

RESEARCH ARTICLE

# Interobserver Variability in Chest CT and Whole Body FDG-PET Screening for Distant Metastases in Head and Neck Cancer Patients

Asaf Senft,<sup>1</sup> Remco de Bree,<sup>1</sup> Richard P. Golding,<sup>2</sup> Emile F. I. Comans,<sup>3</sup>  
Jan-Hein T. M. Van Waesberghe,<sup>2</sup> J. Dirk Kuik,<sup>4</sup> Otto S. Hoekstra,<sup>3</sup> C. René Leemans<sup>1</sup>

<sup>1</sup>Department of Otolaryngology/Head and Neck Surgery, VU University Medical Center, De Boelelaan 1117, 1081 HV Amsterdam, PO Box 7057, 1007 MB Amsterdam, The Netherlands

<sup>2</sup>Department of Radiology, VU University Medical Center, Amsterdam, The Netherlands

<sup>3</sup>Department of Nuclear Medicine and PET Research, VU University Medical Center, Amsterdam, The Netherlands

<sup>4</sup>Department of Clinical Epidemiology and Biostatistics, VU University Medical Center, Amsterdam, The Netherlands

## Abstract

**Purpose:** The aim of the study was to assess the interobserver variability in chest computed tomography (CT) and whole body 2-deoxy-2-[<sup>18</sup>F]fluoro-D-glucose positron emission tomography (FDG-PET) screening for distant metastases in head and neck squamous cell carcinoma (HNSCC) patients.

**Procedure:** Chest CT and whole body FDG-PET of 69 HNSCC patients with high-risk factors who underwent screening for distant metastases were analyzed. All scans were independently read by two experienced radiologists or nuclear physicians who were blinded to the other examinations and follow-up results.

**Results:** A kappa of 0.516 was found for assessment of size on CT. Kappa values for origin and susceptibility of 0.406 and 0.512 for CT and 0.834 and 0.939 for PET were found, respectively. The overall conclusions had a kappa of 0.517–0.634 for CT and 0.820–1.000 for PET.

**Conclusions:** In screening for distant metastases in HNSCC patients with high-risk factors, chest CT readings had a reasonable to substantial agreement, while PET readings showed an almost perfect agreement. These findings suggest that for optimal assessment in clinical practice, PET most often can be scored by one observer, but CT should probably more often be scored by different observers in consensus or combined with PET.

**Key words:** CT, FDG-PET, Interobserver agreement, Distant metastases, Head and neck cancer

## Introduction

Head and neck squamous cell carcinomas (HNSCC) grow locally invasive and have a proclivity to metastasize to regional lymph nodes rather than to spread hematogenously. However, the presence of distant metastases influences prognosis and choice of treatment in patients with HNSCC. Patients with HNSCC and distant metastases

are generally not considered curable and are treated mostly palliatively.

In both clinical and autopsy studies, the lungs are the most frequent site of distant metastases in patients with head and neck cancer [1–3]. Moreover, lung metastases occur in 61–91% in combination with distant metastases at other sites. Distant metastases at other sites without simultaneous lung metastases are found in only 6–25% [2]. Because of the high incidence of lung metastases and the frequent combination of distant metastases at other sites, examination of the thorax is most important in screening for distant metastases.

The diligence with which technique the lungs should be screened remains controversial. Computed tomography (CT) is more sensitive in the detection of pulmonary nodules than plain chest radiography, because of the superiority of CT in detecting small nodules [1, 4, 5].

In a previous study, it was concluded that chest CT was the single most important diagnostic technique for pretreatment screening for distant metastases [1]. However, despite negative screening by chest CT and locoregional tumour control some patients develop distant metastases [6]. These distant metastases must have been present at diagnostic work-up, but were apparently below the detection limit of screening tests.

In screening for distant metastases second primary tumours can occasionally be detected at the same time, a potential secondary gain in this group of patients. Second primary tumours also have impact on survival and may alter the selection of therapy in HNSCC patients. The cumulative risk for second primary tumours in HNSCC patients is 3% per year. Synchronous second primary tumours are diagnosed in about 4% of the HNSCC patients. Although the head and neck region is the most frequent site, synchronous primary tumours also occur below the clavicles: lungs, oesophagus and other sites [7]. Therefore, the detection of second primary tumours during initial work-up is important.

In a multicenter prospective study we found that whole body positron emission tomography (PET) using the radio-labelled glucose analog 2-deoxy-2-[<sup>18</sup>F]fluoro-D-glucose (FDG) has additional value in screening for distant metastases and second primary tumours, if applied to the subset of patients at substantial risk [8]. An assessment of imaging examinations is usually based on a determination of their accuracy rates and sensitivity and specificity values. However, the clinical utility of an imaging study also depends on the reliability or the consistency with which the study is interpreted in the same way by different observers. The consistency of observations made by different observers in interpreting the same studies is termed interobserver reliability or agreement. Although the accuracy rates of CT and PET for screening on distant metastases in HNSCC patients have been determined and compared in several studies, the interobserver reliabilities of these diagnostic techniques have not been measured. The extent to which these accuracy results found by individual observers can be generalised, and thereby foresee the applicability of CT and PET for this patient group in daily clinical practice, tends to depend on the degree of uniform reporting by different observers. This study was performed to evaluate the interobserver variability in reporting of CT and PET for screening on distant metastases in HNSCC patients.

## Materials and Methods

Chest CT and whole body FDG-PET of 69 HNSCC patients (18 women and 51 men) with high-risk factors who underwent screening for distant metastases in our institute were analyzed. The protocol was approved by the institutional ethics committee.

Since these examinations are performed in routine clinical practice no informed consent was asked.

The mean age was 59 years and ranged from 40 to 81 years. Primary tumour sites included the oral cavity ( $n=12$ ), oropharynx ( $n=25$ ), hypopharynx ( $n=16$ ), larynx ( $n=10$ ), cervical oesophagus ( $n=4$ ) and lymph node metastases of unknown primary tumour ( $n=12$ ). Eight patients had two or more synchronous primary tumours. Indications (based on palpation, CT, MRI, and/or ultrasound-guided fine-needle aspiration cytology) for screening for distant metastases were three or more lymph nodes metastases ( $n=8$ ), bilateral lymph metastases ( $n=19$ ), lymph node metastases of 6 cm or larger ( $n=16$ ), low jugular lymph node metastases ( $n=2$ ), regional tumour recurrence ( $n=8$ ) and second primary tumours ( $n=21$ ). Some patients had more than one indication for screening. All were candidates for extensive treatment with curative intent: surgery and/or radiotherapy with or without chemotherapy.

In 67 of the 69 patients a chest CT, which was performed to screen for lung metastases, mediastinal lymph node metastases and second primary bronchogenic carcinoma, was available for review. Spiral CT scans were obtained with a fourth-generation Siemens Somatom Plus (Siemens AG, Erlangen, Germany after intravenous administration of contrast medium (Ultravist, Schering AG, Berlin, Germany). Contiguous axial scanning planes were used with a 5-mm slice thickness without interslice gap. All images were reviewed on PACS. Size was measured with manual electronic measurement. The volume of intravenous contrast was 100 ml at 3 ml/s with a delay of 25 to 30 s. Radiological criteria for lung metastases were: smoothly defined and subpleurally located lesions, multiple and located at the end of a blood vessel; for bronchogenic carcinoma, solitary, spiculated and centrally located lesions; and for mediastinal lymph node metastases, a short axial diameter of more than 10 mm [9].

All 67 chest CT scans were independently read by two experienced radiologists (RPG, JHW) who were blinded to the other examinations and follow-up results. On special forms location, long-axis diameter (<1, 1–2, 2–3, >3 cm), origin (metastasis, second primary, benign), and a five-point ordinal Likert scale score (1=definitively benign, 2=probably benign, 3=equivocal, 4=probably malignant, 5=definitely malignant) of the most suspected lesions (with a maximum of 5) were scored. Finally a conclusion had to be made for the presence (yes, no or equivocal) of metastases or second primary tumour. Spiculations were included in the determination of the long-axis diameter. If a nodule was visible on several adjacent images, the largest diameter was selected.

All 69 patients underwent FDG-PET after a 6-h fast. At 90 min after the intravenous administration of 10 mCi (370 MBq) FDG imaging of the body (trajectory: knee-skull) was performed using a dedicated PET scanner (Siemens HR plus). Any focal abnormality suspicious for malignancy was reported. Although the primary goal was screening for distant metastases, second primary tumours were additionally scored as an event.

As with CT, PET images of the 69 patients were scored by two independent experienced nuclear physicians (OSH, EFC). FDG uptake was considered abnormal in cases of enhanced uptake incompatible with its physiological biodistribution. The interpreters used special forms to register the location and aspect ('focal' or 'diffuse') of lesions in PET scans, and to assign a Likert score to grade their suspicion of malignancy (of the most suspected lesions). Finally, a conclusion had to be drawn for the presence (yes, no or equivocal) of metastases or second primary tumour. The 'aspect' of lesions was included since this is one of the elements that helps with interpretation: areas of diffusely enhanced uptake are more

likely to be inflammatory than focal ones. Like CT, differentiation between primary and secondary lesions can be difficult with PET; in the present study the reviewers classified central pulmonary lesions in PET scans as suspicious of primary tumours, and peripheral lesions as metastases unless there were additional lesions in PET scans (e.g. mediastinal foci) suggesting the presence of a second primary tumour and its metastases. No standard uptake value was calculated. No axis of a lesion was measured because, PET does not reliably estimate tumour size.

The interobserver agreement was determined and expressed in a weighted or unweighted kappa which corrects for agreement by chance. The higher the kappa, the higher the agreement, with a maximum of 1.0: <math>0</math>=no agreement, 0.0–0.19=poor agreement, 0.20–0.39=fair agreement, 0.40–0.59=moderate agreement, 0.60–0.79=substantial agreement, 0.80–1.00=almost perfect agreement [10].

In case of disagreement between the two observers a final consensus reading was performed. Any change in scoring was reported.

To correct for difference in scanning separate analysis was performed for lesions inside the thorax. To examine the role of the spatial resolution separate analysis was performed for lesions <math><1</math> and

## Results

In 39 of the 67 patients, no suspected lesions were found by chest CT. In the remaining 28 patients, a total number of 109 lesions on CT were scored (62 by observer 1 and 47 by

observer 2). In 43 of the 69 patients, no suspected lesions were found by PET. In the remaining 26 patients a total number of 94 lesions on PET were scored (47 by observer 1 and 47 by observer 2). The scorings of the observers and the kappa values are shown in Table 1.

The kappa value for long-axis diameter on CT was 0.516 (95% confidence interval (CI) 0.357–0.675). For origin, Likert scale score, malignancy, metastasis and second primary tumour the values were 0.406 (95% CI; 0.277–0.534), 0.512 (95% CI; 0.384–0.640), 0.634 (95% CI; 0.387–0.881), 0.523 (95% CI; 0.226–0.780) and 0.517 (95% CI; 0.236–0.798), respectively. The long-axis diameter cannot be measured on PET. The kappa values for origin, Likert scale score, malignancy, metastasis and second primary tumour were 0.834 (95% CI; 0.699–0.969), 0.961 (95% CI; 0.909–1.000), 1.000 (95% CI; 1.000–1.000), 0.820 (95% CI; 0.648–0.992) and 0.826 (95% CI; 0.633–1.000), respectively.

Initial disagreement in overall conclusions between the examiners occurred in eight CT examinations. The examiners could reach consensus in all cases. After consensus reading, observer 1 changed his diagnosis five times: three times from second primary to no malignant lesion and two times from no malignancy to metastasis. Observer 2 changed his diagnosis four times: two times from no malignancy to metastasis, one time from equivocal to metastasis and one time from metastasis to no malignancy.

**Table 1.** Scorings of the observers with interobserver agreement as kappa values

|                        | CT           |              | Kappa              | PET          |              | Kappa              |
|------------------------|--------------|--------------|--------------------|--------------|--------------|--------------------|
|                        | Observer 1   | Observer 2   |                    | Observer 1   | Observer 2   |                    |
| Per lesion             | <i>N</i> =62 | <i>N</i> =47 |                    | <i>N</i> =47 | <i>N</i> =47 |                    |
| Long-axis diameter     |              |              | 0.516              | N/A          | N/A          | N/A                |
| <math><1</math> cm     | 34           | 26           |                    |              |              |                    |
| 1–2 cm                 | 21           | 17           |                    |              |              |                    |
| 2–3 cm                 | 4            | 1            |                    |              |              |                    |
| >3 cm                  | 3            | 3            |                    |              |              |                    |
| Origin                 |              |              | 0.406 <sup>a</sup> |              |              | 0.834 <sup>a</sup> |
| Metastasis             | 30           | 19           |                    | 15           | 14           |                    |
| Second primary         | 9            | 12           |                    | 11           | 11           |                    |
| Benign                 | 23           | 16           |                    | 21           | 22           |                    |
| Likert                 |              |              | 0.512              |              |              | 0.939              |
| Definitively benign    | 5            | 6            |                    | 0            | 0            |                    |
| Probably benign        | 17           | 10           |                    | 10           | 12           |                    |
| Equivocal              | 10           | 10           |                    | 9            | 7            |                    |
| Probably malignant     | 13           | 12           |                    | 21           | 21           |                    |
| Definitively malignant | 17           | 9            |                    | 7            | 7            |                    |
| Overall conclusion     | <i>N</i> =67 | <i>N</i> =67 |                    | <i>N</i> =69 | <i>N</i> =69 |                    |
| Malignancy             |              |              | 0.634 <sup>a</sup> |              |              | 1.000 <sup>a</sup> |
| Yes                    | 13           | 10           |                    | 15           | 15           |                    |
| No                     | 54           | 57           |                    | 54           | 54           |                    |
| Metastasis             |              |              | 0.523              |              |              | 0.820              |
| Yes                    | 8            | 6            |                    | 11           | 9            |                    |
| Equivocal              | 6            | 6            |                    | 3            | 4            |                    |
| No                     | 53           | 55           |                    | 55           | 56           |                    |
| Second primary tumour  |              |              | 0.517              |              |              | 0.826              |
| Yes                    | 6            | 6            |                    | 7            | 8            |                    |
| Equivocal              | 2            | 3            |                    | 2            | 2            |                    |
| No                     | 59           | 58           |                    | 60           | 59           |                    |

<sup>a</sup>Unweighted kappa

Initial disagreement in overall conclusions between the examiners occurred in five PET examinations. Also for PET the examiners could reach consensus in all cases. After consensus reading observer 1 changed his diagnosis three times: two times from metastasis to second primary tumour and one time from metastasis to unclear. Observer 2 changed his diagnosis two times: both times from metastasis to second primary tumour.

### *Lesions Outside CT Scanning Range*

Seven lesions were observed outside of the thorax. Three lesions were localized in the rectum and two lesions in the colon. According to both observers, all of these lesions were not suspicious for malignancy (focal polyps). One lesion was localized in the liver. This lesion was scored as probably malignant by both observers. One lesion was localized in the lumbar spine and was scored as being definitively malignant by both observers. If lesions outside of the thorax were left out, the kappa for PET interobserver agreement were as follows: origin 0.811 (95% CI; 0.637–0.985); Likert 0.971 (95% CI; 0.916–1.000); malignancy 1.000; metastases 0.740 (95% CI; 0.521–0.959) and second primary tumour 0.858 (95% CI; 0.698–1.000).

### *Nodules <10 mm*

In a total of 18 patients, lesions <10 mm on CT were reported. In 11 out of 18 patients in whom lesions <10 mm were reported on CT no lesions were seen on PET (both observers negative). For lesions <10 mm (as measured by CT observer 1) the kappa values for CT interobserver agreement were: origin 0.308 (95% CI; 0.009–0.606); Likert 0.411 (95% CI; 0.150–0.671); malignancy 0.558 (95% CI; 0.411–0.705); metastases 0.444 (95% CI; 0.156–0.733) and second primary tumour 0.627 (95% CI; 0.383–0.870). For PET these figures were 1.000 for all variables. For the other lesions ( $\geq 10$  mm) the kappa values for CT interobserver agreement were: origin 0.535 (95% CI; 0.227–0.843); Likert 0.469 (95% CI; 0.178–0.760); malignancy 0.524 (95% CI; 0.387–0.661); metastases 0.509 (95% CI; 0.267–0.752) and second primary tumour 0.339 (95% CI; 0.102–0.576). For PET these figures were: origin 0.811 (95% CI; 0.659–0.963); Likert 0.955 (95% CI; 0.894–1.000); malignancy 1.000; metastases 0.801 (95% CI; 0.630–0.972) and second primary tumour 0.898 (95% CI; 0.784–1.000).

## **Discussion**

To be consistently useful, interpretation of imaging techniques must be reproducible. Ideally, both physicians with and without special expertise in a particular area will provide consistent interpretations. Although some accuracy data of chest CT and PET in screening for distant metastases have been determined and compared, the interobserver variability of CT and PET has not been measured [1, 4–8, 11].

CT is extremely sensitive in the detection of pulmonary nodules but is frequently indeterminate in diagnosis. Increasing numbers of pulmonary nodules are being detected, in large part due to the recent developments in CT imaging techniques. While specific patterns of calcification or the presence of fat in pulmonary nodules on CT can be used to determine if a nodule is benign, most nodules lack benign characteristics and are therefore considered indeterminate for malignancy. These non-calcified nodules represent a diagnostic challenge [12]. Interobserver agreement for the detection of individual pulmonary nodules on CT is reported to be relatively poor. Wormanns et al. [13] reported that, of a total of 286 nodules, 103 nodules were found by both readers. Leader et al. [14] scored 293 low-dose chest CT scans as to their probability of being benign or malignant nodule-based and examination based interobserver agreement among the three radiologist was poor: highest kappa values in paired comparison 0.120 and 0.458, respectively. In the present study a substantial amount of agreement (kappa 0.634) was found for scoring the presence or absence of malignancy using CT, whereas the agreement for this scoring was optimal (kappa 1.000) using PET. Also a five-point ordinal Likert scale was used to classify the level of susceptibility for malignancy. The interobserver agreement for CT findings was moderate (kappa 0.512), whereas for PET a high agreement (kappa 0.939) was found using five-point ordinal scoring. These findings emphasize the difficulty in interpretation of pulmonary nodules on CT. As with CT, reading a PET scan requires weighing several factors to arrive at a diagnostic probability. There is no mathematical formula to cover them all. After detection, the interpretation process of a lesion involves several observer-dependent components, and this was one of the reasons for studying the observer variation. In this study, we described and implemented rules which are applied in our clinical practice. As in the present study Joshi et al. [15] found a very high interobserver agreement for the evaluation of pulmonary nodules by PET as assessed with intraclass correlation coefficients of 0.93 (range from 0 to 1). On PET images lesions are more or less 'present' or 'absent' and therefore probably less susceptible for variation in interpretation. In the presented study this is reflected in the facts that PET detected fewer lesions <10 mm, but the lesions which were seen were scored with an optimal interobserver agreement (kappa 1.0).

On CT, differentiation between a solitary pulmonary metastases and a second primary bronchiogenic carcinoma may be difficult. Therefore, most studies report on intrathoracic malignancies without separating metastases from primary tumours. In the present study the origin of lesions were scored by both CT and PET observers. The agreement on origin for the CT observers was moderate (kappa 0.406) and for PET observers high (kappa 0.834). Also the agreement in overall conclusion if pulmonary metastases were present was higher with PET as compared with CT observers (kappa 0.820 versus 0.523, respectively). Also for

the conclusion if a primary bronchiogenic carcinoma was present or not, a higher interobserver agreement was found for PET than CT (kappa 0.826 and 0.517, respectively).

In certain clinical settings, accurate assessment of the size of pulmonary nodules is important. In screening for distant metastases not the size but the nature (benign or malignant) and type (metastases or primary tumour) of the lesions are important. Only for detection of growth of small equivocal pulmonary nodules at follow-up suggestive of malignancy exact size measurement is warranted. Reports describing interobserver agreement for sizing nodules have been mixed. Hopper et al. [16] evaluating interobserver variability in the measurements of metastases to the lung and liver on CT demonstrated statistically significant interobserver variability of 15%. Bogot et al. [17] found a statistically significant interobserver variability in measuring pulmonary nodule volumes. Revel et al. [18] found that both intra- and interobserver agreement for measurement of nodule size (long-axis diameter) on CT scans was poor. This is especially true for irregular and poorly defined tumour foci [16]. On the contrary, Wormanns et al. [13] assessed the interobserver variability in size determination of pulmonary nodules at spiral CT. In 23 patients with known pulmonary nodules diagnostic confidence and size in exact size measurement and categorization into three size classes ( $\leq 5$ , 6–10,  $>10$  mm) were scored by two observers. A good correlation (Pearson's correlation coefficient 0.89–0.95) of measurements in millimetres was found. A good interobserver agreement in categories (kappa 0.74) was reported [11]. In the present study, a moderate amount (kappa 0.516) of agreement was found in categorization of size classes using CT. This agreement may be slightly different in newer generation CT scanners. In automated volume measurements interobserver agreement is less relevant.

In the present study, reading in consensus changed the diagnosis (metastasis or second primary tumour) in 6% for CT and 7% for PET. This implies that probably in a subset of scans reading by two observers may be helpful.

In the present study in all categories the interobserver agreement of PET was higher as compared with CT. PET detected 47 lesions in 26 patients, while CT detected 69 lesions in 28 patients. Tumour size is an important determinant of the ability of PET to detect smaller lung malignancies. While no absolute size criteria is established, it is generally accepted that lesions less than 10 mm are predisposed to false negative results on PET due to limited spatial resolution or low overall tumour volume. The limited spatial resolution of PET together with nodule motion from respiration at image acquisition may also impact the accurate detection of small pulmonary nodules [19]. If visualized by PET the nature of the lesion is probably less difficult to determine than by CT, which depicts much smaller lesions. Scoring CT is probably more difficult because more lesions are visualized. It is anticipated that the use of newer generation CT scanners and software, e.g. computer-aided detection, yield an increase in detection of (small) lung lesions [20]. These technical improvements may result in a

higher sensitivity. However, as is shown in this study, the detection of smaller lesions is accompanied by a lower interobserver agreement. Combined reading of CT and PET may be helpful in lesions with a size that can theoretically be visualized by PET. In those lesions PET can aid in adding certainty in scoring the level of malignancy.

Because the data were acquired before PET-CT was widely available and became the standard, in the present study stand-alone PET rather than PET-CT has been used. However, we think that the most findings are still of relevance. PET and CT were compared in a head-to-head comparison. Even though PET-CT is becoming more prevalent now, and some comparative issues encountered with stand-alone systems will be become less problematic, we feel that the first step of interpretation of PET-CT images should be an independent review of PET and CT. Combined readings thereafter will allow a joint estimate of the probability of disease.

## Conclusion

In screening for distant metastases in HNSCC patients with high-risk factors chest CT readings had a reasonable to substantial agreement for size, origin and susceptibility of lesions, while PET readings showed an almost perfect agreement for lesion characteristics. These findings suggest that for optimal assessment in clinical practice PET most often can be scored by one observer, but CT should probably more often be scored by different observers in consensus or combined with PET.

*Acknowledgement.* This study is financially supported by ZonMw Health Care Efficiency Research Program grant 945-11-014.

*Open Access.* This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any non-commercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

## References

1. de Bree R, Deurloo EE, Snow GB, Leemans CR (2000) Screening for distant metastases in patients with head and neck cancer. *Laryngoscope* 110:397–401
2. Ferlito A, Rinaldo A, Buckley JG, Mondin V (2001) General considerations on distant metastases from head and neck cancer. *ORL J Otorhinolaryngol Relat Spec* 63:189–191
3. Nishijima W, Takooda S, Tokita N, Takayama S, Sakura M (1993) Analyses of distant metastases in squamous cell carcinoma of the head and neck and lesions above the clavicle at autopsy. *Arch Otolaryngol Head Neck Surg* 119:65–68
4. Chalmers N, Best JJ (1991) The significance of pulmonary nodules detected by CT but not by chest radiography in tumour staging. *Clin Radiol* 44:410–412
5. Warner GC, Cox GJ (2001) Evaluation of chest radiography versus chest computed tomography in screening for pulmonary malignancy in advanced head and neck cancer. *J Otolaryngol* 32:107–109
6. Brouwer J, de Bree R, Hoekstra OS et al (2005) Screening for distant metastases in patients with head and neck cancer: is chest CT sufficient? *Laryngoscope* 115:1808–1812
7. Drooghe IJ, de Vos M, Van Cauwenberge PB (1998) Multiple primary tumours in head and neck cancer: results of a prospective study and future perspectives. *Laryngoscope* 108:250–256

8. Senft A, de Bree R, Hoekstra OS et al (2008) Screening for distant metastases in head and neck cancer patients by chest CT or whole body FDG-PET: a prospective multicenter trial. *Radiat Oncol* 33:221–229
9. Milne EN, Zerhouni EA (1987) Blood supply of pulmonary metastases. *J Thorac Imaging* 2:15–23
10. Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. *Biometrics* 33:159–174
11. Keyes JW, Chen MYM, Watson NE, Greven KM, McGuirt WF, Williams DW (2000) FDG PET evaluation of head and neck cancer: value of imaging the thorax. *Head Neck* 22:105–110
12. Erasmus JJ, Connolly JF, McAdams HP, Roggli VL (2000) Solitary pulmonary nodules: part 1. Morphologic evaluation for differentiation of benign and malignant lesions. *Radiographics* 20:43–58
13. Wormanns D, Diederich S, Lentschig MG, Winter F, Heindel W (2000) Spiral CT of pulmonary nodules: interobserver variation in assessment of lesion size. *Eur Radiol* 10:710–713
14. Leader JK, Warfel TE, Fuhrman CR et al (2005) Pulmonary nodule detection with low-dose CT of the lung: agreement among radiologist. *AJR Am J Roentgenol* 185:973–978
15. Joshi U, Rajmakers PGHM, van Lingen A et al (2005) Evaluation of pulmonary nodules: comparison of a prototype dual crystal dual head coincidence camera and full ring positron emission tomography (PET). *Eur J Radiol* 55:250–254
16. Hopper KD, Kasales CJ, Van Slyke MA, Schwartz TA, Tenhave TR, Jozefiak JA (1996) Analysis of interobserver and intraobserver variability in CT tumor measurements. *AJR Am J Roentgenol* 167:851–854
17. Bogot NR, Kazerooni EA, Kelly AM, Quint LE, Desjardins B, Nan B (2005) Interobserver and intraobserver variability in the assessment of pulmonary nodule size on CT using film and computer display methods. *Acad Radiol* 12:948–956
18. Revel MP, Bissery A, Biennu M, Aycard L, Lefort C, Fria G (2004) Are two-dimensional CT measurements of small noncalcified pulmonary nodules reliable? *Radiology* 231:453–458
19. Gilman MD, Aquino SL (2005) State-of-the-art FDG-PET imaging of lung cancer. *Semin Roentgenol* 40:143–145
20. Beyer F, Zierott L, Fallenberg EM et al (2007) Comparison of sensitivity and reading time for the use of computer-aided detection (CAD) of pulmonary nodules at MDCT as concurrent or second reader. *Eur Radiol* 17:2941–2947