ORIGINAL ARTICLE

CrossMark

What is the prognostic impact of FDG PET in locally advanced head and neck squamous cell carcinoma treated with concomitant chemo-radiotherapy? A systematic review and meta-analysis

Pierluigi Bonomo[1] (iD) · A. Merlotti[2] · E. Olmetto[1] · A. Bianchi[3] · I. Desideri[1] · A. Bacigalupo[4] · P. Franco[5] · C. Franzese[6] · E. Orlandi[7] · L. Livi[1] · S. Caini[8]

Abstract

Purpose Evidence is conflicting on the prognostic value of ^{18}F-fluorodeoxyglucose (FDG) positron emission tomography (PET) in head and neck squamous cell carcinoma. The aim of our study was to determine the impact of semiquantitative and qualitative metabolic parameters on the outcome in patients managed with standard treatment for locally advanced disease.

Methods A systematic review of the literature was conducted. A meta-analysis was performed of studies providing estimates of relative risk (RR) for the association between semiquantitative metabolic parameters and efficacy outcome measures.

Results The analysis included 25 studies, for a total of 2,223 subjects. The most frequent primary tumour site was the oropharynx (1,150/2,223 patients, 51.7%). According to the available data, the majority of patients had stage III/IV disease (1,709/1,799, 94.9%; no information available in four studies) and were treated with standard concurrent chemoradiotherapy (1,562/2,009 patients, 77.7%; only one study without available information). A total of 11, 8 and 4 independent studies provided RR estimates for the association between baseline FDG PET metrics and overall survival (OS), progression-free survival (PFS) and locoregional control (LRC), respectively. High pretreatment metabolic tumour volume (MTV) was significantly associated with a worse OS (summary RR 1.86, 95% CI 1.08–3.21), PFS (summary RR 1.81, 95% CI 1.14–2.89) and LRC (summary RR 3.49, 95% CI 1.65–7.35). Given the large heterogeneity ($I^2 > 50\%$) affecting the summary measures, no cumulative threshold for an unfavourable prognosis could be defined. No statistically significant association was found between SUV_{max} and any of the outcome measures.

Conclusion FDG PET has prognostic relevance in the context of locally advanced head and neck squamous cell carcinoma. Pretreatment MTV is the only metabolic variable with a significant impact on patient outcome. Because of the heterogeneity and the lack of standardized methodology, no definitive conclusions on optimal cut-off values can be drawn.

Keywords Head and neck cancer · Radiotherapy · ^{18}F-Fluorodeoxyglucose (FDG) positron emission tomography · Metabolic tumour volume

✉ Pierluigi Bonomo
  bonomopierlu@gmail.com

[1] Radiation Oncology, Azienda Ospedaliero – Universitaria Careggi, University of Florence, largo Brambilla 3, 50134 Florence, Italy

[2] Radiation Oncology, Azienda Ospedaliera S.Croce e Carle, Cuneo, Italy

[3] Nuclear Medicine Department, Azienda Ospedaliera S.Croce e Carle, Cuneo, Italy

[4] Radiation Oncology Department, Ospedale Policlinico San Martino, Genoa, Italy

[5] Department of Oncology, Radiation Oncology, University of Turin, Turin, Italy

[6] Department of Radiotherapy and Radiosurgery, Humanitas Cancer Center and Research Hospital, Rozzano, Italy

[7] Radiotherapy 2 Unit, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy

[8] Cancer Risk Factors and Lifestyle Epidemiology Unit, Cancer Research and Prevention Institute (ISPO), Florence, Italy

Introduction

Head and neck cancer is the sixth most common malignant tumour, with increasing incidence worldwide [1]. In over 95% of cases, the disease arises from the epithelial layer of the mucosa lining the upper aerodigestive tract. Due to the absence of anatomical barriers, the abundant lymphatic drainage of the neck and the usually infiltrative pattern of growth of head and neck squamous cell carcinoma (HNSCC), in about 60% of patients the diagnosis is made at an advanced locoregional stage. In order to maximize the likelihood of disease cure, multimodality treatment is usually needed. Therapeutic management is often challenging: both primary radical surgery and concurrent chemoradiotherapy are burdened with a high rate of posttreatment complications, acute and long-term toxicities [2] and a marked detrimental effect on quality of life. Notwithstanding the refinement of treatment strategies that has taken place in last 20 years, the prognosis of HNSCC remains severe, with a cumulative 5-year overall survival (OS) rate of 45–55% [3] in patients with locally advanced disease. The prevalent pattern of failure in the overall population is locoregional: about 50% of first events of relapse occur at the primary tumour site and/or in the neck, in the vast majority (about 90%) within the first 2 years after treatment.

Taking into account that the patient's outlook can be substantially influenced by clinical factors with large variability existing among the different subsites of disease, a series of common features contribute to the severe prognosis of locally advanced HNSCC; these include the suboptimal efficacy of the standard "one size fits all" multimodal approach, the large proportion of frail patients who are noncompliant with intensive therapy, and the absence of biomarkers. In this regard, the only notable exception is the human papillomavirus (HPV). In the last 15 years, a major epidemiological shift has taken place in western countries due to the rising incidence of HPV-associated oropharyngeal cancer [4], reducing the dominance of the classical phenotype of HNSCC resulting from alcohol and tobacco-induced field cancerization. A positive HPV status was recognized as an independent favourable prognostic factor in a series of correlative prospective studies and in an unplanned secondary analysis of the randomized phase 3 RTOG 0129 trial [5]. Overall, HPV positivity is associated with a reduction in the risk of death and disease progression of about 60%.

Although major progress has been achieved in unravelling key molecular pathways involved in HNSCC pathogenesis [6], at present no biomarkers are available in clinical practice apart from HPV status. Prognostic information is therefore critically lacking in the management of patients affected by HNSCC. Next to individual genomic profiling, an alternative strategy which has been explored in recent years is to integrate molecular imaging into precision oncology care, exploiting the potential of imaging as a biomarker. The possibility of linking the information obtained from medical images with personalized treatment forms the core of "theragnostics", an term that has been used particularly in the context of radiation therapy [7]. In a hallmark review published in 2000, Ling et al. [8] suggested that the evolution of molecular imaging could facilitate the development of customized dose delivery in the era of intensity-modulated radiotherapy (IMRT). As foreseen by Ling and colleagues, in the last 15 years molecular imaging has been increasingly implemented in the management of HNSCC, in particular ^{18}F-fluorodeoxyglucose (FDG) positron emission tomography (PET). The fundamental prerequisite is the ability to image physiopathological processes occurring within a tumour or its microenvironment. The use of FDG allows the characterization of the metabolic activity of a defined tumour burden. In HNSCC, available evidence supports the role of FDG PET in primary target definition for radiotherapy planning [9], staging [10] and posttreatment response assessment [11]. However, its potential impact on patient outcomes is an unresolved issue. The aim of this work was to define the relevance of semiquantitative and qualitative FDG PET features as prognostic biomarkers in the curative setting of locally advanced head and neck cancer.

Materials and methods

In accordance with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement [12], a systematic review of the literature was conducted. Relevant articles were identified in two databases (MEDLINE and Embase) over a 10-year period (1 January 2007 to 28 February 2017) using the appropriate terminology as described in Appendix 1 of the Supplementary material. Conference proceedings of main international conferences (ASCO, ASTRO, ESMO, ESTRO, ECCO) were also searched. The reference lists of the articles reviewed as full texts were also searched manually. The literature search strategy was based on the PICO methodology [13], as discussed in the following sections.

Population

The target population of our analysis consisted of adult patients (>18 years of age) treated with curatively intended radiotherapy, concurrent chemoradiotherapy or radiotherapy combined with targeted therapy for locally advanced HNSCC. Primary surgery and induction chemotherapy were not allowed. In view of the known heterogeneity among different head and neck subsites, we sought to assess whether the impact of metabolic parameters could be observed in specific disease entities or in HNSCC taken as a whole. In addition, information on the radiotherapy technique used and the schedule of systemic therapy

administered was collected when available. To provide evidence-based support for the analysis, the published literature was categorized according to the type of study design: all case series except those with fewer than 20 patients, literature reviews and consensus statements were eligible. Only studies in the English language were included.

Interventions

Upon inclusion in the analysis, adequate information on FDG PET metrics (semiquantitative parameters and/or qualitative scores) had to be retrieved from the studies analysed. Studies focusing on tracers other than FDG and on integrated PET/MRI were excluded. Since the main aim of this review was to investigate the potential impact of specific metabolic data on HNSCC prognosis, the following parameters were considered as main interventions: standardized uptake values (SUV_{max}, SUV_{mean}, SUV_{peak}), metabolic tumour volume (MTV) and total lesion glycolysis (TLG). These parameters were defined according to reference guidelines [14], as follows:

– SUV (body-weighted): the concentration of FDG in a given region of interest (ROI) or volume of interest (VOI; expressed in kilobecquerels per millilitre) divided by the ratio between administered activity (corrected for radioactive decay at the time of scanning) and the body weight of the patient
– SUV_{max}: the highest SUV of pixels (or voxels) in a given ROI (or VOI)
– SUV_{mean}: the mean SUV of pixels (or voxels) in a given ROI (or VOI)
– SUV_{peak}: SUV_{mean} within a 1-cm^3 spherical VOI centred on the voxels with the highest uptake
– MTV: the VOI segmented using a fixed threshold (usually 41% or 50%) of FDG-avid lesions
– TLG: the product of the VOI average SUV (SUV_{mean}) and the corresponding MTV

Standardized qualitative interpretations of FDG PET scans were also considered interventions, if rigorously defined. In addition, the included studies were further analysed according to the timing of the FDG PET scans, whether performed before, during or after treatment.

Comparators

When available, different clinical factors other than the metabolic FDG PET parameters discussed above were defined as "comparators" if analysed as potential prognostic biomarkers.

Outcomes

Ultimately, we sought to assess whether intrinsic features on FDG PET retain prognostic significance in terms of outcome. Therefore, we searched for a potential correlation between the interventions (as described above) and locoregional control (LRC), progression-free survival (PFS) and OS at a minimum follow-up of 1 year. These outcome measures were defined as follows:

– LRC: the time from randomization (or study initiation) to local and/or regional disease progression
– PFS: the time from randomization (or study initiation) to disease progression or death
– OS: the time from randomization (or study initiation) to death from any cause

Studies in which the main outcome measure was not consistent with the definition of the prespecified efficacy endpoints were excluded. Studies performed to assess the diagnostic accuracy of FDG PET as well as "in-silico" radiotherapy planning analyses were also excluded.

Statistical analysis

Baseline demographics, patient and disease characteristics, treatment features and outcome data were collected by three authors (P.B., A.M., E.O.), verified by two reviewers (I.D., S.C.) and summarized using descriptive statistics. From all studies included in the literature review, we extracted the most adjusted estimate of relative risk (RR), including odds ratio and hazard ratio (HR), for the association between each of the metabolic parameters (SUV_{max}, SUV_{mean}, SUV_{peak}, MTV and TLG) and each of the patient outcomes (OS, PFS and LRC). When there were two or more independent RR estimates, these were transformed into logRR and the corresponding variance using the formula of Greenland [15] and pooled using random effects models to obtain a summary RR (SRR) and corresponding 95% confidence intervals (CI). We assessed the heterogeneity between studies using the I^2 statistic, which is interpreted as the percentage of the variability that is attributable to true heterogeneity rather than chance. Larger values of I^2 denote greater between-estimate heterogeneity; values of I^2 below 50% are considered acceptable. We did not perform subgroup analysis and meta-regression because of the limited sample size. Finally, we evaluated the presence of publication bias using the funnel plot of Begg and Mazumdar [16] and the regression test of Egger et al. [17]. The meta-analysis was conducted using the *metan* command in Stata version 14 (Stata Corp, College Station, TX).

Results

Data collection and analysis

Two authors (P.B., A.M.) independently examined the titles and abstracts of each search record, and retrieved the full text articles for potentially eligible studies. The full texts were further examined according to the inclusion criteria. Discrepancies were resolved by consensus. Data were extracted by the two authors using a data collection form. Overall, of 180 studies identified using the predefined search criteria, 81 were screened by assessment of the abstracts (Fig. 1). Of these screened studies, 42 were evaluated for eligibility, and 25 [18–44] satisfied the inclusion criteria and were therefore analysed fully. The whole reference lists of the eligible studies and the reasons for exclusion are available in Appendix 2 of the Supplementary material. In terms of study design, most included studies (21/25, 84%) were retrospective. Two papers
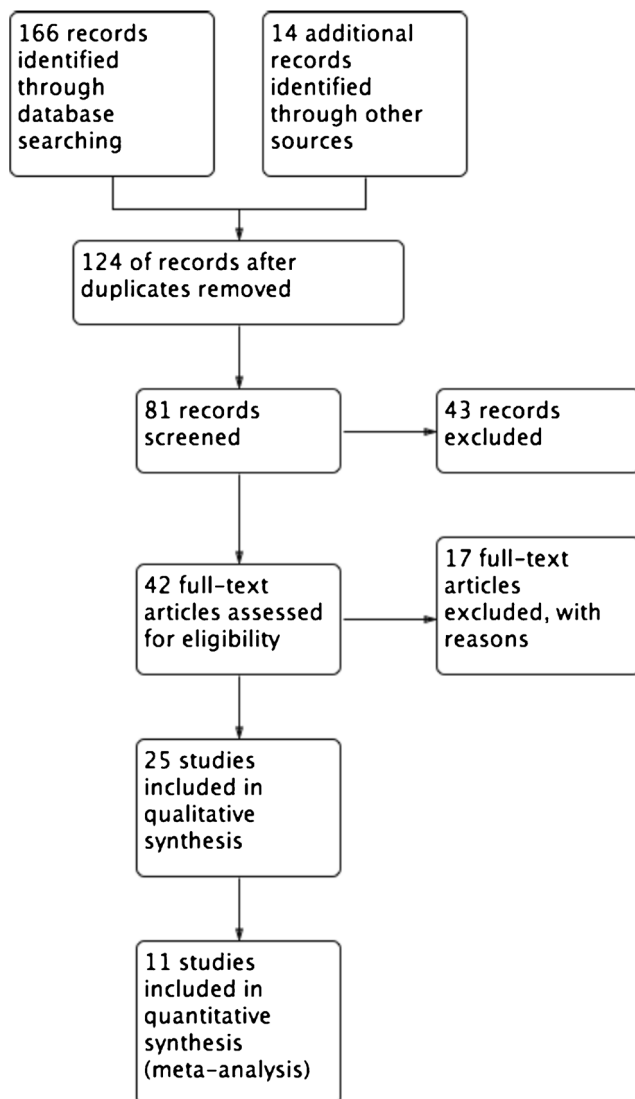
were initially retrieved in abstract form [33, 41] and updated as soon as the full versions became available [34, 42]. One study [29] had limited data on the disease and treatment characteristics collected in most patients, but provided adequate information on FDG PET variables and outcomes.

Patient characteristics

The overall population consisted of 2,223 patients (Table 1). The median age of the whole cohort was 59 years (range 48–68 years). Most patients (1,875/2,223, 84.3%) were men. Only 3 of the 25 studies [31, 32, 40] provided information about tobacco exposure in terms of pack-years. Generic information on subjects with a smoking history was reported in six additional studies [22, 28–30, 39, 43] (81.4%, 82.6%, 67.3%, 80%, 90.9% and 72% of patients were current or former smokers, respectively). In addition, data on baseline ECOG Performance Status were reported for only 29% of the whole cohort (646/2,223).

Disease-related features

The most frequent primary tumour site was the oropharynx (1,150/2,223 patients, 51.7%), followed by the hypopharynx (377, 16.9%), larynx (345, 15.6%), nasopharynx (197, 8.9%), oral cavity (98, 4.4%), and others (56, 2.5%; Table 2). Information on HPV status was available for fewer than half of those with oropharyngeal tumour (508/1,150, 44.1%), and of these (as extrapolable from 7 of the 25 studies) 247 (48.6%) were p16/HPV-positive. Overall, in the majority of patients (1,709/1,799, 94.9%; no information available in four studies) the disease was in aggregated stage III/IV (Table 3).

Treatment-related features

Most patients (1,467/1,544, 95%) were treated with IMRT, while 77 (5%) received 3D-conformal radiotherapy (3DCRT). No information on the radiotherapy technique used was available in seven studies (Table 4). The most adopted radiotherapy regimen consisted of conventional fractionation of 1.8 or 2 Gy per fraction for a total dose of 66–72 Gy in the majority of cases (22/23 papers; no available data in two studies). Concurrent chemoradiotherapy was the most frequent treatment schedule in our analysis, being used in 1,562/2,009 patients (77.7%; no available information in only one study). Standard three-weekly 100 mg/m^2 cisplatin was the chosen regimen in almost half of the included studies (11/24). Finally, a very small group of patients received induction chemotherapy before radiotherapy (181/2,223, 8.1%) in seven studies. On the basis that these studies were not excluded by our entry search criteria, they were retained in the analysis.

Fig. 1 Flow chart of the literature search

Boxes in flow chart:
- 166 records identified through database searching
- 14 additional records identified through other sources
- 124 of records after duplicates removed
- 81 records screened → 43 records excluded
- 42 full-text articles assessed for eligibility → 17 full-text articles excluded, with reasons
- 25 studies included in qualitative synthesis
- 11 studies included in quantitative synthesis (meta-analysis)

Table 1 Design of the studies analysed and patient characteristics

Reference	Year	Study design	Patient characteristics						
			Total number	Age (years)		Gender, n (%)		Smoking (pack-years)	ECOG performance status, grade (%)
				Median	Range	Male	Female		
[18]	2015	Retrospective	62	57	23–83	44 (71)	18 (29)	ns	ns
[19]	2012	Retrospective	26	63	41–79	23 (89)	3 (11)	ns	ns
[20]	2017	Retrospective	122	61	ns	101 (83)	21 (17)	ns	ns
[21]	2014	Prospective	51	52	36–69	48 (94)	3 (6)	ns	ns
[22]	2013	Retrospective	70	52	37–86	66 (94)	4 (6)	ns	ns
[23]	2009	Retrospective	82	53.8	11–70	69 (84)	13 (16)	ns	ns
[24]	2012	Retrospective	88	59	26–83	74 (84)	14 (16)	ns	ns
[25]	2015	Retrospective	70	65	21–91	61 (87)	9 (13)	ns	ns
[26]	2016	Retrospective	78	62	24–79	63 (81)	15 (19)	ns	0/1 (98.8)
[27]	2014	Retrospective	108	67	43–85	93 (86)	15 (14)	ns	ns
[28]	2017	Retrospective	75	59	39–80	67 (89)	8 (11)	ns	ns
[29]	2014	Retrospective	214	58	ns	175 (82)	39 (18)	ns	ns
[30]	2017	Prospective	35	67.6	50–80	32 (91)	3 (9)	ns	0/1 (100)
[31]	2016	Retrospective	69	61	39–81	61 (88)	8 (12)	>10: 40 (60.9%)	ns
[32]	2015	Retrospective	72	60	39–75	61 (85)	11 (15)	>10: 41 (56.9%)	ns
[34]	2017	Retrospective	85	66	43–79	81 (95)	4 (5)	ns	ns
[35]	2011	Retrospective	47	55.1	15–86.1	39 (83)	8 (17)	ns	0/1 (95)
[36]	2016	Prospective	86	50	40–60	80 (93)	6 (7)	ns	ns
[37]	2013	Retrospective	81	65	34–81	74 (91)	7 (9)	ns	ns
[38]	2015	Retrospective	287	64	33–89	221 (77)	66 (23)	ns	0 (63)
[39]	2014	Retrospective	100	56	27–81	86 (86)	14 (14)	ns	ns
[40]	2015	Prospective	74	56	42–73	65 (88)	9 (12)	8.75 (median)	0/1 (100)
[42]	2016	Prospective	125	59	ns	93 (74)	32 (26)	ns	0/1 (100)
[43]	2014	Retrospective	40	48	21–69	30 (75)	10 (25)	ns	ns
[44]	2017	Retrospective	76	55	42–76	68 (89)	8 (11)	ns	ns

ns not stated

Prognostic impact of FDG PET: descriptive analysis

The timing of FDG PET was different among the studies included in the analysis (Table 5). A single baseline assessment time-point was present in almost half of the studies (12/25) while a combination of pretreatment, interim (during treatment) and posttreatment scans was described in four (pretreatment plus interim), seven (pretreatment plus posttreatment) and two (pretreatment plus interim plus posttreatment) studies. Among those studies providing data on more than a single time-point, a time-weighted analysis exploring changes over time ("delta") of specific metabolic semiquantitative or qualitative features was additionally reported in seven. As a single variable, MTV and SUV_{max} were the main metabolic parameters addressed in nine and seven studies, respectively. A qualitative analysis was used in three studies [29, 31, 42]. Zschaeck et al. [44] determined SUV_{mean} in irradiated normal mucosa tissue to explore the impact of off-target hypermetabolism and

its change over time. Only a limited number of alternative prognostic biomarkers (comparators) were reported in parallel with the metabolic evaluation (nine studies). The median overall follow-up time for all studies was 23.6 months (range 15–55.8 months). In terms of threshold or cut-off values to discriminate worse from better outcomes, a large variability was observed for each intervention. Finally, a large heterogeneity characterized the prognostic information which could be extracted from each paper.

Prognostic impact of pretreatment FDG PET: meta-analysis

A total of 11 [20–22, 25, 34–37, 40, 43, 45], 8 [20–23, 35, 36, 38, 40] and 4 [21, 25, 37, 40] independent studies provided RR estimates for the association between baseline FDG PET and OS, PFS and LRC, respectively. These studies were thus included in the meta-analysis aiming to assess the potential

Table 2 Disease features: tumour site

Reference	Number of patients	Oropharynx		Larynx	Hypopharynx	Oral cavity	Nasopharynx	Other
		Total	p16/HPV-positive					
[18]	62	14 (22%)	ns	10 (16%)	12 (20%)	3 (5%)	14 (22%)	9 (15%)
[19]	26	12 (46%)	ns	9 (35%)	2 (8%)	0	3 (11%)	0
[20]	122	122 (100%)	32 (26%)	0	0	0	0	0
[21]	51	20 (39%)	ns	0	21 (41%)	0	10 (20%)	0
[22]	70	70 (100%)	13 (19%)	0	0	0	0	0
[23]	82	13 (16%)	ns	0	6 (7%)	0	63 (77%)	0
[24]	88	58 (66%)	ns	15 (17%)	0	1 (1%)	7 (8%)	7 (8%)
[25]	70	25 (36%)	ns	0	36 (51%)	0	9 (13%)	0
[26]	78	47 (61%)	ns	3 (4%)	19 (24%)	5 (6%)	0	4 (5%)
[27]	108	28 (26%)	ns	29 (27%)	34 (31%)	17 (16%)	0	0
[28]	75	56 (75%)	ns	11 (15%)	5 (6%)	3 (4%)	0	0
[29]	214	135 (63%)	123 (57%)	40 (19%)	0	11 (5%)	0	28: (13%)
[30]	35	9 (26%)	ns	11 (31%)	12 (34%)	3 (9%)	0	0
[31]	69	41 (59%)	ns	20 (30%)	5 (7%)	3 (4%)	0	0
[32]	72	47 (66%)	ns	16 (22%)	6 (8%)	3 (4%)	0	0
[34]	85	0	ns	35 (41%)	50 (59%)	0	0	0
[35]	47	21 (45%)	ns	7 (15%)	4 (8%)	2 (4%)	13 (28%)	0
[36]	86	45 (52%)	ns	0	41 (48%)	0	0	0
[37]	81	0	ns	57 (70%)	24 (30%)	0	0	0
[38]	287	129 (45%)	ns	44 (15%)	55 (19%)	29 (10%)	30 (11%)	0
[39]	100	100 (100%)	14 (14%)	0	0	0	0	0
[40]	74	58 (78%)	25 (34%)	9 (12%)	7 (10%)	0	0	0
[42]	125	69 (56%)	37 (30%)	21 (17%)	11 (9%)	8 (6%)	8 (6%)	8 (6%)
[43]	40	0	0	0	0	0	40 (100%)	0
[44]	76	31 (41%)	3 (4%)	8 (10%)	27 (36%)	10 (13%)	0	0

ns not stated

prognostic impact of pretreatment metabolic features on patient outcome. Of note, Castelli et al. [45] performed a secondary analysis in the same patient population analysed in a previous work [20] providing additional data with distinct RR estimates that were therefore worthy of inclusion. Instead, the results reported by Zschaeck et al. [44] were not considered, since the prognostic value of uptake in mucosa soft tissue was not investigated in any of the other included studies. In terms of baseline FDG PET parameters, the analysis was limited to MTV and SUV_{max}, for which there were six and seven RR estimates for OS (Figs. 2 and 3), three and seven for PFS (Figs. 4 and 5), and two and three for LRC (Figs. 6 and 7), respectively. Higher MTV values for the primary or primary and nodal disease combined were significantly associated with a worse OS (SRR 1.86, 95% CI 1.08–3.21), PFS (SRR 1.81, 95% CI 1.14–2.89) and LRC (SRR 3.49, 95% CI 1.65–7.35), Instead, we found no statistically significant association between SUV_{max} and any of OS, PFS and LRC. Given the large between-study heterogeneity ($I^2 > 50\%$) that affected the summary measures, no effort was made to define a cumulative threshold value for an unfavourable prognosis. Finally, apart from an unclear or high risk of bias in terms of patient selection (21/25 studies, 84%) because of the predominantly retrospective nature of the included studies, according to the QUADAS-2 tool [46] the overall quality was good with low risks of bias and concerns regarding applicability in the remaining domains (Supplementary Table 6).

Discussion

In the era of precision oncology, the lack of prognostic biomarkers has hindered the evolution of standard-of-care management in HNSCC. Apart from HPV status, no molecular stratification is currently available for use in daily practice. In the last two decades, steady technological progress has highlighted the potential of imaging as a comprehensive tumour biomarker [47]. In the field of functional imaging, FDG PET is the most widespread, easily accessible modality that is able to provide surrogate metabolic information on tumour

Table 3 Disease features: stage

Reference	Tx	T1	T2	T3	T4	N0	N1	N2	N3	I	II	III	IV	III/IV
[18]	0	8	20	14	20	ns	ns	ns	ns	0	10 (16%)	18 (29%)	34 (55%)	52 (84%)
[19]	0	3	9	6	8	3	7	16	0	0	1 (3%)	7 (27%)	18 (70%)	25 (97%)
[20]	0	7	36	52	27	14	21	80	7	0	0	ns	ns	122 (100%)
[21]	ns	ns	ns	ns	ns	ns	ns	ns	ns	0	0	16 (31%)	35 (69%)	51 (100%)
[22]	0	0	0	20	50	ns	ns	ns	ns	0	8 (11%)	43 (52%)	19 (27%)	62 (89%)
[23]	0	22	25	17	18	10	22	43	7	4 (5%)	12 (15%)	30 (36%)	36 (44%)	66 (80%)
[24]	ns	ns	ns	ns	ns	ns	ns	ns	ns	3 (3%)	1 (1%)	15 (16%)	70 (80%)	85 (96%)
[25]	0	0	34	16	20	29	10	18	13	ns	ns	ns	ns	ns
[26]	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	23 (30%)	55 (70%)	78 (100%)
[27]	0	26	37	13	32	51	14	40	3	18 (17%)	20 (18%)	19 (18%)	51 (47%)	70 (65%)
[28]	0	6	31	27	11	0	11	59	5	0	0	10 (13%)	65 (87%)	75 (100%)
[29]	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
[30]	0	0	13	16	6	13	4	16	2	0	4 (11%)	10 (29%)	21 (60%)	31 (89%)
[31]	0	4	28	27	10	16	10	39	4	0	6 (9%)	18 (26%)	45 (65%)	63 (91%)
[32]	0	6	25	31	10	9	11	47	5	0	0	18 (25%)	54 (75%)	72 (100%)
[34]	0	0	19	49	17	26	15	44	0	0	0	33 (39%)	52 (61%)	85 (100%)
[35]	ns	ns	ns	ns	ns	ns	ns	ns	ns	0	2 (4%)	11 (23%)	34 (73%)	45 (96%)
[36]	ns	ns	ns	ns	ns	ns	ns	ns	ns	0	0	4 (5%)	82 (95%)	86 (100%)
[37]	0	2	11	43	25	28	15	38	3	0	0	32 (39%)	49 (61%)	81 (100%)
[38]	0	32	92	78	85	30	32	190	35	0	0	54 (19%)	233 (81%)	287 (100%)
[39]	0	14	39	23	24	4	12	80	4	ns	ns	ns	ns	ns
[40]	0	0	29	26	19	0	0	ns	6	0	0	ns	ns	74 (100%)
[42]	7	ns	ns	ns	ns	0	0	119	6	0	0	ns	ns	125 (100%)
[43]	0	14	ns	ns	ns	0	19	ns	ns	7	ns	ns	ns	ns
[44]	0	2	10	27	37	12	7	54	3	1 (1%)	1 (1%)	11 (15%)	63 (83%)	74 (98%)

ns not stated

burden. The aim of our work was to define whether distinct FDG PET features can be intrinsically associated with prognostic relevance in the context of nonmetastatic HNSCC. We acknowledge several limitations which have to be taken into account when interpreting the data presented. First, most studies included in our systematic review were retrospective. Although a strict search methodology was followed, their potential heterogeneity in terms of patient selection, treatment administration and outcome measures may have affected the consistency of our analysis. Second, the technical variability in the performance of FDG PET scans is also a factor that cannot be ignored with a retrospective study design; only a prospective design can ensure that consensus acquisition recommendations [14] are rigorously adopted. Third, among the included studies the methods used to calculate the FDG PET metrics were not consistent. Heterogeneity in their definition has to be taken into account particularly for SUV_{max} and MTV, for which several threshold values were shown to be significant in discriminating patients with different outcomes.

Renewed interest in the role of FDG PET in the management of HNSCC was recently prompted by the publication of the PET-NECK trial [11]. The findings of this large prospective, multicentre phase 3 trial are practice-changing, since the study provided definitive evidence in favour of a response evaluation centred on the high negative predictive value (NPV) of a 12-week posttreatment FDG PET scan. However, the study had two main limitations that prevented the clarification of other relevant issues on the role of FDG PET in the management of HNSCC. First, none of the 564 patients enrolled in the trial underwent a baseline FDG PET scan; a qualitative comparison between pretreatment and posttreatment scans was therefore not performed. Second, FDG PET semiquantitative metrics could not be evaluated due to nonuniform calibration among the different scanners. From this perspective, the PET-NECK trial did not add any new data to the available low-level body of evidence on the prognostic role of specific FDG PET semiquantitative and qualitative features in HNSCC. Although many investigators have focused on this topic in the last 15 years [48], the literature is characterized by inconclusive and heterogeneous findings [49].

Table 4 Treatment-related features

Reference	Radiotherapy					Induction chemotherapy			Concurrent chemotherapy		
	Number of patients		Regimen[a]	Dose (Gy)		Number of patients	Type of chemotherapy	No. of cycles (median)	Number of patients	Type of chemotherapy	No. of cycles (median)
	3DCRT	IMRT		Total (median)	Per fraction						
[18]	32	30	Conventional	70	2	20	Docetaxel/cisplatin	ns	35	Cisplatin 40 mg/m^2 every 7 days	6
[19]	0	26	Conventional	70	1.8	0	Not administered	0	26	Cetuximab	ns
[20]	0	122	Conventional	70	2	0	Not administered	0	122	Cisplatin 100 mg/m^2 every 21 days	ns
[21]	0	51	Conventional	70	ns	0	Not administered	0	51	Cisplatin 100 mg/m^2 every 21 days	ns
[22]	ns	ns	Conventional	72	2	0	Not administered	0	44	Cisplatin 100 mg/m^2 every 21 days	ns
[23]	ns	ns	Conventional	72	2	0	Not administered	0	68	Cisplatin 100 mg/m^2 every 21 days	ns
[24]	0	88	Conventional	70	2	0	Not administered	0	74	Cisplatin + 5-fluorouracil	ns
[25]	0	70	Conventional	66	1.8	0	Not administered	0	70	Cisplatin 100 mg/m^2 every 21 days	ns
[26]	ns	ns	Conventional	70	2	0	Not administered	0	70	Other	ns
[27]	0	108	Conventional	70	2	48	Docetaxel/cisplatin	2	44	Cisplatin + 5-fluorouracil	6
[28]	0	75	Conventional	70	2	0	TPF	1	75	Cisplatin 100 mg/m^2 every 21 days	ns
[29]	ns	ns	ns	ns	ns	0	Not administered	ns	ns	ns	ns
[30]	40	0	Conventional	70	2	0	Not administered	0	35	Cisplatin 30 mg/m^2 every 7 days	ns
[31]	0	69	Conventional	70	2	15	ns	ns	46	Cisplatin 40 mg/m^2 every 7 days	ns
[32]	0	72	Conventional	70	2	15	ns	ns	40	Cisplatin 30 mg/m^2 every 7 days	6
[34]	ns	ns	Conventional	66	2	0	Not administered	0	85	Other	6
[35]	5	42	Conventional	66	2	0	Not administered	0	47	Cisplatin + 5-fluorouracil	ns
[36]	0	86	Conventional	72	2	0	Not administered	0	86	Other	ns
[37]	ns	ns	Conventional	70	2	0	Not administered	0	47	Cisplatin + 5-fluorouracil	ns
[38]	0	287	Conventional	66	2	26	ns	ns	86	Other	ns
[39]	0	100	Accelerated	69.96	2.12	0	Not administered	0	100	Cisplatin 100 mg/m^2 every 21 days	ns
[40]	ns	ns	ns	ns	ns	0	Not administered	0	74	Cisplatin 100 mg/m^2 every 21 days	ns

Table 4 (continued)

Reference	Radiotherapy						Induction chemotherapy				Concurrent chemotherapy		
	Number of patients		Regimen[a]	Dose (Gy)		No. of cycles (median)	Number of patients	Type of chemotherapy	No. of cycles (median)		Number of patients	Type of chemotherapy	No. of cycles (median)
	3DCRT	IMRT		Total (median)	Per fraction								
[42]	0	125	Conventional	70	2	0	0	Not administered	0		125	Cisplatin 100 mg/m^2 every 21 days	ns
[43]	0	40	Conventional	70.2	1.8	0	40	Docetaxel/cisplatin	3		36	Cisplatin 100 mg/m^2 every 21 days	ns
[44]	0	76	Conventional	72	2	0	0	Not administered	0		76	Cisplatin 100 mg/m^2 every 21 days	ns

ns not stated, *3DCRT* 3-D conformal radiotherapy, *IMRT* intensity-modulated radiotherapy, *TPF* docetaxel, cisplatin, 5-fluorouracil

[a] Conventional: conventional once-daily fractionation. Accelerated: accelerated fractionation

A crucial aspect that needs again to be underlined is the strict dependence of FDG PET information on the image acquisition modality, which in turn may be influenced by a series of factors, ranging from the technical parameters of the scanner to the timing of the scan with respect to treatment. As also demonstrated in our descriptive analysis (Table 5), there is significant variability in the correlation between semiquantitative metrics and outcome measures in HNSCC. We have already pointed out that in the posttreatment scenario a negative PET scan at 12 weeks after chemoradiation is a prognostic biomarker of long-term complete remission based on level 1 evidence. However, standardized interpretation of response to treatment is lacking. In this context, the Hopkins criteria are the only proposed scoring system for qualitative interpretation of FDG PET in HNSCC. Marcus et al. [29] showed that a five-point scale based on prespecified qualitative descriptors is accurate in discriminating complete from incomplete responses. The application of the Hopkins criteria resulted in a high NPV of 91.1% with an overall diagnostic accuracy of 86.9%. Notably, the results of the ECLYPS study [42] prospectively confirmed the reliability of the Hopkins criteria applied 12 weeks after the end of treatment, with an overall NPV of 92.1% and a very low number of equivocal reports. As accurately described by Garibaldi et al. [50] in a recent systematic review, the potential prognostic and predictive relevance of an interim FDG PET scan (scan acquired during treatment) is a controversial matter. At present, no firm conclusions can be drawn as to the ideal metabolic parameter to analyse early in treatment, the most informative threshold value, or the best time to re-scan the patient.

Taking all together, the use of FDG PET in patients with HNSCC provides prognostic information through standardized qualitative assessment at a minimum of 12 weeks after chemoradiation, but no added value during its delivery. It is therefore a rational approach to investigate before treatment whether baseline semiquantitative metrics are intrinsically able to characterize the outcome in patients with locally advanced disease. Conflicting evidence is available from the literature. Pak et al. [51] performed a systematic review and meta-analysis of 13 studies (1,180 patients) to assess the prognostic role of MTV and TLG before treatment. The authors found that high values of both volumetric parameters correlated significantly with a worse outcome. The pooled HRs for OS were 3.51 (95% CI 2.62–4.72, $p < 0.00001$) and 3.14 (95% CI 2.24 – 4.40, $p < 0.00001$) for MTV and TLG, respectively. However, the generalizability of these results is open to question. First, loose criteria were followed in the literature search strategy and inclusion of articles. Second, for both parameters no threshold values portending a worse outcome were defined, thus preventing further analysis of the data.

In a prospective study in 77 patients affected by stage II–IV HNSCC, Schinagl et al. [52] consistently applied five different segmentation methods for coregistered CT and FDG PET scans

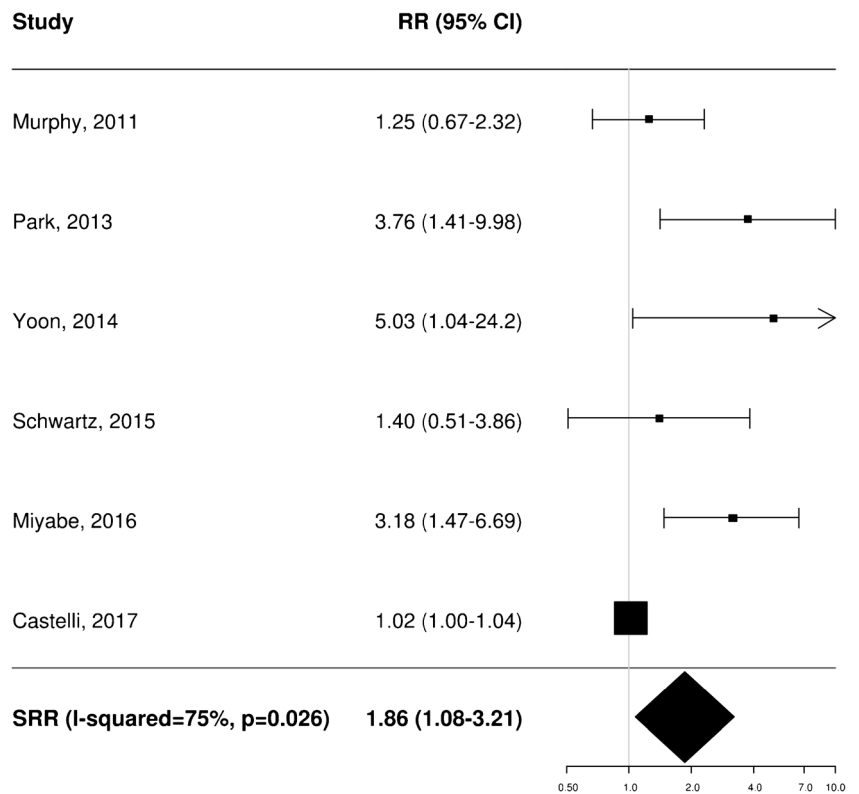**Table 5** Prognostic impact of FDG PET: descriptive analysis

Reference	Timing of FDG PET	Main metabolic parameter	Significant threshold/cut-off value	Prognostic comparator	Median follow-up (months)	Primary outcome measure	Main message
[18]	Pretreatment	MTV	14 ml	ns	18	LRC, PFS	3-year LRFS and DFS lower in patients with MTV ≥14 ml
[19]	Pretreatment Interim Posttreatment Analysis of change over time	SUV$_{max}$	ns	ns	29.2	PFS, DSS	Metabolic response on posttreatment PET correlated with 2-year PFS and DSS
[20]	Pretreatment	MTV	ns	T stage, N stage, HPV status	30.5	LRC, OS	PET-based nomogram: MTV as a continuous variable correlated with 2-year OS
[21]	Pretreatment Interim Analysis of change over time	SUV$_{max}$	ns	ns	23	DFS, OS	SUV$_{max}$ reduction ratio <0.64 associated with inferior 2-year OS and DFS
[22]	Pretreatment	TLG	121.9 g	Uniformity (texture), HPV status	27	PFS, DSS, OS	TLG >121.9 g and uniformity ≤0.138 associated with inferior PFS, DSS and OS
[23]	Pretreatment	MTV	40 ml	ns	ns	DFS	Worse short-term outcome and shorter DFS in patients with MTV >40 ml
[24]	Pretreatment	SUV$_{mean}$	ns	ns	15	DFS	SUV$_{mean}$ >7 (median of cohort) correlated with inferior 2-year DFS
[25]	Pretreatment Posttreatment	SUV$_{max}$	5	Haemoglobin level	38.4	LRC, OS	Posttreatment SUV$_{max}$ <5 and pretreatment haemoglobin >12 g/dl correlated with superior LRC and OS
[26]	Pretreatment Posttreatment	SUV$_{max}$	4.4	ns	52.7	PFS, OS	Posttreatment SUV$_{max}$ <4.4 correlated with superior 3-year PFS and OS
[27]	Pretreatment	MTV, uptake pattern	20 ml	ns	36.4	DFS, DSS	MTV >20 ml and qualitative uptake pattern (ring-shape) correlated with inferior DFS and DSS
[28]	Pretreatment Interim Analysis of change over time	TLG, MTV	2.95 g/ml	ns	28	LRC, DFS, OS	Index node SUV$_{mean}$ on interim PET <2.95 g/ml and TLG, MTV reduction >50% on interim PET correlated with superior LRC, DFS and OS
[29]	Pretreatment Posttreatment Analysis of change over time	Hopkins five-point scale	ns	HPV status	27	PFS, OS	Hopkins five-point qualitative response interpretation and HPV status able to discriminate PFS and OS
[30]	Pretreatment	SUV$_{max}$	ns	ns	ns	LRC	

Table 5 (continued)

Reference	Timing of FDG PET	Main metabolic parameter	Significant threshold/cut-off value	Prognostic comparator	Median follow-up (months)	Primary outcome measure	Main message
	Posttreatment Analysis of change over time						SUV_{max} reduction ratio <1.04 associated with inferior LRC
[31]	Pretreatment Interim Posttreatment Analysis of change over time	Visual grading	ns	ns	28	LRC, DFS, OS	Visual grading response interpretation able to discriminate 2-year LRC, DFS and OS
[32]	Pretreatment, Interim	TLG	9.4 g	ns	25	LRC, DFS, OS	TLG on interim PET <9.4 g correlated with superior 2-year LRC, DFS and OS
[34]	Pretreatment	MTV	28.7 ml	T stage	ns	LRC, OS	MTV >28.7 ml correlated with inferior 3-year LRC and OS
[35]	Pretreatment Posttreatment	MTV	ns	ns	34	DFS, OS	Increase in MTV2.0 (volume above SUV threshold of 2) of 21 ml associated with inferior DFS and OS
[36]	Pretreatment	SUV_{max}	19.4	k_{ep}-tumour, v_e-node (MR parameters)	28	PFS, OS	k_{ep}-tumour, v_e-node, SUV_{max} independently able to discriminate 3-year PFS and OS
[37]	Pretreatment	MTV	18 ml	Primary tumour site	40.4	LRC, OS	MTV >18 ml correlated with inferior 3-year LRC and OS
[38]	Pretreatment	SUV_{max}	13	Gross tumour volume, cisplatin delivery, smoke	32	DFS, OS	SUV_{max} <13 (median of cohort) correlated with superior DFS and OS
[39]	Pretreatment	MTV	9.7 ml	ns	55.8	LRC, PFS, OS	MTV <9.7 ml (median of cohort) correlated with superior 5-year LRC, PFS and OS
[40]	Pretreatment Posttreatment	MTV	ns	ns	50	LRC, PFS	Pretreatment MTV above the median correlated with inferior LRC and PFS
[42]	Pretreatment Interim Analysis of change over time	Hopkins five-point scale	ns	HPV status	20.4	PFS, OS	Hopkins five-point qualitative response interpretation and HPV status able to discriminate PFS and OS
[43]	Pretreatment	MTV	ns	ns	32.5	OS	MTV3.0 (volume above a SUV threshold of 3) >23.01 ml associated with inferior OS
[44]	Pretreatment Interim	SUV_{mean} (in MST)	ns	ns	29.3	LRC, OS	SUV_{mean} of MST on interim PET >2.3 g/ml (median of cohort) correlated with superior LRC and OS

ns not stated, *MST* mucosa and submucosa soft tissues, k_{ep} efflux rate constant of primary tumour on dynamic contrast-enhanced MR imaging, v_e relative volume of extracellular extravascular space of largest metastatic lymph node on diffusion-weighted MR imaging, *LRC* locoregional control, *PFS* progression-free survival, *DSS* disease-specific survival, *OS* overall survival, *DFS* disease-free survival

Fig. 2 Impact of pretreatment MTV on overall survival

Study	RR (95% CI)
Murphy, 2011	1.25 (0.67-2.32)
Park, 2013	3.76 (1.41-9.98)
Yoon, 2014	5.03 (1.04-24.2)
Schwartz, 2015	1.40 (0.51-3.86)
Miyabe, 2016	3.18 (1.47-6.69)
Castelli, 2017	1.02 (1.00-1.04)
SRR (I-squared=75%, p=0.026)	**1.86 (1.08-3.21)**

at baseline. Among the different metrics obtained, only the gross tumour volume (GTV) visually delineated on FDG PET images was significantly correlated with outcome in oral cavity and oropharyngeal tumours, while all isocontour-based volumes, SUV_{mean} and SUV_{max}, were not. A large single-centre [38] retrospective study on 287 patients receiving IMRT-based treatment

Fig. 3 Impact of pretreatment SUV_{max} on overall survival

Study	RR (95% CI)
Cheng, 2013	0.56 (0.16-1.97)
Chen, 2014	1.16 (0.37-4.15)
Yoon, 2014	1.08 (0.25-7.71)
Katahira-Suzuki, 2015	1.24 (0.28-5.52)
Schwartz, 2015	0.38 (0.12-1.20)
Ng, 2016	3.48 (1.50-8.03)
Castelli, 2017	1.43 (1.23-1.65)
SRR (I-squared=49%, p=0.435)	**1.22 (0.74-2.00)**
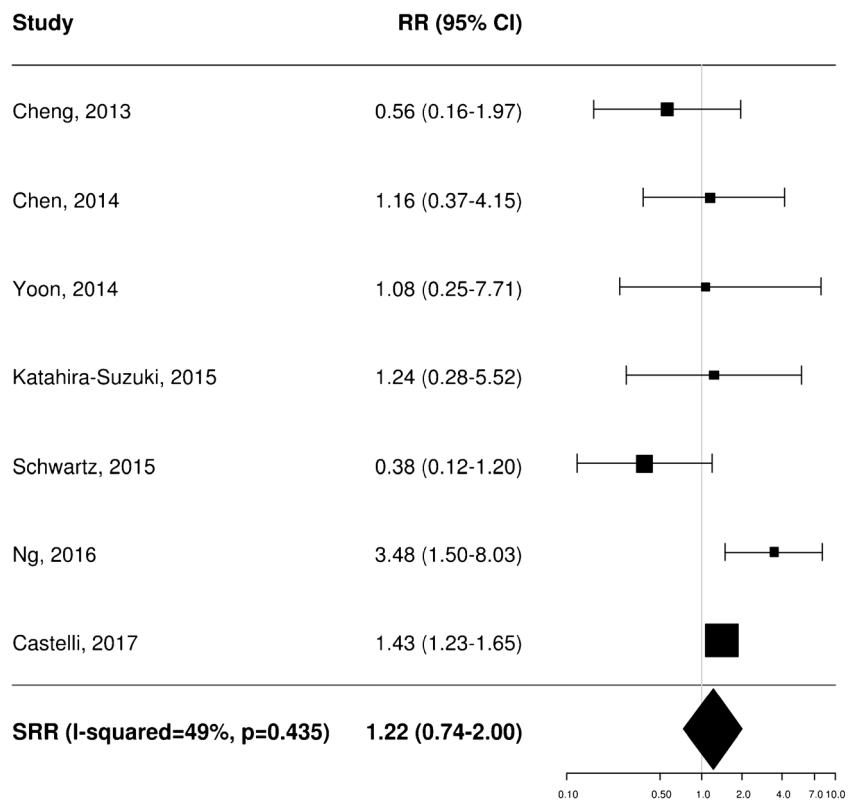
Fig. 4 Impact of pretreatment MTV on progression-free survival

Study	RR (95% CI)
Chung, 2009	3.42 (1.04-11.3)
Murphy, 2011	1.42 (0.93-2.17)
Schwartz, 2015	2.34 (1.02-5.37)
SRR (I-squared=23%, p=0.012)	1.81 (1.14-2.89)
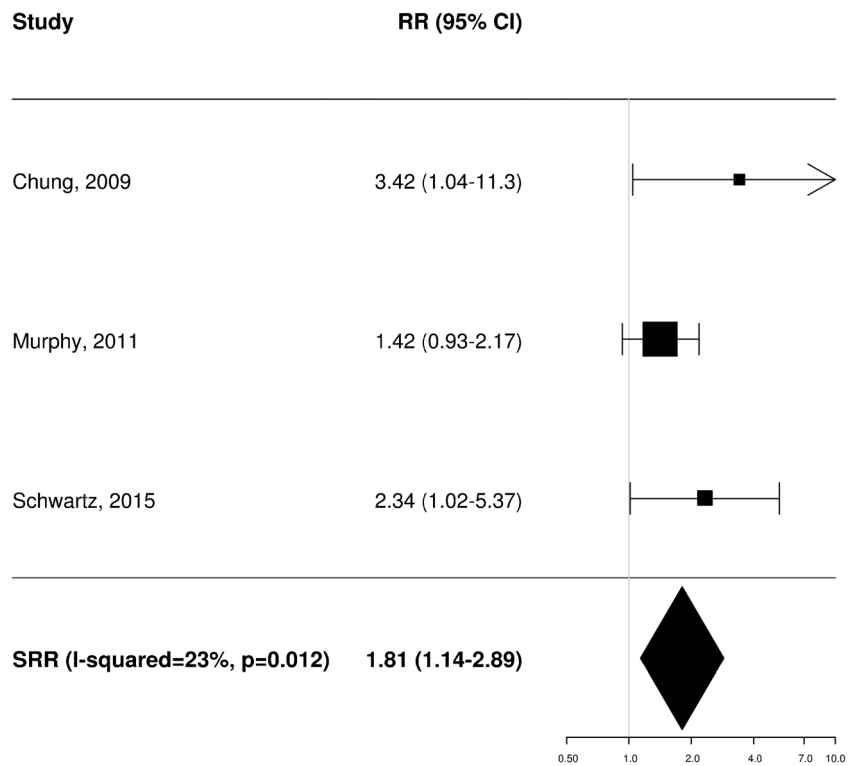
showed different results. In a univariate analysis, increasing values of SUV_{max} (as a logarithmic variable) yielded a HR of 1.72 (95% CI 1.34–2.19) for a worse disease-free survival (DFS) and OS. Multivariate analysis showed an additive effect of increasing GTV (HR 1.74, 95% CI 1.33–2.27; increase in interquartile range from 25% to 75% corresponding to an increase in GTV from 27.4 cm^3 to 95.8 cm^3) and increasing SUV_{max} (HR 1.34, 95% CI 1.01–1.77; increase in interquartile range from

Fig. 5 Impact of pretreatment SUV_{max} on progression-free survival

Study	RR (95% CI)
Chung, 2009	0.76 (0.28-2.07)
Cheng, 2013	0.99 (0.24-4.01)
Chen, 2014	1.19 (0.38-3.74)
Rasmussen, 2015	1.34 (1.01-1.77)
Schwartz, 2015	0.30 (0.12-0.75)
Ng, 2016	2.53 (1.12-5.72)
Castelli, 2017	1.43 (1.23-1.65)
SRR (I-squared=59%, p=0.296)	1.19 (0.86-1.65)
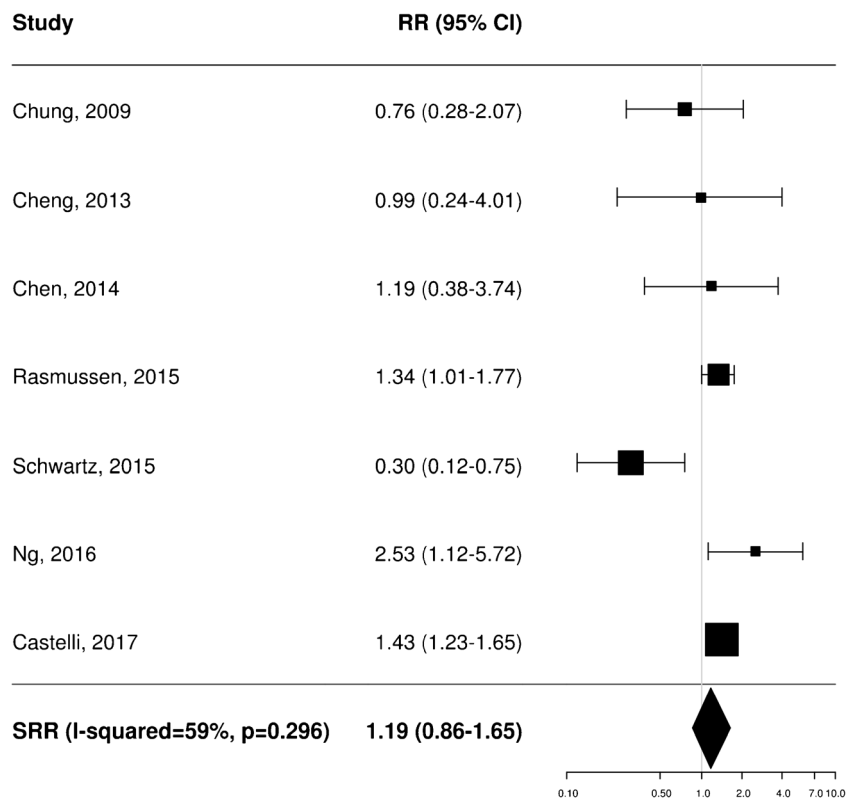
Fig. 6 Impact of pretreatment
MTV on locoregional control

Study	RR (95% CI)
Park, 2013	3.14 (1.17-8.40)
Schwartz, 2015	4.01 (1.28-12.5)
SRR (I-squared=0%, p=0.001)	3.49 (1.65-7.35)
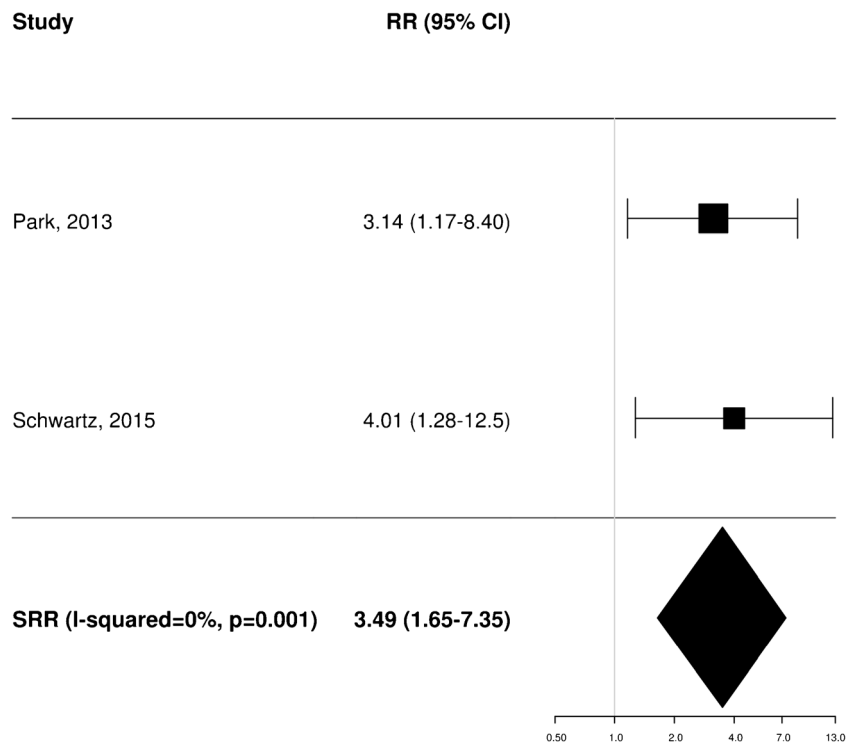
25% to 75% corresponding to an increase in SUV_{max} from 9.6 to 16.8) for a worse prognosis.

The link between FDG avidity and tumour volume has been further explored by different groups focusing on MTV. In this regard, the correlative, prospective imaging study of the randomized phase 3 RTOG 0522 trial [40] is noteworthy. Of the whole sample of 940 patients enrolled, 74 from 19 different centres provided both pretreatment and posttreatment FDG PET scans, as mandated upon inclusion. A prespecified acquisition imaging protocol was followed in all patients. Excellent centralized interobserver agreement (intraclass correlation coefficient ≥0.80) on semiquantitative metrics was reported. Based

Fig. 7 Impact of pretreatment
SUV_{max} on locoregional control

Study	RR (95% CI)
Chen, 2014	2.67 (0.59-12.2)
Katahira-Suzuki, 2015	0.86 (0.22-3.32)
Schwartz, 2015	0.31 (0.10-0.97)
SRR (I-squared=60%, p=0.755)	0.82 (0.24-2.79)
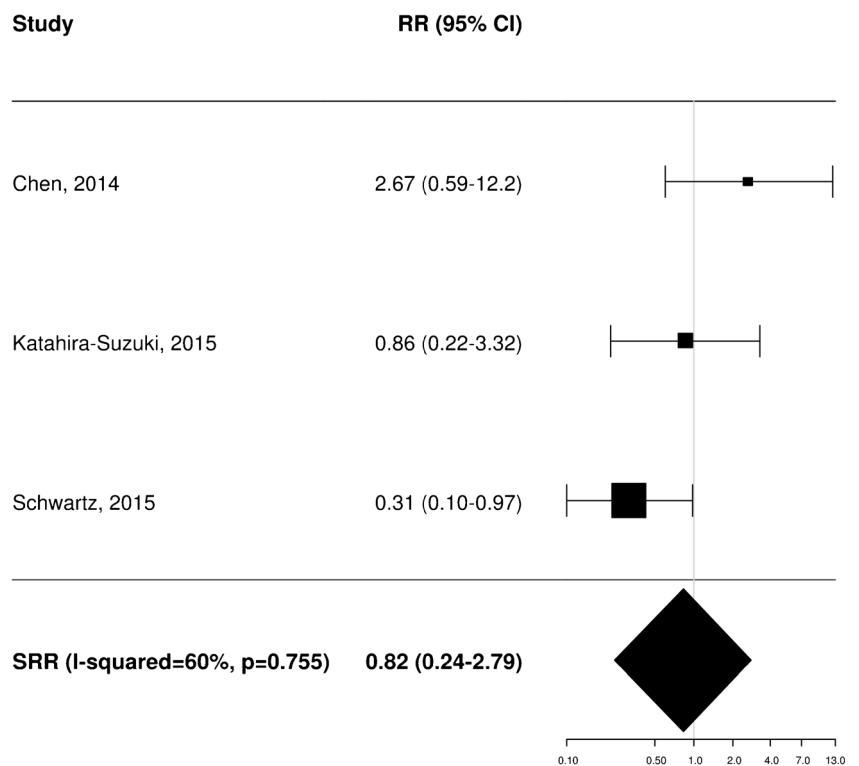
on voxels with a minimum of 40% SUV$_{max}$, baseline primary MTV above the median was the strongest prognosticator of worse LRC (HR 4.01, 95% CI 1.28–12.52, $p = 0.2$). Other retrospective studies [23, 27, 37] have underlined the prognostic value of baseline MTV, reporting different cut-off values as most significant for a worse outcome (combined primary and nodal MTV >40 ml, >20 ml and >18 ml correlating with worse DFS [23], LRC and OS [27], and disease-specific survival [37], respectively). The prognostic value of MTV analysed as a continuous variable has also been reported.

In a single-centre retrospective analysis in 83 patients, Tang et al. [53] found that an increase in primary baseline MTV of 17 ml (from the 25th to the 75th percentile) was associated with a doubling of the risk of disease progression ($p = 0.0002$) and of death ($p = 0.0048$). Of note, combined primary and nodal MTV (as a continuous variable) was also associated with a shorter PFS (HR 4.23, $p < 0.0001$; CI not reported) and OS (HR 3.21, $p < 0.0029$; CI not reported) in the subgroup of 64 patients with p16-positive oropharyngeal cancer. In a larger cohort of 122 patients with oropharyngeal cancer, Castelli et al. [45] assessed whether the use of different absolute and relative thresholds of SUV$_{max}$ result in different discriminatory power of MTV. Using a 51% relative SUV$_{max}$ threshold, combined primary and nodal MTV was the only significant factor in a multivariate analysis predicting OS (HR 1.43 per 10 ml, CI 1.23–1.65, $p < 0.001$) and DFS (HR 1.43 per 10 ml;,CI 1.23–1.65, $p = 0.03$). The optimal cut-off value for MTV 51% was 22.7 ml, which was able to discriminate 2-year DFS with rates of 63.3% versus 32.9% and LRC with rates of 68% versus 35.3%.

The absence of a consensus methodology on VOI delineation is clearly a limitation when comparing different datasets on the prognostic relevance of MTV, since no single optimal cut-off value is recognized. In line with previous experience, our data reinforce the prognostic role of pretreatment MTV as the most informative semiquantitative metabolic feature. In line with our search inclusion criteria, the patient population analysed was extremely homogeneous (about 95% of the whole sample size) in terms of disease stage, radiotherapy technique used and schedule of concomitant chemoradiotherapy. With all due limitations, our analysis provides further evidence on the predominant impact of pretreatment MTV on HNSCC outcome compared with all other available FDG PET metrics. Further consideration of its role also as a predictive biomarker may be generated by pattern-of-failure data correlating baseline FDG PET and radiation dose distribution in HNSCC. Due et al. [54] performed a retrospective analysis in 304 HNSCC patients with the aim of correlating the pattern of disease failure with FDG uptake on pretreatment PET scans. By performing a deformable registration of CT scans acquired at the time of recurrence with the planning PET/CT scan, the authors showed that 96% of relapses (95% CI 86–99%) occurred in the high-dose region. In addition, they found that recurrence density was higher in the central part of the

target volume ($p < 0.0001$), with a significant correlation with increasing FDG avidity ($p = 0.036$). In a smaller cohort of 44 patients enrolled in a prospective phase 2 trial, Leclerc et al. [55] showed that all ten recurrences arose in areas receiving >95% of the dose determined on PET-based plans. A similar finding was reported by Mohamed et al. [56], who hypothesized that a 1-cm margin in addition to the 50% SUV$_{max}$ isocontour on pretreatment FDG PET scans would cover the majority of type A recurrences (according to the authors' definition, those that arise in the central high-dose area).

Once again, it has to be underlined that, among others, the main limitations of FDG in HNSCC are its suboptimal specificity and the large variability in segmentation methods. Potentially, it could be hypothesized that hypoxia PET [57] and diffusion-weighted magnetic resonance imaging [58] may be more refined imaging biomarkers in the field of HNSCC. However, conclusive results on their prognostic impact have long been awaited, mainly due to the lack of reproducibility and cost issues preventing their adoption on a large scale. In our opinion FDG PET will remain the most widespread functional imaging modality used in clinical practice for many years to come.

Conclusion

The absence of prognostic biomarkers is a critical limitation in the management of locally advanced HNSCC. With all due limitations, our analysis showed that MTV defined from pretreatment FDG PET scans has the strongest impact on patient outcome after standard concurrent chemoradiotherapy. Prospective studies to corroborate this finding through standardized FDG PET acquisition and segmentation methods are warranted.

Compliance with ethical standards

References

1. Kamangar F, Dores GM, Anderson WF. Patterns of cancer incidence, mortality, and prevalence across five continents: defining priorities to reduce cancer disparities in different geographic regions of the world. J

Clin Oncol. 2006;24:2137–50. https://doi.org/10.1200/JCO.2005.05.2308.

2. Machtay M, Moughan J, Trotti A, Garden AS, Weber RS, Cooper JS, et al. Factors associated with severe late toxicity after concurrent chemoradiation for locally advanced head and neck cancer: an RTOG analysis. J Clin Oncol. 2008;26:3582–9. https://doi.org/10.1200/JCO.2007.14.8841.

3. Pulte D, Brenner H. Changes in survival in head and neck cancers in the late 20th and early 21st century: a period analysis. Oncologist. 2010;15:994–1001. https://doi.org/10.1634/theoncologist.2009-0289.

4. Gillison ML, Zhang Q, Jordan R, Xiao W, Westra WH, Trotti A, et al. Tobacco smoking and increased risk of death and progression for patients with p16-positive and p16-negative oropharyngeal cancer. J Clin Oncol. 2012;30:2102–11. https://doi.org/10.1200/JCO.2011.38.4099.

5. Ang KK, Harris J, Wheeler R, Weber R, Rosenthal DI, Nguyen-Tan PF, et al. Human papillomavirus and survival of patients with oropharyngeal cancer. N Engl J Med. 2010;363:24–35. https://doi.org/10.1056/NEJMoa0912217.

6. Cancer Genome Atlas Network. Comprehensive genomic characterization of head and neck squamous cell carcinomas. Nature. 2015;517(7536):576–82. https://doi.org/10.1038/nature14129.

7. Bentzen SM. Theragnostic imaging for radiation oncology: dose-painting by numbers. Lancet Oncol. 2005;6:112–7.

8. Ling CC, Humm J, Larson S, Amols H, Fuks Z, Leibel S, et al. Towards multidimensional radiotherapy (MD-CRT): biological imaging and biological conformality. Int J Radiat Oncol Biol Phys. 2000;47:551–60.

9. Daisne J-F, Duprez T, Weynand B, Lonneux M, Hamoir M, Reychler H, et al. Tumor volume in pharyngolaryngeal squamous cell carcinoma: comparison at CT, MR imaging, and FDG PET and validation with surgical specimen. Radiology. 2004;233:93–100.

10. Lonneux M, Hamoir M, Reychler H, Maingon P, Duvillard C, Calais G, et al. Positron emission tomography with [18F]fluorodeoxyglucose improves staging and patient management in patients with head and neck squamous cell carcinoma: a multicenter prospective study. J Clin Oncol. 2010;28:1190–5. https://doi.org/10.1200/JCO.2009.24.6298.

11. Mehanna H, Wong WL, 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(15):1444–54. https://doi.org/10.1056/NEJMoa1514493.

12. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. J Clin Epidemiol. 2009;62(10):e1–34. https://doi.org/10.1016/j.jclinepi.2009.06.006.

13. Aslam S, Emmanuel P. Formulating a researchable question: a critical step for facilitating good clinical research. Indian J Sex Transm Dis. 2010;31(1):47–50. https://doi.org/10.4103/0253-7184.69003.

14. Boellaard R, Delgado-Bolton R, Oyen WJG, Giammarile F, Tatsch K, Eschner W, et al. FDG PET/CT: EANM procedure guidelines for tumour imaging: version 2.0. Eur J Nucl Med Mol Imaging. 2014;42:328–54. https://doi.org/10.1007/s00259-014-2961-x.

15. Greenland S. Quantitative methods in the review of epidemiologic literature. Epidemiol Rev. 1987;9:1–30.

16. Begg C, Mazumdar M. Operating characteristics of a rank correlation test for publication bias. Biometrics. 1994;50(4):1088–101.

17. Egger M, Davey Smith G, Schneider M, et al. Bias in meta-analysis detected by a simple, graphical test. BMJ. 1997;315(7109):629–34.

18. Akagunduz OO, Savas R, Yalman D, Kocacelebi K, Esassolak M. Can adaptive threshold-based metabolic tumor volume (MTV) and lean body mass corrected standard uptake value (SUL) predict prognosis in head and neck cancer patients treated with definitive radiotherapy/chemoradiotherapy? Nucl Med Biol. 2015;42:899–904. https://doi.org/10.1016/j.nucmedbio.2015.06.007.

19. Castaldi P, Rufini V, Bussu F, Miccichè F, Dinapoli N, Autorino R, et al. Can "early" and "late"18F-FDG PET-CT be used as prognostic factors for the clinical outcome of patients with locally advanced head and neck cancer treated with radio-chemotherapy? Radiother Oncol. 2012;103:63–8. https://doi.org/10.1016/j.radonc.2012.03.001.

20. Castelli J, Depeursinge A, Ndoh V, Prior JO, Ozsahin M, Devillers A, et al. A PET-based nomogram for oropharyngeal cancers. Eur J Cancer. 2017;75:222–30. https://doi.org/10.1016/j.ejca.2017.01.018.

21. Chen S-W, Hsieh T-C, Yen K-Y, Yang S-N, Wang Y-C, Chien C-R, et al. Interim FDG PET/CT for predicting the outcome in patients with head and neck cancer. Laryngoscope. 2014;124:2732–8. https://doi.org/10.1002/lary.24826.

22. Cheng N-M, Dean Fang Y-H, Tung-Chieh Chang J, Huang C-G, Tsan D-L, Ng S-H, et al. Textural features of pretreatment 18F-FDG PET/CT images: prognostic significance in patients with advanced T-stage oropharyngeal squamous cell carcinoma. J Nucl Med. 2013;54:1703–9. https://doi.org/10.2967/jnumed.112.119289.

23. Man KC, Jeong HS, Sang GP, Jeon YJ, Son YI, Joon YC, et al. Metabolic tumor volume of [18F]-fluorodeoxyglucose positron emission tomography/computed tomography predicts short-term outcome to radiotherapy with or without chemotherapy in pharyngeal cancer. Clin Cancer Res. 2009;15:5861–8. https://doi.org/10.1158/1078-0432.CCR-08-3290.

24. Higgins KA, Hoang JK, Roach MC, Chino J, Yoo DS, Turkington TG, et al. Analysis of pretreatment FDG-PET SUV parameters in head-and-neck cancer: tumor SUV mean has superior prognostic value. Int J Radiat Oncol Biol Phys. 2012;82:548–53. https://doi.org/10.1016/j.ijrobp.2010.11.050.

25. Katahira-Suzuki R, Hata M, Tateishi U, Taguchi T, Takano S, Omura-Minamisawa M, et al. Definitive chemo-radiotherapy for squamous cell carcinoma of the pharynx: impact of baseline low hemoglobin level (<12 g/dL) and post-radiation therapy F-18 FDG-PET/CT. Ann Nucl Med. 2015;29:37–45. https://doi.org/10.1007/s12149-014-0907-9.

26. Kim R, Ock C-Y, Keam B, Kim TM, Kim JH, Paeng JC, et al. Predictive and prognostic value of PET/CT imaging post-chemoradiotherapy and clinical decision-making consequences in locally advanced head & neck squamous cell carcinoma: a retrospective study. BMC Cancer. 2016;16:116. https://doi.org/10.1186/s12885-016-2147-y.

27. Koyasu S, Nakamoto Y, Kikuchi M, Suzuki K, Hayashida K, Itoh K, et al. Prognostic value of pretreatment 18F-FDG PET/CT parameters including visual evaluation in patients with head and neck squamous cell carcinoma. AJR Am J Roentgenol. 2014;202:851–8. https://doi.org/10.2214/AJR.13.11013.

28. Lin P, Min M, Lee M, Holloway L, Forstner D, Bray V, et al. Nodal parameters of FDG PET/CT performed during radiotherapy for locally advanced mucosal primary head and neck squamous cell carcinoma can predict treatment outcomes: SUVmean and response rate are useful imaging biomarkers. Eur J Nucl Med Mol Imaging. 2017;44:801–11. https://doi.org/10.1007/s00259-016-3584-1.

29. 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. https://doi.org/10.2967/jnumed.113.136796.

30. Matoba M, Tuji H, Shimode Y, Kondo T, Oota K, Tonami H. The role of changes in maximum standardized uptake value of FDG PET-CT for post-treatment surveillance in patients with head and neck squamous cell carcinoma treated with chemoradiotherapy: preliminary findings. Br J Radiol. 2017;90(1071):20150404. https://doi.org/10.1259/bjr.20150404.

31. Min M, Lin P, Lee M, Ho Shon I, Lin M, Forstner D, et al. Prognostic value of 2-[18F] fluoro-2-deoxy-D-glucose positron emission tomography-computed tomography scan carried out

during and after radiation therapy for head and neck cancer using visual therapy response interpretation criteria. Clin Oncol. 2016;28:393–401. https://doi.org/10.1016/j.clon.2015.12.026.

32. Min M, Lin P, Lee MT, Shon IH, Lin M, Forstner D, et al. Prognostic role of metabolic parameters of 18F-FDG PET-CT scan performed during radiation therapy in locally advanced head and neck squamous cell carcinoma. Eur J Nucl Med Mol Imaging. 2015;42:1984–94. https://doi.org/10.1007/s00259-015-3104-8.

33. Miyabe J, Hanamoto A, Inohara H. Prognostic value of metabolic tumor volume in head and neck cancer. Otolaryngol Head Neck Surg. 2009;155(1S).

34. Miyabe J, Hanamoto A, Tatsumi M, Hamasaki T, Takenaka Y, Nakahara S, et al. Metabolic tumor volume of primary tumor predicts survival better than T classification in the larynx preservation approach. Cancer Sci. 2017;108:2030–8. https://doi.org/10.1111/cas.13345.

35. Murphy JD, La TH, Chu K, Quon A, FIschbein NJ, Maxim PG, et al. Postradiation metabolic tumor volume predicts outcome in head-and-neck cancer. Int J Radiat Oncol Biol Phys. 2011;80(2):514–21. https://doi.org/10.1016/j.ijrobp.2010.01.057.

36. Ng SH, Liao CT, Lin CY, Chan SC, Lin YC, Yen TC, et al. Dynamic contrast-enhanced MRI, diffusion-weighted MRI and 18F-FDG PET/CT for the prediction of survival in oropharyngeal or hypopharyngeal squamous cell carcinoma treated with chemo-radiation. Eur Radiol. 2016;26:4162–72. https://doi.org/10.1007/s00330-016-4276-8.

37. Park GC, Kim JS, Roh J-L, Choi S-H, Nam SY, Kim SY. Prognostic value of metabolic tumor volume measured by 18F-FDG PET/CT in advanced-stage squamous cell carcinoma of the larynx and hypopharynx. Ann Oncol. 2013;24:208–14. https://doi.org/10.1093/annonc/mds247.

38. Rasmussen JH, Vogelius IR, Fischer BM, Friborg J, Aznar MC, Persson GF, et al. Prognostic value of 18F-fludeoxyglucose uptake in 287 patients with head and neck squamous cell carcinoma. Head Neck. 2015;37(9):1274–81. https://doi.org/10.1002/hed.23745.

39. Romesser PB, Lim R, Spratt DE, Setton J, Riaz N, Lok B, et al. The relative prognostic utility of standardized uptake value, gross tumor volume, and metabolic tumor volume in oropharyngeal cancer patients treated with platinum based concurrent chemoradiation with a pre-treatment [(18)F] fluorodeoxyglucose positron emission tomography scan. Oral Oncol. 2014;50(9):802–8. https://doi.org/10.1016/j.oraloncology.2014.06.018.

40. Schwartz DL, Harris J, Yao M, Rosenthal DI, Opanowski A, Levering A, et al. Metabolic tumor volume as a prognostic imaging-based biomarker for head-and-neck cancer: pilot results from Radiation Therapy Oncology Group protocol 0522. Int J Radiat Oncol Biol Phys. 2015;91:721–9. https://doi.org/10.1016/j.ijrobp.2014.12.023.

41. Van Den Wyngaert T, Helsen N, Carp L, De Bree R, Martens MJ, Van Laer C, et al. ECLYPS: multicenter trial of FDG-PET/CT to detect residual nodal disease in locally advanced head-and-neck squamous cell carcinoma (LAHNSCC) after chemoradiotherapy (CRT). J Clin Oncol. 2016;34(15_suppl):6021. https://doi.org/10.1200/JCO.2016.34.15_suppl.6021.

42. Van Den Wyngaert T, Helsen N, Carp L, Hakim S, Martens MJ, Hutsebaut I, et al. Fluorodeoxyglucose-positron emission tomography/computed tomography after concurrent chemoradiotherapy in locally advanced head-and-neck squamous cell cancer: the ECLYPS study. J Clin Oncol. 2017;35:3458–64. https://doi.org/10.1200/JCO.2017.73.5845.

43. Yoon Y-H, Lee S-H, Hong S-L, Kim S-J, Roh H-J, Cho K-S. Prognostic value of metabolic tumor volume as measured by fluorine-18-fluorodeoxyglucose positron emission tomography/computed tomography in nasopharyngeal carcinoma. Int Forum Allergy Rhinol. 2014;4:845–50. https://doi.org/10.1002/alr.21363.

44. Zschaeck S, Löck S, Leger S, Haase R, Bandurska-Luque A, Appold S, et al. FDG uptake in normal tissues assessed by PET during treatment has prognostic value for treatment results in head and neck squamous cell carcinomas undergoing radiochemotherapy. Radiother Oncol. 2017;122:437–44. https://doi.org/10.1016/j.radonc.2017.01.019.

45. Castelli J, Depeursinge A, De Bari B, Devillers A, De Crevoisier R, Bourhis J, et al. Metabolic tumor volume and total lesion glycolysis in oropharyngeal cancer treated with definitive radiotherapy which threshold is the best predictor of local control? Clin Nucl Med. 2017;42:e281–5. https://doi.org/10.1097/RLU.0000000000001614.

46. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011;155(8):529–36.

47. O'Connor JP, Aboagye EO, Adams JE, Aerts HJ, Barrington SF, Beer AJ, et al. Imaging biomarker roadmap for cancer studies. Nat Rev Clin Oncol. 2017;14(3):169–86. https://doi.org/10.1038/nrclinonc.2016.162.

48. Halfpenny W, Hain SF, Biassoni L, Maisey MN, Sherman JA, McGurk M. FDG-PET. A possible prognostic factor in head and neck cancer. Cancer. 2002;86:512–6. https://doi.org/10.1038/sj.bjc.6600114.

49. Thorwarth D. Functional imaging for radiotherapy treatment planning: current status and future directions – a review. Br J Radiol. 2015;88(1051):20150056. https://doi.org/10.1259/bjr.20150056.

50. Garibaldi C, Ronchi S, Cremonesi M, Gilardi L, Travaini L, Ferrari M, et al. Interim 18F-FDG PET/CT during chemoradiation therapy in the management of head and neck cancer patients: a systematic review. Radiat Oncol Biol. 2017;98:555–73. https://doi.org/10.1016/j.ijrobp.2017.02.217.

51. Pak K, Cheon GJ, Nam H-Y, Kim S-J, Kang KW, Chung J-K, et al. Prognostic value of metabolic tumor volume and total lesion glycolysis in head and neck cancer: a systematic review and meta-analysis. J Nucl Med. 2014;55:884–90. https://doi.org/10.2967/jnumed.113.133801.

52. Schinagl DA, Span PN, Oyen WJ, Kaanders JH. Can FDG PET predict radiation treatment outcome in head and neck cancer? Results of a prospective study. Eur J Nucl Med Mol Imaging. 2011;38(8):1449–58. https://doi.org/10.1007/s00259-011-1789-x.

53. Tang C, Murphy JD, Khong B, La TH, Kong C, Fischbein NJ, et al. Validation that metabolic tumor volume predicts outcome in head-and-neck cancer. Radiat Oncol Biol. 2012;83:1514–20. https://doi.org/10.1016/j.ijrobp.2011.10.023.

54. Due AK, Vogelius IR, Aznar MC, Bentzen SM, Berthelsen AK, Korreman SS, et al. Recurrences after intensity modulated radiotherapy for head and neck squamous cell carcinoma more likely to originate from regions with high baseline [18F]-FDG uptake. Radiother Oncol. 2014;111:360–5. https://doi.org/10.1016/j.radonc.2014.06.001.

55. Leclerc M, Lartigau E, Lacornerie T, Daisne J, Kramar A, Grégoire V. Primary tumor delineation based on 18 FDG PET for locally advanced head and neck cancer treated by chemo-radiotherapy. Radiother Oncol. 2015;116:87–93. https://doi.org/10.1016/j.radonc.2015.06.007.

56. Mohamed ASR, Cardenas CE, Garden AS, Awan MJ, Rock CD, Westergaard SA, et al. Patterns-of-failure guided biological target volume definition for head and neck cancer patients: FDG-PET and dosimetric analysis of dose escalation candidate subregions. Radiother Oncol. 2017;124:248–55. https://doi.org/10.1016/j.radonc.2017.07.017.

57. Welz S, Mönnich D, Pfannenberg C, Nikolaou K, Reimold M, La C, et al. Prognostic value of dynamic hypoxia PET in head and neck cancer: results from a planned interim analysis of a randomized phase II hypoxia-image guided dose escalation trial. Radiother Oncol. 2017;124:526–32. https://doi.org/10.1016/j.radonc.2017.04.004.

58. Lambrecht M, Van Calster B, Vandecaveye V, De Keyzer F, Roebben I, Hermans R, et al. Integrating pretreatment diffusion weighted MRI into a multivariable prognostic model for head and neck squamous cell carcinoma. Radiother Oncol. 2014;110(3):429–34. https://doi.org/10.1016/j.radonc.2014.01.004.