Alzheimer disease: operating characteristics of PET. A meta-analysis
Patwardhan M B, McCrory D C, Matchar D B, Samsa G P, Rutschmann O T

CRD summary
This review found that the accuracy of positron emission tomography is unclear given limitations in the primary studies in terms of study design and patient characteristics. This review was generally well conducted and clearly reported, and these cautious conclusions are supported by the data presented.

Authors' objectives
To determine the accuracy of fluorine-18 fluorodeoxyglucose (FDG) positron emission tomography (PET) for the diagnosis of Alzheimer's disease (AD).

Searching
MEDLINE, CINAHL and HealthSTAR were searched from 1989 to 2003. The search terms, which were listed, included a diagnostic filter. The references of retrieved articles were screened to identify additional studies. Only studies published in English in peer-reviewed journals were included.

Study selection
Study designs of evaluations included in the review
Diagnostic accuracy studies that included at least 12 patients with AD were eligible for inclusion.

Specific interventions included in the review
Studies of FDG PET in which a dedicated scanner was used and the resolution was reported were eligible for inclusion. The included studies used a range of different scanners. The criteria for positivity also varied and included: frontoparietal hypometabolism; symmetrical parietotemporal hypometabolism; hypometabolism index; any hypometabolism; parietotemporal hypometabolism; metabolic ratio; any deficit; out of mean +/- 2 standard deviations; and parietotemporal and frontal hypometabolism.

Reference standard test against which the new test was compared
Studies in which the reference standard for the diagnosis of AD was either a clinical diagnosis according to standard criteria of the National Institute of Neurological and Communicative Disorders/Alzheimer's Disease and Related Disorders Association or the Diagnostic and Statistical Manual, or histopathological diagnosis were eligible for inclusion. All studies used clinical examination to diagnose dementia; some studies also used histopathological diagnosis after a period of follow-up in all or some of the included patients.

Participants included in the review
Studies that included patients with AD defined using standard criteria were eligible for inclusion. Most of the studies were conducted in tertiary care settings. Patients classified having AD included those with probable disease, possible disease, mild dementia, moderate dementia, severe dementia and mild cognitive impairment. Patients classified as controls included healthy controls, patients with multi-infarct dementia, dementia with Lewy bodies, dementia not caused by AD, questionable or mild dementia, and vascular dementia.

Outcomes assessed in the review
Studies had to provide sufficient data to construct a 2x2 table of test performance. The outcomes reported in the review were the sensitivity and specificity.

How were decisions on the relevance of primary studies made?
Two reviewers, one methodologist and one content expert, reviewed the abstracts of all studies identified by the searches. Studies considered relevant by at least one reviewer were obtained. Two reviewers then independently reviewed the full-text articles; any disagreements were resolved through consensus.
Assessment of study quality
Studies were assessed for methodological quality using the following eight criteria developed specifically for this review: scanner details; clear description of setting and selection of the population; inclusion of a representative sample of patients with an appropriate spectrum of disease; categorisation of results by disease severity; use of standard criteria for image interpretation; reference standard involved standard criteria and were applied on the basis of long-term follow-up of 1 year or more; follow-up complete (absence of verification bias); and blinded application of the index test and reference standard. Studies were assigned a score of 0 or 1 according to whether each criterion was fulfilled, and the scores summed to give a final quality score.

Two reviewers, one methodologist and one content expert, performed the quality assessment; any disagreements were resolved through consensus.

Data extraction
Two reviewers, one methodologist and one content expert, independently extracted the data onto forms developed for the review and agreed on the data extracted. The results data were extracted as 2x2 tables of test performance.

Methods of synthesis
How were the studies combined?
Sensitivity and specificity were pooled using a random-effects model. A summary receiver operating characteristic (ROC) curve was constructed using the Moses-Littenberg model. Individual study results were plotted in ROC space.

How were differences between studies investigated?
The studies were divided according to whether they compared patients with AD to healthy controls or to patients with dementia not caused by AD. Heterogeneity was investigated by separately plotting studies that used healthy controls in ROC space according to whether or not they fulfilled the following criteria: patients with mild AD included; patients with probable AD included; standard criteria for the diagnosis of AD applied; use of quantitative criteria for interpretation of PET; blinded interpretation of PET images; and association of operating characteristics with the quality score.

Results of the review
Fifteen studies were included in the review (total number of patients unclear).

The quality scores ranged from 1 to 6 out of a maximum of 8. Four studies included patients with a representative sample of mild to severe dementia. Two studies included patients with mild cognitive impairment. Twelve studies included patients with a clinical diagnosis of probable AD according to clinical criteria. Thirteen studies used clinical criteria as the reference standard, while the remaining two used histopathological criteria.

Comparison of PET in patients with AD and patients with dementia arising from other causes (6 studies).

The sensitivity ranged from 75 to 94% and the specificity from 18 to 86%. A meta-analysis was not conducted for these studies.

Comparison of PET in patients with AD and healthy controls (9 studies).

The sensitivity ranged from 61 to 100% and the specificity from 54 to 100%. The pooled sensitivity and specificity were 86% (95% confidence interval, CI: 76, 93) and 86% (95% CI: 72, 93), respectively.

Estimates of specificity were higher among studies that included a healthy control group than among those that included a control group of patients with other forms of dementia. Sensitivity was similar between the groups.

None of the factors investigated explained the observed differences in estimates of sensitivity and specificity between studies.
Authors' conclusions
The accuracy of PET, in terms of its sensitivity and specificity, is limited by both study design and patient characteristics. The clinical value of PET is unclear.

CRD commentary
This review addressed a focused question that was supported by clearly defined inclusion criteria. The search included a diagnostic filter, which means that relevant studies are likely to have been missed. No attempts were made to locate unpublished studies, so the review may be subject to publication bias. Full details of the review process were reported and these included appropriate steps to minimise bias. A detailed quality assessment was conducted and the results of this were reported and considered in the synthesis of results.

The methods used to pool the studies were acceptable, although a more robust analysis based on more sophisticated models would have been preferable. It is unclear why the results were pooled for studies that included a healthy control group but not for those that used patients with dementia from other causes as the controls, especially as these would be less likely to produce biased estimates than those that included healthy controls. The authors' conclusions appear to be based mainly on the studies that included a healthy control group and for which meta-analysis was conducted. However, these conclusions are suitably cautious and are supported by the data presented.

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

Research: The authors stated that future studies should be reported in accordance with the Standards for the Reporting of Diagnostic Accuracy Studies (STARD) guidelines and future researchers should consider the utility of their study before undertaking it. The authors also made more specific suggestions for future studies: the patients should be a representative sample of those likely to receive PET in practice; clear criteria should be used to define a positive scan; long-term follow-up should be undertaken to determine whether or not a patient has AD; and the results should be analysed according to disease severity.

Bibliographic details

PubMedID
15068942

DOI
10.1148/radiol.2311021620

Original Paper URL
http://radiology.rsnajnls.org/

Indexing Status
Subject indexing assigned by NLM

MeSH
Alzheimer Disease /classification /diagnosis; Fluorodeoxyglucose F18; Humans; Image Interpretation, Computer-Assisted; ROC Curve; Radiopharmaceuticals; Sensitivity and Specificity; Severity of Illness Index; Tomography, Emission-Computed /methods

AccessionNumber
12004000720
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.