FDG-PET, CT, MRI for diagnosis of local residual or recurrent nasopharyngeal carcinoma, which one is the best: a systematic review

Liu T, Xu W, Yan W L, Ye M, Bai Y R, Huang G

CRD summary
This review found that positron emission tomography is more accurate than computed tomography and magnetic resonance imaging in the detection of patients with local residual or recurrent carcinoma. These conclusions appear reliable, but should be interpreted with some degree of caution given the limitations of the search and lack of details of the individual included studies.

Authors' objectives
To compare the performance of fluorodeoxyglucose positron emission tomography (FDG-PET), computed tomography (CT) and magnetic resonance imaging (MRI) in the diagnosis of local residual or recurrent nasopharyngeal carcinoma.

Searching
MEDLINE and EMBASE were searched from 1990 to May 2007; the search terms, which were reported, included a diagnostic filter. CBM-disc (Chinese language studies), ScienceDirect, SpringerLink, Scopus, the Cochrane Database of Systematic Reviews and a database of Chinese Technological Journals were also searched. The reference lists of all retrieved articles were screened. The review was restricted to English and Chinese articles.

Study selection
Studies that assessed FDG-PET, CT or MRI in the diagnosis of local residual or recurrent nasopharyngeal carcinoma, and that provided sufficient data to construct a 2x2 table on a per-patient basis, were eligible for inclusion. The reference standard had to consist of histopathologic analysis and/or close clinical and imaging follow-up of at least 6 months. Studies in which the results were only presented for combined imaging modalities, i.e. they could not be extracted separately for individual imaging modalities, were excluded. Studies that scored 9 points or less on the quality assessment were also excluded.

Most of the studies performed the imaging evaluation at least 3 to 4 months after treatment (range: 2 weeks to 12 years). One study of PET evaluated PET and CT, all others evaluated PET alone. For PET, the amount of tracer ranged from 259 to 444 megabecquerel, and studies used qualitative and/or quantitative analysis. For CT, studies used single-section helical, non-helical, dual-section helical or four-section helical; most used a section thickness of 5 mm, although 2.5, 3, 4 and 4.25 mm were also reported. All but one study used a contrast agent. For MRI, studies used 0.5, 1.0 or 1.5 tesla (T); all but one study used a contrast agent of 0.1 mmol/kg gado-pentate dimeglumine. The patients were aged from 12 to 76 years. Eight studies used histopathologic analysis as the reference standard and 25 used a combination of histopathologic analysis and clinical follow-up of greater than 6 months (average 12 months). The outcomes reported in the review were the sensitivity, specificity, diagnostic odds ratio (DOR) and $Q^*$ index.

Two reviewers assessed studies for inclusion.

Assessment of study quality
The studies were assessed for methodological quality using the Quality Assessment of Diagnostic Accuracy Studies (QUADAS) tool. They were assigned a score out of 14 according to the number of items rated as 'yes'.

The authors did not state how many reviewers performed the quality assessment.

Data extraction
The results data were extracted as 2x2 tables of test performance. Where data were reported for different readers, multiple observations per reader, or for multiple CT and MRI systems and/or techniques, each was counted as a separate data set.
Two reviewers independently extracted the data using a standardised form. Any discrepancies were resolved through referral to a third reviewer.

**Methods of synthesis**

The pooled sensitivity, specificity and DOR were calculated separately for PET, CT and MRI. Summary receiver operating characteristic (SROC) curves were calculated and $Q^*$ (the point on the SROC curve where sensitivity and specificity are equal) was calculated. Z tests were used to compare estimated summary measures across imaging modalities.

Meta-regression was used to investigate heterogeneity. The SROC model was extended to include variables for year of publication, sample size, reference standard, language, imaging modality and each QUADAS item. Initially, a single factor regression analysis was carried out. All items that showed a significant association ($p<0.05$) were then included in a backward stepwise multivariate regression model, with items showing significant associations ($p<0.05$) being retained in the model.

Subgroup analysis was used to investigate technical differences in the way each of the imaging modalities were performed. A subgroup analysis for PET compared studies that used a qualitative analysis to those that used both qualitative and quantitative analyses. Subgroup analysis for CT compared type of scanner (non- and single-section helical versus dual- and multi-section helical) and section thickness (<5 mm versus 5 mm). Subgroup analysis for MRI compared MRI field strength (<1.5 T versus ≥1.5 T).

**Results of the review**

Twenty-one studies (1,813 patients) reporting 33 sets of 2x2 data were included.

A quality assessment was not possible for one study as the data were only reported in an abstract. Of the remaining 32 data sets, 29 included an appropriate patient spectrum and described selection criteria; 31 included an appropriate reference standard; all included an appropriate time between the index test and reference standard; 31 applied the reference standard to all patients; 7 applied the same reference standard irrespective of the index test result; 31 performed the reference standard independently of the index test; 21 provided sufficient details of the index test execution; 10 provided sufficient details of the reference standard execution; 31 interpreted index test results blind to the results of the reference standard; none of the studies interpreted the reference standard blind to the index test result; 24 reported that the same data were available to the person interpreting the index test as would be available in clinical practice; 28 reported uninterpretable or intermediate results; and 25 reported reasons for withdrawals.

PET (11 data sets; 578 patients): the pooled sensitivity was 95% (95% confidence interval, CI: 90, 97) and the pooled specificity 90% (95% CI: 87, 93). Qualitative analysis showed significantly higher pooled sensitivity compared with combined qualitative and quantitative analysis (98% versus 91%); specificity was unchanged.

CT (13 data sets; 681 patients): the pooled sensitivity was 76% (95% CI: 70, 81) and the pooled specificity 59% (95% CI: 55, 63). Dual-section helical and multi-section helical showed significantly higher sensitivity (91% versus 67%) and specificity (67% versus 56%) compared with non-helical or single-section helical CT. Studies that reported a section thickness of 5 mm reported significantly higher sensitivity (85% versus 71%) than studies that reported a section thickness of less than 5 mm, but specificity was significantly lower (56% versus 74%).

MRI (9 data sets; 470 patients): the pooled sensitivity was 78% (95% CI: 71, 84) and the pooled specificity 76% (95% CI: 71, 80). There was no difference in sensitivity and specificity based on magnetic field strength.

The pooled sensitivity and specificity of PET were both significantly higher ($p<0.001$) than those of CT and MRI. Estimates of sensitivity were similar for CT and MRI ($p>0.05$), but specificity was higher for MRI ($p<0.01$). This was supported by visual analysis of the SROC curves: the curve for PET was further to the upper left hand corner than those for CT and MRI, suggesting greater accuracy of this technique. The SROC plots suggested considerable heterogeneity in estimates of sensitivity and specificity for CT and MRI; heterogeneity was less for PET. The only item remaining in the multivariate regression model was imaging modality ($p=0.002$).
Authors' conclusions
PET is more accurate than CT and MRI in the detection of patients with local residual or recurrent carcinoma.

CRD commentary
This review addressed a focused question with inclusion criteria defined in terms of the intervention, reference standard and outcomes. A variety of databases were searched, but the search included a diagnostic filter, was restricted to English and Chinese studies, and did not seek unpublished studies. It is therefore likely that relevant studies have been missed. A detailed quality assessment was carried out using appropriate criteria, and the results were clearly presented and considered in the analysis. However, the authors also estimated summary scores and used these as a basis for inclusion in the review. This is not appropriate for the QUADAS tool which specifically recommends against the use of summary scores. It would have been preferable to have included all studies and then investigated the effects of quality within these studies (as was done for all other included studies). The analysis was appropriate and included an investigation of heterogeneity, although the use of more sophisticated models such as the bivariate/hierarchical SROC model would have been more likely to produce reliable results. Only aggregate data from the included studies was presented, with very little details on the individual included studies, especially in terms of population. The authors state that all reported sets of 2x2 data were treated as separate data sets which can be problematic, especially as it is unclear exactly which studies contributed multiple data sets. Overall, the authors' conclusions appear reliable, but should be interpreted with some degree of caution given the limitations of the search and lack of details of the individual included studies.

Implications of the review for practice and research
Practice: The authors did not state any implications for practice.

Research: The authors stated that further ideally designed studies are needed on PET and CT, more than 4-section helical CT and 3 T MRI.

Funding
Science and Technology Commission of Shanghai Municipality Funds (No. 04JC1104).

Bibliographic details
Liu T, Xu W, Yan W L, Ye M, Bai Y R, Huang G. FDG-PET, CT, MRI for diagnosis of local residual or recurrent nasopharyngeal carcinoma, which one is the best: a systematic review. Radiotherapy and Oncology 2007; 85(3): 327-335

PubMedID
18037523

DOI
10.1016/j.radonc.2007.11.002

Indexing Status
Subject indexing assigned by NLM

MeSH
Fluorodeoxyglucose F18; Humans; Magnetic Resonance Imaging; Nasopharyngeal Neoplasms /diagnosis; Neoplasm Recurrence, Local; Neoplasm, Residual /diagnosis; Positron-Emission Tomography; Sensitivity and Specificity; Tomography, X-Ray Computed

AccessionNumber
12007004181

Date bibliographic record published
09/08/2008
Date abstract record published
03/11/2008

Record Status
This is a critical abstract of a systematic review that meets the criteria for inclusion on DARE. Each critical abstract contains a brief summary of the review methods, results and conclusions followed by a detailed critical assessment on the reliability of the review and the conclusions drawn.