Fluorodeoxyglucose positron emission tomography-computerized tomography and lung cancer: a significant referral bias exists

Ikenna C. Okereke, Sidhu P. Gangadharan, Michael S. Kent, Saila P. Nicotera, Malcolm M. DeCamp

Department of Cardiothoracic Surgery, Indiana University Medical Center, Indianapolis, IN, USA
Division of Cardiothoracic Surgery, Beth Israel Deaconess Medical Center, Boston, MA, USA

Received 2 February 2010; received in revised form 5 July 2010; accepted 12 July 2010; Available online 15 September 2010

Abstract

Objective: 18F]Fluorodeoxyglucose positron emission tomography-computerized tomography (PET-CT) scan is a tool widely used in the diagnosis and staging of lung cancer. Referral bias is present when the results of a diagnostic study affect the decision to proceed with definitive testing. This bias artificially increases the sensitivity and decreases the specificity, and may inappropriately alter the decision to undergo definitive testing. The accuracy of PET-CT scan in suspected lung cancer and the role of referral bias were investigated.

Methods: From January 2005 through June 2007, 584 consecutive patients undergoing PET-CT scan for suspected lung cancer were studied. Endpoints measured included qualitative and quantitative results of PET-CT scans and pathologic results from patients, who underwent invasive procedures for diagnosis, staging and/or therapy. A positive PET scan was defined as one in which the standard uptake value (SUV) was greater than 2.5. A standard mathematical model, based on overall results of PET-CT scan in all patients, was used to create adjustments to account for the effect of referral bias.

Results: A total of 414 (71%) of PET-CT scans were reported as positive, while 170 (29%) were negative. Attempt at tissue diagnosis occurred in 417 patients (71%); surgical intervention was performed in 246 (42%). Whereas 86% (355/414) of patients with a positive PET-CT scan underwent tissue sampling, only 36% (62/170) with a negative PET-CT scan had an attempt (p < .001). In patients with a positive study, a higher SUV was associated with an attempt at tissue diagnosis (p < .001). Before adjustment, the sensitivity and specificity of the PET-CT scan for any cancer were 87% and 43%. After adjustment, sensitivity fell to 75%, while specificity rose to 64%. The unadjusted sensitivity of PET-CT scan for detecting mediastinal disease was only 79%.

Conclusions: The estimate of the accuracy of PET-CT scan is significantly influenced by referral bias, and a strong reliance is placed on its results. Furthermore, patients with a positive PET-CT scan are more likely to undergo tissue sampling as the SUV increases. Given the relatively low sensitivity of the PET-CT scan in detecting disease and its poor performance in the mediastinum, the PET-CT scan may have too high an impact on the decision to undergo further invasive diagnostic procedures. Patients should not be deferred from tissue sampling based solely on a negative PET-CT scan.

Keywords: Lung cancer; Oncology; Diagnostic accuracy

1. Introduction

Lung cancer is the most lethal malignancy in the USA. Approximately 160 000 people die each year from lung cancer. As such, the ability to diagnose and stage lung cancer accurately is critical for public health. 18F]Fluorodeoxyglucose (FDG) positron emission tomography-computerized tomography (PET-CT) scan, which measures the amount and trapping of radio-labeled glucose uptake by tissues, has become widely used as part of the evaluation in patients with suspected lung cancer.

Previous studies have shown PET-CT scan to be quite proficient in the detection of lung cancer [1]. These studies, however, focused only on patients who underwent surgery, rather than all patients who underwent a PET-CT scan. This type of study design creates an inherent referral bias, which may significantly alter the results of the study.

Referral bias occurs when the results of a diagnostic study affect the decision to proceed with definitive, or gold standard, testing. Because this tendency underestimates both the true negatives and the false negatives of a study, sensitivity will be artificially elevated, while specificity will be falsely lowered.

The effect of referral bias has been examined in other studies. Prostate-specific antigen (PSA) screening was seen to be relatively ineffective in diagnosing prostate cancer, after adjusting for referral bias [2]. Similarly, adjustments...
for this bias led to the determination that exercise echocardiography for the evaluation of cardiac disease is not as sensitive as previously thought [3]. Further, a study, which looked at stand-alone PET scan for lung cancer, concluded that accuracy was significantly overestimated due to referral bias [4].

Previous studies have not examined the effect of referral bias on integrated PET-CT scan in the evaluation of patients suspected to have lung cancer. In the evaluation of a patient with suspected lung cancer, many factors, such as age, smoking history, and overall health, are included in the overall decision to proceed with definitive testing. Our goals were to determine the impact that PET-CT results have on this overall decision, to determine the effect of referral bias related to PET-CT results, and to establish the true accuracy of integrated PET-CT scanning.

2. Materials and methods

2.1. Patients

From January 2005 through June 2007, all 584 patients, who underwent an integrated PET-CT scan as part of the work-up of suspected lung cancer at Beth Israel Deaconess Medical Center, were included in the study. All patients had their imaging and subsequent care at the Beth Israel Deaconess Medical Center. Patients, who were proven to have small-cell lung cancer or lymphoma, were excluded from the study. Patients were also excluded from the study if they had proven metastatic disease to the chest from an extrathoracic site. Patient demographics, such as age, gender, performance status, and co-morbidities were recorded. Patients, who ultimately underwent mediastinoscopy and/or surgery, were used to create the unadjusted results. In addition, patients, who had at least some attempt at tissue diagnosis, were also recorded. The remaining patients, who had PET-CT scan without further intervention, were used to create the adjustment for referral bias. Adjustments were created by using both the Diamond method [5,6] and the Begg and Greenes method [7], validated techniques of adjustment previously described in other studies.

2.2. Outcomes

The results of the integrated PET-CT scan, including standard uptake value (SUV) measure, mass size, presence of lymphadenopathy, and location of positivity, were recorded. Similarly, the clinical and pathologic stage was determined, using standard tumor-node-metastases (TNM) classifications [8], in patients who underwent surgery. Survival data were obtained through electronic medical records and verified using the Social Security Death Index.

2.3. PET-CT scan

All patients fasted for more than 4 h before the scan. Blood glucose levels were determined before administration of 10 Ci 18FDG. One hour after administration of 18FDG, PET and CT scans were obtained from the skull base to the level of the hips. All integrated PET-CT scans were reviewed by radiologists, who specialized in nuclear medicine techniques and were blinded to pathologic results and patient outcomes. All PET-CT scans were performed at the Beth Israel Deaconess Medical Center, using the single PET-CT scanner at our institution. Mass size, presence of lymphadenopathy by CT criteria (>1 cm in short axis dimension), SUV levels of the primary mass, lymph nodes, and satellite lesions were recorded. A PET-CT scan was interpreted as positive if the maximum SUV of the primary lesion or locoregional lymph nodes exceeded 2.5 [9].

2.4. Statistical analysis

Continuous variables are summarized by mean and standard deviations and categorical variables are summarized by percentage. Two-sample T test was used to compare continuous variables and the chi-square test was used to compare categorical variables. The Kaplan–Meier method was used to estimate the survival curves for time to death since examination among subjects with different PET-CT stages. The difference was compared using Log-rank chi-square test. Cox proportional hazard model was used to assess the effect of mass SUV on survival.

Unadjusted sensitivity and specificity were calculated based on PET-CT results among those whose pathological results (gold standard) are available. We also tried to adjust the potential referral bias in the unadjusted estimates by employing two methods, the Diamond method and the Begg and Greenes method. The key assumption of these two ‘adjusted’ methods is that conditional on the test result (i.e., PET-CT in our case) and possibly other measured characteristics, the decision on obtaining a formal diagnosis (gold standard) does not depend on the disease status. This assumption is important because the probability of being diseased will be estimated for the entire samples, with or without tissue sampling, based a logistic model obtained from the subgroup of samples that underwent tissue sampling. Begg and Greenes incorporated both test result and other characteristics to correct for both the post- and pre-test referral bias. The Diamond method is a simplification of this method by only using the test result for the correction of referral bias, and only corrects for the post-test referral bias. From the theoretical standpoint, the Begg and Greenes estimate is likely to be less biased than the Diamond approach if the decision of obtaining formal diagnosis does depend on some measured characteristics of the patients. In our case, age seems to be relevant on the decision of tissue attempt (p-value < 0.05 in a logistic regression model with tissue attempt as the outcome and sex, smoking status and PET-CT stage as the other covariates). Therefore, we consider Begg and Greenes a more reliable method for the correction of bias. Because the Begg and Greenes method requires estimates of the probability of being diseased for those in whom no formal diagnosis has been made, we fit a logistic regression model to subjects with formal diagnosis with age, sex, PET-CT stage, and smoking status as the covariates, and the estimated model was used to calculate the probabilities for those without formal diagnosis. Finally, we generated 1000 bootstrap samples to obtain the 95% confidence intervals of the accuracy measure of the PET-CT test. All statistical analyses were carried out in SAS 9.1 (SAS Institute, Inc., Cary, NC, USA).
3. Results

3.1. Demographics

A total of 584 patients were included in the study. Table 1 lists the demographics. The average age was 67 years. Some previous smoking history was reported in 83% of patients.

3.2. Tissue diagnosis

As many as 417 of the 584 patients underwent at least some attempt at tissue diagnosis. Procedures performed to obtain tissue included percutaneous biopsy, endobronchial biopsy, thoracentesis, mediastinoscopy, and surgical resection.

3.3. PET-CT scan

PET-CT scan results are listed in Table 2. In the entire study, 197 of the 584 patients (34%) were determined to have malignant disease. As many as 36% of patients were assigned a PET stage of III or higher. PET-CT scan was positive in the mediastinum in 28% of patients.

The unadjusted sensitivity and specificity of PET-CT scan was 87% and 43%, respectively (Table 3). After correction for referral bias by the Diamond method, the sensitivity fell to 75% and the specificity rose to 64%. The Begg and Greenes method led to a sensitivity of 80% and a specificity of 58%. Patients with a positive PET-CT scan were more likely to be recommended for further tissue attempt than patients with a negative PET-CT scan. This referral bias resulted in an underrepresentation of false negative, which inflated sensitivity.

3.4. Referral bias, PET-CT scan

In patients who were PET Stage 0, only 62 of 170 patients (36%) underwent some attempt at tissue diagnosis. Conversely, in patients who were PET Stage I and higher, 355 of 414 patients (86%) underwent an attempt at tissue procurement ($p < 0.001$).

3.5. Referral bias, SUV

Among patients with a positive PET-CT scan, those who underwent an attempt at tissue procurement ($n = 355$) had a

### Table 1. Demographics.

<table>
<thead>
<tr>
<th></th>
<th>All</th>
<th>With tissue attempt</th>
<th>Without tissue attempt</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patients</td>
<td>584</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age, mean</td>
<td>67.0 ± 11.9</td>
<td>66.7 ± 11.7</td>
<td>67.4 ± 12.4</td>
<td>0.54</td>
</tr>
<tr>
<td>Male</td>
<td>47%</td>
<td>46%</td>
<td>50%</td>
<td>0.38</td>
</tr>
<tr>
<td>ECOG 2</td>
<td>16%</td>
<td>17%</td>
<td>13%</td>
<td>0.26</td>
</tr>
<tr>
<td>Smoking, current</td>
<td>21%</td>
<td>22%</td>
<td>17%</td>
<td>0.04</td>
</tr>
<tr>
<td>Smoking, past</td>
<td>62%</td>
<td>63%</td>
<td>60%</td>
<td>0.7</td>
</tr>
<tr>
<td>Pk-yrs</td>
<td>53.9 ± 28.4</td>
<td>54.4 ± 29.3</td>
<td>51.9 ± 24.9</td>
<td>0.7</td>
</tr>
<tr>
<td>Smoking, never</td>
<td>17%</td>
<td>15%</td>
<td>17%</td>
<td></td>
</tr>
</tbody>
</table>

* Continuous variables are summarized by mean ± standard deviation. Categorical variables are summarized by percentages.

** $p$-value for the hypothesis that the proportions of current smoking, past smoking and never smoking are the same between group with tissue attempt and group without tissue attempt.

### Table 2. PET-CT scan results.

<table>
<thead>
<tr>
<th></th>
<th>All</th>
<th>With tissue attempt</th>
<th>Without tissue attempt</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>PET Stage 0</td>
<td>29%</td>
<td>15%</td>
<td>65%</td>
<td></td>
</tr>
<tr>
<td>Mass size, cm</td>
<td>1.3 (0.7)</td>
<td>1.4 (0.9)</td>
<td>1.2 (0.7)</td>
<td>0.03</td>
</tr>
<tr>
<td>Attempt at tissue procurement</td>
<td>36%</td>
<td>34%</td>
<td>17%</td>
<td></td>
</tr>
<tr>
<td>PET Stage I</td>
<td>29%</td>
<td>34%</td>
<td>17%</td>
<td></td>
</tr>
<tr>
<td>Mass size, cm</td>
<td>2.0 (1.4)</td>
<td>2.2 (1.3)</td>
<td>1.6 (1.4)</td>
<td>0.002</td>
</tr>
<tr>
<td>Mass SUV</td>
<td>5.3 (6.5)</td>
<td>6.2 (6.6)</td>
<td>3.0 (1.7)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Attempt at tissue procurement</td>
<td>83%</td>
<td>7%</td>
<td>3%</td>
<td></td>
</tr>
<tr>
<td>PET Stage II</td>
<td>6%</td>
<td>7%</td>
<td>3%</td>
<td></td>
</tr>
<tr>
<td>Mass size, cm</td>
<td>2.4 (1.9)</td>
<td>2.6 (1.8)</td>
<td>1.5 (0.5)</td>
<td>0.05</td>
</tr>
<tr>
<td>Mass SUV</td>
<td>6.7 (10.1)</td>
<td>9.7 (11.4)</td>
<td>3.8 (0.2)</td>
<td>0.09</td>
</tr>
<tr>
<td>Attempt at tissue procurement</td>
<td>86%</td>
<td>25%</td>
<td>10%</td>
<td></td>
</tr>
<tr>
<td>PET Stage III</td>
<td>21%</td>
<td>25%</td>
<td>10%</td>
<td></td>
</tr>
<tr>
<td>Mass size, cm</td>
<td>2.5 (2.8)</td>
<td>2.7 (3.2)</td>
<td>1.8 (1.0)</td>
<td>0.002</td>
</tr>
<tr>
<td>Mass SUV</td>
<td>11.2 (9.6)</td>
<td>11.3 (9.2)</td>
<td>6.5 (7.2)</td>
<td>0.24</td>
</tr>
<tr>
<td>Attempt at tissue procurement</td>
<td>86%</td>
<td>19%</td>
<td>5%</td>
<td></td>
</tr>
<tr>
<td>PET Stage IV</td>
<td>15%</td>
<td>19%</td>
<td>5%</td>
<td></td>
</tr>
<tr>
<td>Mass size, cm</td>
<td>3.1 (3.0)</td>
<td>3.2 (2.9)</td>
<td>1.7 (5.0)</td>
<td>0.47</td>
</tr>
<tr>
<td>Mass SUV</td>
<td>10.3 (10.4)</td>
<td>10.8 (9.5)</td>
<td>5.5 (9.4)</td>
<td>0.64</td>
</tr>
<tr>
<td>Attempt at tissue procurement</td>
<td>91%</td>
<td>35%</td>
<td>11%</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

Continuous variables are summarized by median (IQR). Categorical variables are summarized by percentages.

Wilcoxon rank sum test was used to compare the mass size and mass SUV between patients with and without tissue attempt for different PET stages.

* $p$-value is for the test that the proportions of various PET stages are the same for the group with and without tissue attempt.
median SUV of the mass of 8.7, with an interquartile range (IQR) of 9.8. In the same group of patients with a positive PET-CT, however, the median SUV of the mass in those who did not undergo attempt at tissue procurement \((n = 59)\) was only 4.2 (IQR = 3.5) \((p < 0.001,\) Wilcoxon rank sum test).

### 3.6. Referral bias, mass size

Among patients with a positive PET-CT scan, those who underwent an attempt at tissue diagnosis had a median mass size of 2.5 cm (IQR = 2.5). In the same group of patients with a positive PET-CT scan, conversely, the median mass size in those who did not undergo attempt at tissue procurement was 1.6 cm (IQR = 1.2, \(p < 0.001,\) Wilcoxon rank sum test).

### 3.7. Survival

The mean follow-up time since examination was 1.23 years, during which 168 (29%) patients died. Fig. 1 shows survival curves stratified by PET stage. There was a statistically significantly correlation between PET stage and survival \((p < 0.001)\). Furthermore, among patients with a positive PET-CT scan, increasing SUV was associated with increased mortality (hazard ratio 1.05, 95% confidence interval (CI): 1.03–1.07).

In this study, 15 patients, who were initially PET Stage 0, were found to have died. Of those, seven were ultimately confirmed to have died of lung cancer.

### 4. Discussion

This study was developed as a means to determine the accuracy of integrated PET-CT scan in detecting and staging lung cancer. Though the unadjusted sensitivity was 87%, this value fell to 80% when the effects of referral bias were included. This trend matches the expected effects of referral bias and those seen in previous studies [10]. In determining the true accuracy of any diagnostic modality, it is necessary to include the impact of referral bias.

Our study is limited in that it is a retrospective study, and referrals toward definitive surgery may be subject to biases not captured in the analysis. Nevertheless, we feel that the size of the cohort and the disparity in results based on PET-CT scan is enough to make certain conclusions.

From our results, it is clear that PET-CT scan by itself is an inadequate non-invasive diagnostic tool. Yet, it is our belief that some use PET-CT scan in this manner, with an over-reliance on its positivity in determining whether a patient is referred for definitive testing. The fact that there exists such a dramatic disparity in referral rates of 86% versus 36%, based upon PET-CT results, supports this contention. Though it is expected that patients with a positive PET-CT scan would have an increased rate of referral, the observed difference is excessive. PET-CT scan is not a replacement for definitive testing in appropriate clinical scenarios [11].

A previous study evaluating the role of PET scan in lung cancer looked primarily at whether the PET scan was positive in determining referral patterns [4]. The present study allowed us to examine other aspects of integrated PET-CT scan. In this study, it was seen that as the size of the mass increased, a patient was more likely to be referred for definitive testing. Though mass size is part of the TNM classification, caution must be taken not to ignore smaller lesions in patients with appropriate risk factors. Of those who underwent attempt at tissue procurement, 10% had primary lesions less than 1 cm. As prior literature has shown that PET scan is less sensitive for lesions less than 8–10 mm [12], it is debatable whether this represents an appropriate percentage of ‘small’ lesions in our cohort, or whether more subcentimeter lesions should have been referred.

Increasing SUV was associated with decreased survival in this study. This was true both in all patients and in patients with a positive PET-CT scan. An increase in SUV of 1 was associated with a 5% increased likelihood of death. Similarly, other studies have demonstrated decreased survival with higher SUV values [13,14].

We saw a referral bias associated with the SUV of the primary lesion. Among those patients with a positive PET-CT scan, as the SUV increased, a patient was more likely to be referred for tissue procurement. Though SUV appears to be associated with worse survival, it should be noted that many lesions, especially carcinoid and bronchoalveolar carcinoma,
may be associated with lower SUV levels [15]. In fact, almost 11% of patients in our study with tissue-proven lung cancer had SUV levels below 4. As such, it is important not to decline a patient for referral solely based on a ‘marginal’ SUV level.

PET-CT scan will remain an important tool in the management of lung cancer. Its prognostic utility, as detailed above, is useful in both counseling and treating patients. Its ability to detect distant disease can help to stage patients appropriately and spare unnecessary morbidity. PET-CT scan is not, however, a good stand-alone diagnostic tool and must be used with other patient characteristics in determining whether to refer a patient for tissue procurement.

References