Investigations and research

Improving prediction of radiotherapy response and optimizing target definition by using FDG-PET for lung cancer patients

R.J.H.M. Steenbakkers
G.R. Borst
M. van Herk
H. Bartelink

The outcome of patients with non-small-cell lung cancer (NSCLC) treated with radiotherapy alone used to be very poor, mainly due to advanced tumor stage and poor performance status [1]. The addition of chemotherapy to radiation has improved prognosis, especially for the more advanced stages of NSCLC [2].

Another way to improve treatment outcome is to optimize the radiotherapy itself. Until the late 1990’s, X-ray images were used to depict the radiation fields. Square radiation fields around the primary tumor and mediastinum were drawn manually. To avoid the possibility of a geographical miss, these radiation fields were large. Consequently, it was not possible to apply larger doses than 50 – 60 Gy without the risk of excessive toxicity (radiation pneumonitis and/or radiation esophagitis).

With the development of computer tomography (CT), defining the target for radiotherapy became more accurate. Now, most radiotherapy departments have there own CT scanner.

Although defining pathological tissue on CT images is easier than on conventional radiographs, estimation of the extent of a tumor still seems to be difficult. In particular, identification of pathological lymph nodes and separating tumor from atelectasis is still challenging when using CT alone [3, 4]. Consequently, if several radiation oncologists and radiologists are asked to define the Gross Tumor Volume (GTV) of a single patient on CT, there is likely to be a large variability will be seen [3, 4], which could easily lead to a geographical miss and underdosage of the tumor [5].

Recently, implementation of a 2-[18F]fluoro-2-deoxy-D-glucose Positron Emission Tomography (FDG-PET) in the diagnostic procedure of NSCLC has become the standard of care. Compared to other image modalities, FDG-PET is superior in detecting pathological lymph nodes in the mediastinum and detecting distant metastasis [6].

Several authors have already demonstrated that implementing FDG-PET in the radiotherapy planning has a large impact on treated volume [7, 8]. However, there has been no large-scale study among radiation oncologists evaluating the reproducibility of defining/delineating the GTV using the combination of CT and FDG-PET.

FDG-PET is a biological image modality. Neoplastic cells regulate their glucose metabolism upwards in order to proliferate. The measurement of the standard uptake value (SUV) is a semi-quantitative method for assessing the uptake of FDG and hence the glucose metabolism in the tumor. Higher SUV’s are correlated with higher proliferation rates. For operable NSCLC patients, the SUV is a prognostic factor for survival [9,10]. The question is whether SUV is also a predictive factor in terms of treatment response and survival for patients treated with high-dose radiotherapy.

This article is in two parts. First, we investigated the impact of matched FDG-PET with CT, compared with CT alone, on target definition in a multi-institutional study. Secondly, we evaluated whether SUV can be used to predict the outcome of high-dose radiotherapy.

Materials and methods

Target definition study

For the target definition study, 22 patients with NSCLC (stage I – IIIB) were selected. Separate CT scans and PET scans were made for each patient (no dedicated PET-CT scanner was available at that time). The CT and FDG-PET scans were matched automatically.

Eleven experienced radiation oncologists from five different radiotherapy departments in the
Netherlands, considered to be experts in lung cancer treatment, were first asked to delineate the GTV of each patient on CT. More than one year later they were asked to delineate the GTV again, but now on the matched FDG-PET and CT scans.

For the first delineation round, delineation software was designed containing most of the delineation possibilities that are available in commercial planning systems. For the second delineation round, the delineation software was adapted to display the matched FDG-PET CT scans and optimized to assist the radiation oncologists. One example is that the radiation oncologists were able to see the CT in mediastinal and lung window settings at the same time.

When the radiation oncologists finished delineating, the contours were sent to our institute via the internet for analysis. The variability of the contours was analyzed in 3-D as described elsewhere [11].

Planning Target Volume (PTV) margins should incorporate all geometrical uncertainties such as setup variation, organ motion and variation in delineation. To evaluate the impact of these geometrical uncertainties on the PTV margin, we used the margin recipe described by van Herk et al. [12, 13]. The PTV margin around the GTV should be 2.5 times the SD of the systematic error plus 0.7 times the random error minus 3 mm in order to irradiate the GTV with at least 98% EUD (equivalent uniform dose) in 90% of the cases.

**SUV study**

To measure SUV, 51 medically or technically inoperable patients with NSCLC (stage I – III B) underwent a FDG-PET scan. Both transmission and emission scans were required in order to analyze the SUV retrospectively. Recorded characteristics were age, gender, performance status, histology, tumor volume, weight loss, treatment response, and tumor and lymph node stage. Treatment response was scored three months after treatment and defined according to the Response Evaluation Criteria in Solid Tumors (RECIST) criteria [14].

The mean radiotherapy dose delivered to the primary tumor and FDG-PET positive lymph nodes was 77 Gy (range 60 – 94.5 Gy). The majority of the patients (n = 34) were derived from a Phase I/II dose escalation study [15]. Ten patients received either induction or concurrent chemotherapy.

The SUV max was defined as the maximum tumor concentration of FDG divided by the injected dose and corrected for the body weight of the patient: 
\[
\text{SUV} = \frac{\text{maximum activity concentration}}{\left(\frac{\text{injected dose}}{\text{body weight}}\right)}.
\]

For statistical analysis, overall survival and disease-specific survival were measured. Survival probabilities were estimated using the Kaplan-Meier method. Logistic regression analysis was performed to assess which patients’ characteristics (including SUV) were significant for predicting treatment response.

**Results**

**Target definition study**

The mean delineated volume of all delineated GTVs using the matched FDG-PET CT scans was reduced from 69 cm³ to 62 cm³ (p = 0.041, paired Students’ t-test) compared to the GTV based on CT alone. Using CT alone, in four patients not a single point was found in the CT scan that was included in the GTV by all radiation oncologists, indicating that there was complete disagreement between the radiation oncologists as to where the tumor was located (Figure 1). In the second delineation round, using the matched FDG-PET CT scans, there was always at least one common point found.
that was included in the GTV by all radiation oncologists.

For the 3D analysis, the variability in delineation among radiation oncologists using the matched FDG-PET CT was 0.42 cm (1 SD). This is significantly smaller than the 1.02 cm (1 SD) for the delineations based on CT alone (Table 1). For 18 patients out of 22, the variability was reduced when matched FDG-PET CT was used. For two patients, the variability remained unchanged. For the remaining two patients, a significantly worse variability was observed. This increase is probably caused by the fact that FDG-PET revealed pathological lymph nodes, whereas with CT alone these patients were staged as lymph node negative.

Using CT alone, the largest variability was found at the atelectasis region (1.91 cm, 1 SD). With the addition of FDG-PET, the variability was reduced to 0.48 cm. In three out of five patients with atelectasis, the atelectasis could be separated from the primary tumor using the matched FDG-PET CT. In two patients, the atelectasis also had increased FDG-uptake. For these patients the radiation oncologists were instructed that atelectasis with increased FDG-uptake had to be included into the GTV. In the case of lymph nodes, the variability remains large, even with the addition of FDG-PET (0.82 cm, 1 SD). This is probably because it is still difficult to identify the extent of FDG-PET positive lymph nodes on CT without contrast enhancement.

### Table 1. Delineation variability measured in 3D for delineations based on CT alone and on matched FDG-PET CT per anatomical region.

<table>
<thead>
<tr>
<th>Anatomical regions</th>
<th>CT alone</th>
<th>Matched FDG-PET CT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tumor – lung</td>
<td>0.59</td>
<td>0.33</td>
</tr>
<tr>
<td>Tumor – mediastinum</td>
<td>0.74</td>
<td>0.44</td>
</tr>
<tr>
<td>Tumor – chest wall</td>
<td>0.40</td>
<td>0.37</td>
</tr>
<tr>
<td>Lymph nodes</td>
<td>1.46</td>
<td>0.82</td>
</tr>
<tr>
<td>Tumor – atelectasis</td>
<td>1.91</td>
<td>0.48</td>
</tr>
<tr>
<td>All</td>
<td>1.02</td>
<td>0.42</td>
</tr>
</tbody>
</table>

### Table 2. Theoretical PTV margin for upper and lower lobe tumor with a motion amplitude (peak-peak) of 0.2 cm and 1.2 cm, respectively, and delineation based on CT alone and matched FDG-PET – CT, per anatomical region.

<table>
<thead>
<tr>
<th>Anatomical regions</th>
<th>Upper lobe tumor</th>
<th>Lower lobe tumor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tumor – lung</td>
<td>1.4</td>
<td>0.8</td>
</tr>
<tr>
<td>Tumor – mediastinum</td>
<td>1.8</td>
<td>1.1</td>
</tr>
<tr>
<td>Tumor – chest wall</td>
<td>1.0</td>
<td>0.9</td>
</tr>
<tr>
<td>Lymph nodes</td>
<td>3.6</td>
<td>2.0</td>
</tr>
<tr>
<td>Tumor – atelectasis</td>
<td>4.7</td>
<td>1.2</td>
</tr>
<tr>
<td>All</td>
<td>2.5</td>
<td>1.0</td>
</tr>
</tbody>
</table>

* based on margin recipe described by van Herk et al [12,13] (the margin around the GTV should be 2.5 times the SD of the systematic error plus 0.7 times the SD of the random error minus 0.3 cm), incorporating:
* setup variation (systematic error: SD = 0.14 cm, random error: SD = 0.27 cm) [16]
* tumor motion (systematic and random error: SD = 0.33 x peak-peak amplitude) [13, 17]

### Table 3. Response, median survival and 2-year survival data of all patients.

<table>
<thead>
<tr>
<th>Response</th>
<th>Median survival (months)</th>
<th>2-year Survival (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complete (n = 17)</td>
<td>38</td>
<td>50</td>
</tr>
<tr>
<td>Partial (n = 25)</td>
<td>14</td>
<td>26</td>
</tr>
<tr>
<td>Stable / progression (n = 9)</td>
<td>10</td>
<td>33</td>
</tr>
<tr>
<td>All (n = 51)</td>
<td>17</td>
<td>43</td>
</tr>
</tbody>
</table>
Table 2 lists the required Planning Target Volume (PTV) margins incorporating setup variation, organ motion and variation in delineation for an upper and lower lobe tumor. The PTV margin needed using the delineation based on the matched CT – FDG-PET is much smaller than that based on CT alone.

**SUV study**

The overall median follow-up was 17 months (range: 3 to 57 months). Seventeen patients (33%) experienced a complete response with a median survival of 38 months. Twenty-five patients (50%) had a partial response with a median survival of 14 months. Patients with stable disease (8%) and progressive disease (10%) had a median survival of 10 and 9 months, respectively (Table 3). Only one patient in the complete response group died due to lung cancer, compared with 14 patients in the partial response group.

The median SUV\textsubscript{max} was 15 (mean = 17, SD = 9, range: 3 to 41). No statistical correlation was observed between tumor volume and SUV\textsubscript{max} (\(r = 0.17, p = 0.2\)), or between lymph node status and SUV\textsubscript{max} (\(r = 0.15, p = 0.3\)). Stage I and II patients had a mean SUV\textsubscript{max} of 14. Stage III patients had a mean SUV\textsubscript{max} of 16.

When a median SUV\textsubscript{max} of 15 was used, a sensitivity of 77 % and specificity 84 % in predicting complete response could be achieved. Using univariate analysis, stage (\(p = 0.02\)), performance status (\(p = 0.04\)) and positive lymph nodes (\(p = 0.04\)) were significant factors correlated with complete response. In contrast, chemotherapy, dose, tumor volume, weight loss and histology were not significant.

To assess whether SUV\textsubscript{max} was an independent predictive factor for treatment response, logistic regression analysis was used. Stage (\(p = 0.02\)), performance status (\(p = 0.01\)) and SUV\textsubscript{max} (\(p = 0.05\)) were independently associated with complete response.

For the disease-specific survival (DSS) and overall survival (OS), Kaplan Meier plots were generated (Figure 3). Large survival differences were observed when patients were separated into groups with SUV\textsubscript{max} < 15 (n = 26) and SUV\textsubscript{max} \(\geq\) 15 (n = 25). The median DSS for the patients with a SUV\textsubscript{max} < 15 was not reached and for patients with a SUV\textsubscript{max} \(\geq\) 15 this was 12 months (\(p < 0.001\)). The median OS was 38 months and 12 months for the patients a SUV\textsubscript{max} < 15 and SUV\textsubscript{max} \(\geq\) 15, respectively (\(p < 0.001\)). At two years the DSS and OS for the patients with an
As with the DSS, a one-unit increase of SUV\(_{\text{max}}\) status (p = 0.06) remained significant factors. For OS, SUV\(_{\text{max}}\) (p = 0.001) and performance status were 1.6 and 3.86, respectively. The Cox proportional hazards model was used to analyze variables to predict DSS and OS. For DSS, SUV\(_{\text{max}}\) (p = 0.01), performance status (p = 0.008) and stage (p = 0.04) remained significant. The hazard ratio (HR) of SUV\(_{\text{max}}\) was 1.06, which indicated that a one-unit increase of SUV\(_{\text{max}}\) corresponds to a 6% increase of hazard of lung cancer related death. The HRs for stage and performance status were 1.6 and 3.86, respectively. For OS, SUV\(_{\text{max}}\) (p = 0.001) and performance status (p = 0.06) remained significant factors. As with the DSS, a one-unit increase of SUV\(_{\text{max}}\) corresponded with a 6% increase of hazard of death due to any cause.

### Discussion

In this paper we have demonstrated that implementing FDG-PET in the radiotherapy treatment planning has a major impact on accurate target definition and the prediction of treatment outcome.

For the study concerning target definition, high-precision radiotherapy based on target definition using CT alone appears to be unacceptable, due to very large variability in tumor delineation. The PTV margin required to correct for this geometrical uncertainty becomes too large (Table 2). The two other main geometrical uncertainties, i.e., setup variation and breathing motion, are usually much smaller. In a modern radiotherapy department using at least a portal imaging device (EPID), setup variation (systematic and random) is limited to about 1 to 3 mm (1 SD) [16]. For a lower lobe lung tumor, the breathing motion can reach an amplitude of 2.0 cm [17]. About one third of this amplitude translates into a systematic or random error [13]. Furthermore, with the use of modern techniques, such as respiratory correlated CT for planning and respiratory correlated cone beam CT for verification, the systematic error of breathing motion can be significantly reduced [18]. Variation in delineation is by definition a systematic error, so that reduction of delineation variation is warranted.

In our study we showed that a large reduction of delineation variation can be achieved by using matched FDG-PET CT. This leads to a significantly reduction in the PTV margins around the GTV required to correct for the main geometrical uncertainties (Table 2). Consequently, a safe dose escalation can be achieved with avoidance of geographical misses. This might lead to a higher cure rate and reduction of toxicity, because the radiation dose to healthy normal lung tissue can be reduced significantly. Nevertheless, the variability in delineation is still large compared to setup variation and organ motion. Therefore improvement of delineation accuracy is warranted.

As we have seen in three patients, tumor can be discriminated from atelectasis with matched FDG-PET CT. Unfortunately, some patients also had increased FDG uptake in the atelectasis, probably due to inflammation. For these patients it is impossible to distinguish between inflammation and tumor with FDG-PET. Perhaps in the future more specific tracers such as 11C-tyrosine or 11C-choline might solve this issue. For the time being, we recommend including the increased FDG uptake in atelectasis in the target volume in order to remain on the safe side. During the course of the treatment, a new CT (and PET) scan can be made, to see whether the atelectasis and/or inflammation disappears. In such cases the treatment plan might be adjusted accordingly.

With FDG-PET the presence of pathological lymph nodes can be easily seen, but our study also showed that, even with matched CT, estimation of the extent of these lymph nodes remained difficult. This was probably caused by the fact that no contrast enhancement was used for the CT scan. With the use of contrast enhancement, lymph nodes and tumor in or near the mediastinum can be better discriminated from blood vessels.

Recently, a project has been started at our institute to correlate CT and FDG-PET images with operated lung tumors [19]. In this study, patients with operable lung cancer have a planning CT scan and an FDG-PET just before surgery (< 1 week). After operation the surgical specimen is matched (including warping) with the CT and FDG-PET. In this way GTV defined
from the CT and FDG-PET scans can be correlated with the pathology. From some preliminary results can be concluded that the volumes defined by the different modalities can be quite different. Furthermore, with some patients a serious underestimation of the GTV extent was found [19].

For the FDG SUV study, the FDG SUV$_{\text{max}}$ was predictive for treatment response, and the multivariate survival analysis showed that SUV$_{\text{max}}$ was an explanatory prognostic factor for both disease-specific (DSS) and overall (OS) survival.

Most previous studies concerning the prognostic value of SUV were performed using operable patients [9, 10] with better survival rates than those in our study. It is therefore difficult to compare these studies with ours. Only one other study [20] evaluated the prognostic value of SUV$_{\text{max}}$ for radiotherapy patients. They found that a cut-off value of 5 (median SUV$_{\text{max}}$ was 8) provided the most significant survival difference. Their radiotherapy patients had a high 2-year OS of 71 %, which was not different from the 2-year OS of surgically treated patients.

This might explain why we found a higher median SUV$_{\text{max}}$ accompanied with a lower 2-year OS.

Another problem in comparing the prognostic value of SUV between different studies and institutions is the difference in PET scanning techniques. Differences in injected FDG-dose, scanning time (i.e. time after injection), reconstruction algorithms, filters, scanner characteristics, sinogram noise and quantification methods might lead to (structural) inter-institutional SUV differences [21].

In this study we demonstrated that besides the traditional prognostic factors, such as performance status, stage and weight loss, SUV adds predictive and prognostic information to the treatment outcome of inoperable NSCLC patients. Measurement of SUV could therefore help to make a better selection of patients who might benefit from more aggressive therapy such as concurrent chemo-radiotherapy and dose escalation, or those who would be better be treated in a more palliative setting. However, in order to really use SUV for selecting patients for a certain therapy, confirmation of our results by prospective clinical data is required.

Conclusions:
Implementing FDG-PET in radiotherapy treatment planning has a major impact on accurate target definition and the prediction of treatment outcome. For high precision radiotherapy, target definition based on CT alone results in too great a variability between radiation oncologists. With the addition of FDG-PET, radiotherapy can be delivered more accurately, allowing safe dose escalation. The SUV measurement from FDG-PET is a strong predictor of prognosis and treatment outcome. It may offer a better selection of those patients who might benefit from more aggressive therapy.
References


