Prognostic value of FDG uptake in early stage non-small cell lung cancer

François-Xavier Hanin a, Max Lonneux a, Julien Cornet b, Philippe Noirhomme b, Corinne Coulon b, Julien Distexhe a, Alain J. Poncelet b,*

aDepartment of Nuclear Medicine, U.C.L. Saint-Luc Hospital, Brussels, Belgium
bDepartment of Cardiovascular and Thoracic Surgery, U.C.L. Saint-Luc Hospital, Brussels, Belgium

Received 30 August 2007; received in revised form 25 January 2008; accepted 1 February 2008

Abstract

Background: Non-small cell lung cancer (NSCLC) has a poor prognosis even for early stages of the disease (stage I and II). We studied the prognostic value of PET FDG in patients with completely resected stage I and II NSCLC. Methods: Retrospective study of 96 patients with NSCLC whose staging included 18F-FDG PET (fluoro deoxy glucose positron emission tomography). Histopathological stage was either stage I (75) or stage II (n = 21). FDG uptake was measured as maximal standardized uptake value for body weight (SUVmax). Mean follow-up was 45 ± 30 months (1–142 months). Overall and cancer-free survival rates were recorded. Results: SUVmax were higher for stage II than for stage I (10.5 ± 4.5 vs 8.5 ± 5, p = 0.04). Mean tumor volumes were equivalent for both stages (33 cm³, p = 0.18), excluding a partial volume effect. The median SUVmax in the whole study population was 7.8. The median survival was significantly longer in patients with a lower (SUVmax ≤ 7.8) FDG uptake (127 months vs 69 months, p = 0.001). For stage I tumors (n = 75), high FDG uptake was significantly associated with reduced median survival: 127 months if SUVmax ≤ 7.8 and 69 months if SUVmax > 7.8 (p = 0.001). For stage II tumors (n = 21), no statistical difference was observed: 72 months vs 40 months for SUVmax ≤ 7.8 and for SUVmax > 7.8, respectively (p = 0.11), although there was a clear trend towards reduced survival for highly metabolic tumors. Disease-free survival was also significantly better for lower metabolic tumors: 96.1 months vs 87.7 months (p = 0.01). Conclusion: High FDG uptake is associated with reduced overall survival and disease-free survival of patients with completely resected stage I–II NSCLC. Whether patients with highly metabolic tumors should undergo a closer postoperative surveillance or adjuvant chemotherapy has to be addressed in a properly designed prospective trial.

Keywords: Non-small cell lung cancer; PET FDG; SUV; Prognosis; Thoracic surgery

1. Introduction

Non-small cell lung cancer (NSCLC) staging represents a key part in the algorithm of cancer management, where down- and up-staging can represent higher mortality and morbidity [1]. In this field, FDG PET has shown its ability to accurately detect mediastinal lymph nodes and metastases [2,3], its cost-effectiveness [4], its performance in evaluating the response to therapy [5] and its performance in detecting recurrent disease [6]. FDG PET is now included in NSCLC treatment guidelines of NSCLC [7].

Recent data available in early stage (stage I and II) NSCLC patients arises from the JBR.10 trial where 5-year survival rates of 69% and 54% were reported, whether or not cisplatin-based adjuvant chemotherapy was given. Analysis of stage Ib and II subgroups revealed that only stage II patients benefited from adjuvant treatment and those results have set the current standard of practice in early stage lung cancer patients. Despite all current efforts, there still remains a significant number of patients who will relapse in those completely resected early stage (stage I and II) NSCLC that drives us to pursue efforts in order to better understand tumor behavior and identify additional prognostic factors.

Moreover, several data suggest that FDG PET also provides a prognostic information: highly metabolic tumors (as measured by the semi-quantitative index standardized uptake value (SUV)) have a more aggressive evolution. This has been shown in a large cohort representing all stages [8], as well as surgically treated patients [9–11], before neo-adjuvant treatment [12] or in inoperable stages [13].

For those early stage patients, if proven accurate, such preoperative clinical and biological sub-staging could help to identify groups of patients most at risk for locoregional recurrence or metastasis, and therefore might guide clinicians in using adjuvant chemotherapy.
The purpose of this single-center study was to assess the prognostic significance of preoperative FDG PET SUVmax in a very selected subgroup of patients with pathologically proven stage I or II NSCLC, with respect to cancer recurrence and overall survival.

2. Materials and methods

2.1. Patients characteristics

The database of our cardiothoracic surgery department was searched for all patients operated on for a NSCLC stage I and II, who had an FDG PET performed in our center as part of the preoperative workup within one month of resection. Attenuation-corrected PET was mandatory in order to be able to calculate the SUV. Ninety-six patients were eligible for analysis: 75 patients with stage I and 21 patients with stage II, according to the last revision of the international system for staging lung cancer [14]. There were 23 females and 73 males included in the study. Their clinico-pathological data are summarized in Table 1.

Before surgery, all patients had a standard staging workup including fibroscopy, chest and abdominal CT-scan, brain MRI or CT, and FDG PET. When appropriate, additional diagnostic procedures were performed to assess the mediastinum (EUS, mediastinoscopy or EBUS).

The Committee on Human Rights in Research (Institutional Review Board) of Cliniques Universitaires Saint-Luc approved this study.

2.2. PET FDG staging preoperatively

FDG PET images were obtained 60 min after tracer injection (300–400 MBq FDG), on an ECAT HR+ PET system (Siemens, Erlangen, Germany). Whole-body images were collected and reconstructed using an iterative algorithm, with correction for attenuation [15]. Patients were fasted 6 h prior to injection.

A region of interest (ROI) was drawn over the primary tumor, with the help of the CT scan for localization when necessary. Maximal standardized uptake value (SUVmax, corrected for body weight), i.e. the pixel with the maximal uptake within the ROI, was calculated for each tumor. The median SUVmax was used as cut-off threshold for defining groups of patients.

2.3. Surgical strategies and R (residual) status

The surgical procedures consisted of 78 lobectomies (81.2%), six bilobectomies (6.2%), 11 pneumonectomies (11.4%). Segmentectomy was performed in one patient. Complete homolateral mediastinal lymph node dissection was performed in all cases and node stations were labeled according to the American Thoracic Society guidelines [16]. There were 1596 lymph nodes available for pathological examination with a median number of 16 resected lymph nodes per patient (range 3–37).

Resection was classified R0 (macro- and microscopically complete) in all patients. No patient was classified R1 (microscopically incomplete) or R2 (macroscopically incomplete) [17].

2.4. Pathologic evaluation

Tumor was evaluated in all 96 patients according to the World Health Organization classification for NSCLC [18] both for histology and grade by one of the authors (BW). After localization and size measurement, the specimens were planed with a cryostat, serially sectioned (3–4 mm) and embedded, and then stained by standard hematoxylin and eosin. All resected lymph nodes were fixed either with formalin (later period) or Bouin (earlier period).

Histology revealed squamous cell carcinoma in 47 patients (49%), adenocarcinoma in 46 patients (48%) and undifferentiated carcinoma in three patients (3.1%). According to the grade of differentiation, there were 26 well differentiated (27%), 51 moderately (53.1%) and 19 poorly differentiated tumors (19.8%).

2.5. Selection of predictor variables

Predictor variables studied for their potential impact on prognosis were age, sex, the pathology of tumor, the grade of differentiation, T status, N status, disease’s pathological stage, tumor diameter in T2 tumors (less or greater than 5 cm) and preoperative SUVmax.

2.6. Data collection and patients follow-up

Follow-up was performed by an oncologist or pulmonary physician at 3, 6, 9, 12, 15, 18 months in outpatient clinic, then every 6 months until 60 months, then yearly. Standard follow-up consisted of chest X-ray and clinical examination, chest CT-scan being only performed in case of suspicious radiological or clinical finding.
All the information necessary for the study was collected from the patient operative reports, the hospitalization charts and our thoracic surgery database. Follow-up was completed from September 2006 to January 2007 and was done through contacts with the referring pulmonary physician, the primary care physician, patient’s family when appropriate, and cross-checked with the national registry database. Follow-up was 100% complete (96 patients) until closure of the study as of 31 January 2007. Forty patients had died and the cause of death was determined in all patients.

2.7. Statistical analysis

The association of preoperative SUVmax with clinico-pathological factors was analyzed using the two-tailed Pearson’s chi-square or Fisher’s exact test as appropriate. Differences in groups for SUVmax values were tested using one-way ANOVA, nonparametric ANOVA or Mann–Whitney test as appropriate. Survival from the date of operation was calculated using the Kaplan–Meier survival analysis method. Differences in observed survival between groups were tested for significance using the log rank test [19]. Differences were considered significant when the p value was less than 0.05. For patients presenting recurrences, the number of days was calculated from the date of pulmonary resection to the first documentation of either locoregional or distant recurrence.

Statistical analysis and Kaplan–Meier curves were performed with GraphPad Prism and GraphPad Instat software (GraphPad Software, Inc., San Diego, USA).

3. Results

3.1. Standardized uptake value (SUV) and clinico-pathological variables

The mean SUVmax was lower for adenocarcinoma (7.2 ± 4.1), than for squamous cell carcinoma (10.4 ± 5.5) whereas undifferentiated carcinoma had the highest FDG uptake (12.9 ± 1.8) (Nonparametric ANOVA, p = 0.0014). The mean SUVmax was (7.4 ± 4.8) for T1, (9.9 ± 5.2) for T2 and (12 ± 3.2) for T3 lesions, respectively (nonparametric ANOVA, p = 0.011). Mean SUVmax was significantly lower for well-differentiated tumors (7.0 ± 5.3) than for moderately/poorly differentiated ones (9.6 ± 5.2 and 9.9 ± 3.6) that we analyzed together since they had similar SUVmax (non-parametric ANOVA, p = 0.011).

SUVmax were higher for stage II than for stage I (mean ± SD 10.5 ± 4.5 vs. 8.5 ± 5) (Mann–Whitney test, p = 0.04). Partial volume effect, i.e. the systematic underestimation of the activity within small lesions, could not be held responsible for this as there was no statistical difference in the mean tumor volumes between stage I and stage II tumors (both 33 cm³, p = 0.18, Mann–Whitney test).

3.2. Overall survival

The median SUVmax for the whole group was 7.8 and patients were dichotomized according to this threshold. The mean follow-up time was 45 ± 30 months (1–142 months).

Forty patients died during follow-up. There were 21 cancer unrelated deaths and 19 cancer-related deaths.

Fig. 1 shows the Kaplan–Meier survival curves of the two groups of patient with SUV ≤ 7.8 or SUV > 7.8. The median survival was significantly different, respectively 127 months and 69 months (p = 0.001).

We analyzed subgroups according to their pathological stage. For stage I patients (n = 76), the median survival was 127 months for SUVmax ≤ 7.8 and 69 months for SUVmax > 7.8 (p = 0.001) as shown in Fig. 2A. For stage II patients (n = 21), no statistical difference was found between median survival of SUVmax ≤ 7.8 (72 mo) and
median survival of SUVmax > 7.8 (40 months) (\( p = 0.11 \)). However, the ratio of median survivals was 1.8 with 95% confidence intervals of 1.58–2.016, suggesting that there is a difference between the groups but that the small number of patients (\( n = 21 \)) prevents reaching statistical significance (Fig. 2B).

3.3. Disease-free survival (DFS) (Fig. 3)

Median DFS was 96.1 months for all patients with SUVmax ≤ 7.8 and 87.7 months for SUVmax > 7.8 (\( p = 0.01 \)). For stage I patients (\( n = 76 \)), the median DFS was 96 months for SUVmax ≤ 7.8 and 87.7 months for SUVmax > 7.8 (\( p = 0.058 \)). For stage II patients (\( n = 21 \)), median DFS was 51.5 months for SUV > 7.8, and undefined for SUV ≤ 7.8 (too small sample, \( n = 6 \)).

4. Discussion

Our results show that the metabolism of tumors as measured by FDG PET, and expressed as SUVmax, yields a strong prognostic information in patients with completely resected NSCLC. Indeed, when SUVmax is higher than 7.8, the overall survival is significantly decreased (69 months) compared to the survival of patients with lower metabolically active tumors (127 months). This is also true for disease-free survival: highly metabolic tumors relapse earlier (DFS of 87 months vs 96 months).

In our patient population, the prognostic information appears stronger for stage I tumors. For stage II patients, the overall survival was indeed longer for tumors with low SUV (72 months vs 40 months), without reaching statistical significance (\( p = 0.11 \)). However, the ratio of median survivals is 1.8 with 95% confidence intervals of 1.58–2.02, suggesting that there is a difference between low and highly metabolic stage II tumors, but that the small number of patients (\( n = 21 \)) prevented us from reaching statistical significance.

Our results for stage I NSCLC confirm previous works observed in early stage adenocarcinoma [11]. In our study design, we mixed all histological subgroups. We are also in accordance with the results published by Cerfolio et al. [20] who studied 315 patients with NSCLC and found that SUVmax was an independent prognostic predictor. In that study, patients with stage I–IV disease were included, which differs from our study strictly limited to stages I and II. We did not perform a multivariate analysis because at univariate analysis, no other variable tested (histologic subtype, stage, differentiation) was found significant, and could thus be entered in a multivariate model. We cannot compare our cut-off SUVmax with the ones used by Cerfolio et al. [20].

Indeed, stage grouping analysis was performed but the median SUVmax for each stage group (especially Ib, II) was unfortunately not given in the article, although the figures show that differences in survival and DFS were observed according to the metabolic activity.

The threshold SUVmax value of 7.8 was determined a posteriori using the median value observed in our population. It is important for the clinician to understand that SUV is a semi-quantitative index and may vary from one PET center to another. Especially, the SUV is not stable in time since the tumor uptake usually does not reach a plateau until 2 or 3 h after tracer injection. In this study, patients were scanned 60 min after injection, which is almost a standard procedure in clinical PET centers. Besides the interval between injection and data acquisition, SUV may vary according to the reconstruction algorithm, the region of interest drawing method, or blood glucose level [21]. Differences in any of these parameters might account for the differences in threshold SUV between studies. This implies that clinical PET centers must somehow standardize the method for measuring SUV before applying threshold published in the literature.

Despite the standardization problems, it remains that the metabolic activity of tumors has been repetitively shown to add significant information in terms of prognosis [6,8,10–13]. For stage I patients, this raises the question of using the PET information to decide whether or not the patient might benefit from adjuvant chemotherapy.

Indeed, whether chemotherapy can be useful as adjuvant treatment in stage I NSCLC is still unclear. It is for now not recommended [22]. Since PET FDG is part of the standard pre-therapeutic workup of NSCLC [7], SUVmax of the primary tumor can be easily obtained. We believe that this information could be used to select patients who would benefit from at least a closer postoperative surveillance, as suggested by the significantly reduced mean disease-free survival associated with highly metabolic tumors. Introducing systematic adjuvant chemotherapy for stage I tumors with high FDG uptake will require additional prospective studies. In particular, the issue of partial volume correction will have to be taken into account, since many of those stage I tumors are small lesions, which can lead to an underestimation of their FDG uptake and introduce a major bias by classifying small lesions into the lower metabolic group. In that view, the use of integrated PET-CT scanners will facilitate corrections for the partial volume effect.

References


[2] Birim O, Kappetein AP, Stijnen T, Bogers AJ. Meta-analysis of positron emission tomographic and computed tomographic imaging in detecting...


Appendix A. Conference discussion

Dr C. Tan (London, United Kingdom): Your results correlating SUVmax with survival is consistent with the literature. To add to the studies you quoted in your presentation, we have also done our own review of just under 500 patients and found similar results that the SUV correlates the survival.

Robert Cerfolio from Alabama in 2005 published a study of 315 patients and came to similar conclusions, and last year there was a review in the Journal of Thoracic Oncology of 10 papers coming to similar conclusions. This is not surprising since PET measures metabolic activity in the tumor.

So how does your data translate into clinical practice and how can we use this in reality? As you have indicated, the cutoff SUV is difficult between centres because of various factors. In your study, you derived your cutoff SUVmax from a median so that you can dichotomize into high and low SUVmax groups. But in reality, as you have demonstrated, the SUV is a continuum of value so that most of the patients with very low SUVmax do well and those with very high SUVs do not do as well. With the group in the middle, some do well and some do not, but it is more difficult to differentiate this group of patients. What would you do with an SUV of 7.7? And what would you do with an SUV of 7.9?

Looking at your subanalysis of the patients with stage II Lung cancer, you did not find a significant difference in survival between the high and the low SUV groups. As you said, this is based on the 21 patients, so I think this is more due to a statistical power effect than an actual difference effect.

The evidence for adjuvant chemotherapy in Lung cancer is that it only makes a difference of 5%, with the worst cancer groups probably getting the biggest difference. So how much benefit will stage I lung cancer patients get with adjuvant chemotherapy? How much difference are you going to make?

Dr Poncelet: Number one, the cutoff of 7.8 was, obviously, to dichotomize the group. However, in our work, we did all the studies and statistical analysis for different threshold values. And so any center could calculate its own threshold value, and go down as low as to see the time where statistical significance is reached. Because in our study, of course, at 7.8 we were highly significant. But going down to SUV threshold of 6.0 or to 5.5, there was still statistical significance between the two groups. So I think that you can go down on the SUV for each center and decide at which time it starts to be significant, and so it basically gives you where you can stratify your group.

Secondly 21 patients for stage II is obviously not enough and so we think this is more due to a statistical power effect than an actual difference effect.

The evidence for adjuvant chemotherapy in Lung cancer is that it only makes a difference of 5%, with the worst cancer groups probably getting the biggest difference. So how much benefit will stage I lung cancer patients get with adjuvant chemotherapy? How much difference are you going to make?

Dr Poncelet: Number one, the cutoff of 7.8 was, obviously, to dichotomize the group. However, in our work, we did all the studies and statistical analysis for different threshold values. And so any center could calculate its own threshold value, and go down as low as to see the time where statistical significance is reached. Because in our study, of course, at 7.8 we were highly significant. But going down to SUV threshold of 6.0 or to 5.5, there was still statistical significance between the two groups. So I think that you can go down on the SUV for each center and decide at which time it starts to be significant, and so it basically gives you where you can stratify your group.

Secondly 21 patients for stage II is obviously not enough and so we think this is more due to a statistical power effect than an actual difference effect.