Systematic review and meta-analysis of strategies for the diagnosis of suspected pulmonary embolism

Pierre-Marie Roy, Isabelle Colombet, Pierre Durieux, Gilles Chatellier, Hervé Sors, Guy Meyer

Abstract

Objectives To assess the likelihood ratios of diagnostic strategies for pulmonary embolism and to determine their clinical application according to pretest probability.


Study selection Studies that evaluated diagnostic tests for confirmation or exclusion of pulmonary embolism.

Data extracted Positive likelihood ratios for strategies that confirmed a diagnosis of pulmonary embolism and negative likelihood ratios for diagnostic strategies that excluded a diagnosis of pulmonary embolism.

Data synthesis 48 of 1012 articles were included. Positive likelihood ratios for diagnostic tests were: high probability ventilation perfusion lung scan 18.3 (95% confidence interval 10.3 to 32.5), spiral computed tomography 24.1 (12.4 to 46.7), and ultrasonography of leg veins 16.2 (5.6 to 46.7). In patients with a moderate or high pretest probability, these findings were associated with a greater than 85% post-test probability of pulmonary embolism. Negative likelihood ratios were: normal or near normal appearance on lung scan 0.05 (0.03 to 0.10), a negative result on spiral computed tomography along with a negative result on ultrasonography 0.04 (0.03 to 0.06), and a d-dimer concentration < 500 μg/l measured by quantitative enzyme linked immunosorbent assay 0.08 (0.04 to 0.18). In patients with a low or moderate pretest probability, these findings were associated with a post-test probability of pulmonary embolism below 5%. Spiral computed tomography alone, a low probability ventilation perfusion lung scan, magnetic resonance angiography, a quantitative latex d-dimer test, and haemagglutination d-dimers had higher negative likelihood ratios and can therefore only exclude pulmonary embolism in patients with a low pretest probability.

Conclusions The accuracy of tests for suspected pulmonary embolism varies greatly, but it is possible to estimate the range of pretest probabilities over which each test or strategy can confirm or rule out pulmonary embolism.

Introduction

Pulmonary embolism is a common and serious disease. Clinical signs and symptoms allow the clinician to determine the pretest probability of someone having pulmonary embolism (the clinical probability) but are insufficient to diagnose or rule out the condition. Laboratory tests and imaging are thus required in all patients with suspected pulmonary embolism. Since 1990 a large number of diagnostic tests and strategies have been evaluated for pulmonary embolism. As the design, clinical setting, and reference methods differ between studies, the diagnostic value of most tests may seem inconsistent. Although several reviews have been published on this topic, systematic reviews that may clarify the role of the different diagnostic tests are lacking.

We carried out a systematic review to assess the likelihood ratios of the diagnostic tests used for suspected pulmonary embolism. For clinical purposes, we estimated the range of pretest probabilities over which each test can accurately confirm or exclude pulmonary embolism.

Materials and methods

We searched Medline, Embase, and Pascal Biomed for studies published from January 1990 to September 2003 using the search terms ((pulmonary embol* or pulmonary thromboembol*) and (diagnosis or diagnostic) and (angiography or arteriography or (follow adj up) or followup or (management adj study)) and (PY = 1990-2003) and (study or studies or trial) and (LA = ENGLISH). We also manually searched published bibliographies and our own personal libraries. We retained only studies published in English. We excluded abstracts, editorials, reviews, case reports, and case series.

Data selection

Two reviewers (PMR, GM) independently selected potentially relevant studies. Studies were included if they evaluated tests or strategies aimed at confirming or excluding pulmonary embolism (confirmation or exclusion diagnostic strategies, respectively) and they met the following criteria: the reference method was pulmonary angiography for confirmation strategies and clinical follow-up or pulmonary angiography for exclusion strategies; the study was prospective; participants were recruited consecutively; and the test being evaluated and the reference test were interpreted independently.

We excluded retrospective studies; follow-up studies with more than 5% of patients lost to follow-up or those that used additional imaging to pulmonary angiography in patients with a negative experimental test result; studies in which crude data could not be extracted for the calculation of positive and negative likelihood ratios; and studies that had specific populations. Each study was graded according to the reference method and the characteristics of the patients (see table A on
For studies with multiple publications, we used data from the most recent publication.

Data extraction

Two investigators (PMR, GM) independently abstracted data on the design; study size; setting; characteristics of the patients; type of reference standard; and the number of true positive, true negative, false positive, and false negative test results. When we used follow-up as the reference method, we considered all the patients with a negative test result to have a false negative result if they developed deep vein thrombosis or pulmonary embolism during the three month follow-up period.

We classified deaths believed to be caused by pulmonary embolism as thromboembolic events. When we could not extract data from published articles, we contacted the authors. Discrepancies in data abstraction between investigators were resolved by a third author (PD).

Statistical analysis

We calculated the positive likelihood ratio for confirmation diagnostic strategies and the negative likelihood ratio for exclusion diagnostic strategies. We used the adjusted Wald method to calculate 95% confidence intervals. Summary estimates of the likelihood ratios were calculated as a weighted average, and we

<table>
<thead>
<tr>
<th>Table 1 Summary of studies evaluating tests or strategies aimed at confirming pulmonary embolism</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagnostic strategy</td>
</tr>
<tr>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>High probability ventilation-perfusion lung scan</td>
</tr>
<tr>
<td>Perfusion lung scan compatible with pulmonary embolism</td>
</tr>
<tr>
<td>Positive imaging result:</td>
</tr>
<tr>
<td>Spiral computed tomography</td>
</tr>
<tr>
<td>Leg vein ultrasonography</td>
</tr>
<tr>
<td>Magnetic resonance angiography</td>
</tr>
</tbody>
</table>

Studies of confirmation strategies

- Lung scan
  - PIOPED10
  - Miniati11

- Spiral computed tomography
  - Quandt12
  - Nilsson13
  - Remy-Jardin14
  - Remy-Jardin15
  - Van Rossum16
  - Stone17
  - Pooled positive likelihood ratio

- Leg vein ultrasonography
  - Quinn18
  - Van Beek19
  - Christiansen20
  - Turkstra21
  - Pooled positive likelihood ratio

- Echocardiography
  - Miniati22
  - Bova23
  - Pooled positive likelihood ratio

- Magnetic resonance angiography
  - Grist24
  - Loubeyre25
  - Meaney26
  - Gupta27
  - Oudkerk28
  - Pooled positive likelihood ratio

Fig 1 Positive likelihood ratios (squares) and 95% confidence intervals for strategies used to confirm a diagnosis of pulmonary embolism. Size of square is related to variance of study. Broken line represents pooled positive likelihood ratio, and limits of diamond represents 95% confidence intervals of pooled ratios
calculated the confidence intervals using the DerSimonian and Laird random effects method. Homogeneity tests were carried out to evaluate the consistency of findings across the studies. We used Cochran’s Q heterogeneity statistic and the quantity I² to determine the percentage of total variation across the studies due to heterogeneity rather than to chance. When I² was more than 0%, we explored possible reasons for heterogeneity, such as patient populations (selected or unselected patients) and the nature of the reference method (angiography or composite reference standard), using subgroup analysis based on the three categories for study quality (see table A on bmj.com).

Analyses were carried out in STATA (release 6).

Clinical practice perspectives

We considered that a confirmation strategy was accurate enough to diagnose pulmonary embolism when the post-test probability was above 85%, and that an exclusion strategy was accurate enough to exclude pulmonary embolism when the post-test probability was below 5%. We used Bayes’s theorem to calculate the probability of pulmonary embolism, conditioned by the likelihood ratio as a function of the pretest probability.

Results

We identified 1012 potentially eligible articles. After scanning the abstracts and titles we screened 93 for possible retrieval. We selected 66 articles for more detailed evaluation; 48 of these were included in the final analysis (see figure on bmj.com). The studies totalled 11 004 patients with suspected pulmonary embolism. The condition was confirmed in 3329 patients and excluded in 7675 (prevalence 30%). We did not analyse studies that used electron beam computed tomography as this technique is no longer used.  See tables B-D on bmj.com for characteristics of the included studies.

Confirmation diagnostic strategies

Table 1 and figure 1 summarise the confirmation diagnostic strategies and their pooled positive likelihood ratios.

Two studies evaluated lung scintigraphy. The prospective investigation of pulmonary embolism diagnosis study assessed the performances of ventilation and perfusion lung scans. Miniati et al studied the value of a perfusion lung scan without ventilation. We were unable to pool the results of these two studies as they used different diagnostic criteria and evaluated two different techniques.

We found significant heterogeneity among the five studies on magnetic resonance angiography.

Exclusion diagnostic strategies

Table 2 and figure 2 summarise the exclusion diagnostic strategies and their pooled negative likelihood ratios.

Nine studies analysed the value of a negative result on spiral computed tomography for excluding pulmonary embolism; however, one used a specific definition for negative results. We detected significant heterogeneity in the study group, but not in the two grade A studies.

We found heterogeneity in the group of ultrasonography studies. Five of the six studies were carried out in patients with a non-diagnostic ventilation and perfusion lung scan and one in patients selected on the basis of clinical probability and n-dimer testing. Wells et al studied the negative diagnostic value of serial ultrasonography after a non-diagnostic ventilation and perfusion lung scan.

Table 2 and figure 3 summarise the studies that evaluated n-dimers for the exclusion of pulmonary embolism (see also table D on bmj.com). In the analysis we included 12 studies that evaluated three different quantitative n-dimer enzyme linked immunosorbent assays, including two classic microplate methods and one rapid quantitative method.

### Table 2 Summary of studies evaluating tests or strategies aimed at excluding pulmonary embolism

<table>
<thead>
<tr>
<th>Diagnostic strategy</th>
<th>Study Grade</th>
<th>Heterogeneity</th>
<th>Pooled random negative likelihood ratio (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
<td>B</td>
<td>C</td>
</tr>
<tr>
<td>Normal or near normal radioisotopic lung scan</td>
<td>0</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Subgroup of grade B studies</td>
<td>5</td>
<td>1841</td>
<td>0.14</td>
</tr>
<tr>
<td>Subgroup of grade C studies</td>
<td>4</td>
<td>1329</td>
<td>0.90</td>
</tr>
<tr>
<td>Low probability ventilation perfusion radioisotopic lung scan</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Perfusion lung scan not compatible with pulmonary embolism</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Negative spiral computed tomography</td>
<td>2</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>Subgroup of grade A studies</td>
<td>1</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>Subgroup of grade C studies</td>
<td>2</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>Negative test results:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ultrasonography and spiral computed tomography</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Leg vein ultrasonography</td>
<td>0</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>Echocardiography</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Magnetic resonance angiography</td>
<td>0</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>Quantitative enzyme linked immunosorbent assay n-dimer &lt;500 μg/l</td>
<td>0</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>Subgroup of grade B studies</td>
<td>—</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Subgroup of grade C studies</td>
<td>—</td>
<td>—</td>
<td>7</td>
</tr>
<tr>
<td>Quantitative latex n-dimer test &lt;500 μg/l</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Semiquantitative latex n-dimer test &lt;500 μg/l</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Negative hemagglutination n-dimer test</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>
Studies of exclusion strategies

Abnormal lung scan
PIOPED (low probability)\(^5\)
Miniati (not compatible with pulmonary embolism)\(^7\)

Normal or near normal lung scan
PIOPED\(^10\)
Kruit\(^20\)
Van Beek\(^23\)
De Groot\(^33\)
Miron\(^32\)
Huil\(^33\)
Perrier\(^34\)
Wells\(^45\)
Leclercq\(^46\)
Pooled negative likelihood ratio

Spiral computed tomography
Quanadli\(^12\)
Nilsson\(^13\)
Van Strijen\(^37\)
Remy-Jardin\(^14\)
Remy-Jardin\(^15\)
Van Rossum\(^16\)
Ferretti\(^38\)
Ost\(^39\)
Stone\(^17\)
Pooled negative likelihood ratio

Spiral computed tomography and leg vein ultrasonography
Van Strijen\(^37\)
Musset\(^40\)
Perrier\(^41\)
Pooled negative likelihood ratio

Leg vein ultrasonography
Quinn\(^18\)
Van Beek\(^19\)
Christiansen\(^38\)
Turkstra\(^21\)
Perrier\(^42\)
Kruit\(^30\)
Pooled negative likelihood ratio

Magnetic resonance angiography
Grist\(^24\)
Loubeyre\(^25\)
Meaney\(^24\)
Gupta\(^27\)
Oudkerk\(^28\)
Pooled negative likelihood ratio

Echocardiography
Miniati\(^22\)
Bove\(^23\)
Pooled negative likelihood ratio

Fig 2 Negative likelihood ratios (squares) and 95% confidence intervals for strategies used to exclude a diagnosis of pulmonary embolism. Size of square related to variance of study. Broken line represents pooled negative likelihood ratio, and limits of diamond represents 95% confidence intervals of pooled ratios.
study used a different cut-off threshold so we excluded it from the calculation of summary negative likelihood ratios.36 We detected significant heterogeneity in the study group, but we found no heterogeneity in the grade B or grade C studies.

Studies that used seven different quantitative D-dimer latex agglutination assays met our inclusion criteria.13 36 40 49 50 53 55 Two studies evaluated several latex D-dimer tests in the same patients so we excluded them from the calculation of summary negative likelihood ratios.43 49 54 One study used a different cut-off value so we excluded that from the calculation of the summary negative likelihood ratio too.53 Three studies could be pooled.13 36 55

Two studies that evaluated a semiquantitative agglutination latex assay had significant heterogeneity.49 54 A whole blood agglutination D-dimer assay was evaluated in three studies, with no significant heterogeneity.31 35 51

Clinical practice perspectives
For each strategy we calculated the post-test probability as a function of the pretest probability (figs 4 and 5). For each diagnostic strategy we express the accuracy of diagnostic decisions as a function of the pretest probability (fig 6).

Relation to pretest probability
Confirmation of pulmonary embolism
In patients with a high pretest probability; a positive result with spiral computed tomography, ultrasonography, echocardiography, or magnetic resonance angiography; or a high probability ventilation perfusion lung scan are associated with a post-test probability of over 85%, allowing pulmonary embolism to be accurately diagnosed. Patients with a moderate pretest probability require additional imaging after a positive echocardiography result. In patients with a low pretest probability, the post-test probability was below 5%, and further investigations would not be needed to rule out pulmonary embolism.

Exclusion of pulmonary embolism
In patients with a low clinical probability; negative test results for D-dimers or with spiral computed tomography or magnetic resonance angiography; or a normal or low probability lung scan are associated with a post-test probability of below 5%. In this situation, additional testing would not be needed to rule out pulmonary embolism. Conversely, patients with a negative echocardiography result and a normal venous ultrasonography result would require additional testing to rule out pulmonary embolism.
embolism, even when the clinical probability was low. In patients with a moderate pretest probability, a negative quantitative D-dimer enzyme linked immunosorbent assay result, a normal or near normal lung scan, or a combination of normal spiral computed tomography results and normal venous ultrasonography results accurately exclude pulmonary embolism. In patients with a high pretest probability, the residual post-test probability remained above 5% for all diagnostic tests (fig 6). In these patients, additional testing would be required to confidently exclude pulmonary embolism.

Discussion

Large differences exist in the accuracy of diagnostic tests used to confirm or rule out pulmonary embolism. Ventilation perfusion lung scanning, spiral computed tomography, and ultrasonography of the leg veins all had positive likelihood ratios above 10. When these tests are positive in patients with a moderate or high clinical probability of pulmonary embolism they provide a post-test probability greater than 85%. A normal or near normal ventilation perfusion lung scan result, a combination of spiral computed tomography and ultrasonography, and quantitative D-dimer enzyme linked immunosorbent assay results had negative likelihood ratios below 0.10 and can exclude pulmonary embolism in patients with a low or moderate pretest probability. Spiral computed tomography alone, a low probability ventilation perfusion lung scan, magnetic resonance angiography, the latex Tinaquant D-dimer test, and the haemagglutination D-dimer test have higher negative likelihood ratios and can exclude pulmonary embolism only in patients with a low clinical probability. Echocardiography and ultrasonography seem unable to exclude pulmonary embolism.

The most straightforward approach for determining the accuracy of a diagnostic test is to carry out a cross sectional study in unselected patients, with independent, blinded assessments of test and reference methods. Our literature search identified only three studies that used such a stringent design in patients with suspected pulmonary embolism. Pulmonary angiography is the reference method for the diagnosis of pulmonary embolism, but it has the limitations of being an invasive procedure with associated risks, and physicians are reluctant to carry it out in all patients. Clinical follow-up of untreated patients with negative test results is considered a valuable alternative to this risky reference method, as the number of symptomatic thromboembolic events during a three month follow-up period without anticoagulant treatment reflects the number of false negative tests. Nevertheless, inclusion of follow-up studies in our analysis is associated with some drawbacks. Blinding is not maintained and some false negative test results may be undetected. In addition, in most of these studies a positive angiogram was not used to confirm the diagnosis of pulmonary embolism, and the rate of false positive test results may have been miscalculated. However, the criteria used to confirm pulmonary embolism (positive results with computed tomography or ultrasonography, high probability ventilation perfusion lung scan) are widely accepted.

We expressed test performance as likelihood ratios. According to Bayes’s theorem, the likelihood ratio indicates the extent of change in the odds of disease after a test result. Likelihood ratios can be calculated irrespective of the format of the
Fig 5  Post-test probability according to pre-test probability and pooled values (solid line) or limit of 95% confidence intervals (broken lines) of the negative likelihood ratio
test result: dichotomous (spiral computed tomography, ultrasonography, magnetic resonance angiography, qualitative D-dimer test), ordinal (ventilation perfusion lung scan), or continuous (quantitative D-dimer tests).

Some of our findings are limited by the low number of studies meeting our quality criteria, especially those of lung scanning. Systematic reviews of diagnostic studies are hampered by the heterogeneity of results, even when attempts to define the most homogeneous set of studies are made, as in our study. In studies dealing with qualitative D-dimer enzyme linked immunosorbent assay results, the negative likelihood ratio of grade B studies were homogeneous but lower than those of grade C studies. This discrepancy may be explained by differences in the study population and by the use of pulmonary angiography as the reference method in most grade C studies as opposed to follow-up in most grade B studies. Pulmonary angiography is difficult to interpret in patients selected on the basis of a previous lung scan result (grade C studies), leading to a high risk of misclassification. These patients are also likely to have small pulmonary emboli, which may increase the rate of false negative D-dimer test results.

As proposed by Kearon, we assumed that the diagnosis of pulmonary embolism was accurate when the post-test probability was above 85% and that pulmonary embolism could be safely ruled out when the risk of venous thromboembolism was below 5%. We defined a range of pretest probabilities at which each test can confirm or rule out pulmonary embolism, with an acceptable risk of misdiagnosis (fig 6). As a general rule, the results suggest that discordance between clinical probability and the diagnostic test result requires additional studies.

Our findings allow the calculation of the post-test probability of pulmonary embolism provided that the pretest probability has been estimated before the test. Our results also suggest that the performances of several diagnostic tests remain poorly defined and need additional evaluation.

Contributors: GM had the idea for the study; he is guarantor for the paper. PMR, GM, and PD designed the study. PMR and GM assessed the studies for inclusion, extracted the data, and drafted the paper. IC, GC, and HS contributed to data extraction. All authors revised the paper critically and approved the final manuscript.

Ethical approval: Not required.

Competing interests: None declared.

Funding: None.

Accepted 3 December 2004. Published Online First on 28 July 2005.

---

**What is already known on this topic**

The accuracy of diagnostic tests for suspected pulmonary embolism varies largely between studies, and the appropriate clinical setting for their use is unclear.

**What this study adds**

When the clinical probability is moderate or high, pulmonary embolism is confirmed by a high probability lung scan and a positive result on spiral computed tomography or venous ultrasonography.

When clinical probability is low, these results require confirmation by pulmonary angiography.

In patients with a low or moderate clinical probability, the condition can be excluded by a negative quantitative D-dimer test result, a normal or near normal lung scan, or normal findings on spiral computed tomography and venous ultrasonography.

When clinical probability is high, these results require confirmation by pulmonary angiography.


