Meta-analysis of positron emission tomographic and computed tomographic imaging in detecting mediastinal lymph node metastases in nonsmall cell lung cancer

Birim O, Kappetein A P, Stijnen T, Bogers A J

CRD summary
This review compared the accuracy of positron emission tomography (PET) and computed tomography (CT) for the detection of lymph node metastases in patients with non-small-cell lung cancer. The authors' conclusion, that PET is more accurate than CT for this purpose, was supported by the results but should be interpreted with caution given the limitations in the searches.

Authors' objectives
To compare the accuracy of 2-(18F)-fluoro-2-deoxy-D-glucose (FDG) positron emission tomography (PET) and computed tomography (CT) in detecting mediastinal lymph node metastases in patients with non-small-cell lung cancer (NSCLC).

Searching
MEDLINE was searched up to January 2003; the keywords were reported. In addition, the reference lists from all identified studies were screened for additional studies. Only studies published in the English language were eligible for inclusion. Studies published only as abstracts were excluded.

Study selection
Study designs of evaluations included in the review
Studies that involved at least 15 patients were eligible for inclusion. Both prospective and retrospective studies were included.

Specific interventions included in the review
Studies that included both PET and CT were eligible for inclusion.

Reference standard test against which the new test was compared
The studies had to include a reference standard to detect the presence of mediastinal lymph node metastases, although no specific inclusion criteria were specified. The reference standards used in the included studies were thoracotomy, mediastinoscopy and imaging follow-up with CT.

Participants included in the review
Studies restricted to patients with NSCLC were eligible for inclusion. Where reported, the proportion of male patients in the included studies ranged from 18 to 84% and the mean age ranged from 61 to 65 years.

Outcomes assessed in the review
The studies had to report sufficient data to enable the calculation of the sensitivity and specificity for malignancy.

How were decisions on the relevance of primary studies made?
Two reviewers independently screened studies for inclusion. Any disagreements were resolved through discussion.

Assessment of study quality
Eight items from the Standards for the Reporting of Diagnostic Accuracy Studies (STARD) guidelines (see Other Publications of Related Interest no.1) were assessed: description of study population; cohort assembly; study design; clear description of the CT technique; clear description of the PET technique; technical quality of the reference standard; clear definition of cut-off levels; and interpretation of both PET and CT scans independent of each other and without knowledge of histology. The studies were given a score of between 0 and 2 for each of these items (2 =
adequate, 1 = partial, 0 = not performed or not mentioned), giving a maximum possible score of 16. Two reviewers independently scored the studies for methodological quality. Any disagreements were resolved through discussion.

**Data extraction**
For each study, the sensitivity, specificity and diagnostic odds ratio (DOR) were calculated from the 2x2 table data. To prevent division by 0 when calculating the DOR, 0.5 was added to each cell in the 2x2 tables. The authors did not report how many reviewers carried out the data extraction.

**Methods of synthesis**

**How were the studies combined?**
The DOR was analysed using a random-effects meta-analysis for the logarithm of the DOR, leading to a pooled estimate of the DOR and the corresponding 95% confidence interval (CI). The bivariate approach of van Houwelingen and colleagues was used to pool sensitivity and specificity using the logit transformation (see Other Publications of Related Interest no.2). A summary receiver operating characteristic (ROC) curve analysis was also undertaken, using the Moses-Littenberg equally weighted least-squares method (see Other Publications of Related Interest no.3). The point Q*, which represents the maximum joint sensitivity and specificity, was calculated. Funnel plots were constructed to assess the possibility of publication bias.

**How were differences between studies investigated?**
The authors did not report how differences between the studies were formally investigated. However, data on heterogeneity were reported in the 'Results' section.

**Results of the review**

Seventeen studies (n=833) were included.

The quality scores ranged from 10 to 16 (maximum 16). The most poorly scored item was the description of the study population.

**PET.**
The sensitivity ranged from 66 to 100% and the specificity from 81 to 100%. The pooled sensitivity was 83% (95% CI: 77, 87) and the pooled specificity was 92% (95% CI: 89, 95). There was no statistical evidence of heterogeneity.

**CT.**
The sensitivity ranged from 20 to 81% and the specificity from 44 to 100%. The pooled sensitivity was 59% (95% CI: 50, 67) and the pooled specificity was 78% (95% CI: 70, 84). There was no statistical evidence of heterogeneity.

All studies reported a greater accuracy of PET compared with CT; for 10 studies these differences were considered statistically significant (p<0.05). The pooled DOR was 5.4 (95% CI: 3.3, 8.8) for CT compared with 76.4 (95% CI: 41.2, 141.7) for PET. There was no statistical evidence of heterogeneity in either DOR. An analysis of summary ROC curves also showed greater accuracy of PET compared with CT, with the summary ROC curve for PET much further towards the upper left hand quarter than that of CT. Q* was 70% (95% CI: 65, 75) for CT and 90% (95% CI: 86, 95) for PET; this difference was highly significant (p<0.0001).

There was no evidence of publication bias based on funnel plots and linear regression tests on the symmetry of them.

**Authors' conclusions**
FDG PET was more accurate than CT for the detection of lymph node metastases.

**CRD commentary**
This review addressed a focused review question that was supported by clearly defined inclusion criteria. The literature search was limited to MEDLINE and the review was restricted to published, English language studies. It is therefore likely that relevant studies might have been missed and the review might be subject to language and publication bias. Some methods of the review process were reported and these included appropriate steps to minimise bias. A formal quality assessment was undertaken.

The statistical analysis undertaken used rigorous methods and was appropriate given the data. Relevant study details were tabulated clearly and discussed in the text. The authors' conclusions were supported by the data presented, but should be interpreted with some degree of caution given the limitations of the search.

**Implications of the review for practice and research**

Practice: The authors stated that 'as more accurate staging should lead to more appropriate therapy, FDG PET may lead to improvements in both patient quality of life and in savings of healthcare costs'.

Research: The authors did not report any implications for research.

**Bibliographic details**


**PubMedID**

15620991

**DOI**

10.1016/j.athoracsur.2004.06.041

**Other publications of related interest**


**Indexing Status**

Subject indexing assigned by NLM

**MeSH**

Adult; Aged; Aged, 80 and over; Carcinoma, Non-Small-Cell Lung /radiography /radionuclide imaging /secondary; False Negative Reactions; False Positive Reactions; Female; Fluorodeoxyglucose F18; Humans; Lung Neoplasms /pathology /radiography /radionuclide imaging; Lymphatic Metastasis /radiography /radionuclide imaging; Male; Mediastinum /radiography /radionuclide imaging; Middle Aged; Positron-Emission Tomography; Prospective Studies; ROC Curve; Radiopharmaceuticals; Research Design; Retrospective Studies; Sensitivity and Specificity; Tomography, X-Ray Computed

**AccessionNumber**

12005004084

**Date bibliographic record published**

31/08/2007

**Date abstract record published**
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.