



FDG-PET/CT for treatment response assessment in head and neck squamous cell carcinoma: a systematic review and meta-analysis of diagnostic performance

Nils Helsen^{1,2} · Tim Van den Wyngaert^{1,2} · Laurens Carp^{1,2} · Sigrid Stroobants^{1,2}

Received: 26 September 2017 / Accepted: 12 February 2018 / Published online: 24 February 2018
© Springer-Verlag GmbH Germany, part of Springer Nature 2018

Abstract

Purpose 18-fluorodeoxyglucose positron emission tomography combined with computed tomography (FDG-PET/CT) is increasingly used to evaluate treatment response in head and neck squamous cell carcinoma (HNSCC). This analysis assessed the diagnostic value of FDG-PET/CT in detecting nodal disease within 6 months after treatment, considering patient and disease characteristics.

Methods A systematic review was performed using the MEDLINE and Web of Knowledge databases. The results were pooled using a bivariate random effects model of the sensitivity and specificity.

Results Out of 22 identified studies, a meta-analysis of 20 studies (1293 patients) was performed. The pooled estimates of sensitivity, specificity and diagnostic odds ratio (with 95% CI) were 85% (76–91%), 93% (89–96%) and 76 (35–165), respectively. With the prevalence set at 10%, the positive and negative predictive values were 58% and 98%. There was significant heterogeneity between the trials ($p < 0.001$). HPV positive tumors were associated with lower sensitivity (75% vs 89%; $p = 0.01$) and specificity (87% vs 95%; $p < 0.005$).

Conclusion FDG-PET/CT within 6 months after (chemo)radiotherapy in HNSCC patients is a reliable method for ruling out residual/recurrent nodal disease and obviates the need for therapeutic intervention. However, FDG-PET/CT may be less reliable in HPV positive tumors and the optimal surveillance strategy remains to be determined.

Keywords FDG-PET/CT · Head and neck cancer · Chemoradiotherapy · HPV · Surveillance

Introduction

Head and neck cancer is the sixth most common malignancy worldwide [1]. About 60% of patients are diagnosed with locally advanced disease. Treatment of these patients usually consists of a combination of radiotherapy with chemotherapy

or targeted agents [2, 3]. For many years, planned neck dissection and radiotherapy was the preferred treatment for patients with nodal involvement [4]. Nowadays, neck dissection is no longer routinely performed in patients with a complete response after chemoradiation of initially N1 disease, and is more and more questioned in patients with initial N2–N3 disease at diagnosis [5]. The recurrence rate for node positive disease is still about 35–45%, and most nodal relapses occur within 2 years after treatment [6, 7]. Therefore, the use of reliable imaging methods for identifying residual nodal disease allowing for timely salvage surgery may support the waning use of neck dissections in patients with a complete response. However, this remains a matter of debate as others argue that neck dissection improves disease control regardless of the response in the neck [8].

Functional imaging with ¹⁸F-fluorodeoxyglucose (FDG) positron emission tomography (PET) is less hampered by the altered anatomy compared to CT and MRI, and has proved

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00259-018-3978-3>) contains supplementary material, which is available to authorized users.

✉ Nils Helsen
nils.helsen@uza.be

¹ Department of Nuclear Medicine, Antwerp University Hospital, Wilrijkstraat 10, 2650 Edegem, Belgium

² Faculty of Medicine and Health Sciences, University of Antwerp, Wilrijk, Belgium

to be valuable in distinguishing residual nodal disease from therapy-induced changes. Previous meta-analyses have established the favorable diagnostic test characteristics of FDG-PET and FDG-PET/CT for evaluating residual nodal disease, even though considerable heterogeneity was present between studies [9, 10]. However, the technique has improved considerably since its introduction due to technical advances enabling integration of PET and CT devices, improvements in detector capabilities yielding higher image resolution, optimization of head and neck images acquisition parameters, and combining PET with diagnostic resolution contrast enhanced CT scans [11–19]. Likewise, the identification of the human papilloma virus (HPV) as an important cause and prognostic factor of oropharyngeal cancers has greatly increased the understanding of the underlying tumor biology. Taken together, these recent advances warrant a reappraisal of the current literature focusing on integrated FDG-PET/CT in HNSCC within 6 months after treatment, with special emphasis on factors that may impact the technique (scanning parameters, patient characteristics, HPV status and follow-up).

Methods

Studies were identified through a systematic electronic search of PubMed and Web of Science databases on 14/11/2016. The following search criteria were used: “(pet/ct or pet-ct or (“Positron-Emission Tomography” and “Tomography, X-Ray Computed”)) and (“head and neck” or “Head and Neck Neoplasms”) and (response or evaluation or therapy or surveillance)”. We also examined the references from all studies that were retrieved in full text. Abstracts or posters were excluded.

The eligibility criteria were: 1) the use of FDG-PET/CT for the detection of recurrent/residual nodal disease within 6 months after therapy (based on the reported central tendency measures), 2) patients with squamous cell carcinoma of the head and neck treated with radiotherapy with or without chemotherapy or targeted agents (surgery prior to radiotherapy was allowed as used in oral cavity tumors), 3) 2×2 tables should be extractable for the neck nodes. Simulation was used to estimate the proportion of patients who had PET/CT imaging beyond 6 months after the end of treatment. Briefly, with the reported or estimated mean and standard deviation from the individual studies, the normal distribution was used to generate datasets with the same size as the original studies and the proportion of simulated interval times exceeding 6 months was calculated [20]. After 50,000 iterations, the 99% confidence interval of this proportion was calculated and the upper boundary was used to estimate the potential impact of these patients on the analysis results. Studies with less than 15 patients, patients below 18 years or studies in languages other than English or French were excluded.

Study selection was performed by the first author, but in case of doubt, a consensus was sought among all authors.

For every study, the following variables were extracted: patient demographics (age, sex), disease characteristics (nodal status, HPV status, initial vs. recurrent disease), study characteristics (retro- vs. prospective), number of patients, treatment, follow-up, reference standard, prevalence, and imaging protocol (scanning parameters, number of weeks after end of treatment, use of contrast enhanced CT scans and acquisition of dedicated head and neck images). In case of missing data an attempt was made to contact the investigators for further information. If HPV prevalence was not reported, the country-specific HPV prevalence was estimated using data from Stein et al. [21]. Using the reported or estimated prevalence of HPV and the number of patients with an oropharyngeal malignancy, the percentage of patients with an HPV associated malignancy relative to the entire study population was estimated for every study.

The QUADAS 2 (quality assessment tool for diagnostic accuracy studies) system was used to assess the methodological quality of selected studies, but was not used to exclude studies [22–24].

From the individual studies, the reported 2×2 tables of FDG-PET/CT outcomes versus the reference standard were extracted to allow the bivariate estimation of the pooled sensitivity and specificity. This method uses a random effects model for both sensitivity and specificity to compensate for observed heterogeneity beyond chance caused by varying clinical and methodological aspects of the selected studies. Also, this approach adjusts for any differences in study size and the possible negative correlation between sensitivity and specificity of FDG-PET/CT that may be caused by varying thresholds in image interpretation. Model diagnostics were performed as required [25, 26]. From the pooled estimates of sensitivity and specificity, the mean positive and negative likelihood ratios were calculated. A hierarchical summary receiver operator characteristics curve (HSROC) was constructed to estimate the pooled area under the curve (AUC) of FDG-PET/CT.

Small study effects (e.g. publication bias) were assessed by constructing a scatter plot of the inverse of the square root of the effective sample size versus the log of the diagnostic odds ratio. Meta-regression was performed if more than 10 studies were identified and if the overall I^2 exceeded 50%.

Results

Systematic review

A total of 1483 references were identified, of which 1417 were discarded based on the title or abstract, because they were not related to the evaluation of FDG-PET/CT in head and neck

cancer (Fig. 1). Another 45 studies were discarded based on the abstract, as they did not meet the eligibility criteria. An additional 3 were excluded because they were not in English or French. In all, 66 references were retrieved in full text and analyzed for potential inclusion. One additional study, performed at our hospital, was included in this analysis [27].

In all, 22 studies reporting on a total of 1423 patients were included in this analysis (Table 1) [11–13, 15–17, 19, 27–41]. The mean prevalence of oropharyngeal tumors was 70.3% and 6 out of 22 studies (27.2%) reported the prevalence of HPV. The median percentage of HPV positive patients in the included studies was estimated to be 38.5%. The primary and secondary endpoints of the included studies were diverse and included the evaluation of the diagnostic performance of FDG-PET/CT, the comparison of FDG-PET/CT with conventional imaging techniques, the ability to avoid neck dissections, or assessed which patient population would derive most benefit from a post-treatment FDG-PET/CT scan. While the reported medians and means of the time between the end of treatment and FDG-PET/CT imaging was well below the cut-off of 6 months for all included studies, the upper ranges for this variable crossed the 6 month threshold in 3 studies [29, 30, 36]. However, it was estimated that less than 0.5% of patients in the pooled dataset were scanned beyond 6 months and it was considered justified to keep these studies in the analysis.

A detailed overview of the QUADAS 2 scores is reported in the supplementary appendix (Table S1). Briefly, there was a moderate risk of bias due to the exclusion of patients (“flow and timing” domain). Also, confirmation of positive nodal disease at the time of diagnosis and prior to therapy remains an important weakness in most included studies (“patient selection” domain), which may hamper answering the research question of this review (“applicability”).

Meta-analysis

A funnel plot was created to assess the presence of publication bias (Fig. S2), revealing no evidence of small study effects ($p = 0.56$). Crude pooled estimates of sensitivity, specificity, positive and negative likelihood ratio, diagnostic odds ratio, and AUC were 82% (95% CI 72–89%), 92% (95% CI 87–95%), 10.3 (95% CI 5.9–18.1), 0.19 (95% CI 0.12–0.31), 54 (95% CI 22–131), and 0.93 (95% CI 0.90–0.95). However, model diagnostics identified the studies of Gourin et al. and Vainstein et al. as outliers (Cook’s distance of 1.98 and 2.22, respectively) [11, 40]. The meta-analysis was subsequently performed without these studies, yielding a total of 20 studies and 1293 patients [12, 13, 15–17, 19, 27–39, 41].

The pooled estimate without outliers of sensitivity, specificity, positive and negative likelihood ratio, diagnostic odds ratio, and AUC were 85% (95% CI 76–91%), 93% (95% CI 89–96%), 12.4 (95% CI 7.4–20.8), 0.16 (95% CI 0.10–0.27), 76 (95% CI 35–165), and 0.94 (95% CI 0.91–0.95) (Fig. 2). The negative and positive predictive value (NPV and PPV) are considered clinically most useful, yet depend on the probability of residual nodal disease, i.e. disease prevalence. Figure 3 represents a nomogram-style plot of the pooled estimate of NPV and PPV as a function of disease prevalence. Given a pre-test probability of disease of 10%, the NPV and PPV are estimated at 98% and 58%, respectively. Moreover, Fig. 3 demonstrates the important increase in positive predictive value (58, 76 and 84%) with increasing pre-test probabilities (10, 20 and 30%). In contrast, the negative predictive value remained rather stable (98, 96 and 93%) across a broad range of pre-test probabilities (10, 20 and 30%).

The heterogeneity in reported diagnostic test characteristics between the different trials was larger than can be explained by chance alone, both overall ($I^2 = 80\%$; 95% CI 56–100%; $p = 0.004$) and for specificity ($I^2 = 77\%$; 95% CI 66–87%; $p <$

Fig. 1 Flow chart of the study for the selection of the systematic review

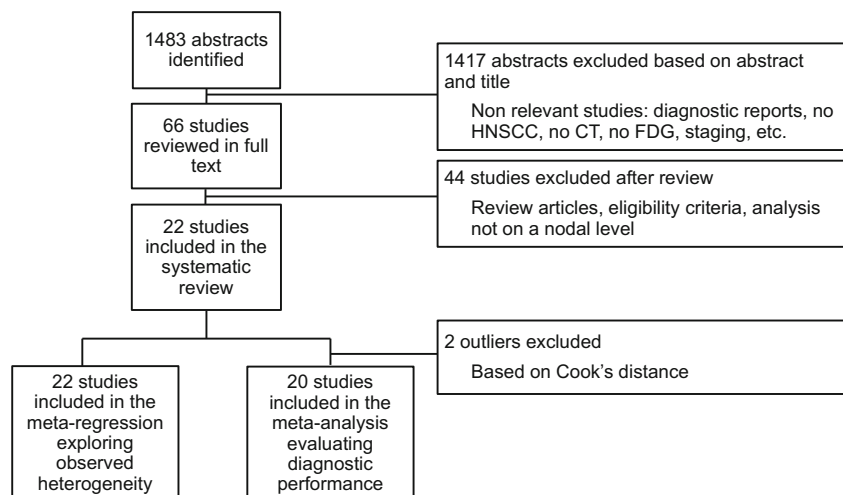


Table 1 Characteristics of the studies

Author, year	Study aim	Design	N° of patients	N+ (%)	Treatment-PET interval (weeks)	Follow-up (months)	PET LN interpretation criteria	Reference standard
Bird, 2016 [28]	Treatment response assessment for advanced oropharynx HNSCC	Retro	146	90	12.4 ^b (4–22)	25 ^b	Qualitative assessment	Pos pathology (14) Clinical follow-up
Chan, 2012 [17]	Treatment response assessment in HPV+ patients	Retro	67	100	13 ^a	26 ^b	SUV ≥ 2.5	Pos pathology (5) Clinical follow-up of at least 12 months
Chen, 2006 [19]	Comparison PET/CT and CT	Pro	30	97	7 ^a (2–24)	>9	Qualitative assessment with comparison to background	NR Clinical follow-up of at least 6 months
Cho, 2009 [16]	Avoidance of neck dissection in N2 patients	Retro	48	100	>8	26 ^b	Qualitative assessment	Pos pathology (6) Neck dissection (10) and yearly follow-up (28)
Comnel, 2007 [29]	Treatment response assessment	Pro	28	NR	14 ^b (6–27)	28 ^b	Qualitative assessment	Pos pathology FP: no progression within 3 months TN: clinical follow-up of at least 12 months Pos pathology (5) Clinical follow-up
Fahkry, 2006 [13]	Treatment response assessment for advanced HNSCC	Pro	61	75	14 ^a (9–21)	7 ^a	Qualitative assessment	Pos pathology (10) Neg pathology (22)
Gourin, 2009 [11]	Treatment response assessment for advanced HNSCC	Retro	32	100	(8–11)	20 ^a	Qualitative assessment with comparison to background	Pos pathology (8) Clinical follow-up of at least 12 months
Gupta, 2010 [30]	Treatment response assessment	Pro	57	65	9 ^b (5–29)	26 ^b	Qualitative assessment with comparison to background	Pos pathology (13), imaging (5), clinical progression (3) Clinical follow-up of 12 months Pos pathology (12) Clinical follow-up
Helsen, 2017 [27]	Treatment response assessment	Retro	99	63	13 ^b (5–19)	28 ^a	Qualitative assessment	NR Clinical follow-up Pos pathology (4) Clinical follow-up
Keski-Säntti, 2014 [31]	Treatment response assessment for clinically N+ HNSCC	Retro	54	100	13 ^b (10–18)	26 ^b	Qualitative assessment	Pos pathology (10) Clinical follow-up of at least 5 months
Loo, 2011 [32]	Avoidance of neck dissection in N2 patients	Retro	34	100	13	39 ^b	Qualitative assessment	NR Clinical follow-up Pos pathology (4) Clinical follow-up
Malone, 2009 [33]	Treatment response assessment for advanced HNSCC	Retro	31	68	5.8 ^b (5.5–8.0)	24 ^b	Qualitative assessment with comparison to background	Pos pathology (10) Clinical follow-up or negative pathology
Moeller, 2009 [12]	Treatment response assessment for advanced HNSCC	Pro	98	84	8 (5–12)	21 ^b	Qualitative interpretation	Pos pathology (5) clinical progression (3) Clinical follow-up of at least 5 months
Nayak, 2007 [15]	Treatment response assessment for advanced HNSCC	Pro	43	100	(8.5–21)	18 ^b	Qualitative assessment	Pos pathology (8) or clinical progression (1) Clinical follow-up of at least 12 months
Porceddu, 2011 [34]	Avoidance of neck dissection in N+ patients	Pro	112	100	12 ^b (10–14)	28 ^b	Qualitative assessment with comparison to background	Clinical progression (4) or pos pathology (1) Clinical follow-up
Prestwich, 2012 [35]	Treatment response assessment for advanced HNSCC	Retro	44	97	17 ^b (9–24)	14 ^b	Qualitative assessment with comparison to background	Pos pathology (6), PET/CT follow-up (4) Clinical follow-up of at least 9 months or serial PET/CT
Rabalais, 2009 [36]	Treatment response assessment for advanced HNSCC	Retro	52	100	11.8 ^b (3.7–29.4)	18 ^a	Qualitative assessment	Pos pathology (3) Absence of pos histology + regression size LN
Schouten, 2015 [37]	Treatment response assessment in N+ patients	Retro	58	100	13 ^a	9	Qualitative assessment with comparison to background	

Table 1 (continued)

Author, year	Study aim	Design	N° of patients	N+ (%)	Treatment-PET interval (weeks)	Follow-up (months)	PET LN interpretation criteria	Reference standard
Sjövall, 2016 [38]	Treatment response assessment	Retro	105	100	6 ^b (5–12)	42 ^b	Qualitative assessment with comparison to background	Pos pathology Clinical follow-up
Slevin, 2015 [39]	Treatment response assessment	Retro	93	100	17.4 ^b (9–24)	25 ^b	Qualitative assessment with comparison to background	Clinicopathological evidence Clinical follow-up
Vainshtein, 2013 [40]	Treatment response assessment for HPV+ patients	Retro	98	97	13 ^b	44 ^b	Qualitative assessment	Pos pathology Clinical follow-up
Zundel, 2011 [41]	Comparison PET and physical examination	Retro	52	69	(17–26)	58 ^b	SUV ≥ 2.5 + node on CT	NR Clinical follow-up of at least 29 months

^a Mean

^b Median

^c The values between brackets in the reference column are the numbers of patients with that reference standard. The reference above/under the solid line is the reference standard for positive and negative cases, respectively

N+: patients with positive lymph nodes, NR: not reported, LN: Lymph node, TN: true negative, FP: false positive, HPV: human papillomavirus, SUV: Standardized uptake value

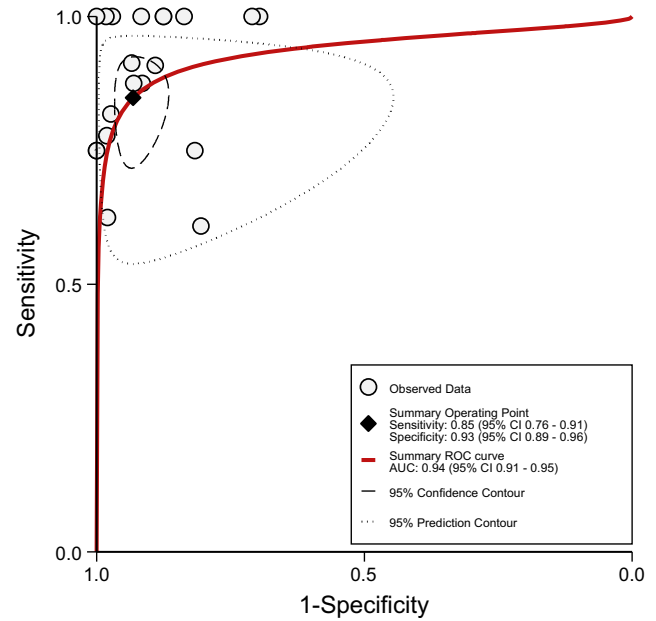


Fig. 2 Hierarchical summary receiver-operating curve (HSROC) along with the summary operating point and 95% confidence and prediction region

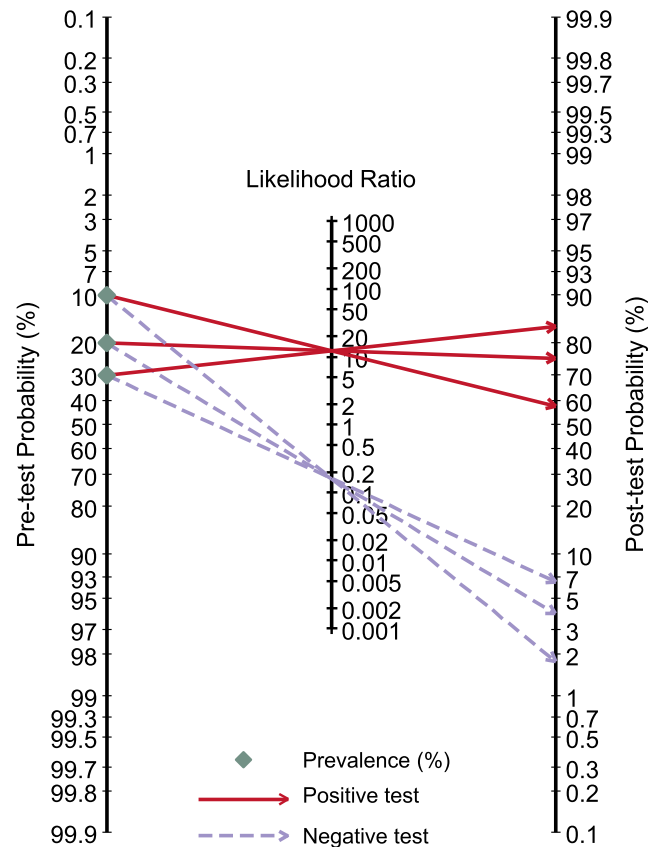


Fig. 3 The post-test probability of residual/recurrent nodal disease after a positive/negative FDG-PET/CT result with the prevalence of nodal disease set at 10, 20 and 30%. The red line shows that a patient will have a 58% chance of having residual nodal disease in the neck if the FDG-PET/CT scan is positive and 2% if the scan is negative if the pre-test probability is 10%

0.001). A subsequent meta-regression was performed using all available studies ($n = 22$) in an attempt to identify contributing factors. Of all study quality associated factors, those related to the index test (e.g. timing, acquisition, and reporting of the FDG-PET/CT study) were found to be associated with reported specificity, with lower values in studies with a higher risk of bias (77% vs 94%; $p < 0.005$) or a lower degree of applicability (82% vs 93%; $p = 0.01$). With respect to patient related items, both a higher percentage of oropharyngeal tumors and HPV positive malignancies were associated with lower sensitivity (76% vs 87% and 75% vs 89%; both $p = 0.01$) and specificity (85% vs 96% and 87% vs 95%; both $p < 0.005$). Interestingly, the differential effect of HPV status on sensitivity and specificity remained when analyzed separately in studies with a prevalence of oropharyngeal tumors below (between HPV group difference in sensitivity = 9% and specificity = 9%) and above (between HPV group difference in sensitivity = 17% and specificity = 7%) the overall median. In contrast, within the HPV group, differences across anatomic sites did not exceed 6% for either sensitivity or specificity. This suggests that HPV status - rather than anatomic site - drives the observed differences in diagnostic test characteristics. No other patient- or disease-related parameters were found to impact the reported study characteristics (including patient characteristics, scanning parameters, study design and follow-up).

Discussion

This meta-analysis demonstrates that FDG-PET/CT is a reliable technique for detecting residual nodal neck disease within the first 6 months after treatment, with a pooled sensitivity and specificity of 85% (95% CI, 76–91%) and 93% (95% CI, 89–96%). In contrast to the meta-analyses of Isles and Gupta, this analysis excluded studies that imaged patients more than 6 months after therapy to increase the clinical applicability of the results [9, 10]. The most important clinically relevant finding was the high negative predictive value. A negative PET/CT scan is therefore highly indicative for the absence of disease obviating further therapeutic interventions. Recently the PET-NECK trial, a randomized phase III trial, compared FDG-PET/CT guided active surveillance with standard neck dissections in 564 patients with stage N2 and N3 disease [42]. The trial was able to demonstrate non-inferior survival outcomes compared to routine neck dissections, with only 20% of the patients in the surveillance arm receiving a neck dissection. The use of FDG-PET/CT surveillance also resulted in less serious adverse events and was cost-effective compared to routine neck dissection. While the PET-NECK study has provided remarkable proof-of-principle for FDG-PET/CT surveillance, it did not specifically address optimal image acquisition techniques and interpretation criteria, nor

identify subgroups who would benefit most from FDG-PET/CT surveillance, and was limited to node positive disease only.

Impact of integrated multimodality imaging

Compared to the meta-analysis of Isles et al. based on standalone FDG-PET, the present review included only studies using integrated multimodality PET/CT scanners and found higher point estimates for sensitivity (85% vs 74%) and specificity (93% vs 88%), even though the 95% confidence intervals overlap (Table 2) [9, 10]. Fakhry et al. reported a higher accuracy for FDG-PET/CT than standalone FDG-PET, attributable to a higher specificity because of improved anatomical localization, which decreased the number of false positive and equivocal findings [43]. In the study of Goshen et al. FDG-PET/CT decreased the number of equivocal PET findings with 60% in the initial staging and evaluation of suspected recurrent HNSCC [44].

In contrast, the meta-analysis of Gupta et al., which included studies on FDG-PET with and without CT, did not find significant differences between devices for the evaluation of the nodes in the neck ($p = 0.8$) [10]. Therefore, other advances may have contributed to the increased diagnostic performance over time, including optimized scanning protocols, technical advances in scanning equipment (e.g. TOF-PET, dedicated head and neck protocols), and the increasing experience of nuclear medicine physicians in this field. It remains an open question how much each factor has contributed to this improvement.

Effect of timing after therapy

The timing of FDG-PET/CT imaging after completion of therapy is important, as FDG-PET/CT may be less reliable when performed in the first weeks after treatment [9, 10]. The magnitude and dynamics of this effect have recently been elaborated in the study of Helsen et al. [27] which confirmed that the diagnostic performance increases until 11 weeks after treatment and reaches a plateau thereafter. However, the current analysis did not demonstrate an association between the number of weeks after therapy and the sensitivity or specificity of FDG-PET/CT. This may be due to ambiguous reporting of this parameter between the studies and to the low variability of this parameter in the current dataset (mean 12 weeks; 95% CI 10–13).

Impact of HPV status

Our results suggest a lower diagnostic performance of FDG-PET/CT in HPV positive tumors. At the extreme end, the

Table 2 Comparison of the current meta-analyses

Technique	N° of patients	Eligibility criteria	Sensitivity (95% CI)	Specificity (95% CI)	NPV (95% CI)	PPV (95% CI)
Isles [9] Standalone PET	658	FDG-PET after (chemo)radiotherapy for HNSCC	74% (50–89%)	88% (74–95%)	96% ^a (84–99%)	49% ^a (29–70%)
Gupta [10] Standalone PET + Integrated PET/CT	1525	FDG-PET/CT after (chemo)radiotherapy for HNSCC	73% (67–78%)	88% (86–89%)	95% ^b (93–96%)	52% ^b (47–58%)
Current analysis Integrated PET/CT	1293	FDG-PET/CT within 6 months after (chemo)radiotherapy for HNSCC	85% (76–91%)	93% (89–96%)	98% ^c (97–99%)	58% ^c (52–63%)

^a Reported prevalence of 13%

^b Reported prevalence of 15%

^c For a prevalence of 10%

study by Vainstein et al. reported a sensitivity and specificity of only 25% (3–65%) and 82% (73–90%), respectively [40]. Data reported by Moeller et al. show a similar trend with a lower accuracy of FDG-PET/CT in low risk patients, which included the HPV status [12]. Although there are distinct morphologic and glycolytic differences between HPV positive and negative tumors, FDG avidity of both cancers is comparable and cannot explain the difference in diagnostic performance [45–47]. HPV-positive patients have a better outcome compared to HPV-negative ones irrespective of the treatment choice, but are also more radiosensitive [48]. Therefore, repopulation of resistant cells may take longer before they can be detected by PET imaging, resulting in lower sensitivities early after the end of radiotherapy. The lower specificity can be explained by the increased cytotoxic T-cell based immune response reported in HPV-positive tumors, resulting in the presence of inflamed nodes that take longer to involute [49, 50]. The recent PET-NECK trial, however, did not find a difference in overall survival between surveillance and neck dissection, when analyzed separately based on HPV status [42]. While it is reassuring that the potentially lower sensitivity and specificity of FDG-PET/CT in HPV positive patients did not result in inferior survival, an alternate timing of surveillance may prove more appropriate in these patients. The latter would reduce post-treatment inflammation during FDG-PET/CT surveillance and could impact the number of neck dissections. Taken together, these results provide compelling evidence that the identification of HPV status as an important prognostic factor in oropharyngeal squamous cell carcinoma may have implications beyond optimal treatment selection, but may also affect subsequent imaging surveillance strategies.

Standardization of FDG-PET/CT

Issues relating to technical details of the FDG-PET/CT study were the other important source of heterogeneity in this analysis. The way of reading scans varied greatly, with some centers using standardized uptake value (SUV), while others only interpreted the images qualitatively with or without comparison to a predefined background region (Table 1). Moreover, there is no consensus on the cut-off value used when performing a quantitative analysis. In an effort to standardize image interpretation, the Hopkins criteria have been proposed, using a 5-point response interpretation comparing the SUV value of the lesion with the SUV value of the liver and the internal jugular vein [51]. Implementing such a system may contribute to reducing the variability in reporting between centers and increase the interreader agreement. Similar efforts to standardize tracer uptake time, scan acquisition parameters, and reconstruction settings can be expected to optimize the PET/CT technique, as exemplified by the EANM EARL FDG-PET/CT accreditation programme [52].

Effect of pre-test probability

Even though the negative predictive value of FDG-PET/CT was very high across a broad range of prevalences, the positive predictive value was found to vary more widely with changing pre-test probability. This confirms the need for confirmation of positive findings on FDG-PET/CT scan by cytology or biopsy, or to proceed to neck dissection as was done in the PET-NECK trial [42].

Conclusion

FDG-PET/CT within the first 6 months post-treatment is reliable for detecting residual nodal disease in patients with HNSCC and a negative scan obviates the need for further therapeutic intervention. However, in HPV positive tumors FDG-PET/CT may be less reliable and the optimal surveillance strategy in this patient population remains to be determined. Also, further standardization of the PET technique is required.

Acknowledgements We would like to thank the corresponding authors of the selected studies who provided additional information that contributed significantly to this meta-analysis.

This study was supported by a grant by the Flemish agency for innovation by science and technology (IWT-90867).

Funding This study was supported by a grant by the Flemish agency for innovation by science and technology (IWT-90867).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethics approval and consent to participate This article does not contain any studies with human participants or animals performed by any of the authors.

References

- Parkin DM, Bray F, Ferlay J, Pisani P. Global cancer statistics, 2002. *CA Cancer J Clin*. 2005;55:74–108.
- Marur S, Forastiere AA. Head and neck cancer: changing epidemiology, diagnosis, and treatment. *Mayo Clin Proc*. 2008;83:489–501.
- Petrelli F, Coinu A, Riboldi V, Borgonovo K, Ghilardi M, Cabiddu M, et al. Concomitant platinum-based chemotherapy or cetuximab with radiotherapy for locally advanced head and neck cancer: a systematic review and meta-analysis of published studies. *Oral Oncol*. 2014;50:1041–8.
- Barkley HT, Fletcher GH, Jesse RH, Lindberg RD. Management of cervical lymph node metastases in squamous cell carcinoma of the tonsillar fossa, base of tongue, supraglottic larynx, and hypopharynx. *Am J Surg*. 1972;124:462–7.
- Hamoir M, Ferlito A, Schmitz S, Hanin FX, Thariat J, Weynand B, et al. The role of neck dissection in the setting of chemoradiation therapy for head and neck squamous cell carcinoma with advanced neck disease. *Oral Oncol*. 2012;48:203–10.
- Posner M, Hershock D, Blajman C, Mickiewicz E, Winquist E, Gorbounova V, et al. Cisplatin and fluorouracil alone or with docetaxel in head and neck cancer. *N Engl J Med*. 2007;357:1705–15.
- Lavertu P, Adelstein D, Saxton J, Sestic M, Wanamaker J, Eliachar I, et al. Management of the neck in a randomized trial comparing concurrent chemotherapy and radiotherapy with radiotherapy alone in resectable stage III and IV squamous cell head and neck cancer. *Head Neck*. 1997;19:559–66.
- McHam SA, Adelstein DJ, Rybicki LA, Lavertu P, Esclamado RM, Wood BG, et al. Who merits a neck dissection after definitive chemoradiotherapy for N2-N3 squamous cell head and neck cancer? *Head Neck*. 2003;25:791–8.
- Isles MG, McConkey C, Mehanna HM. A systematic review and meta-analysis of the role of positron emission tomography in the follow up of head and neck squamous cell carcinoma following radiotherapy or chemoradiotherapy. *Clin Otolaryngol*. 2008;33:210–22.
- Gupta T, Master Z, Kannan S, Agarwal JP, Ghosh-Laskar S, Rangarajan V, et al. Diagnostic performance of post-treatment FDG PET or FDG PET/CT imaging in head and neck cancer: a systematic review and meta-analysis. *Eur J Nucl Med Mol Imaging*. 2011;38:2083–95.
- Gourin CG, Boyce BJ, Williams HT, Herdman AV, Bilodeau PA, Coleman TA. Revisiting the role of positron-emission tomography/computed tomography in determining the need for planned neck dissection following chemoradiation for advanced head and neck cancer. *Laryngoscope*. 2009;119:2150–5.
- Moeller BJ, Rana V, Cannon BA, Williams MD, Sturgis EM, Ginsberg LE, et al. Prospective risk-adjusted [18F]fluorodeoxyglucose positron emission tomography and computed tomography assessment of radiation response in head and neck cancer. *J Clin Oncol*. 2009;27:2509–15.
- Fakhry N, Barberet M, Paris J, Jacob T, Deveze A, Mundler O, et al. Intérêt de la TEP au 18FDG couplée à la TDM dans la surveillance post-thérapeutique des carcinomes épidermoïdes des voies aérodigestives supérieures. *Ann Otolaryngol Chir Cervicofac*. 2006;123:167–74.
- Ito K, Yokoyama J, Kubota K, Morooka M, Shiibashi M, Matsuda H. 18F-FDG versus 11C-choline PET/CT for the imaging of advanced head and neck cancer after combined intra-arterial chemotherapy and radiotherapy: the time period during which PET/CT can reliably detect non-recurrence. *Eur J Nucl Med Mol Imaging*. 2010;37:1318–27.
- Nayak JV, Walvekar RR, Andrade RS, Daamen N, Lai SY, Argiris A, et al. Deferring planned neck dissection following chemoradiation for stage IV head and neck cancer: the utility of PET-CT. *Laryngoscope*. 2007;117:2129–34.
- Cho AH, Shah S, Ampil F, Bhartur S, Nathan C-AO. N2 disease in patients with head and neck squamous cell cancer treated with chemoradiotherapy: is there a role for posttreatment neck dissection? *Arch Otolaryngol Head Neck Surg*. 2009;135:1112–8.
- Chan JYK, Sanguineti G, Richmon JD, Marur S, Gourin CG, Koch W, et al. Retrospective review of positron emission tomography with contrast-enhanced computed tomography in the posttreatment setting in human papillomavirus-associated oropharyngeal carcinoma. *Arch Otolaryngol Head Neck Surg*. 2012;138:1040–6.
- Ng SH, Chan SC, Yen TC, Liao CT, Lin CY, Tung-Chieh Chang J, et al. PET/CT and 3-T whole-body MRI in the detection of malignancy in treated oropharyngeal and hypopharyngeal carcinoma. *Eur J Nucl Med Mol Imaging*. 2011;38:996–1008.
- Chen AY, Vilaseca I, Hudgins PA, Schuster D, Halkar R. PET-CT vs contrast-enhanced CT: what is the role for each after chemoradiation for advanced oropharyngeal cancer? *Head Neck*. 2006;28:487–95.

20. Wan X, Wang W, Liu J, Tong T. Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range. *BMC Med Res Methodol*. 2014;14.
21. Stein AP, Saha S, Kraninger JL, Swick AD, Yu M, Lambert PF, et al. Prevalence of human papillomavirus in oropharyngeal cancer: a systematic review. *Cancer J*. 2015;21:138–46.
22. Whiting P, Rutjes A, Westwood M, Mallett S, Deeks J, Reitsma J, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 2011;155:529–36.
23. Jüni P, Witschi A, Bloch R, Egger M. The hazards of scoring the quality of clinical trials for meta-analysis. *JAMA*. 1999;282:1054–60.
24. Møller B, Weedon-Fekjaer H, Haldorsen T. Empirical evaluation of prediction intervals for cancer incidence. *BMC Med Res Methodol*. 2005;5:21.
25. Harbord RM, Deeks JJ, Egger M, Whiting P, Sterne JA. A unification of models for meta-analysis of diagnostic accuracy studies. *Biostatistics*. 2007;8:239–51.
26. Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol*. 2005;58:982–90.
27. Helsen N, Roothans D, Van Den Heuvel B, Van den Wyngaert T, van den Weyngaert D, Carp L, et al. 18F-FDG-PET/CT for the detection of disease in patients with head and neck cancer treated with radiotherapy. *PLoS One*. 2017;12:e0182350.
28. Bird T, Barrington S, Thavaraj S, Jeannon JP, Lyons A, Oakley R, et al. 18F-FDG PET/CT to assess response and guide risk-stratified follow-up after chemoradiotherapy for oropharyngeal squamous cell carcinoma. *Eur J Nucl Med Mol Imaging*. 2016;43:1239–47.
29. Connell CA, Corry J, Milner AD, Hogg A, Hicks RJ, Rischin D, et al. Clinical impact of, and prognostic stratification by, F-18 FDG PET/CT in head and neck mucosal squamous cell carcinoma. *Head Neck*. 2007;29:986–95.
30. Gupta T, Jain S, Agarwal JP, Rangarajan V, Purandare N, Ghosh-Laskar S, et al. Diagnostic performance of response assessment FDG-PET/CT in patients with head and neck squamous cell carcinoma treated with high-precision definitive (chemo)radiation. *Radiother Oncol*. 2010;97:194–9.
31. Keski-Säntti H, Mustonen T, Schildt J, Saarilahti K, Mäkitie A. FDG-PET/CT in the assessment of treatment response after oncologic treatment of head and neck squamous cell carcinoma. *Clin Med. Insights ear. Nose Throat*. 2014;19:25–9.
32. Loo SW, Geropantans K, Beadsmoore C, Montgomery PQ, Martin WM, Roques TW. Neck dissection can be avoided after sequential chemoradiotherapy and negative post-treatment positron emission tomography-computed tomography in N2 head and neck squamous cell carcinoma. *Clin Oncol*. 2011;23:512–7.
33. Malone JP, Gerber MA, Vasireddy S, Hughes LF, Rao K, Shevlin B, et al. Early prediction of response to chemoradiotherapy for head and neck cancer: reliability of restaging with combined positron emission tomography and computed tomography. *Arch Otolaryngol Head Neck Surg*. 2009;135:1119–25.
34. Porceddu SV, Pryor DI, Burmeister E, Burmeister BH, Poulsen MG, Foote MC, et al. Results of a prospective study of positron emission tomography-directed management of residual nodal abnormalities in node-positive head and neck cancer after definitive radiotherapy with or without systemic therapy. *Head Neck*. 2011;33:1675–82.
35. Prestwich RJ, Subesinghe M, Gilbert A, Chowdhury FU, Sen M, Scarsbrook AF. Delayed response assessment with FDG-PET-CT following (chemo)radiotherapy for locally advanced head and neck squamous cell carcinoma. *Clin Radiol*. 2012;67:966–75.
36. Rabalais AG, Walvekar R, Nuss D, McWhorter A, Wood C, Fields R, et al. Positron emission tomography-computed tomography surveillance for the node-positive neck after chemoradiotherapy. *Laryngoscope*. 2009;119:1120–4.
37. Schouten CS, de Graaf P, Alberts FM, Hoekstra OS, Comans EF, Bloemena E, et al. Response evaluation after chemoradiotherapy for advanced nodal disease in head and neck cancer using diffusion-weighted MRI and 18F-FDG-PET-CT. *Oral Oncol*. 2015;51:541–7.
38. Sjövall J, Bitzén U, Kjellén E, Nilsson P, Wahlberg P, Brun E. Qualitative interpretation of PET scans using a Likert scale to assess neck node response to radiotherapy in head and neck cancer. *Eur J Nucl Med Mol Imaging*. 2016;43:609–16.
39. Slevin F, Subesinghe M, Ramasamy S, Sen M, Scarsbrook AF, Prestwich RJD. Assessment of outcomes with delayed 18F-FDG PET-CT response assessment in head and neck squamous cell carcinoma. *Br J Radiol*. 2015;88:20140592.
40. Vainshtein JM, Spector ME, Stenmark MH, Bradford CR, Wolf GT, Worden FP, et al. Reliability of post-chemoradiotherapy F-18-FDG PET/CT for prediction of locoregional failure in human papillomavirus-associated oropharyngeal cancer. *Oral Oncol*. 2014;50:234–9.
41. Zundel MT, Michel MA, Schultz CJ, Maheshwari M, Wong SJ, Campbell BH, et al. Comparison of physical examination and fluorodeoxyglucose positron emission tomography/computed tomography 4–6 months after radiotherapy to assess residual head-and-neck cancer. *Int J Radiat Oncol Biol Phys*. 2011;81:825–32.
42. Mehanna H, Wong W-L, McConkey CC, Rahman JK, Robinson M, Hartley AG, et al. PET-CT surveillance versus neck dissection in advanced head and neck cancer. *N Engl J Med*. 2016;374:1444–54.
43. Fakhry N, Lussato D, Jacob T, Giorgi R, Giovanni A, Zanaret M. Comparison between PET and PET/CT in recurrent head and neck cancer and clinical implications. *Eur Arch Otorhinolaryngol*. 2007;264:531–8.
44. Goshen E, Davidson T, Yahalom R, Talmi YP, Zwas ST. PET/CT in the evaluation of patients with squamous cell cancer of the head and neck. *Int J Oral Maxillofac Surg*. 2006;35:332–6.
45. Krupar R, Robold K, Gaag D, Spanier G, Kreutz M, Renner K, et al. Immunologic and metabolic characteristics of HPV-negative and HPV-positive head and neck squamous cell carcinomas are strikingly different. *Virchows Arch*. 2014;465:299–312.
46. Tahari AK, Alluri KC, Quon H, Koch W, Wahl RL, Subramaniam RM. FDG PET/CT imaging of oropharyngeal squamous cell carcinoma: characteristics of human papillomavirus-positive and -negative tumors. *Clin Nucl Med*. 2014;39:225–31.
47. Clark J, Jeffery CC, Zhang H, Cooper T, O’Connell DA, Harris J, et al. Correlation of PET-CT nodal SUVmax with p16 positivity in oropharyngeal squamous cell carcinoma. *J Otolaryngol Head Neck Surg*. 2015;44.
48. Ang K, Harris J, Wheeler R, Weber R, Rosenthal D, Nguyen-Tân P, et al. Human papillomavirus and survival of patients with oropharyngeal cancer. *N Engl J Med*. 2010;363:24–35.
49. Mirghani H, Amen F, Tao Y, Deutsch E, Levy A. Increased radiosensitivity of HPV-positive head and neck cancers: molecular basis and therapeutic perspectives. *Cancer Treat Rev*. 2015;41:844–52.
50. Huang SH, O’Sullivan BO, Xu W, Zhao H, Chen D, Ringash J, et al. Temporal nodal regression and regional control after primary radiation therapy for N2-N3 head-and-neck cancer stratified by HPV status. *Radiat Oncol Biol Elsevier Inc*. 2013;87:1078–85.
51. Marcus C, Ciarallo A, Tahari AK, Mena E, Koch W, Wahl RL, et al. Head and neck PET/CT: therapy response interpretation criteria (Hopkins criteria)-interreader reliability, accuracy, and survival outcomes. *J Nucl Med*. 2014;55:1411–6.
52. Aide N, Lasnon C, Veit-Haibach P, Sera T, Sattler B, Boellaard R. EANM/EARL harmonization strategies in PET quantification: from daily practice to multicentre oncological studies. *Eur J Nucl Med Mol Imaging*. 2017;44:17–31.