The Journal of Laryngology & Otology

cambridge.org/jlo

Systematic review and meta-analysis of the diagnostic effectiveness of positron emission tomography-computed tomography versus magnetic resonance imaging in the post-treatment surveillance of head and neck squamous cell carcinoma

Y Zhu[1], O McLaren[1], J Hardman[2], J Evans[3] and R Williams[1]

[1]Department of ENT, University Hospitals Plymouth NHS Trust, [2]Head and Neck Unit, The Royal Marsden NHS Foundation Trust, London and [3]Musculoskeletal Research Unit, University of Bristol, UK

Review Article

Dr Y Zhu takes responsibility for the integrity of the content of the paper

Cite this article: Zhu Y, McLaren O, Hardman J, Evans J, Williams R. Systematic review and meta-analysis of the diagnostic effectiveness of positron emission tomography-computed tomography versus magnetic resonance imaging in the post-treatment surveillance of head and neck squamous cell carcinoma. *J Laryngol Otol* 2023;**137**:22–30. https://doi.org/10.1017/S0022215122000317

Accepted: 24 January 2022
First published online: 28 January 2022

Key words:
Squamous Cell Carcinoma of Head And Neck; Positron Emission Tomography Computed Tomography; Magnetic Resonance Imaging; Diffusion Magnetic Resonance Imaging; Follow-Up Studies

Author for correspondence:
Dr Y Zhu, Department of ENT, University Hospitals Plymouth NHS Trust, Derriford Rd, Plymouth PL6 8DH, UK
E-mail: yinan.zhu@nhs.net

Abstract

Objective. There is currently no consensus on the ideal protocol of imaging for post-treatment surveillance of head and neck squamous cell carcinoma. This study aimed to consolidate existing evidence on the diagnostic effectiveness of positron emission tomography-computed tomography versus magnetic resonance imaging.

Method. Systematic electronic searches were conducted using Medline, Embase and Cochrane Library (updated February 2021) to identify studies directly comparing positron emission tomography-computed tomography and magnetic resonance imaging scans for detecting locoregional recurrence or residual disease for post-treatment surveillance.

Results. Searches identified 3164 unique records, with three studies included for meta-analysis, comprising 176 patients. The weighted pooled estimates of sensitivity and specificity for scans performed three to six months post-curative treatment were: positron emission tomography-computed tomography, 0.68 (95 per cent confidence interval, 0.49–0.84) and 0.89 (95 per cent confidence interval, 0.84–0.93); magnetic resonance imaging, 0.72 (95 per cent confidence interval, 0.54–0.88) and 0.85 (95 per cent confidence interval, 0.79–0.89), respectively.

Conclusion. Existing studies do not provide evidence for superiority of either positron emission tomography-computed tomography or magnetic resonance imaging in detecting locoregional recurrence or residual disease following curative treatment of head and neck squamous cell carcinoma.

Introduction

Head and neck squamous cell carcinoma (SCC) is one of the 10 most common malignancies in the UK.[1] Following treatment with curative intent, optimal surveillance for survivors of head and neck SCC is an essential element of patient care.[2] Locoregional recurrence is highest during the first three years post-curative treatment and is the greatest cause of mortality in this period.[3] Early identification of recurrence improves the chance of salvage surgery being a treatment option, which can achieve a 5-year disease free survival rate as high as 39 per cent. Primary site or nodal recurrence may be hidden beneath intact mucosa in anatomically distorted areas post-irradiation or reconstruction, making identification on clinical examination challenging.[3]

Radiological investigations can add vital early information on the response to treatment and the recurrence of disease often before it may be clinically detectable. Positron emission tomography-computed tomography (PET-CT) and magnetic resonance imaging (MRI) are commonly used. Positron emission tomography-computed tomography with 2-deoxy-2-[fluorine-18]fluoro-D-glucose (FDG) guided surveillance can reduce the need for salvage surgery following oncological treatment and is more cost effective compared with elective neck dissection alone, with similar survival outcomes.[4] It has consistently demonstrated a high sensitivity and negative predictive value for the presence of recurrent or residual disease.[5,6] MRI offers a superior delineation of soft tissue compared with other imaging modalities[6] without radiation exposure. This also aids surgical or radiotherapy treatment planning. Newer diffusion-weighted MRI sequences generate better image contrast between post-treatment tissue inflammation or fibrosis and tumour recurrence or persistence, and it is increasingly employed.[7,8]

Although several studies have investigated the different imaging modalities, to date there have been no systematic reviews or meta-analyses performed to directly compare PET-CT and MRI. There is no consensus in either the National Comprehensive Cancer Network guidance[9] or UK National Multidisciplinary Guidelines[2] on which

© The Author(s), 2022. Published by Cambridge University Press on behalf of J.L.O. (1984) LIMITED

imaging modality is better for the post-treatment surveillance of head and neck SCC. This study aimed to consolidate existing evidence to identify if PET-CT or MRI is superior at detecting locoregional recurrence or residual disease in the post-treatment surveillance of head and neck SCC.

Materials and methods

Protocol and registration

This systematic review was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses ('PRISMA') statement with the diagnostic test accuracy extension[10] as well as guidance from the Cochrane diagnostic test accuracy protocol.[11] The protocol was registered with Prospero (CRD42021219840) before the search was conducted.

Eligibility criteria

Participants

Adults with histologically proven primary head and neck SCC who had undergone any treatment modality with curative intent were included. Studies were excluded if primarily reviewing patients with nasopharyngeal or non-squamous head and neck cancer. Studies focusing on palliative treatment or those with incomplete treatment were also excluded.

Setting

All countries and health systems were considered.

Index tests

FDG PET-CT and MRI performed on the same cohort and directly compared were included. Studies evaluating the role of PET-CT or MRI imaging in patients with clinically suspected recurrence were excluded as these patients represent a different patient cohort with a higher prevalence of recurrence.

Reference standard

Histological confirmation was used for a positive index test. Histological confirmation or clinical follow up was used for at least six months for a negative index test. Ideally, the reference standard across all imaging modalities compared should be a 'complete pathological response', that is, histological confirmation of head and neck SCC. However, invasive procedures to obtain this come with operative risks, and it would be difficult to justify in the case of a low suspicion of recurrence. Hence, a complete clinical response, as defined by the National Comprehensive Cancer Network guidance,[9] that includes no visible or palpable residual neck disease and the absence of concerning findings on imaging can be considered a suitable standard for a negative index test. We further define a duration of a minimum of six months[12] for such a follow up to be considered a negative index test.

Target condition

The target condition was recurrent or residual head and neck SCC, including the primary site and regional neck nodal disease. Local recurrence was defined as regrowth of the tumour at the primary tumour site or surgical bed, and regional recurrence was defined as regrowth of the tumour within cervical lymph nodes.[13] Residual tumour was defined as a tumour left behind after definitive treatment.

Study design

All types of experimental and observational studies were considered, including retrospective and prospective designs.

Report characteristics

Articles in English or with English translation available with no limitations on dates or periods of study or recruitment were considered.

Search and information sources

Sources searched included the following databases: Medline and PubMed via the Ovid search platform as well as Cochrane Library. A scoping Boolean search was conducted with terms related to 'head and neck cancer', 'MRI', 'PET-CT' and 'surveillance of residual or recurrent disease'. These terms included a combination of free text and Medical Subject Headings adapted for each individual database searched. Searches were conducted in February 2021 with the full search strategy detailed in Appendix 1 and 2.

Study selection and data collection

The titles and abstracts of all studies were screened against the eligibility criteria independently by two authors (YZ and OM) on the Rayyan platform.[14] Full texts were sought when the study could not be screened by the title and abstract alone. Where any uncertainty or disagreement was encountered, the senior author (RW) was consulted for a final decision. One author (YZ) used a pre-planned data extraction proforma on Microsoft Excel® to extract data from eligible studies. This was vetted by a second author (OM) and final approval was given by the senior author (RW).

Risk of bias and applicability

The Quality Assessment of Diagnostic Accuracy Studies-2 tool[15] was used to assess risk of bias and applicability of these studies. Because the studies included were comparative diagnostic accuracy studies, the newer unpublished Quality Assessment of Diagnostic Accuracy Studies-2 Comparison tool was applied to enhance the quality screening and assessment of the included studies. Quality Assessment of Diagnostic Accuracy Studies-2 and Quality Assessment of Diagnostic Accuracy Studies-2 Comparison tools were tailored to fit this systematic review as intended by its authors, with the following changes.

Quality Assessment of Diagnostic Accuracy Studies-2 changes were: (1) each modality (PET-CT and MRI) was assessed separately; and (2) specified section 4.4 was: 'Were all patients for the respective imaging modality included in the final analysis comparing PET-CT *vs* MRI?'

Quality Assessment of Diagnostic Accuracy Studies-2 Comparison changes were: (1) questions C1.3 and C1.4 (randomisation not applicable) were removed; and (2) specified section C4.2 was: 'Was the interval between the index tests less than 1 month apart?'

Diagnostic accuracy measures

Where available, the diagnostic accuracy for both PET-CT and MRI was reported for each unit of assessment. This encompassed rates for sensitivity and specificity. We also report absolute numbers for true positives, false positives, false negatives

https://doi.org/10.1017/S0022215122000317 Published online by Cambridge University Press

and true negatives to allow for pooled analysis. Where these absolute numbers were not reported, they were deduced from the reported diagnostic accuracy rates and number of patients. The authors accepted a broad spectrum of definitions for the unit of assessment (per-primary tumour, per-hemi neck or per-node for nodal metastases to the neck), providing a direct comparison was made between PET-CT and MRI.

Meta-analysis

Data for individual studies fitting the inclusion criteria were summarised in a 2×2 table for both PET-CT and MRI. The derived rates for sensitivity and specificity for each imaging modality were calculated and pooled together using the inverse variance method, with the DerSimonian–Laird estimator for Tau,[2] Freeman–Tukey double arcsine transformation and Clopper–Pearson confidence interval for individual studies. This was performed using R statistical computing software (version 4.1.0; The R Foundation). A fixed effects model was used because of the small number of included studies (fewer than 10).

Sensitivity and specificity of a diagnostic test are linked and can be interpreted in conjunction. A summary receiver operator characteristic curve was plotted. The hierarchical bivariate binominal model was selected because it models the sensitivity and specificity of studies directly accounting for variation within and between studies.[16] This was performed using R software (version 4.1.0), as described by Cochrane Methods,[17] with Revman software (version 5.4.1; ReviewManager).

Results

Following the database searches based on the inclusion and exclusion criteria, the study selection process is outlined in Figure 1.[18] Study characteristics are presented in Tables 1 and 2.

Risk of bias and applicability

The risk of bias and applicability assessment was performed independently for each imaging modality using the Quality Assessment of Diagnostic Accuracy Studies-2 and is presented in Figure 2 and Figure 3. Figure 4 shows the Quality Assessment of Diagnostic Accuracy Studies-2 Comparison assessment in the comparison of PET-CT and MRI within a study. In general, although there were no major concerns with the applicability of the included papers, patient selection methods were vague in four out of the six included studies and did not mention if a consecutive sample of patients were enrolled. Only two studies explicitly mentioned blinding of observers to both the other index test and other observers.[5,21]

Individual study results

Qualitatively, the six studies[5,7,19–22] covered a range of anatomical sites for the primary cancer and treatment modalities used. Only two studies mentioned the status of human papilloma virus (HPV) in their patient characteristics.[5,7] Three of the six studies used diffusion-weighted MRI for analysis in addition to routine MRI protocols.[5,7,21] The timing of PET-CT and MRI scans performed were within six months of curative treatment for all six studies, and no more than three months apart within individual studies.

Three studies[5,19,21] reported the use of a scoring system to classify lesions suspicious of malignancy for both PET-CT and MRI. With the exception of the Hopkins criteria for PET-CT interpretation used in one study,[5] all papers used their own scoring system. Two studies explored the effects of using different cut-off points for index test positivity and found that a sensitive reading (positive index test for equivocal readings) produced the best combination of sensitivity and specificity for the detection of nodal disease using diffusion-weighted MRI.[5,21] Two studies[5,21] highlighted issues with inter-observer agreement and the role of consensus in the interpretation of images. One study[21] showed higher inter-observer variation for MRI as compared with PET-CT, whereas the other[5] showed the opposite. Both papers agree that negative agreement is higher than positive agreement regardless of the modality, reaching very good kappa values of more than 0.80. Schouten *et al.*[21] published data from individual observers, and it is noted that consensus does not necessarily improve diagnostic test accuracy compared with the single most accurate observer for both PET-CT and MRI.

Quantitatively, 3 studies were ultimately included in the meta-analysis, with 176 patients analysed for comparison. Ghanooni *et al.*[19] was excluded because the unit of assessment 'n' used was inconsistent for PET-CT and MRI. In this study, the target condition was defined as recurrence at various anatomical sites for primary tumour, adjacent extensions and lymph node regions. Although the total number of patients who underwent both scans were the same and the amalgamation of anatomical sites itself does not preclude exclusion, the fact that each anatomical site was not specified and directly compared for PET-CT versus MRI makes it so. The study by Yu *et al.*[7] was excluded because the number of patients was different for PET-CT and MRI. The patient number in the PET-CT group is a subset of the MRI group because 9 patients did not undergo PET-CT for various reasons. Unfortunately, there is no data for direct comparison. Breik *et al.*[22] was excluded because different sets of patients within the same study group underwent PET-CT and MRI at three months and six months and no direct head-to-head comparison was made at either of those intervals. Data within one study was merged. Noij *et al.*[5] presented two datasets, one for the imaging assessment of the most suspicious lymph node and another for assessment of the primary tumour. These were compared for PET-CT versus MRI directly and were merged for the meta-analysis. A visual representation of the individual true positives, false positives, false negatives, true negatives, sensitivity and specificity extracted from each included study are seen in Figure 5.

Analysis

Because of the small number of studies, the fixed effects model was used to calculate a pooled sensitivity and specificity. The weighted pooled estimates of sensitivity and specificity for PET-CT were 0.68 (95 per cent CI, 0.49 to 0.84) and 0.89 (95 per cent CI, 0.84 to 0.93), whereas for MRI they were 0.72 (95 per cent CI, 0.54 to 0.88) and 0.85 (95 per cent CI, 0.79 to 0.89), respectively. These are shown in Figure 6.

Summary receiver operator characteristic curve

Individual summary receiver operator characteristic curves were plotted for PET-CT and MRI, with each data point in the figure representing a separate study and paired data linked

https://doi.org/10.1017/S0022215122000317 Published online by Cambridge University Press

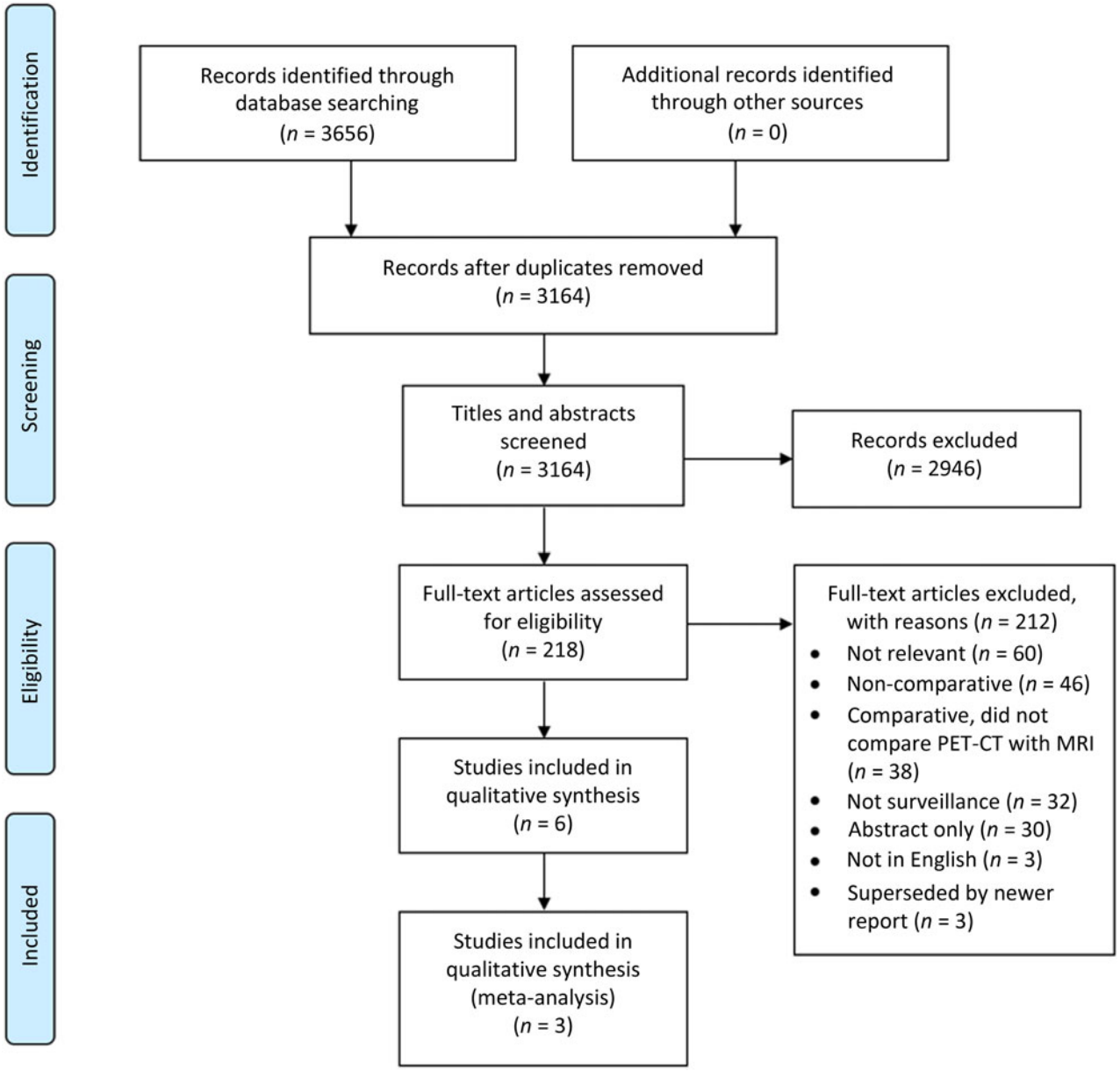

Fig. 1. Preferred reporting items for systematic reviews and meta-analyses ('PRISMA') flowchart with results of the database searches, screening and reasons for exclusion. PET-CT = positron emission tomography-computed tomography; MRI = magnetic resonance imaging.

with dotted lines. This is shown in Figure 7. The best operating point for MRI (red dot) is sensitivity 0.71 (95 per cent CI, 0.52 to 0.85) and specificity 0.84 (95 per cent CI, 0.73 to 0.91) and for PET-CT (black dot) is sensitivity 0.78 (95 per cent CI, 0.35 to 0.96) and specificity 0.89 (95 per cent CI, 0.82 to 0.94).

Discussion

Summary of evidence

There is overlap in the 95 per cent confidence intervals of weighted mean pooled estimates of both sensitivity for PET-CT (0.68; 95 per cent CI, 0.49–0.84) versus MRI (0.72; 95 per cent CI, 0.54–0.88) as well as specificity for (PET-CT, 0.89; 95 per cent CI, 0.84–0.93) versus MRI (0.85; 95 per cent CI, 0.79–0.89) of the two imaging modalities compared. Given the small number of studies, the shapes of the summary receiver operator characteristic curves for PET-CT and MRI are not useful, and the best operating point cannot be meaningfully interpreted. There is insufficient evidence to recommend one over the other for the role of surveillance imaging in recurrent or residual head and neck cancer.

The included studies also shed light on human and imaging factors affecting comparative diagnostic accuracy in the two modalities compared. First, in terms of inter-observer agreement for a single imaging modality, with a maximum kappa value of 1 implying perfect agreement, most of the inter-observer kappa values for PET-CT and MRI in our included studies fell within the moderate agreement category (kappa, 0.40–0.60). Post-treatment imaging interpretation is considered to be one of the most difficult aspects of head and neck radiology, and together with the subjective nature of qualitative imaging,[5] it may be difficult to obtain consensus even for experienced observers. In fact, it might be natural to assume that consensus would improve diagnostic accuracy. However, a consensus report may be vulnerable to factors such as groupthink and dominance by seniority of a more experienced observer,[23] leading to the contrary.

Future studies should report variability between observers and state if the study setting reflects clinical practice for

https://doi.org/10.1017/S0022215122000317 Published online by Cambridge University Press

https://doi.org/10.1017/S0022215122000317 Published online by Cambridge University Press

Table 1. Study characteristics 1

Study	Origin	Funding	Conflict of interest	Design	Patients (*n*)	Age (mean (range); years)	Primary cancer site (type)	Prior treatment	Patient presentation
Ghanooni *et al.* 2011[19]	Belgium	FRS-FNRS and National Lottery Belgium	No information	Prospective	32	59 (40–76)	Oropharynx, hypopharynx, larynx, oral cavity (SCC)	Surgery, CT, RT or combination	Not specified
Pellini *et al.* 2014[20]	Italy	No funding declared	No conflict declared	Prospective	36	61.4 (42–71)	Oropharynx (SCC)	Concurrent CRT	Surveillance
Schouten *et al.* 2015[21]	Netherlands	No funding declared	No conflict declared	Retrospective	58	57.6 (23–73)	Oropharynx, hypopharynx, larynx, nasopharynx, oral cavity (SCC)	CRT	Surveillance
Yu *et al.* 2017[7]	USA	No information	No information	Retrospective	41	58 (43–73)	Oropharynx (SCC)	Concurrent CRT	Surveillance
Noij *et al.* 2018[5]	Netherlands	Netherlands Organisation for Health Research & Development	No conflict declared	Retrospective	82	59.2 (43–81)	Oropharynx, hypopharynx, larynx, oral cavity, unknown primary (SCC)	RT with or without CT	Surveillance
Breik *et al.* 2020[22]	UK	No funding declared	No conflict declared	Retrospective	140	68	Oral cavity, oropharynx, hypopharynx (SCC)	Surgery, CT, RT or combination	Surveillance

FRS-FNRS = Fund for Scientific Research; SCC = squamous cell carcinoma; CT = chemotherapy; RT = radiotherapy; CRT = chemoradiotherapy

Table 2. Study characteristics 2

Study	Target condition	Modalities compared	Suitability for meta-analysis	Timing of PET-CT (range)	Timing of MRI (range)	PET-CT interpretation	MRI interpretation	Reference standard
Ghanooni *et al.* 2011[19]	Locoregional recurrence	18-F FDG PET-CT *vs* MRI	No	4 months	4 months	Visual and semi-quantitative interpretation (SUV_{max}) by 2 nuclear medicine physicians	Visual interpretation by 1 practitioner	Histopathology and/or clinical follow up (minimum 18 months)
Pellini *et al.* 2014[20]	Neck nodal recurrence	18-F FDG PET-CT *vs* MRI	Yes	12.4 (12–13) weeks	12.4 (12–13) weeks	Visual interpretation by one nuclear medicine physician with consensus by a radiologist	Visual interpretation by 4 radiologists in consensus	Histopathology from neck dissection 28 days post-scan
Schouten *et al.* 2015[21]	Neck nodal recurrence	18-F FDG PET-CT *vs* DW MRI	Yes	92.2 days	86.8 days	Visual interpretation by 2 nuclear medicine physicians	Visual interpretation by 2 radiologists	Histopathology and/or clinical follow up (9 months post-treatment clinically and/or radiologically)
Yu *et al.* 2017[7]	Neck nodal recurrence	18-F FDG PET-CT *vs* DW MRI	No	14 (7–43) weeks	8.6 (4–16) weeks	Visual and semi-quantitative interpretation (SUV_{max}) with personnel unspecified	Visual and semi-quantitative interpretation (ADC) with personnel unspecified	Histopathology and/or clinical follow up for 120 (19–260) weeks
Noij *et al.* 2018[5]	Primary tumour and most suspicious lymph node	18-F FDG PET-CT *vs* DW MRI	Yes	3–6 months	3–6 months	Visual interpretation by 1 nuclear medicine physician and 1 resident	Visual interpretation by 2 radiologists and quantitative interpretation (ADC) by a researcher and radiologist	Clinical follow up with targeted investigations if indicated for 26 (IQR 9–39) months
Breik *et al.* 2020[22]	Locoregional/ metastatic disease	PET-CT *vs* MRI	No	3–6 months	3–6 months	Unspecified	Unspecified	Minimum of 2.5 years of clinical follow up

18-F FDG = 2-deoxy-2-[fluorine-18]fluoro-D-glucose; PET-CT = positron emission tomography-computed tomography; MRI = magnetic resonance imaging; SUV = standard uptake value; max = maximum; DW = diffusion weighted; ADC = apparent diffusion coefficient; IQR = interquartile range

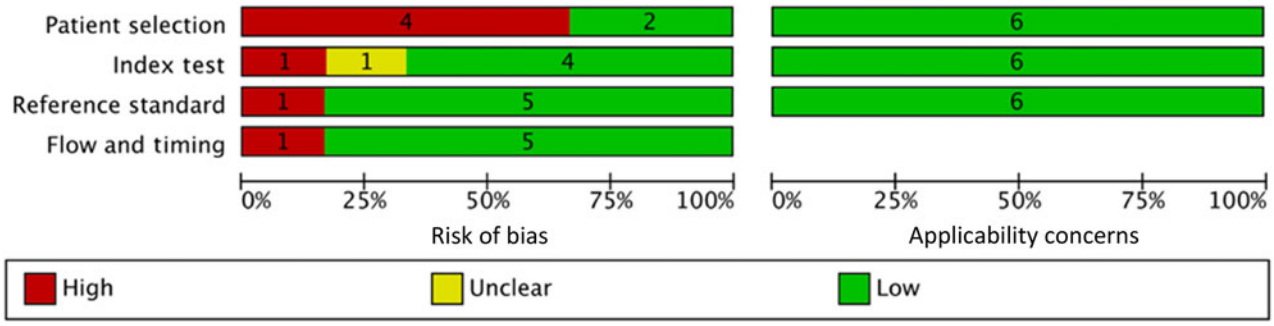

Fig. 2. Results of Quality Assessment of Diagnostic Accuracy Studies-2 tool for positron emission tomography-computed tomography.

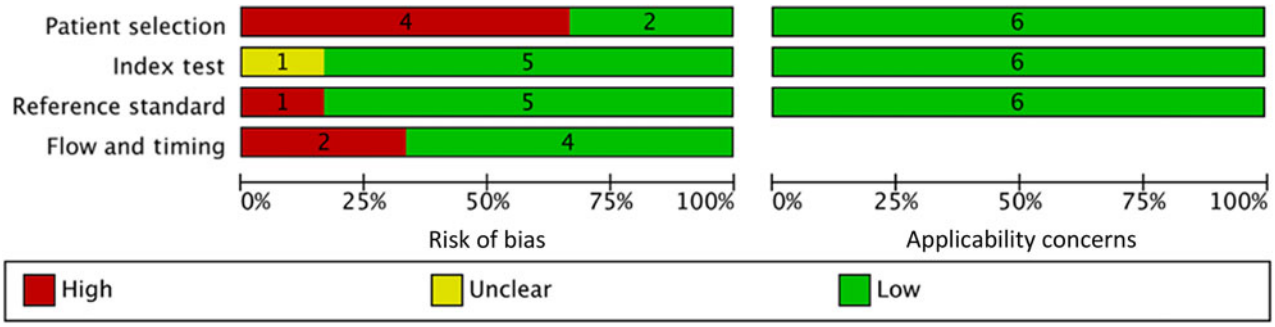

Fig. 3. Results of Quality Assessment of Diagnostic Accuracy Studies-2 tool for magnetic resonance imaging.

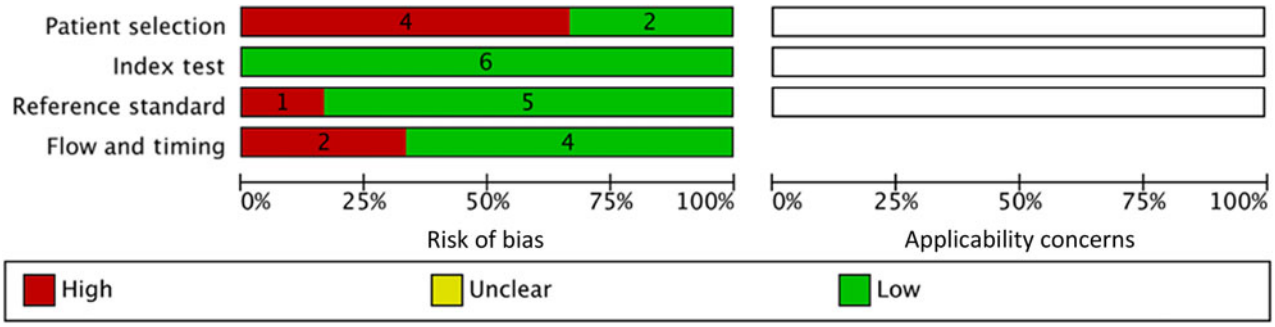

Fig. 4. Results of Quality Assessment of Diagnostic Accuracy Studies-2 Comparison tool for positron emission tomography-computed tomography versus magnetic resonance imaging.

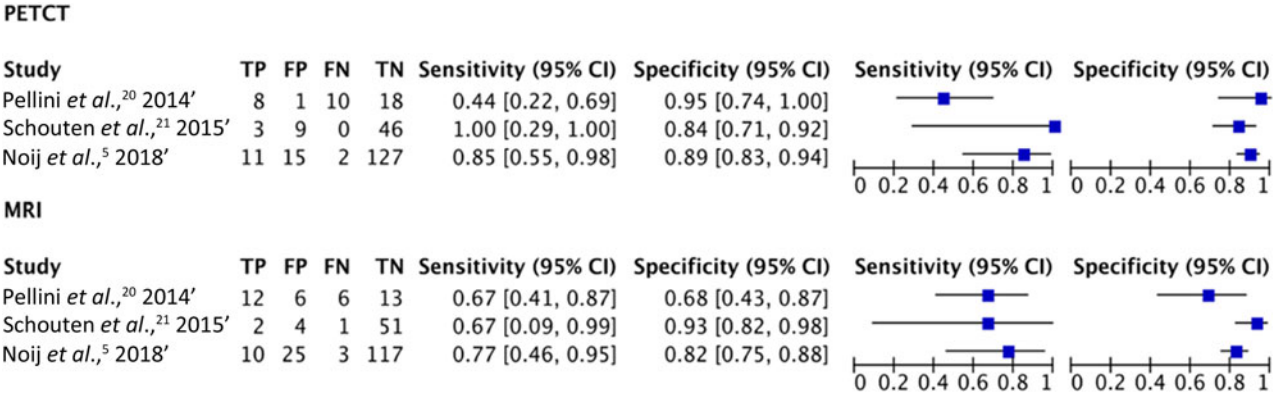

Fig. 5. Forest plot of individual studies included in meta-analysis. PET-CT = Positron emission tomography-computed tomography; TP = true positive; FP = false positive; FN = false negative; TN = true negative; CI = confidence interval; MRI = magnetic resonance imaging

more realistic and applicable results.[23] Next, observer blinding must also be stated clearly and should ideally be from the different imaging modalities, and if there is more than one person, to each other. Blinding of a radiologist from one modality to another is difficult to enforce in prospective studies. The use of one imaging modality would not prohibit the

https://doi.org/10.1017/S0022215122000317 Published online by Cambridge University Press

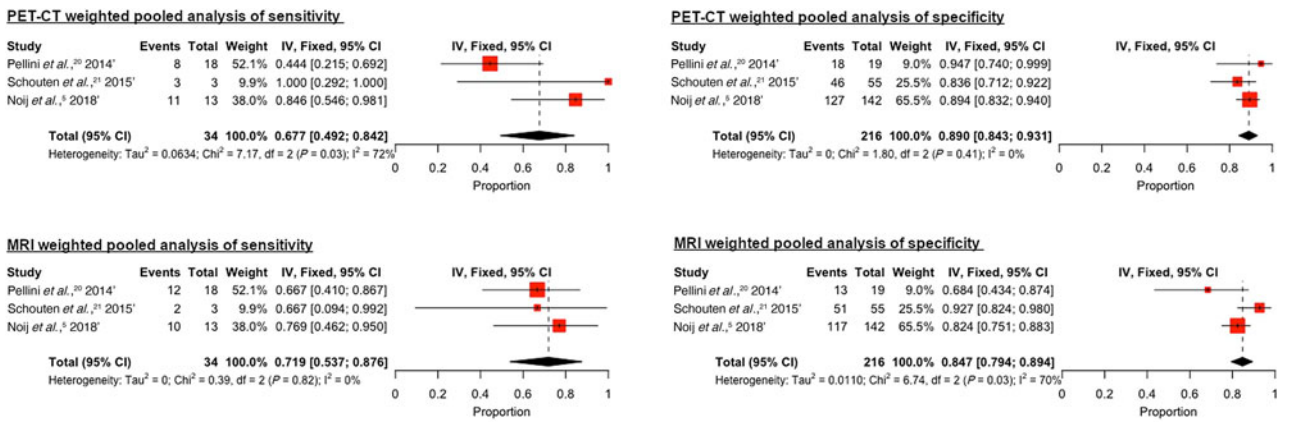

Fig. 6. Positron emission tomography-computed tomography (PET-CT) and magnetic resonance imaging (MRI) weighted pooled analysis of specificity and sensitivity using the fixed effects model. IV = inverse variance; CI = confidence interval

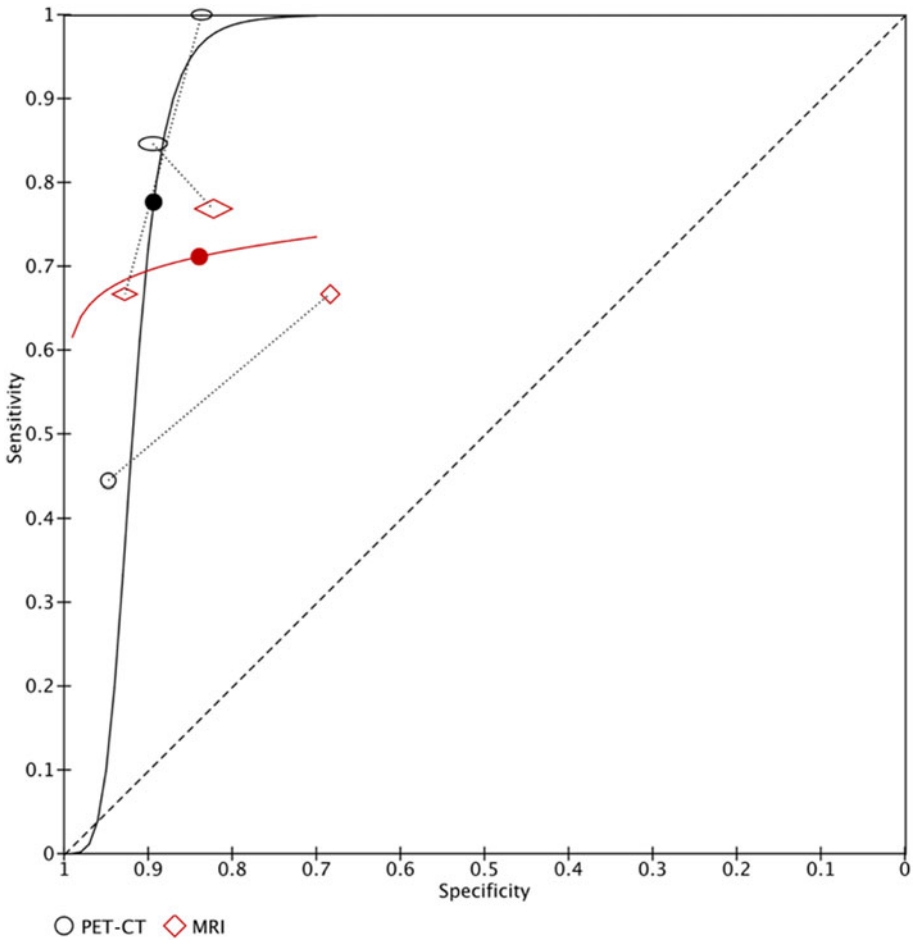

Fig. 7. Summary receiver operating characteristic curves for positron emission tomography-computed tomography (PET-CT) and magnetic resonance imaging (MRI).

use of another in clinical practice. In fact, the PET-CT and MRI scans may present unique complementary information that should not be concealed from the clinician. However, the lack of such blinding may mean that there is an additional element of confirmatory bias which may skew accuracy of the interpretation of index tests.[24]

Strengths

A key strength of our paper is that we included only studies performing a direct comparison of PET-CT and MRI within the same patient group to ensure a more homogeneous cohort and limit selection bias in line with the Cochrane guidance.[11] The other key feature is that we excluded studies with imaging performed for suspected recurrence rather than surveillance. When comparing with a non-comparative systematic review of PET-CT by Sheikhbahaei *et al.*[12] in 2015 that included studies with suspected recurrences, their sensitivity of 0.92 is higher and their specificity of 0.87 about the same. This shows that the sensitivity of PET-CT may be comparatively lower when used for surveillance. Although retrospective data obtained from imaging results for patients who had undergone a biopsy or resection was useful because of a consistent histopathological reference standard used, such as the

https://doi.org/10.1017/S0022215122000317 Published online by Cambridge University Press

protocol used by Kim *et al.*,[13] the prevalence of true positive disease in patient groups with suspected recurrence and surveillance is different and limits comparison. When comparing our findings to a non-comparative systematic review of PET-CT in a surveillance setting performed in 2011 by Gupta *et al.*,[25] their results of sensitivity and specificity of 0.73 and 0.88, respectively, for recurrence of neck disease are similar.

Limitations

One major limitation of this systematic review and meta-analysis is that because of the small numbers of studies analysed, further subgroup analysis for effects of patient characteristics including age, HPV status, site of primary cancer, modality of treatment, and study design characteristics including thresholds for index test positivity and reference standards used, is not possible. Each of these characteristics could affect diagnostic test accuracy. For example, the modality of treatment used (e.g. surgery, chemotherapy or radiotherapy) can present a different challenge to interpreting images.[26]

Another limitation is that studies that performed diffusion-weighted MRI were grouped together with MRI studies for comparison with PET-CT. For diffusion-weighted MRI, additional quantitative analysis can be performed with apparent diffusion coefficient values. A low apparent diffusion coefficient value represents increased cellularity and higher impedance of water molecules through tissues associated with a tumour.[7] Although all 3 diffusion-weighted MRI studies also included traditional MRI sequences, 80 per cent (140 of 176) of the patient population included in the meta-analysis undertook diffusion-weighted MRI imaging, and the results would more accurately reflect a comparison between diffusion-weighted MRI and PET-CT. The paper is also heterogeneous in terms of defining the unit of assessment 'n' that was compared. Such units of assessment can include individual lymph nodes, hemi-neck levels or even individual patients, as along as a direct comparison is made in the same group of patients. Because of the small number of studies, although the comparative accuracy between PET-CT and MRI is not affected, instances where a particular index test may be better at assessing a specific 'n' are overlooked.

One deviation was made from the preregistered protocol: studies reporting solely on nasopharyngeal carcinomas were excluded as this was seen to be a unique subset of head and neck SCC with its own histopathological spectrum, geographical distribution and distinctive risk profile,[27] contributing to additional heterogeneity in the meta-analysis.

Conclusion

This was the first systematic review and meta-analysis to consider direct comparison of PET-CT and MRI in the same patients in the post-treatment surveillance of head and neck SCC without clinical suspicion of residual or recurrent disease. Existing studies do not provide evidence for superiority of either PET-CT or MRI in detecting locoregional recurrence or residual disease following curative intent treatment of head and neck SCC. Future imaging studies should focus on direct comparison of index tests, with appropriate subgroup analysis for the relevant patient and study design characteristics mentioned above. In addition, other factors including patient selection methods, blinding and consensus methods of observers need to be clearly specified to reduce risk of bias.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/S0022215122000317.

Acknowledgements. This research was undertaken as partial fulfilment of the requirements for the MSc in Surgical Sciences at The University of Edinburgh. Many thanks to search strategists Mary Smith (University of Exeter), Thomas Arnold (University of Plymouth) and Marshall Dozier (University of Edinburgh) for giving valuable feedback for the search strategy.

Competing interests. None declared

References

1 Cancer incidence for common cancers. In: https://www.cancerresearchuk.org/health-professional/cancer-statistics/incidence/common-cancers-compared#heading-Zero [20 July 2021]
2 Simo R, Homer J, Clarke P, Mackenzie K, Paleri V, Pracy P *et al.* Follow-up after treatment for head and neck cancer: United Kingdom National Multidisciplinary Guidelines. *J Laryngol Otol* 2016;**130**:S208–S11
3 Imbimbo M, Alfieri S, Botta L, Bergamini C, Gloghini A, Calareso G *et al.* Surveillance of patients with head and neck cancer with an intensive clinical and radiologic follow-up. *Otolaryngol Head Neck Surg* (*United States*) 2019;**161**:635–42
4 Mehanna H, Wong W-L, McConkey CC, Rahman JK, Robinson M, Hartley AGJ *et al.* PET-CT surveillance versus neck dissection in advanced head and neck cancer. *New England J Med* 2016;**374**:1444–54
5 Noij DP, Martens RM, Koopman T, Hoekstra OS, Comans EFI, Zwezerijnen B *et al.* Use of diffusion-weighted imaging and 18f-fluorodeoxyglucose positron emission tomography combined with computed tomography in the response assessment for (chemo)radiotherapy in head and neck squamous cell carcinoma. *Clin Oncol* 2018;**30**:780–92
6 Zhao X, Rao S. Surveillance imaging following treatment of head and neck cancer. *Seminars Oncol* 2017;**44**:323–9
7 Yu Y, Mabray M, Silveira W, Shen PY, Ryan WR, Uzelac A *et al.* Earlier and more specific detection of persistent neck disease with diffusion-weighted MRI versus subsequent PET/CT after definitive chemoradiation for oropharyngeal squamous cell carcinoma. *Head Neck* 2017;**39**:432–8
8 Vandecaveye V, De Keyzer F, Nuyts S, Deraedt K, Dirix P, Hamaekers P *et al.* Detection of head and neck squamous cell carcinoma with diffusion weighted MRI after (chemo)radiotherapy: correlation between radiologic and histopathologic findings. *Int J Radiation Oncol Biol Physics* 2007;**67**:960–71
9 Pfister DG, Spencer S, Adelstein D, Adkins D, Anzai Y, Brizel DM *et al.* Head and neck cancers, version 2.2020, NCCN Clinical Practice Guidelines in Oncology. *J Natl Compr Canc Netw* 2020;**18**:873–98
10 PRISMA for diagnostic test accuracy. In: http://www.prisma-statement.org/Extensions/DTA [24 June 2021]
11 Deeks JJ, Wisniewski S, Davenport C. Guide to the contents of a Cochrane Diagnostic Test Accuracy Protocol. In: https://methods.cochrane.org/sites/methods.cochrane.org.sdt/files/public/uploads/Ch04_Sep2013.pdf [19 July 2021]
12 Sheikhbahaei S, Taghipour M, Ahmad R, Fakhry C, Kiess AP, Chung CH *et al.* Diagnostic accuracy of follow-up FDG PET or PET/CT in patients with head and neck cancer after definitive treatment: a systematic review and meta-analysis. *Am J Roentgenol* 2015;**205**:629–39
13 Kim ES, Yoon DY, Moon JY, Baek S, Han YM, Seo YL *et al.* Detection of loco-regional recurrence in malignant head and neck tumors: a comparison of CT, MRI, and FDG PET-CT. *Acta Radiologica* 2019;**60**:186–95
14 Rayyan Intelligent Systematic Review. In: https://www.rayyan.ai/ [14 July 2021]
15 Whiting PF, Rutjes A, Westwood M, Mallett S, Deeks J, Reitsma J. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Internal Med* 2011;**155**:529
16 Cochrane Methods. Analysing and presenting results. In: http://srdta.cochrane.org/ [25 July 2021]
17 Cochrane Methods. Meta-analysis of test accuracy studies in R: a summary of user-written programs and step-by-step guide to using glmer. In: http://methods.cochrane.org/sdt/ [27 July 2021]
18 Welcome to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) website. In: http://www.prisma-statement.org/ [19 July 2021]
19 Ghanooni R, Delpierre I, Magremanne M, Vervaet C, Dumarey N, Remmelink M *et al.* 18F-FDG PET/CT and MRI in the follow-up of

head and neck squamous cell carcinoma. *Contrast Media Mol Imag* 2011;**6**:260–6
20 Pellini R, Manciocco V, Turri-Zanoni M, Vidiri A, Sanguineti G, Marucci L *et al.* Planned neck dissection after chemoradiotherapy in advanced oropharyngeal squamous cell cancer: the role of US, MRI and FDG-PET/TC scans to assess residual neck disease. *J Craniomaxillofac Surg* 2014;**42**:1834–9
21 Schouten CS, Graaf PD, Alberts FM, Hoekstra OS, Comans EFI, 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
22 Breik O, Kumar A, Birchall J, Mortimore S, Laugharne D, Jones K. Follow up imaging of oral, oropharyngeal and hypopharyngeal cancer patients: comparison of PET-CT and MRI post treatment. *J Craniomaxillofacial Surg* 2020;**48**:672–9
23 Bankier AA, Levine D, Halpern EF, Kressel HY. Consensus interpretation in imaging research: is there a better way? *Radiol* 2010;**257**:14–7
24 Jadvar H, Colletti PM, Delgado-Bolton R, Esposito G, Krause BJ, Iagaru AH *et al.* Appropriate use criteria for 18F-FDG PET/CT in restaging and treatment response assessment of malignant disease. *J Nuclear Medicine* 2017;**58**:2026–37
25 Gupta T, Master Z, Kannan S, Agarwal JP, Ghsoh-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 Imag* 2011;**38**:2083–95
26 Vandecaveye V, Nuyts S, Delgado Bolton RC. Response assessment and follow-up by imaging in head and neck tumours. In: Beets-Tan R OW, Valentini V, ed. *Imaging and Interventional Radiology for Radiation Oncology*. New York: Springer, Cham, 2020;405–16
27 Abdulamir AS, Hafidh RR, Abdulmuhaimen N, Abubakar F, Abbas KA. The distinctive profile of risk factors of nasopharyngeal carcinoma in comparison with other head and neck cancer types. *BMC Public Health* 2008;**8**:400

https://doi.org/10.1017/S0022215122000317 Published online by Cambridge University Press