Beyond diagnostic accuracy. Applying and extending methods for diagnostic test research
Glas, A.S.

Citation for published version (APA):
Abstract

Introduction - It is an empirical question to what extent physicians adjust their initial uncertainty after diagnostic information becomes available. The aim of this study is to investigate the interpretation of V/Q lung scan results by physicians in the diagnosis of pulmonary embolism (PE).

Methods - Data were obtained in a prospective multi-center study in patients with suspected PE. Treating physicians were asked to assign the clinical probability of PE before and after V/Q scans were performed. This made it possible to calculate the subjective likelihood ratio (LR) of scan results. The objective LR was calculated from a blinded comparison of test results with a conjoint reference standard (independent interpretation of V/Q scan and pulmonary angiography). Calibration curves were estimated using the clinical probability estimates.

Results - The mean subjective LR of a normal V/Q scan result was 0.056 (95% CI: 0.038 to 0.082) compared to an objective LR of 0.046 for the same result (95% CI: 0.012 to 0.19). These values were 26 (CI:16 to 41) versus 20 (CI: 8.8 to 43) for a high probability scan and 1.1 (CI: 0.77 to 1.5) versus 0.43 (CI: 0.31 to 0.62) for a non-diagnostic scan. No significant center differences were found for non-diagnostic scan results, although centers more or less adjust for center differences in the diagnostic worth of a non-diagnostic scan result. Overall calibration was good.

Conclusion - In general physicians are well calibrated and they update prior probabilities according to the inherent information in a normal and high-probability lung scan result. A non-diagnostic scan is interpreted inconsistent to its diagnostic worth.
**Introduction**

The diagnosis of pulmonary embolism represents a clinical challenge, due to the limited specificity of presenting signs and symptoms and the less-than-perfect accuracy of additional diagnostic tests. A number of research efforts have aimed at the development of safe and cost-effective diagnostic strategies for suspected pulmonary embolism.\(^{14}\) In many of these strategies, the ventilation-perfusion (V/Q) lung scan plays a crucial role, despite its limited overall sensitivity and specificity.\(^{13}\) The hitch of the lung scan lies in the non-diagnostic test result, obtained in about half of the patients suspected for pulmonary embolism, of which 10 to 30% actually have the disease when subjected to further testing.\(^6\)

The limitations of all imaging modalities for pulmonary embolism - lung scanning, spiral computerized tomography, pulmonary angiography and venous leg imaging - have necessitated the use of these tests in algorithms and in combination with clinical pretest probability assessment in diagnostic management algorithms\(^2\). Rodger and Wells considered clinicians to be adept at assigning such pretest probabilities after overall clinical assessment, and incorporating them in the further workup of pulmonary embolism suspected patients. The clinical strength of suspicion after clinical assessment will vary, depending on the patient's complaints, risk factors, and the likelihood of other conditions. Yet pretest probabilities are susceptible to error.\(^7\) Multiple formal clinical models, explicit prediction and decision rules have been developed to improve the accuracy of pretest probability assessment.\(^5\)\(^6\)\(^8\)

Expressing one's clinical strength of suspicion in probability form can be done both before and after testing. Bayes' theorem puts a formal link between these two probabilities: the posttest odds should equal the pretest odds for pulmonary embolism multiplied by the likelihood ratio of the test result. The likelihood ratio of a test result expresses how much more likely that test result is in patients with pulmonary embolism compared to those without it.\(^9\)

It remains an empirical question how well physicians adjust their uncertainty based on the information in the test result. Highly informative test results, with likelihood ratios that are either close to zero or far above unity, should lead to a more pronounced reduction in uncertainty compared to test results with likelihood ratios closer to unity. In addition, the reduction in uncertainty should be proportional to the strength of the prior suspicion. Imperfect pre-test probabilities and incorrect likelihood ratios can lead to flawed post-test probabilities and, subsequently, to faulty therapy decisions.\(^10\)

The purpose of this study was to investigate the judgment of physicians in interpreting lung scan results in the diagnosis of pulmonary embolism, to see to
what degree they update prior probabilities. A consecutive series of patients with a clinical suspicion for pulmonary embolism underwent a standard diagnostic algorithm, starting with V/Q scanning and followed, if necessary, by further testing. Physicians could express pre-test and post-test probabilities of pulmonary embolism on a visual analogue scale, immediately before and after lung scanning. The calibration of respective probabilities were evaluated and the physician's subjective likelihood ratios - obtained from the pre and posttest probabilities - were compared with the objective likelihood ratios of the V/Q lung scan - obtained by comparing the lung scan results with the final results of the reference standard.

Patients and methods

Patients
Data were obtained as part of a prospective study in six Dutch teaching hospitals, reported in detail elsewhere. From May 1997 through March 1998, consecutive in- and outpatients with a clinical suspicion of pulmonary embolism (PE) were eligible for the study. Patients were excluded if they were younger than 18 years of age, were pregnant, had already undergone objective diagnostic tests for PE, had an indication for acute thrombolytic therapy, or if there was an expected inability to complete the study protocol within 48 hours of presentation. The Institutional Review Boards of all participating hospitals had approved the study protocol. Informed and consenting patients were included in the study.

Diagnostic Investigations
Within 24 hours of study inclusion a detailed medical history, compression ultrasonography and a ventilation-perfusion (V/Q) lung scan were performed in all patients. Spiral CT was only performed in patients with an abnormal perfusion scan. Pulmonary angiography was performed in patients with a non-diagnostic lung scan and if a high probability lung scan was followed by a normal spiral CT. The maximum allowed time for completion of this diagnostic protocol was 48 hours.

A six-view perfusion lung scintigraphy was obtained after the administration of 100 Mbq of 99mTechnetium-labelled macro-aggregates of albumin. If segmental or larger perfusion defects were seen, ventilation lung scintigraphy was added using 81mKrypton gas. In each center, experienced nuclear medicine physicians interpreted the lung scans by using a lung segment reference chart. Lung scans were classified according to previously described criteria as either normal (no perfusion defects), high probability (at least one segmental or larger perfusion
defect with locally normal ventilation) or non-diagnostic (ventilation-perfusion abnormalities not meeting the criteria for normal or high probability) (12, 13). Lung scans of poor quality due to technical failure were classified as non-interpretable.

Pulmonary angiography was performed using a digital subtraction technique with the catheter positioned selectively in the right and left pulmonary artery and was interpreted independently by two radiologists according to accepted criteria. If there was no consensus, the independent interpretation of a third reader was considered decisive.

A central panel re-interpreted independently all lung scans and pulmonary angiograms, without knowledge of the outcomes of ultrasonography, locally interpreted V/Q scan or, if applicable, the spiral CT. The panel’s conclusion was used as the reference standard. Patients were classified as having PE in case of a high probability lung scan or abnormal angiogram, while the diagnosis was refuted if the lung scan or angiogram was normal. Only those patients in whom the diagnosis of PE was confirmed or refuted per-protocol were included in the analysis.

**Probability estimates**

Upon presentation of a patient and prior to all other diagnostic investigations, the treating physician was asked to assign a clinical probability of pulmonary embolism using a non-linear, logistic visual analog scale, ranging from 0 to 100 percent. This estimate could solely be based on symptoms and signs and results of routine tests (arterial blood gas analysis, electrocardiogram and chest X-ray), if available. A second scale, identical to the first, was used in assigning a post-test clinical probability, after the results of the locally interpreted V/Q scan had become available. The probability estimates were not communicated to the central panel.

**Data analysis**

If subjective probabilities are well calibrated the assigned probability of PE should equal the observed proportion of patients with PE. For example, of all patients assigned a 10% probability of PE, 10% is expected to harbor a pulmonary embolism.

To study calibration, the probability estimates were ranked and assigned to ten groups, based on deciles. Calibration curves were drawn for the pre-test and post-test probabilities, plotting the average probability as assigned by the physicians in each of the ten groups against the proportion of patients with a pulmonary embolism.

Likelihood ratios of the results of the V/Q scan were estimated by comparing
the results of the locally interpreted V/Q scan with the results from the conjoint reference standard, as assigned by the panel. For the purpose of this paper, these likelihood ratios will be referred to as objective likelihood ratios. 95% Confidence intervals for the objective likelihood ratio were calculated using Simel’s method.16

For each pair of a pre-test and post-test probability, assigned to the same patient, a subjective likelihood ratio was calculated by transforming the pre-test and post-test probability into odds, and taking the ratio of post-test odds against pre-test odds. To avoid undefined results probabilities of 0 and 100 were replaced by probabilities of 0.001 and 99.999 respectively. Since the estimated subjective likelihood ratios do not follow a normal distribution, a logarithmic transformation was performed. The average subjective likelihood ratio of each test result of the V/Q scan was then calculated as the back-transformed (geometric) mean of the logarithmically transformed subjective likelihood ratios. Confidence intervals (CI) were calculated on a log-scale as mean lnLR ±1.96*SE_{\text{mean}} and then back-transformed to the original scale.

To explore whether physicians modified for center differences in (objective) likelihood ratio, we compared them with the subjective likelihood ratio per center. All computations were done with the statistical software package SPSS version 10.

Figure 1 Patient flow-chart. *) Patients were excluded for the following reasons: protocol expected to be >48 hours n=104, prior diagnostic testing n=43, age <18 years n=16, pregnancy n=11, indication for acute thrombolytic therapy n=5. 1) Reasons pre-test probability not available: physician unreachable n=31, result lung scan already available n=27, physician refusal n=7, not specified n=39. Reasons post-test probability not available: physician unreachable n=11, result angiography available n=2, not specified n=19, not applicable n=106.
### Results

A total of 1162 consecutive patients with clinically suspected pulmonary embolism were screened, of whom 179 had to be excluded on the basis of one of the predefined criteria. (See flowchart, figure 1.) Of the 983 eligible patients, 627 (64%) gave written informed consent. A reference diagnosis regarding the absence or presence of pulmonary embolism could not be reached in 110 of these 627 patients, due to withdrawal of informed consent, clear evidence for an alternative diagnosis, medical reasons, or technical failure. For 242 of the 517 remaining patients a pretest or posttest probability was missing.

The diagnosis of pulmonary embolism was eventually confirmed in 114 patients of the remaining 275 (41%). In this group, 23% had a normal scan (n=63), 43% a non-diagnostic scan (n=119), 32% a high probability scan (n=89) and 1.5% a non-interpretable scan (n=4). The prevalence of pulmonary embolism in these 4 groups was 2/63=0.032, 28/119=0.24, 83/89=0.93 and 1/4=0.25 respectively. A total of 140 different physicians were involved in assigning the probabilities, of which 10% were estimated by specialists. Table 1 presents the characteristics of the patients used in the analyses.

<table>
<thead>
<tr>
<th>Table 1. Clinical and demographical characteristics.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Study patients</strong></td>
</tr>
<tr>
<td>(n=627)</td>
</tr>
<tr>
<td>Males</td>
</tr>
<tr>
<td>270 (43%)</td>
</tr>
<tr>
<td>Mean age, years (SD)</td>
</tr>
<tr>
<td>53 (18)</td>
</tr>
<tr>
<td>Outpatients</td>
</tr>
<tr>
<td>490 (78%)</td>
</tr>
<tr>
<td><strong>Median duration of symptoms, days</strong></td>
</tr>
<tr>
<td>(IQR)</td>
</tr>
<tr>
<td>3 (1-9)</td>
</tr>
<tr>
<td>Symptoms of DVT†</td>
</tr>
<tr>
<td>43/624 (6.9%)</td>
</tr>
<tr>
<td>Surgery</td>
</tr>
<tr>
<td>111/626 (18%)</td>
</tr>
<tr>
<td>Trauma</td>
</tr>
<tr>
<td>38/623 (6.1%)</td>
</tr>
<tr>
<td>Immobilization</td>
</tr>
<tr>
<td>100/622 (16%)</td>
</tr>
<tr>
<td>Previous VTE †</td>
</tr>
<tr>
<td>98/622 (16%)</td>
</tr>
<tr>
<td>Family history of VTE</td>
</tr>
<tr>
<td>122/551 (20%)</td>
</tr>
</tbody>
</table>

*) Patients of which data available for calculation of the subjective LR.
† DVT = deep vein thrombosis
‡ VTE = venous thromboembolism
Probabilities and calibration

The pre-test probability estimates were distributed across the whole range, with an inter-quartile range (IQR) from 30% to 75%. The overall calibration for the pre-test probability was good but not perfect, as the probabilities do not tend to match the observed proportions of pulmonary embolism. The calibration curve shows that physicians tended to overestimate the risk of pulmonary embolism when they assigned higher probabilities. (Figure 2).

Figure 3 shows scatter plots of the pre-test versus post-test probability estimate for the three local lung scan result categories. Overall low post-test probabilities were assigned to patients with a normal scan (IQR: 1% to 5%) and high probabilities to patients with a high-probability scan (IQR: 87% to 100%). The probabilities after a non-diagnostic scan show far more variability, with a substantial number of patients with a very low or high post-test probability estimate of pulmonary embolism (IQR 30% to 80%). In some patients a high pre-test probability was followed by a very low post-test probability after a non-diagnostic scan. In others, the reverse could be observed. This implies that not all physicians interpret a non-diagnostic lung scan result in an identical way.

The calibration curve for the post-test probabilities shows good calibration for the categories 0-25 and 75-100%. (Figure 4) These categories consist mainly of patients with a normal or high probability V/Q scan result. Physicians were less well calibrated when assigning post-test probabilities to patients with a non-diagnostic scan, most of which ended up in the middle range of the probability scale. No significant differences were found in the mean pre-and post probabilities between residents and specialists.

![Figure 2: Calibration curve comparing physician’ pre-test probability estimate with the actual rate of pulmonary embolism (PE) for the 275 patients. For each point estimate, the 95% confidence interval is given. The dashed line indicates perfect calibration between observed rate of pulmonary embolism and the pre-test probability estimate.](image-url)
Physicians’ interpretation of ventilation-perfusion lung scans

Figure 3 Scatter plots of pre-test and post-test probability estimates per test result for an overall of 271 patients. A) normal scan, B) non-diagnostic scan and C) high probability scan. The circles are from patients with confirmed pulmonary embolism and the cubes are from patients with no pulmonary embolism. (Non-interpretable scan, n=4, not shown)

Figure 4 Calibration curve comparing physician post-test probability estimate with the actual rate of pulmonary embolism (PE) for the 275 patients. For each point estimate, the 95% confidence interval is given. The dashed line indicates perfect calibration between observed rate of pulmonary embolism and the post-test probability estimate. Note that this figure consists of 9 groups instead of the 10 groups, as two deciles had to be combined in a single group.
Subjective likelihood ratio versus objective likelihood ratio

We compared the mean subjective likelihood ratios, calculated from pairs of post-test and pre-test probabilities, with the objective likelihood ratios, as obtained from a comparison of test results with those from the reference standard. (Table 2) The objective likelihood ratio for a normal scan was 0.046 (95% CI: 0.012 to 0.19). The mean of the subjective likelihood ratio was almost identical: 0.056 (CI: 0.038 to 0.082). For a high probability scan both likelihood ratios were far from unity: 20 (CI: 8.8 to 43) (objective) versus 26 (CI: 16 to 41) (subjective). The objective likelihood ratio for a non-diagnostic scan was 0.43 (CI: 0.31 to 0.62), significantly lower compared to the subjective likelihood ratio of 1.1 (CI: 0.77 to 1.5) for this result (p<0.01).

Table 2 The geometric mean of the subjective likelihood ratio versus the objective likelihood ratio per category of V/Q test result.

<table>
<thead>
<tr>
<th></th>
<th>Subjective Likelihood Ratio</th>
<th>Objective Likelihood Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>LR</td>
</tr>
<tr>
<td>Normal</td>
<td>63</td>
<td>0.056</td>
</tr>
<tr>
<td>Non-diagnostic</td>
<td>119</td>
<td>1.1</td>
</tr>
<tr>
<td>High probability</td>
<td>89</td>
<td>26</td>
</tr>
<tr>
<td>Non-interpretable</td>
<td>4</td>
<td>0.75</td>
</tr>
</tbody>
</table>

The tendency in physicians to judge a non-diagnostic scan as literally being non-informative - although the objective LR suggests that it should lower the strength of pulmonary embolism suspicion - can also observed from figure 5, which illustrates the likelihood ratios for a non-diagnostic scan per center. Though not significant it seems that the subjective LR overall is higher than the objective LR. In center 1 the subjective likelihood ratio was significantly lower than unity (0.54 (CI: 0.33 to 0.90)) comparable to its corresponding objective likelihood ratios (0.42 (CI: 0.22 to 0.79)). This result shows that these physicians adjust for center-specific differences in definitions of the category non-diagnostic test result. The subjective likelihood ratios in the other centers ranged from 0.77 to 9.4, with objective likelihood ratios ranging from 0.16 to 0.72. There were no statistically significant differences between residents and specialists.
Discussion

This study evaluated physicians interpreting ventilation/perfusion lung scan results in the workup of patients with suspected pulmonary embolism. The pre-test probabilities of pulmonary embolism, based on information from history, physical and routine testing, varied widely but were well calibrated. Pre-test probabilities were appropriately updated if the results of the lung scan were either negative (normal scan) or clearly positive (high-probability scan). Substantially more variability was observed for non-diagnostic scans, sometimes leading to an increased and on other occasions to a decreased level of suspicion.

The pre-test calibration curve observed here is rather typical for physicians who in general underestimate the prevalence of pulmonary embolism in the low suspicion patients and overestimate the risk in patients with a high suspicion of pulmonary embolism. The calibration for the post-test probabilities was almost perfect for clearly negative or positive results of the lung scan.

A subjective likelihood ratio is an expression of the diagnostic impact of a lung scan result. For the normal and high probability scan the subjective likelihood ratios were similar to the objective likelihood ratios, indicating that the various physicians in this study updated the pre-test probabilities according to the diagnostic worth of these scan results. This was not the case for the non-diagnostic scan result. The mean subjective likelihood ratio was close to unity, indicating that on average physicians remain uncertain of the diagnosis of pulmonary embolisms. Yet in reality the non-diagnostic scan has diagnostic worth, as expressed by the objective likelihood ratio, which was significantly lower than one.

Variability in the probabilities attached to a non-diagnostic result could not be explained by differences between the centers. Here also the subjective
likelihood ratio tended to be higher than the objective likelihood ratio. Only for center 1, the subjective likelihood was consistent with the objective likelihood ratio. Both were significantly less than unity, indicating less pulmonary embolism patients were assigned a non-diagnostic scan result than non-pulmonary embolism patients. In this center physicians, knowing of earlier cases, did assign weight to a non-diagnostic result.

Overall one can say that physicians intuitively updated probabilities in accordance with the diagnostic worth of the lung scan. This finding underscores Frybacks view that subjective probabilities are not just guesses, but the result of incorporating all relevant information with the appropriate weight into an opinion.21 The physicians' judgment refers to an intellectual process with incorporation of all subjective and objective evidence with their appropriate weight using known physiological and anatomical models relevant for the disease in question. Nevertheless the conclusion that the physicians in this study properly adjust their uncertainty can be challenged. Maybe they did so explicitly. Lung scanning is not a new technique and it is possible that some physicians were using previously published likelihood ratios. On the other hand, the quality of their judgment may well be due to the large amount of short-term feedback they receive of the final pulmonary embolism diagnosis. Such feedback mechanisms make physicians perform better on familiar tests compared to novel tests.22

Probability estimates were available for only 275 patients. Although, in principle, it is possible that physicians refrained from assigning probabilities to more difficult patients, we do not think that this selection has caused a bias. On the contrary almost all patients for which the reason for not assigning a posttest probability was 'not applicable' had a normal scan. Possible in these cases a pulmonary embolism as the cause for the complaints was refuted and another diagnosis was considered giving the idea to the physicians that probabilities were not necessary anymore. In one center 51% of the pre-test and 63% of the post-test probability expressions were missing, indicative of the difficulties in obtaining probability expressions. The prevalence of pulmonary embolism in this particular group of patients with a normal scan as well as the overall prevalence are consequently overestimated. Yet we think this did not affect our conclusions. Comparing the subjective likelihood ratio to the objective likelihood ratio based on the 517 patients who had a final diagnosis gave the same results.

A number of studies have previously reported objective likelihood ratios for lung scan results. The objective likelihood ratios that can be derived from the PIOPED study are comparable to the ones obtained in the study reported here.6,23

A small series of studies looked at the interpretation of lung scans by physicians.
One study found that a normal scan was misinterpreted: 31% of physicians remained uncertain about the diagnosis. In another study by the same authors, published in 1998, physicians were reported to interpret lung scans quite well if odds were used to express the likelihood of pulmonary embolism. The authors’ finding that physicians adjusted the odds of pulmonary embolism according to the pre-test probability is in line with our findings. A recent study found a good agreement between 3 observers in assigning probabilities of pulmonary embolism after evaluating lung scans and clinical information.

Are subjective probability estimates to be preferred over decision rules within strategies for diagnosing pulmonary embolism? It is plausible to say that objective methods do not suffer from the variation that does exist between physicians. Nevertheless it is also known that in general decision rules do not travel well between populations and poor implementation may hamper the accuracy of formal decision rules. In addition various authors confirmed that (aided) subjective probability estimates perform as well as decision rules. With respect to strategies for diagnosing pulmonary embolism it has been shown that two formal decision rules expanded by Wells performed poorly. Another study showed that a decision rule performed more accurate when combined with clinical judgment. Although physicians’ were reasonable calibrated we did not study the discriminating abilities of their assigned pre-test probability of pulmonary embolism. Considering these arguments subjective probability assessment may be the right choice for use within diagnostic strategies for pulmonary embolism.

In conclusion we can say that pre-test probability estimates are properly adjusted to well-calibrated post-test probabilities by physicians responsible for the workup of patients suspected for pulmonary embolism. Physicians tend to interpret a non-diagnostic scan as a indifferent result despite the diagnostic worth it has in reality.

**Acknowledgments**

This study was supported from a grant from the Dutch Health Insurance Council.

References

Physicians’ interpretation of ventilation-perfusion lung scans
