>0.82</td>
<td>(0.68-0.96)</td>
<td>0.85</td>
<td>(0.69-1.01)</td>
<td>0.89</td>
<td>(0.76-1.01)</td>
<td>0.77</td>
<td>(0.61-0.93)</td>
<td>0.58</td>
</tr>
<tr>
<td>Flamen et al</td>
<td>2000</td>
<td>0.19</td>
<td>(0.00-0.38)</td>
<td>0.85</td>
<td>(0.65-1.04)</td>
<td>0.60</td>
<td>(0.17-1.03)</td>
<td>0.46</td>
<td>(0.02-0.90)</td>
<td>0.55</td>
</tr>
<tr>
<td>Meltzer et al</td>
<td>2000</td>
<td>0.43</td>
<td>(0.27-0.59)</td>
<td>0.83</td>
<td>(0.62-1.04)</td>
<td>0.88</td>
<td>(0.73-1.04)</td>
<td>0.33</td>
<td>(0.11-0.56)</td>
<td>0.74</td>
</tr>
<tr>
<td>Jager et al</td>
<td>2001</td>
<td>0.08</td>
<td>(-0.07-0.24)</td>
<td>1.00</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.35</td>
</tr>
<tr>
<td>Junginger et al</td>
<td>2002</td>
<td>0.38</td>
<td>(0.17-0.59)</td>
<td>1.00</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-0.06-0.53</td>
</tr>
<tr>
<td>Kato et al</td>
<td>2002</td>
<td>0.67</td>
<td>(0.43-0.91)</td>
<td>0.88</td>
<td>(0.73-1.04)</td>
<td>0.63</td>
<td>(0.62-1.04)</td>
<td>0.75</td>
<td>(0.51-1.00)</td>
<td>0.47</td>
</tr>
<tr>
<td>Wren et al</td>
<td>2002</td>
<td>0.71</td>
<td>(0.38-1.05)</td>
<td>0.86</td>
<td>(0.67-1.04)</td>
<td>0.71</td>
<td>(0.38-1.05)</td>
<td>0.86</td>
<td>(0.60-1.12)</td>
<td>0.33</td>
</tr>
<tr>
<td>Yoon et al</td>
<td>2003</td>
<td>0.64</td>
<td>(0.49-0.79)</td>
<td>0.69</td>
<td>(0.55-0.83)</td>
<td>0.66</td>
<td>(0.51-0.81)</td>
<td>0.67</td>
<td>(0.53-0.62)</td>
<td>0.48</td>
</tr>
<tr>
<td>Pooled estimate</td>
<td></td>
<td>0.51</td>
<td>(0.34-0.69)</td>
<td>0.84</td>
<td>(0.76-0.91)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

### Table 5. Parameters of Diagnostic Accuracy of FDG-PET for the detection of Distant Lymph Node and Organ Metastases (M stage)

<table>
<thead>
<tr>
<th>Study</th>
<th>Year</th>
<th>Sensitivity</th>
<th>95% CI</th>
<th>Specificity</th>
<th>95% CI</th>
<th>Positive Predictive Value</th>
<th>95% CI</th>
<th>Negative Predictive Value</th>
<th>95% CI</th>
<th>Prevalence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Block et al</td>
<td>1997</td>
<td>0.65</td>
<td>(0.42-0.87)</td>
<td>0.97</td>
<td>(0.90-1.03)</td>
<td>0.92</td>
<td>(0.78-1.07)</td>
<td>0.83</td>
<td>(0.62-1.04)</td>
<td>0.36</td>
</tr>
<tr>
<td>Kole et al</td>
<td>1998</td>
<td>1.00</td>
<td>-</td>
<td>0.95</td>
<td>(0.85-1.05)</td>
<td>0.78</td>
<td>(0.33-1.17)</td>
<td>1.00</td>
<td>-</td>
<td>0.13</td>
</tr>
<tr>
<td>Rankin et al</td>
<td>1998</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Kobori et al</td>
<td>1999</td>
<td>0.87</td>
<td>(0.70-1.04)</td>
<td>0.94</td>
<td>(0.84-1.05)</td>
<td>0.93</td>
<td>(0.79-1.06)</td>
<td>0.90</td>
<td>(0.73-1.05)</td>
<td>0.45</td>
</tr>
<tr>
<td>Choi et al</td>
<td>2000</td>
<td>0.56</td>
<td>(0.32-0.81)</td>
<td>1.00</td>
<td>-</td>
<td>1.00</td>
<td>-</td>
<td>0.82</td>
<td>(0.73-1.05)</td>
<td>0.33</td>
</tr>
<tr>
<td>Flamen et al</td>
<td>2000</td>
<td>0.74</td>
<td>(0.59-0.88)</td>
<td>0.90</td>
<td>(0.81-0.99)</td>
<td>0.86</td>
<td>(0.74-0.99)</td>
<td>0.80</td>
<td>(0.65-0.95)</td>
<td>0.46</td>
</tr>
<tr>
<td>Meltzer et al</td>
<td>2000</td>
<td>0.70</td>
<td>(0.42-0.98)</td>
<td>0.92</td>
<td>(0.83-1.01)</td>
<td>0.70</td>
<td>(0.42-0.98)</td>
<td>0.92</td>
<td>(0.75-1.09)</td>
<td>0.22</td>
</tr>
<tr>
<td>Jager et al</td>
<td>2001</td>
<td>0.80</td>
<td>(0.45-1.51)</td>
<td>1.00</td>
<td>-</td>
<td>1.00</td>
<td>-</td>
<td>0.93</td>
<td>(0.68-1.18)</td>
<td>0.28</td>
</tr>
<tr>
<td>Junginger et al</td>
<td>2002</td>
<td>0.33</td>
<td>(0.07-0.60)</td>
<td>1.00</td>
<td>-</td>
<td>1.00</td>
<td>-</td>
<td>0.64</td>
<td>(0.17-1.11)</td>
<td>0.46</td>
</tr>
<tr>
<td>Kato et al</td>
<td>2002</td>
<td>0.71</td>
<td>(0.48-0.95)</td>
<td>1.00</td>
<td>-</td>
<td>-</td>
<td>1.00</td>
<td>0.82</td>
<td>(0.58-1.06)</td>
<td>0.44</td>
</tr>
<tr>
<td>Wren et al</td>
<td>2002</td>
<td>0.67</td>
<td>(0.40-0.93)</td>
<td>0.92</td>
<td>(0.76-1.07)</td>
<td>0.89</td>
<td>(0.68-1.09)</td>
<td>0.73</td>
<td>(0.44-1.02)</td>
<td>0.50</td>
</tr>
<tr>
<td>Yoon et al</td>
<td>2003</td>
<td>0.43</td>
<td>(0.06-0.80)</td>
<td>0.99</td>
<td>(0.96-1.01)</td>
<td>0.75</td>
<td>(0.33-1.17)</td>
<td>0.95</td>
<td>(0.73-1.17)</td>
<td>0.09</td>
</tr>
<tr>
<td>Pooled estimate</td>
<td></td>
<td>0.67</td>
<td>(0.58-0.76)</td>
<td>0.97</td>
<td>(0.90-1.0)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
The results of this systematic review should be interpreted with caution because of several limitations. First, the included studies have limited methodological quality. In these studies it was not clear whether the reference test was interpreted independently of the index test, which might lead to diagnostic bias. In general, this leads to overestimation of the diagnostic accuracy. The retrospective design in four studies as well as the interpretation of FDG-PET with other available clinical information available further decreased the methodological quality. In three studies, there was verification bias, because the reference test was assessed on patients selected by the index test results, which can lead to overestimation of the sensitivity. While common in clinical practice, diagnostic studies should avoid this preferential ordering of a gold standard test. Another type of bias related to patient selection is spectrum bias, which occurs when the test performance in the research population differs from that seen in day-to-day customary care. This type of bias was present in three studies, which did not include all stages of disease. To improve the methodological quality in reporting of diagnostic studies, the new standards for reporting of diagnostic accuracy (STARD) checklist will be a useful and perhaps an obligatory resource.

Unfortunately, most studies did not stage their patients according to the TNM classification. Many studies did not enable to stage all patients consistently as having locoregional or distant lymph node metastases. Celiac trunk metastases were considered as N disease in some studies, but M disease in others. Therefore, analysis was performed excluding the 4 studies that did not enable staging according to the UICC classification, yielding a pooled sensitivity and specificity for N stage of 0.47 (95% CI, 0.22 to 0.71) and 0.83 (95% CI, 0.71 to 0.95), respectively. Sensitivity and specificity regarding M stage were 0.61 (95% CI, 0.48 to 0.75) and 0.97 (95% CI, 0.84 to 1.0), respectively. This analysis showed a slightly lower sensitivity for detection of locoregional (N stage) and distant metastases (M stage). Studies reporting on the staging of esophageal cancer should therefore use the UICC classification to prevent confusion, especially about distant lymphatic disease.

The pooled sensitivity and specificity of FDG-PET for the detection of locoregional metastases were low, 0.51 and 0.84, respectively. Reasons for this may be the inhomogenous tracer uptake in the primary tumor, masking of adjacent lymph nodes by tracer accumulation in the primary tumor and the occurrence of false-positive findings due to chronic inflammation. In both day-to-day clinical practice and in ongoing clinical trials, the presence or absence of lymph node metastases may have an impact on systemic treatment selection. Because of the moderate sensitivity for the detection of lymph node metastases, FDG-PET alone does not appear to be suitable for the allocation of neoadjuvant therapy in these patients. Currently, EUS combined with FNA is the first choice modality to assess locoregional lymph node involvement.

The pooled sensitivity and specificity for the detection of distant lymph node metastases and hematogenous metastases was 0.67 and 0.97, respectively. The interpretation of these reasonable results is complicated by the heterogeneity of the M stage in this review, as mentioned earlier. Of the 11 included studies, 2 had conspicuously low sensitivity for the detection of distant metastases. Yoon et al attributed the low sensitivity in their study to the inclusion criteria. More cases of early-stage disease, and therefore more patients with only microscopic metastatic foci, might have been included in their study. Junginger et al argued that the uptake of FDG is low in smaller tumors and in tumors obtaining most of their energy from metabolic pathways other than glycolysis, which might explain the low
sensitivity they reported. Exclusion of both outliers from the present meta-analysis results in a pooled sensitivity of 0.72 and pooled specificity of 0.95. In contrast to the N stage of disease, M stage directly influences the clinical management of esophageal cancer. The presence of distant hematogenous metastases excludes a curative intended surgical, leading to a palliative nonsurgical treatment as the only treatment option. The optimal management of patients with positive celiac trunk nodes is still a matter of debate.

For many years, CT has been the first-line method to detect distant metastases in EC with a sensitivity of 37% to 66%. Later on, EUS was introduced and became the most reliable method for determining T stage and identifying pathological involvement of regional lymph nodes. EUS combined with FNA enabled selective aspiration of echographically suspected nodes, including those at the celiac trunk. Sensitivity, specificity, positive predictive value and negative predictive value for the assessment of celiac lymph nodes are ranging from 53% to 98%, 77% to 100%, 79% to 100%, and 82% to 100%, respectively.

Different studies have shown a high accuracy of FDG-PET. However, the hallmark for implementation in diagnostic work-up is the ability to change patient management due to more accurate staging. Of the included studies, the change in patient management ranged from 3% to 20% due to addition of FDG-PET to preoperative work-up. However, these studies involved only a limited number of patients.

PET may be cost-effective in the prevention of noncurative surgery by detecting metastases not identified by conventional staging modalities. Recently, the spatial resolution of CT has increased by a multislice technique. Therefore, in future studies FDG-PET should preferably be compared to multislice CT after an obligate EUS-FNA examination.

In this systematic review, FDG-PET was found to have moderate sensitivity and specificity in the detection of locoregional lymph node metastases, with considerable heterogeneity across the included studies. In the detection of distant nodal and hematogenous metastases, FDG-PET has reasonable sensitivity and specificity, with a lower degree of heterogeneity. As M stage determines patient management, we feel that the potential contribution of FDG-PET to staging should carry more weight than its role in N staging when deciding whether or not to implement FDG-PET in the standard preoperative work-up of patient with esophageal cancer. Larger prospective studies should quantify to what extent the routine use of FDG-PET leads to changes in management and better health care for these patients.

ACKNOWLEDGMENT

This study was supported by a ZonMw program for Health Care Efficiency Research.
REFERENCES

CHAPTER 2


