Systematic review: accuracy of imaging tests in the diagnosis of recurrent laryngeal carcinoma after radiotherapy

Brouwer J, Hooft L, Hoekstra OS, Riphagen II, Castelijns JA, de Bree R, Leemans CR

CRD summary
This review assessed the accuracy of imaging tests for recurrent laryngeal carcinoma after radiotherapy and concluded that 18-fluorodeoxyglucose positron emission tomography was sufficiently promising for a randomised controlled trial comparing it with conventional work-up. The included studies were small and few, which created wide confidence intervals around the sensitivity and specificity; further accuracy data might be useful before any trial.

Authors' objectives
To determine the diagnostic accuracy of imaging tests, which were computed tomography (CT), magnetic resonance imaging (MRI), thallium-201 scintigraphy, and 18-fluorodeoxyglucose (FDG) positron emission tomography (PET), in patients with a suspicion of recurrent laryngeal carcinoma after radiotherapy.

Searching
MEDLINE and EMBASE were searched for articles from January 1990 to April 2006 and the search terms were reported. Only studies reported in English, German, French, or Dutch were included. The bibliographies of the included studies were screened for additional articles.

Study selection
Studies of at least seven patients, who had undergone radiotherapy to cure a primary laryngeal carcinoma and who were tested during follow-up, by CT, MRI, thallium-201 scintigraphy, or 18-fluorodeoxyglucose (FDG) positron emission tomography (PET), were eligible for inclusion. Studies of patients treated with chemoradiation were excluded. Studies of CT or MRI published before 1990 were also excluded, as older scans were judged to be of inferior quality; the approximate introduction of FDG PET into clinical practice was 1990.

The included studies were of patients, with a wide range of tumour stages, and their age and gender distributions were often not reported. As a reference standard, a follow-up (between four and 45 months), histopathology, or both were used to determine the presence or absence of recurrence. Two studies used a dual-head camera and all the others used a full-ring PET camera. Positive PET scans were identified by qualitative visual analyses or the determination of standardised uptake values and the criteria were reported in full for each study.

Two reviewers independently assessed studies for inclusion and disagreements were resolved by consensus.

Assessment of study quality
The methodological quality of the included studies was assessed using the criteria recommended by the Cochrane Methods Group on Screening and Diagnostic Tests, including assessment of their internal validity, external validity (generalisability), and the detail reported on the imaging index test.

Methodological quality was assessed by two reviewers and disagreements were resolved by consensus.

Data extraction
Data were extracted to populate two-by-two contingency tables of test performance; the numbers of true positive, false negative, false positive and true negative test results. These data were used to calculate the test sensitivity and specificity with 95% confidence intervals. A correction factor of 0.5 was added to zero cells.

Two reviewers independently extracted the data, using a standardised form, and disagreements were resolved by consensus.
Methods of synthesis
Pooled estimates of sensitivity and specificity were calculated, where between-study heterogeneity was not found. Heterogeneity was assessed visually in forest plots and statistically using the Cochran Q test (heterogeneity was rejected if p was over 0.10). In the presence of mild heterogeneity, which was not defined, a random-effects model was used to generate the pooled estimates and studies were weighted by sample size. The estimates of sensitivity and specificity, for each study, were plotted in receiver operating characteristic space and summary receiver operating characteristic curves, estimated using the Moses and Littenberg model.

Results of the review
Eight studies, with 191 patients (range seven to 75), were included in the review. The prevalence of recurrent tumour varied from 29% to 63%. All eight studies were on FDG PET; three of them were head-to-head comparisons with CT, MRI, or both and one study compared it with 18-fluorothymidine (FLT) PET. No studies evaluating CT, MRI, or thallium-201 scintigraphy were identified.

In two of the eight studies, all patients underwent a valid reference test, five studies were prospective, and only one study reported missing data. Six of the eight studies included only patients with laryngeal carcinomas and only two studies included consecutive participants. One study reported the availability of clinical data and results of previous tests to clinicians interpreting the FDG PET. None described patient comorbidities and only one reported the duration of suspicion of recurrence prior to imaging. Three studies did not provide an adequate description of the FDG PET scan procedure.

One case-control study, using a dual-head coincidence gamma camera, was excluded from the meta-analyses, due to its different design. This study reported 100% sensitivity and specificity, with FDG PET, in 11 patients. For FDG PET, using a random-effects model, the pooled estimate of sensitivity was 89% (95% CI 80 to 94) and specificity was 74% (95% CI 64 to 83). Sensitivity ranged from 80% to 100% (Cochran Q test p=0.73) and specificity ranged from 63% to 100% (p=0.06).

Two studies, in the same group of patients, compared FDG PET with CT and MRI. In one of these studies, 12 out of 13 patients underwent MRI or CT in addition to FDG PET (it was unclear how imaging was assigned) and recurrence was correctly identified by MRI or CT in five cases, whilst seven were equivocal. In the other study, 19 of 31 scans were correct, six were equivocal, and six were incorrect. A study of 23 patients found a sensitivity of 58% and a specificity of 100% for CT, compared with a sensitivity of 80% and a specificity of 81% for FDG PET, in the same patients. The study comparing FLT with FDG PET found that both methods identified all five recurrences, but FLT PET had one false-positive result while FDG PET had none.

Authors’ conclusions
The diagnostic accuracy of FDG PET was promising and a randomised trial comparing conventional diagnostic work-up with FDG PET was justified.

CRD commentary
This review addressed a clearly stated research question, which was defined by broad, but appropriate, inclusion criteria. No inclusion criteria were defined for the reference standard, but those used in the included studies were fully described. The restriction of the included studies to four European languages introduced the possibility of language bias, and could have resulted in the omission of relevant data. No attempt to identify unpublished studies was reported, so the possibility of publication bias cannot be ruled out. The methodological quality of the included studies was comprehensively assessed and reported in full. Measures to minimise errors and bias were included in the review process. There was between-study heterogeneity, which means the value of pooled estimates of sensitivity and specificity is questionable. The included studies were few and small, which created wide confidence intervals around the estimates of sensitivity and specificity.

This means that the authors’ conclusion that these data were sufficient to justify a randomised controlled trial might be an overstatement and further accuracy data might be useful before a trial.
Implications of the review for practice and research

Practice: The authors made no recommendations for practice.

Research: The authors stated that the accuracy estimates generated by this review justified a randomised trial, in which conventional diagnostic work up would be compared with FDG PET to determine its clinical utility and cost-effectiveness.

Funding
Not stated.

Bibliographic details

PubMedID
18213716

DOI
10.1002/hed.20790

Original Paper URL

Indexing Status
Subject indexing assigned by NLM

MeSH
Carcinoma, Squamous Cell /diagnosis /therapy; Confidence Intervals; Diagnostic Imaging /methods; Female; Fluorodeoxyglucose F18; Humans; Laryngeal Neoplasms /diagnosis /therapy; Magnetic Resonance Imaging /methods; Male; Neoplasm Recurrence, Local /diagnosis /radiotherapy; Observer Variation; Positron-Emission Tomography /methods; Probability; ROC Curve; Risk Assessment; Sensitivity and Specificity; Tomography, X-Ray Computed /methods

AccessionNumber
12008105199

Date bibliographic record published
29/04/2009

Date abstract record published
14/07/2010

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.