National Institute for Health and Care Excellence

Draft for consultation

Cancer of the upper aerodigestive tract: assessment and management in people aged 16 and over

[A] Evidence reviews for treatment of advanced disease

NICE guideline NG36

Evidence reviews

March 2018

Draft for Consultation

These evidence reviews were developed by the NICE Guideline Updates Team

Disclaimer

The recommendations in this guideline represent the view of NICE, arrived at after careful consideration of the evidence available. When exercising their judgement, professionals are expected to take this guideline fully into account, alongside the individual needs, preferences and values of their patients or service users. The recommendations in this guideline are not mandatory and the guideline does not override the responsibility of healthcare professionals to make decisions appropriate to the circumstances of the individual patient, in consultation with the patient and/or their carer or guardian.

Local commissioners and/or providers have a responsibility to enable the guideline to be applied when individual health professionals and their patients or service users wish to use it. They should do so in the context of local and national priorities for funding and developing services, and in light of their duties to have due regard to the need to eliminate unlawful discrimination, to advance equality of opportunity and to reduce health inequalities. Nothing in this guideline should be interpreted in a way that would be inconsistent with compliance with those duties.

NICE guidelines cover health and care in England. Decisions on how they apply in other UK countries are made by ministers in the Welsh Government, Scottish Government, and Northern Ireland Executive. All NICE guidance is subject to regular review and may be updated or withdrawn.

Copyright

ISBN:

Contents

Management of nodal metastasis in head and neck cancer after chemoradiotherapy

Review question

What is the comparative effectiveness of PET-CT-guided decision making versus planned neck dissection in the management of nodal metastasis in head and neck cancer after chemoradiotherapy?

Introduction

The NICE guideline on upper aerodigestive cancer does not currently consider the use of PET-CT scans as part of the assessment as to whether neck dissection surgery is needed following chemoradiotherapy. PET-CT has the potential to prevent people having unnecessary surgery, and to save money for the NHS. Therefore, this update will review the evidence on effectiveness of PET-CT-guided decision making versus planned neck dissection in the management of nodal metastasis in head and neck cancer after chemoradiotherapy.

PICO table

Population	People aged 16 and over with squamous-cell carcinoma of the oropharynx, hypopharynx, larynx, oral cavity, or an unknown primary site in the head or neck and with nodal disease that has been treated with chemoradiotherapy
Intervention	PET-CT-guided decision making
Comparator	Planned neck dissection
Outcomes	• Recurrence rates • Overall survival • Quality of life (see core symptoms and domains in Appendix B) • Surgical complications • Adverse events

Methods and process

This evidence review was developed using the methods and process described in Developing NICE guidelines: the manual. Methods specific to this review question are described in the review protocol in Appendix A and the methods section in Appendix B.

Declarations of interest were recorded according to NICE's 2014 conflicts of interest policy.

Clinical evidence

Included studies

A systematic search was carried out to identify randomised controlled trials (RCTs) and systematic reviews of RCTs, which found 1,545 references (see Appendix C for the literature search strategy). Evidence identified in the surveillance review was also reviewed (2 references). In total, 1,547 references were identified for screening at title and abstract level. 1,545 were excluded based on their titles and abstracts and 2 references were ordered for screening based on their full texts. Both references were included based on their relevance to the review protocol (Appendix A). The clinical evidence study selection is presented as a PRISMA diagram in Appendix D.

See Appendix L for a list of references for included studies.

Excluded studies

No studies were excluded at full text screen.

Summary of clinical studies included in the evidence review

Two references reporting on the same RCT (PET-NECK) were included. See Appendix E for full evidence tables and Appendix F for forest plots of the results. Forest plots were only presented for outcomes with subgroup analyses. Additional data on particular subgroups of interest was requested from the researchers of the PET-NECK trial, and the data provided can be seen in Appendix N.

The PET-NECK trial included 564 people with head and neck squamous cell carcinoma with advanced neck metastasis treated by radical chemoradiotherapy. Participants were randomised to FDG PET-CT guided surveillance (n=282 participants) or to neck dissection (n=282 participants). Appendix E has more details on the characteristics of the study and the participants and the quality assessment of the study.

Age and sex were not listed in the protocol as relevant subgroups. However, we reported overall mortality by age and sex because both are protected characteristics under the Equality Act 2010, and were therefore agreed to be important to consider.

Quality assessment of clinical studies included in the evidence review

See Appendix H for full GRADE tables.

Economic evidence

Included studies

A single search was conducted for economic evidence relating to both review questions, by applying standard health economic filters to the shared population/intervention terms from the search strategy. Details are provided in Appendix C. In total, 855 records were returned, of which 840 studies could be confidently excluded on sifting of titles and/or abstracts. The remaining 15 studies were reviewed in full text: 2 were considered as duplicates and 11 were found not to be relevant. Therefore, 2 unique cost–utility analyses (CUAs) were included.

One older CUA came from outside the UK and explored the cost-effectiveness of PET-CT before neck dissection compared with neck dissection for all patients with head and neck squamous cell carcinoma (HNSCC).

Sher et al. (2010) developed a Markov model from a US payer's perspective. The model predicted costs and health benefits for 5 years (the majority of clinical trials present 5-year survival data), using 5 health states: distant metastasis, local recurrence, salvage (dissection/local surgery), nodal recurrence and death (disease caused or other causes). The authors collected data from multiple sources. Utility values for head and neck cancer were taken from Hollenbeak et al. (2001). Metastasis and local recurrence states' utility were obtained from 2 other studies exploring metastatic oesophageal cancer. Costs were sourced from Medicare payment schedules. The authors found that PET-CT followed by neck dissection for residual disease was cost-saving and associated with more QALYs than either dissecting all patients or using CT to define residual disease. The authors did not report the expected cost and QALYs in each arm. The study was judged to be partially applicable with potentially serious limitations.

The second study representing directly applicable evidence was PET-NECK, a multicentre randomised trial by Mehanna et al. (2017). It is a UK study that included within-trial economic evaluation and also model-based analysis.

The economic evaluation alongside the clinical trial had a 2-year time frame, adopting an NHS and PSS perspective. The study aimed to compare the cost effectiveness of an FDG PET-CT guided "watch and wait" policy with standard practice of planned neck dissection. Health-related utility was measured by EQ-5D questionnaires completed by participants over specific time points during the 2-year follow-up and valued by the UK population using time trade-off (TTO) methods using the standard UK tariff. Survival duration was then weighted by the utility values to derive QALYs. Multiple imputation methods were followed to address missing utility values or costs. Uncertainty was addressed by deterministic sensitivity analyses and performing non-parametric bootstrapping on the mean QALYs and costs at each arm. The authors found that FDG PET-CT was cost saving (-£1,513) and produced slightly more QALYs (0.07) compared with planned neck dissection. If QALYs are valued at £20,000 each, the probability that FDG PET-CT is the optimal approach was 99%.

The model-based economic evaluation used a state-transition model to extrapolate the PET-NECK trial data and estimate the lifetime costs and health benefits of patients with HNSCC. The model structure was split into 2 phases: an initial 6-month treatment phase sourced from the trial, in which patients received chemo and radio therapy (CRT) followed either by neck dissection directly or FDG PET-CT then (potential) neck dissection (ND); and a follow-up phase where patients may go on to recover (disease free, DF), develop local recurrence (LR) or experience distant recurrence (DR). Simulated patients were at risk of death during the follow-up stage. During the first 6 months, the utility value assigned to the DF state was taken from the trial as there were sufficient data to estimate this. Utility values assigned to patients in the LR or DR states were obtained from a Canadian study, using standard gamble and visual analogue scales for preference elicitation. The cost of the DF state was derived from the trial data (the average monthly cost in each arm between 6 and 24 months). The initial cost applied in the first cycle once the patient moved to LR or DR was also taken from the trial data. However, the ongoing supportive care cost was obtained from the literature. Probabilities of progression from DF to LR or DR were derived from trial data using the recurrence-free survival Kaplan-Meier. The proportion of LR vs DR, assumed to be constant over time, was used to derive the transition probability to each type of recurrence, which was assumed to be possible only within the first 5 years. The trial data could not inform probabilities of a subsequent recurrence; these were derived from existing literature. Probabilistic sensitivity analysis was performed to address the imprecision in the model parameters; one-way sensitivity analysis was performed to assess the impact of key parameters. Further scenario analyses were also performed to check the impact of using different utility values, incorporating additional patient-reported costs and simulating additional recurrences beyond 5 years.

The lifetime analysis base-case findings showed that, compared with planned neck dissection, FDG PET-CT saves money (−£1,485 per patient) and increases quality-adjusted life expectancy (generating an extra 0.16 QALYs). In probabilistic analysis, the probability of FDG PET-CT being cost-saving was 96%, whereas the probability that it is more effective than planned neck dissection was 66%. The results appeared to be robust in the types of sensitivity analyses performed. In one-way sensitivity analysis the parameter with the largest impact on results was the primary recurrence rate when altered by 25%. Increasing the rate of primary recurrences in the PET–CT surveillance arm resulted in the PET–CT watch-and-wait strategy no longer being cost-effective. The study was judged to be directly applicable with minor limitations.

For more details of these studies, please see the economic evidence profiles in Appendix I.

Excluded studies

Details are provided in Appendix K.

Evidence statements

RCT evidence (based on TNM7 staging criteria)

- Low-quality evidence from 1 RCT containing 564 people diagnosed with head and neck squamous cell carcinoma of the oropharynx, larynx or hypopharynx with advanced nodal metastases could not detect a difference in recurrence rates at 2 years between people randomised to FDG PET–CT-guided active surveillance compared to people randomised to planned neck dissection.
- Low- to moderate-quality evidence from 1 RCT containing 564 people diagnosed with head and neck squamous cell carcinoma of the oropharynx, larynx or hypopharynx with advanced nodal metastases could not detect a difference in overall mortality over 36 months between people randomised to FDG PET–CT-guided active surveillance compared to people randomised to planned neck dissection. The same result was found in subgroups of participants with the following characteristics: males; age; tumour stage T1, T2, T3, or T4; nodal stage N2a, N2b, N2c, or N3; cancer site including oral cavity, oropharynx, larynx, or hypopharynx; HPV status.
- High-quality evidence from 1 subgroup analysis of 104 women from an RCT containing 564 people diagnosed with head and neck squamous cell carcinoma of the oropharynx, larynx or hypopharynx with advanced nodal metastases found that fewer women randomised to FDG PET–CT-guided active surveillance died over 36 months compared to women randomised to planned neck dissection.
- High-quality evidence from 1 RCT containing 564 people diagnosed with head and neck squamous cell carcinoma of the oropharynx, larynx or hypopharynx with advanced nodal metastases found that fewer people randomised to FDG PET–CT-guided active surveillance had surgical complications and serious adverse events over 36 months compared to people randomised to planned neck dissection.
- Data on quality of life at 2 years for PET–CT-guided active surveillance versus planned neck dissection were reported for a number of instruments, but it was not possible to construct confidence intervals for these data, and therefore assess the clinical importance of these results.

Economic evidence

A directly applicable cost–utility analysis with minor limitations found that, compared with neck dissection for all patients, introducing FDG PET-CT-guided management results in an incremental cost saving of £1,485 and an expected gain of 0.13 QALYs. If QALYs are valued at £20,000 each, the probability that FDG PET-CT-guided management represents good value for money is 75%.

A further cost–utility analysis, representing partially applicable evidence with potentially serious limitations also concluded that PET-CT-guided management was cost saving and produced more QALYs compared with planned neck dissection for people with HNSCC.

Accuracy of PET-CT to diagnose residual nodal disease

Review question

What is the accuracy of PET-CT to diagnose residual nodal disease in people after chemoradiotherapy?

Introduction

The guideline on cancer of the upper aerodigestive tract does not currently consider PET-CT as a means to determine if residual tumour cells remain after completion of chemoradiotherapy. Alternative strategies to confirm or rule out residual nodal disease are invasive, such as biopsy or dissection. By comparing different sites in the head and neck, different diagnostic accuracies could become evident. As diagnostic accuracy of head and neck cancer has been considered as one site, when evaluating individual patients and sites, the overall diagnostic accuracy will not be directly applicable to that site. This update will review the diagnostic accuracy of PET-CT for residual nodal disease in all sites and individual sites in comparison to other sites in head and neck cancer after chemoradiotherapy.

PICO table

Population	People aged 16 and over with squamous-cell carcinoma of the oropharynx, hypopharynx, larynx, oral cavity, or an unknown primary site in the head or neck who have been treated with chemoradiotherapy
Index test	PET-CT
Reference standard	• Histology including neck dissection, biopsy/surgical resection of tissue • Pathological confirmation of recurrence • Ultrasound scan (USS) / magnetic resonance imaging (MRI) scan / computerised tomography (CT) scan
Outcomes	• Sensitivity • Specificity • Positive likelihood ratio • Negative likelihood ratio

Methods and process

This evidence review was developed using the methods and process described in Developing NICE guidelines: the manual. Methods specific to this review question are described in the review protocol in Appendix A and the methods section in Appendix B.

There was not sufficient evidence to conduct separate analyses for each of the subgroups specified and therefore meta-regression was carried out to assess the effect of different subgroups on the diagnostic accuracy (Appendix G).

Declarations of interest were recorded according to NICE's 2014 conflicts of interest policy.

Clinical evidence

Included studies

A systematic search was carried out to identify diagnostic accuracy studies and systematic reviews of diagnostic accuracy studies and cross-sectional studies, which found 1,236 references (see Appendix C for the literature search strategy). 2 duplicate references were removed. In total, 1,233 references were identified for screening at title and abstract level. Based on their titles and abstracts 1,187 were excluded and 46 references were ordered for full text screening. Thirty three references were excluded as they were not relevant to the review protocol, 2 were excluded during data extraction and 1 was not available to exclude. Ten references were included based on their relevance to the review protocol (Appendix A). The clinical evidence study selection is presented as a PRISMA diagram in Appendix D.

Patients with primary sites at base of tongue and tonsil were considered to have had oropharyngeal cancer and data from these sites were incorporated into the oropharyngeal analyses. Retrospective studies were included in the final analysis as the committee agreed that a correctly designed retrospective study can be representative of clinical practice. Retrospective studies were included only if the studies avoided a case control study design to reflect clinical practice and were blinded. To assess if the inclusion of retrospective studies affects the results, prospective studies for all sites will also be analysed separately. All studies included FDG in conjunction with PET-CT as the marker for tissue metabolism.

The original protocol for this review question (based on the PET-NECK study) did not include people with nasopharyngeal cancer. PET-NECK trial excluded people with nasopharyngeal cancer because this type of cancer is highly sensitive to radiotherapy and should not be treated by neck dissection. For completeness, this evidence was included within this section on diagnostic accuracy.

See Appendix L for a list of references for included studies.

Excluded studies

See Appendix K for a list of excluded studies and the reason for excluding.

Summary of clinical studies included in the evidence review

The 10 studies reported proportions on the following:

- Primary site (10)
- Oropharynx (10)
- Hypopharynx (8)
- Larynx (6)
- Nasopharynx (4)
- Oral cavity (3)
- Unknown (3)
- Nodal stage by site (1)
- HPV status (2)
- Cancer staging (4)

Of the 10 included studies, 2 reported only on the following specific sites

- Oropharynx (2)

Quality assessment of clinical studies included in the evidence review

See Appendix H for full GRADE tables.

Economic evidence

See the section on economics evidence under management of nodal metastasis in head and neck cancer after chemoradiotherapy.

Evidence statements

Diagnostic test accuracy

Reference standards varied between studies including histopathology, clinical assessment, endoscopy with or without biopsy, and confirmation of residual disease with two imaging modalities.

Results that indicate a patient has increased probability of residual nodal disease after completion of chemoradiotherapy (based on positive likelihood ratios)

Very large increase in the probability of residual nodal disease, compared to a negative test result:

- Positive FDG PET-CT result for residual nodal disease in all primary sites in the head and neck, prospective studies only (low quality, 95% CI ranges from large increase to very large increase)

Large increase in probability of residual nodal disease, compared to a negative test result:

- Positive FDG PET-CT result for residual nodal disease in all primary sites in the head and neck, prospective and retrospective studies (very low quality, 95% CI ranges from large increase to very large increase)
- Positive FDG PET-CT result for residual nodal disease when oropharynx is the primary site, retrospective studies (low quality, 95% CI ranges from moderate increase to very large increase)

Results that indicate a patient has decreased probability of residual nodal disease after completion of chemoradiotherapy (based on negative likelihood ratios)

Moderate decrease in probability of residual nodal disease, compared to a positive test result:

- Negative FDG PET-CT result for residual nodal disease in all primary sites in the head and neck, prospective and retrospective studies (very low quality, 95% CI from ranges from moderate decrease to moderate decrease)
- Negative FDG PET-CT result for residual nodal disease in all primary sites in the head and neck, prospective studies only (very low quality, 95% CI ranges from slight decrease to moderate decrease)

Slight decrease in probability of residual nodal disease, compared to a positive test result:

- Negative FDG PET-CT result for residual nodal disease when oropharynx is the primary site, retrospective studies (very low quality, 95% CI ranges from slight decrease to moderate decrease)

Meta-regression

Meta-regression was based on sensitivity and specificity data from 10 studies containing 764 participants. Quality of studies was from very-low to low. Reference standards also varied between studies including histopathology, clinical assessment, endoscopy with or without

biopsy, and confirmation of residual disease with two imaging modalities. The results of the meta-regression analysis found that FDG PET-CT had lower sensitivity in oropharyngeal cancer than other sites and no evidence of differences for specificity. However, there was considerable heterogeneity between studies, and there was a correlation between site of cancer and the risk of bias in the included studies, and therefore there was considerable doubt as to the robustness of these results.

See Table 9 in Appendix G which shows the results of the meta-regression.

Recommendations

Response assessment after chemoradiotherapy[a]

A1. Offer FDG PET-CT to guide management for people treated with radical chemoradiotherapy who have:

- an oropharyngeal primary cancer site and
- 2 or more positive nodes in the neck, all of which are less than 6 cm across. [2018]

A2. Consider FDG PET-CT to guide management for people treated with radical chemoradiotherapy who have:

- an oropharyngeal primary site with 1 positive node in the neck that is less than 6 cm across **or**
- an oropharyngeal primary site with 1 or more positive nodes larger than 6 cm across in the neck **or**
- a hypopharyngeal or laryngeal primary site with 1 or more positive nodes in the neck.

A3. For people having an FDG PET-CT scan after chemoradiotherapy, perform the scan 3 to 6 months after chemoradiotherapy has finished.

A4. Do not offer neck dissection to people with no abnormal FDG uptake or residual soft tissue mass on an FDG PET-CT scan.

Research recommendations

A5. What are the long-term outcomes for people with an indeterminate FDG PET-CT scan result (no abnormal FDG uptake or residual mass) after radical chemoradiotherapy?

A6. What are the most appropriate investigations for people with an indeterminate FDG PET-CT scan result (no abnormal FDG uptake or residual mass) after radical chemoradiotherapy?

A7. What is the effectiveness and cost-effectiveness of using FDG PET-CT to guide follow-up after treatment for people with head and neck cancer?

A8. What is the optimal management strategy of nodal metastasis in nasopharynx cancer after chemoradiotherapy?

Rationale and impact

Why the committee made the recommendations

Overall, the evidence showed that recurrence rates and overall mortality for FDG PET-CT-guided management after radical chemoradiotherapy were similar to those for neck dissection. In addition, the evidence showed that FDG PET-CT was cost-saving compared

[a] The term 'radical chemoradiotherapy' refers to treatment aiming to cure cancer rather than to relieve symptoms (palliative treatment). It is used here to reflect the evidence these recommendations are based on.

with neck dissection, and would prevent unnecessary surgeries, surgical complications, and adverse events.

The committee agreed to make recommendations only for people with oropharyngeal, laryngeal and hypopharyngeal primary sites, because these were the main focus of the evidence. Most of the people in the studies had an oropharyngeal primary site and 1 or more positive nodes under 6 cm across in the neck, and the evidence was strongest for this population. Therefore, the committee agreed that they should be offered an FDG PET-CT scan.

The evidence was weaker for people with an oropharyngeal primary site and more severe disease (1 or more positive node larger than 6 cm across in the neck) and for people with laryngeal or hypopharyngeal primary sites. To reflect this, FDG PET-CT scanning could be considered for these groups.

The evidence did not include people with an oropharyngeal primary site and less severe disease (only 1 positive node of less than 6 cm across). However, the committee agreed that it is particularly important that FDG PET-CT scans are considered for this population to avoid unnecessary surgery. These people are likely to be at a lower risk of recurrence and so the benefits of neck dissection are lower.

The committee noted that new classifications for head and neck cancer (TNM classification of malignant tumours, 8th edition) have been introduced, which are different to those used in the evidence. They decided to describe the stage of cancer for these recommendations in terms of the number and size of positive nodes to avoid confusion.

The timing of FDG PET-CT scans (3 to 6 months after completion of radical chemoradiotherapy) is in line with current Royal College of Radiologists guidelines. Earlier scans are more likely to give a false-positive result, due to the residual effects of treatment.

The committee decided to be specific that neck dissection should not be offered to people with no abnormal FDG uptake or residual soft tissue mass, to give clear advice about how to interpret a ‘negative’ FDG PET-CT result.

The committee noted several areas in which future research would be helpful, such as management for people with indeterminate test results (see research recommendations A5 and A6), the role of FDG PET-CT for people with nasopharyngeal cancer (see research recommendation A8) and the effectiveness of FDG PET-CT to guide follow-up (see research recommendation A7).

Impact of the recommendations on practice

There may an increase in the number of FDG PET-CT scans performed and a reduction in surgical procedures. However, the evidence showed that the amount of money saved from unnecessary surgery is likely to be considerably higher than the cost of the additional scans.

The committee’s discussion of the evidence

Interpreting the evidence

The outcomes that matter most

The committee agreed that recurrence rates and overall survival were the critical outcomes for FDG PET-CT-guided management after chemoradiotherapy. Complications following neck dissection surgery and serious adverse events were also important outcomes because these could have an impact on function. The committee noted that if survival and recurrence rates were not worse for FDG PET-CT guided management, it was highly unlikely quality of

life would be worse for that group, as a reduction in the number of unnecessary surgeries would be expected to have a positive impact on quality of life.

The quality of the evidence

The committee noted that the RCT evidence available came from a high-quality UK study, and the only reason for downgrading the evidence was imprecision. In particular, the thresholds for clinical significance specified by the committee were narrower than the non-inferiority margin specified in the PET-NECK study. This led to a lower rating of whether we could be certain the trial demonstrated equivalence of the interventions. However, the committee agreed that on the evidence available, FDG PET-CT guided management was likely to be non-inferior to planned neck dissection for both survival and recurrence rates.

The committee found it difficult to assess the clinical importance of the quality of life outcomes reported in the trial because it was not possible to construct confidence intervals for the reported differences (some of the subscales of the quality of life instruments included are not continuous or equally spaced). The committee agreed however that there was no systematic pattern of difference in quality of life that would raise concern about the effectiveness of FDG PET-CT, and noted the trial did collect EQ-5D data on quality of life which was used as part of the economic analysis.

Most of the participants included in the PET-NECK trial had oropharyngeal cancer, which is the second most common head and neck cancer in the UK but the most common cancer treated with primary radical chemoradiotherapy. The committee agreed that HPV status is a relevant consideration for oropharyngeal cancer as it may be correlated with outcomes, but this subgroup (oropharyngeal cancer, stratified by HPV status) was not reported for any of the outcomes in the trial. The committee also noted there were only sufficient numbers of participants in the PET-NECK study with oropharyngeal, laryngeal and hypopharyngeal cancer. Therefore, the committee felt that it was not appropriate to make the same inferences to other groups of head and neck cancer.

The included studies on the diagnostic accuracy of FDG PET-CT for residual nodal disease had heterogeneous disease prevalence, populations and study design, and potential bias that resulted in both inconsistent results and downgrading of the quality of evidence. The committee agreed the difference in prevalence of different head and neck squamous cell carcinoma (HNSCC) subtypes between studies conducted in the UK and studies of other countries in the review (e.g. Taiwan) may mean the overall population is not representative of the UK. The longer interval for performing response assessment with FDG PET-CT found in some studies may have resulted in overestimating diagnostic accuracy, compared to performing a scan at an appropriate earlier stage to inform management. Although people with nasopharyngeal cancer were not included in the PET-NECK trial due to its pathology and treatment making its management distinct from other HNSCC, the committee agreed that diagnostic accuracy of residual nodal disease is not affected by these factors and were included in the relevant question.

The committee agreed retrospective studies are likely to be at risk of bias due to difficulties in blinding image reviewers to other test results. However, they noted there were only minor differences between the results for prospective and retrospective studies, so this did not appear to be a major issue. There was a lack of reporting on the interval between FDG PET-CT and reference testing in all but one study. The committee noted that qualitative image interpretation was used in the majority of the studies, which is routine in clinical practice in the UK, but that variability in practice may arise due to a lack of standardisation in FDG PET-CT reviewing. FDG PET-CT results for residual disease are often classified as positive, negative and indeterminate, but the definition of indeterminate may be different between centres, and there may be different policies as to whether these people are treated as positives or negatives for the purpose of management. In particular, people with residual soft tissue mass but no abnormal FDG uptake may be classified differently by different centres.

The committee noted that all the included studies were post-2007, and agreed that methods used after this time are sufficiently modern to be comparable. Only limited subgroup analysis was possible as the patient characteristics was not reported consistently across the studies. The analyses did find that moderate risk of bias studies and/or oropharynx as the primary site may lower sensitivity but could not discern which had the greatest or any effect. The committee agreed the diagnostic accuracy evidence was not sufficiently robust as to enable extrapolation of the results of the PET-NECK study outside of the trial population.

The committee noted that the diagnostic accuracy analysis produced results that were, on face value, surprising. Specifically, they found that positive PET-CT results (measured by positive likelihood ratios) provided more diagnostic accuracy that negative results (measured by negative likelihood ratios), which differed from their clinical experience that PET-CT has a high negative predictive value but a low positive predictive value. However, it was noted this apparent contradiction is actually a result of the low prevalence of residual nodal disease after chemoradiotherapy (around 20% in the included diagnostic accuracy studies, and likely to be lower in practice due to selection bias in some of the studies). Setting a prevalence of approximately 12% in the cohort gives a negative predictive value around 95%, in line with their clinical experience. Therefore, it was concluded that although a positive PET-CT result does, taken in isolation, give more information than a negative PET-CT result, a testing strategy combining a negative PET-CT with an already low pre-test probability of residual nodal disease gives a higher negative predictive value than a positive PET-CT result combined with a low pre-test probability gives as a positive predictive value.

Benefits and harms

Recurrence rates and overall survival after chemoradiotherapy were not significantly different between FDG PET-CT-guided management and planned neck dissection, with the point estimate for survival favouring FDG PET-CT. There was also evidence of reduced surgical complications and serious adverse events when the decision to offer neck dissection was based on FDG PET-CT results. Therefore, the committee agreed to make a strong positive recommendation for FDG PET-CT for people with >1 positive node all <6cm in the neck with an oropharyngeal primary site. This represented the majority of people (around 60%) in the PET-NECK trial, and was therefore the population in which the confidence in the evidence was most strong. The committee agreed it was appropriate to phrase the recommendation in these terms, rather than in terms of staging, because of the ongoing transition between the TNM classification 7th and 8th Editions for head and neck cancer. The PET-NECK study was conducted using the 7th Edition rather than more recent versions, and therefore using old classifications is likely to prove confusing over time.

The committee made ‘consider’ recommendations for two other populations within the PET-NECK study. These were people with more severe disease (at least 1 positive node >6cm in the neck with an oropharyngeal primary site) and people with laryngeal or hypopharyngeal primary sites. These people were included within PET-NECK, but the number of participants from these groups was lower, and therefore the committee agreed the confidence in the evidence for this group was also lower.

The committee also agreed it was appropriate to make a consensus ‘consider’ recommendation for people with less severe disease (only 1 positive node). Although these participants were not included in the PET-NECK study, they are likely to be at lower risk of recurrence than those included, and therefore it is even more desirable to avoid unnecessary surgery in this population, as the benefits from neck dissection are lower.

The committee agreed that FDG PET-CT-guided management should be performed 3 to 6 months after completion of radical chemoradiotherapy (in line with current Royal College of Radiologists 2016 PET-CT guidelines) because earlier scans are more likely to give a false positive result, due to the residual effects of treatment. The committee also agreed to make a ‘do not offer’ recommendation for neck dissection for people with no abnormal FDG uptake or residual soft tissue mass. It was agreed important to be specific about this population

(again, defined by the criteria of the PET-NECK study) due to the differences in the interpretation of what counts as a 'negative FDG PET-CT result.

Cost effectiveness and resource use

The committee reviewed the included economic evidence. It agreed that cost-effectiveness analysis from the PET-NECK trial provided directly applicable evidence. The committee agreed that PET-NECK trial results can be generalised to the UK NHS settings. The committee noted that people with head and neck squamous cell carcinoma were routinely followed-up in secondary care settings. Thus, the cost-effectiveness analysis provided by PET-NECK trial, not including primary care resource uses in its base-case, would not miss significant costs. The committee was confident in drawing the conclusion that the use of FDG PET-CT-guided management by people with head and neck squamous cell carcinoma with positive lymph nodes represented a good value of the UK NHS resources, as it was cost saving and slightly more effective.

The economic evidence provided by PET-NECK trial was agreed to be sufficient to underpin strong recommendations in favour of offering FDG PET-CT-guided management to people with more than 1 positive node, all of which are less than 6cm, in the neck with an oropharyngeal primary site. The committee also noted that if the clinical effectiveness of PET-CT was similar in other population of people (e.g. those with laryngeal and hypopharyngeal cancer), then the same findings for costs and therefore cost-effectiveness would be expected.

Other factors the committee took into account

The committee agreed that it is unusual for people >70 years to be offered chemotherapy for head and neck cancer, but age was not an exclusion criterion in the PET-NECK trial. Age subgroup analyses were reported by the PET-NECK trial, but no significant differences were found between FDG PET-CT and planned neck dissection for any subgroup. The committee also noted that whilst choices of treatment may differ by age, this does not mean the follow-up strategies would be different for those people who have had chemoradiotherapy. The PET-NECK trial also reported subgroup analyses by sex showing significantly fewer deaths in women allocated to FDG PET-CT compared with planned neck dissection. The total number of deaths in women was 18, which was considered to be insufficient to draw strong conclusions, but the committee noted this as a relevant finding for future research. The committee agreed that since a positive recommendation for FDG PET-CT was made, this would be equally applicable for both men and women.

The committee agreed that FDG PET-CT is widely available at the UK including mobile vans, though there may be issues with capacity at certain centres. The committee noted that although this review question and the evidence from the PET-NECK trial only covered people undergoing chemoradiotherapy, it was likely people would, in practice, extrapolate those results to people having undergone only radiotherapy.

The committee noted there were a number of future research issues raised by the PET-NECK results. The first of these is around how indeterminate test results should be managed, where there is considerable variability in practice, with some centres performing a second scan in these people. The committee agreed there was value in research both on the natural history of people with indeterminate results, and on what future investigation are best able to resolve that uncertainty and guide management. Second, the committee agreed there was uncertainty as to the role of FDG PET-CT in people with nasopharyngeal cancer, who were not contained within the PET-NECK trial. Finally, the committee agreed the PET-NECK trial only covered one aspect of the potential value of FDG PET-CT in guiding treatment (response assessment after chemoradiotherapy), and there were other points in the pathway where an assessment of the value of FDG PET-CT to guide decision making could be evaluated. The committee considered relevant to look at the effect of FDG PET-CT-guided

management on overall survival in the subgroups by sex and HPV status because the number of deaths in women was significantly lower for FDG PET-CT compared with planned neck dissection. Data on overall survival in the subgroups by sex and HPV status has been requested from the researchers of the PET-NECK trial.

Appendices

Appendix A – Review protocols

Review protocol for the management of nodal metastasis in head and neck cancer after chemoradiotherapy

Field (based on PRISMA-P)	Content
Review question	What is the comparative effectiveness of PET-CT-guided decision making versus planned neck dissection in the management of nodal metastasis in head and neck cancer after chemoradiotherapy?
Type of review question	Intervention
Objective of the review	To compare the effectiveness of PET-CT-guided decision making versus planned neck dissection in people with squamous-cell carcinoma of the oropharynx, hypopharynx, larynx, oral cavity, or an unknown primary site in the head or neck and with nodal disease that has been treated with chemoradiotherapy.
Eligibility criteria – population	People aged 16 and over with squamous-cell carcinoma of the oropharynx, hypopharynx, larynx, oral cavity, or an unknown primary site in the head or neck and with nodal disease that has been treated with chemoradiotherapy.

Eligibility criteria – intervention	PET-CT-guided decision making
Eligibility criteria – comparator	Planned neck dissection
Outcomes and prioritisation	• Recurrence rates • Overall survival • Quality of life (see core symptoms and domains in Appendix B) • Surgical complications • Adverse events
Measures	Recurrence rates • Hazard ratios Overall survival • Hazard ratios Quality of life measures that have overall scores or include any of the core symptoms and domains listed in Appendix B • Mean difference Surgical complications • Relative risk Adverse events • Relative risk
Eligibility criteria – study design	• Randomised controlled trials (RCTs)
Other inclusion/exclusion criteria	Other inclusion criteria: • English language • Published studies only

	Other exclusion criteria: • Observational studies • Studies without extractable data • Conference abstracts • Subscales of quality of life measures of symptoms or domains which were not identified as core symptoms or domains
Proposed sensitivity/sub-group analysis, or meta-regression	All of these subgroups will be reported regardless of heterogeneity: • Cancer staging • HPV status • Cancer site • Nodal stage by site
Selection process – duplicate screening/selection/analysis	10% of the abstracts were reviewed by two reviewers, with any disagreements resolved by discussion or, if necessary, a third independent reviewer. If meaningful disagreements were found between the different reviewers, a further 10% of the abstracts were reviewed by two reviewers, with this process continued until agreement is achieved between the two reviewers. From this point, the remaining abstracts will be screened by a single reviewer.
Data management (software)	See Appendix B
Information sources – databases and dates	**Sources to be searched** Clinical searches: • MEDLINE (Ovid) • MEDLINE In-Process (Ovid) • PubMed (NLM)

	• EMBASE (Ovid) • Cochrane Database of Systematic Reviews (Wiley) • Cochrane Central Register of Controlled Trials (Wiley) • Database of Abstracts of Reviews of Effects (Wiley) (legacy records) Economic searches: • MEDLINE (Ovid) • MEDLINE In-Process (Ovid) • EMBASE (Ovid) • NHS Economic Evaluation Database (Wiley) (legacy records) • Health Technology Assessment Database (Wiley) • Econlit (Ovid) Economic evaluations and quality of life filters applied **Supplementary search techniques** • None identified **Limits** • Studies reported in English • Study design - RCT filters • Animal studies will be excluded from the search results • Conference abstracts will be excluded from the search results • No date limit will be set
Identify if an update	N/A
Author contacts	Guideline update

Highlight if amendment to previous protocol	For details please see section 4.5 of Developing NICE guidelines: the manual
Search strategy – for one database	For details please see Appendix C
Data collection process – forms/duplicate	A standardised evidence table format will be used, and published as Appendix D (clinical evidence tables) or I (economic evidence tables).
Data items – define all variables to be collected	For details please see evidence tables in Appendix E (clinical evidence tables) or J (economic evidence tables).
Methods for assessing bias at outcome/study level	See Appendix B
Criteria for quantitative synthesis	See Appendix B
Methods for quantitative analysis – combining studies and exploring (in)consistency	See Appendix B
Meta-bias assessment – publication bias, selective reporting bias	See Appendix B
Confidence in cumulative evidence	See Appendix B
Rationale/context – what is known	For details please see the introduction to the evidence review in the main file.
Describe contributions of authors and guarantor	A multidisciplinary committee developed the evidence review. The committee was convened by the NICE Guideline Updates Team and chaired by Steve Pilling in line with section 3 of Developing NICE guidelines: the manual.

	Staff from the NICE Guideline Updates Team undertook systematic literature searches, appraised the evidence, conducted meta-analysis and cost-effectiveness analysis where appropriate, and drafted the evidence review in collaboration with the committee. For details please see Developing NICE guidelines: the manual.
Sources of funding/support	The NICE Guideline Updates Team is an internal team within NICE.
Name of sponsor	The NICE Guideline Updates Team is an internal team within NICE.
Roles of sponsor	The NICE Guideline Updates Team is an internal team within NICE.
PROSPERO registration number	N/A

1 **Review protocol for the accuracy of PET-CT to diagnose residual nodal disease after radiotherapy or chemoradiotherapy**

Field (based on PRISMA-P)	Content
Review question	What is the accuracy of PET-CT to diagnose residual nodal disease in people after chemoradiotherapy?
Type of review question	Diagnostic accuracy
Objective of the review	To determine diagnostic accuracy of PET-CT for the diagnosis of residual nodal disease in people with squamous-cell carcinoma of the oropharynx, hypopharynx, larynx, oral cavity, or an unknown primary site in the head or neck who have been treated with chemoradiotherapy

Eligibility criteria – population	People aged 16 and over with squamous-cell carcinoma of the oropharynx, hypopharynx, larynx, oral cavity, or an unknown primary site in the head or neck who have been treated with chemoradiotherapy
Eligibility criteria – index test	PET-CT
Eligibility criteria – reference standard	• Histology including neck dissection, biopsy/surgical resection of tissue • Pathological confirmation of recurrence • Ultrasound scan (USS) / magnetic resonance imaging (MRI) scan / computerised tomography (CT) scan
Outcomes and prioritisation	• Sensitivity • Specificity • Positive likelihood ratio • Negative likelihood ratio
Eligibility criteria – study design	• Cross-sectional studies • Systematic reviews of cross-sectional studies
Other inclusion/exclusion criteria	Other inclusion criteria: • English language • Published studies only Other exclusion criteria: • Retrospective studies

	• Studies from which a 2x2 table cannot be calculated • Conference abstracts
Proposed sensitivity/sub-group analysis, or meta-regression	All of these subgroups will be reported regardless of heterogeneity: • Cancer site • Cancer staging • HPV status • Nodal stage by site
Selection process – duplicate screening/selection/analysis	10% of the abstracts were reviewed by two reviewers, with any disagreements resolved by discussion or, if necessary, a third independent reviewer. If meaningful disagreements were found between the different reviewers, a further 10% of the abstracts were reviewed by two reviewers, with this process continued until agreement is achieved between the two reviewers. From this point, the remaining abstracts will be screened by a single reviewer.
Data management (software)	See Appendix B
Information sources – databases and dates	**Sources to be searched** Clinical searches: • MEDLINE (Ovid) • MEDLINE In-Process (Ovid) • PubMed (NLM) • EMBASE (Ovid) • Cochrane Database of Systematic Reviews (Wiley) • Cochrane Central Register of Controlled Trials (Wiley) • Database of Abstracts of Reviews of Effects (Wiley) (legacy records) Economic searches:

	• MEDLINE (Ovid) • MEDLINE In-Process (Ovid) • EMBASE (Ovid) • NHS Economic Evaluation Database (Wiley) (legacy records) • Health Technology Assessment Database (Wiley) • Econlit (Ovid) Economic evaluations and quality of life filters applied **Supplementary search techniques** • None identified **Limits** • Studies reported in English • Study design - diagnostic accuracy filters • Animal studies will be excluded from the search results • Conference abstracts will be excluded from the search results • No date limit will be set
Identify if an update	N/A
Author contacts	Guideline update
Highlight if amendment to previous protocol	For details please see section 4.5 of Developing NICE guidelines: the manual
Search strategy – for one database	For details please see Appendix C
Data collection process – forms/duplicate	A standardised evidence table format will be used, and published as Appendix D (clinical evidence tables) or H (economic evidence tables).

Data items – define all variables to be collected	For details please see evidence tables in Appendix E (clinical evidence tables) or I (economic evidence tables).
Methods for assessing bias at outcome/study level	See Appendix B
Criteria for quantitative synthesis	See Appendix B
Methods for quantitative analysis – combining studies and exploring (in)consistency	See Appendix B
Meta-bias assessment – publication bias, selective reporting bias	See Appendix B
Confidence in cumulative evidence	See Appendix B
Rationale/context – what is known	For details please see the introduction to the evidence review in the main file.
Describe contributions of authors and guarantor	A multidisciplinary committee developed the evidence review. The committee was convened by the NICE Guideline Updates Team and chaired by Steve Pilling in line with section 3 of Developing NICE guidelines: the manual. Staff from the NICE Guideline Updates Team undertook systematic literature searches, appraised the evidence, conducted meta-analysis and cost-effectiveness analysis where appropriate, and drafted the evidence review in collaboration with the committee. For details please see Developing NICE guidelines: the manual.
Sources of funding/support	The NICE Guideline Updates Team is an internal team within NICE.

Name of sponsor	The NICE Guideline Updates Team is an internal team within NICE.
Roles of sponsor	The NICE Guideline Updates Team is an internal team within NICE.
PROSPERO registration number	N/A

1

Appendix B – Methods

Evidence of effectiveness of interventions

Evidence synthesis and meta-analyses

Where possible, meta-analyses were conducted to combine the results of studies for each outcome. For mean differences, where change from baseline data were reported in the trials and were accompanied by a measure of spread (for example standard deviation), these were extracted and used in the meta-analysis. Where measures of spread for change from baseline values were not reported, the corresponding values at study end were used and were combined with change from baseline values to produce summary estimates of effect. These/All studies were assessed to ensure that baseline values were balanced across the treatment/comparison groups; if there were significant differences in important confounding variables at baseline these studies were not included in any meta-analysis and were reported separately.

Quality assessment

Individual RCTs and quasi-randomised controlled trials were quality assessed using the Cochrane Risk of Bias Tool. Each individual study was classified into one of the following three groups:

- Low risk of bias – The true effect size for the study is likely to be close to the estimated effect size.
- Moderate risk of bias – There is a possibility the true effect size for the study is substantially different to the estimated effect size.
- High risk of bias – It is likely the true effect size for the study is substantially different to the estimated effect size.

Each individual study was also classified into one of three groups for directness, based on if there were concerns about the population, intervention, comparator and/or outcomes in the study and how directly these variables could address the specified review question. Studies were rated as follows:

- Direct – No important deviations from the protocol in population, intervention, comparator and/or outcomes.
- Partially indirect – Important deviations from the protocol in one of the population, intervention, comparator and/or outcomes.
- Indirect – Important deviations from the protocol in at least two of the following areas: population, intervention, comparator and/or outcomes.

Methods for combining intervention evidence

Meta-analyses of interventional data were conducted with reference to the Cochrane Handbook for Systematic Reviews of Interventions (Higgins et al. 2011).

Where different studies presented continuous data measuring the same outcome but using different numerical scales (e.g. a 0-10 and a 0-100 visual analogue scale), these outcomes were all converted to the same scale before meta-analysis was conducted on the mean differences. Where outcomes measured the same underlying construct but used different instruments/metrics, data were analysed using standardised mean differences (Hedges' g).

A pooled relative risk was calculated for dichotomous outcomes (using the Mantel–Haenszel method). Both relative and absolute risks were presented, with absolute risks calculated by applying the relative risk to the pooled risk in the comparator arm of the meta-analysis.

Fixed- and random-effects models (der Simonian and Laird) were fitted for all syntheses, with the presented analysis dependent on the degree of heterogeneity in the assembled evidence. Fixed-effects models were the preferred choice to report, but in situations where the assumption of a shared mean for fixed-effects model were clearly not met, even after appropriate pre-specified subgroup analyses were conducted, random-effects results are presented. Fixed-effects models were deemed to be inappropriate if one or both of the following conditions was met:

- Significant between study heterogeneity in methodology, population, intervention or comparator was identified by the reviewer in advance of data analysis. This decision was made and recorded before any data analysis was undertaken.
- The presence of significant statistical heterogeneity in the meta-analysis, defined as $I^2 \geq 50\%$.

In any meta-analyses where some (but not all) of the data came from studies at high risk of bias, a sensitivity analysis was conducted, excluding those studies from the analysis. Results from both the full and restricted meta-analyses are reported. Similarly, in any meta-analyses where some (but not all) of the data came from indirect studies, a sensitivity analysis was conducted, excluding those studies from the analysis.

Meta-analyses were performed in Cochrane Review Manager v5.3.

Minimal clinically important differences (MIDs)

The Core Outcome Measures in Effectiveness Trials (COMET) database was searched to identify published minimal clinically important difference thresholds relevant to this guideline. Identified MIDs were assessed to ensure they had been developed and validated in a methodologically rigorous way, and were applicable to the populations, interventions and outcomes specified in this guideline. In addition, the Guideline Committee were asked to prospectively specify any outcomes where they felt a consensus MID could be defined from their experience. In particular, any questions looking to evaluate non-inferiority (that one treatment is not meaningfully worse than another) required an MID to be defined to act as a non-inferiority margin.

MIDs found through this process and used to assess imprecision in the guideline are given in Table 1. For other continuous outcomes not specified in the table below, the line of no effect was used to assess imprecision.

Table 1 Identified MIDs

Outcome	MID	Source
Overall survival	0.8, 1.25	The committee specified an absolute difference of 5% in survival as being clinically meaningful. The PET-NECK study assumed a baseline 2-year probability of 75% overall survival with planned neck dissection. Therefore, an absolute difference of 5% corresponds to a hazard ratio of [ln(0.7)/ln(0.75) = 1.240; and 1/1.24=0.807 for the lower bound]. For convenience, this was rounded to MIDs for hazard ratios of (0.80, 1.25)
Recurrence rates	0.8, 1.25	The committee specified the same absolute difference of 5% as being meaningful as for overall survival, which corresponds to the same MIDs for hazard ratios of (0.80, 1.25)

For quality of life outcomes (Table 2), the COMET database provided a list of recommended core symptoms and domains of quality of life for head and neck clinical trials. This list also includes cross-cutting symptoms that apply to all cancer patients.

Table 2 Identified core outcomes

Outcome	Source
Head and neck specific symptoms • Swallowing • Pain/oral • Skin changes • Dry mouth • Dental health • Opening mouth/trismus • Taste • Excess/thick mucus/saliva • Shoulder disability/motion • Voice/hoarseness Domains • Social • Functional	Chera BS, Eisbruch A, Murphy BA, et al. (2014) Recommended patient-reported core set of symptoms to measure in head and neck cancer treatment trials. JNCI: Journal of the National Cancer Institute, 106(7).
Cross-cutting symptoms • Weight loss/appetite • Pain/general • Nausea/vomiting • Anxiety • Dyspnoea • Fatigue • Depression/mood	Chera BS, Eisbruch A, Murphy BA, et al. (2014) Recommended patient-reported core set of symptoms to measure in head and neck cancer treatment trials. JNCI: Journal of the National Cancer Institute, 106(7).

For standardised mean differences where no other MID was available, an MID of 0.2 was used, corresponding to the threshold for a small effect size initially suggested by Cohen et al. (1988). For relative risks where no other MID was available, a default MID interval for dichotomous outcomes of 0.8 to 1.25 was used.

When decisions were made in situations where MIDs were not available, the 'Evidence to Recommendations' section of that review should make explicit the committee's view of the expected clinical importance and relevance of the findings. In particular, this includes consideration of whether the whole effect of a treatment (which may be felt across multiple independent outcome domains) would be likely to be clinically meaningful, rather than simply whether each individual sub outcome might be meaningful in isolation.

GRADE for pairwise meta-analyses of interventional evidence

GRADE was used to assess the quality of evidence for the selected outcomes as specified in 'Developing NICE guidelines: the manual (2014)'. Data from RCTs was initially rated as high quality and the quality of the evidence for each outcome was downgraded or not from this initial point. If non-RCT evidence was included for intervention-type systematic reviews then these were initially rated as either moderate quality (quasi-randomised studies) or low quality (cohort studies) and the quality of the evidence for each outcome was further downgraded or not from this point, based on the criteria given in Table 3.

Table 3: Rationale for downgrading quality of evidence for intervention studies

GRADE criteria	Reasons for downgrading quality
Risk of bias	Not serious: If less than 33.3% of the weight in a meta-analysis came from studies at moderate or high risk of bias, the overall outcome was not downgraded. Serious: If greater than 33.3% of the weight in a meta-analysis came from studies at moderate or high risk of bias, the outcome was downgraded one level. Very serious: If greater than 33.3% of the weight in a meta-analysis came from studies at high risk of bias, the outcome was downgraded two levels. Outcomes meeting the criteria for downgrading above were not downgraded if there was evidence the effect size was not meaningfully different between studies at high and low risk of bias.
Indirectness	Not serious: If less than 33.3% of the weight in a meta-analysis came from partially indirect or indirect studies, the overall outcome was not downgraded. Serious: If greater than 33.3% of the weight in a meta-analysis came from partially indirect or indirect studies, the outcome was downgraded one level. Very serious: If greater than 33.3% of the weight in a meta-analysis came from indirect studies, the outcome was downgraded two levels. Outcomes meeting the criteria for downgrading above were not downgraded if there was evidence the effect size was not meaningfully different between direct and indirect studies.
Inconsistency	Concerns about inconsistency of effects across studies, occurring when there is unexplained variability in the treatment effect demonstrated across studies (heterogeneity), after appropriate pre-specified subgroup analyses have been conducted. This was assessed using the I^2 statistic. N/A: Inconsistency was marked as not applicable if data on the outcome was only available from one study. Not serious: If the I^2 was less than 33.3%, the outcome was not downgraded. Serious: If the I^2 was between 33.3% and 66.7%, the outcome was downgraded one level. Very serious: If the I^2 was greater than 66.7%, the outcome was downgraded two levels. Outcomes meeting the criteria for downgrading above were not downgraded if there was evidence the effect size was not meaningfully different between studies with the smallest and largest effect sizes.
Imprecision	If an MID other than the line of no effect was defined for the outcome, the outcome was downgraded once if the 95% confidence interval for the effect size crossed one line of the MID, and twice if it crosses both lines of the MID. If the line of no effect was defined as an MID for the outcome, it was downgraded once if the 95% confidence interval for the effect size crossed the line of no effect (i.e. the outcome was not statistically significant), and twice if the sample size of the study was sufficiently small that it is not plausible any realistic effect size could have been detected. Outcomes meeting the criteria for downgrading above were not downgraded if the confidence interval was sufficiently narrow that the upper and lower bounds would correspond to clinically equivalent scenarios.

The quality of evidence for each outcome was upgraded if any of the following three conditions were met:

- Data from non-randomised studies showing an effect size sufficiently large that it cannot be explained by confounding alone.
- Data showing a dose-response gradient.

- Data where all plausible residual confounding is likely to increase our confidence in the effect estimate.

Publication bias

Publication bias was assessed in two ways. First, if evidence of conducted but unpublished studies was identified during the review (e.g. conference abstracts, trial protocols or trial records without accompanying published data), available information on these unpublished studies was reported as part of the review. Secondly, where 10 or more studies were included as part of a single meta-analysis, a funnel plot was produced to graphically assess the potential for publication bias.

Evidence statements

Evidence statements for pairwise intervention data are classified in to one of four categories:

- Situations where the data are only consistent, at a 95% confidence level, with an effect in one direction (i.e. one that is 'statistically significant'), and the magnitude of that effect is most likely to meet or exceed the MID (i.e. the point estimate is not in the zone of equivalence). In such cases, we state that the evidence showed that there is an effect.
- Situations where the data are only consistent, at a 95% confidence level, with an effect in one direction (i.e. one that is 'statistically significant'), but the magnitude of that effect is most likely to be less than the MID (i.e. the point estimate is in the zone of equivalence). In such cases, we state that the evidence could not demonstrate a meaningful difference.
- Situations where the data are consistent, at a 95% confidence level, with an effect in either direction (i.e. one that is not 'statistically significant') but the confidence limits are smaller than the MIDs in both directions. In such cases, we state that the evidence demonstrates that there is no difference.
- In all other cases, we state that the evidence could not differentiate between the comparators.

Diagnostic test accuracy evidence

In this guideline, diagnostic test accuracy (DTA) data are classified as any data in which a feature – be it a symptom, a risk factor, a test result or the output of some algorithm that combines many such features – is observed in some people who have the condition of interest at the time of the test and some people who do not. Such data either explicitly provide, or can be manipulated to generate, a 2x2 classification of true positives and false negatives (in people who, according to the reference standard, truly have the condition) and false positives and true negatives (in people who, according to the reference standard, do not).

The 'raw' 2x2 data can be summarised in a variety of ways. Those that were used for decision making in this guideline are as follows:

- **Positive likelihood ratios** describe how many times more likely positive features are in people with the condition compared to people without the condition. Values greater than 1 indicate that a positive result makes the condition more likely.
  - LR^+ = (TP/[TP+FN])/(FP/[FP+TN])
- **Negative likelihood ratios** describe how many times less likely negative features are in people with the condition compared to people without the condition. Values less than 1 indicate that a negative result makes the condition less likely.
  - LR^- = (FN/[TP+FN])/(TN/[FP+TN])
- **Sensitivity** is the probability that the feature will be positive in a person with the condition.

- sensitivity = TP/(TP+FN)
- **Specificity** is the probability that the feature will be negative in a person without the condition.
  - specificity = TN/(FP+TN)

The following schema, adapted from the suggestions of Jaeschke et al. (1994), was used to interpret the likelihood ratio findings from diagnostic test accuracy reviews.

Table 4 Interpretation of likelihood ratios

Value of likelihood ratio	Interpretation
LR ≤ 0.1	**Very large** decrease in probability of disease
0.1 < LR ≤ 0.2	**Large** decrease in probability of disease
0.2 < LR ≤ 0.5	**Moderate** decrease in probability of disease
0.5 < LR ≤ 1.0	**Slight** decrease in probability of disease
1.0 < LR < 2.0	**Slight** increase in probability of disease
2.0 ≤ LR < 5.0	**Moderate** increase in probability of disease
5.0 ≤ LR < 10.0	**Large** increase in probability of disease
LR ≥ 10.0	**Very large** increase in probability of disease

The schema above has the effect of setting a minimal important difference for positive likelihoods ratio at 2, and a corresponding minimal important difference for negative likelihood ratios at 0.5. Likelihood ratios (whether positive or negative) falling between these thresholds were judged to indicate no meaningful change in the probability of disease.

Quality assessment

Individual studies were quality assessed using the QUADAS-2 tool, which contains four domains: patient selection, index test, reference standard, and flow and timing. Each individual study was classified into one of the following two groups:

- Low risk of bias – Evidence of non-serious bias in zero or one domain.
- Moderate risk of bias – Evidence of non-serious bias in two domains only, or serious bias in one domain only.
- High risk of bias – Evidence of bias in at least three domains, or of serious bias in at least two domains.

Each individual study was also classified into one of three groups for directness, based on if there were concerns about the population, index features and/or reference standard in the study and how directly these variables could address the specified review question. Studies were rated as follows:

- Direct – No important deviations from the protocol in population, index feature and/or reference standard.
- Partially indirect – Important deviations from the protocol in one of the population, index feature and/or reference standard.
- Indirect – Important deviations from the protocol in at least two of the population, index feature and/or reference standard.

Methods for combining diagnostic test accuracy evidence

Meta-analysis of diagnostic test accuracy data was conducted with reference to the Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy (Deeks et al. 2010).

Where applicable, diagnostic syntheses were stratified by:

- Presenting symptomatology (features shared by all participants in the study, but not all people who could be considered for a diagnosis in clinical practice).
- The reference standard used for true diagnosis.
- Where five or more studies were available for all included strata, a bivariate model was fitted using the `mada` package in R v3.4.0, which accounts for the correlations between positive and negative likelihood ratios, and between sensitivities and specificities. Where sufficient data were not available (2-4 studies), separate independent pooling was performed for positive likelihood ratios, negative likelihood ratios, sensitivity and specificity, using Microsoft Excel. This approach is conservative as it is likely to somewhat underestimate test accuracy, due to failing to account for the correlation and trade-off between sensitivity and specificity (see Deeks 2010).

Random-effects models (der Simonian and Laird) were fitted for all syntheses, as recommended in the Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy (Deeks et al. 2010).

In any meta-analyses where some (but not all) of the data came from studies at high risk of bias, a sensitivity analysis was conducted, excluding those studies from the analysis. Results from both the full and restricted meta-analyses are reported. Similarly, in any meta-analyses where some (but not all) of the data came from indirect studies, a sensitivity analysis was conducted, excluding those studies from the analysis.

Modified GRADE for diagnostic test accuracy evidence

GRADE has not been developed for use with diagnostic studies; therefore a modified approach was applied using the GRADE framework. GRADE assessments were only undertaken for positive and negative likelihood ratios, as the MIDs used to assess imprecision were based on these outcomes, but results for sensitivity and specificity are also presented alongside those data.

Cross-sectional and cohort studies were initially rated as high-quality evidence if well conducted, and then downgraded according to the standard GRADE criteria (risk of bias, inconsistency, imprecision and indirectness) as detailed in Table 5below.

Table 5 Rationale for downgrading quality of evidence for diagnostic questions

GRADE criteria	Reasons for downgrading quality
Risk of bias	Not serious: If less than 33.3% of the weight in a meta-analysis came from studies at moderate or high risk of bias, the overall outcome was not downgraded. Serious: If greater than 33.3% of the weight in a meta-analysis came from studies at moderate or high risk of bias, the outcome was downgraded one level. Very serious: If greater than 33.3% of the weight in a meta-analysis came from studies at high risk of bias, the outcome was downgraded two levels. Outcomes meeting the criteria for downgrading above were not downgraded if there was evidence the effect size was not meaningfully different between studies at high and low risk of bias.

GRADE criteria	Reasons for downgrading quality
Indirectness	Not serious: If less than 33.3% of the weight in a meta-analysis came from partially indirect or indirect studies, the overall outcome was not downgraded. Serious: If greater than 33.3% of the weight in a meta-analysis came from partially indirect or indirect studies, the outcome was downgraded one level. Very serious: If greater than 33.3% of the weight in a meta-analysis came from indirect studies, the outcome was downgraded two levels. Outcomes meeting the criteria for downgrading above were not downgraded if there was evidence the effect size was not meaningfully different between direct and indirect studies.
Inconsistency	Concerns about inconsistency of effects across studies, occurring when there is unexplained variability in the treatment effect demonstrated across studies (heterogeneity), after appropriate pre-specified subgroup analyses have been conducted. This was assessed using the I^2 statistic. N/A: Inconsistency was marked as not applicable if data on the outcome was only available from one study. Not serious: If the I^2 was less than 33.3%, the outcome was not downgraded. Serious: If the I^2 was between 33.3% and 66.7%, the outcome was downgraded one level. Very serious: If the I^2 was greater than 66.7%, the outcome was downgraded two levels. Outcomes meeting the criteria for downgrading above were not downgraded if there was evidence the effect size was not meaningfully different between studies with the smallest and largest effect sizes.
Imprecision	If the 95% confidence interval for a positive likelihood ratio spanned 2, the outcome was downgraded one level, as the data were deemed to be consistent with a meaningful increase in risk and no meaningful predictive value. Similarly, negative likelihood ratios that spanned 0.5 led to downgrading for serious imprecision. Any likelihood ratios that spanned both 0.5 and 2 were downgraded twice, as suffering from very serious imprecision. Outcomes meeting the criteria for downgrading above were not downgraded if the confidence interval was sufficiently narrow that the upper and lower bounds would correspond to clinically equivalent scenarios.

The quality of evidence for each outcome was upgraded if either of the following conditions were met:

- Data showing an effect size sufficiently large that it cannot be explained by confounding alone.
- Data where all plausible residual confounding is likely to increase our confidence in the effect estimate.

Publication bias

Publication bias was assessed in two ways. First, if evidence of conducted but unpublished studies was identified during the review (e.g. conference abstracts or protocols without accompanying published data), available information on these unpublished studies was reported as part of the review. Secondly, where 10 or more studies were included as part of a single meta-analysis, a funnel plot was produced to graphically assess the potential for publication bias.

Incorporating published systematic reviews

For the review question on diagnostic accuracy, systematic reviews containing cross-sectional studies were also included. All included studies from those systematic reviews were

screened to identify any additional relevant primary studies not found as part of the initial search.

Quality assessment

Individual systematic reviews were quality assessed using the ROBIS tool, with each classified into one of the following three groups:

- High quality – It is unlikely that additional relevant and important data would be identified from primary studies compared to that reported in the review, and unlikely that any relevant and important studies have been missed by the review.
- Moderate quality – It is possible that additional relevant and important data would be identified from primary studies compared to that reported in the review, but unlikely that any relevant and important studies have been missed by the review.
- Low quality – It is possible that relevant and important studies have been missed by the review.

Each individual systematic review was also classified into one of three groups for its applicability as a source of data, based on how closely the review matches the specified review protocol in the guideline. Studies were rated as follows:

- Fully applicable – The identified review fully covers the review protocol in the guideline.
- Partially applicable – The identified review fully covers a discrete subsection of the review protocol in the guideline (for example, some of the factors in the protocol only).
- Not applicable – The identified review, despite including studies relevant to the review question, does not fully cover any discrete subsection of the review protocol in the guideline.

Using systematic reviews as a source of data

If systematic reviews were identified as being sufficiently applicable and high quality, and were identified sufficiently early in the review process (for example, from the surveillance review or early in the database search), they were used as the primary source of data, rather than extracting information from primary studies. The extent to which this was done depended on the quality and applicability of the review, as defined in Table 6. When systematic reviews were used as a source of primary data, and unpublished or additional data included in the review which is not in the primary studies was also included. Data from these systematic reviews was then quality assessed and presented in GRADE tables as described below, in the same way as if data had been extracted from primary studies. In questions where data was extracted from both systematic reviews and primary studies, these were cross-referenced to ensure none of the data had been double counted through this process.

Table 6: Criteria for using systematic reviews as a source of data

Quality	Applicability	Use of systematic review
High	Fully applicable	Data from the published systematic review were used instead of undertaking a new literature search or data analysis. Searches were only done to cover the period of time since the search date of the review.
High	Partially applicable	Data from the published systematic review were used instead of undertaking a new literature search and data analysis for the relevant subsection of the protocol. For this section, searches were only done to cover the period of time since the search date

Quality	Applicability	Use of systematic review
		of the review. For other sections not covered by the systematic review, searches were undertaken as normal.
Moderate	Fully applicable	Details of included studies were used instead of undertaking a new literature search. Full-text papers of included studies were still retrieved for the purposes of data analysis. Searches were only done to cover the period of time since the search date of the review.
Moderate	Partially applicable	Details of included studies were used instead of undertaking a new literature search for the relevant subsection of the protocol. For this section, searches were only done to cover the period of time since the search date of the review. For other sections not covered by the systematic review, searches were undertaken as normal.

Health economics

Literature reviews seeking to identify published cost–utility analyses of relevance to the issues under consideration were conducted for all questions. In each case, the search undertaken for the clinical review was modified, retaining population and intervention descriptors, but removing any study-design filter and adding a filter designed to identify relevant health economic analyses. In assessing studies for inclusion, population, intervention and comparator, criteria were always identical to those used in the parallel clinical search; only cost–utility analyses were included. Economic evidence profiles, including critical appraisal according to the Guidelines manual, were completed for included studies.

Economic studies identified through a systematic search of the literature are appraised using a methodology checklist designed for economic evaluations (NICE guidelines manual; 2014). This checklist is not intended to judge the quality of a study per se, but to determine whether an existing economic evaluation is useful to inform the decision-making of the committee for a specific topic within the guideline.

There are 2 parts of the appraisal process. The first step is to assess applicability (that is, the relevance of the study to the specific guideline topic and the NICE reference case); evaluations are categorised according to the criteria in Table 7.

Table 7 Applicability criteria

Level	Explanation
Directly applicable	The study meets all applicability criteria, or fails to meet one or more applicability criteria but this is unlikely to change the conclusions about cost effectiveness
Partially applicable	The study fails to meet one or more applicability criteria, and this could change the conclusions about cost effectiveness
Not applicable	The study fails to meet one or more applicability criteria, and this is likely to change the conclusions about cost effectiveness. These studies are excluded from further consideration

In the second step, only those studies deemed directly or partially applicable are further assessed for limitations (that is, methodological quality); see categorisation criteria in Table 8.

Table 8 Methodological criteria

Level	Explanation
Minor limitations	Meets all quality criteria, or fails to meet one or more quality criteria but this is unlikely to change the conclusions about cost effectiveness
Potentially serious limitations	Fails to meet one or more quality criteria and this could change the conclusions about cost effectiveness
Very serious limitations	Fails to meet one or more quality criteria and this is highly likely to change the conclusions about cost effectiveness. Such studies should usually be excluded from further consideration

Where relevant, a summary of the main findings from the systematic search, review and appraisal of economic evidence is presented in an economic evidence profile alongside the clinical evidence.

Appendix C – Literature search strategies

Medline strategy

1 exp Neoplasms, Squamous Cell/

2 exp "Head and Neck Neoplasms"/

3 Neoplasms, Unknown Primary/

4 (("head and neck" or "head adj neck" or "upper aero-digestive" or "upper aerodigestive" or "upper airway*" or "upper respiratory" or UAT or UADT) adj3 (neoplas* or cancer* or carcinoma* or adenocarcinom* or tumour* or tumor* or malignan* or metasta* or angiosarcoma* or sarcoma* or teratoma* or lymphoma* or blastoma* or microcytic* or carcino* or leiomyosarcoma* or lump*)).tw.

5 ((squamous or epidermoid or planocellular) adj3 (neoplas* or cancer* or carcinoma* or adenocarcinom* or tumour* or tumor* or malignan* or metasta* or angiosarcoma* or sarcoma* or teratoma* or lymphoma* or blastoma* or microcytic* or carcino* or leiomyosarcoma* or lump*)).tw.

6 ((mouth or oral or intra-oral or intraoral or oral mucos* or lip* or tongue or cheek* or gingiva* or gum* or palat* or buccal or buccal mucosa* or maxilla* or tonsil*) adj2 (neoplas* or cancer* or carcinoma* or adenocarcinom* or tumour* or tumor* or malignan* or metasta* or angiosarcoma* or sarcoma* or teratoma* or lymphoma* or blastoma* or microcytic* or carcino* or leiomyosarcoma* or lump*)).tw.

7 ((throat* or pharynx* or salivar* gland or parotid* gland* or sublingual* gland* or submandibular* gland* or nose* or nasal* or paranasal* or nasosinus* or sininasal* or sinus* or odontogenic* or face or facial or maxilla* or pharyngeal*) adj2 (neoplas* or cancer* or carcinoma* or adenocarcinom* or tumour* or tumor* or malignan* or metasta* or angiosarcoma* or sarcoma* or teratoma* or lymphoma* or blastoma* or microcytic* or carcino* or leiomyosarcoma* or lump*)).tw.

8 ((oropharyn* or retromolar trigone) adj2 (neoplas* or cancer* or carcinoma* or adenocarcinom* or tumour* or tumor* or malignan* or metasta* or angiosarcoma* or sarcoma* or teratoma* or lymphoma* or blastoma* or microcytic* or carcino* or leiomyosarcoma* or lump*)).tw.

9 ((hypopharyn* or laryngopharyn* or nasopharyn*) adj2 (neoplas* or cancer* or carcinoma* or adenocarcinom* or tumour* or tumor* or malignan* or metasta* or angiosarcoma* or sarcoma* or teratoma* or lymphoma* or blastoma* or microcytic* or carcino* or leiomyosarcoma* or lump*)).tw.

10 ((laryn* or glotti* or epiglotti* or subglotti* or supraglotti* or vocal cord* or vocal fold* or voice box* or cordal) adj2 (neoplas* or cancer* or carcinoma* or adenocarcinom* or tumour* or tumor* or malignan* or metasta* or angiosarcoma* or sarcoma* or teratoma* or lymphoma* or blastoma* or microcytic* or carcino* or leiomyosarcoma* or lump*)).tw.

11 or/1-10

12 exp Positron-Emission Tomography/

13 exp Fluorodeoxyglucose F18/

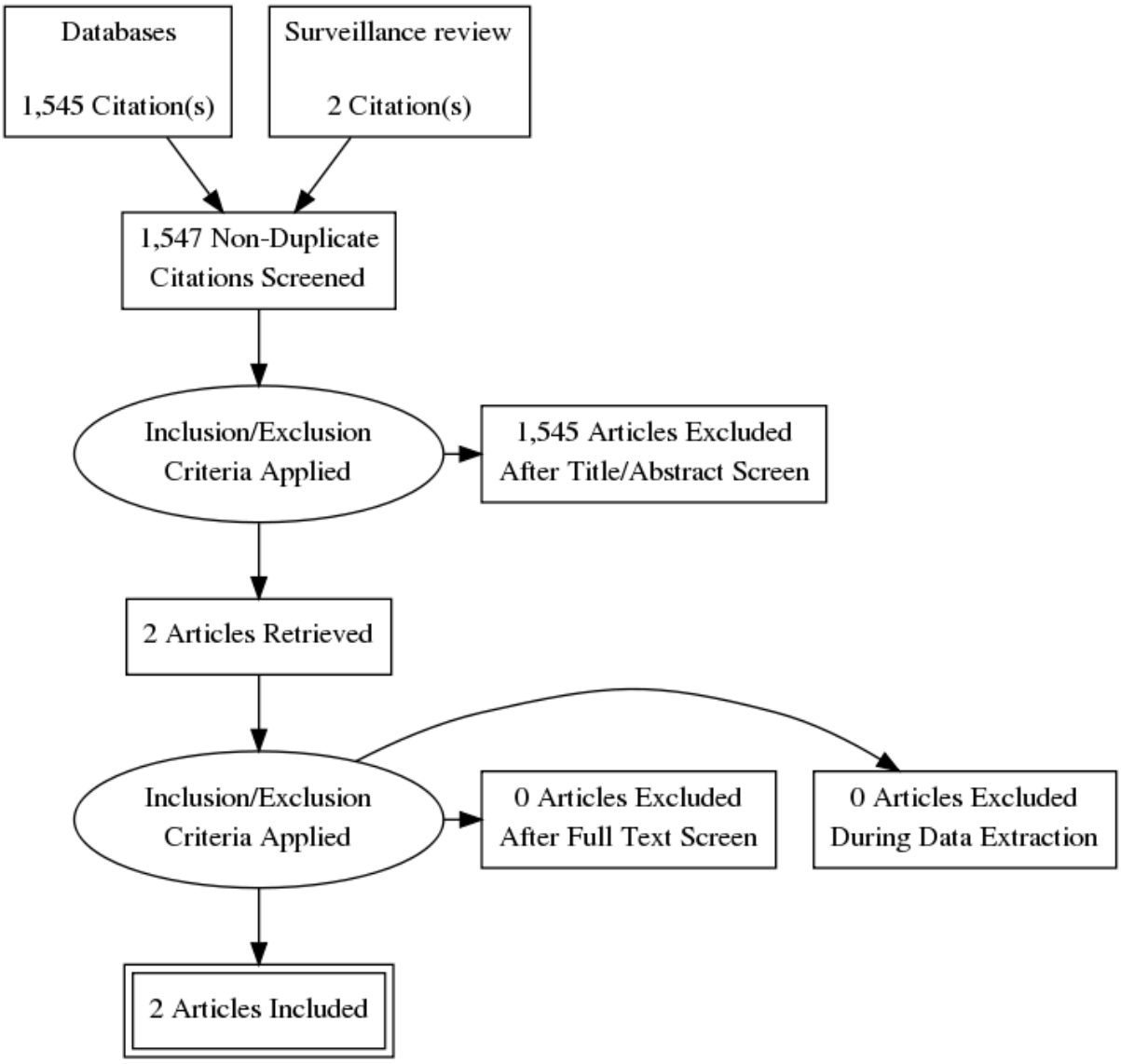
14 (PET-CT* or Positron Emission Tomography or FDG PET* or 18F-FDG* or 18F-fluorodeoxyglucose*).tw.

1 15 or/12-14
2 16 11 and 15
3 17 Randomized Controlled Trial.pt.
4 18 Controlled Clinical Trial.pt.
5 19 Clinical Trial.pt.
6 20 exp Clinical Trials as Topic/
7 21 Placebos/
8 22 Random Allocation/
9 23 Double-Blind Method/
10 24 Single-Blind Method/
11 25 Cross-Over Studies/
12 26 ((random$ or control$ or clinical$) adj3 (trial$ or stud$)).tw.
13 27 (random$ adj3 allocat$).tw.
14 28 placebo$.tw.
15 29 ((singl$ or doubl$ or trebl$ or tripl$) adj (blind$ or mask$)).tw.
16 30 (crossover$ or (cross adj over$)).tw.
17 31 or/17-30
18 32 specificity.tw.
19 33 Cross-Sectional Studies/
20 34 cross sectional.tw.
21 35 or/32-34
22 36 16 and 31
23 37 16 and 35
24 38 LETTER/ or EDITORIAL/ or NEWS/ or COMMENT/ or exp HISTORICAL ARTICLE/ or
25 CASE REPORT/
26 39 (editorial or case reports or clinical conference).pt.
27 40 animals/ not humans/
28 41 or/38-40
29 42 36 not 41
30 43 limit 42 to english language
31 44 37 not 41
32 45 limit 44 to english language

1

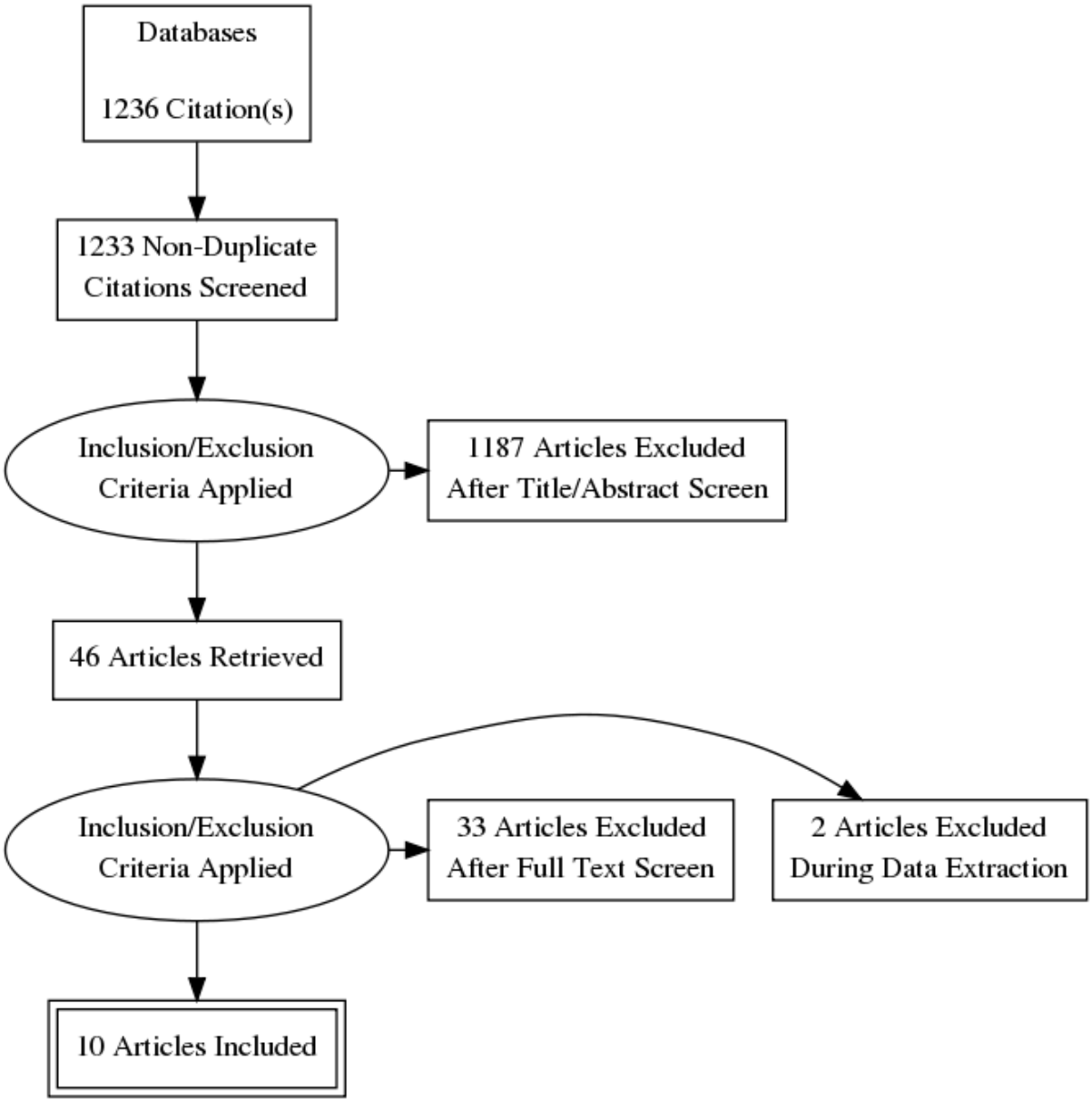
Appendix D – Clinical evidence study selection

Management of nodal metastasis in head and neck cancer after chemoradiotherapy



Diagnostic accuracy of PET-CT to diagnose residual nodal disease after radiotherapy or chemoradiotherapy

46 articles were ordered and 45 have been retrieved. 1 was unavailable for exclusion.



1 Appendix E – Clinical evidence tables

2 Management of nodal metastasis in head and neck cancer after chemoradiotherapy

Author (year)	Title	Study details	Quality assessment
Mehanna (2017)	PET-NECK: a multicentre randomised Phase III non-inferiority trial comparing a positron emission tomography-computerised tomography-guided watch-and-wait policy with planned neck dissection in the management of locally advanced (N2/N3) nodal metastases in patients with squamous cell head and neck cancer.	Related publications • Mehanna 2016 Study type • Randomised controlled trial Study details • Study location UK • Study setting H&N cancer-treating centres throughout UK NHS hospital trusts • Study dates Recruitment took place between October 2007 and August 2012 • Duration of follow-up Median time to follow-up was 36 months • Sources of funding National Institute for Health Research (NIHR) Health Technology Assessment programme Inclusion criteria • Histological diagnosis of oropharyngeal, laryngeal, oral, hypopharyngeal or occult HNSCC • Clinical and CT/MRI imaging evidence of nodal metastases, stage N2 (a, b or c) or N3 Patients with N2 or N3 histologically and/or cytologically proven squamous cell carcinoma and an occult primary (after EUA and PET–CT scan) were eligible for the trial if they were going to be treated with CRT • Multidisciplinary team (MDT) decision to receive curative radical	Random sequence generation • Low risk of bias Allocation concealment • Low risk of bias Blinding of participants and personnel • Low risk of bias Blinding of outcome assessment • Low risk of bias Incomplete outcome data • Low risk of bias Selective reporting • Low risk of bias Other sources of bias • Low risk of bias Overall risk of bias • Low

Author (year)	Title	Study details	Quality assessment
		concurrent CRT for primary • Indication to receive one of the CRT regimens approved by the study • Fit for ND surgery • ND was technically feasible to perform to remove nodal disease For example: no carotid encasement, no direct extension between tumour and nodal disease • Aged ≥ 18 years • Able to provide written informed consent Exclusion criteria • Tumours that were not squamous cell carcinomas histologically • Undergoing resection for their primary tumour, for example resection of the tonsil or base of tongue with flap reconstruction Diagnostic tonsillectomy was not considered an exclusion criteria • N1 stage nodal metastasis • Receiving neoadjuvant CRT with no concomitant chemotherapy • Receiving adjuvant chemotherapy • Undergoing chemotherapy with or without radiotherapy for palliative purposes • Undergoing radiotherapy alone This is not an optimal treatment for neck node disease • Distant metastases to the chest, liver, bones or other sites • Unfit for surgery or CRT • Received previous treatment for HNSCC • Primary nasopharyngeal carcinoma • Being pregnant • Another cancer diagnosis in the past 5 years With the exception of basal cell carcinoma or carcinoma of the cervix in situ Sample characteristics • Sample size 564 • Split between study groups	Directness • Directly applicable

Author (year)	Title	Study details	Quality assessment
		FDG-PET-CT-guided active surveillance: 282; Planned neck dissection: 282 • Loss to follow-up FDG-PET-CT-guided active surveillance: 55 out of 282; Planned neck dissection: 72 out of 282 • %female FDG-PET-CT-guided active surveillance: 20.9; Planned neck dissection: 16.0 • Mean age (SD) FDG-PET-CT-guided active surveillance: 57.6 years (7.5); Planned neck dissection: 58.2 years (8.1) • Tumour site % FDG-PET-CT-guided active surveillance: Oral 1.4; Oropharyngeal 85.1; Laryngeal 6.4; Hypopharyngeal 5.3; Occult H&N 1.8; Planned neck dissection: Oral 2.5; Oropharyngeal 83.7; Laryngeal 6.7; Hypopharyngeal 5.0; Occult H&N 2.1 • T stage % FDG-PET-CT-guided active surveillance: T1 17.0; T2 40.4; T3 21.6; T4 19.5; Occult 1.4; Planned neck dissection: T1 18.4; T2 38.3; T3 18.4; T4 22.7; Occult 2.1 • N stage % FDG-PET-CT-guided active surveillance: N2a 19.1; N2b 29.2; N2c 18.4; N3 3.2; Planned neck dissection: N2a 15.6; N2b 63.1; N2c 18.4; N3 2.8 • p16 status % FDG-PET-CT-guided active surveillance: p16 positive 72.6 (164/226); p16 negative 27.4 (62/226); p16 test not done or result not available n=56; Planned neck dissection: p16 positive 77.7 (171/220); p16 negative 22.3 (49/220); p16 test not done or result not available n=62 Interventions • FDG-PET-CT-guided active surveillance FDG PET-CT 12 weeks after completion of chemoradiotherapy • Planned neck dissection	

Author (year)	Title	Study details	Quality assessment
		Planned neck dissection before or after chemoradiotherapy Outcome measure(s) • Recurrence rates Any notification of recurrence within 3 months of radiotherapy was regarded as persistent disease and notifications after that date were regarded as recurrences. Recurrences in the neck nodes were reported as ipsilateral or contralateral in the notification of recurrence • Overall survival Information on death and survival was obtained from centres via death and follow-up forms • Quality of life - EQ-5D: five 3-point scales and one summary 100-point scale. This was used for the health economics evaluation. - European Organisation for Research and Treatment of Cancer (EORTC) Quality of Life Questionnaire for Cancer (with 30 questions) (QLQ-C30): five functional, three symptom and a global scale and six single items for assessment of general quality of life. - EORTC Quality of Life Questionnaire for Cancer head and neck module with 35 questions (QLQ-H&N35): seven scales and 11 single items for H&N cancer-related quality of life. - MD Anderson Dysphagia Inventory (MDADI): an overall function scale and four subscales. • Surgical complications Complications following neck dissection surgery were reported • Adverse events Investigators were required to inform the trials unit immediately of any serious adverse events (SAEs) following chemotherapy, positron emission tomography-computed tomography or neck dissection. Each time the patient was seen in clinic he or she was asked if any SAEs had occurred. The occurrence of SAEs was based on information provided by either patients or their carers. The following adverse events were considered serious: - death - life-threatening disease - hospitalisation or prolongation of hospitalisation - congenital abnormality - persistent disability - other medically significant event	

Author (year)	Title	Study details	Quality assessment
		Subgroups • Staging Tumour stage: T1 or T2; T3 or T4; occult disease; Nodal stage: N2a or N2b; N2c or N3 • HPV status p16 status: positive, negative, not known • Cancer site Oral cavity; oropharynx; larynx; hypopharynx; occult disease	

1 **Diagnostic accuracy of PET-CT to diagnose residual nodal disease after radiotherapy or chemoradiotherapy**

Author (year)	Title	Study details	Quality assessment
Helsen (2017)	18F-FDG-PET/CT for the detection of disease in patients with head and neck cancer treated with radiotherapy	**Study type** • Retrospective cohort study **Study location** • Belgium **Study setting** • Hospital **Study dates** • July 2005 - May 2009	**Patient selection** • Low risk of bias **Reference standard** • High risk of bias Unclear if results of reference test done without knowledge of index test results. **Flow and timing** • High risk of bias Unclear on interval between index and reference test. One patient was excluded because of insufficient follow-up. Sum of patients reported for Stage and Therapy in Table 1 n = 104 and not n = 103. **Overall risk of bias** • High Reviewer did not know result of outcome but did know the result of other imaging

Author (year)	Title	Study details	Quality assessment
		Sources of funding • None declared **Sample size** • 103 **%female** • 19.4 **Median age (range)** • 61 (38 - 87) **Patients who received CRT (%)** • 81.6 **Oropharynx (%)** • 38.8 **Nasopharynx (%)** • 2.9	modalities. Unclear if results of reference test done without knowledge of index test results. Unclear on interval between index and reference test. One patient was excluded because of insufficient follow-up. Sum of patients reported for Stage and Therapy in Table 1 n = 104 and not n = 103. **Directness** • Directly applicable

Author (year)	Title	Study details	Quality assessment
		Hypopharynx (%) • 14.6 **Larynx (%)** • 29.1 **Unknown (%)** • 1.9 **Oral cavity (%)** • 12.6 **Patients with residual disease** • 41/103 (39.8%) **Inclusion criteria** • Histologically confirmed SCC of head and neck • Primary treatment of CRT • Minimum follow-up 12 months • Treatment with curative intent **Exclusion criteria** • Distant metastases • Another malignancy 5 years prior to HNSCC	

Author (year)	Title	Study details	Quality assessment
		diagnosis **Index test(s)** • PET-CT **PET-CT procedure** • Patients underwent FDG-PET/CT 5-19 weeks after CRT for detection • Patients were instructed to fast for 6 hours prior to their appointment and blood glucose levels were measured prior to the injection of FDG. • Performed on a Siemens Biograph 6 HIREZ scanner: 60±90 minutes after tracer injection (4 MBq/kg) • Patients received a dedicated head and neck image, with a higher resolution PET (acquisition 10 min per bed position, from vertex to aortic root, reconstructed in a 336x336 matrix; 6 iterations, 16 subsets) **Image interpretation** • The certified nuclear medicine physician was aware of the clinical history and results of other imaging modalities but not of outcome. • FDG-PET images were interpreted qualitatively through visual analysis. • The reports were retrospectively reviewed and	

Author (year)	Title	Study details	Quality assessment
		classified into positive, negative or equivocal for residual disease. • Positive if residual focal FDG-uptake was a greater intensity than background bloodpool activity or surrounding normal tissue and outside normal anatomic structures seen on CT. • Equivocal reports were re-analysed by a nuclear physician and categorised as positive or negative. If doubt remained, the scan was read positive. • No predetermined SUV threshold was used in the analysis. **Reference standard(s)** • Histopathology for all patients • Clinical assessment	
Keski-Santti (2014)	FDG-PET/CT in the Assessment of Treatment Response after Oncologic Treatment of Head and Neck Squamous Cell Carcinoma	**Study type** • Retrospective cohort study **Study location** • Finland **Study setting** • Hospital **Study dates** • 2008 - 2010, chosen to ensure a sufficient	**Patient selection** • Low risk of bias **Reference standard** • Low risk of bias **Flow and timing** • Unclear risk of bias Unclear on interval between index and reference test. **Overall risk of bias** • Moderate

Author (year)	Title	Study details	Quality assessment
		sample size and adequate follow-up time for the analysis. **Sources of funding** • None declared **Sample size** • 88 **%female** • 22 **Patients who received CRT (%)** • 86.0 **Oropharynx (%)** • 44 **Nasopharynx (%)** • 5	Unclear on interval between index and reference test. **Directness** • Directly applicable

Author (year)	Title	Study details	Quality assessment
		Hypopharynx (%) • 23 **Larynx (%)** • 27 **Oral cavity (%)** • 1 **Patients with residual disease** • 17/88 (19.3%) **Inclusion criteria** • Patients with previously untreated HNSCC • Primary treatment of CRT **Exclusion criteria** • PET-CT performed before 10 weeks and after 18 weeks post-treatment completion **Index test(s)** • PET-CT	

Author (year)	Title	Study details	Quality assessment
		PET-CT procedure • PET-CT performed 10-18 weeks after treatment completion. • Patients were instructed to fast for 6 hours prior to their appointment and blood glucose levels were measured prior to the injection of FDG. • 60 minutes between 5 MBq/kg injection of 18F-FDG and imaging • PET-CT performed 10-18 weeks after treatment completion. • Gemini PET/CT scanner (Philips) • PET-CT performed 10-18 weeks after treatment completion. • Full-body images from clavicle to mid-thigh obtained first (CT, 120–140 kV, 50–60 mAs, a section width of 4 mm then PET (PET, 8 cm bed position, 1.5 minutes per frame) • Head and neck images performed with arms down (CT, 120 KeV, 50 mAs, a section width of 3 mm and PET 8 cm bed position, 2.5 minutes per frame) **Image interpretation** • Focal uptake distinguishable from the background, which could not be considered physiologic, reactive, or inflammatory, was interpreted to be pathological uptake. • No predetermined SUV threshold was used in	

Author (year)	Title	Study details	Quality assessment
		the analysis. **Reference standard(s)** • Clinical assessment • Histopathologic data as a reference standard were available for all patients who had had positive FDG-PET/CT or clinical suspicion of residual disease indicating a neck dissection and/or biopsies from the primary tumour area.	
Nayak (2007)	Deferring planned neck dissection following chemoradiation for stage IV head and neck cancer: The utility of PET-CT	**Study type** • Prospective cohort study **Study location** • USA **Study setting** • Hospital **Sources of funding** • None declared **Sample size** • 43	**Patient selection** • Low risk of bias **Reference standard** • Low risk of bias **Flow and timing** • Low risk of bias **Overall risk of bias** • Low **Directness** • Directly applicable

Author (year)	Title	Study details	Quality assessment
		Patients who received CRT (%) • 100 **Oropharynx (%)** • 93 **Nasopharynx (%)** • 2.3 **Base of tongue (%)** • 55.8 **Tonsil (%)** • 23.3 **Epiglottis (%)** • 11.6 **Pyriform sinus (%)** • 2.3	

Author (year)	Title	Study details	Quality assessment
		Unknown (%) • 2.3 **Patients with residual disease** • 8/43 (18.6%) **Inclusion criteria** • Clinical and radiologic stage N2 N3 nodal metastases • de novo cervical equal to or more than N2 naso-, oro- and hypopharyngeal • Primary treatment of CRT **Exclusion criteria** • Distant metastases **Index test(s)** • PET-CT **PET-CT procedure** • Patients were instructed to fast for 6 hours prior to their appointment and blood glucose levels were measured prior to the injection of FDG. • All but one patient received scans 2-5 months post-CRT. One patient returned 6 months post-	

Author (year)	Title	Study details	Quality assessment
		CRT. • Reveal scanner (CTI Medical Systems), which combines LSO Allegra PET and dual-channel CT. • PET-CT imaging from skull base through the abdomen was performed approximately 1 hour following an intravenous injection of 8 to 15 mCi 18F-FDG. • Helical CT (pitch = 1.0, mAs 120–140, kVp 130) was performed immediately preceding the acquisition of PET emission data and with Optiray-350 intravenous contrast. • The PET images were reconstructed with and without CT-based attenuation correction. **Image interpretation** • Axial CT images were reviewed on a PACS workstation. • PET images, fused PET-CT images, and reconstructions of PET, CT, and PET-CT into sagittal and coronal planes were reviewed on a Syngo Fusion Workstation. • Images were reviewed by one of two fellowship-trained CAQ-certified neuroradiologists who spend the majority of their clinical time reading head and neck imaging. • A PET-CT was defined as “positive" if 1) the radiologist recommended nodal tissue biopsy or resection of cervical disease based on increased metabolic activity and suspicious radiographic	

Author (year)	Title	Study details	Quality assessment
		characteristics in the neck • 2) progressive hypermetabolic activity was identified in the neck, but in the setting of distant metastatic disease further surgical intervention was not warranted. **Reference standard(s)** • Histopathologic data as a reference standard were available for all patients who had had positive FDG-PET/CT or clinical suspicion of residual disease indicating a neck dissection and/or biopsies from the primary tumour area.	
Ng (2011)	PET/CT and 3-T whole-body MRI in the detection of malignancy in treated oropharyngeal and hypopharyngeal carcinoma	**Study type** • Prospective cohort study **Study location** • Taiwan **Study setting** • Hospital **Study dates** • June 2006 - June 2009	**Patient selection** • High risk of bias Only patients with high risk for residual/recurrent disease included. **Reference standard** • Low risk of bias **Flow and timing** • Low risk of bias **Overall risk of bias** • High Only patients with high risk for residual/recurrent disease included. **Directness**

Author (year)	Title	Study details	Quality assessment
		Sources of funding • National Science Council (Taiwan) **Sample size** • 79 **%female** • 12.9 **Median age (range)** • 52.4 (33 - 74) **Patients who received CRT (%)** • 100 **Oropharynx (%)** • 41.8 **Hypopharynx (%)** • 58.2	• Directly applicable

Author (year)	Title	Study details	Quality assessment
		Patients with residual disease • 16/79 (20.3%) **Inclusion criteria** • Patients with suspected residual/recurrent disease • Initial diagnosis of stage III-IV **Exclusion criteria** • None reported **Index test(s)** • PET-CT **PET-CT procedure** • Patients were instructed to fast for 6 hours prior to their appointment and blood glucose levels were measured prior to the injection of FDG. • Discovery ST 16 PET-CT system • Helical CT head to proximal thigh performed before PET acquisition (transverse 3.0-mm collimation x 16 modes, 100 kVp, 100mA, 0.5-s tube rotation, 35 mm/s table speed and pitch 1.5). • No oral or intravenous iodinated contrast was administered. • PET images acquired in D for 3 min per table	

Author (year)	Title	Study details	Quality assessment
		position. PET images were reconstructed with CT for attenuation correction using an ordered-subset expectation maximisation iterative reconstruction algorithm. • Performed a mean of 6.5 months (range, 2.8 - 24.6 months) post-CRT • PET performed 50-70mins after F-FDG injection. **Image interpretation** • Any focus of FDG uptake greater than the surrounding background and not attributable to normal FDG biodistribution was assessed. The intensity of FDG uptake was graded using a five-point scale. • Two radiologists and one nuclear medicine physician interpreted the images and were blinded to the results of other imaging techniques but were aware of the protocol. • Retropharyngeal node is considered metastatic if its minimal axial diameter ≥ 5 mm **Reference standard(s)** • Histopathology for all patients *Histology within 12 months*	

Author (year)	Title	Study details	Quality assessment
Pellini (2014)	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	**Study type** • Prospective cohort study **Study location** • Italy **Study setting** • Hospital **Study dates** • January 2006 - February 2009 **Sample size** • 36 **%female** • 36.1 **Median age (range)** • 61.4 (42 - 71)	**Patient selection** • High risk of bias 3 patients excluded because of incomplete response to CRT. **Reference standard** • Low risk of bias **Flow and timing** • Unclear risk of bias Unclear on interval between index and reference test. **Overall risk of bias** • Moderate 3 patients excluded because of incomplete response to CRT. Unclear on interval between index and reference test. **Directness** • Directly applicable

Author (year)	Title	Study details	Quality assessment
		Patients who received CRT (%) • 100 **Oropharynx (%)** • 100 **Patients with residual disease** • 18/37 (48.6%) **Inclusion criteria** • Clinical and radiologic stage N2 N3 nodal metastases • Primary treatment of CRT • Bulky nodal disease with lymph nodes > 3cm at diagnosis • Informed consent of participants **Exclusion criteria** • Distant metastases • Surgery or other treatments prior to CRT • Patients not able to complete the standard treatment protocol • Contraindication the imaging examinations	

Author (year)	Title	Study details	Quality assessment
		Index test(s) • PET-CT **PET-CT procedure** • Median time interval between CRT and PET-CT was 12.4 weeks (range, 12 - 13 weeks). • PET-CT scan performed with 15 mCi of 18F-FDG injected intravenously and performed 45 min after injection • From skull base to upper thighs • Biograph (Siemens Medical Solutions, Inc.) or Discovery LS (GE Healthcare) with images acquired for 4 mins per bed position. • Response to treatment was recorded using an SUV of 2 as a threshold. • PET-CT images interpreted by a nuclear medicine physician with consensus by a radiologist. Readers were blind to the results of other imaging modalities and to the final neck pathology report. **Image interpretation** • PET-CT images interpreted by a nuclear medicine physician with consensus by a radiologist. Readers were blind to the results of other imaging modalities and to the final neck	

Author (year)	Title	Study details	Quality assessment
		pathology report. **Reference standard(s)** • Histopathology for all patients • Tissues were fixed in unbuffered 10% formalin solution and embedded in paraffin. Serial sections of 6 µm thickness were stained with haematoxylin/eosin and PAS. Immunohistochemistry was performed with the biotin avidin complex peroxidase method. • Histology was graded into 5 categories depending on histological-proved metastatic and tumour-free nodes in each neck and initial N-status. Grades 2-4 were considered as CRT failures as residual viable tumour cells remained.	
Prestwich (2012)	Delayed response assessment with FDG-PET-CT following (chemo) radiotherapy for locally advanced head and neck squamous cell carcinoma	**Study type** • Retrospective cohort study **Study location** • UK **Study setting** • Hospital	**Patient selection** • Low risk of bias **Reference standard** • Low risk of bias **Flow and timing** • Low risk of bias **Overall risk of bias** • High Two PET-CT systems were used depending on when the scan was undertaken, Discovery STE PET-CT (GE Healthcare) prior to June 2010 and 64-

Author (year)	Title	Study details	Quality assessment
		Study dates • August 2008 - April 2011 **Sources of funding** • None declared **Sample size** • 44 **%female** • 30 **Median age (range)** • 55 (29-75) **Patients who received CRT (%)** • 56.8 **Oropharynx (%)** • 68	section Philips Gemini TF64 system after June 2010. Unclear on interval between index and reference test. Not all patients received a reference standard and the reference standard varied. **Directness** • Partially directly applicable Only 56.8% of patients received CRT

Author (year)	Title	Study details	Quality assessment
		Hypopharynx (%) • 14 **Larynx (%)** • 7 **Unknown (%)** • 9 **Paranasal sinuses (%)** • 2 **Patients with residual disease** • 7/44 (56.8%) **Inclusion criteria** • Histologically confirmed SCC of head and neck • Initial diagnosis of stage III-IV • Reviewed by a specialist head and neck multidisciplinary team meeting • TNM stage III or IV • Received radical non-surgical treatment • PET-CT performed as a baseline prior to	

Author (year)	Title	Study details	Quality assessment
		treatment **Exclusion criteria** • Surgery or other treatments prior to CRT • Nasopharynx cancer • FDG PET-CT performed only following response assessment with CT and/or MRI **Index test(s)** • PET-CT **PET-CT procedure** • Gemini PET/CT scanner (Philips) • Discovery ST 16 PET-CT system • PET-CT performed a median of 16.8 weeks (range, 9 - 24 weeks) from CRT completion • Acquisition from skull vertex to upper thighs performed 60 mins after 440 MBq of intravenous FDG was administered. • CT was set to 140 kV, 80 mAs, tube rotation time 0.5s per rotation, pith 6, section thickness 3.75mm **Image interpretation** • Images were categorised into positive, equivocal and negative. A positive image	

Author (year)	Title	Study details	Quality assessment
		included focal FDG uptake corresponding to a structural abnormality and being of greater intensity than background liver activity. Uptake was classed as equivocal if focal FDG **Reference standard(s)** • Histopathology for all patients • Clinical assessment	
Seng (2008)	Clinical utility of 18F-FDG PET/CT in assessing the neck after concurrent chemoradiotherapy for locoregional advanced head and neck cancer	**Study type** • Retrospective cohort study **Study location** • USA **Study setting** • Hospital **Study dates** • March 2002 - December 2004 **Sources of funding** • None declared	**Patient selection** • Low risk of bias **Reference standard** • Low risk of bias **Flow and timing** • High risk of bias Interval between index and reference test range 0 - 6 months. **Overall risk of bias** • High Interval between index and reference test range 0 - 6 months. **Directness** • Directly applicable

Author (year)	Title	Study details	Quality assessment
		Sample size • 82 **%female** • 30 **Median age (range)** • 55 (29-75) **Inclusion criteria** • Histologically confirmed SCC of head and neck • Primary treatment of CRT • PET-CT no later than 6 months post-treatment **Exclusion criteria** • Distant metastases • Nasopharynx cancer • Cancer of paranasal sinus • Cancer of salivary glands • Treated with palliative intent **Index test(s)** • PET-CT	

Author (year)	Title	Study details	Quality assessment
		PET-CT procedure • PET-CT scan performed with 15 mCi of 18F-FDG injected intravenously and performed 45 min after injection • From skull base to upper thighs • Biograph (Siemens Medical Solutions, Inc.) or Discovery LS (GE Healthcare) with images acquired for 4 mins per bed position. • 15 mCi of 18F-FDG injected intravenously and PET performed 45 min after injection. • CT data used for attenuation correction and anatomic localisation • Median interval between CRT completion and scan was 12 weeks (range, 8 - 27 weeks). **Image interpretation** • Investigator was unaware of other imaging findings, clinical findings or patient outcome • Whenever available, baseline PET/CT was used for comparison. • The reports were retrospectively reviewed and classified into positive, negative or equivocal for residual disease and then cross-referenced with the original clinical PET/CT report. A second investigator reviewed the scans when disputes arose between the • Images were reviewed on a picture archiving and communication system (PACS) workstation (AWsuite; GE Healthcare).	

Author (year)	Title	Study details	Quality assessment
		• 18F-FDG uptake was considered abnormal when it was focal (rather than diffuse), outside normal anatomic structures seen on companion CT, and of intensity greater than background blood-pool activity or uptake in adjacent normal tissue. • SUVs were obtained for lesions with focal 18F-FDG uptake and background SUV were measured for these lesions from the contralateral nromal neck side and the treated disease site. **Reference standard(s)** • Viable tumour cells were defined as those epithelial cells present within a lymph node, adjacent fibroadipose tissue, skeletal muscle, or other structures, which were morphologically identifiable and recognizable as squamous.	
Sjovall (2014)	Radiotherapy response in head and neck cancer - evaluation of the primary tumour site	**Study type** • Prospective cohort study **Study location** • Sweden **Study setting** • Hospital	**Patient selection** • Low risk of bias **Reference standard** • Low risk of bias **Flow and timing** • Low risk of bias **Overall risk of bias** • Low

Author (year)	Title	Study details	Quality assessment
		Sources of funding • Swedish Foundation Acta Oto-Laryngoloica **Sample size** • 82 **%female** • 24 **Median age (range)** • 62 (34 - 89) **Oropharynx (%)** • 85 **Hypopharynx (%)** • 6 **Larynx (%)** • 8	**Directness** • Directly applicable

Author (year)	Title	Study details	Quality assessment
		HPV +ve (%) • 69 **Inclusion criteria** • Histologically confirmed SCC of head and neck • Primary treatment of CRT • Treatment with curative intent **Exclusion criteria** • Distant metastases • Not undergone baseline PET-CT • Lack of visible hypermetabolism on baseline PET-CT **Index test(s)** • PET-CT **PET-CT procedure** • Gemini PET/CT scanner (Philips) • Patients were injected intravenously with 4 MBq/kg body weight of FDG to a maximum dose of 400 MBq after 4h fasting and images were acquired 1h after FGD injected. • Images interpreted by visual inspection and FDG uptake above background is described as hypermetabolism, no hypermetabolism or	

Author (year)	Title	Study details	Quality assessment
		equivocal **Image interpretation** • Images interpreted by visual inspection and FDG uptake above background is described as hypermetabolism, no hypermetabolism or equivocal **Reference standard(s)** • Histopathology for all patients • Endoscopy with or without biopsy (biopsy in 66 tumours or 65 patients)	
Taghipour (2017)	Post-treatment 18F-FDG-PET/CT versus contrast-enhanced CT in patients with oropharyngeal squamous cell carcinoma: comparative effectiveness study	**Study type** • Retrospective cohort study **Study location** • USA **Study setting** • Hospital **Study dates** • 2000 - 2013	**Patient selection** • Low risk of bias **Reference standard** • Low risk of bias **Flow and timing** • Unclear risk of bias **Overall risk of bias** • Moderate Not all patients received a reference standard and the reference standard varied. Unclear on interval between index and reference test.

Author (year)	Title	Study details	Quality assessment
		Sources of funding • NIH T32 grant **Sample size** • 110 **%female** • 13.8 **Oropharynx (%)** • 100 **Inclusion criteria** • Histologically confirmed SCC of head and neck • PET-CT no later than 6 months post-treatment **Exclusion criteria** • None reported **Index test(s)** • PET-CT	**Directness** • Directly applicable

Author (year)	Title	Study details	Quality assessment
		PET-CT procedure • All patients were instructed to fast for 4h before scanning. 18F-FDG was injected at a dose of 5.55 MBq/kg. • After 60min uptake, patients were scanned by a whole-body PET from clavicle to the mid-thigh with arms above head, followed by a dedicated head and neck PET from the top of the head to carnia, with arms by the side. • Non-contrast CT scan performed after each PET scan for attenuation correction and anatomical coregistration purposes. • Images obtained by using Discovery LS (2D)(GE Healthcare). The CT for the attenuation correction was performed at 120kV, 20-200mA, 8.0 noise indec, a 512x512 matrix with a beam collimation of 10mm and a pitch of 0.984. • The PET images were obtained at 4.15min/bed position, slice thickness 3.75mm, matrix 128x128 with a field of view of 50cm for the whole-body exam and 25cm for the head and neck exam. **Reference standard(s)** • Histopathology for all patients • Clinical assessment	
Van Den Wyngaert (2017)	Fluorodeoxyglucose-positron emission tomography/computed tomography after concurrent chemoradiotherapy in locally	**Study type** • Prospective cohort study	**Patient selection** • Low risk of bias **Reference standard**

Author (year)	Title	Study details	Quality assessment
	advanced head-and-neck squamous cell cancer: The ECLYPS study	**Study location** • Belgium • Netherlands **Study setting** • Hospital **Study dates** • March 2011 - February 2014 **Sources of funding** • Flemish Agency for Innovation by Science and Technology **Sample size** • 125 **%female** • 25.6 **Median age (range)** • 59 (IQR 11)	• Low risk of bias **Flow and timing** • Low risk of bias **Overall risk of bias** • Moderate Unclear if results of index test done without knowledge of reference test results. **Directness** • Directly applicable

Author (year)	Title	Study details	Quality assessment
		Oropharynx (%) • 55.2 **Nasopharynx (%)** • 6.4 **Hypopharynx (%)** • 11 **Larynx (%)** • 16.8 **HPV +ve (%)** • 53.6 **Oral cavity (%)** • 6.4 **Inclusion criteria** • Histologically confirmed SCC of head and neck • 18y + • Clinical and radiologic stage N2 N3 nodal metastases	

Author (year)	Title	Study details	Quality assessment
		• Primary treatment of CRT **Exclusion criteria** • Distant metastases • Nonsquamous cell histology • Upfront inoperable neck disease • Inability to undergo neck dissection • History of another malignancy • Concurrent second primary tumour requiring systemic treatment • Poorly controlled diabetes or serious concomitant illness precluding CCRT. **Index test(s)** • PET-CT **PET-CT procedure** • Performed according to the European Association of Nuclear Medicine procedure guideline. • Performed 12 weeks after CCRT completion • For PET acquisitions, Hopkins criteria used, compares lesion uptake to internal jugular vein and liver as background blood pool reference. 1-3 regarded benign uptake, 4 and 5 regarded as	

Author (year)	Title	Study details	Quality assessment
		malignant. **Image interpretation** • Locally assessed by qualified nuclear medicine physician using a 5-point scale based on the surrounding background blood of internal jugular vein and liver. 1-2 were negative and 3-5 positive. **Reference standard(s)** • Histopathologic data as a reference standard were available for all patients who had had positive FDG-PET/CT or clinical suspicion of residual disease indicating a neck dissection and/or biopsies from the primary tumour area. • Otherwise residual disease was confirmed with two imaging modalities.	

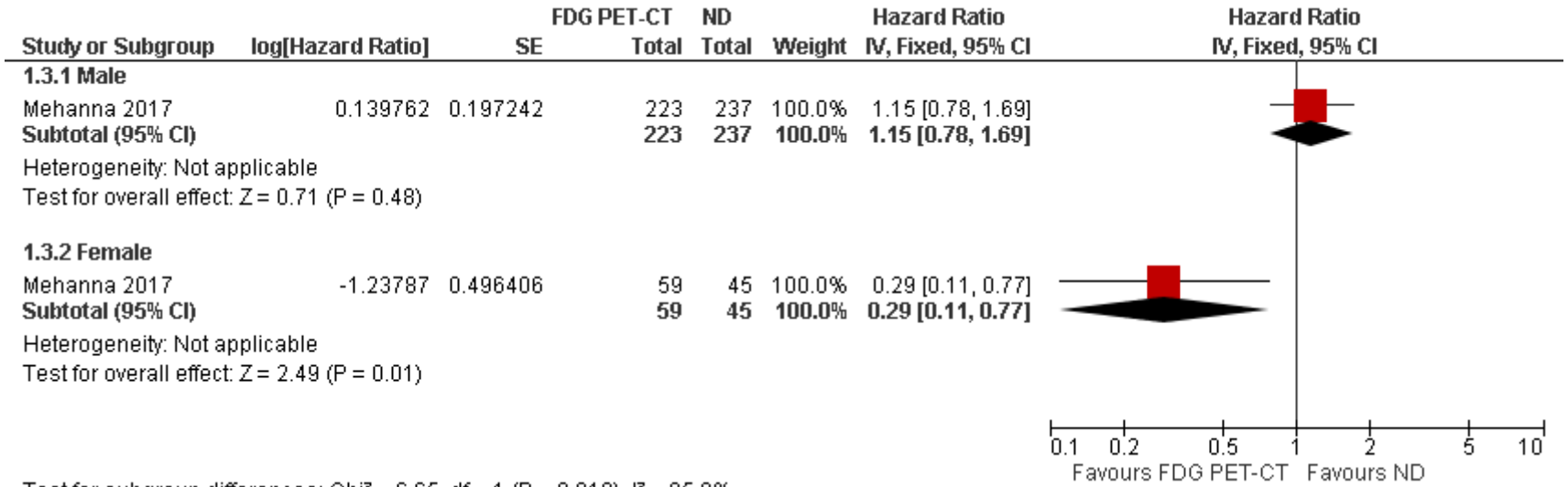
1

Appendix F – Forest plots

Management of nodal metastasis in head and neck cancer after chemoradiotherapy

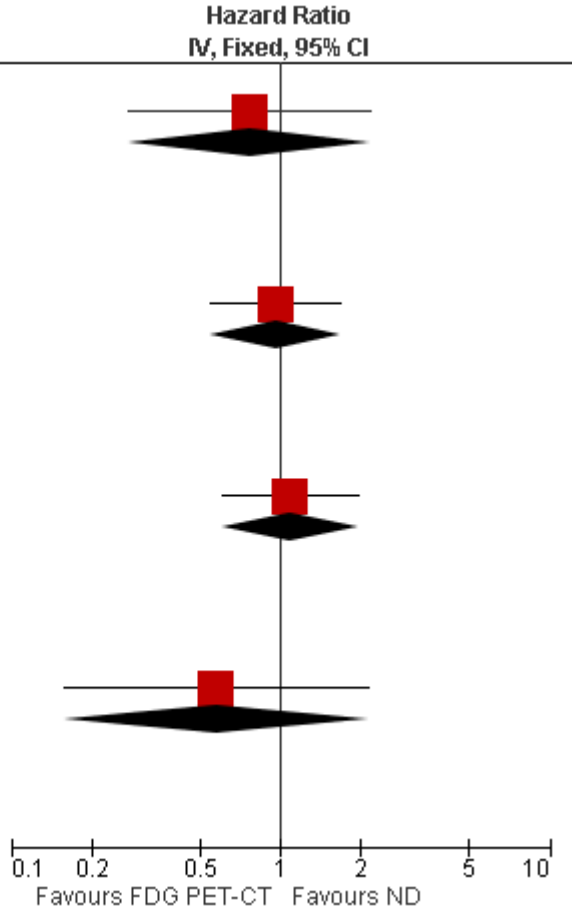
PET–CT-guided active surveillance (FDG PET-CT) compared to planned neck dissection (ND)

Outcome: overall mortality (number of deaths) over 36 months



1 ***Outcome: overall mortality (number of deaths) over 36 months***

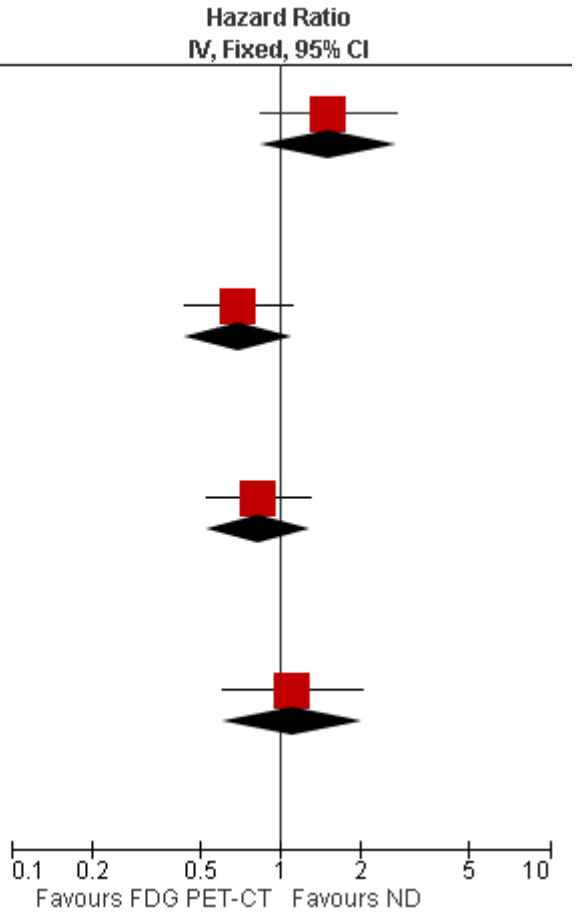
Study or Subgroup	log[Hazard Ratio]	SE	FDG PET-CT Total	ND Total	Weight	Hazard Ratio IV, Fixed, 95% CI
1.4.3 Age <50 years						
Mehanna 2017	-0.26136	0.529286	37	43	100.0%	0.77 [0.27, 2.17]
Subtotal (95% CI)			**37**	**43**	**100.0%**	**0.77 [0.27, 2.17]**
Heterogeneity: Not applicable						
Test for overall effect: Z = 0.49 (P = 0.62)						
1.4.4 Age 50 to 59 years						
Mehanna 2017	-0.04082	0.284855	135	113	100.0%	0.96 [0.55, 1.68]
Subtotal (95% CI)			**135**	**113**	**100.0%**	**0.96 [0.55, 1.68]**
Heterogeneity: Not applicable						
Test for overall effect: Z = 0.14 (P = 0.89)						
1.4.5 Age <60 to 69 years						
Mehanna 2017	0.086178	0.296461	90	107	100.0%	1.09 [0.61, 1.95]
Subtotal (95% CI)			**90**	**107**	**100.0%**	**1.09 [0.61, 1.95]**
Heterogeneity: Not applicable						
Test for overall effect: Z = 0.29 (P = 0.77)						
1.4.6 Age ≥70 years						
Mehanna 2017	-0.54473	0.661578	20	19	100.0%	0.58 [0.16, 2.12]
Subtotal (95% CI)			**20**	**19**	**100.0%**	**0.58 [0.16, 2.12]**
Heterogeneity: Not applicable						
Test for overall effect: Z = 0.82 (P = 0.41)						



2 Test for subgroup differences: Chi² = 0.93, df = 3 (P = 0.82), I² = 0%

1 ***Outcome: overall mortality (number of deaths) over 36 months***

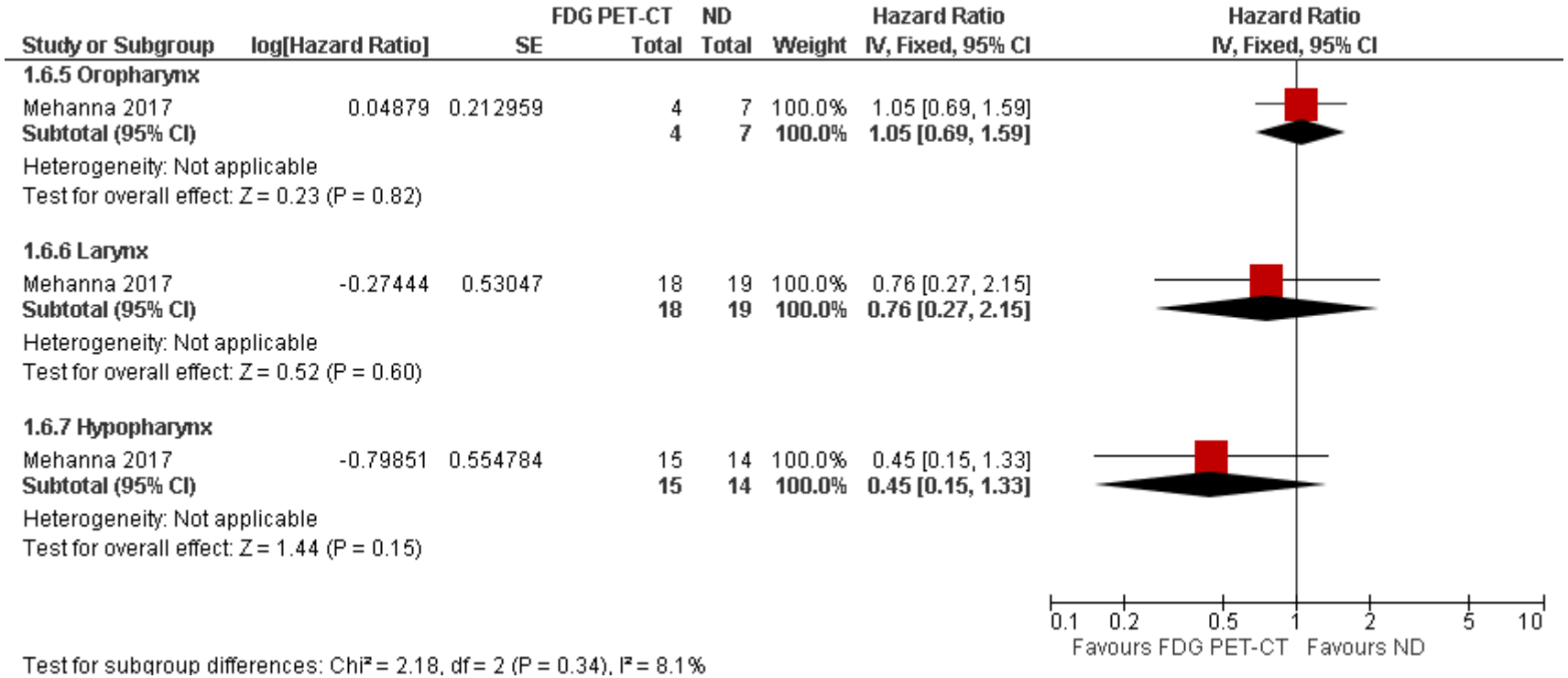
Study or Subgroup	log[Hazard Ratio]	SE	FDG PET-CT Total	ND Total	Weight	Hazard Ratio IV, Fixed, 95% CI
1.5.7 T1 or T3 stage						
Mehanna 2017	0.405465	0.295962	162	160	100.0%	1.50 [0.84, 2.68]
Subtotal (95% CI)			**162**	**160**	**100.0%**	**1.50 [0.84, 2.68]**
Heterogeneity: Not applicable						
Test for overall effect: Z = 1.37 (P = 0.17)						
1.5.8 T3 or T4 stage						
Mehanna 2017	-0.35667	0.236056	116	116	100.0%	0.70 [0.44, 1.11]
Subtotal (95% CI)			**116**	**116**	**100.0%**	**0.70 [0.44, 1.11]**
Heterogeneity: Not applicable						
Test for overall effect: Z = 1.51 (P = 0.13)						
1.5.9 N2a or N2b stage						
Mehanna 2017	-0.18633	0.226918	221	222	100.0%	0.83 [0.53, 1.29]
Subtotal (95% CI)			**221**	**222**	**100.0%**	**0.83 [0.53, 1.29]**
Heterogeneity: Not applicable						
Test for overall effect: Z = 0.82 (P = 0.41)						
1.5.10 N2c or N3 stage						
Mehanna 2017	0.10436	0.304192	61	60	100.0%	1.11 [0.61, 2.01]
Subtotal (95% CI)			**61**	**60**	**100.0%**	**1.11 [0.61, 2.01]**
Heterogeneity: Not applicable						
Test for overall effect: Z = 0.34 (P = 0.73)						



2 Test for subgroup differences: Chi² = 4.65, df = 3 (P = 0.20), I² = 35.4%

3 There were not enough events for occult stage. Therefore, hazard ratios could not be calculated for this subgroup.

1 ***Outcome: overall mortality (number of deaths) over 36 months***



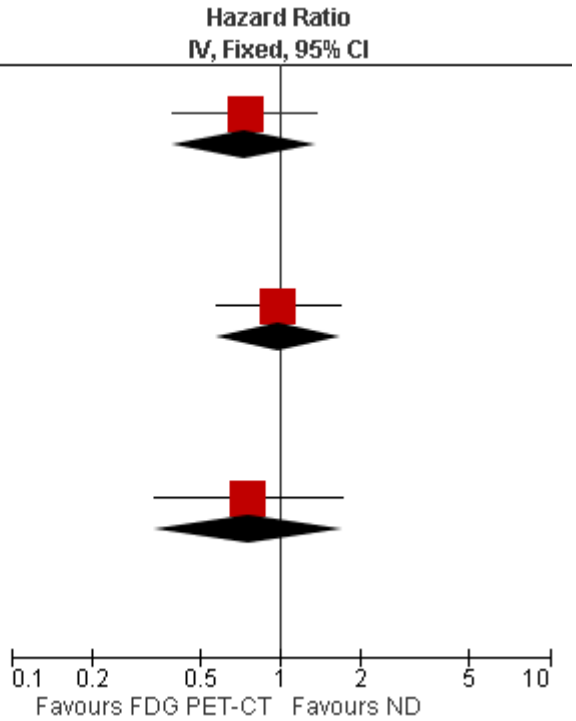
2 Test for subgroup differences: $Chi^2 = 2.18$, df = 2 (P = 0.34), $I^2 = 8.1\%$

3 There were not enough events for oral cavity and occult disease. Therefore, hazard ratios could not be calculated for any of these subgroups.

4

1 ***Outcome: overall mortality (number of deaths) over 36 months***

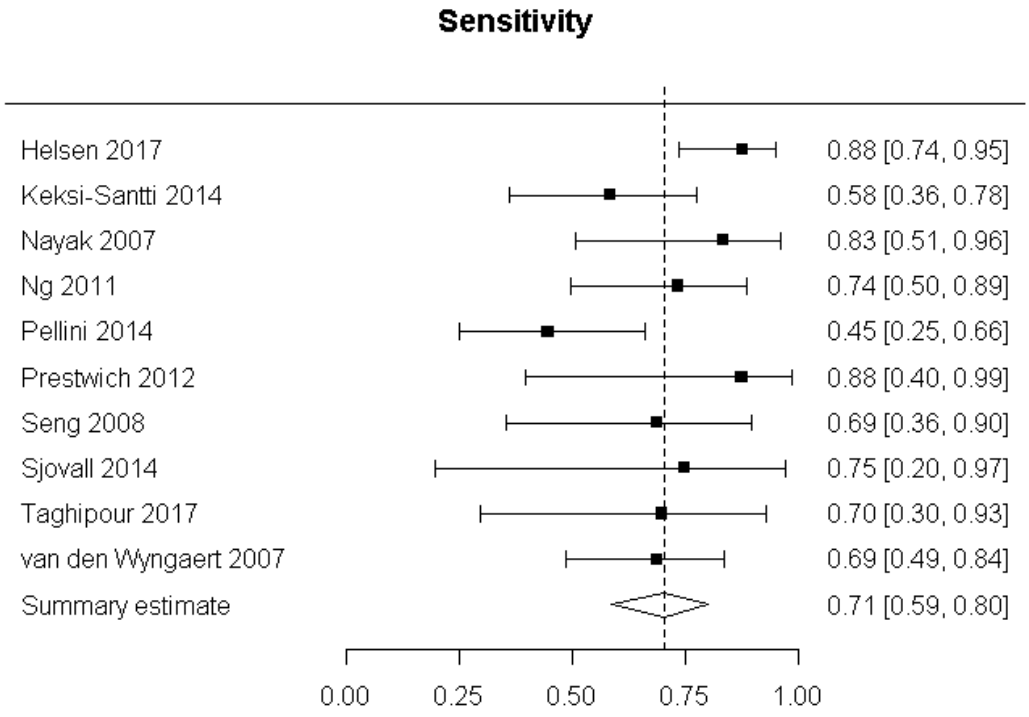
Study or Subgroup	log[Hazard Ratio]	SE	FDG PET-CT Total	ND Total	Weight	Hazard Ratio IV, Fixed, 95% CI
1.7.8 p16 positive						
Mehanna 2017	-0.30111	0.314056	164	171	100.0%	0.74 [0.40, 1.37]
Subtotal (95% CI)			**164**	**171**	**100.0%**	**0.74 [0.40, 1.37]**
Heterogeneity: Not applicable						
Test for overall effect: Z = 0.96 (P = 0.34)						
1.7.9 p16 negative						
Mehanna 2017	-0.0202	0.268251	62	49	100.0%	0.98 [0.58, 1.66]
Subtotal (95% CI)			**62**	**49**	**100.0%**	**0.98 [0.58, 1.66]**
Heterogeneity: Not applicable						
Test for overall effect: Z = 0.08 (P = 0.94)						
1.7.10 p16 status not known						
Mehanna 2017	-0.27444	0.410571	56	62	100.0%	0.76 [0.34, 1.70]
Subtotal (95% CI)			**56**	**62**	**100.0%**	**0.76 [0.34, 1.70]**
Heterogeneity: Not applicable						
Test for overall effect: Z = 0.67 (P = 0.50)						

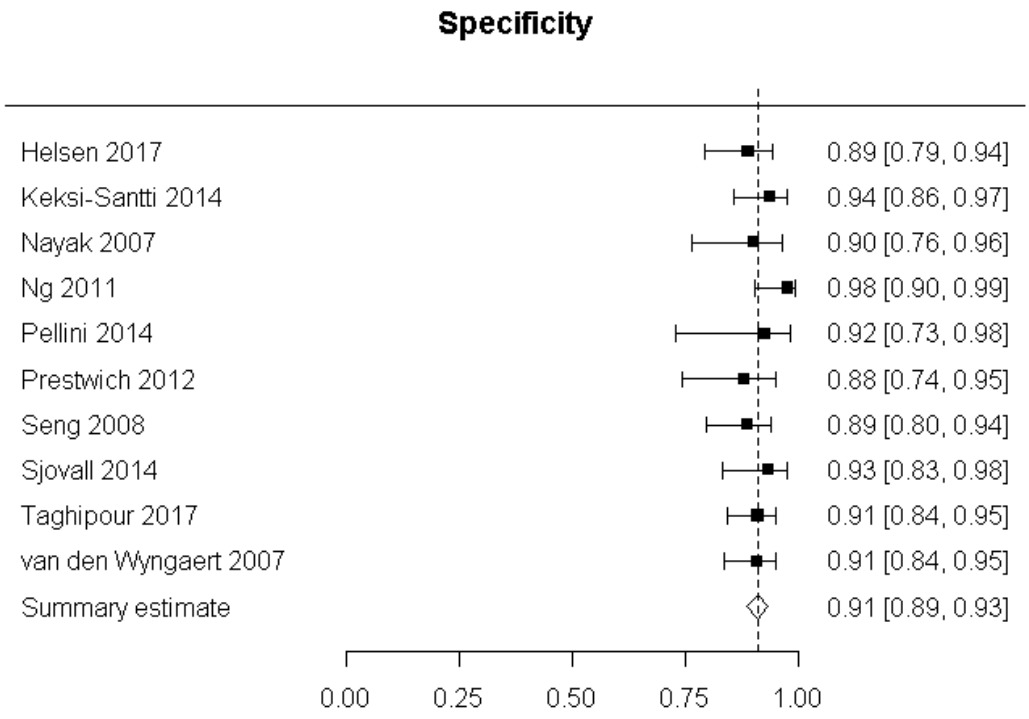


2 Test for subgroup differences: Chi² = 0.55, df = 2 (P = 0.76), I² = 0%

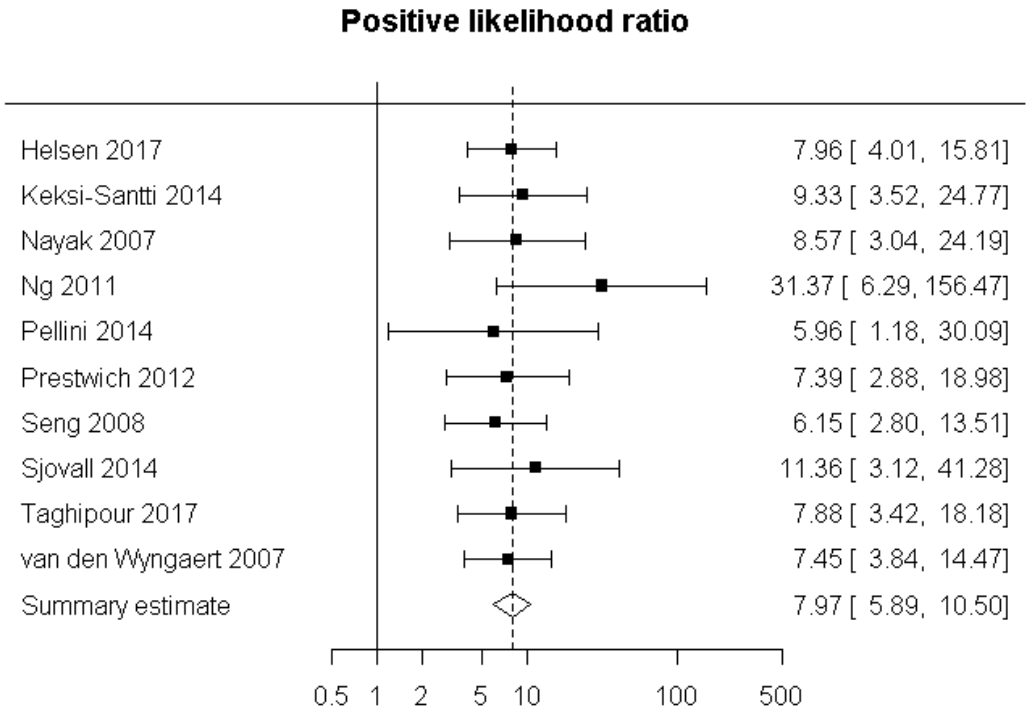
Diagnostic accuracy of PET-CT to diagnose residual nodal disease after radiotherapy or chemoradiotherapy

All studies



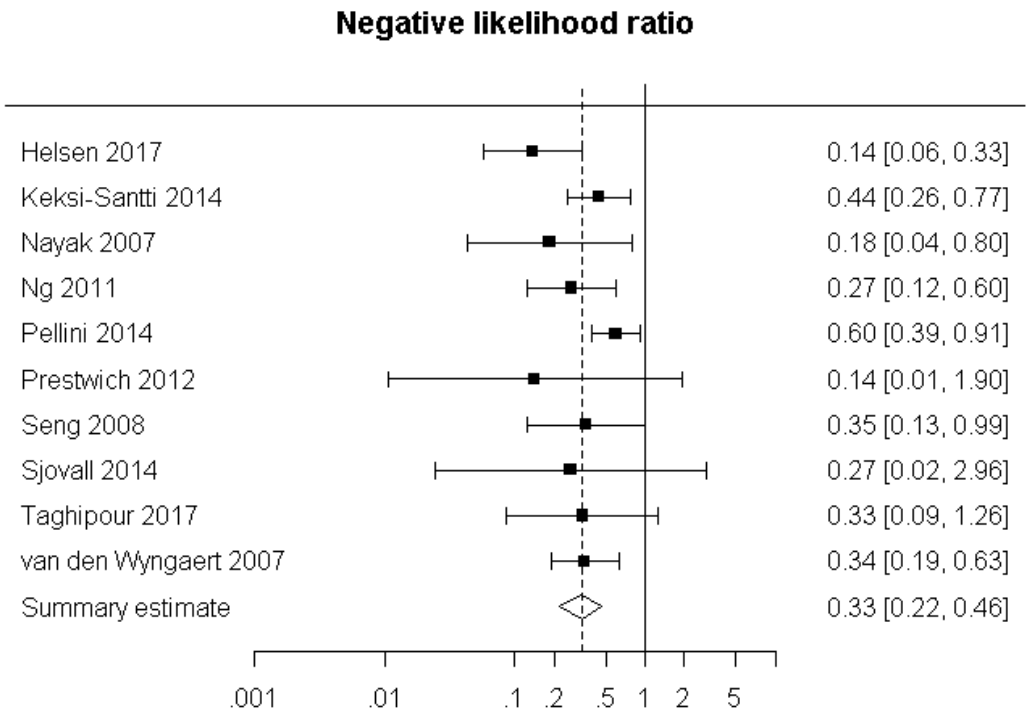

Specificity
Helsen 2017
Keksi-Santti 2014
Nayak 2007
Ng 2011
Pellini 2014
Prestwich 2012
Seng 2008
Sjovall 2014
Taghipour 2017
van den Wyngaert 2007
Summary estimate
0.89 [0.79, 0.94]
0.94 [0.86, 0.97]
0.90 [0.76, 0.96]
0.98 [0.90, 0.99]
0.92 [0.73, 0.98]
0.88 [0.74, 0.95]
0.89 [0.80, 0.94]
0.93 [0.83, 0.98]
0.91 [0.84, 0.95]
0.91 [0.84, 0.95]
0.91 [0.89, 0.93]
0.00
0.25
0.50
0.75
1.00

1


Positive likelihood ratio
Helsen 2017
Keksi-Santti 2014
Nayak 2007
Ng 2011
Pellini 2014
Prestwich 2012
Seng 2008
Sjovall 2014
Taghipour 2017
van den Wyngaert 2007
Summary estimate
7.96 [4.01, 15.81]
9.33 [3.52, 24.77]
8.57 [3.04, 24.19]
31.37 [6.29, 156.47]
5.96 [1.18, 30.09]
7.39 [2.88, 18.98]
6.15 [2.80, 13.51]
11.36 [3.12, 41.28]
7.88 [3.42, 18.18]
7.45 [3.84, 14.47]
7.97 [5.89, 10.50]
0.5
1
2
5
10
100
500

2

3

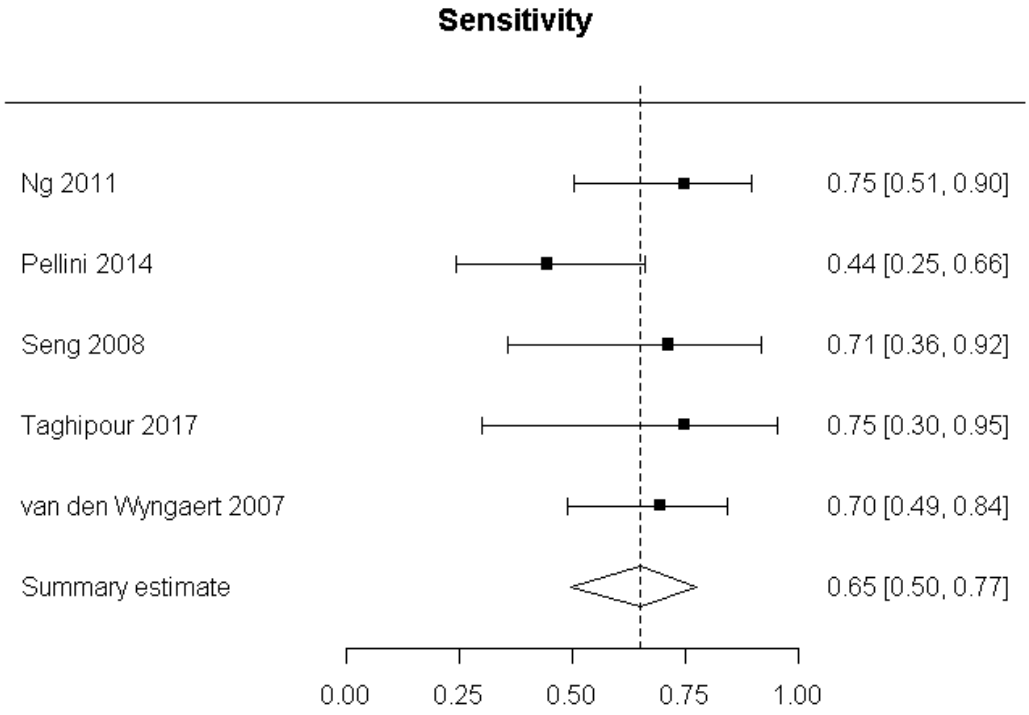

Negative likelihood ratio
Helsen 2017
Keksi-Santti 2014
Nayak 2007
Ng 2011
Pellini 2014
Prestwich 2012
Seng 2008
Sjovall 2014
Taghipour 2017
van den Wyngaert 2007
Summary estimate
0.14 [0.06, 0.33]
0.44 [0.26, 0.77]
0.18 [0.04, 0.80]
0.27 [0.12, 0.60]
0.60 [0.39, 0.91]
0.14 [0.01, 1.90]
0.35 [0.13, 0.99]
0.27 [0.02, 2.96]
0.33 [0.09, 1.26]
0.34 [0.19, 0.63]
0.33 [0.22, 0.46]
.001
.01
.1
.2
.5
1
2
5

1

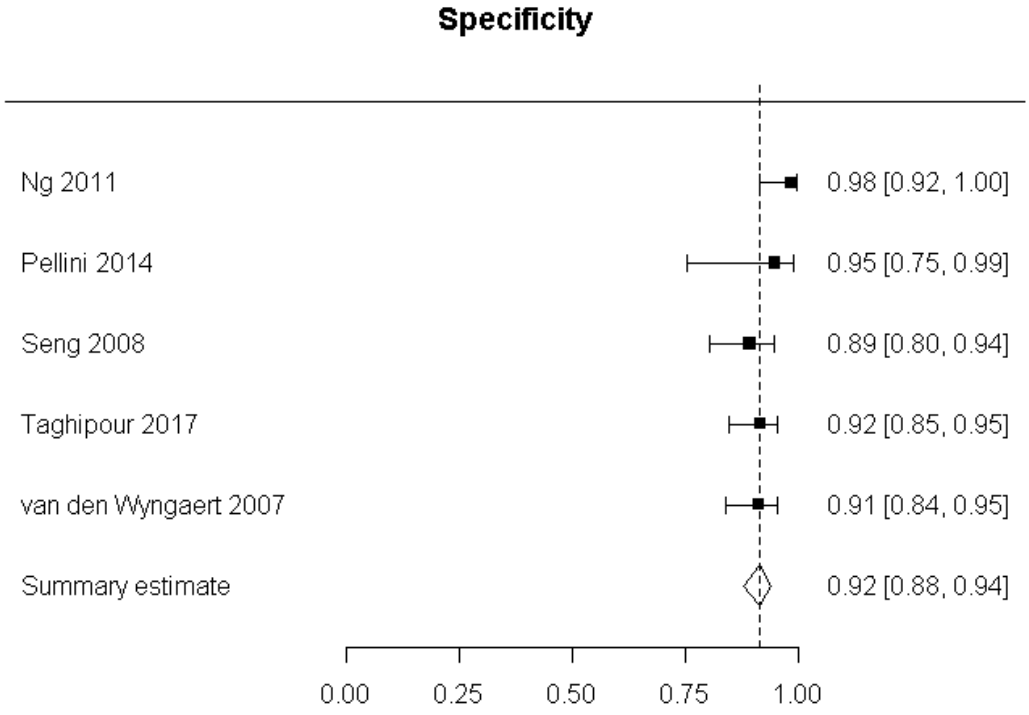
2

3

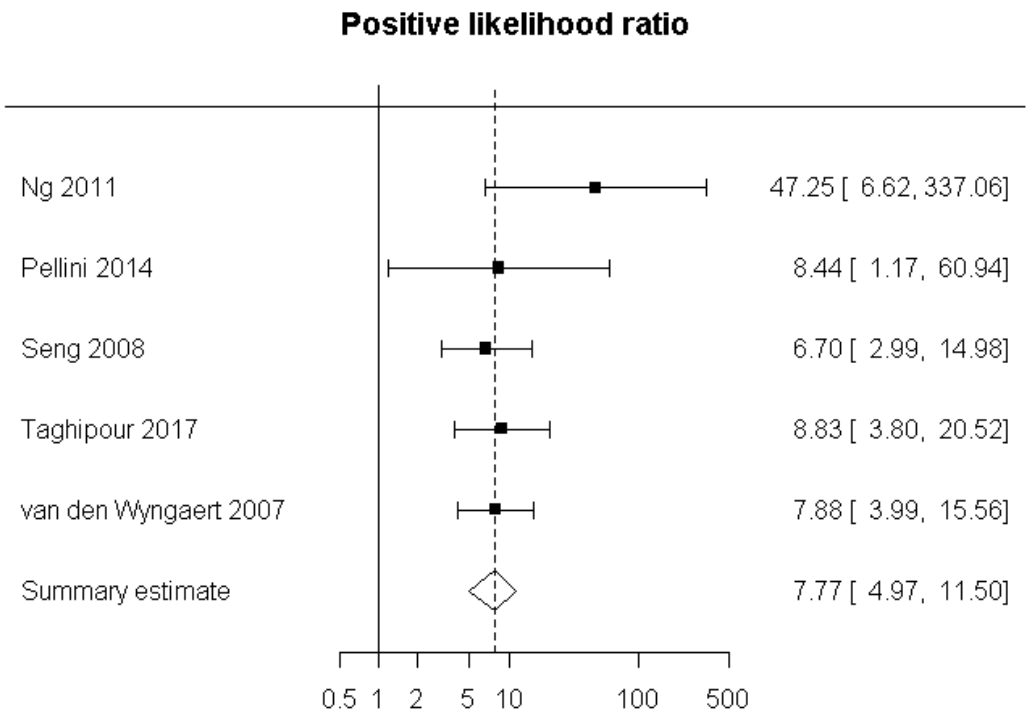
1 **Prospective studies only**

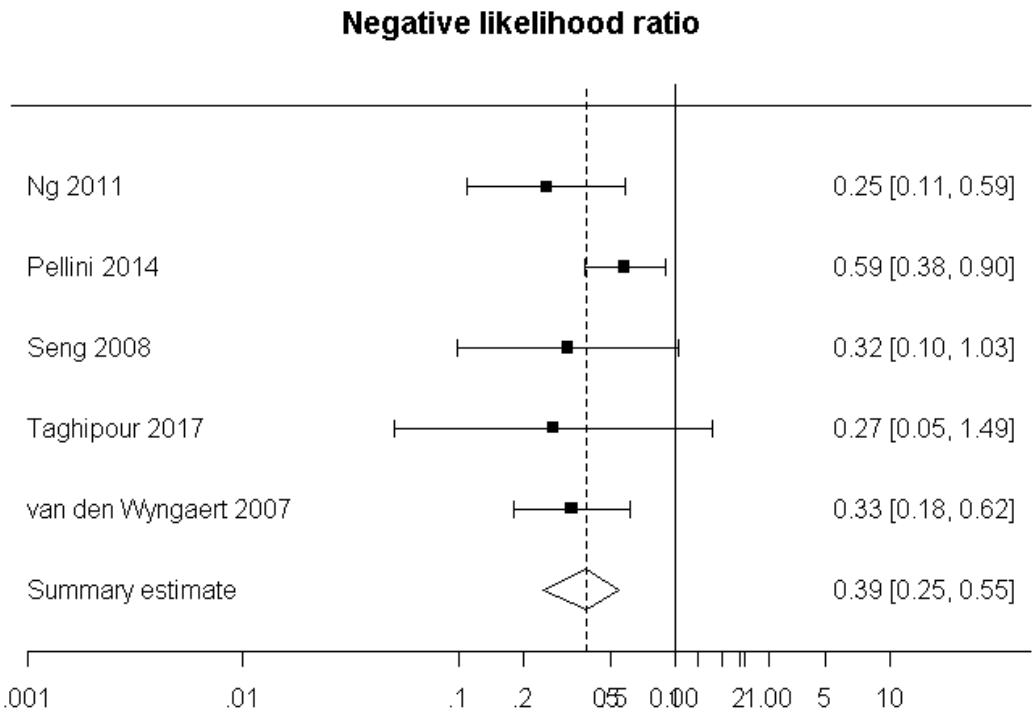


2



3


Positive likelihood ratio
Ng 2011
47.25 [6.62, 337.06]
Pellini 2014
8.44 [1.17, 60.94]
Seng 2008
6.70 [2.99, 14.98]
Taghipour 2017
8.83 [3.80, 20.52]
van den Wyngaert 2007
7.88 [3.99, 15.56]
Summary estimate
7.77 [4.97, 11.50]
0.5 1 2 5 10 100 500


Negative likelihood ratio
Ng 2011
0.25 [0.11, 0.59]
Pellini 2014
0.59 [0.38, 0.90]
Seng 2008
0.32 [0.10, 1.03]
Taghipour 2017
0.27 [0.05, 1.49]
van den Wyngaert 2007
0.33 [0.18, 0.62]
Summary estimate
0.39 [0.25, 0.55]
.001
.01
.1
.2
.5
1
2
5
10

Appendix G – Meta-regression

Diagnosing residual nodal disease in head and neck after radiotherapy or chemoradiotherapy

To estimate the diagnostic accuracy of PET-CT at different sites, random effects meta-regression with study level subgroups was performed. This was done as too few studies reported results for subgroups to compare results between individual studies, and there was significant heterogeneity in the results.

Single sites were analysed where more than 5 studies reported what proportion of the cohort had the site as the primary site of disease. An exception was made with respect to moderate risk of bias to assess the effect of bias on outcomes. Bivariate meta-regression was carried out for single sites that had a significantly different fit to the all sites model ($p < 0.2$; high risk of bias, $p = 0.115$; moderate risk of bias, $p = 0.052$). Low risk of bias was also included in the bivariate analyses as bias to assess the effect of bias ($p = 0.716$). Oropharynx ($p = 0.381$) was also included to assess if the effect of lowering sensitivity (59.6%, 95% CI 37.9 - 78.1) is associated with oropharynx as a primary site or moderate risk of bias because both oropharynx only papers have a moderate risk of bias. Meta-regression was not performed for oral cavity, unknown primary site, cancer staging, HPV status or nodal site by stage as there was not a sufficient number of studies reporting on these characteristics.

The effect on diagnostic accuracy by adding the subgroups on the all sites, all studies model is presented as sensitivity and specificity. The Akaike information criterion (AIC) was used to estimate the quality of the model relative to the all sites, all studies model (AIC for model without covariates = -41.616).

The analysis was carried out in RStudio Version 1.0.143 with imported diagnostic data from Excel 2013. The following code was used for the univariate analysis without the site in question:

```
library(readxl)
data = read_excel("test").xlsx")
model = madad(data)
model
model2 = reitsma(data,formula=cbind(tsens, tfpr) ~ 1)
summary(model2)
model3 = reitsma(data,formula=cbind(tsens, tfpr) ~ data$Oropharynx)
summary(model3)
```

and for univariate analysis with the site in question:

```
NotOro=1-data$Oropharynx
model2 = reitsma(data,formula=cbind(tsens, tfpr) ~ 1)
summary(model2)
model3 = reitsma(data,formula=cbind(tsens, tfpr) ~ NotOro)
summary(model3)
```

Effects of covariates in a bivariate model was performed. A model without two covariates was constructed and the proportion of patients without disease at site A, e.g. oropharynx, and the proportion of patients at site B, e.g. nasopharynx, were added to produce data on the model without site A. The reverse was done to obtain the results for site B. This method was preferred to entering the proportion of sites for both covariates into the same model as it produces values for the confidence intervals that do not represent the 95% specified.

A bivariate model without both oropharynx and nasopharynx was achieved by the following:

```
library(readxl)
data = read_excel("test).xlsx")
model = madad(data)
model
NotOro=1-data$Oropharynx
NotNaso=1-data$Nasopharynx
model2 = reitsma(data,formula=cbind(tsens, tfpr) ~ 1)
summary(model2)
model3 = reitsma(data,formula=cbind(tsens, tfpr) ~ data$Oropharynx+Nasopharynx)
summary(model3)
```

For the effect of oropharynx and not nasopharynx:

```
library(readxl)
NotOro=1-data$Oropharynx
NotNaso=1-data$Nasopharynx
model2 = reitsma(data,formula=cbind(tsens, tfpr) ~ 1)
summary(model2)
model3 = reitsma(data,formula=cbind(tsens, tfpr) ~ data$Nasopharynx+NotOro)
summary(model3)
```

For the effect of nasopharynx and not oropharynx:

```
library(readxl)
data = read_excel("test).xlsx")
model = madad(data)
model
NotOro=1-data$Oropharynx
NotNaso=1-data$Nasopharynx
model2 = reitsma(data,formula=cbind(tsens, tfpr) ~ 1)
summary(model2)
```

```
model3 = reitsma(data,formula=cbind(tsens, tfpr) ~ data$Oropharynx+NotNaso)
summary(model3)
```

The coefficients were transformed to produce the sensitivity and false positive rate for models without covariates by:

=EXP(intercept)/(EXP(intercept)+1)

Outputs created values for sensitivity and false positive rate. Therefore, specificity was calculated by:

Specificity = 1 – false positive rate

AIC values represent the quality of the model's fit and was used to assess the effect of the covariates on the model. A lower number denotes a better model fit.

Table 9 Univariate and bivariate meta-regression of primary sites

Site(s)	Sensitivity without effect of site(s) (95% CI)	Effect on sensitivity (95% CI)	Specificity without effect of site(s) (95% CI)	Effect on specificity (95% CI)	AIC
Oropharyngeal					
	84.3 (57.2, 52.5)	59.6 (37.9, 78.1)	92.0 (81.4, 96.7)	90.4 (5.8, 85.6)	-42.112
Hypopharyngeal					
	68.2 (50.6, 81.8)	70.8 (22.8, 98.4)	89.4 (84.9, 92.7)	98.2 (83.0, 99.8)	-40.441
Larynx					
	66.0 (47.4, 80.7)	91.5 (16.8, 99.8)	91.6 (87.5 94.4)	86.4 (36.3, 98.6)	-44.248
Nasopharynx					
	70.1 (52.4, 83.4)	69.5 (53.4, 81.8)	91.0 (87.5, 93.7)	91.1 (87.5, 93.6)	-40.510
Low risk of bias (LRoB)					
	69.3 (56.4, 79.8)	58.5 (30.7, 81.8)	91.0 (88.1, 93.2)	92.4 (87.0, 95.7)	-40.056
Moderate risk of bias (MRoB)					
	80.7 (69.9, 88.3)	58.8 (46.4, 70.2)	90.5 (86.1, 93.3)	91.6 (87.9, 94.3)	-42.225
High risk of bias (HRoB)					
	61.4 (49.7, 72.0)	80.5 (68.7, 88.7)	91.7 (88.5, 94.1)	89.9 (85.1, 93.3)	-40.732
Oropharyngeal (Oro) & LRoB					
	86.9 (63.6, 96.1)	Oro 52.4 (30.5, 73.4)	92.3 (90.1, 97.0)	Oro 90.3 (83.9, 94.3)	-42.018
		LRoB 95.7 (65.4, 99.6)		LRoB 93.5 (85.5, 98.4)	
Oropharyngeal (Oro) & MRoB					
	87.4 (70.1, 95.3)	Oro 73.3 (51.2, 87.8)	91.5 (81.0, 96.5)	Oro 89.9 (83.3, 94.0)	-41.985
		MRoB 72.9 (43.5, 90.4)		MRoB 92.7 (82.1, 97.2)	
Oropharyngeal (Oro) & HRoB					
	73.6 (41.7, 91.6)	Oro 55.9 (38.5, 72.0)	93.7 (83.2, 97.8)	Oro 91.0 (86.2, 94.3)	-40.561
		HRoB 85.8 (67.5, 94.6)		MRoB 91.8 (81.7, 96.6)	
HRoB & MRoB					

Site(s)	Sensitivity without effect of site(s) (95% CI)	Effect on sensitivity (95% CI)	Specificity without effect of site(s) (95% CI)	Effect on specificity (95% CI)	AIC
	81.6 (48.9, 95.4)	HRoB 80.6 (68.7, 88.7)	92.0 (84.1, 96.1)	HRoB 89.9 (85.1, 93.3)	
		MRoB 58.8 (46.4, 70.2)		MRoB 91.6 (87.9, 94.3)	

AIC: Akaike Information Criterion; CI: confidence interval; HRoB: high risk of bias; LRoB: low risk of bias; MRoB: moderate risk of bias; Oro: oropharyngeal.

1

Appendix H – GRADE tables

Management of nodal metastasis in head and neck cancer after chemoradiotherapy

PET–CT-guided active surveillance (FDG PET-CT) compared to planned neck dissection (ND)

Outcome: recurrence rates

No. of studies	Study design	Sample size	Effect size (95% CI)	Absolute risk: ND	Absolute risk: FDG PET-CT (95% CI)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
Recurrence rates – disease was apparent >3 months after radiotherapy (lower values favour FDG PET-CT-guided active surveillance)										
1 (Mehanna 2017)	RCT	564	RR 1.07 (0.73, 1.58)[1]	148 per 1,000	158 per 1,000 (108, 234)	Not serious	N/A	Not serious	Very serious[2]	Low

1. Mehanna 2017 did not report hazard ratios for recurrence. Only raw data was provided to calculate relative risk
2. 95% confidence interval crosses both ends of a defined MID interval

CI: confidence interval; HR: hazard ratio; ND: neck dissection; N/A: not applicable; RCT: randomised controlled trial.

Outcome: overall mortality

No. of studies	Study design	Sample size	Effect size (95% CI)	Absolute risk: ND	Absolute risk: FDG PET-CT (95% CI)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
Overall mortality – number of deaths (lower values favour FDG PET-CT-guided active surveillance)										
1 (Mehanna 2017)	RCT	564	HR 0.92 (0.65, 1.32)	219 per 1,000	204 per 1,000 (149, 279)	Not serious	N/A	Not serious	Very serious[1]	Low
Overall mortality – number of deaths, subgroup by sex: male (lower values favour FDG PET-CT-guided active surveillance)										
1 (Mehanna 2017)	RCT	460	HR 1.15 (0.78, 1.69)	210 per 1,000	238 per 1,000 (168, 329)	Not serious	N/A	Not serious	Very serious[1]	Low
Overall mortality – number of deaths, subgroup by sex: female (lower values favour FDG PET-CT-guided active surveillance)										
1 (Mehanna 2017)	RCT	104	HR 0.29 (0.11, 0.77)	266 per 1,000	86 per 1,000 (33, 212)	Not serious	N/A	Not serious	Not serious	High
Overall mortality – number of deaths, subgroup by age: <50 years (lower values favour FDG PET-CT-guided active surveillance)										

No. of studies	Study design	Sample size	Effect size (95% CI)	Absolute risk: ND	Absolute risk: FDG PET-CT (95% CI)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
1 (Mehanna 2017)	RCT	80	HR 0.77 (0.27, 2.15)	209 per 1,000	165 per 1,000 (61, 396)	Not serious	N/A	Not serious	Very serious[1]	Low
Overall mortality – number of deaths, subgroup by age: 50 to 59 years (lower values favour FDG PET-CT-guided active surveillance)										
1 (Mehanna 2017)	RCT	248	HR 0.96 (0.55, 1.68)	203 per 1,000	196 per 1,000 (117, 317)	Not serious	N/A	Not serious	Very serious[1]	Low
Overall mortality – number of deaths, subgroup by age: <60 to 69 years (lower values favour FDG PET-CT-guided active surveillance)										
1 (Mehanna 2017)	RCT	197	HR 1.09 (0.61, 1.95)	224 per 1,000	241 per 1,000 (143, 390)	Not serious	N/A	Not serious	Very serious[1]	Low
Overall mortality – number of deaths, subgroup by age: ≥70 years (lower values favour FDG PET-CT-guided active surveillance)										
1 (Mehanna 2017)	RCT	39	HR 0.58 (0.16, 2.14)	315 per 1,000	197 per 1,000 (58, 556)	Not serious	N/A	Not serious	Very serious[1]	Low
Overall mortality – number of deaths, subgroup by tumour stage: T1 or T2 (lower values favour FDG PET-CT-guided active surveillance)										
1 (Mehanna 2017)	RCT	322	HR 1.50 (0.84, 2.68)	112 per 1,000	163 per 1,000 (95, 273)	Not serious	N/A	Not serious	Serious[2]	Moderate
Overall mortality – number of deaths, subgroup by tumour stage: T3 or T4 (lower values favour FDG PET-CT-guided active surveillance)										
1 (Mehanna 2017)[3]	RCT	232	HR 0.70 (0.44, 1.11)	362 per 1,000	269 per 1,000 (179, 392)	Not serious	N/A	Not serious	Serious[2]	Moderate
Overall mortality – number of deaths, subgroup by nodal stage: N2a or N2b (lower values favour FDG PET-CT-guided active surveillance)										
1 (Mehanna 2017)[3]	RCT	443	HR 0.83 (0.53, 1.29)	189 per 1,000	159 per 1,000 (105, 237)	Not serious	N/A	Not serious	Very serious[1]	Low
Overall mortality – number of deaths, subgroup by nodal stage: N2c or N3 (lower values favour FDG PET-CT-guided active surveillance)										
1 (Mehanna 2017)	RCT	121	HR 1.11 (0.61, 2.01)	333 per 1,000	362 per 1,000 (219, 557)	Not serious	N/A	Not serious	Very serious[1]	Low
Overall mortality – number of deaths, subgroup by cancer site: oral cavity (lower values favour FDG PET-CT-guided active surveillance)										
1 (Mehanna 2017)	RCT	11	RR 1.75 (0.15, 21.0)[4]	142 per 1,000	236 per 1,000 (22, 960)	Not serious	N/A	Not serious	Very serious[1]	Low
Overall mortality – number of deaths, subgroup by cancer site: oropharynx (lower values favour FDG PET-CT-guided active surveillance)										
1 (Mehanna 2017)	RCT	476	HR 1.05 (0.69, 1.59)	177 per 1,000	185 per 1,000 (126, 267)	Not serious	N/A	Not serious	Very serious[1]	Low

No. of studies	Study design	Sample size	Effect size (95% CI)	Absolute risk: ND	Absolute risk: FDG PET-CT (95% CI)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
Overall mortality – number of deaths, subgroup by cancer site: larynx (lower values favour FDG PET-CT-guided active surveillance)										
1 (Mehanna 2017)	RCT	37	HR 0.76 (0.27, 2.16)	421 per 1,000	339 per 1,000 (137, 692)	Not serious	N/A	Not serious	Very serious[1]	Low
Overall mortality – number of deaths, subgroup by cancer site: hypopharynx (lower values favour FDG PET-CT-guided active surveillance)										
1 (Mehanna 2017)	RCT	29	HR 0.45 (0.15, 1.32)	571 per 1,000	317 per 1,000 (119, 673)	Not serious	N/A	Not serious	Very serious[1]	Low
Overall mortality – number of deaths, subgroup by HPV status: p16 positive (lower values favour FDG PET-CT-guided active surveillance)										
1 (Mehanna 2017)	RCT	335	HR 0.74 (0.40, 1.37)	134 per 1,000	101 per 1,000 (56, 179)	Not serious	N/A	Not serious	Very serious[1]	Low
Overall mortality – number of deaths, subgroup by HPV status: p16 negative (lower values favour FDG PET-CT-guided active surveillance)										
1 (Mehanna 2017)	RCT	111	HR 0.98 (0.58, 1.66)	510 per 1,000	503 per 1,000 (338, 694)	Not serious	N/A	Not serious	Very serious[1]	Low
Overall mortality – number of deaths, subgroup by HPV status: p16 status not known (lower values favour FDG PET-CT-guided active surveillance)										
1 (Mehanna 2017)	RCT	118	HR 0.76 (0.34, 1.70)	225 per 1,000	176 per 1,000 (83, 352)	Not serious	N/A	Not serious	Very serious[1]	Low

1. 95% confidence interval crosses both ends of a defined MID interval
2. 95% confidence interval crosses one end of a defined MID interval
3. Data was taken from Mehanna 2016
4. Relative risks were calculated based on raw data from Mehanna 2017

CI: confidence interval; HR: hazard ratio; ND: neck dissection; N/A: not applicable; RCT: randomised controlled trial.

1 ***Outcome: complications***

No. of studies	Study design	Sample size	Effect size (95% CI)	Absolute risk: ND	Absolute risk: FDG PET-CT (95% CI)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
Complications – number of patients with complications following neck dissection surgery (lower values favour FDG PET-CT-guided active surveillance) at the follow-up visit at 2 weeks post neck dissection surgery										
1 (Mehanna 2017)	RCT	564	RR 0.27 (0.17, 0.41)[1]	294 per 1,000	79 per 1,000 (50, 121)	Not serious	N/A	Not serious	Not serious	High

No. of studies	Study design	Sample size	Effect size (95% CI)	Absolute risk: ND	Absolute risk: FDG PET-CT (95% CI)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
1. Relative risks were calculated based on raw data from Mehanna 2017 CI: confidence interval; HR: hazard ratio; ND: neck dissection; N/A: not applicable; RCT: randomised controlled trial.										

1 ***Outcome: serious adverse events (SAEs)***

No. of studies	Study design	Sample size	Effect size (95% CI)	Absolute risk: ND	Absolute risk: FDG PET-CT (95% CI)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
SAEs – number of patients with at least 1 serious adverse event (lower values favour FDG PET-CT-guided active surveillance) at 2 years follow-up										
1 (Mehanna 2017)	RCT	564	RR 0.67 (0.56, 0.79)[1]	599 per 1,000	401 per 1,000 (335, 473)	Not serious	N/A	Not serious	Not serious	High
1. Relative risks were calculated based on raw data from Mehanna 2017 CI: confidence interval; HR: hazard ratio; ND: neck dissection; N/A: not applicable; RCT: randomised controlled trial; SAEs: serious adverse events.										

2 ***Outcome: quality of life***

No. of studies	Study design	Sample size	Effect size (95% CI)	Absolute risk: ND	Absolute risk: FDG PET-CT (95% CI)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
Quality of life – overall scores										
EORTC's QLQ-C30 global health status (positive difference favour FDG PET-CT-guided active surveillance) at 2 years follow-up										
1 (Mehanna 2017)	RCT	346[1]	MD[2] -0.81	N/A	N/A	Not serious	N/A	Not serious	Very serious[3]	Low
EQ-5D overall health status (positive difference favour FDG PET-CT-guided active surveillance) at 2 years follow-up										
1 (Mehanna 2017)	RCT	331[1]	MD[2] 0.02	N/A	N/A	Not serious	N/A	Not serious	Very serious[3]	Low
Quality of life – Head & neck specific outcomes										
EORTC's H&N35 swallowing (positive difference favour FDG PET-CT-guided active surveillance) at 2 years follow-up										
1 (Mehanna 2017)	RCT	348[1]	MD[2] -3.08	N/A	N/A	Not serious	N/A	Not serious	Very serious[3]	Low
MDADI dysphagia total (positive difference favour FDG PET-CT-guided active surveillance) at 2 years follow-up										

No. of studies	Study design	Sample size	Effect size (95% CI)	Absolute risk: ND	Absolute risk: FDG PET-CT (95% CI)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
1 (Mehanna 2017)	RCT	333[1]	MD[2] -0.64	N/A	N/A	Not serious	N/A	Not serious	Very serious[3]	Low
EORTC's H&N35 pain/mouth, jaw or throat (positive difference favour FDG PET-CT-guided active surveillance) at 2 years follow-up										
1 (Mehanna 2017)	RCT	348[1]	MD[2] -3.67	N/A	N/A	Not serious	N/A	Not serious	Very serious[3]	Low
EORTC's H&N35 problems with teeth (positive difference favour FDG PET-CT-guided active surveillance) at 2 years follow-up										
1 (Mehanna 2017)	RCT	340[1]	MD[2] -2.9	N/A	N/A	Not serious	N/A	Not serious	Very serious[3]	Low
EORTC's H&N35 problems opening mouth wide (positive difference favour FDG PET-CT-guided active surveillance) at 2 years follow-up										
1 (Mehanna 2017)	RCT	346[1]	MD[2] 5.75	N/A	N/A	Not serious	N/A	Not serious	Very serious[3]	Low
EORTC's H&N35 sticky saliva (positive difference favour FDG PET-CT-guided active surveillance) at 2 years follow-up										
1 (Mehanna 2017)	RCT	345[1]	MD[2] -3.47	N/A	N/A	Not serious	N/A	Not serious	Very serious[3]	Low
EORTC's H&N35 speech problems (positive difference favour FDG PET-CT-guided active surveillance) at 2 years follow-up										
1 (Mehanna 2017)	RCT	346[1]	MD[2] -1.21	N/A	N/A	Not serious	N/A	Not serious	Very serious[3]	Low
Quality of life – Domains										
EORTC's H&N35 trouble with social eating (positive difference favour FDG PET-CT-guided active surveillance) at 2 years follow-up										
1 (Mehanna 2017)	RCT	342[1]	MD[2] 0.23	N/A	N/A	Not serious	N/A	Not serious	Very serious[3]	Low
EORTC's H&N35 trouble with social contact (positive difference favour FDG PET-CT-guided active surveillance) at 2 years follow-up										
1 (Mehanna 2017)	RCT	346[1]	MD[2] 0.12	N/A	N/A	Not serious	N/A	Not serious	Very serious[3]	Low
EORTC's QLQ-C30 social functioning (positive difference favour FDG PET-CT-guided active surveillance) at 2 years follow-up										
1 (Mehanna 2017)	RCT	339[1]	MD[2] -5.57	N/A	N/A	Not serious	N/A	Not serious	Very serious[3]	Low
EORTC's QLQ-C30 physical functioning (positive difference favour FDG PET-CT-guided active surveillance) at 2 years follow-up										

No. of studies	Study design	Sample size	Effect size (95% CI)	Absolute risk: ND	Absolute risk: FDG PET-CT (95% CI)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
1 (Mehanna 2017)	RCT	349[1]	MD[2] -0.9	N/A	N/A	Not serious	N/A	Not serious	Very serious[3]	Low
EORTC's QLQ-C30 role functioning (positive difference favour FDG PET-CT-guided active surveillance) at 2 years follow-up										
1 (Mehanna 2017)	RCT	346[1]	MD[2] -0.71	N/A	N/A	Not serious	N/A	Not serious	Very serious[3]	Low
EORTC's QLQ-C30 emotional functioning (positive difference favour FDG PET-CT-guided active surveillance) at 2 years follow-up										
1 (Mehanna 2017)	RCT	346[1]	MD[2] 0.86	N/A	N/A	Not serious	N/A	Not serious	Very serious[3]	Low
EORTC's QLQ-C30 cognitive functioning (positive difference favour FDG PET-CT-guided active surveillance) at 2 years follow-up										
1 (Mehanna 2017)	RCT	336[1]	MD[2]0.15	N/A	N/A	Not serious	N/A	Not serious	Very serious[3]	Low
Quality of life – Cross-cutting symptoms										
EORTC's H&N35 weight loss (positive difference favour FDG PET-CT-guided active surveillance) at 2 years follow-up										
1 (Mehanna 2017)	RCT	338[1]	MD[2] 1.49	N/A	N/A	Not serious	N/A	Not serious	Very serious[3]	Low
EORTC's QLQ-C30 appetite loss (positive difference favour FDG PET-CT-guided active surveillance) at 2 years follow-up										
1 (Mehanna 2017)	RCT	346[1]	MD[2] 4.35	N/A	N/A	Not serious	N/A	Not serious	Very serious[3]	Low
EORTC's QLQ-C30 pain/general (positive difference favour FDG PET-CT-guided active surveillance) at 2 years follow-up										
1 (Mehanna 2017)	RCT	341[1]	MD[2] 3.98	N/A	N/A	Not serious	N/A	Not serious	Very serious[3]	Low
EORTC's QLQ-C30 nausea and vomiting (positive difference favour FDG PET-CT-guided active surveillance) at 2 years follow-up										
1 (Mehanna 2017)	RCT	347[1]	MD[2] -1.17	N/A	N/A	Not serious	N/A	Not serious	Very serious[3]	Low
EORTC's QLQ-C30 dyspnoea (positive difference favour FDG PET-CT-guided active surveillance) at 2 years follow-up										
1 (Mehanna 2017)	RCT	349[1]	MD[2] 1.34	N/A	N/A	Not serious	N/A	Not serious	Very serious[3]	Low
EORTC's QLQ-C30 fatigue (positive difference favour FDG PET-CT-guided active surveillance) at 2 years follow-up										

No. of studies	Study design	Sample size	Effect size (95% CI)	Absolute risk: ND	Absolute risk: FDG PET-CT (95% CI)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
1 (Mehanna 2017)	RCT	350[1]	MD[2] 4.52	N/A	N/A	Not serious	N/A	Not serious	Very serious[3]	Low

1. Number of participants at 24 months after randomisation
2. Mean treatment difference was defined as the mean change from baseline in the PET–CT-guided active surveillance arm minus the mean change from baseline in the planned neck dissection arm
3. 95% confidence interval of the effect could not be estimated

CI: confidence interval; CRT: chemoradiotherapy; EORTC's QLQ-C30: European Organisation for Research and Treatment of Cancer's Quality of Life Questionnaire for Cancer with 30 questions; EQ-5D: EuroQol Group measure for health status; HR: hazard ratio; MDADI: MD Anderson Dysphagia Inventory; ND: neck dissection; N/A: not applicable; RCT: randomised controlled trial; SAEs: serious adverse events

1 **Diagnosing residual nodal disease in head and neck after radiotherapy or chemoradiotherapy**

No. of studies	Study design	Sample size	Sensitivity (95%CI)	Specificity (95%CI)	Effect size (95%CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	Quality
PET-CT – all sites, all studies										
10	Retrospective/ prospective cohort	764	70.5 (58.5, 80.2)	91.0 (88.5, 93.2)	LR+ 8.25 (6.17, 11.03)	Very serious[1]	Not serious	Not serious	Not serious	Low[2]
					LR- 0.32 (0.23, 0.47)	Very serious[1]	Not serious	Serious[3]	Not serious	Very low[2]
PET-CT – all sites, prospective studies only										
5	Prospective cohort	322	66.0 (49.8, 79.1)	92.3 (88.2, 95.0)	LR+ 10.87 (6.28, 18.80)	Serious[4]	Not serious	Not serious	Not serious	Very low
					LR- 0.38 (0.22, 0.65)	Serious[4]	Not serious	Not serious	Serious[5]	Very low
PET-CT – oropharynx										
2	Retrospective cohort	157	50.5 (27.2, 73.5)	91.9 (85.6, 95.6)	LR+ 8.77 (4.04, 19.05)	Serious[4]	Not serious	Not serious	Not serious	Low
					LR- 0.56 (0.37, 0.84)	Serious[4]	Not serious	Not serious	Serious[5]	Very low

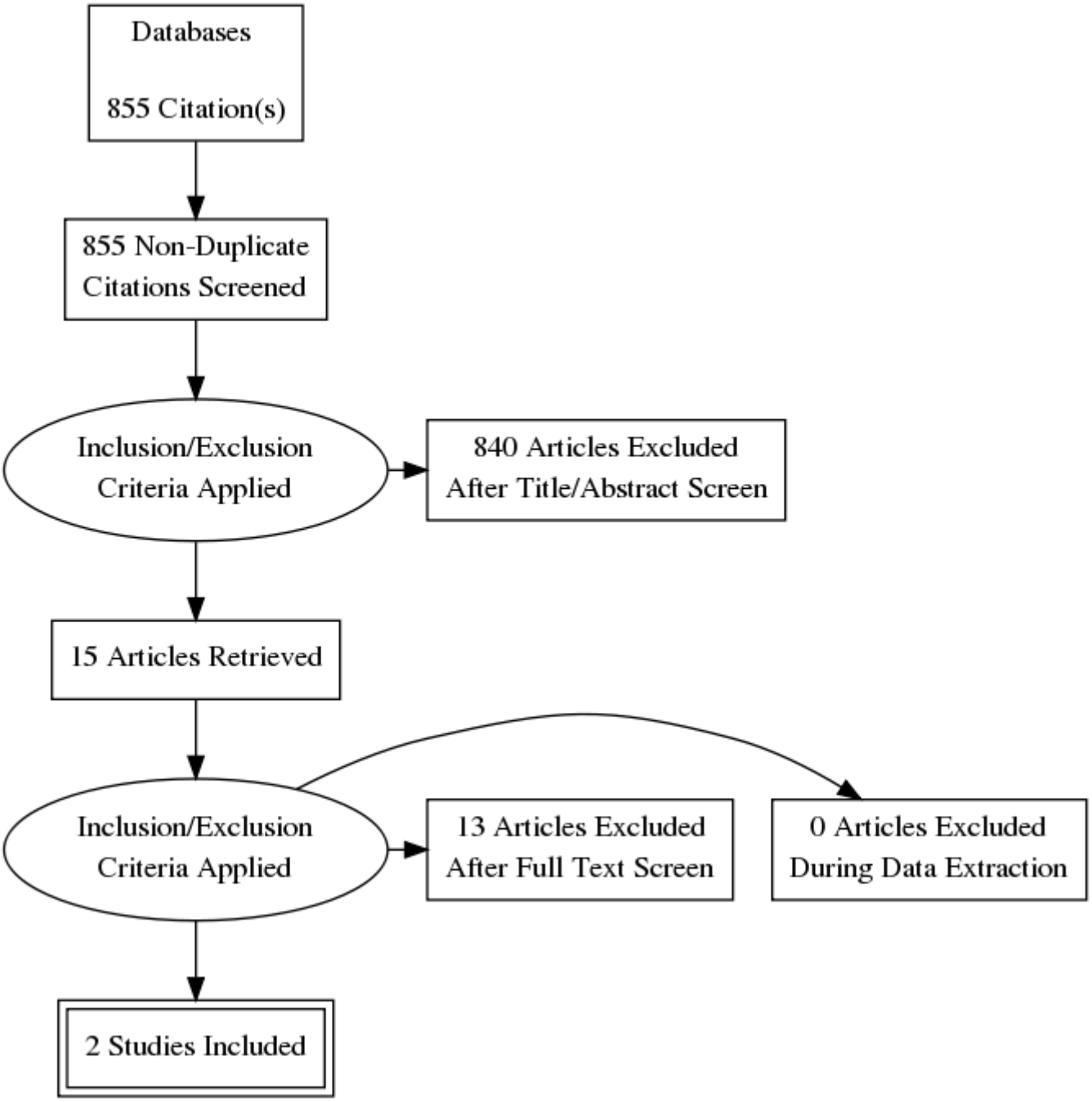
1. >33.3% of weighted data from studies at high risk of bias

No. of studies	Study design	Sample size	Sensitivity (95%CI)	Specificity (95%CI)	Effect size (95%CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	Quality

2. The only two studies not rated as being at high risk of bias were conducted solely in patients with oropharyngeal cancer
3. i-squared >33.3%
4. >33.3% of weighted data from studies at moderate or high risk of bias
5. 95% confidence interval for likelihood ratio crosses one end of a defined MID interval – (0.5, 2)

1

Appendix I – Economic evidence study selection



Appendix J – Health economic evidence profiles

Study, population, country and quality	Data sources	Other comments	Results			Conclusions	Uncertainty
			Incremental cost	Incremental effect (QALY)	ICER (/QALY)		

Study, population, country and quality	Data sources	Other comments	Results			Conclusions	Uncertainty
			Incremental cost	Incremental effect (QALY)	ICER (/QALY)		
Sher et al 2010 Patients with node positive head and neck squamous cell carcinoma (HNSCC). US focus to analysis **Partially applicable** [a, b, c] **Potentially serious limitations** [d, e]	Effects: Quality adjusted life years Costs: adopting the perspective of Medicare and obtained from the available schedules Utilities: obtained from an existing study for health states within head and neck cancer. Utilities for patients in the metastatic state were obtained from a preference elicitation from metastatic esophageal cancer patients. For patients who have undergone salvage surgery for an LR, utility values were taken from a preference study for palliative therapies for patients with inoperable oesophageal cancer	Comparing PET-CT detecting ND vs planned ND (all) A Markov model predicting costs and health benefits for 5 years and consisting of five health states: distant metastasis, local recurrence, salvage (dissection/local surgery), nodal recurrence and death (disease caused or other causes). Discounting at 3% per year was applied	The authors did not report the expected cost and QALYs in each arm. It was only reported that PET-CT was dominant option			PET-CT detecting ND for patients with RD was the dominant strategy over a wide range of realistic and exaggerated assumptions. Probabilistic sensitivity analyses confirmed that the PET-CT strategy was almost certainly cost-effective at a rang of societal willingness-to-pay thresholds from $25,000 to $500 000/QALY.	One-way sensitivity analyses and PSA were performed. Only one scenario where the upper extreme value assigned to health related utility for those having ND led to ND being more effective with ICER at $380,000/QALY

Study, population, country and quality	Data sources	Other comments	Results			Conclusions	Uncertainty
			Incremental cost	Incremental effect (QALY)	ICER (/QALY)		
PET-Neck Mehanna et al 2017 Patients with node positive head and neck squamous cell carcinoma (HNSCC). UK study **Directly applicable** **Minor limitations** [f, g]	**<u>Within Trial Analysis</u>** <u>Effects:</u> within-RCT measurement of EQ-5D with area-under-the-curve calculation of QALYs, with various assumptions tested in sensitivity analysis <u>Costs:</u> within-RCT NHS resource-use (missing values multiply imputed); unit costs from BNF, NHS RefCosts. Scenarios with a wider societal perspective including patients' loss in productivity and out of pocket expenses were considered. Also, scenario with including multiple imputations for the very small sample of patients who reported data on their primary and community care resource use was considered.	PET–CT-guided active surveillance (watch-and-wait) policy vs the current practice of planned ND Two year time horizon Second year costs and outcomes discounted at 3.5%	Base case			The probability of PET-CT being dominant option vs ND for all patients (saving cost and producing more QALYs) is 99% at a threshold of £20,000 per additional QALY	Sensitivity analyses including societal costs, complete case analysis or MI including the primary and social care services Mean costs and QALYs derived from the within trial analysis were bootstrapped 10,000 times. The study's results appear to be robust in the all sensitivity analyses performed in within trial analysis.
			-£1,513	**0.07**	**Dominant**		

Study, population, country and quality	Data sources	Other comments	Results			Conclusions	Uncertainty
			Incremental cost	Incremental effect (QALY)	ICER (/QALY)		
PET-Neck Mehanna et al 2017 Patients with node positive head and neck squamous cell carcinoma (HNSCC). UK study **Directly applicable** **Minor limitations** [f, g, h]	**Model-based Analysis** Effects: for the first six months data were taken from the PET-NECK trial. Utility value assigned to the DF state was taken also from the trial as the sample size allows. Utility values, however assigned to patients in the LR or DR states were obtained from another Canadian study, using SG and VAS for preference elicitation. Costs: DF cost was derived from the trial data (the average monthly cost in each arm between 6 and 24 months). The initial cost applied in the first cycle once the patient moved to LR or DR was also taken from the trial data. However, the ongoing supportive care cost was obtained from the literature	PET–CT-guided active surveillance (watch-and-wait) policy vs the current practice of planned ND Lifetime costs and health benefits Costs/QALYs discounted from second year and onward at a rate of 3.5% a year.	**-£1,485**	**0.13**	**Dominant**	The probability of PET-CT being cost effective option vs ND for all patients is 75% at a threshold of £20,000 per additional QALY	One way sensitivity analysis and PSA were performed in the model-based analysis. The key driving parameter that may influence the results when altered by 25% is the primary recurrence rate. The study's results appear to be robust in the all sensitivity analyses performed in the model-based analysis.

a) Not a UK study
b) Different discount rate from the NICE reference case
c) Non-EQ5D utility
d) The time horizon is not sufficiently long to reflect all important differences in costs and outcomes
e) Expected costs and QALYs in each arm were not reported
f) The population within the study was predominated by the N2 nodal type (almost 80% of patients), which is the most common type. Very few patients had N3 nodal type.
g) Likewise, in terms of tumour site, patients with oropharyngeal cancer represented about 85% of the patients in both arms. This may affect the generalisability of the study results on the least common subgroups of the HNSCC
h) The unit cost used for the salvage surgery doesn't match (£7722 used in the analyses, whereas the elective price in the reference cost return for 13/14 is £4378 (CA93A)). However, this cost is used in the model-based analysis and assigned to patients coming either from the PET-CT arm or the ND arm and experienced local recurrence, (i.e. affecting the two study arms)

1 Appendix K – Excluded studies

2 Clinical studies

3 Management of nodal metastasis in head and neck cancer after chemoradiotherapy

4 No studies were excluded during full text screening.

5 Diagnostic accuracy of PET-CT to diagnose residual nodal disease after radiotherapy or 6 chemoradiotherapy

Short Title	Title	Reason for exclusion
Bird 2016	18F-FDG PET/CT to assess response and guide risk-stratified follow-up after chemoradiotherapy for oropharyngeal squamous cell carcinoma	PET-CT only pre-treatment, follow-up done by unspecified imaging and histopathology. Prediction of residual tumour post-treatment by pre-treatment PET-CT scans, not PET-CT scans diagnosing nodal disease post-treatment.
Chen 2006	PET-CT vs contrast-enhanced CT: what is the role for each after chemoradiation for advanced oropharyngeal cancer?	Clinical examination only but also used the index tests as a secondary reference.
Cheung 2016	Detecting Residual/Recurrent Head Neck Squamous Cell Carcinomas Using PET or PET/CT: Systematic Review and Meta-analysis	Does not include a population who underwent chemoradiotherapy as primary treatment. Not possible to create a 2x2 table from data.
Evangelista 2014	Comparison between anatomical cross-sectional imaging and 18F-FDG PET/CT in the staging, restaging, treatment response, and long-term surveillance of squamous cell head and neck cancer: a systematic literature overview	PET-CT and PET not separated in analysis.
Ghanooni 2011	18F-FDG PET/CT and MRI in the follow-up of head and neck squamous cell carcinoma	Does cover patients who have undergone chemoradiotherapy but also radiotherapy only. The proportion of patients who received chemoradiotherapy was low (40.6%).
Gourin 2006	Utility of positron emission tomography-computed tomography in identification of residual nodal disease after chemoradiation for advanced head and neck cancer	Retrospective study that is case controlled.
Gourin 2009	Revisiting the role of positron-emission tomography/computed tomography in determining the need for planned neck dissection following chemoradiation for advanced head and neck cancer	Retrospective study that is case controlled.
Gupta 2010	Diagnostic performance of response assessment FDG-PET/CT in patients with head and neck squamous cell carcinoma treated with high-precision definitive (chemo)radiation	Does not discern between chemoradiotherapy patients and radiotherapy patients in analysis. Does not discern between PET and PET-CT in analysis.
Gupta 2011	Diagnostic performance of post-treatment FDG PET or FDG PET/CT	Does not discern between chemoradiotherapy patients and radiotherapy patients in analysis. Does not

Short Title	Title	Reason for exclusion
	imaging in head and neck cancer: a systematic review and meta-analysis	discern between PET and PET-CT in analysis.
Haerle 2010	18F-FET PET/CT in Advanced Head and Neck Squamous Cell Carcinoma: an Intra-individual Comparison with 18F-FDG PET/CT	Does not include a population who underwent chemoradiotherapy.
Kim 2011	Evaluation of 18F-FDG PET/CT and CT/MRI with histopathologic correlation in patients undergoing salvage surgery for head and neck squamous cell carcinoma	The proportion of patients who received chemoradiotherapy was low (33.3%).
Kim 2016	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	Retrospective study that is not blinded. Outcome includes more than residual/recurrent and is too board for the scope of this review.
Loo 2011	Neck dissection can be avoided after sequential chemoradiotherapy and negative post-treatment positron emission tomography-computed tomography in N2 head and neck squamous cell carcinoma.	Only one patient had increased FDG uptake on PET-CT and underwent neck dissection. The patients with negative PET-CT results did not undergo neck dissection, only observation. Retrospective study that is not blinded.
Lyford-Pike 2009	Limitations of PET/CT in determining need for neck dissection after primary chemoradiation for advanced head and neck squamous cell carcinoma	Retrospective study that is not blinded.
Malone 2009	Early prediction of response to chemoradiotherapy for head and neck cancer: reliability of restaging with combined positron emission tomography and computed tomography	Retrospective study that is not blinded.
Marcus 2014	Head and neck PET/CT: therapy response interpretation criteria (Hopkins Criteria)-interreader reliability, accuracy, and survival outcomes	Retrospective study that is not blinded.
Matoba 2015	Lesion regression rate based on RECIST: prediction of treatment outcome in patients with head and neck cancer treated with chemoradiotherapy compared with FDG PET-CT	PET-CT performed for pre-treatment only, CT only post-treatment. Not possible to create a 2x2 table from data presented I study.
Mehanna 2016	PET-CT Surveillance versus Neck Dissection in Advanced Head and Neck Cancer	Incorrect study design as an RCT
Ng 2010	Comprehensive imaging of residual/recurrent nasopharyngeal carcinoma using whole-body MRI at 3 T compared with FDG-PET-CT	CRT completion - scan interval range too broad
Noij 2017	Detection of residual head and neck squamous cell carcinoma after (chemo)radiotherapy: a pilot study assessing the value of diffusion-weighted magnetic resonance imaging	Retrospective study that is case controlled.

Short Title	Title	Reason for exclusion
	as an adjunct to PET-CT using 18F-FDG	
Rabalais 2009	Positron Emission Tomography–Computed Tomography Surveillance for the Node-Positive Neck After Chemoradiotherapy	Retrospective study that is not blinded.
Sagardoy 2016	Accuracy of (18) FDG PET-CT for treatment evaluation 3 months after completion of chemoradiotherapy for head and neck squamous cell carcinoma: 2-year minimum follow-up	Not possible to calculate a 2x2 table from data presented in the study.
Schouten 2015	Response evaluation after chemoradiotherapy for advanced nodal disease in head and neck cancer using diffusion-weighted MRI and 18F-FDG-PET-CT	Retrospective study that is case controlled.
Seitz 2009	18F-Fluorodeoxyglucose-PET/CT to evaluate tumor, nodal disease, and gross tumor volume of oropharyngeal and oral cavity cancer: Comparison with MR imaging and validation with surgical specimen	Measures recurrent disease only and not residual nodal disease.
Sharma 2013	Utility of 18F-FDG PET-CT in staging and restaging of patients with malignant salivary gland tumours: A single-institutional experience	Does not specify which treatments the patients received. Patients undergoing initial staging also included in analysis.
Slevin 2015	Assessment of outcomes with delayed (18)F-FDG PET-CT response assessment in head and neck squamous cell carcinoma	Retrospective study that is not blinded.
Slevin 2017	Accuracy of [18Fluorine]-Fluoro-2-Deoxy-d-Glucose Positron Emission Tomography-Computed Tomography Response Assessment Following (Chemo)radiotherapy for Locally Advanced Laryngeal/Hypopharyngeal Carcinoma	Retrospective study that is not blinded.
Taghipour 2015	The value of follow-up FDG-PET/CT in the management and prognosis of patients with HPV-positive oropharyngeal squamous cell carcinoma	Does not contain a population who received chemoradiotherapy. The reviewers were not blinded to result of reference test.
Taghipour 2016	Comparative effectiveness study: post treatment FDG PET/CT versus contrast enhanced ct in patients with oropharyngeal squamous cell carcinoma	Conference abstract
Taghipour 2016	FDG PET/CT in Patients With Head and Neck Squamous Cell Carcinoma After Primary Surgical Resection With or Without Chemoradiation Therapy	All patients had surgery during primary treatment
Uzel 2013	Is FDG -PET-CT a valuable tool in prediction of persistent disease in head and neck cancer	Unclear if PET-CT was used during post-chemoradiotherapy but a mixture of histological and imaging techniques were used. Retrospective study that is not blinded.

Short Title	Title	Reason for exclusion
Wei 2016	Comparison of 18F-FDG PET/CT, MRI and SPECT in the diagnosis of local residual/recurrent nasopharyngeal carcinoma: A meta-analysis	Combines data from PET and PET-CT studies into one analysis.
Wong 2017	Changes in multimodality functional imaging parameters early during chemoradiation predict treatment response in patients with locally advanced head and neck cancer	Reference standard in study does not match that specified in protocol.
Yao 2017	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	Reference standard in study does not match that specified in protocol. Retrospective study that is not blinded.
Zhang 2011	The benefit of early PET/CT surveillance in HPV-associated head and neck squamous cell carcinoma	Does not specify which treatment(s) were used for the patients, no data in baseline characteristics table. Mentions that patients with HPV-+ve carcinomas have better response to CRT but does not say anywhere in the methods which treatment was used.

1 Economic studies

Short Title	Title	Reason for exclusion
Greuter et al 2017	Cost-effectiveness of response evaluation after chemoradiation in patients with advanced oropharyngeal cancer using 18F-FDG-PET-CT and/or diffusion-weighted MRI	It is based on a trial of 46 patients in the Netherlands. Used data to determine the spec of the diagnosis strategies from another study in India 86 pts to assess in SA. Not a CUA; proportion of correctly diagnosed patients and cost per patient
Van Hooren et al 2009	The cost-effectiveness of 18FDG-PET in selecting patients with suspicion of recurrent laryngeal carcinoma after radiotherapy for direct laryngoscopy	Out of the scope; the study assesses the CE of PET-CT in selecting patients for laryngoscopy
Annunziata et al 2014	Cost-effectiveness of Fluorine-18-Fluorodeoxyglucose positron emission tomography in tumours other than lung cancer: A systematic review	Systematic review reporting already existing studies
Buck et al 2010	Economic evaluation of PET and PET/CT in oncology: evidence and methodologic approaches	Review reporting results from an already existing study
Anonymous	Rapid HTA on the use of Positron Emission Tomography (PET) in the diagnosis, staging and re-staging of head and neck cancers	Structured abstract
Kurien et al 2011	Cost-effectiveness of positron emission tomography/computed tomography in the management of advanced head and neck cancer	Costing study
Pryor et al 2013	Economic analysis of FDG-PET-guided management of the neck after primary chemoradiotherapy for node-positive head and neck squamous cell carcinoma	Cost minimisation analysis

Short Title	Title	Reason for exclusion
Rabalais et al 2012	A cost-effectiveness analysis of positron emission tomography-computed tomography surveillance versus up-front neck dissection for management of the neck for N2 disease after chemoradiotherapy	Not CUA; outcome is measured by number of patients who are free of disease at 1 year; USA settings
Smith et al 2015	Cost-effectiveness of positron emission tomography-CT in the evaluation of cancer of unknown primary of the head and neck	Canadian study reporting the outcome as cost per life year gained. The 2 comparative groups were: (1) PET-CT followed by panendoscopy versus (2) panendoscopy alone.
Yen et al 2009	The Cost-utility Analysis of 18-Fluoro-2- Deoxyglucose Positron Emission Tomography in the Diagnosis of Recurrent Nasopharyngeal Carcinoma	Not applicable evidence (Chinese study)
Smith et al 2017	Cost-effectiveness analysis of PET-CT-guided management for locally advanced head and neck cancer	Different publication, reporting same findings as an included study
Mehanna et al 2016	PET-CT Surveillance versus Neck Dissection in Advanced Head and Neck Cancer	No economic evaluation reported in this study
Hollenbeak et al 2001	The cost-effectiveness of fluorodeoxyglucose 18-F positron emission tomography in the N0 neck. Cancer	Population with no nodal diseases

1

Appendix L – References

Included clinical studies

Management of nodal metastasis in head and neck cancer after chemoradiotherapy

Mehanna H, McConkey CC, Rahman JK, Wong WL, Smith AF, Nutting C, Hartley AG, Hall P, Hulme C, Patel DK, Zeidler SV, Robinson M, Sanghera B, Fresco L, and Dunn JA (2017) PET-NECK: a multicentre randomised Phase III non-inferiority trial comparing a positron emission tomography-computerised tomography-guided watch-and-wait policy with planned neck dissection in the management of locally advanced (N2/N3) nodal metastases in patients with squamous cell head and neck cancer.. Health technology assessment (Winchester, and England) 21(17), 1-122

Mehanna H, Wong WL, McConkey CC, Rahman JK, Robinson M, Hartley AG, Nutting C, Powell N, Al-Booz H, Robinson M, Junor E, Rizwanullah M, von Zeidler SV, Wieshmann H, Hulme C, Smith AF, Hall P, and Dunn J (2016) PET-CT Surveillance versus Neck Dissection in Advanced Head and Neck Cancer.. The New England journal of medicine 374(15), 1444-54

Diagnostic accuracy of PET-CT to diagnose residual nodal disease after radiotherapy or chemoradiotherapy

Helsen N, Roothans D, Van Den Heuvel B, Van den Wyngaert T, Van den Weyngaert D, Carp L, and Stroobants S (2017) 18F-FDG-PET/CT for the detection of disease in patients with head and neck cancer treated with radiotherapy. PloS one 12(8), e0182350

Keski-Santti H, Mustonen T, Schildt J, Saarilahti K, and Makitie A (2014) FDG-PET/CT in the Assessment of Treatment Response after Oncologic Treatment of Head and Neck Squamous Cell Carcinoma. Clinical medicine insights. Ear, and nose and throat 7, 25-9

Nayak JV, Walvekar RR, Andrade RS, Daamen N, Lai SY, Argiris A, Smith RP, Heron DE, Ferris RL, Johnson JT, and Branstetter IVBF (2007) Deferring planned neck dissection following chemoradiation for stage IV head and neck cancer: The utility of PET-CT. Laryngoscope 117(12), 2129-2134

Ng SH, Chan SC, Yen TC, Liao CT, Lin CY, Chang JTC, Ko SF, Wang HM, Chang KP, and Fan KH (2011) PET/CT and 3-T whole-body MRI in the detection of malignancy in treated oropharyngeal and hypopharyngeal carcinoma. European Journal of Nuclear Medicine and Molecular Imaging 38(6), 996-1008

Pellini R, Manciocco V, Turri-Zanoni M, Vidiri A, Sanguineti G, Marucci L, Sciuto R, Covello R, Sperduti I, Kayal R, Anelli V, Pichi B, Mercante G, and Spriano G (2014) 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. Journal of cranio-maxillo-facial surgery: official publication of the European Association for Cranio-Maxillo-Facial Surgery 42(8), 1834-9

Prestwich RJD, Subesinghe M, Gilbert A, Chowdhury FU, Sen M, and Scarsbrook AF (2012) Delayed response assessment with FDG-PET-CT following (chemo) radiotherapy for locally advanced head and neck squamous cell carcinoma. Clinical radiology 67(10), 966-75

Ong SC, Schoder H, Lee NY, Patel SG, Carlson D, Fury M, Pfister DG, Shah JP, Larson SM, and Kraus DH (2008) Clinical utility of 18F-FDG PET/CT in assessing the neck after concurrent chemoradiotherapy for locoregional advanced head and neck cancer. Journal of Nuclear Medicine 49(4), 532-540

Sjovall J, Brun E, Almquist H, Kjellen E, and Wahlberg P (2014) Radiotherapy response in head and neck cancer - evaluation of the primary tumour site. Acta oto-laryngologica 134(6), 646-51

Taghipour M, Mena E, Kruse MJ, Sheikhbahaei S, and Subramaniam RM (2017) Post-treatment 18F-FDG-PET/CT versus contrast-enhanced CT in patients with oropharyngeal squamous cell carcinoma: comparative effectiveness study. Nuclear medicine communications 38(3), 250-258

Van Den Wyngaert T, Helsen N, Carp L, Hakim S, Martens M J, Hutsebaut I, Debruyne PR, Maes ALM, Van Dinther J, Van Laer CG, Hoekstra OS, De Bree R, Meersschout SAE, Lenssen O, Vermorken JB, Van Den Weyngaert D, and Stroobants S (2017) Fluorodeoxyglucose-positron emission tomography/computed tomography after concurrent chemoradiotherapy in locally advanced head-and-neck squamous cell cancer: The ECLYPS study. Journal of Clinical Oncology 35(30), 3458-3464

Included economic studies

Sher DJ, Tishler RB, Annino D, Punglia RS. Cost-effectiveness of CT and PET-CT for determining the need for adjuvant neck dissection in locally advanced head and neck cancer. Annals of oncology. 2009 Oct 15;21(5):1072-7.

Mehanna H, McConkey CC, Rahman JK, Wong WL, Smith AF, Nutting C, Hartley AG, Hall P, Hulme C, Patel DK, von Zeidler SV. PET-NECK: a multicentre randomised Phase III non-inferiority trial comparing a positron emission tomography–computerised tomography-guided watch-and-wait policy with planned neck dissection in the management of locally advanced (N2/N3) nodal metastases in patients with squamous cell head and neck cancer.

Excluded clinical studies

Management of nodal metastasis in head and neck cancer after chemoradiotherapy

No studies were excluded at full text screen.

Diagnostic accuracy of PET-CT to diagnose residual nodal disease after radiotherapy or chemoradiotherapy

Bird T, Barrington S, Thavaraj S, Jeannon J P, Lyons A, Oakley R, Simo R, Lei M, Guerrero Urbano T (2016) 18F-FDG PET/CT to assess response and guide risk-stratified follow-up after chemoradiotherapy for oropharyngeal squamous cell carcinoma. European Journal of Nuclear Medicine and Molecular Imaging 43(7), 1239-1247

Chen AY, Vilaseca I, Hudgins PA, Schuster D, and Halkar R (2006) PET-CT vs contrast-enhanced CT: what is the role for each after chemoradiation for advanced oropharyngeal cancer?. Head & neck 28(6), 487-95

Cheung PKF, Chin RY, and Eslick GD (2016) Detecting Residual/Recurrent Head Neck Squamous Cell Carcinomas Using PET or PET/CT: Systematic Review and Meta-analysis. Otolaryngology--head and neck surgery : official journal of American Academy of Otolaryngology-Head and Neck Surgery 154(3), 421-32

Evangelista L, Cervino AR, Chondrogiannis S, Marzola MC, Maffione AM, Colletti PM, Muzzio PC, and Rubello D (2014) Comparison between anatomical cross-sectional imaging and 18F-FDG PET/CT in the staging, restaging, treatment response, and long-term surveillance of squamous cell head and neck cancer: a systematic literature overview. Nuclear medicine communications 35(2), 123-34

Ghanooni R, Delpierre I, Magremanne M, Vervaet C, Dumarey N, Remmelink M, Lacroix S, Trotta N, Hassid S, and Goldman S (2011) 18F-FDG PET/CT and MRI in the follow-up of head and neck squamous cell carcinoma. Contrast media & molecular imaging 6(4), 260-6

Gourin CG, Boyce BJ, Williams HT, Herdman AV, Bilodeau PA, and Coleman TA (2009) Revisiting the role of positron-emission tomography/computed tomography in determining the need for planned neck dissection following chemoradiation for advanced head and neck cancer. The Laryngoscope 119(11), 2150-5

Gourin CG, Williams HT, Seabolt WN, Herdman AV, Howington JW, and Terris DJ (2006) Utility of positron emission tomography-computed tomography in identification of residual nodal disease after chemoradiation for advanced head and neck cancer. The Laryngoscope 116(5), 705-10

Gupta T, Jain S, Agarwal J P, Rangarajan V, Purandare N, Ghosh-Laskar S, and Dinshaw KA (2010) Diagnostic performance of response assessment FDG-PET/CT in patients with head and neck squamous cell carcinoma treated with high-precision definitive (chemo)radiation. Radiotherapy and oncology: journal of the European Society for Therapeutic Radiology and Oncology 97(2), 194-9

Gupta T, Master Z, Kannan S, Agarwal JP, Ghsoh-Laskar S, Rangarajan V, Murthy V, and Budrukkar A (2011) Diagnostic performance of post-treatment FDG PET or FDG PET/CT imaging in head and neck cancer: a systematic review and meta-analysis. European journal of nuclear medicine and molecular imaging 38(11), 2083-95

Haerle SK, Fischer DR, Schmid DT, Ahmad N, Huber GF, and Buck A (2010) 18F-FET PET/CT in Advanced Head and Neck Squamous Cell Carcinoma: an Intra-individual Comparison with 18F-FDG PET/CT. Molecular Imaging and Biology , 1-7

Kim R, Ock CY, Keam B, Kim TM, Kim JH, Paeng JC, Kwon SK, Hah JH, Kwon TK, Kim DW, Wu HG, Sung MW, and Heo DS (2016) 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 16(1), 116

Kim SY, Kim JS, Yi JS, Lee JH, Choi SH, Nam SY, Cho KJ, Lee SW, Kim SB, and Roh JL (2011) Evaluation of 18F FDG PET/CT and CT/MRI with histopathologic correlation in patients undergoing salvage surgery for head and neck squamous cell carcinoma. Annals of Surgical Oncology 18(9), 2579-2584

Loo SW, Geropantas K, Beadsmoore C, Montgomery PQ, Martin WM, and Roques TW (2011) Neck dissection can be avoided after sequential chemoradiotherapy and negative post-treatment positron emission tomography-computed tomography in N2 head and neck squamous cell carcinoma.. Clinical oncology (Royal College of Radiologists (Great Britain)) 23(8), 512-7

Lyford-Pike S, Ha PK, Jacene HA, Saunders JR, and Tufano RP (2009) Limitations of PET/CT in determining need for neck dissection after primary chemoradiation for advanced head and neck squamous cell carcinoma. ORL, and journal for oto-rhino-laryngology and its related specialties 71(5), 251-6

Malone JP, Gerberi Michael AT, Vasireddy S, Hughes LF, Rao K, Shevlin B, Kuhn M, Collette D, Tennenhouse J, and Robbins KT (2009) Early prediction of response to chemoradiotherapy for head and neck cancer: reliability of restaging with combined positron emission tomography and computed tomography. Archives of otolaryngology--head & neck surgery 135(11), 1119-25

Marcus C, Ciarallo A, Tahari AK, Mena E, Koch W, Wahl RL, Kiess AP, Kang H, and Subramaniam RM (2014) Head and neck PET/CT: therapy response interpretation criteria

(Hopkins Criteria)-interreader reliability, accuracy, and survival outcomes. Journal of nuclear medicine : official publication, and Society of Nuclear Medicine 55(9), 1411-6

Matoba M, Tuji H, Shimode Y, Kondo T, Oota K, and Tonami H (2015) Lesion regression rate based on RECIST: prediction of treatment outcome in patients with head and neck cancer treated with chemoradiotherapy compared with FDG PET-CT. Journal of radiation research 56(3), 553-60

Mehanna H, Wong W-L, McConkey CC, Rahman JK, Robinson M, Hartley AGJ, Nutting C, Powell N, Al-Booz H, Robinson M, Junor E, Rizwanullah M, V von Zeidler S, Wieshmann H, Hulme C, Smith AF, Hall P, and Dunn J (2016) PET-CT Surveillance versus Neck Dissection in Advanced Head and Neck Cancer. New England Journal of Medicine 374(15), 1444-1454

Ng SH, Chan SC, Yen TC, Liao CT, Chang JTC, Ko SF, Wang HM, Lin CY, Chang KP, and Lin YC (2010) Comprehensive imaging of residual/recurrent nasopharyngeal carcinoma using whole-body MRI at 3 T compared with FDG-PET-CT. European Radiology 20(9), 2229-2240

Noij DP, Jagesar VA, de Graaf P, de Jong MC, Hoekstra OS, de Bree R, and Castelijns JA (2017) Detection of residual head and neck squamous cell carcinoma after (chemo)radiotherapy: a pilot study assessing the value of diffusion-weighted magnetic resonance imaging as an adjunct to PET-CT using 18F-FDG. Oral surgery, oral medicine, and oral pathology and oral radiology 124(3), 296-305.e2

Rabalais AG, Walvekar R, Nuss D, McWhorter A, Wood C, Fields R, Mercante DE, and Pou AM (2009) Positron emission tomography-computed tomography surveillance for the node-positive neck after chemoradiotherapy. The Laryngoscope 119(6), 1120-4

Sagardoy T, Fernandez P, Ghafouri A, Digue L, Haaser T, de Clermont-Galleran H, Castetbon V, de Mones E (2016) Accuracy of (18) FDG PET-CT for treatment evaluation 3 months after completion of chemoradiotherapy for head and neck squamous cell carcinoma: 2-year minimum follow-up. Head & neck 38 Suppl 1, E1271-6

Schouten CS, Graaf PD, Alberts FM, Hoekstra OS, Comans EFI, Bloemena E, Witte BI, Sanchez E, Leemans CR, Castelijns JA, De Bree R (2015) Response evaluation after chemoradiotherapy for advanced nodal disease in head and neck cancer using diffusion-weighted MRI and 18F-FDG-PET-CT. Oral Oncology 51(5), 541-547

Seitz O, Chambron-Pinho N, Middendorp M, Sader R, Mac KM, Vogl TJ, and Bisdas S (2009) 18F-Fluorodeoxyglucose-PET/CT to evaluate tumor, nodal disease, and gross tumor volume of oropharyngeal and oral cavity cancer: Comparison with MR imaging and validation with surgical specimen. Neuroradiology 51(10), 677-686

Sharma P, Jain TK, Singh H, Suman SK. C, Faizi NA, Kumar R, Bal C, and Malhotra A (2013) Utility of 18F-FDG PET-CT in staging and restaging of patients with malignant salivary gland tumours: A single-institutional experience. Nuclear Medicine Communications 34(3), 211-219

Slevin F, Ermis K, Vaidyanathan S, Sen M, Scarsbrook AF, and Prestwich RJD (2017) Accuracy of [18Fluorine]-Fluoro-2-Deoxy-d-Glucose Positron Emission Tomography-Computed Tomography Response Assessment Following (Chemo)radiotherapy for Locally Advanced Laryngeal/Hypopharyngeal Carcinoma. Clinical Medicine Insights. Oncology 11, 1179554917713005

Slevin F, Subesinghe M, Ramasamy S, Sen M, Scarsbrook AF, and Prestwich RJD (2015) Assessment of outcomes with delayed (18)F-FDG PET-CT response assessment in head and neck squamous cell carcinoma. The British journal of radiology 88(1052), 20140592

Taghipour M, Marcus C, Califano J, Fakhry C, and Subramaniam RM (2015) The value of follow-up FDG-PET/CT in the management and prognosis of patients with HPV-positive oropharyngeal squamous cell carcinoma. Journal of medical imaging and radiation oncology 59(6), 681-6

Taghipour M, Mena E, Kruse M, Sheikhbahaei S, and Subramaniam R (2016) Comparative effectiveness study: post treatment FDG PET/CT versus contrast enhanced ct in patients with oropharyngeal squamous cell carcinoma. Clinical nuclear medicine. Conference: 2016 annual meeting and society of nuclear medicine and molecular imaging, SNMMI mid-winter meeting of the american college of nuclear medicine, and ACNM. United states 41(10), 823

Taghipour M, Sheikhbahaei S, Wray R, Agrawal N, Richmon J, Kang H, and Subramaniam RM (2016) FDG PET/CT in Patients With Head and Neck Squamous Cell Carcinoma After Primary Surgical Resection With or Without Chemoradiation Therapy. AJR. American journal of Roentgenology 206(5), 1093-100Uzel EK, Ekmekcioglu O, Elicin O, Halac M, and Uzel OE (2013) Is FDG -PET-CT a valuable tool in prediction of persistent disease in head and neck cancer. Asian Pacific journal of cancer prevention: APJCP 14(8), 4847-51

Wei J, Pei S, and Zhu X (2016) Comparison of 18F-FDG PET/CT, MRI and SPECT in the diagnosis of local residual/recurrent nasopharyngeal carcinoma: A meta-analysis. Oral oncology 52, 11-7

Wong KH, Panek R, Dunlop A, McQuaid D, Riddell A, Welsh LC, Murray I, Koh DM, Leach MO, Bhide SA, Nutting CM, Oyen WJ, Harrington KJ, and Newbold KL (2017) Changes in multimodality functional imaging parameters early during chemoradiation predict treatment response in patients with locally advanced head and neck cancer. Eur J Nucl Med Mol Imaging.

Yu Y, Mabray M, Silveira W, Shen PY, Ryan WR, Uzelac A, and Yom SS (2017) 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 39(3), 432-438

Zhang I, Branstetter IBF, Beswick DM, Maxwell JH, Gooding WE, and Ferris RL (2011) The benefit of early PET/CT surveillance in HPV-associated head and neck squamous cell carcinoma. Archives of Otolaryngology - Head and Neck Surgery 137(11), 1106-1111

Studies unavailable for retrieval

Huang SH, Chuang HC, Chien C, Chang YH, Hung BT, Fang FM, and Chang CC (2016) The proposed physiology-based FDG PET/CT criteria in reducing false-positive results in advanced head and neck cancer after chemoradiotherapy. The quarterly journal of nuclear medicine and molecular imaging: official publication of the Italian Association of Nuclear Medicine (AIMN).

Excluded economic studies

Greuter MJ, Schouten CS, Castelijns JA, de Graaf P, Comans EF, Hoekstra OS, de Bree R, Coupé VM. Cost-effectiveness of response evaluation after chemoradiation in patients with advanced oropharyngeal cancer using 18 F–FDG-PET-CT and/or diffusion-weighted MRI. BMC cancer. 2017 Dec;17(1):256.

van Hooren AC, Brouwer J, de Bree R, Hoekstra OS, Leemans CR, Uyl-de Groot CA. The cost-effectiveness of 18FDG-PET in selecting patients with suspicion of recurrent laryngeal carcinoma after radiotherapy for direct laryngoscopy. European Archives of Oto-Rhino-Laryngology. 2009 Sep 1;266(9):1441-8.

1 Annunziata S, Caldarella C, Treglia G. Cost-effectiveness of Fluorine-18-Fluorodeoxyglucose
2 positron emission tomography in tumours other than lung cancer: A systematic review. World
3 journal of radiology. 2014 Mar 28;6(3):48.

4 Buck AK, Herrmann K, Stargardt T, Dechow T, Krause BJ, Schreyögg J. Economic
5 evaluation of PET and PET/CT in oncology: evidence and methodologic approaches. Journal
6 of nuclear medicine technology. 2010 Mar 1;38(1):6-17.

7 Kurien G, Hu J, Harris J, Seikaly H. Cost-effectiveness of positron emission
8 tomography/computed tomography in the management of advanced head and neck cancer.
9 Journal of Otolaryngology--Head & Neck Surgery. 2011 Dec 1;40(6).

10 Pryor DI, Porceddu SV, Scuffham PA, Whitty JA, Thomas PA, Burmeister BH. Economic
11 analysis of FDG-PET–guided management of the neck after primary chemoradiotherapy for
12 node-positive head and neck squamous cell carcinoma. Head & neck. 2013 Sep
13 1;35(9):1287-94.

14 Rabalais A, Walvekar RR, Johnson JT, Smith KJ. A cost-effectiveness analysis of positron
15 emission tomography-computed tomography surveillance versus up-front neck dissection for
16 management of the neck for N2 disease after chemoradiotherapy. The Laryngoscope. 2012
17 Feb 1;122(2):311-4.

18 Smith KA, Dort JC, Hall SF, Rudmik L. Cost-effectiveness of positron emission tomography-
19 CT in the evaluation of cancer of unknown primary of the head and neck. Head & neck. 2015
20 Dec 1;37(12):1781-7.

21 Yen RF, Yen MF, Hong RL, Tzen KY, Chien CR, Chen TH. The cost-utility analysis of 18-
22 fluoro-2-deoxyglucose positron emission tomography in the diagnosis of recurrent
23 nasopharyngeal carcinoma. Academic radiology. 2009 Jan 1;16(1):54-60.

24 Smith AF, Hall PS, Hulme CT, Dunn JA, McConkey CC, Rahman JK, McCabe C, Mehanna
25 H. Cost-effectiveness analysis of PET-CT-guided management for locally advanced head
26 and neck cancer. European Journal of Cancer. 2017 Nov 1;85:6-14.

27 Mehanna H, Wong WL, McConkey CC, Rahman JK, Robinson M, Hartley AG, Nutting C,
28 Powell N, Al-Booz H, Robinson M, Junor E. PET-CT surveillance versus neck dissection in
29 advanced head and neck cancer. New England Journal of Medicine. 2016 Apr
30 14;374(15):1444-54.

1 Appendix M – Research recommendations

2 FDG PET-CT after radical chemoradiotherapy (long term outcomes)

Research recommendations A5a	What are the long term outcomes for people with an indeterminate FDG PET-CT scan result (no abnormal FDG uptake or residual mass) after radical chemoradiotherapy?
Population	People aged 16 and over with squamous-cell carcinoma of the nasopharynx, oropharynx, hypopharynx, larynx, oral cavity, or an unknown primary site in the head or neck and with nodal disease with indeterminate result on FDG PET-CT following primary radical chemoradiotherapy.
Predictive factor	Indeterminate result on FDG PET-CT following primary radical chemoradiotherapy.
Outcomes	• Recurrence rates • Overall survival • Quality of life (see core symptoms and domains in Appendix B) • Surgical complications • Adverse events
Measures	Adjusted • Hazard ratios • Risk ratios
Study design	• Randomised controlled trials • Prospective cohort studies

3

Potential criterion	Explanation
Importance to patients, service users or the population	The committee agreed that there is a variation in clinical practice about the follow-up of people with indeterminate result on FDG PET-CT following primary radical chemoradiotherapy.
Relevance to NICE guidance	High priority: the research is essential to inform future updates of key recommendations in the guideline.
Current evidence base	The committee highlighted that there is no current evidence about outcomes and investigations in people with indeterminate result on FDG PET-CT following primary radical chemoradiotherapy.
Equality	No specific equality concerns are relevant to this research recommendation.
Feasibility	There is evidence on positive and negative results of FDG PET-CT following primary radical chemoradiotherapy but no evidence on indeterminate results.

4 FDG PET-CT after radical chemoradiotherapy (additional investigations)

Research recommendations A5b	What are the most appropriate investigations for people with an indeterminate FDG PET-CT scan result (no abnormal FDG uptake or residual mass) after radical chemoradiotherapy?
Population	People aged 16 and over with squamous-cell carcinoma of the nasopharynx, oropharynx, hypopharynx, larynx, oral cavity, or an unknown primary site in the head or neck and with nodal disease with indeterminate result on FDG PET-CT following primary radical chemoradiotherapy.
Assessment tools	• Interval FDG PET-CT • Ultrasound +/- biopsy • Multi-parametric MRI • Serial imaging
Outcomes	• Recurrence rates

Research recommendations A5b	What are the most appropriate investigations for people with an indeterminate FDG PET-CT scan result (no abnormal FDG uptake or residual mass) after radical chemoradiotherapy?
	• Overall survival • Quality of life (see core symptoms and domains in Appendix B) • Surgical complications • Adverse events
Measures	• Sensitivity/specificity (preferred outcomes) • c-statistic, • Hazard ratios • Model fit (e.g. r-squared)
Study design	• Randomised controlled trials • Prospective cohort studies

1

Potential criterion	Explanation
Importance to patients, service users or the population	The committee agreed that there is a variation in clinical practice about the follow-up of people with indeterminate result on FDG PET-CT following primary radical chemoradiotherapy.
Relevance to NICE guidance	High priority: the research is essential to inform future updates of key recommendations in the guideline.
Current evidence base	The committee highlighted that there is no current evidence about outcomes and investigations in people with indeterminate result on FDG PET-CT following primary radical chemoradiotherapy.
Equality	No specific equality concerns are relevant to this research recommendation.
Feasibility	There is evidence on positive and negative results of FDG PET-CT following primary radical chemoradiotherapy but no evidence on indeterminate results.

2 Effectiveness of FDG PET-CT to guide follow-up

Research recommendations A6	What is the effectiveness and cost-effectiveness of FDG PET-CT to guide follow-up after treatment for people with head and neck cancer?
Population	People aged 16 and over with squamous-cell carcinoma of the nasopharynx, oropharynx, hypopharynx, larynx, oral cavity, or an unknown primary site in the head or neck.
Intervention	FDG PET-CT
Comparator	Usual care
Outcomes	• Recurrence rates • Overall survival • Quality of life (see core symptoms and domains in Appendix B)
Study design	• Randomised controlled trials • Observational studies • Model
Subgroups	• HPV status

3

Potential criterion	Explanation
Importance to patients, service	Regular follow-up after treatment is important to monitor the success of earlier treatment and guide treatment planning. The committee agreed that there is uncertainty about the effectiveness of FDG PET-CT to guide follow-

Potential criterion	Explanation
users or the population	up. The committee suggested to look for evidence in the subgroup of HPV status because HPV positive patients could have de-escalation of follow-up. Other low-risk categories could have the same.
Relevance to NICE guidance	High priority: the research is essential to inform future updates of key recommendations in the guideline.
Current evidence base	There is no evidence about the effectiveness of FDG PET-CT to guide follow-up.
Equality	No specific equality concerns are relevant to this research recommendation.
Feasibility	There is evidence on FDG PET-CT guiding follow-up after primary radical chemoradiotherapy but no evidence on FDG PET-CT for follow-up after other head and neck cancer treatments.

1 **Management of nodal metastasis in nasopharynx cancer after chemoradiotherapy**

Research recommendations A7	What is the optimal management strategy of nodal metastasis in nasopharynx cancer after chemoradiotherapy?
Population	People aged 16 and over with squamous-cell carcinoma of the nasopharynx and with nodal disease that has been treated with chemoradiotherapy
Intervention	FDG PET-CT-guided decision making
Comparator	Usual care
Outcomes	• Recurrence rates • Overall survival • Quality of life (see core symptoms and domains in Appendix B) • Adverse events
Study design	Randomised controlled trial

2

Potential criterion	Explanation
Importance to patients, service users or the population	The committee acknowledge that squamous-cell carcinoma of the nasopharynx is different to other cancer sites of the head and neck. Therefore, evidence on nasopharynx is crucial to improve care.
Relevance to NICE guidance	High priority: the research is essential to inform future updates of key recommendations in the guideline.
Current evidence base	There is evidence for other sites of head and neck cancer but not on nasopharynx.
Equality	No specific equality concerns are relevant to this research recommendation.
Feasibility	The PET-NECK trial excluded people with primary nasopharynx carcinoma because it has different biological behaviours and natural history compared with other head and neck cancer sites. The PET-NECK researchers highlighted that nasopharynx carcinoma is highly sensitive to radiotherapy and should not be treated by neck dissection surgery.

3

1 Appendix N – Additional data from PET-NECK trial

2 Baseline data on nodal status by site

Tumour site	N stage				Total
	N2a	N2b	N2c	N3	
Oral	4 (36.4%)	5 (45.5%)	2 (18.2%)	0	11
Oropharyngeal	85 (17.9%)	297 (62.4%)	80 (16.8%)	14 (2.9%)	476
Laryngeal	2 (5.4%)	22 (59.5%)	13 (35.1%)	0	37
Hypopharyngeal	4 (13.8%)	16 (55.2%)	7 (24.1%)	2 (6.9%)	29
Occult H&N	3 (27.3%)	5 (45.5%)	2 (18.2%)	1 (9.1%)	11
Total	98 (17.4%)	345 (61.2%)	104 (18.4%)	17 (3.0%)	564

3 Baseline data on HPV status by gender

p16 status	Gender		Total
	Male	Female	
Positive	275 (82.1%)	60 (17.9%)	335
Negative	88 (79.3%)	23 (20.7%)	111
Not available	97 (82.2%)	21 (17.8%)	118
Total	460 (81.6%)	104 (18.4%)	564

4