Accuracy of positron emission tomography for diagnosis of pulmonary nodules and mass lesions: a meta-analysis
Gould M K, Maclean C C, Kuschner W G, Rydzak C E, Owens D K

Authors' objectives
To estimate the diagnostic accuracy of 18-fluorodeoxyglucose positron emission tomography (FDG-PET) for malignant focal pulmonary lesions.

Searching
MEDLINE and Cancerlit were searched from January 1966 to September 2000. The reference lists of retrieved studies, and abstracts from conference proceedings, were reviewed for additional studies. Experts in the field were also contacted for additional studies. Studies published in any language were eligible for inclusion. Abstracts were only included when the study authors provided full reports of their methods and results.

Study selection
Study designs of evaluations included in the review
The studies had to include at least 10 participants, of which at least 5 had to have malignant lesions.

Specific interventions included in the review
Studies that examined FDG-PET or FDG with a gamma camera in coincidence mode for the diagnosis of pulmonary lesions or mass lesions, were included.

Reference standard test against which the new test was compared
The authors do not report details of the reference standard tests.

Participants included in the review
Patients with pulmonary nodules or masses. Some studies included only patients with known lung cancer. Where stated, the mean age of the patients ranged from 55.5 to 70.8 years. The proportion of men in the studies ranged from 39 to 99%. The median prevalence of malignancy was 72.5%.

Outcomes assessed in the review
The studies had to present sufficient data to allow the sensitivity and specificity for malignancy to be calculated.

How were decisions on the relevance of primary studies made?
Two investigators independently reviewed potential English language studies for inclusion. Any disagreements were resolved through discussion. One reviewer evaluated non-English language studies. The reviewers of the English language studies were blinded to the journal, author, institutional affiliation and date of publication.

Assessment of study quality
The studies were assessed according to the following criteria: technical quality of the index test; technical quality of the reference test; independence of test interpretation; description of the study population cohort assembly; sample size; and unit of data analysis. Two investigators independently assessed the English language studies for methodological quality, and subsequently resolved any disagreements by discussion. One reviewer assessed the quality of the non-English language studies.

Data extraction
One investigator extracted study data regarding the demographic characteristics of the participants, size distribution and the number of pulmonary nodules or mass lesions, and the study inclusion and exclusion criteria. Two investigators independently abstracted data regarding the prevalence of malignancy and the sensitivity and specificity of the imaging
test for malignancy. When possible, the test performance for pulmonary nodules measuring less than 3 to 4 cm in diameter was tabulated separately. Data from studies that used semi-quantitative methods for interpreting FDG-PET images, and from studies that evaluated FDG imaging with a gamma camera in coincidence mode, were also tabulated separately. When necessary, study authors who published more than one study were contacted to establish whether they reported the results of overlapping patient populations.

For each study, 2x2 contingency tables were constructed. The sensitivity, specificity and the log diagnostic odds ratio (DOR) were calculated; 0.5 was added to each cell in any 2x2 table that contained one or more zero values.

Methods of synthesis
How were the studies combined?
Summary receiver operating characteristic curves (sROC) were constructed using meta-analytic methods. When the sROC curve was symmetrical it was assumed that it could be described by a common or summary log DOR. When this condition was met, the ORs were pooled using the Mantel-Haenszel method. The maximum joint sensitivity and specificity was calculated. Publication bias was investigated by creating inverted funnel plots of individual study log ORs plotted against sample size.

How were differences between studies investigated?
A number of sensitivity analyses were conducted. The studies were ordered chronologically, and a cumulative meta-analysis was performed to investigate whether the diagnostic accuracy of FDG-PET had improved over time. To determine whether study quality affected diagnostic accuracy, studies which did and did not satisfy each quality criterion were compared. Studies that satisfied at least 70% of the quality criteria were also compared to those that did not. To evaluate whether the use of semi-quantitative methods to interpret FDG-PET images improves accuracy, studies that used qualitative and semi-quantitative methods of interpretation were compared. To investigate the influence of excluding abstracts for which full reports could not be obtained, data from these studies were included in a secondary analysis. To make statistical comparisons between groups of studies, log ORs were compared using unpaired t-tests or the Mann-Whitney U test, as appropriate.

Results of the review
Forty diagnostic accuracy studies (n=1,474) were included.

None of the studies met all of the criteria for study quality. Fourteen studies satisfied 70 to 80% of the criteria, while 19 satisfied 50 to 69% of the criteria.

The sensitivity ranged from 83% (with a specificity of 90%) to 100% (with a specificity ranging between 0 and 100). The specificity ranged from 0 (with a sensitivity of 97 to 100%) to 100% (with a sensitivity ranging from 87 to 100%). The mean sensitivity and specificity were 96 and 73.5%, respectively.

For lesions of any size, the summary log DOR for FDG-PET was 4.68 (95% confidence interval, CI: 4.21, 5.14). This corresponded to a maximum joint sensitivity and specificity of 91.2% (95% CI: 89.1, 92.9).

There was no significant difference in the accuracy of FDG-PET for pulmonary nodules compared with pulmonary lesions of any size (p=0.43). The diagnostic accuracy was similar for studies that used semi-quantitative methods of image interpretation and those that used qualitative methods (p=0.52). Accuracy appeared to be similar for studies of FDG-PET and studies of FDG imaging with a modified gamma camera. The cumulative meta-analysis showed that the summary log DOR gradually improved from 3.81 (95% CI: 1.24, 6.37) in 1990 to 5.09 (95% CI: 4.40, 5.78) in 1997, after which time it fell slightly.

Diagnostic accuracy was better in the 14 studies that met at least 70% of the criteria for methodological quality, compared with the remaining studies (p=0.007). The only aspect of study quality that had a statistically-significant effect on diagnostic accuracy was the blinding of FDG-PET image readers to the final diagnosis, with a higher log DOR reported in studies in which the readers were blinded. The summary log DOR for data from abstracts did not differ significantly from the summary log DORs for published articles. An inverted funnel plot did not suggest evidence of publication bias.
Cost information
The authors reported that, at present, Medicare reimbursement for FDG-PET imaging is approximately $1,912. In comparison, reimbursement for non-contrast computed tomography of the thorax is $276 and reimbursement for computed tomography-guided needle biopsy is approximately $560.

Authors' conclusions
PET with FDG is an accurate noninvasive imaging test for the diagnosis of pulmonary nodules and larger mass lesions, although few data exist for nodules smaller than 1 cm in diameter. In current practice, FDG-PET has high sensitivity and intermediate specificity for malignancy.

CRD commentary
This was a good review of the area. A reasonable literature search was conducted, although a more extensive search may have identified additional relevant articles. No language restrictions were applied and publication bias was investigated. The inclusion criteria were clearly stated, and the review methods were appropriate and described in detail. A formal quality assessment was performed and the effects of quality on the results were investigated. Relevant study details were presented in tabular format. However, the reference standard used in the studies was neither reported in the tables nor mentioned in the text. This information would be essential in order to fully interpret these results. Appropriate methods were used to pool the study results, but since heterogeneity was not formally investigated, it is unclear whether it was appropriate to pool the study results. A number of subgroup analyses were performed that revealed few differences between the studies.

The authors' conclusions are supported by the evidence but should be interpreted with some degree of caution due to the limitations highlighted, in particular, the lack of details on the reference standards used.

Implications of the review for practice and research
Practice: The authors state 'policy-level decisions regarding dissemination of FDG-PET must consider not only diagnostic accuracy but also clinical outcomes and costs'.

Research: The authors state 'formal cost-effectiveness studies are needed to determine if diagnostic strategies that include FDG-PET represent a good value for the health care dollar'.

Funding
Department of Veterans Affairs Cooperative Study in Health Services No. 27.

Bibliographic details

PubMedID
11180735

Original Paper URL
http://jama.ama-assn.org/

Indexing Status
Subject indexing assigned by NLM

MeSH
Glucose-6-Phosphate /analogs & derivatives; Humans; Lung Neoplasms /radionuclide imaging; ROC Curve;
Radiopharmaceuticals; Sensitivity and Specificity; Solitary Pulmonary Nodule /radionuclide imaging; Tomography, Emission-Computed

**AccessionNumber**
12001008317

**Date bibliographic record published**
31/01/2003

**Date abstract record published**
31/01/2003

**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.