Ventilation-perfusion scanning and helical CT in suspected pulmonary embolism: meta-analysis of diagnostic performance
Hayashino Y, Goto M, Noguchi Y, Fuku T

CRD summary
This review compared helical computed tomography (CT) and ventilation-perfusion (V-P) scanning in the detection of pulmonary embolism (PE). The authors concluded that helical CT is better than V-P scanning using a 'normal threshold' to exclude PE, while helical CT and V-P scanning with 'high probability threshold' have similar discriminatory power. The review was generally well conducted and the conclusions are likely to be reliable.

Authors' objectives
To determine the accuracy of helical computed tomography (CT) and ventilation-perfusion (V-P) scanning in the detection of acute pulmonary embolism (PE).

Searching
MEDLINE and EMBASE were searched from January 1985 to May 2003 for studies published in the English language; the search terms were reported. Studies of CT scanning were only included if they were published after 1990. References from retrieved articles were screened and experts in the area were contacted for additional relevant studies.

Study selection
Study designs of evaluations included in the review Diagnostic accuracy studies were included in the review.

Specific interventions included in the review
Studies of helical CT or V-P scanning for PE were eligible for inclusion. Studies in which non-comparable CT methods (e.g. electron beam CT) were used, or in which helical CT was performed after anticoagulant therapy or surgery for PE, were excluded. CT findings were classified as positive if an intraluminal filling defect or complete non-filling of a pulmonary artery was found. V-P findings were classified according to the probability of PE (high, intermediate, low, or near normal or normal). Collimation for CT ranged from 2 to 5 mm and pitch ranged from 1 to 2. Studies of V-P scanning used different ventilation radionuclides (details listed), but all used 99m-Tc-macroaggregated albumin as the perfusion radionuclide.

Reference standard test against which the new test was compared
Studies that used pulmonary angiography as the reference standard for the diagnosis of PE, and in which the time interval between the test and reference standard was 48 hours or less, were eligible for inclusion. Studies in which pulmonary angiography combined with any other modality was used as the reference standard were excluded.

Participants included in the review
Studies of patients being evaluated for acute PE were eligible for inclusion. Studies of chronic PE or septic embolism were excluded. In the included studies, the mean age of the participants ranged from 34 to 63 years and 4 to 60% were women.

Outcomes assessed in the review
The studies had to report sufficient data to extract a 2x2 table of test performance to be included in the review. The outcomes reported in the review were sensitivity, specificity, summary receiver operating characteristic (SROC) curves and estimates of the post-test probability of disease.

How were decisions on the relevance of primary studies made?
The authors did not state how the papers were selected for the review, or how many reviewers performed the selection.
Assessment of study quality
The studies were assessed for study design (including criteria used to define a positive result), patient recruitment, extent of blinding between readers, information on extent of disease, and any evidence of verification or test interpretation bias. Two independent reviewers extracted data on quality-related items. Any disagreements were resolved through consensus.

Data extraction
Two reviewers independently extracted the data from the included studies. Any disagreements were resolved by discussion and consensus. The data were extracted as 2x2 tables of test performance. Three different thresholds were evaluated for V-P scanning:

- high-findings classified as positive and all other findings classified as negative (threshold 1);
- high- and intermediate-probability findings classified as positive and all other findings classified as negative (threshold 2); and
- normal and/or near normal findings classified as negative and all other findings classified as positive (threshold 3).

For articles in which the results were tabulated for different observers, data for the first observer were extracted unless one observer was emphasised in the study.

Methods of synthesis
How were the studies combined?
The Spearman correlation coefficient was used to assess the correlation between sensitivity and specificity in order to determine the suitability of using an SROC curve analysis. The sensitivities and specificities were pooled using a random-effects model weighted according to sample size. SROC curves were estimated using the method of variance-weighted least-squares regression, with 0.5 added to each cell of the 2x2 table to account for cells with zero events. The final SROC curves were restricted to the range of observed sensitivity and specificity. After the sensitivities and specificities had been pooled, the post-test probability of disease was estimated based on different pre-test probabilities of disease: low (3%), moderate (27%) and high (78%).

How were differences between studies investigated?
The heterogeneity of sensitivity and specificity values across studies were assessed. The SROC regression model was extended to adjust for clinical variable, and a dummy variable for type of diagnostic examination was added to enable a comparison of the tests. The effect of publication year, mean age, prevalence of PE, duration of tests and some specified quality criteria were investigated in a combined model of CT and V-P scanning. The following variables were analysed separately in each model, retaining the type of diagnostic test in each model: proportion of women, collimation, size of PE for helical CT model, and type of radionuclide used for V-P scanning. Studies of CT scanning published before 1995 were excluded from a further analysis, the results of which were compared with the analysis of all studies.

Results of the review
Twelve studies were included. Nine studies (n=520) assessed helical CT and 5 studies (n=1,269) assessed V-P scanning; 2 studies assessed both techniques.

Eleven studies were prospective in design and one was retrospective. Four of the prospective studies reported that the enrolment of patients was consecutive. Verification bias was judged to be likely in 4 studies, while interpretation bias was thought to be absent in all but 2 studies.

Accuracy of helical CT (n=9): the sensitivity ranged from 53 to 100% and the specificity from 75 to 100%. The pooled sensitivity was 86% (95% confidence interval, CI: 80, 92) and the pooled specificity was 94% (95% CI: 91, 96).

Accuracy of V-P scanning, threshold 1 (n=5): the specificity ranged from 96 to 100%; the range in sensitivity was not
reported. The pooled sensitivity was 39% (95% CI: 37, 41) and the pooled specificity was 97% (95% CI: 96, 98).

Accuracy of V-P scanning, threshold 2 (n=5): the sensitivity ranged from 54 to 100%; the range in specificity was not reported. The pooled sensitivity was 86% (95% CI: 83, 89) and the pooled specificity was 46% (95% CI: 44, 47).

Accuracy of V-P scanning, threshold 3 (n=5): the sensitivity ranged from 98 to 100%; the range in specificity was not reported. The pooled sensitivity was 98% (95% CI: 97, 100) and the pooled specificity was 4.8% (95% CI: 4.7, 4.9).

No significant predictors were found in the separate univariate analysis for CT or V-P scanning, and none remained significant in the final multivariate model that included the dummy variable for test type. The test comparison showed that helical CT was superior to V-P scanning at threshold 2 (P<0.001), and suggested that it was also superior at threshold 3 (P=0.05) and that there was no significant difference between helical CT and V-P scanning at threshold 1 (P=0.457). The sensitivity analysis, in which studies of CT published before 1995 were excluded, found similar results when these studies were included.

**Authors' conclusions**

Helical CT has greater discriminatory power than V-P scanning with normal and/or near normal threshold to exclude PE, while helical CT and V-P scanning with high-probability threshold have similar discriminatory power in the diagnosis of PE.

**CRD commentary**

This was a well conducted and reported review. It addressed a focused review question supported by clearly defined inclusion criteria. The literature search was reasonable but, since the review was limited to publications in English, language and publication bias are a potential problem. The authors acknowledged the possibility of publication bias. Several key quality items were assessed, and the effects of these were evaluated in the multivariate analysis. Some study details were tabulated, but further information on the types of patient included and the study designs of the included studies would have been helpful. The methods used to synthesise the results were appropriate but, given the small number of studies included, the value of the multivariate analysis is questionable. Although there was the potential for publication and language bias in this review, overall, the authors' conclusions are likely to be reliable.

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

**Practice:** The authors recommended several strategies. Confirm PE with moderate to high pre-test probability, and use either helical CT or V-P scanning, according to high-probability threshold. Exclude PE with low pre-test probability, as helical CT is a better test than V-P scanning. If helical CT is not available, V-P scanning with normal and/or near normal threshold could be an alternative technique. To exclude PE with moderate or high pre-test probability to confirm PE with low pre-test probability, avoid V-P scanning by using low-probability threshold.

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

**Bibliographic details**


**PubMedID**

15734930

**DOI**

10.1148/radiol.2343031009

**Original Paper URL**

http://radiology.rsna.org
Other publications of related interest
This additional published commentary may also be of interest. Kaatz S, Buckley J. Review: Helical CT has better discriminatory power than ventilation-perfusion scan to exclude pulmonary embolism. ACP J Club 2005;143:52.

Indexing Status
Subject indexing assigned by NLM

MeSH
Humans; Pulmonary Embolism /radiography /radionuclide imaging; ROC Curve; Tomography, Spiral Computed; Ventilation-Perfusion Ratio

AccessionNumber
12005009601

Date bibliographic record published
30/09/2006

Date abstract record published
30/09/2006

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.