

Ruling Out Pulmonary Embolism in Primary Care: Comparison of the Diagnostic Performance of "Gestalt" and the Wells Rule

Janneke M. T. Hendriksen, MD, PhD¹

Wim A. M. Lucassen, MD, PhD²

Petra M. G. Erkens, MD, PhD³

Henri E. J. H. Stoffers, MD, PhD³

Henk C. P. M. van Weert, MD, PhD²

Harry R. Büller, MD, PhD⁴

Arno W. Hoes, MD, PhD⁴

Karel G. M. Moons, PhD⁴

Geert-Jan Geersing, MD, PhD¹

¹Department of Epidemiology, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, The Netherlands

²Department of General Practice, Academic Medical Center, Amsterdam, The Netherlands

³Department of Family Medicine, CAPHRI School for Public Health and Primary Care, Maastricht University, Maastricht, The Netherlands

⁴Department of Vascular Medicine, Academic Medical Center, Amsterdam, The Netherlands



Conflicts of interest: authors report none.

CORRESPONDING AUTHOR

Janneke M. T. Hendriksen, MD, PhD
Julius Center for Health Sciences and Primary Care, Str 6.131
University Medical Center Utrecht
PO Box 85500
3508 GA Utrecht, The Netherlands
j.m.t.hendriksen-9@umcutrecht.nl

ABSTRACT

PURPOSE Diagnostic prediction models such as the Wells rule can be used for safely ruling out pulmonary embolism (PE) when it is suspected. A physician's own probability estimate ("gestalt"), however, is commonly used instead. We evaluated the diagnostic performance of both approaches in primary care.

METHODS Family physicians estimated the probability of PE on a scale of 0% to 100% (gestalt) and calculated the Wells rule score in 598 patients with suspected PE who were thereafter referred to secondary care for definitive testing. We compared the discriminative ability (c statistic) of both approaches. Next, we stratified patients into PE risk categories. For gestalt, a probability of less than 20% plus a negative point-of-care D-dimer test indicated low risk; for the Wells rule, we used a score of 4 or lower plus a negative D-dimer test. We compared sensitivity, specificity, efficiency (percentage of low-risk patients in total cohort), and failure rate (percentage of patients having PE within the low-risk category).

RESULTS With 3 months of follow-up, 73 patients (12%) were confirmed to have venous thromboembolism (a surrogate for PE at baseline). The c statistic was 0.77 (95% CI, 0.70-0.83) for gestalt and 0.80 (95% CI, 0.75-0.86) for the Wells rule. Gestalt missed 2 out of 152 low-risk patients (failure rate = 1.3%; 95% CI, 0.2%-4.7%) with an efficiency of 25% (95% CI, 22%-29%); the Wells rule missed 4 out of 272 low-risk patients (failure rate = 1.5%; 95% CI, 0.4%-3.7%) with an efficiency of 45% (95% CI, 41%-50%).

CONCLUSIONS Combined with D-dimer testing, both gestalt using a cutoff of less than 20% and the Wells rule using a score of 4 or lower are safe for ruling out PE in primary care. The Wells rule is more efficient, however, and PE can be ruled out in a larger proportion of suspected cases.

Ann Fam Med 2016;14:227-234. doi: 10.1370/afm.1930.

INTRODUCTION

Pulmonary embolism (PE) can be considered in patients with a wide variety of (pulmonary) symptoms, such as shortness of breath, coughing, or pain on inspiration, but the diagnosis will be confirmed in only 10% to 30% of patients in whom it is suspected.¹ Many diagnostic procedures are therefore performed when PE is not present. To reduce the number of these unnecessary procedures, guidelines recommend first identifying those patients having such a low probability of the condition that referral or further diagnostics can safely be withheld.^{2,3} This risk stratification can be based on either an implicit physician's estimate ("gestalt") or a formal diagnostic prediction model (such as the Wells rule or the [revised] Geneva score).⁴⁻⁶ In patients identified as having a low probability of PE, a negative D-dimer test result can safely rule out the condition.^{7,8}

Nowadays, formal prediction models are often regarded as a more accurate way to estimate disease probability when compared with gestalt. As they rely on predefined items, prediction models are easy to use, and results are independent of the level of experience. On the other hand,

gestalt enables incorporation of individual characteristics, such as the patient-specific context, that are not covered by prediction models. Many physicians maintain that a standardized prediction model, sometimes even referred to as cookbook medicine, does not allow for such individually tailored diagnostics as much as gestalt does.⁹

The diagnostic performance of gestalt and prediction models in cases of suspected PE has been compared in several studies in secondary care, yet with conflicting results because of substantial heterogeneity across studies.¹⁰ In primary care, however, evidence on the performance of gestalt in PE diagnosis is lacking altogether. The results from studies on gestalt performed in secondary care cannot directly be generalized to a primary care setting because family physicians do not come across patients with PE on a daily basis and thus inherently have less experience in recognizing the condition as compared with hospital specialists. Moreover, hospital specialists often have access to some basic laboratory and imaging tests (eg, blood gas analysis, chest radiograph, electrocardiogram) before making a gestalt estimate. These tests are usually not readily available in most primary care settings. Nevertheless, family physicians do have much experience in distinguishing severe disease (such as PE) from mild illnesses. The ability to do so is often a result of the contextual knowledge of their patients, which is slowly constructed during the long-standing

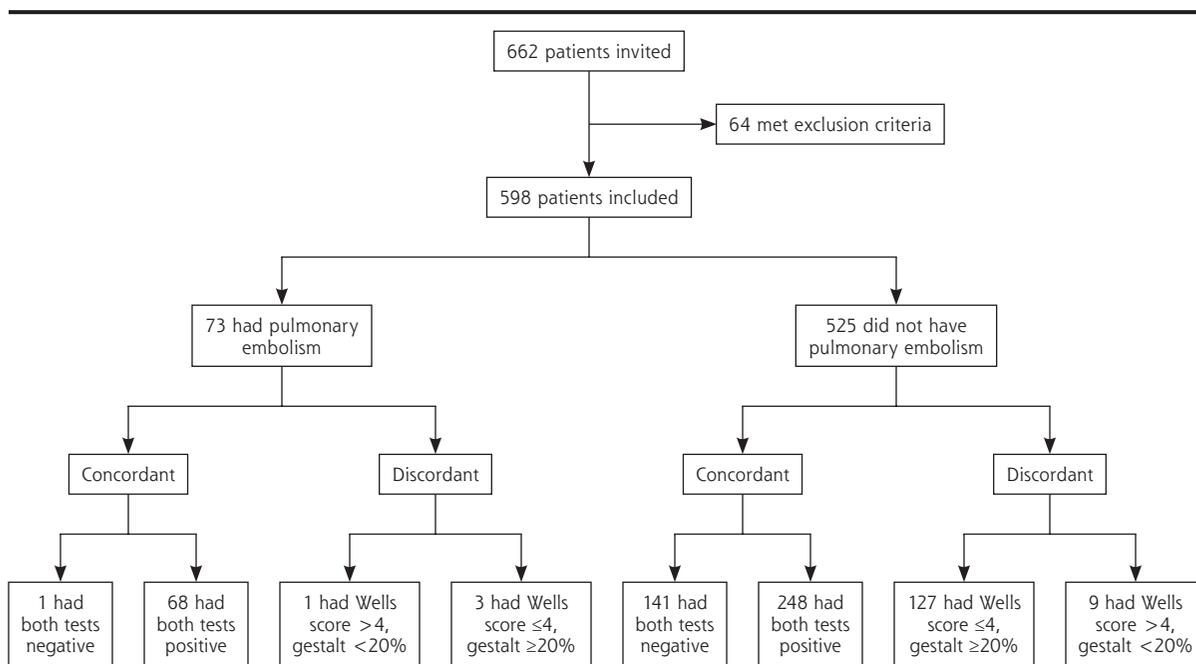
relationship that family physicians have with most of their patients.¹¹

Taking into account these possible merits and drawbacks of gestalt in primary care, as well as the mixed results regarding the comparison of gestalt vs diagnostic decision rules from studies in secondary care, the aim of this study was to compare the diagnostic performance of gestalt and the Wells rule for safely and efficiently ruling out PE in a large primary care population with symptoms suggestive of PE.

METHODS

We used prospectively collected data from the Dutch AMUSE 2 (Amsterdam, Maastricht, Utrecht Study on thromboEmbolism) cohort.⁷ This cohort study was initially designed to evaluate the diagnostic performance of the Wells rule combined with a qualitative D-dimer test in a Dutch primary care setting. A total of 662 consecutive adult patients seeking care at any of the 300 participating family physicians with symptoms raising suspicion of PE were invited to participate. Sixty-four of these patients met 1 of the predefined exclusion criteria, leaving 598 patients for further evaluation (Figure 1). Further details of the cohort are described elsewhere.⁷ The study protocol was assessed by the local institutional review board and exempted from formal reviewing; nevertheless, informed consent was obtained from all participants.

Figure 1. Flow of patients in the study.



Note: Gestalt refers to the family physician's estimation of the probability of pulmonary embolism.

Diagnostic Strategies

In all participants, family physicians assessed relevant information on general health, as well as specific cardiopulmonary or deep venous thrombosis signs and symptoms by systematically filling out a prespecified case record form. The physicians were asked to provide an estimated probability of PE being present using a visual analogue scale with a range from 0% to 100% (gestalt) (Supplemental Appendix Figure 1, available at <http://www.annfammed.org/content/14/3/227/suppl/DC1>). In addition, they calculated a Wells rule score for each patient; possible scores range from 0 to 12.5, with higher scores indicating a greater probability of PE (Supplemental Appendix Table 1, available at <http://www.annfammed.org/content/14/3/227/suppl/DC1>). Finally, a qualitative point-of-care D-dimer test (Clearview Simplify) was performed in all patients.

Outcome Assessment

The study protocol required referral of all patients with suspected PE to secondary care immediately after assessment in primary care, and the standard diagnostic pathway according to current local hospital guidelines was followed, with no explicit blinding to the family physician's findings. This process usually entailed a combination of estimated probability and quantitative laboratory-based D-dimer testing and, if indicated, a spiral computed tomography scan of the pulmonary arteries. The primary outcome was presence or absence of venous thromboembolism, based on a combined reference standard of all diagnostic imaging tests performed in the hospital (spiral computed tomography, ventilation-perfusion scanning, pulmonary angiography, leg ultrasonography, and clinical probability assessment as performed in secondary care, with or without D-dimer testing) and including any occurrence of venous thromboembolic events during 3 months of follow-up in primary care.⁷ Venous thromboembolism at this time point thus served as a surrogate for PE at baseline.

Statistical Analyses

Some patients were missing data for diagnostic variables: 0.5% were missing data for heart rate exceeding 100 beats/min, 2.7% for results of the D-dimer test, and 16% for gestalt probability. To minimize the effect of bias associated with selectively ignoring these patients, we imputed all missing data using multiple imputation techniques before undertaking analyses.^{12,13}

The diagnostic performance of gestalt and the Wells rule were compared using various methods. First, we quantified and compared the *c* statistic—the area under the curve (AUC)—of the receiver operating characteristic curve for gestalt and for the

Wells rule. An AUC of 0.5 reflects no discriminative ability, whereas an AUC of 1.0 indicates perfect discrimination.

Second, we stratified patients into groups of high or low predicted PE probability based on the gestalt and on the Wells rule estimate, each combined with the point-of-care D-dimer test. For gestalt, no common cutoff exists. We chose to apply a threshold for low probability of gestalt of less than 20% estimated probability and a negative D-dimer test, in line with prior research in the field of venous thrombosis.^{14,15} In sensitivity analyses, we assessed the impact of varying this threshold between 10% and 30%. The low-risk threshold used for the Wells rule was a score of 4 or lower and a negative D-dimer test, in accordance with previous publications.^{1,5} We calculated the efficiency (patients in the low PE probability category as a proportion of total number of patients) and failure rate (proportion of patients with PE in this low probability category) of both strategies, along with sensitivity, specificity, positive and negative predictive value (PPV and NPV), and overall accuracy. These diagnostic performance measures were compared. With the aim of safely ruling out PE in primary care, we primarily focused on safe exclusion (ie, high sensitivity, high NPV, and low failure rate), and subsequently on efficient exclusion of PE without further objective testing.

Finally, we hypothesized that family physicians using gestalt can correctly identify those patients who are at either very low or very high estimated risk. Nevertheless, they likely have difficulties deciding whom to refer in the group having intermediate (or uncertain) estimated risk. We therefore also assessed a combined approach, using gestalt, the Wells rule, and D-dimer testing. According to this stepped approach, a D-dimer test is performed if the estimated gestalt probability is less than 20%. Then, a negative D-dimer result safely rules out PE, whereas a positive test result is an indication for referral. In the case of an estimated probability exceeding 80%, PE suspicion is so pronounced that the patient is referred immediately, without further procedures. In the remaining group with an estimated probability between 20% and 80%, the Wells rule, followed by a D-dimer test in case of a score of 4 or lower, is applied. We calculated the efficiency and failure rate for this stepped approach as well.

All analyses were performed in IBM SPSS version 20. We assessed 2-sided statistical significance and defined it as a *P* value of less than .05.

RESULTS

The cohort consisted of 598 patients with suspected PE. Twenty-nine percent were male, and the mean age

was 48 years. More baseline characteristics are presented in Table 1.

In all, 73 patients (12%) received a diagnosis of

venous thromboembolic disease (72 had PE and 1 had deep venous thrombosis). The median gestalt estimated probability was 33% (interquartile range [IQR] = 40%)

with a total range from 0% up to 95%. Patients in whom the diagnosis of PE was confirmed ultimately had a median gestalt estimated probability of 70% (IQR = 40%), compared with 30% (IQR = 32%) in those without PE ($P \leq .001$).

As can be seen in Figure 2, both gestalt and the Wells rule had good overall discriminative ability for diagnosing PE, with an AUC of 0.77 (95% CI, 0.70-0.83) and 0.80 (95% CI, 0.75-0.86), respectively.

On the basis of a gestalt estimated probability of less than 20% plus a negative D-dimer test, 152 patients had a low predicted PE probability (efficiency = 25%; 95% CI, 22%-29%) with a failure rate of 1.3% (95% CI, 0.2%-4.7%) (Table 2). The sensitivity and specificity were 97% (95% CI, 90%-99%) and 29% (95% CI, 25%-33%), respectively. A conservative threshold (<10%) in combination with a negative D-dimer result was associated with very low efficiency (44/598 patients, 7%; 95% CI, 5%-10%) and 2 missed cases (failure rate = 4.5%; 95% CI, 0.5%-15.5%).

On the basis of a Wells rule score of 4 or lower plus a negative point-of-care D-dimer test result, 272 patients were stratified into the group having a low predicted PE probability (efficiency = 45%; 95% CI, 42%-50%), of whom 4 patients ultimately received a diagnosis of PE (failure rate = 1.5%; 95% CI, 0.4%-3.7%). Sensitivity and specificity of the Wells rule were 95% (95% CI, 87%-98%) and 51% (95% CI, 47%-55%), respectively.

For both diagnostic strategies, performance measures were substantially lower when the strategy was applied without the information from the point-of-care D-dimer test (Table 3). For example, sensitivity of the Wells rule score fell from 95% to 71%, and sensitivity of gestalt fell from 97% to 90%.

Table 1. Baseline Characteristics of the Study Cohort (N = 598)

| Characteristic | Final Diagnosis | | P Value |
|--|-----------------|-----------------|---------|
| | PE (n = 73) | Other (n = 525) | |
| Age, mean (SD), y | 53 (15.0) | 47 (16.0) | .004 |
| Male, No. (%) | 25 (34.2) | 148 (28.2) | .29 |
| Clinical signs and symptoms of DVT, No. (%) | 26 (35.6) | 31 (5.9) | <.001 |
| Alternative diagnosis less likely, No. (%) | 61 (83.6) | 272 (51.8) | <.001 |
| Heart rate >100/min, No. (%) | 25 (34.2) | 86 (16.4) | <.001 |
| Immobilization, No. (%) | 23 (31.5) | 71 (13.5) | <.001 |
| Previous PE or DVT, No. (%) | 18 (24.7) | 66 (12.6) | .005 |
| Hemoptysis, No. (%) | 5 (6.8) | 16 (3.0) | .19 |
| Malignancy, No. (%) | 5 (6.8) | 21 (4.0) | .42 |
| Gestalt probability, ^a median % (IQR) | 70 (40.0) | 30 (32.0) | <.001 |
| Wells rule score, ^b median (IQR) | 4.5 (4.0) | 3.0 (3.5) | <.001 |
| Positive D-dimer test, ^c No. (%) | 61 (83.6) | 198 (37.7) | <.001 |

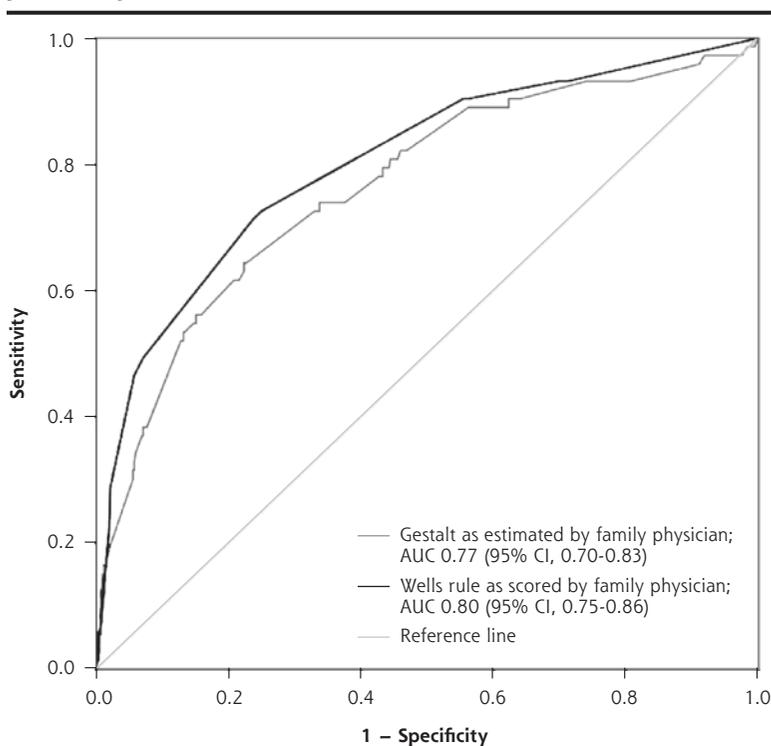
DVT = deep venous thrombosis; IQR = interquartile range; PE = pulmonary embolism.

^a Estimated PE probability by family physician.

^b Possible values range from 0 to 12.5, with higher values indicating a greater probability of PE.

^c Clearview Simplify point-of-care assay.

Figure 2. Receiver operating characteristic curve of the Wells rule and the family physician's gestalt estimated probability of pulmonary embolism.



AUC = area under the curve.

Table 2. Diagnostic Performance of the Wells Rule Score and the Gestalt Probability in Combination With D-Dimer Testing to Rule Out Pulmonary Embolism

| Performance Measure | Wells Rule Score ≤ 4 in Combination With D-Dimer Result | Gestalt Probability $< 20\%$ in Combination With D-Dimer | Sensitivity Analysis: Gestalt in Combination With D-Dimer | |
|---------------------------|--|--|--|------------------------------|
| | | | Gestalt Probability $< 10\%$ | Gestalt Probability $< 30\%$ |
| Sensitivity | 69/73 (95) [87-98] | 71/73 (97) [90-99] | 71/73 (97) [90-99] | 70/73 (96) [88-99] |
| Specificity | 268/525 (51) [47-55] | 150/525 (29) [25-33] | 42/525 (8) [6-11] | 216/525 (41) [37-45] |
| PPV | 69/326 (21) [17-26] | 71/446 (16) [13-20] | 71/554 (13) [10-16] | 70/379 (18) [15-23] |
| NPV | 268/272 (99) [96-100] | 150/152 (99) [95-100] | 42/44 (95) [85-99] | 216/219 (99) [96-100] |
| Accuracy | 337/598 (56) [52-60] | 221/598 (37) [33-41] | 113/598 (19) [16-22] | 286/598 (48) [44-52] |
| Efficiency ^a | 272/598 (45) [42-50] | 152/598 (25) [22-29] | 44/598 (7) [6-10] | 219/598 (37) [33-41] |
| Failure rate ^b | 4/272 (1.5) [0.6-3.7] | 2/152 (1.3) [0.4-4.7] | 2/44 (4.5) [1.3-15.1] | 3/219 (1.4) [0.5-3.9] |

NPV = negative predictive value; PPV = positive predictive value.

Note: Values are numerator/denominator (%) [95% CI]. Gestalt probability was tested at various low-risk cutoffs: $< 10\%$, $< 20\%$, and $< 30\%$.

^a Proportion of low-risk patients in the total cohort (Wells rule score ≤ 4 , gestalt probability of pulmonary embolism $< 10\%$, $< 20\%$, or $< 30\%$, and negative point-of-care D-dimer test result) (ie, patients not referred for objective testing).

^b The proportion of patients with pulmonary embolism in the low-risk group.

The stepped approach combining gestalt, the Wells rule, and D-dimer testing yielded efficiency and failure rates comparable to those seen with the Wells rule plus D-dimer test without gestalt (Table 4 and Figure 3).

DISCUSSION

Main Findings

In this study, we compared the performance of gestalt and a formal diagnostic prediction model (ie, the Wells rule) for ruling out PE in suspected cases in primary care. Both diagnostic strategies had good overall discriminative ability with AUCs of 0.77 and 0.80, respectively. In combination with a qualitative D-dimer test result, both gestalt and the prediction model could safely rule out PE; however, the number of patients who need to be referred for objective testing was substantially lower when using the Wells rule (efficiency = 45% vs 25%), as well as when using the stepped approach.

Comparison With Existing Literature

In 2011, Lucassen et al¹⁰ assessed the performance of gestalt and prediction models for diagnosing PE in a meta-analysis. They found substantial heterogeneity among the studies evaluating gestalt; for instance, the thresholds for low probability used ranged from 10% to 40% across studies. Furthermore, all studies were conducted in a secondary care setting. Our current findings, however, are in line with the main conclusions of that meta-analysis: family physicians do very well in safely excluding patients at very low risk by combining gestalt with D-dimer testing, yet at the price of referring (many) more patients as compared with using a formal decision rule.

Barais et al¹⁶ performed a qualitative study in a French primary care setting. Using semistructured interviews, they aimed to define the process preceding a confirmed diagnosis of PE. For all interviewed family physicians, the diagnostic process was mainly driven by intuitive factors, highlighting the importance of contextual knowledge and evidence in primary care. Given the qualitative nature of this study, however, it does not provide information on whether gestalt or a prediction model is more suitable for efficiently and safely identifying low-risk patients.

Table 3. Diagnostic Performance of the Wells Rule Score and the Gestalt Probability Without Point-of-Care D-Dimer Test

| Performance Measure | Wells Rule Score ≤ 4 | Gestalt Probability $< 20\%$ |
|---------------------|---------------------------|------------------------------|
| Sensitivity | 52/73 (71.2) | 66/73 (90.4) |
| Specificity | 401/525 (76.4) | 189/525 (36.0) |
| Efficiency | 422/598 (70.6) | 196/598 (32.8) |
| Failure rate | 21/422 (5.0) | 7/196 (3.6) |

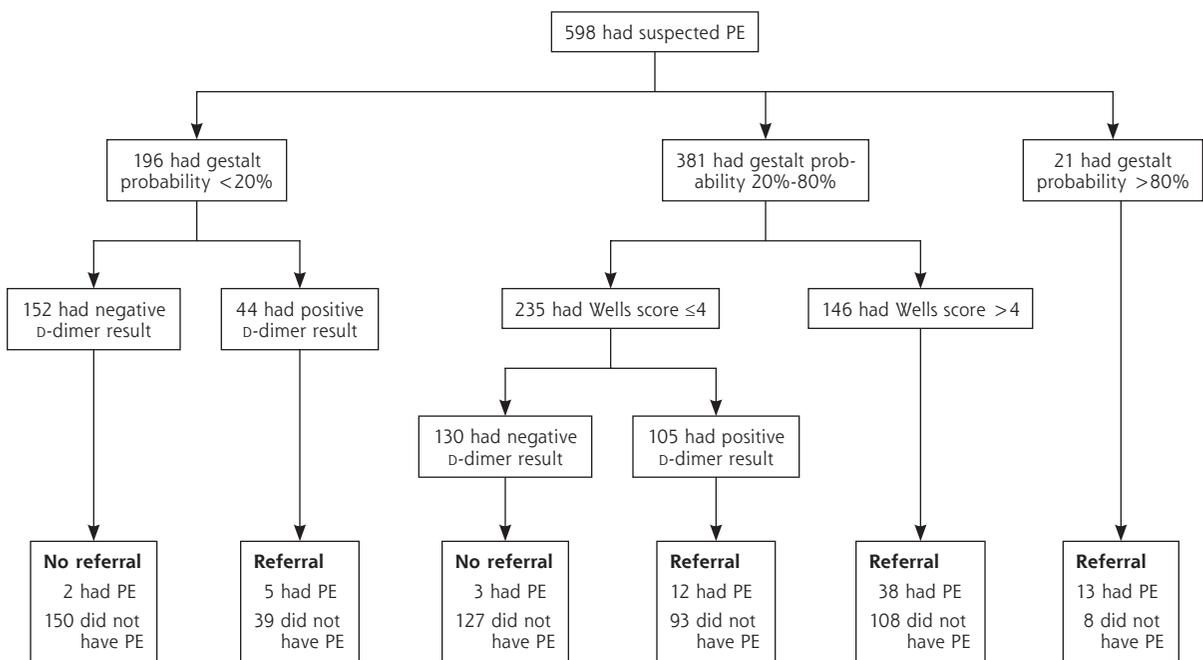
Note: Values are numerator/denominator (%).

Table 4. Diagnostic Performance of the Stepped Approach

| Performance Measure | Values |
|---------------------|-----------------------|
| Sensitivity | 68/73 (93) [85-97] |
| Specificity | 277/525 (53) [48-57] |
| Efficiency | 282/598 (47) [43-51] |
| Failure rate | 5/282 (1.8) [0.7-4.1] |

Note: Values are numerator/denominator (%) [95% CI].

Figure 3. Flow of patients in the study according to the stepped approach whereby gestalt was applied first.



PE = pulmonary embolism.

Note: Further management depends on this estimated probability of pulmonary embolism presence: referral if predicted risk is high, no referral if predicted risk is low. The diagnostic performance measures of the stepped approach are presented in Table 4.

Strengths and Limitations

The main strength of this study is that, to our knowledge, it is the first to assess the diagnostic performance of both gestalt and a PE prediction model in a primary care setting among a large cohort of patients with suspected PE.

Some limitations need to be addressed for full appreciation of our findings, however. First, this is a post hoc analysis of a prospective cohort study that had a main aim of validating the Wells rule in a primary care setting. As such, the estimated probabilities of PE by gestalt and the Wells rule were reported on the same case report form, introducing the chance of contamination of the estimates: gestalt might be influenced by the score of the Wells rule. As a consequence, the results of gestalt and the Wells rule are likely to be more alike, leading to dilution of the difference between the estimates. Nevertheless, we found distinct differences between them in this study, illustrated by the nearly universally overestimated gestalt estimate. As such, we hypothesize that the possible contamination might have merely attenuated our current inferences and that the real-life differences are expected to be larger in favor of the diagnostic prediction model.

Second, no consensus exists on the optimal threshold for low and high estimated PE probability. For this current analysis, we based the threshold on previously used thresholds in this field of research.¹⁴ Furthermore, we evaluated the performance using 3 cutoff values to see how this would affect the diagnostic performance. Rather than being viewed as absolute probabilities, these thresholds should be seen as a reflection of the degree of uncertainty, often categorized as low, intermediate, or high estimated probability.

Third, the Wells rule is a structured diagnostic prediction model, yet the subjective item "PE most likely diagnosis" has a relatively large contribution of 3 points to the score.¹⁷ It could therefore be argued that gestalt and this subjective item of the Wells rule more or less capture the same information. The Wells rule combines this information with other diagnostic variables; hence, it may be expected to give better predictions when compared with only the gestalt information. The discriminative value of both strategies, however, is comparable, with respective AUCs of 0.77 (95% CI, 0.70-0.83) and 0.80 (95% CI, 0.75-0.86). Although this remains speculative, it seems that the objective items of the Wells rule attenuates the (often overestimated) disease probability when using

gestalt alone. The integration of a subjective item with multiple (preselected) objective items may be pivotal to enhancing its value for use in clinical practice. This pattern is also exemplified from psychology research, which has demonstrated that physicians incline toward overestimation of predicted risk, especially if a potentially severe diagnosis such as PE is considered.¹¹ In case of any doubt (ie, intermediate risk), they tend to be hesitant and thus, to be on the safe side, refer the patient or initiate treatment relatively easily. This practice subsequently contributes to overdiagnosis, overtreatment, and increased health care use.^{18,19}

Fourth, scoring of the gestalt estimate is likely to be influenced by levels of experience and individual style. Albeit speculative, experienced family physicians may take prior experience into account and thus might feel more confident in assigning a low disease probability, although the opposite may also be true if this prior experience included a missed case of PE.¹⁶ Unfortunately, this information was not collected in our study. As such, we do not have insight into the individual range of scoring probabilities, years of experience, and level of training of all participating family physicians, and this may be subject for further research.

Fifth, a qualitative point-of-care D-dimer test was used in this study. Point-of-care tests return results within 10 to 15 minutes and are easy to use in the family physician's office. Especially for elderly patients and in rural areas, testing on the spot can be an attractive option. There is increasing experience with the use of the point-of-care tests in a variety of settings, including the Netherlands, United Kingdom, Canada, most countries in Scandinavia, and parts of the United States. These assays may not be available in all primary care settings worldwide, however. Nevertheless, point-of-care testing by itself is not obligatory for ruling out PE. One alternative, for instance, is to send patients to the nearest laboratory facility for an enzyme-linked immunosorbent assay (ELISA) or a latex D-dimer assay, usually offered in or near a hospital. These assays are slightly more sensitive and less specific when compared with point-of-care testing. As a consequence, the safety of both diagnostic strategies (Wells rule and gestalt) will likely be somewhat higher if an ELISA or latex assay is used, but at the expense of a slight decrease in efficiency. In the absence of ready availability of a point-of-care D-dimer test, neither diagnostic strategy alone (ie, without this test) is safe or efficient for excluding PE in primary care (Table 3).

Sixth, to express the safety of the diagnostic strategies in terms of missed cases in those patients identified as having a low risk, we used the term failure rate. A diagnostic failure can also occur in the high-risk group, however; in that group, it reflects the

proportion of patients without PE who are referred for objective testing. Although redundant referral is time consuming, burdensome, and associated with high costs and occasionally even renal failure due to contrast nephropathy, the clinical consequences of a failure in the low-risk group are perhaps more relevant given that it may include fatal consequences of missing PE. To quantify the overall failure rate in both risk groups, however, we presented the overall accuracy in Table 2 (ie, the probability that a patient is correctly classified and thus, the probability of failure is 100% minus accuracy). This overall accuracy is highest for the Wells rule in combination with D-dimer testing.

Finally, we defined our clinical outcome as the occurrence of a venous thromboembolic event, including both PE and deep venous thrombosis, during 3 months of follow-up. This rather conservative definition follows many previous studies in the field of PE and is based on the fact that we used a composite reference standard, implying that not all patients undergo imaging tests such as spiral computed tomography scanning at baseline. As such, finding deep venous thrombosis during 3 months of follow-up is considered surrogate evidence for having PE at baseline.⁸

Clinical Implications

Current guidelines recommend the use of a (structured) estimated disease probability to rule out PE. Our findings support the use of a prediction model, but leave room for relying on gestalt if disease presence or absence is highly likely or unlikely. As we expected, family physicians seem very capable of identifying patients at both ends of the probability spectrum. For a large group of patients at intermediate risk, however, application of the Wells rule and D-dimer testing will optimize the risk stratification better than using gestalt alone.

In conclusion, combined with D-dimer testing, both gestalt and the Wells rule were safe for ruling out PE in this primary care cohort. As compared with intuitive diagnostic reasoning (gestalt), however, the Wells rule is more efficient and enables ruling out of PE in a larger proportion of patients.

To read or post commentaries in response to this article, see it online at <http://www.annfammed.org/content/14/3/227>.

Key words: gestalt; diagnostic prediction models; family practice; pulmonary embolism; practice-based research; primary care; deep venous thrombosis

Submitted July 31, 2015; submitted, revised, January 5, 2016; accepted January 20, 2016.

Funding support: KGMM received a grant from The Netherlands Organization for Scientific Research (ZONMW 918.10.615 and 91208004).

Disclaimer: None of the funding sources had any role in the design, conduct, analyses, or reporting of the study or in the decision to submit the manuscript for publication.

Previous presentation: This work was presented in part in a poster by Hendriksen JMT et al: Diagnostic prediction model versus gestalt in the diagnosis of pulmonary embolism in primary care, International Society on Thrombosis and Haemostasis (ISTH) Scientific Meeting; June 20-25, 2015; Toronto, Canada.

Acknowledgments: We thank AMUSE-2 project members Ruud Oudega, Hugo ten Cate, and Martin H. Prins for their contribution to the design and initiation of the AMUSE 2 cohort.

Supplementary materials: Available at <http://www.AnnFamMed.org/content/14/3/227/suppl/DC1>.

References

1. PIOPED Investigators. Value of the ventilation/perfusion scan in acute pulmonary embolism. Results of the prospective investigation of pulmonary embolism diagnosis (PIOPED). *JAMA*. 1990;263(20):2753-2759.
2. Bates S, Jaeschke R, Stevens SM, et al. Diagnosis of DVT antithrombotic therapy and prevention of thrombosis, 9th ed: American College of Chest Physicians evidence-based clinical practice guidelines. *Chest*. 2012;141(2 Supp):e351S-e418S.
3. Konstantinides SV, Torbicki A, Agnelli G, et al; Task Force for the Diagnosis and Management of Acute Pulmonary Embolism of the European Society of Cardiology (ESC). 2014 ESC guidelines on the diagnosis and management of acute pulmonary embolism. *Eur Heart J*. 2014;35(43):3033-3069, 3069a-3069k.
4. Wells PS, Anderson DR, Rodger M, et al. Derivation of a simple clinical model to categorize patients probability of pulmonary embolism: increasing the models utility with the SimpliRED D-dimer. *Thromb Haemost*. 2000;83(3):416-420.
5. Le Gal G, Righini M, Roy P-M, et al. Prediction of pulmonary embolism in the emergency department: the revised Geneva score. *Ann Intern Med*. 2006;144(3):165-171.
6. Wicki J, Perneger TV, Junod AF, Bounameaux H, Perrier A. Assessing clinical probability of pulmonary embolism in the emergency ward: a simple score. *Arch Intern Med*. 2001;161(1):92-97.
7. Geersing GJ, Erkens PMG, Lucassen WA, et al. Safe exclusion of pulmonary embolism using the Wells rule and qualitative D-dimer testing in primary care: prospective cohort study. *BMJ*. 2012;345:e6564.
8. van Belle A, Büller HR, Huisman MV, et al; Christopher Study Investigators. Effectiveness of managing suspected pulmonary embolism using an algorithm combining clinical probability, D-dimer testing, and computed tomography. *JAMA*. 2006;295(2):172-179.
9. Cervellin G, Borghi L, Lippi G. Do clinicians decide relying primarily on Bayesian principles or on Gestalt perception? Some pearls and pitfalls of Gestalt perception in medicine. *Intern Emerg Med*. 2014;9(5):513-519.
10. Lucassen W, Geersing G-J, Erkens PMG, et al. Clinical decision rules for excluding pulmonary embolism: a meta-analysis. *Ann Intern Med*. 2011;155(7):448-460.
11. Starfield B, Shi L, Macinko J. Contribution of primary care to health systems and health. *Milbank Q*. 2005;83(3):457-502.
12. Janssen KJ, Vergouwe Y, Donders AR, et al. Dealing with missing predictor values when applying clinical prediction models. *Clin Chem*. 2009;55(5):994-1001.
13. Rubin DB, Schenker N. Multiple imputation in health-care databases: an overview and some applications. *Stat Med*. 1991;10(4):585-598.
14. Sanson BJ, Lijmer JG, Mac Gillavry MR, Turkstra F, Prins MH, Büller HR; ANTELOPE-Study Group. Comparison of a clinical probability estimate and two clinical models in patients with suspected pulmonary embolism. *Thromb Haemost*. 2000;83(2):199-203.
15. Geersing GJ, Janssen KJ, Oudega R, et al; AMUSE Study Group. Diagnostic classification in patients with suspected deep venous thrombosis: physicians' judgement or a decision rule? *Br J Gen Pract*. 2010;60(579):742-748.
16. Barais M, Morio N, Cuzon Breton A, et al. "I can't find anything wrong: it must be a pulmonary embolism": diagnosing suspected pulmonary embolism in primary care, a qualitative study. *PLoS One*. 2014;9(5):e98112.
17. Klok FA, Karami Djurabi R, Nijkeuter M, Huisman MV. Alternative diagnosis other than pulmonary embolism as a subjective variable in the Wells clinical decision rule: not so bad after all. *J Thromb Haemost*. 2007;5(5):1079-1080.
18. Harteis C, Billett S. Intuitive expertise: theories and empirical evidence. *Educ Res Rev*. 2013;9(9):145-157.
19. Grover SA, Lowensteyn I, Esrey KL, Steinert Y, Joseph L, Abrahamowicz M. Do doctors accurately assess coronary risk in their patients? Preliminary results of the coronary health assessment study. *BMJ*. 1995;310(6985):975-978.