

RESEARCH ARTICLE

2-Deoxy-2[F-18]FDG-PET for Detection of Recurrent Laryngeal Carcinoma after Radiotherapy: Interobserver Variability in Reporting

L. van der Putten,¹ O. S. Hoekstra,² R. de Bree,¹ D. J. Kuik,³ E. F. I. Comans,² J. A. Langendijk,⁴ C. R. Leemans¹

¹Department of Otolaryngology/Head and Neck Surgery, VU University Medical Center, De Boelelaan 1117, 1081 HV, Amsterdam, The Netherlands

²Department of Nuclear Medicine and PET Research, VU University Medical Center, Amsterdam, The Netherlands

³Department of Clinical Epidemiology and Biostatistics, VU University Medical Center, Amsterdam, The Netherlands

⁴Department of Radiation Oncology, University Medical Center Groningen, Groningen, The Netherlands

Abstract

Purpose: To evaluate accuracy and interobserver variability in the assessment of 2-deoxy-2[F-18]fluoro-D-glucose-positron emission tomography (FDG-PET) for detection of recurrent laryngeal carcinoma after radiotherapy.

Procedures: Eleven experienced nuclear physicians from eight centres assessed 30 FDG-PET scans on the appearance of local recurrence (negative/equivocal/positive). Conservative (equivocal analysed as negative) and sensitive (equivocal analysed as positive) assessment strategies were compared to the reference standard (recurrence within 6 months after PET).

Results: Seven patients had proven recurrences. For the conservative and sensitive strategy, the mean sensitivity was 87% and 97%, specificity 81% and 63%, positive predictive values 61% and 46% and negative predictive values 96% and 99%, respectively. Interobserver variability showed a reasonable relation in comparison to the reference standard ($\kappa = 0.55$).

Conclusions: FDG-PET has acceptable interobserver agreement and yields good negative predictive value for detection of recurrent laryngeal carcinoma. It could therefore be used as first diagnostic step and may reduce futile invasive diagnostics.

Key words: FDG-PET, Laryngeal carcinoma, Radiotherapy, Interobserver variability, Detection, Local recurrence

Introduction

Laryngeal cancer is the most common primary cancer site of the head and neck, with annually about 700 new cases in The Netherlands [1]. Because of the relatively low incidence and the specialised care, the vast majority of

laryngeal cancers in The Netherlands are treated in the eight recognised head and neck cancer centres of the Dutch Head and Neck Oncology Cooperative Group.

Parallel to developments elsewhere, non-surgical treatment modalities in our country are more common than in the past due to improved results of altered fractionation schedules in radiation therapy and the addition of chemotherapy to radiation [2–4]. The aims of non-surgical treatments are organ preservation and improvement of quality of life [5, 6]. Especially, in this group of patients, early detection of

This project is financially supported by the Netherlands Organisation for Health Research and Development (ZonMw), grant 945-04-311, RELAPS.

Correspondence to: R. de Bree; e-mail: r.bree@vumc.nl

residual or recurrent tumour is of critical importance because prompt salvage surgery improves control of disease.

The diagnostic accuracy of the currently available diagnostic techniques to diagnose persistent or recurrent disease after (chemo)-radiotherapy is currently limited. Post-irradiation inflammation, oedema and necrosis can hamper the detection of residual or recurrent local tumour. CT and MRI rely on structural changes and show a limited accuracy, with reported sensitivities ranging from 50% to 58% and specificities from 33% to 100% for detection of recurrent laryngeal carcinoma [7–14]. Direct laryngoscopy under general anaesthesia with biopsies runs the risk of complications (e.g., inducing necrosis, infection and further oedema) when biopsies are taken from irradiated tissue and was false negative in 31% of the initial laryngoscopies in previous research [15].

2-Deoxy-2-[F-18]fluoro-D-glucose-positron emission tomography (FDG-PET) is a promising technique for tumour detection after (chemo)radiotherapy. FDG-PET seems more accurate for the detection of recurrent head and neck carcinomas than other diagnostic methods [10, 16–22]. The reported sensitivity of FDG-PET for the detection of recurrent carcinoma after (chemo)radiotherapy varies between 80% and 100% and the specificity varies between 63% and 93% [12–14, 23–28].

Evaluation of PET in these studies is typically by visual interpretation, which is prone to observer variation. To the best of our knowledge, no studies have been conducted to investigate observer variation in a multi-centre setting for this indication.

To evaluate the value of FDG-PET in the detection of recurrent laryngeal carcinoma after radiotherapy, a randomised controlled multi-centre trial was started recently within the framework of the Dutch Head and Neck Oncology Cooperative Group [29]. The extent to which the future results can be generalised, and thereby foresee the applicability of PET in daily clinical practice, tends to depend on the degree of agreement among the observers. Therefore, we evaluated the interobserver variability in reporting among 11 observers involved in this trial, with a set of FDG-PET scans of patients suspected of having recurrent laryngeal carcinoma after radiotherapy.

Materials and Methods

From the VU University Medical Center (VUmc) PET database, we identified 30 FDG-PET scans of consecutive patients with a clinical suspicion of persistent or recurrent laryngeal carcinoma after radiotherapy from the year 1998 to 2000. This suspicion was based on clinical symptoms, office laryngoscopy or diagnostic imaging, other than FDG-PET. Patients' T stages prior to radiotherapy were T1 ($n = 4$), T2 ($n = 16$), T3 ($n = 5$) and T4 ($n = 5$), N stages were N0 ($n = 27$) and N1 ($n = 3$) and none of the patients had distant metastases. The mean age at time of the PET scan was 60.9 ± 11.1 years. The median interval between the last

radiation fraction and the PET scan was 8.7 months (range 2.4–32.1 months). As reference standard, we used the results of biopsies and histology or the absence of signs of tumour within 6 months after the PET scan.

PET Imaging

All patients underwent FDG-PET after at least 6-h fasting. Blood glucose levels were measured before scanning. All patients were non-diabetic. The median blood glucose level measured with a glucotouch stick was 5.9 ± 1.6 mmol/l. Sixty minutes after intravenous injection of 370 MBq of ^{18}F -FDG, imaging was performed in a full-ring bismuth germanate oxide PET scanner (ECAT EXACT HR+; CRI/Siemens, Erlangen, Germany). Five patients received a dose of 555 MBq because of higher weight. A scanning track from the base of skull to the clavicles was used, i.e. two bed positions per patient, with an acquisition time of 5 min per bed position. PET imaging was done with 2D acquisition using Ordered Subset Expectation Maximization (OSEM 2-16) to reconstruct the images. Attenuation correction was not applied in our clinical practice because of the results of the systemic review by Joshi et al. [30]. The acquired images were viewed on a local PC of the participating centre with the PETViewer 2.0.10.570 (Microsoft Windows XP Professional Service Pack 2 (build 2600)).

Data Analysis

The panel of observers consisted of 11 experienced nuclear medicine physicians from the eight Dutch Head and Neck Oncology Cooperative Group medical centres. We recorded the experience of the observers with FDG-PET for this indication in terms of the estimated number of FDG-PET scans for laryngeal cancer they had assessed. Images were reviewed at each site on available PC running under Microsoft Windows (XP/2000). The scans were presented with a PET-viewer that allowed variable gamma and window tuning.

The observers were requested to interpret the PET scans as being indicative for the presence of local residue or recurrence and to classify the result as negative, equivocal or positive. Only increased uptake at the site of the initial primary tumour was used for further analyses. The observers were not provided with specific criteria for determining positivity and negativity. Evaluation was made on an overall basis rather than on a per-lesion basis.

The observers read the scans independently, and clinical information (regarding TNM stage and site of the primary tumour, last radiotherapy fraction, symptoms and the result of other diagnostic tests) was provided to imitate the clinical setting. Correlative anatomic imaging (CT or MRI) was not provided nor were the reports of these scans. There was no time restriction for the assessment. Observers were blinded for the final clinical classification until they had reported all scans; thereafter, the investigator (LvdP) provided them with these results to improve standardised reading during the following prospective randomised trial. Cases that were scored incorrectly (when compared to the reference) or equivocal by at least nine observers were regarded as difficult cases and are described in the results.

Table 1. Classification of the interobserver variability with kappa

Kappa (κ)	Agreement
<0	No agreement other than agreement based on coincidence
0.01–0.19	Slight agreement
0.20–0.39	Fair agreement
0.40–0.59	Moderate agreement
0.60–0.79	Substantial agreement
0.80–0.99	Almost perfect agreement
1	Perfect agreement

The observers were asked which criteria they used for their interpretation of the FDG-PET scan.

Statistical Analysis

Data were obtained for a sensitive and a conservative PET reading strategy: The PET results were dichotomised by assigning equivocal scores to either the PET-positive or to the PET-negative classifications, respectively. Mean sensitivity, specificity, positive predictive value and negative predictive value were determined for either strategy. A Bayesian plot was used to show the probability of proven tumour within 6months for varying prevalences of tumour. To illustrate the probability of tumour in time after the PET scan for the different strategies and for different observers, a Kaplan Meier analysis was performed. Correlation between experience and diagnostic performance (measured with both conservative and sensitive strategies) and percentage of equivocal scores was evaluated.

To analyse the interobserver variability, we used agreement statistics (κ) with a classification according to Landis et al. [31] (Table 1). Linear-weighted kappa was used to determine interobserver variability of the 11 observers compared to the reference standard and pairwise compared to each other, for both conservative and sensitive strategy.

Results

Within 6 months after the FDG-PET scan, a local recurrence was histologically proven in seven patients (23%).

For both conservative (equivocal considered negative) and sensitive strategies (equivocal considered positive), the accuracy was determined per observer and depicted in a box plot (Fig. 1). The mean data of the accuracy of the 11 observers are shown in Table 2. For the conservative reading strategy, mean sensitivity of 87% (range 57–100%), specificity of 81% (range 65–96%), positive predictive value of 61% (range 43–80%) and negative predictive value of 96% (range 88–100%) were found. For the sensitive reading strategy, mean sensitivity of 97% (range 86–100%), specificity of 63% (range 39–87%), positive predictive value of 46% (range 33–70%) and negative predictive value of 99% (range 93–100%) were found.

In the Bayesian plot (Fig. 2), the two strategies mainly differ in the intermediate ranges of the prior probability of proven recurrence. Also, these differences are larger for the false negatives (lower corner) than for the true (and thus false) positives.

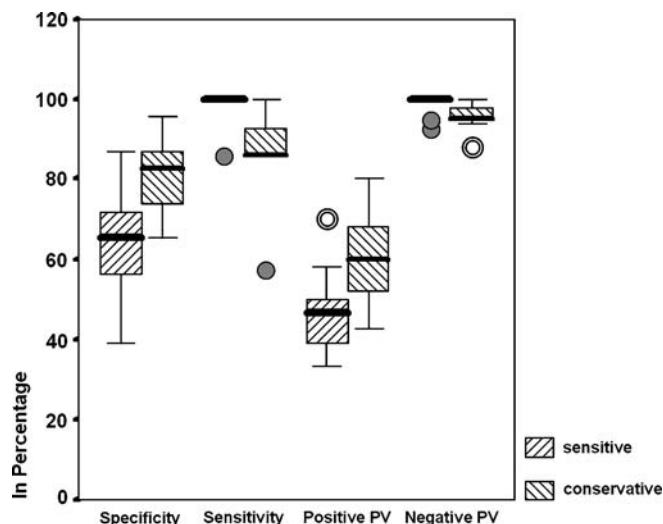


Fig. 1. Accuracy for conservative and sensitive reading strategies depicted in a box-and-whisker plot. The boxes contain the central half of the measurements (heavy line indicating the median). The dots are the values that are extremely far from the central box (outliers).

In a Kaplan Meier analysis (Fig. 3), as expected, an observer with a high accuracy (versus the reference) predicted the prognosis for local disease-free control more accurately than an observer with a low accuracy. With the conservative strategy, for both observers, a curve was established with significantly more local recurrences in the PET-positive than in the PET-negative group. Seven recurrences (71%) manifested within 6months after PET, no recurrences were diagnosed between 6 and 12 months and the remaining two recurrences were seen between 12 and 24 months.

The estimated total number of FDG-PET scans for suspected recurrent laryngeal cancer of the observers had assessed previously varied between 0 and 300 (experience with dual head gamma cameras included). There was no statistically significant association between this experience and the number of ‘equivocal’ scores ($p = 0.610$, Table 3). The equivocal category was reported at a mean of five (out of 30 cases) per observer (17%, range 1–10). Furthermore, we found no significant correlation between the experience and diagnostic performance (conservative $p = 0.360$, sensitive $p = 0.528$, Table 3).

The interobserver variability in comparison to the reference (local recurrence within 6months after the PET

Table 2. Mean pooled accuracy for conservative (equivocal = negative) and sensitive (equivocal = positive) strategies

	Conservative strategy (%)	Sensitive strategy (%)
Sensitivity	87	97
Specificity	81	63
Positive predictive value	61	46
Negative predictive value	96	99

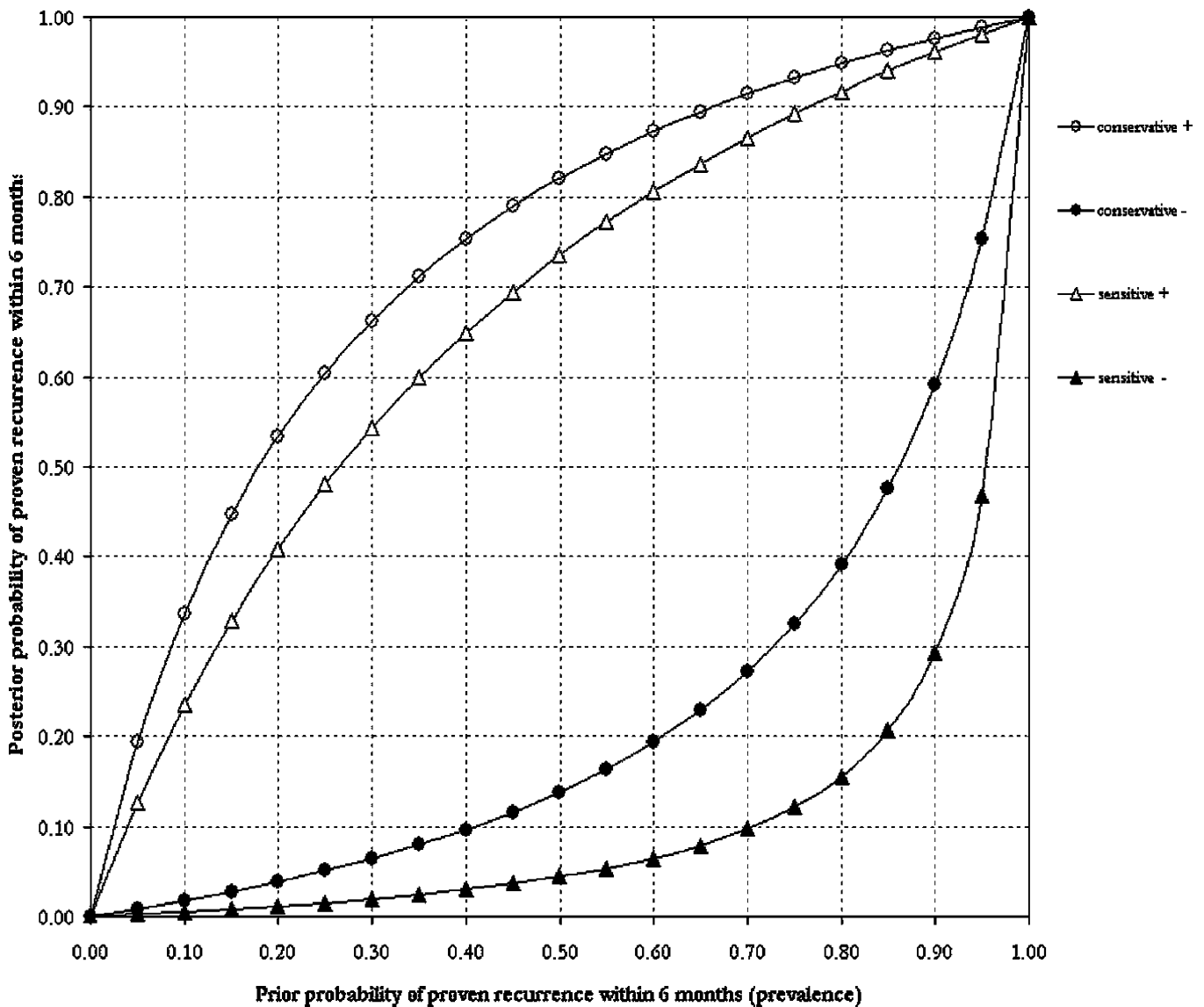


Fig. 2. Bayesian plot with the prior and posterior probability of proven recurrence within 6 months after FDG-PET for the conservative and the sensitive strategies.

scan) showed a moderate relation [$\kappa = 0.55$; 95% confidence interval (CI): 0.33–0.76]. The interobserver variability as pairwise comparison of the observers, which expresses the consistency between observers, also showed a moderate relation ($\kappa = 0.54$; 95% CI: 0.42–0.67).

When reducing the data from a three- to a two-point scale, the conservative strategy proved to result in a better interobserver agreement in comparison to the reference ($\kappa = 0.59$; 95% CI: 0.38–0.79) than the sensitive one ($\kappa = 0.43$; 95% CI: 0.22–0.63). The same was true for the pairwise comparison of the observers (conservative: $\kappa = 0.58$; 95% CI: 0.44–0.71, versus $\kappa = 0.51$; 95% CI: 0.37–0.65 for SR).

There were two difficult cases, with much discrepancy between the report of the observers and the reference (Fig. 4; patients #5 and #9). Both cases were negative according to

the reference. Patient #5 underwent FDG-PET 5 months after the last radiotherapy fraction for a left-sided T3N1 supraglottic laryngeal carcinoma. Clinical suspicion of recurrence was based on unexplained otalgia. A MRI of the neck showed diffuse paraglottic swelling on both sides (mainly on the right side), which could either be post-irradiation effects or recurrent tumour according to the radiologist (Fig. 5). Five observers scored the PET scan (Fig. 5) as equivocal and six as positive. At direct laryngoscopy, irregular tissue at the left aryepiglottic fold and the epiglottis was seen, but biopsy revealed no malignancy. Three years and 2 months after PET, the patient died with lung metastases, but a local recurrence was never detected.

Patient #9 had a PET scan 2 years after completion of radiotherapy for a left-sided T2N0 glottic carcinoma. The PET scan (Fig. 6) was indicated because the left side of the

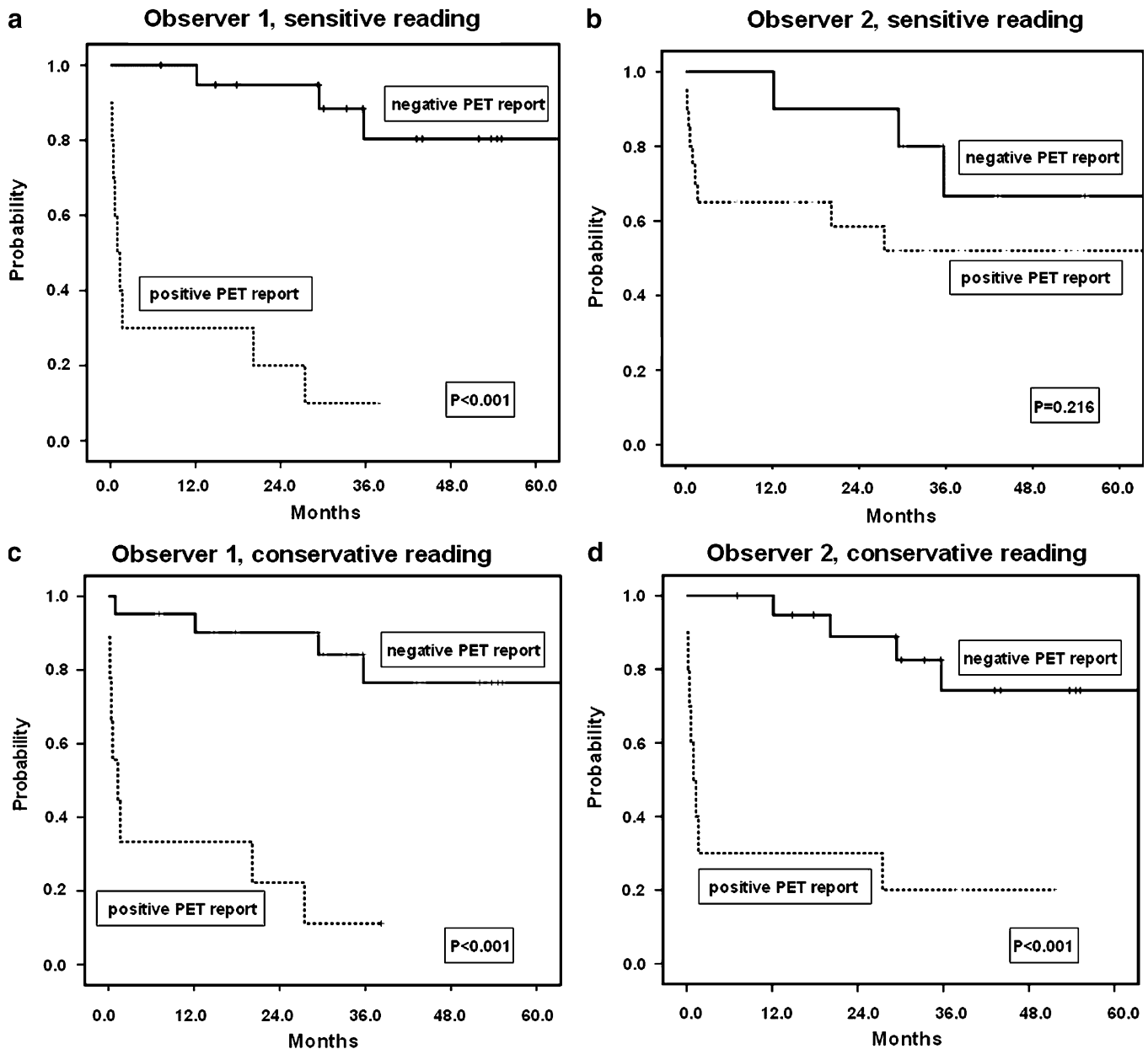


Fig. 3. Kaplan Meier analysis of a proven recurrence after PET, with stratification of the negative and positive assessed patients using the sensitive (a, b) and conservative (c, d) strategies by an observer with high accuracy (observer 1; a, c) and an observer with low accuracy (observer 2; b, d).

glottis appeared suspicious at indirect laryngoscopy. CT scan of the neck showed a suspect area just ventral to the lesion described on the PET scan. Seven observers scored the PET scan as positive, three as equivocal and one as negative. Clinical follow-up was uneventful until 2 years, later direct laryngoscopy revealed squamous cell carcinoma at the original tumour site. The laryngectomy specimen contained a squamous cell carcinoma of 1.5 cm in diameter located in the glottis with tumour extension into the thyroid cartilage.

The criteria the observers used for their interpretation were information derived from the PET scan, such as

localisation, (a)symmetry and aspect of suspicious areas, diffuse versus focal lesions and the intensity of the suspicious areas compared to the intensity of the background, in combination with the clinical data (localisation of primary tumour, interval between radiation and PET).

Discussion

In the present study, we analysed the performance of 11 observers from the eight head and neck cancers centres in The Netherlands for the assessment of FDG-PET scans from

Table 3. Correlations with log-experience between experience of the observers and the number of equivocal score, the accuracy with equivocal as negative (conservative) or positive (sensitive)

	<i>r</i>	<i>p</i>	95% CI of <i>r</i>
Number of equivocals	0.175	0.610	-0.464–0.694
Accuracy (conservative)	0.307	0.360	-0.347–0.760
Accuracy (sensitive)	0.215	0.528	-0.499–0.669

r Correlation coefficient, *p* significance, 95% CI 95% confidence interval

patients who were suspected of having a local recurrence of laryngeal carcinoma after primary radiotherapy.

We found a reasonable chance-corrected proportional observer agreement, both in comparison to the reference standard and pairwise. It is difficult to predict how the agreement would change if a larger sample size was studied. To the best of our knowledge, this is the first study that examines the interobserver variability of more than two observers from different institutes in the detection of recurrent laryngeal carcinoma with FDG-PET. Many authors

stress the importance of interobserver agreement [32, 33]. Fakhry et al. [34] studied the interobserver variability between two observers of FDG-PET in the detection of recurrent head and neck squamous cell carcinoma and found a good agreement (intraclass correlation coefficient >90). A substantial agreement was also described for metastatic disease. Bohdiewicz et al. [35] found a 90% agreement between two observers who reviewed FDG-PET scans for metastatic disease in the spinal cord, and Lim et al. [36] found a kappa of 0.68 for three observers who reviewed FDG-PET scans for peritoneal metastases. Hashimoto et al. [37] evaluated lung nodules with two observers of FDG-PET and found a kappa of 0.65. In a study performed by Zijlstra et al. [38], 11 observers reviewed FDG-PET scans for suspicion of recurrent lymphoma in 82% to 94% of the tumour-positive patients and 45% of the tumour-negative patients, which were in accordance with the experts.

Because the observer panel in this study consisted of nuclear physicians from all Dutch Head and Neck Cancer Centres, the results give a good impression of the overall diagnostic performance of PET for suspected laryngeal

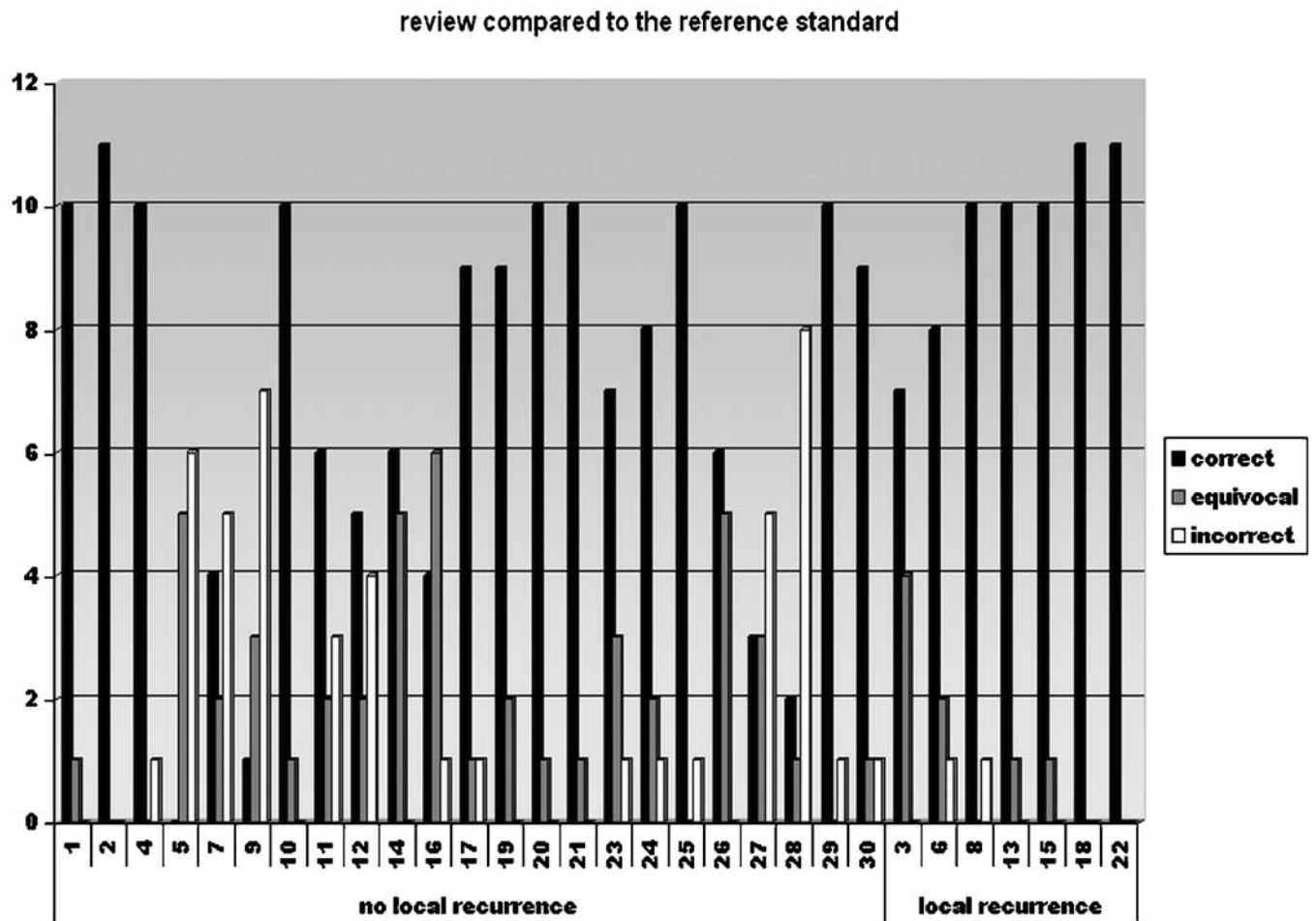


Fig. 4. All 30 cases and the results of the review (correct, equivocal, incorrect) compared to the reference standard (numbers of observers).

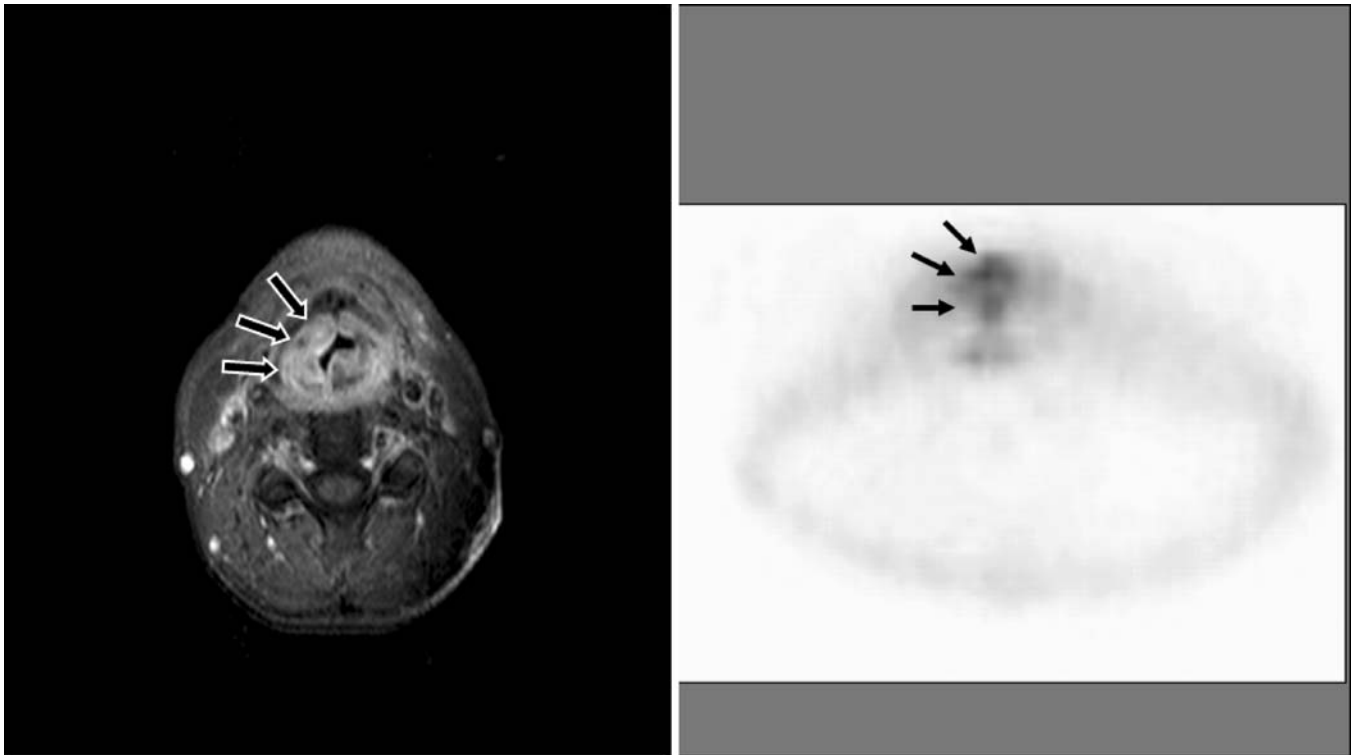


Fig. 5. Patient 5: MRI (STIR, axial) with diffuse paraglottic swelling on both sides, mainly right (*arrows*). PET (axial) with abnormal supraglottic ventral uptake, on the right side more than on the left side (*arrows*). *Below the arrows* is a region with abnormal uptake, probably caused by uptake in the crico-arythenoid muscle.

recurrence after radiotherapy in The Netherlands. Especially, since the interobserver agreement was reasonable, an acceptable reproducibility and thereby a general applicability of these results is assumed.

As expected, sensitivity and specificity varied inversely with the threshold of test positivity. Our data (mean sensitivity ranging from 87% to 97%, specificity from 63% to 81%) appear to reflect the distribution of such measures reported in

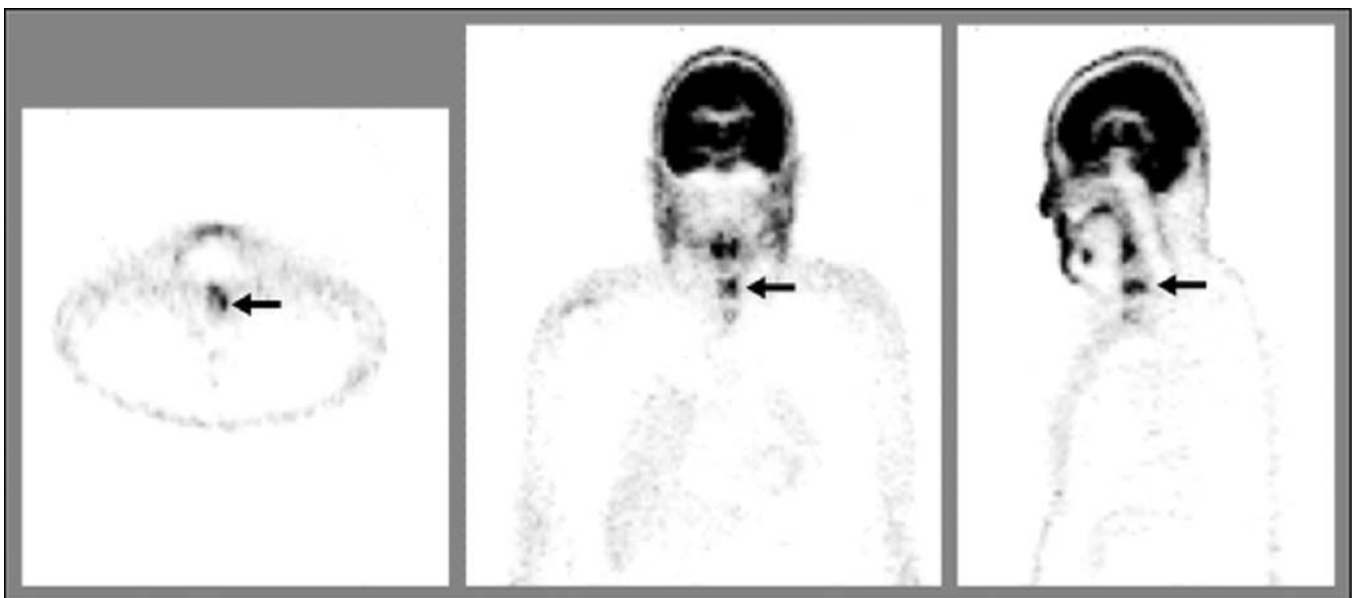


Fig. 6. FDG-PET scan of case 9 (axial, coronal and sagittal images) reviewed as tumour positive by seven observers, equivocal by three observers and negative by one observer (*arrows* indicate region suspected of tumour).

the literature, with sensitivities ranging from 80% to 100% and specificities from 63% to 100% [12, 26–28, 39, 40].

Remarkably, no significant correlation between the accuracy and the experience of the observer was found. Also, no correlation was found between experience of the observers and the number of non-conclusive reports. At first glance, these findings may suggest that no specific experience is needed with FDG-PET for laryngeal carcinoma. Another explanation for this finding could be the lack of clinical feedback during daily practice in which a learning curve cannot be established. Therefore, regular feedback during daily practice seems essential also in situations where proof of presence or absence of disease may be obtained several months after PET. Finally, we recognise that the sample size was relatively small and that some observers reported that they were unfamiliar with interpretation of images without attenuation correction. However, the performance of these observers was not clearly different from the others. Moreover, considering the 95% confidence intervals of the correlation coefficients, it seems unlikely that a larger sample size would change these findings. Unfortunately, it was not possible to compare the correlations between accuracy and the experience of the observer or the number of equivocal scores with previous studies, as previous studies used two or three experienced nuclear physicians without differentiation of the level of experience or the relation with equivocal scores. Zijlstra et al. [38] reported that the experts did not have any equivocal scores, while the less experienced observers did have equivocal scores.

A variable amount of cases were scored equivocal (a median of 17% equivocal reports per observer). Although the number of non-conclusive scores differed greatly per observer (range 0–10), this indicates that in contrast to how data are typically reported, dichotomous results of FDG-PET for recurrent carcinoma may be regarded as an artificial and unwanted simplification. To explore the effect on diagnostic performance of this phenomenon, we dichotomised the data. The conservative strategy in which the equivocal scores were analysed as negative resulted in a better overall accuracy and a better interobserver agreement (κ 0.59 and 0.58) than the sensitive strategy (κ 0.43 and 0.51).

In our population, the prevalence of histologically proven recurrence was 23%. Because the mean reported prevalence is 50% [14, 25, 27], we compared these prevalences in a Bayesian plot. When the prevalence is 50%, the difference between the two strategies for a negative PET scan, in favour of the sensitive strategy, is larger as compared to the prevalence in the present study.

We assume that in clinical practice, the sensitive reading is used if FDG-PET is used to select patients suspected of recurrent laryngeal carcinoma after radiotherapy for direct laryngoscopy under general anaesthesia. For the physician, the risk of missing a recurrence probably outweighs a futile direct laryngoscopy because early detection of a recurrence can be important for salvage surgery and clinical outcome. An inherent disadvantage of sensitive reading is the higher

percentage of false positives and subsequently futile direct laryngoscopies under general anaesthesia and more interobserver variability. It can be expected that the interobserver agreement has improved by the feedback received after the assessment.

We used a disease-free follow-up of 6 months as reference standard of patients without recurrence because we assume that local disease manifest itself within this period. Extending this period carries the risk to include recurrent disease that developed after the PET scan. If local recurrences were not detected within the first 6 months, these were diagnosed at least 21 months after the PET scan. It seems highly unlikely that the lead-time of PET would be that long, but we admit that we cannot exclude the possibility of a very slow growing recurrence.

As was shown in the Kaplan Meier analyses (Fig. 3), a negative PET scan was highly predictive for local control, especially in the first 12 months. This suggests that patients with a negative PET scan might be spared a futile laryngoscopy under general anaesthesia and that regular follow-up might be sufficient.

While false-positive reading tends to be a problem, the negative predictive value of FDG-PET is high in both the conservative (96%) and the sensitive strategy (99%). The negative predictive value is, of course, dependent on the prevalence of disease. In the present study, the prevalence was only 23%. In the PET literature, the mean prevalence appears to be about 50% (manuscript in preparation), and high negative predictive values for different prevalence are reported [41]. Therefore, it can be anticipated that a negative FDG-PET excludes recurrent disease with a high certainty. With this unique characteristic, FDG-PET may be safely used as the first diagnostic step of triage for invasive procedures in patients suspected of recurrent laryngeal tumour. By filtering the patients with a negative PET scan out of the further diagnostic process, the percentage of futile diagnostic laryngoscopies can probably be diminished.

To further investigate the potential of FDG-PET for this indication, a prospective study with more patients is recommended. In the current study, the images were not attenuation-corrected. In the future, the fused PET-CT will probably take over the PET alone. Besides the anatomical information of the CT, it also offers the possibility to easily determine the ‘standard uptake values’ for objective assessment. For the present indication, selection of patients with suspicion of recurrent laryngeal carcinoma after radiotherapy for direct laryngoscopy under general anaesthesia, detailed anatomical information is probably not essential. Uptake in the laryngeal area, which indicates further examination, can be assessed on PET alone. For this indication, no literature is available about the diagnostic value of PET-CT in comparison with PET alone. PET-CT may yield slightly different results, and this will be subject of further study [29]. Another relative disadvantage of the present study is the varying interval between the last radiation and the PET scan, with a minimum interval of 2.4 months. McGuirt et al. [41] and

Ryan et al. [28] concluded that the accuracy of PET is significantly higher for an interval more than 3 months compared to 1 month.

Conclusions

While acknowledging that additional confirmation is necessary, we propose in view of the acceptable interobserver agreement that FDG-PET yields good negative predictive value for the detection of recurrent laryngeal carcinoma after radiotherapy. It could therefore be used as a first diagnostic step and may reduce the percentage of futile invasive diagnostics.

Acknowledgements. Due to the participation of the following nuclear physicians, it was possible to conduct a unique study on interobserver variability. We wish to thank the following observers: A.I.J. Arens, MD (AZM), J.W. van Isselt, MD, PhD (UMCU), B.L.R. Kam, MD (Erasmus MC), W. Oyen, MD, PhD (UMCN St. Radboud), J. Pruijm, MD, PhD (UMCG), M.P.M. Stokkel, MD, PhD (LUMC), G.J.J. Teule, MD, PhD (AZM), R.A. Valdes Olmos, MD, PhD (NKI/AvL) and R. Valkema, MD, PhD (Erasmus MC).

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

- Siesling S, van Dijk JA, Visser O, Coebergh JW (2003) Trends in incidence of and mortality from cancer in The Netherlands in the period 1989–1998. *Eur J Cancer* 39:2521–2530
- Bourhis J, Overgaard J, Audry H et al (2006) Hyperfractionated or accelerated radiotherapy in head and neck cancer: a meta-analysis. *Lancet* 368:843–854
- Pignon JP, Bourhis J, Domenge C, Designe L (2000) Chemotherapy added to locoregional treatment for head and neck squamous-cell carcinoma: three meta-analyses of updated individual data. MACH-NC Collaborative Group. *Meta-Analysis of Chemotherapy on Head and Neck Cancer*. *Lancet* 355:949–955
- Moyer JS, Wolf GT, Bradford CR (2004) Current thoughts on the role of chemotherapy and radiation in advanced head and neck cancer. *Curr Opin Otolaryngol Head Neck Surg* 12:82–87
- Hillman RE, Walsh MJ, Fisher SG, Wolf GT, Hong WK (1998) Functional outcomes following treatment for advanced laryngeal cancer. Part I—voice preservation in advanced laryngeal cancer. Part II—laryngectomy rehabilitation: the state of the art in the VA system. *Ann Otol Rhinol Laryngol* 107:2–27
- Terrell JE, Fisher SG, Wolf GT (1998) Long-term quality of life after treatment of laryngeal cancer. *Arch Otolaryngol Head Neck Surg* 124:964–971
- Di Martino E, Nowak B, Hassan HA et al (2000) Diagnosis and staging of head and neck cancer—a comparison of modern imaging modalities (positron emission tomography, computed tomography, color-coded duplex sonography) with panendoscopic and histopathologic findings. *Arch Otolaryngol Head Neck Surg* 126:1457–1461
- Gussack GS, Hudgins PA (1991) Imaging modalities in recurrent head and neck tumors. *Laryngoscope* 101:119–124
- Harnsberger HR, Mancuso AA, Muraki AS, Parkin JL (1983) The upper aerodigestive tract and neck—CT evaluation of recurrent tumors. *Radiology* 149:503–509
- McGuirt WF, Greven K, Williams D et al (1998) PET scanning in head and neck oncology: a review. *Head Neck* 20:208–215
- Nowak B, Di Martino E, Janicke S et al (1999) Diagnostic evaluation of malignant head and neck cancer by F-18-FDG PET compared to CT/MRI. *Nuklearmedizin* 38:312–318
- Bongers V, Terhaard CJ, van Isselt JW, Hordijk GJ, van Rijk PP (2000) Dual-head FDG-PET for the detection of recurrent laryngeal cancer compared with histopathological biopsy results and minimally 1 year clinical follow-up. *J Nucl Med* 41:287P
- Wong RJ, Lin DT, Schoder H et al (2002) Diagnostic and prognostic value of [(18)F]fluorodeoxyglucose positron emission tomography for recurrent head and neck squamous cell carcinoma. *J Clin Oncol* 20:4199–4208
- Greven KM, Williams DW, Keyes JW, McGuirt WF, Watson NE, Case LD (1997) Can positron emission tomography distinguish tumor recurrence from irradiation sequelae in patients treated for larynx cancer? *Cancer J Sci Am* 3:353–357
- Brouwer J, Bodar EJ, de Bree R et al (2004) Detecting recurrent laryngeal carcinoma after radiotherapy: room for improvement. *Eur Arch Otorhinolaryngol* 261:417–422
- Sigg MB, Steinert H, Gratz K, Hugenin P, Stoeckli S, Eyrych GK (2003) Staging of head and neck tumors: [F-18]fluorodeoxyglucose positron emission tomography compared with physical examination and conventional imaging modalities. *J Oral Maxillofac Surg* 61:1022–1029
- Fischbein NJ, Aassar OS, Caputo GR et al (1998) Clinical utility of positron emission tomography with F-18-fluorodeoxyglucose in detecting residual/recurrent squamous cell carcinoma of the head and neck. *ANJR Am J Neuroradiol* 19:1189–1196
- Hanasono MM, Kunda LD, Segall GM, Ku GH, Terris DJ (1999) Uses and limitations of FDG positron emission tomography in patients with head and neck cancer. *Laryngoscope* 109:880–885
- Lowe VJ, Boyd JH, Dunphy FR et al (2000) Surveillance for recurrent head and neck cancer using positron emission tomography. *J Clin Oncol* 18:651–658
- Muylle K, Castaigne C, Flamen P (2005) F-18-fluoro-2-deoxy-D-glucose positron emission tomographic imaging: recent developments in head and neck cancer. *Curr Opin Oncol* 17:249–253
- Kitagawa Y, Nishizawa S, Sano K et al (2003) Prospective comparison of F-18-FDG PET with conventional imaging modalities (MRI, CT, and Ga-67 scintigraphy) in assessment of combined intraarterial chemotherapy and radiotherapy for head and neck carcinoma. *J Nucl Med* 44:198–206
- Rusthoven KE, Koshy M, Paulino AC (2005) The role of PET–CT fusion in head and neck cancer. *Oncology (Williston Park)* 19:241–246
- Goerres GW, Schmid DT, Bandhauer F et al (2004) Positron emission tomography in the early follow-up of advanced head and neck cancer. *Arch Otolaryngol Head Neck Surg* 130:105–109
- Lowe VJ, Dunphy FR, Varvares M et al (1997) Evaluation of chemotherapy response in patients with advanced head and neck cancer using [F-18]fluorodeoxyglucose positron emission tomography. *Head Neck* 19:666–674
- McGuirt WF, Greven KM, Williams DW, Keyes JW, Watson N (1998) Laryngeal radionecrosis versus recurrent cancer: a clinical approach. *Ann Otol Rhinol Laryngol* 107:293–296
- Stokkel MP, Terhaard CH, Mertens IJ, Hordijk GJ, van Rijk PP (1998) Fluorine-18-FDG detection of laryngeal cancer postradiotherapy using dual-head coincidence imaging. *J Nucl Med* 39:1385–1387
- Terhaard CH, Bongers V, van Rijk PP, Hordijk GJ (2001) F-18-fluoro-deoxy-glucose positron-emission tomography scanning in detection of local recurrence after radiotherapy for laryngeal/pharyngeal cancer. *Head Neck* 23:933–941
- Ryan WR, Fee WE Jr, Le QT, Pinto HA (2005) Positron-emission tomography for surveillance of head and neck cancer. *Laryngoscope* 115:645–650
- de Bree R, van der Putten L, Hoekstra OS et al (2007) A randomized trial of PET scanning to improve diagnostic yield of direct laryngoscopy in patients with suspicion of recurrent laryngeal carcinoma after radiotherapy. *Contemp Clin Trials* 28:705–712
- Joshi U, Raijmakers PG, Riphagen II, Teule GJ, van LA, Hoekstra OS (2007) Attenuation-corrected vs. nonattenuation-corrected 2-deoxy-2-[F-18]fluoro-D-glucose-positron emission tomography in oncology: a systematic review. *Mol Imaging Biol* 9:99–105
- Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. *Biometrics* 33:159–174
- Jones TD, Cheng L (2006) Papillary urothelial neoplasm of low malignant potential: evolving terminology and concepts. *J Urol* 175:1995–2003
- Christensen JA, Fletcher JG, Fidler JL et al (2004) Intraductal papillary mucinous neoplasms of the pancreas: CT patterns of recurrence and

- multiobserver performance in detecting recurrent neoplasm after surgical resection. *AJR Am J Roentgenol* 183:1367–1374
34. Fakhry N, Lussato D, Jacob T, Giorgi R, Giovanni A, Zanaret M (2007) Comparison between PET and PET/CT in recurrent head and neck cancer and clinical implications. *Eur Arch Otorhinolaryngol* 264:531–538
 35. Bohdiewicz PJ, Wong CY, Kondas D, Gaskill M, Dworkin HJ (2003) High predictive value of F-18 FDG PET patterns of the spine for metastases or benign lesions with good agreement between readers. *Clin Nucl Med* 28:966–970
 36. Lim JS, Kim MJ, Yun MJ et al (2006) Comparison of CT and 18F-FDG pet for detecting peritoneal metastasis on the preoperative evaluation for gastric carcinoma. *Korean J Radiol* 7:249–256
 37. Hashimoto Y, Tsujikawa T, Kondo C et al (2006) Accuracy of PET for diagnosis of solid pulmonary lesions with 18F-FDG uptake below the standardized uptake value of 2.5. *J Nucl Med* 47:426–431
 38. Zijlstra JM, Comans EF, van Lingen A et al (2007) FDG PET in lymphoma: the need for standardization of interpretation. An observer variation study. *Nucl Med Commun* 28:798–803
 39. Alvarez Perez RM, Borrego DI, Vazquez AR, Ruiz Franco-Baux J, Ceballo Pedraja JM, Esteban OF (2006) Evaluation of efficacy and clinical impact of positron emission tomography with 18F fluoro-deoxyglucose (FDG) in patients with suspicion of recurrent laryngeal carcinoma. *Acta Otorrinolaringol Esp* 57:134–139
 40. Schoder H, Yeung HWD (2004) Positron emission imaging of head and neck cancer, including thyroid carcinoma. *Sem Nucl Med* 34:180–197
 41. McGuiert WF, Greven KM, Keyes JW et al (1995) Positron emission tomography in the evaluation of laryngeal carcinoma. *Ann Otol Rhinol Laryngol* 104:274–278