A systematic review on the role of FDG-PET/CT in tumour delineation and radiotherapy planning in patients with esophageal cancer

Christina T. Muijs, Jannet C. Beukema, Jan Pruim, Veronique E. Mul, Henk Groen, John Th. Plukker, Johannes A. Langendijk

Purpose: FDG-PET/CT has proven to be useful in the staging process of esophageal tumours. This review analysed the role of FDG-PET/CT in tumour delineation and radiotherapy planning in comparison with CT alone among patients with esophageal cancer. Thereby we focused on the detection of the primary tumour and lymph nodes by FDG-PET/CT, changes in target volume (TV) delineation based on FDG-PET/CT and its validity, changes in inter- and intra-observer variability in TV delineation, consequences for radiotherapy treatment planning with regard to either target volumes or organs at risk and finally on the validation of FDG-PET/CT-based TVs in terms of treatment outcome.

Methods: A literature search was performed in MEDLINE and Cochrane library databases for studies concerning the current value of FDG-PET/CT in tumour detection and delineation and radiotherapy-planning procedures among patients with esophageal cancer. Both prospective and retrospective studies were included.

Results: Fifty publications met the eligibility criteria, of which 19 were review papers and one was a case report. The remaining 30 publications reported on the results of original studies. FDG-PET was able to identify most primary tumours, with a sensitivity and specificity for the detection of metastatic lymph nodes of 30–93% and 79–100%. The use of FDG-PET/CT resulted in changes of target volumes, and consequently in changes in treatment planning. However, evidence supporting the validity of the use of FDG-PET/CT in the tumour delineation process is very limited. Only three studies reported a significant positive correlation between FDG-PET-based tumour lengths and pathological findings. There were two studies that tested the influence of FDG-PET/CT to the inter- and intra-observer variability. One of them found a significant decrease in inter- and intra-observer variability, while the others did not. Furthermore, there are no studies demonstrating the use of PET/CT in terms of improved locoregional control or survival.

Conclusion: Since the literature is very limited standard implementation of FDG-PET/CT into the tumour delineation process for radiation treatment seems unjustified and needs further clinical validation first.

Corresponding author. Address: Department of Radiation Oncology, University Medical Center Groningen, University of Groningen, P.O. Box 30001, 9300 RB Groningen, The Netherlands.

E-mail address: ct.muijs@rt.umcg.nl (C.T. Muijs).

0167-8140/$ - see front matter © 2010 Elsevier Ireland Ltd. All rights reserved. Radiotherapy and Oncology 97 (2010) 165–171
coverage and seems to spare normal tissues in various tumour types, in particular in non-small cell lung cancer (NSCLC) [8,38].

Several studies showed that FDG-PET has been very useful in the staging process of esophageal tumours [37]. FDG-PET is superior in detecting distant metastases and seems to improve nodal staging as well, especially for non-adjacent lymph nodes. Furthermore, FDG-PET can be an effective tool for restaging esophageal malignancies after (neo-adjuvant) treatment [41,42].

This study focuses on the additional value of FDG-PET/CT in the tumour delineation process in comparison with CT alone among patients with esophageal cancer. More specifically, the purpose of this review were (1) to analyse the ability of FDG-PET/CT to detect the primary tumour and/or pathologic lymph nodes; (2) to determine if, and to what extent, the addition of FDG-PET changes target volume delineation; (3) to assess the validity of FDG-PET/CT with regard to GTV delineation; (4) to assess if the addition of FDG-PET improves inter-observer and intra-observer variability in target volume delineation; (5) to determine the consequences for radiotherapy treatment planning with regard to either target volumes or organs at risk, and finally; (6) to assess the validation of FDG-PET/CT-based tumour delineation on treatment outcome.

Methods

Search strategy and selection criteria

A literature search was performed to retrieve articles concerning the detection of esophageal tumours and pathologic lymph nodes using the following keywords:

- synonyms for esophageal cancer
- synonyms for PET/CT
- synonyms for detection or synonyms for visualisation

These keywords were combined using ‘AND’. The search was carried out in MEDLINE and Cochrane Library. In addition, references lists of papers were screened in order to retrieve additional relevant papers.

A similar search was performed for studies concerning the current value of FDG-PET/CT in tumour delineation and radiotherapy-planning procedures among patients with esophageal cancer. The following keywords were used:

- synonyms for esophageal cancer
- synonyms for PET/CT
- synonyms for tumour delineation or synonyms for radiotherapy

To be selected for this review, studies had to fulfil the following eligibility criteria: (1) squamous cell or adenocarcinoma of the esophagus or the gastro-esophageal junction; (2) eligible for curative treatment; (3) use FDG-PET in conjunction with CT; (4) FDG used as a tracer in PET.

Both prospective and retrospective studies were included. Studies only available in abstract form were excluded from this review. Articles in languages other than English were excluded as well.

The selection process of both search strategies together is summarized in Fig. 1.

Results

Literature search

Using the search strategies described, we were able to identify 114 publications. However, only 50 publications met the eligibility criteria, of which 19 were review papers. In these papers, the use of FDG-PET/CT in tumour staging process or in radiotherapy in general was reviewed. One publication reported on a single case.

In the 30 studies, a total number of 1222 patients were included. Table 1 summarizes the original studies concerning the detection of esophageal cancer. Table 2 summarizes the original studies concerning the tumour delineation process.

For reasons of clarity, the abbreviation FDG-PET/CT refers to both software fusion-based PET added to CT and PET/CT images acquired with an integrated PET/CT scanner.

FDG-avidity of the primary tumour

Several studies have investigated the detection of the primary tumour by FDG-PET. Increased uptake of FDG was seen in 68–100% of the esophageal tumours (Table 1). Undetected tumours are mostly stages T1 and T2 tumours. Especially T1a tumours, remaining within the submucosa, are difficult to detect by FDG-PET [3,17].

Kato et al. [16] found a significant relationship between the intensity of the primary tumour FDG-uptake, expressed as SUV, and the depth of the tumour invasion. However, Flamen et al. [10] found no correlation between SUV and pT-stage.

Detection of locoregional metastatic lymph nodes

The ability of different imaging modalities to identify metastatic lymph nodes has been widely investigated. However, the literature is not very consistent when it comes to the ability of CT, FDG-PET or FDG-PET/CT to identify pathologic locoregional lymph nodes. The sensitivity of CT and FDG-PET varied widely; 11–93% vs. 30–93%. There was less variation regarding the specificity of CT and FDG-PET; 71–100% vs. 79–100%, respectively (Table 1).

Recently, van Vliet et al. [36] published a meta-analysis on staging investigations for esophageal cancer. In this analysis, they found a pooled sensitivity for the detection of regional lymph node metastases by CT and FDG-PET of 0.50 (95% CI 0.41–0.60) and 0.57 (95% CI 0.43–0.70), respectively. The pooled specificity of CT was 0.83 (95% CI 0.77–0.89) vs. 0.85 (95% CI 0.76–0.95) for FDG-PET. The pooled sensitivity of EUS was significantly higher than both CT and FDG-PET, but showed a similar diagnostic performance.

Four other studies evaluated the use of FDG-PET/CT for the detection of locoregional lymph node metastases. Yuan et al. [46] found a significant improvement of the sensitivity (93%), accuracy (92%) and negative predictive value (98%) in the assessment of locoregional lymph nodes for esophageal cancer by the use of integrated FDG-PET/CT, compared to the use of FDG-PET alone. Silho et al. [33] also compared the ability of CT and FDG-PET to identify metastatic lymph nodes to integrated FDG-PET/CT. In this study,
the sensitivity was 42%, 35% and 50% for CT, FDG-PET and integrated FDG-PET/CT, respectively. Although FDG-PET/CT improved the sensitivity, it remained significantly lower than that for EUS ($p = 0.001$). Both FDG-PET and FDG-PET/CT had a specificity of 100%, while CT showed a specificity of 82%. Schreurs et al. [32] compared the detection of locoregional lymph node metastases...
on fused FDG-PET/CT to the detection of CT and FDG-PET side by side in 18 patients. They found that the use of fused FDG-PET/CT images improved the sensitivity (87% vs. 80%) and the specificity (87% vs. 83%). Kato et al. [15] also evaluated the additional value of PET/CT over PET for the detection of a metastatic lymph node group. This study demonstrated that PET/CT had a higher sensitivity in lower thoracic regions than both PET and CT (p < 0.05).

Target volume modifications

In nine studies, the use of FDG-PET/CT, based on either software fusion [25,26,21,40] or integrated FDG-PET/CT [9,11,14,21,39], resulted in changes of target volumes.

Gondi et al. found that the addition of FDG-PET resulted in GTV reduction with more than 5% in 10 of 16 patients (62.5%). The mean overlap between FDG-PET/CT- and CT-based GTVs was low, only 46%. The observers delineated the GTVs independently.

Moureau-Zabotto et al. [25] evaluated the impact of the addition of FDG-PET to CT on GTV and PTV delineation among 34 patients treated with 3D-conformal radiotherapy. In this study, the target volume was initially based on CT images. The corresponding FDG-PET data were used as an overlay to the CT data to define the GTV. This resulted in tumour length reduction in 12 patients (35%) and an increase in 12 patients (35%). Modifications in GTV delineation resulted in a mean reduction of the GTV of 21.3% and a mean increase of 20.0%. These changes in GTV affected the planning target volume (PTV) in 18 patients (53%).

Leong et al. [21] also evaluated the impact of FDG-PET on CT-based target volume delineation. They found an exclusion of FDG-PET-avid disease in 11 of 16 patients (69%) when the GTV was based on CT alone, with a median exclusion of FDG-PET/CT-based GTV of 38%. In five patients (31%), the FDG-PET/CT-based GTV was located outside the CT-based PTV, with a median exclusion of 5%. Modifications based on FDG-PET/CT were mainly seen in longitudinal direction.

Vrieze et al. [40] focused on the detection of pathologic lymph node regions, instead of primary tumour extension and suggested that the GTV should not be reduced based on negative FDG-PET findings given the low sensitivity of FDG-PET in esophageal cancer. Therefore, they stated that the GTV could only be enlarged based on positive FDG-PET findings. In this study, the irradiated volume was based on conventional imaging modalities, including CT and endoscopic ultrasound (EUS). In 6 patients (20%), 8 pathologic lymph node regions were detected on FDG-PET which were not visible on conventional imaging. In three of these patients (10%) this would have led to an increase of the GTV.

In most studies visual interpretation of FDG-PET images was used for target volume delineation. Only a few studies used automatic contouring, based on certain intensity levels, such as the standard uptake value (SUV) or the source-to-background ratio (SBR) [20,22,47].

Pathological validation of FDG-PET findings

As described above, the addition of FDG-PET/CT resulted in changes in the delineation of target volumes in a considerable proportion of patients (20–94%). However, the question that remains to be answered is whether these changes are justified. More specifically, the question arises as to whether the FDG-PET/CT-based changes correspond better with the actual pathological tumour extension. The most ideal method to validate the additional value of FDG-PET/CT would be to compare GTV delineations based on different imaging modalities with the pathological specimens. However, accurate validation is hampered by the difficulties in obtaining pathological material and in comparing resected specimens with pre-surgical imaging. Therefore, reports on pathologic validation are limited.

Mamede et al. [22] evaluated the correlation between tumour length of esophageal carcinoma based on FDG-PET and pathological data (n = 17). These tumour length measurements were taken from fresh tissues obtained by surgical resections, without neo-adjuvant treatment. They reported a significant positive correlation between FDG-PET-based tumour lengths, estimated for different SUV-thresholds, and pathological findings. The best correlation was estimated for a SUV-threshold of 2 – standard deviations (r = 0.74; p < 0.001).

Similar results were found by Zhong et al. [47], comparing FDG-PET-based tumour lengths with pathological specimens (n = 36). In this study, three different methods were used to delineate the GTV: (1) visual interpretation (LengthVis); (2) threshold at SUV 2.5 (Length2.5) and (3) threshold at SUV at 40% of the maximum SUV (Length40). To correct for shrinkage of the surgical specimens, the pathologic tumour length was obtained by measuring the tumour in situ before surgical removal. In addition, after surgical resection, the specimen was longitudinally opened and stretched to the same length as measured in vivo and pinned on a flat board. The FDG-PET-based tumour length correlated well with the pathological tumour length, for all thresholds (Lvis: r = 0.828; L2.5: r = 0.887; L40: r = 0.857; p < 0.001). However, the tumour length at a cutoff of SUV 2.5 seemed most approximate to the pathological tumour length.

Recently, Han et al. [12] compared FLT-PET- and FDG-PET-based tumour lengths with the tumour length in the pathologic specimen of 22 patients. They found that for FDG-PET a SUV cutoff of 2.5 provided the closest estimation of GTV length.

Yu et al. [45] performed a similar study and compared FDG-PET-based tumour lengths with the length of the gross tumour region in vivo in 16 patients. They found no significant difference in absolute value between the CT-, PET- (SUVcut-off: SUVbackground + 20%) and pathology-based tumour lengths. However, regarding the spatial conformity index, the conformity index (CI) of pathology and PET_{SUV+20%} was significantly superior to the CI obtained with pathology and CT.

FDG-PET-based tumour lengths were also found to correlate well with EUS-based tumour lengths [22], the gold standard for clinical T-staging. Konski et al. [20] demonstrated that EUS-based measurements of tumour length closely approximated FDG-PET-based tumour measurements (n = 25), using a SUV-threshold for malignancy of 2.5. The tumour length as determined by PET correlated better with endoscopy than with CT and was significantly shorter as compared to those measured by CT, regardless of the SUV. In this study, the CT-based tumour length was determined on the CT part of the integrated FDG-PET/CT scan. The CT- and FDG-PET-based tumour lengths were determined independently by two different observers.

Inter- and intra-observer variability

Another way to investigate the validity of new imaging techniques in target volume delineation is to test inter- and intraobserver variability, based on the assumption that lower inter- and intra-observer variability represents more accurate delineation.

The effect of the addition of FDG-PET/CT on intra- and interobserver variability in target volume delineation in patients with gastro-esophageal cancer was investigated in two studies. In the study of Vesprini et al. [39], target volumes (n = 10) were delineated based on CT and FDG-PET/CT in ten patients by six radiation oncologists. Combined use of FDG-PET/CT for delineation of the GTV significantly decreased both intra- and inter-observer variability. In 57% of the cases, observers felt more confident with the results of GTV contouring.
Subsequently, these changes may result in both inadequate dose coverage of the PTV and, in some cases, even of the CTV and GTV, and in clinically relevant changes in dose distributions to normal tissues [11,21,25].

There is no doubt that GTV delineation is affected by the diagnostic information used. In clinical practice, all available information, such as physical examination, endoscopy reports, EUS and diagnostic contrast-enhanced CT scans, is applied for delineation of the target volume. In most of the reviewed studies, this information was available for both CT- and FDG-PET/CT-based GTV delineation [11,14,21,39]. However, in other studies, this information, which we consider essential, was disregarded and the delineation of the GTVs was exclusively based on CT- or FDG-PET/CT-information [20,25,40]. This implies that the results of these latter studies may be less representative for use in clinical practice.

The clinical use of FDG-PET/CT may also be hampered by some technical issues. For FDG-PET/CT-based GTVs, two types of contouring methods were used, including visual interpretation (with or without source-to-background correction), or semi-automatic contouring based on different SUV-thresholds. However, these methods are neither objective nor uniform. For visual interpretation, which is the main tool used in clinical practice and used in most of the reviewed studies, image representation can be controlled by changing window-widths and window-levels, which is highly observer dependent, and may result in significant differences in visible tumour volumes. The SUV is, on the other hand, a semi-quantitative parameter for evaluation of the FDG-uptake in tumours. However, many factors, such as patient preparation procedures, scan acquisition, image reconstruction and data analysis settings, affect the outcome of the SUV [5,6]. Even though these factors have small effects individually, accumulation of many of these factors can result in considerable differences in SUV outcome [4]. For this reason, recommendations for standardisation and quantification of FDG-PET studies have been made by Boellaard et al. [4,6]. In the currently available literature, SUV-thresholds to distinguish pathological tissues from normal tissues could not be determined for esophageal cancer. For other tumours, for example, NSCLC, estimation of a reliable automatic contouring method remains difficult as well [27].

The way in which additional FDG-PET information is incorporated in the target delineation process will also influence the changes in target volume. In the reviewed studies, two main methods (Table 3) were used: (1) delineation of the CT-based GTV and modification of this target volume based on the additional PET information; (2) independent delineation of CT- and PET/CT-GTVs. Using the latter method, target volume modifications could partially be explained by intra-observer variability. On the other hand, observers using the first method are more prejudiced by their CT findings. The lack of uniformity in the methods, by which FDG-PET information has been used to contour the GTV, illustrates the difficulty to incorporate FDG-PET information properly into the tumour delineation process.

Another shortcoming is the use of co-registered FDG-PET/CT images for tumour delineation. Software-based co-registration seems less accurate, considering the errors associated with this type of image co-registration. To minimize these errors, all FDG-PET- and CT-scans, used in the reviewed studies, were made in treatment position.

Despite the non-standardised way of the use of FDG-PET for tumour delineation, which might result in an over or under estimation of the GTV modifications, these changes may have an effect on treatment outcome in terms of both the locoregional tumour control and the incidence of radiation-induced side effects.

It is clear that the main advantage of FDG-PET/CT in the management of esophageal cancer is the higher validity with regard to lymph node status and the detection of occult metastases. Integrated FDG-PET/CT has overall a higher sensitivity, specificity and...
Table 3
Changes in target volume by addition of PET(CT) in tumour delineation process.

<table>
<thead>
<tr>
<th>Authors</th>
<th>TV changes</th>
<th>TV of interest</th>
<th>PET interpretation</th>
<th>Contour method</th>
<th>Fusion/integrated PET/CT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vrieze et al.</td>
<td>Overall</td>
<td>Increase</td>
<td>Size</td>
<td>Increase</td>
<td>Size</td>
</tr>
<tr>
<td>[40]</td>
<td>20 (6/30)</td>
<td>10 (3/30)</td>
<td>Not defined</td>
<td>10 (3/30)</td>
<td>Not defined</td>
</tr>
<tr>
<td>Gondi et al.</td>
<td>94 (15/16)</td>
<td>31 (5/16)</td>
<td>Not defined</td>
<td>62.5 (10/16)</td>
<td>&gt;5%</td>
</tr>
<tr>
<td>Hong et al.</td>
<td>84 (21/25)</td>
<td>–</td>
<td>Δ sup/inf extent + 1 cm</td>
<td>–</td>
<td>Δ sup/inf extent + 1 cm</td>
</tr>
<tr>
<td>Moureau-Zabotto</td>
<td>56 (19/34)</td>
<td>21 (7/34)</td>
<td>203 ± 8.7</td>
<td>35 (12/34)</td>
<td>21.3 ± 4.7</td>
</tr>
<tr>
<td>Leong et al.</td>
<td>53 (18/34)</td>
<td>21 (7/34)</td>
<td>22% ± 11</td>
<td>33 (11/34)</td>
<td>9.8% ± 7.4</td>
</tr>
<tr>
<td>Muijs et al.</td>
<td>&gt;69 (11/16)</td>
<td>–</td>
<td>Not defined</td>
<td>–</td>
<td>Not defined</td>
</tr>
<tr>
<td></td>
<td>76 (16/21)</td>
<td>24 (5/21)</td>
<td>&gt;3 mm</td>
<td>52 (11/21)</td>
<td>&gt;3 mm</td>
</tr>
</tbody>
</table>

Abbreviations: TV, target volume; GTV, gross tumour volume; CTV, clinical target volume; PTV, planning target volume; SBR, source-to-background ratio.

Conclusion

FDG-PET is able to detect most esophageal tumours and seems useful for the detection of locoregional lymph nodes. However, evidence supporting the use of FDG-PET/CT in the tumour delineation process and radiotherapy planning is very limited. Tumour length comparison as pathological validation has important shortcomings and seems therefore unreliable. Furthermore, there are no studies demonstrating the use of PET/CT in terms of improved locoregional control or survival. Standard implementation of FDG-PET/CT into the tumour delineation process for radiation treatment seems therefore unjustified at this moment and needs further clinical validation first. This is now subject of a prospective multicenter study in the Netherlands.

Conflict of interest statement

The authors state that the research presented in this manuscript is free of conflicts of interest.

References

computed tomography (FDG-PET/CT) for tumor staging in solid tumors: comparison with CT and PET. J Clin Oncol 2004;22:4357–68.