Comparison of a Clinical Probability Estimate and Two Clinical Models in Patients with Suspected Pulmonary Embolism

Bernd-Jan Sanson1, Jeroen G. Lijmer2, Melvin R. Mac Gillavry3, Franktien Turkstra1,3, Martin H. Prins2, Harry R. Bül1
on behalf of the ANTELOPE-Study Group**

From the 1Department of Vascular Medicine and 2Department of Clinical Epidemiology and Biostatistics, Academic Medical Center, 3Department of Internal Medicine, Slotervaart Hospital, Amsterdam, The Netherlands

Key words

Pulmonary embolism, diagnosis, probability, estimate, model

Summary

Recent studies have suggested that both the subjective judgement of a physician and standardized clinical models can be helpful in the estimation of the probability of the disease in patients with suspected pulmonary embolism (PE). We performed a multi-center study in consecutive in- and outpatients with suspected PE to compare the potential diagnostic utility of these methods. Of the 517 study patients, 160 (31%) were classified as having PE. Of these patients, 14% had a low probability as estimated by the treating physician, while 25 to 36% were categorized as having a low clinical probability with the use of two previously described clinical models. The objectively confirmed prevalence of PE in these three low probability categories was 19%, 28% and 28%, respectively. The three methods yielded comparable predictive values for PE in the other probability categories. We conclude that a physician’s clinical judgement alone and two standardized clinical models, although comparable, perform disappointingly in categorizing the pre-test probability in patients with suspected PE.

Introduction

The introduction of objective testing for the diagnosis of pulmonary embolism dates to the early 1970’s. Before this time the diagnosis was usually based on clinical signs and symptoms alone. The systematic use of objective methods revealed that the clinical diagnosis of pulmonary embolism was highly inaccurate, with only one third of patients with a clinical suspicion actually having the disease (1, 2). Based on these observations, it was recommended that clinical signs and symptoms only be used for raising the suspicion of pulmonary embolism, after which the objective diagnostic work-up needed to be implemented.

In recent years there has been a revived interest in the potential role of clinical judgement in the diagnosis of pulmonary embolism. The PIOPED investigators showed that clinicians were able to categorize patients, with a reasonable degree of accuracy, into groups with a low, moderate and high probability of the disease using clinical judgement alone (3). Perrier and colleagues have recently shown that such a probability estimate appears clinically useful in selecting a subgroup of patients in which further diagnostic testing can be withheld after a non-diagnostic ventilation-perfusion lung scan (4). A potential drawback of the estimate is that it is subjective and may therefore show variation depending on experience and training of the physician which may hamper its generalizability.

Wells and colleagues recently developed a clinical model that uses information from the medical history and physical examination in a structured algorithm in order to come to a classification of the probability of pulmonary embolism (5). This model was subsequently simplified to a limited number of questions (6). These investigators suggested that it appears safe to withhold further diagnostic testing in patients with a non-diagnostic scan and a low or moderate probability according to the clinical model (5). The presumed advantage of the clinical models over the clinical probability estimate is that they may have less observer variability due to the fact that they are based on defined, relatively objective clinical findings (e.g. dyspnea, pleuritic chest pain, heart rate, leg symptoms) and history.

Before either the clinical probability estimate or the clinical models can be widely advocated for routine clinical practice, it is necessary that the utility of these methods be confirmed. In addition, it is well known that most of the available diagnostic techniques are unable to detect smaller subsegmental pulmonary emboli, which may have a different natural history regarding mortality and risk of recurrence, than larger emboli. Whether this is also true for the clinical rules is unknown.

The objective of this multi-center study was to assess and compare the accuracy and variability of both the clinical probability estimate by the attending physician as performed in the PIOPED study and by Perrier and colleagues, and the two clinical models previously described by Wells and colleagues, in a large series of consecutive patients with clinically suspected pulmonary embolism. Furthermore, we investigated the segmental or subsegmental localization of pulmonary emboli in patients categorized as having a low probability.

Patients and Methods

Patients

Six Dutch teaching hospitals participated in a large-scale study, which assessed various diagnostic methods for pulmonary embolism. From May 1997
through March 1998 consecutive in- and outpatients presenting with clinically suspected pulmonary embolism were potentially eligible. Patients were excluded if they were younger than 18 years of age, were pregnant, had an indication for acute thrombolytic therapy, had already undergone objective diagnostic testing for venous thromboembolism, or if there was an expected inability to complete the diagnostic protocol within 48 h of presentation. The latter applied to patients presenting during holidays and long weekends when not all diagnostic facilities were in operation. The Institutional Review Boards of all centers approved the study protocol and informed consent was obtained from all participants.

Clinical Probability and Clinical Models

Upon referral, the attending physician was asked to give an estimate of the clinical probability of pulmonary embolism based on information from the medical history, physical examination and if available chest X-ray, ECG and bloodgas analysis. This was done on a visual analogue scale of 0 to 100 percent. Directly after inclusion, the study physician obtained all relevant information for the models on signs and symptoms and the medical history using a standardized case record form. After conclusion of the study, the result (low, moderate or high probability) of both clinical models was calculated from the obtained information, as described by Wells and colleagues.

Briefly, the simplified model contains seven variables: 1) clinical symptoms of deep vein thrombosis (3.0 points), 2) cancer (1.0 point), 3) heart rate >100 b.p.m. (1.5 points), 4) immobilization or surgery in the previous 4 weeks (1.5 points), 5) previous deep vein thrombosis and/or pulmonary embolism (1.5 points), 6) hemoptysis (1.0 point), 7) no alternative diagnosis for the presenting chest symptoms (3.0 points). Patients scoring less than a total of 15 points were assigned a low, those with between 2 and 6 points a moderate, and those with more than 6 points a high probability for pulmonary embolism (6).

The extended clinical model consists of an algorithm by which the presence of severe, typical or atypical symptoms according to pre-defined criteria, the consideration of an alternative diagnosis, and the presence of established risk factors for venous thromboembolism together results in a classification of a low, moderate or high probability of pulmonary embolism (5). A slightly modified version of this model was performed since the original contains information concerning the blood oxygen saturation and the ECG, which were not collected in this study. Besides disregarding these two variables, the model was performed according to the same methods as Wells and colleagues. This modification can theoretically lead only to an underestimation of the probability derived from the model.

Diagnostic Investigations

A six-view perfusion lung scintigraphy was performed within 24 h of referral using 50-100 MBq of 99mTc-labeled macroaggregates of albumin. If at least one segmental or larger perfusion defect was seen, ventilation scintigraphy, using 133Xe-Krypton gas, was performed. Ventilation-perfusion scans were interpreted by two experienced nuclear medicine physicians using a lung segment reference chart (7). In case of disagreement, the interpretation of a third was decisive. Lung scans were reported as being normal (no perfusion defects), high probability (at least one segmental or larger perfusion defect with local normal ventilation), or non-diagnostic (ventilation-perfusion defects not qualifying as high probability) according to previously described criteria (8). Pulmonary angiography was indicated within 24 h of the lung scan in all patients with a non-diagnostic result. Angiography was performed using standard techniques and was interpreted according to accepted criteria (9, 10). Agreement by two independent radiologists on the result was required. If there was no consensus, a third radiologist was consulted whose judgement was decisive. In all patients with a non-diagnostic or high-probability ventilation-perfusion lung scan, spiral computed tomographic (CT) angiography was performed as previously described (11). In the case of a high-probability lung scan and a normal spiral CT scan, pulmonary angiography was performed.

Pulmonary embolism was considered absent in case of a normal perfusion scan or a normal pulmonary angiography and present in case of a high-probability ventilation-perfusion scan or an abnormal angiography. In all patients categorized as having pulmonary embolism, the largest vessel in which thrombus was visible was scored using the angiography or the spiral CT scan. Emboli were thus divided into those visible in segmental or larger pulmonary arteries and those visible only in subsegmental arteries.

Statistical Analysis

The analysis was performed in the patient group which had the diagnosis of pulmonary embolism established or refuted as defined in the protocol. The clinical probability estimates given by the treating physicians were divided into three categories: less than 20%, 20 to 80% and greater than 80%. These cut-off points were determined on the basis of previous literature (3, 4). The proportion of patients with pulmonary embolism was calculated in each of the three categories of the clinical probability estimates and in each of the three categories of the two clinical models. The likelihood ratio of each result category of the clinical probability estimate and the clinical models was calculated. For each parameter estimate 95% confidence intervals were calculated according to the normal approximation of the binomial distribution. Furthermore, the frequency of segmental (or larger) and subsegmental pulmonary emboli was calculated in the patients with the disease in the low probability category according to the three methods. Finally, observer variability between the six centers was analyzed for the three methods using the Chi-square test (SPSS 8.0).

Results

Patients

A total of 1162 patients with clinically suspected pulmonary embolism were screened. Of these, 179 patients were excluded for the following reasons: expected inability to complete the protocol within 48 h (104), diagnostic testing already performed (43), age less than 18 years (16), pregnancy (11), indication for acute thrombolytic therapy (5). A total of 983 patients were therefore eligible for inclusion in the study of whom 627 (64%) gave informed consent.

Of the 627 participants, 517 (82%) had the diagnosis of pulmonary embolism either established or refuted as defined in the protocol. A final diagnosis was not reached in the remaining 110 patients because of withdrawal of informed consent, clear evidence for an alternative diagnosis for the presenting symptoms, medical reasons or technical failure. The baseline clinical characteristics of the 517 study patients were similar to those of the 110 excluded patients (Table 1). However, these patients were slightly older and more frequently had co-morbid conditions.

| Table 1 | Baseline clinical characteristics of the 517 patients with clinically suspected pulmonary embolism in whom a final diagnosis was obtained, as well as of the 110 excluded patients without a final diagnosis |
|---------------------------------|-----------------|-----------------|-----------------|
| Study patients (n=517) | Excluded patients (n=110) |
| Male | 215 (42%) | 55 (50%) |
| Mean age, years (SD) | 51 (18) | 61 (17) |
| Out-patients | 417 (81%) | 73 (66%) |
| Median duration of symptoms, days (quartiles) | 3 (1, 9) | 3 (1, 10) |
| Previous VTE | 73 (14%) | 25 (23%) |
| Family history of VTE | 105 (20%) | 17 (16%) |
| Risk-period | 103 (37%) | 56 (51%) |
| Active malignancy | 50 (10%) | 21 (19%) |
| Symptoms of DVT | 35 (7%) | 16 (15%) |

1 venous thromboembolism
2 period of immobilization, surgery or trauma in period of 3 months before presentation
3 deep-vein thrombosis
Diagnosis

Of the 517 study patients, 160 (31%) were classified as having pulmonary embolism. Pulmonary embolism was considered present on the basis of 39 abnormal pulmonary angiographies and 121 high-probability ventilation-perfusion scans. In the remaining 357 patients the diagnosis of pulmonary embolism was rejected on the basis of a normal pulmonary angiography (144) or a normal perfusion scan (213).

Clinical Probability and Clinical Models

The performance of the clinical probability estimate as well as both clinical models in categorizing the study patients is shown in Table 2. A clinical probability estimate, as assessed by the treating physician prior to objective testing, was available in 413 (80%) of the 517 study patients. In the remaining 104 patients the estimate was not obtained before the result of the ventilation-perfusion scan was known and were therefore excluded from further analysis. Nearly 20% of patients were categorized as having a probability of greater than 80%. The majority of patients (67%) were considered to have a moderate probability (20-80%) of pulmonary embolism, while 14% were judged to have a low probability (less than 20%). The prevalence of pulmonary embolism in this latter category was 19% and the likelihood ratio associated with this result was 0.53 (95% CI 0.29-0.99).

Complete information for the simplified clinical model was available for 414 (81%) of the 517 study patients. Only 2% of these patients had a high probability for pulmonary embolism. While the proportion of patients in the moderate category remained approximately the same as with the clinical probability estimate, more than one third of patients were categorized as having a low probability of pulmonary embolism with the use of the simplified clinical model. However, the prevalence of pulmonary embolism in this category was 28% and the likelihood ratio of this test result was 0.93 (95% CI 0.69-1.24), indicating almost no discriminatory effect.

The extended clinical model was performed in a subgroup of 237 patients in whom all required variables were collected. A quarter of the patients was categorized as having a low probability using this clinical model. The prevalence of pulmonary embolism in patients with this test result was 28%, while the likelihood ratio of 0.66 (95% CI 0.40 – 1.08) was slightly better than that of the simplified clinical model, albeit not significant.

To examine whether potential differences in the groups in which the various tests were performed were of influence on the observed performance of the tests, we repeated the analysis in the subgroup of patients in whom both the clinical probability estimate and the clinical models were available. This analysis showed similar results as the primary analysis (data not shown).

Emboli Localization

In 7 of the 11 patients with a clinical probability estimate of less than 20%, information about the localization of the emboli was available. Six of these seven patients (86%) had emboli in segmental or larger arteries, while only 1 patient had subsegmental pulmonary emboli. In 38 of the 41 patients in the low probability group according to the simplified clinical model, the localization of emboli was available. In twenty-eight (74%) of these cases the emboli were located in arteries of segmental or larger size. In 11 of the 12 (92%) patients in whom the localization was available in the low probability group as defined by the extended clinical model, the disease was located in segmental or larger arteries.

Observer Variability

Among the 6 participating centers an important variation in the percentages of patients falling into the respective categories of the studied models was observed. This despite the fact that the patient samples within the six centers were fully comparable concerning relevant clinical characteristics. The proportion of patients with a clinical probability estimate of less than twenty percent varied between 2% and 21%, while that of patients with a low probability according to the simplified clinical model ranged from 23% to 47%.

Discussion

This large multicenter study demonstrates that a physician’s clinical judgement alone performs comparably to two standardized clinical models in categorizing the pre-test probability in patients with suspected pulmonary embolism. However, all three methods show a disappointing performance, with the discriminatory potential among the various categories being low (Table 2). Although a low probability result is relatively frequently obtained (in approximately 15 to 35% of patients), the high prevalence of pulmonary embolism in patients in this category, which varied from 19 to 28%, is a limitation for the clinical utility. In particular, since our analysis of the extent of emboli indicates that a large proportion of these patients have thrombi in segmental or larger arteries.

Our results seem in contrast with those reported earlier for the clinical probability estimate and the clinical models (3-6). These previous studies observed a relatively low prevalence of pulmonary embolism (3 to 9%) in patients in the low probability category as compared to a prevalence as high as 78% in patients categorized as having a high probability. It is not fully clear why these methods show a lower discriminatory performance in the present study.

A possible explanation is the difference in the reference test for the diagnosis of pulmonary embolism (12). Our study used a strict protocol with objective diagnostic tests to arrive at a final diagnosis regarding the presence or absence of pulmonary embolism, while the studies of Wells et al. and Perrier et al. also used clinical follow-up as a reference test. Although this undoubtedly is a useful measure, this could lead to an underdiagnosis of (smaller) emboli with a low tendency for recurrence when untreated.

Table 2: The number of patients within each category, the proportion with pulmonary embolism, the likelihood ratios and their 95% confidence intervals of the clinical probability estimate and the two clinical models

<table>
<thead>
<tr>
<th>Clinical Probability</th>
<th>n</th>
<th>PE</th>
<th>LR</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 20%</td>
<td>58</td>
<td>11</td>
<td>0.53</td>
<td>0.29 – 0.99</td>
</tr>
<tr>
<td>20 – 80%</td>
<td>278</td>
<td>80</td>
<td>0.92</td>
<td>0.79 – 1.07</td>
</tr>
<tr>
<td>&gt; 80%</td>
<td>77</td>
<td>35</td>
<td>1.90</td>
<td>1.28 – 2.82</td>
</tr>
<tr>
<td>Total</td>
<td>413</td>
<td>129</td>
<td>1.00</td>
<td>0.66 – 1.55</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Simplified Model</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>147</td>
<td>41</td>
<td>0.93</td>
<td>0.69 – 1.24</td>
</tr>
<tr>
<td>Moderate</td>
<td>259</td>
<td>78</td>
<td>1.03</td>
<td>0.88 – 1.21</td>
</tr>
<tr>
<td>High</td>
<td>8</td>
<td>3</td>
<td>1.44</td>
<td>0.35 – 5.92</td>
</tr>
<tr>
<td>Total</td>
<td>414</td>
<td>122</td>
<td>1.00</td>
<td>0.66 – 1.55</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Extended Model</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>60</td>
<td>17</td>
<td>0.86</td>
<td>0.40 – 1.08</td>
</tr>
<tr>
<td>Moderate</td>
<td>139</td>
<td>54</td>
<td>1.07</td>
<td>0.86 – 1.33</td>
</tr>
<tr>
<td>High</td>
<td>39</td>
<td>18</td>
<td>1.43</td>
<td>0.80 – 2.53</td>
</tr>
<tr>
<td>Total</td>
<td>237</td>
<td>89</td>
<td>1.00</td>
<td>0.66 – 1.55</td>
</tr>
</tbody>
</table>
Another source of differences in discriminatory performance is a difference in the spectrum of patients studied (13). While in the present study the prevalence of pulmonary embolism was 31%, this figure was only 17 and 21% in the studies by Wells et al. and Perrier et al., respectively. This would seem to indicate that the spectrum of patients studied is indeed different. An influx of patients with a low probability of the disease in the earlier studies, leading on the one hand to a larger number of patients, and on the other hand to a lower prevalence of the disease in this category is a likely explanation.

The fact that in the present study there were no clinical consequences of either the estimate or the clinical models may also be of importance. While this could potentially lead to a less dedicated clinical probability estimate, one would not expect this to lead to a difference in the results with the clinical models, which are based on relatively objective criteria from the medical history and physical examination. Finally, it needs to be recognized that there is always a tendency towards overmodeling with multivariate analyses (14). The inclusion of more variables in a prediction model causes it to become a reflection of the original population and consequently less applicable in a new population. Furthermore, the addition of more variables may increase the accuracy, but also results in a worse precision of the model. These factors often lead to less favorable results when the model is validated.

The various potential explanations do not however apply to the differences observed between this investigation and the PIOPED study, which had a similar prevalence, used objective diagnostic outcome assessment and did not attach clinical consequences to the clinical probability estimates.

Some aspects of our study deserve comment. Firstly, a substantial proportion (15%) of screened patients had to be excluded due to presentation at a time at which there was an expected inability to complete the diagnostic protocol within 48 h or because some diagnostic testing had already been performed. Furthermore, a final diagnosis of pulmonary embolism was not reached in 110 of the 627 initially included patients. The baseline clinical characteristics of the 517 study patients are similar to those of these excluded patients (Table 1), although the latter patients were slightly older, were more often in-patients and had more accompanying disease. Nevertheless, we believe the patients studied are representative for patients with suspected pulmonary embolism seen at a large teaching hospital.

Secondly, we modified the extended clinical model by omitting information concerning the blood oxygen saturation and ECG since these variables are not routinely collected at admission. This could lead to a systematic underestimation of the probability for the disease. However, it is unlikely that this has occurred, since the proportion of patients in the low probability category was less than half that observed in the earlier study. Moreover, the extended model performed comparably with the simplified model, which is in line with the same observation by Wells and colleagues.

Finally, a formal training of physicians in the participating centers in the performance of the clinical probability estimate was not undertaken before the start of the study. This may have influenced the accuracy of the probability estimate. It can also be argued, however, that this study in both centers with and without a pre-existing specialized interest in the diagnosis of pulmonary embolism, approaches the clinical reality with regard to the accuracy of this method. This point of criticism does not apply to the clinical models, for which the data were collected in a formalized manner.

What is the clinical utility of the probability estimate and the clinical models in the diagnostic work-up of patients with suspected pulmonary embolism? Although our findings are less positive than previous reports, these methods still may have the potential to guide and thus simplify the diagnosis of pulmonary embolism. It is clear from the present study that they can not be used alone for the exclusion or acceptance of the diagnosis of pulmonary embolism. As suggested previously, they should be used in combination with other diagnostic tests.

Theoretically, there are at least two applications of these methods. The first is to use them to exclude pulmonary embolism in patients who have already undergone an objective diagnostic investigation, as suggested by Wells et al. and Perrier et al. for patients with a non-diagnostic finding upon ventilation-perfusion scintigraphy. The other application would be to use the clinical probability for example with D-dimer testing as the first step in the diagnostic work-up. One could thus envision a strategy in which patients with a low probability of the disease and a normal D-dimer level can be spared further diagnostic investigations. The safety of such strategies needs to be properly evaluated in prospective studies before implementation in routine clinical practice.

Appendix

Participating Centers

Academic Medical Center, Amsterdam:
B. J. Sanson, F. Turkstra, H. R. Büller
Leiden University Medical Center, Leiden:
W. de Monyé, P. M. T. Pattynama, M. V. Huisman
Leyenburgh Hospital, The Hague:
M. J. L. van Strijen, G. J. Kieft
Slotervaart Hospital, Amsterdam:
M. R. Mac Gillavry, D. P. M. Brandjes
University Hospital Vrije Universiteit, Amsterdam:
P. J. Hagen, O. S. Hoekstra, P. E. Postmus
Utrecht University Medical Center, Utrecht:
I. J. C. Hartmann, J. D. Banga, P. F. G. M. van Waes

References


Received July 20, 1999  Accepted September 3, 1999