Diagnostic prediction models for suspected pulmonary embolism: systematic review and independent external validation in primary care

Janneke M T Hendriksen,1 Geert-Jan Geersing,1 Wim A M Lucassen,2 Petra M G Erkens,3 Henri E J H Stoffers,3 Henk C P M van Weert,2 Harry R Büller,4 Arno W Hoes,1 Karel G M Moons1

ABSTRACT

OBJECTIVE
To validate all diagnostic prediction models for ruling out pulmonary embolism that are easily applicable in primary care.

DESIGN
Systematic review followed by independent external validation study to assess transportability of retrieved models to primary care medicine.

SETTING
300 general practices in the Netherlands.

PARTICIPANTS
Individual patient dataset of 598 patients with suspected acute pulmonary embolism in primary care.

MAIN OUTCOME MEASURES
Discriminative ability of all models retrieved by systematic literature search, assessed by calculation and comparison of C statistics. After stratification into groups with high and low probability of pulmonary embolism according to pre-specified model cut-offs combined with qualitative D-dimer test, sensitivity, specificity, efficiency (overall proportion of patients classified as having low probability of pulmonary embolism) and failure rate (proportion of pulmonary embolism cases in group of patients with low probability) were calculated for all models.

RESULTS
Ten published prediction models for the diagnosis of pulmonary embolism were found. Five of these models could be validated in the primary care dataset: the original Wells, modified Wells, simplified Wells, revised Geneva, and simplified revised Geneva models. Discriminative ability was comparable for all models (range of C statistic 0.75–0.80). Sensitivity ranged from 88% (simplified revised Geneva) to 96% (simplified Wells) and specificity from 48% (revised Geneva) to 53% (simplified revised Geneva). Efficiency of all models was between 43% and 48%. Differences were observed between failure rates, especially between the simplified Wells and the simplified revised Geneva models (failure rates 1.2% (95% confidence interval 0.2% to 3.3%) and 3.1% (1.4% to 5.9%), respectively; absolute difference 1.98% (−3.33% to −0.74%). Irrespective of the diagnostic prediction model used, three patients were incorrectly classified as having low probability of pulmonary embolism; pulmonary embolism was diagnosed only after referral to secondary care.

CONCLUSIONS
Five diagnostic pulmonary embolism prediction models that are easily applicable in primary care were validated in this setting. Whereas efficiency was comparable for all rules, the Wells rules gave the best performance in terms of lower failure rates.

Introduction
Pulmonary embolism is a potentially fatal condition if left untreated. Its presentation can be relatively mild, sometimes even mimicking myalgia or a simple cough. This causes pulmonary embolism to be a diagnosis that is easily missed.1 2 As a result, physicians have a low threshold for suspicion and subsequent referral for further diagnostics.3 4 Referred patients will be exposed to the burden, costs, and even potential iatrogenic damage of diagnostic techniques such as spiral computed tomography or contrast nephropathy.5 However, only in a small subset (about 10-15%) of all suspected cases are emboli actually confirmed during diagnostic investigation.6

Several non-invasive diagnostic prediction models have been developed for safe exclusion of pulmonary embolism and are usually followed by D-dimer testing.7 Physicians can use these models as a strategy to enhance the efficiency of the diagnostic process by precluding those patients with a low probability of pulmonary embolism from further diagnostic tests, without compromising on safety (that is, missing cases of pulmonary embolism). Such diagnostic strategies can reduce the number of unnecessary computed tomography scans by 35%, with only 1-2% of missed cases in the group of patients with a low probability of pulmonary embolism.7

In many countries, general practitioners are the first physicians to encounter patients with symptoms suggestive of pulmonary embolism. Risk stratification...
is valuable in deciding which patients to refer. All diagnostic models for safe exclusion of pulmonary embolism have been developed and validated in hospital or acute care settings. However, diagnostic prediction models developed in a particular setting often perform less well when applied in another setting. Therefore, models derived in hospital or acute care settings cannot simply be implemented in primary care. Reasons for this poorer performance include differences in the case mix and the prevalence of pulmonary embolism due to the unselected population, as well as differences in physicians’ experience of patients with suspected pulmonary embolism. Hence, when transferring diagnostic models or strategies across healthcare settings, evaluation of their performance in this other setting is necessary first. This form of external validation is referred to as domain or setting validation, or as quantification of the transportability of prediction models.

The recent AMUSE-2 study (Amsterdam, Maastricht, Utrecht Study on thrombo-Embolism) has been the first to prospectively quantify the transportability of the, perhaps best known, secondary care derived diagnostic prediction model for pulmonary embolism (that is, the Wells pulmonary embolism rule, combined with point of care D-dimer testing) in a primary care setting. Various other diagnostic pulmonary embolism prediction models that may also be valuable for primary care have been developed but have not been validated in a primary care population.

The aim of this study was therefore to assess the clinical performance in a primary care setting of all existing diagnostic models developed for patients with suspected pulmonary embolism. We firstly did a systematic review and critical appraisal of all available diagnostic models for pulmonary embolism, as recommended by guidelines on prediction models research. Next, the diagnostic models easily applicable in primary care were validated in the AMUSE-2 dataset—that is, a large independent prospectively constructed cohort of patients presenting to their general practitioner with complaints suggestive of pulmonary embolism.

Methods
Updated systematic review
For our systematic review and critical appraisal of the existing diagnostic models for pulmonary embolism, we followed the recent methodological guidance by the Prognosis Methods Group of the Cochrane Collaboration.

Firstly, we framed the review question and design by using the CHARMS checklist for systematic reviews of prediction models (see appendix box A). We then repeated the systematic search previously performed for an aggregate meta-analysis by Lucassen et al and used the same study selection criteria. We searched for studies on development and validation of diagnostic prediction models published between January 2010 and October 2014. Details on the search syntax can be found in appendix figure A. We then critically appraised studies on the development of diagnostic prediction models by using the CHARMS checklist (appendix table A). All retrieved papers were examined by two independent reviewers (JH, GJG) and a third independent reviewer (KGMM) in case of disagreement.

Given the scope of our systematic review (see appendix box A), we assessed all diagnostic prediction models for pulmonary embolism, retrieved by our search, on their applicability in a primary care domain. Accordingly, the diagnostic predictors or tests included in the diagnostic model needed to be measurable at the general practitioner’s office. Variables such as signs and symptoms, items from history taking, pulse rate, or blood pressure are easily and quickly obtained in primary care, whereas results from (arterial) blood gas analyses, chest radiographs, or advanced electrocardiograph interpretations generally are not. Diagnostic models with predictors that cannot easily be obtained in primary care were excluded from the main analyses.

Validation cohort
Population characteristics
The AMUSE-2 cohort was designed to prospectively validate the Wells pulmonary embolism rule in a Dutch primary care setting. The study took place between 1 July 2007 and 31 December 2010. In short, it included 662 adult patients presenting at one of the participating general practices with complaints raising suspicion of pulmonary embolism (that is, acute dyspnoea, pain on inspiration, or unexplained cough; all at the discretion of the including physicians). Of these patients, 64 met one of the predefined exclusion criteria of anticoagulant treatment at presentation, pregnancy, or unwillingness or inability to provide written informed consent, leaving 598 patients for further evaluation. More details of this cohort and the sample size calculation are described elsewhere.

Predictors
In all participants, the general practitioner assessed relevant information on general health and specific cardiopulmonary and signs and symptoms of deep venous thrombosis by systematically filling out a pre-specified case record form. Subsequently, a qualitative point of care D-dimer test (Simplify D-dimer; Clearview, Inverness Medical, Bedford, UK) was performed. This test returns a visual dichotomous outcome; a positive test result is indicated by a pink-purple coloured line that appears on the disposable device within 10 minutes of application. This corresponds to a D-dimer concentration above 80 ng/mL. Only a control line will be visible if the test is negative. In case of an inconclusive test result, we classified the result as positive.

All predictors in the validation cohort were assessed blinded for the outcome. Exact definitions and measurement methods of the predictors in the validation cohort have been described previously.

Outcome
The study protocol recommended referral of all patients with suspected pulmonary embolism to secondary care, regardless of the outcome of the Wells
rule or D-dimer test. In secondary care, the regular diagnostic pathway according to local hospital guidelines was followed, with no explicit blinding for the general practitioner’s findings. This usually comprised a combination of estimated probability and quantitative laboratory based D-dimer testing and was followed by diagnostic imaging if indicated. The primary outcome was the presence of venous thromboembolism (either deep venous thrombosis or pulmonary embolism), as based on a composite 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 three months of follow-up in primary care.

Data analysis
For all diagnostic prediction models, we retrospectively calculated the individual score of each included patient on the basis of the presence or absence of the model’s predictors. We compared the overall discriminative ability of the models, using the C statistic (that is, the area under the curve) of the receiver operating characteristics curve, with 95% confidence intervals. We assessed differences between the C statistics with the DeLong method.25

To stratify all participants in the validation cohort into categories of low or high probability of having pulmonary embolism, we used the diagnostic pathway as recommended by guidelines (fig 1): first the stratification based on the cut-off value of each diagnostic model as suggested in the development papers for the model, followed by a D-dimer test in case of a low predicted probability of pulmonary embolism. A subsequent negative D-dimer test implied a low predicted probability of pulmonary embolism and no need for referral to secondary care. In all other cases, the high predicted probability of pulmonary embolism meant referral for further objective testing.

For the Wells rule, two cut-off values have been proposed: low risk (<4) and very low risk (<2). Given the prevalence in this primary care cohort and following previous publications on this topic, we chose to use the first of these cut-off values for all subsequent analyses.19 26 We then calculated the common diagnostic accuracy measures of the models (that is, sensitivity, specificity, and positive and negative predictive values).

From a clinical point of view, a diagnostic prediction model should ideally classify as many patients as possible in the non-referral group, but not at the expense of an increase in pulmonary embolism events missed in this group. Therefore, we evaluated the clinical performance of each model (combined with D-dimer) by focusing on the efficiency and the failure rate. We defined efficiency as the proportion of patients in the whole cohort stratified to the group with low predicted probability of pulmonary embolism (that is: (true negatives (tn)+false negatives (fn))/total cohort). We defined failure rate as the proportion of these patients with low predicted probability of pulmonary embolism ultimately diagnosed as having pulmonary embolism on the basis of our composite reference standard (that is: fn/(tn+fn)). We calculated differences in failure rates between the models, with the surrounding 95% confidence intervals. We then varied the cut-off values as proposed in the model development studies to evaluate the influence of different cut-off values on the outcome measures. Failure rates were displayed in a forest plot, together with the failure rates found in previous validation studies of the different models with D-dimer testing.

Finally, we constructed calibration plots for the diagnostic prediction models. With calibration plots, the agreement between the predicted and observed probability of pulmonary embolism can be visualised. In the absence of a reported intercept for the models, we re-estimated the intercept in the validation cohort by using the linear predictor as offset in a logistic regression model including the model coefficients.

In the dataset, we imputed missing values for predictors by using multiple imputation techniques before our analyses.27 28 Imputation was performed to minimise the effect of the bias associated with selectively ignoring these patients. In case of a high percentage of missing values, no imputation was performed. Instead, the main analysis was carried out with all missing values assigned as the predictor being absent. In a sensitivity analysis, all missing values were assigned as the predictor being present.

We used IBM SPSS version 21 for descriptive statistical analyses. We used R version 3.2.0 for forest plots and the calculation of differences in C statistics (DeLong method) and failure rates.

Reporting
The results of this validation study were reported in adherence to the TRIPOD (Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis) guideline (see appendix table C).29 30
Patient involvement
No patients were involved in setting the research question or the outcome measures; nor were they involved in recruitment or in the design and implementation of the study. There are no plans to involve patients in dissemination.

Results
Our systematic literature search identified four model validation studies but no newly developed models in addition to the 10 models previously found with the search by Lucassen (fig 2). Of these 10 models, we excluded two models (the Pisa rule and the original Geneva score) from further analysis on the basis of the predetermined criteria for validation in our primary care cohort (fig 2). The original, modified, and simplified Wells rules, the revised Geneva scores, the revised Pisa rule, the pulmonary embolism rule-out criteria (PERC) rule, and the Charlotte rule fulfilled the predetermined primary care criteria. However, owing to complete missing information on the predictors oxygen saturation and body temperature in our dataset, we were not able to validate the revised Pisa rule, the PERC rule, and the Charlotte rule, leaving five models for further evaluation. See appendix table B for an overview of all diagnostic models identified in this study, including the five models that were tested in the individual patient dataset of our validation cohort: the original, modified, and simplified Wells rules and the original and simplified revised Geneva scores.

Of the four recently published validation studies identified by our search, only one study reported the results of the combination of the prediction model and D-dimer testing and was used in this current update. The observed failure rates from previous development and validation studies, and this validation of the prediction models, are shown in a forest plot (fig 3).

External validation
In the validation cohort, venous thromboembolism was diagnosed in 73 of 598 patients (72 pulmonary embolism, 1 deep venous thrombosis) during the complete three month follow-up period (prevalence 12%). Table 1 shows the baseline characteristics of the cohort, as well as the baseline characteristics observed in the development studies of the validated diagnostic prediction models. The main differences between the four development cohorts and the current validation cohort include the prevalence of the outcome, mean age, and percentage of male participants.

All models had moderate to good discriminative ability, with a C statistic ranging from 0.75 (simplified revised Geneva score) to 0.80 (original and modified Wells rules) (largest difference P=0.038; simplified revised Geneva v original Wells) (table 2 and appendix figure B).

For the originally suggested thresholds of the three Wells rules (original Wells, modified Wells, and simplified Wells, all with D-dimer testing if low predicted probability of pulmonary embolism), sensitivity was around 95%; it was slightly lower for the Geneva score models (88-90%). All five diagnostic models showed a specificity of approximately 50% (table 3). The simplified revised Geneva model was observed to be the most efficient rule (48%; 287 non-referred patients in cohort of 598 patients with suspected pulmonary embolism), but it was also associated with the highest failure rate (nine missed events in 287 non-referred patients; 3.1%). The largest difference in failure rates was observed between the simplified Wells and simplified revised Geneva score (-1.98%, 95% confidence interval -3.33% to -0.74%) (table 4). Interestingly, three of these nine patients were missed by all rules under study. These three patients are described in more detail in the box.

Overall, a one point lower cut-off (low risk original Wells ≤3 and D-dimer negative) affected the failure rate of each model little (failure rate 0.9-2.9%) but hampered its efficiency, especially for the simplified Wells rule (table 5). Conversely, an increased cut-off was more efficient, ranging from 49% to 54%, but more pulmonary embolism events were missed.
As can be appreciated from appendix figure C, the calibrations of the Wells rules and the revised Geneva score were good and comparable. Calibration of the simplified revised Geneva score was slightly worse, especially in the highest tenth of predicted risk.

**Discussion**

Our systematic review identified 10 previously developed diagnostic prediction models to rule out pulmonary embolism. Of these, we evaluated five models for their transportability to primary care in an independent cohort of 598 patients. We found that all five models could rule out pulmonary embolism in about four in every 10 patients. However, we observed a difference in favour of the three Wells rules compared with the revised Geneva models in terms of safety. The proportion of cases of pulmonary embolism in those patients identified as being at low risk was substantially lower for the Wells rules.
A consensus exists, some people consider an upper 95% confidence interval boundary of a failure rate higher than 3% to be too high, although previous landmark studies in the field of diagnosing pulmonary embolism found failure rates with an upper 95% confidence interval boundary close to 4%. Others focused on the point estimate of the failure rate instead, which should be under 2% given that even the most invasive diagnostic procedure (digital subtraction angiography) cannot diagnose all pulmonary embolism events and misses about 2% of cases. We observed failure rates of the Wells rules of 1.2-1.5%, which is well below this 2% point estimate but with a 95% confidence interval that crosses 3%. Importantly, however, these boundaries for the 95% confidence interval are not necessarily the “true” values but reflect the remaining uncertainty around the point estimate.

For scoring the item “pulmonary embolism most likely,” Wells originally suggested that information on electrocardiography, routine laboratory tests, and

### Table 1 | Baseline characteristics of participants in external validation cohort (AMUSE-2; patients presenting in primary care, with suspected pulmonary embolism), and baseline characteristics of development studies of five diagnostic prediction models. Values are numbers (percentages) unless stated otherwise

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>AMUSE 2 validation cohort (n=598)</th>
<th>Original Wells (n=1260)</th>
<th>Modified/simplified Wells (n=3306)</th>
<th>Original revised Geneva (n=965)</th>
<th>Simplified revised Geneva (n=1049)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male sex</td>
<td>173 (28.9)</td>
<td>NR</td>
<td>1409 (42.6)</td>
<td>403 (41.8)</td>
<td>414 (39.5)</td>
</tr>
<tr>
<td>Mean (SD) age, years</td>
<td>48 (16)</td>
<td>NR</td>
<td>53 (18)</td>
<td>61 (19)</td>
<td>56 (SD NR)</td>
</tr>
<tr>
<td>Venous thromboembolic events</td>
<td>73 (12.2)</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>241 (23.0)</td>
</tr>
<tr>
<td>Pulmonary embolism cases</td>
<td>72 (12.0)</td>
<td>222 (17.6)</td>
<td>674 (20.4)</td>
<td>222 (23.0)</td>
<td>241 (23.0)</td>
</tr>
</tbody>
</table>

**Items Wells rule**

- Clinical signs and symptoms of DVT: 57 (9.5) NR 190 (5.7) NR
- Alternative diagnosis less likely: 333 (55.7) NR 2032 (61.5) NR
- Heart rate >100 bpm: 111 (18.6) NR 867 (26.2) NR
- Immobilisation or recent surgery: 94 (15.7) NR 610 (18.5) 232 (24.0) NR
- Pulmonary embolism or DVT: 84 (14.0) 160 (12.7) 480 (14.5) 166 (17.2) NR
- Haemoptysis: 21 (3.5) NR 176 (5.3) 43 (4.5) NR
- Malignancy: 26 (4.3) NR 476 (14.4) 89 (9.2) NR

**Items revised Geneva score**

- Age >65 years: 103 (17.2) NR NR NR NR
- Heart rate:
  - <75 bpm: 187 (31.3) NR NR NR NR
  - 75-94 bpm: 259 (43.3) NR NR NR NR
  - >94 bpm: 152 (25.4) NR NR NR NR
- Lower limb pain: 96 (16.1) NR NR 138 (14.3) NR
- Oedema/swelling: 48 (8.0) NR NR NR NR
- Pain and swelling*: 22 (3.7) NR NR 51 (5.3) NR
- Pain on lower limb deep venous palpation†: 34 (5.7) NR NR NR NR
- Pain and swelling*‡: 24 (4.0) NR NR NR NR
- Pain on lower limb deep venous palpation‡: 391 (65.4) NR NR NR NR

**D-dimer measurement**

- Qualitative D-dimer positive: 259 (43.3) NR NR NR NR
- Quantitative D-dimer test instead of qualitative point of care D-dimer test performed.

### Table 2 | Head to head comparison of C statistics of five diagnostic prediction models (without D-dimer testing) to rule out pulmonary embolism, validated in primary care AMUSE-2 cohort (n=598), using DeLong method

<table>
<thead>
<tr>
<th>Models compared (model 1 v model 2)</th>
<th>C statistic</th>
<th>Estimated difference (95% CI)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original Wells v simplified Wells</td>
<td>0.801</td>
<td>0.014 (−0.004 to 0.032)</td>
<td>0.132</td>
</tr>
<tr>
<td>Original Wells v modified Wells</td>
<td>0.801</td>
<td>0.003 (−0.001 to 0.007)</td>
<td>0.114</td>
</tr>
<tr>
<td>Original Wells v original revised Geneva</td>
<td>0.801</td>
<td>0.045 (−0.007 to 0.097)</td>
<td>0.091</td>
</tr>
<tr>
<td>Original Wells v simplified revised Geneva</td>
<td>0.801</td>
<td>0.053 (0.003 to 0.103)</td>
<td>0.038</td>
</tr>
<tr>
<td>Simplified Wells v original revised Geneva</td>
<td>0.787</td>
<td>−0.011 (−0.028 to 0.006)</td>
<td>0.203</td>
</tr>
<tr>
<td>Simplified Wells v modified Wells</td>
<td>0.787</td>
<td>0.031 (−0.017 to 0.079)</td>
<td>0.207</td>
</tr>
<tr>
<td>Simplified Wells v original revised Geneva</td>
<td>0.787</td>
<td>0.039 (−0.005 to 0.083)</td>
<td>0.084</td>
</tr>
<tr>
<td>Modified Wells v original revised Geneva</td>
<td>0.798</td>
<td>0.043 (−0.010 to 0.096)</td>
<td>0.113</td>
</tr>
<tr>
<td>Modified Wells v simplified revised Geneva</td>
<td>0.798</td>
<td>0.050 (0.000 to 0.100)</td>
<td>0.048</td>
</tr>
<tr>
<td>Original revised Geneva v simplified revised Geneva</td>
<td>0.756</td>
<td>0.008 (−0.010 to 0.026)</td>
<td>0.388</td>
</tr>
</tbody>
</table>
Table 3 | Diagnostic accuracy measures of five diagnostic prediction models, combined with point of care D-dimer testing, to rule out pulmonary embolism, validated in primary care AMUSE-2 cohort (n=598), with (95% confidence intervals)

<table>
<thead>
<tr>
<th>Measure</th>
<th>Original Wells ≤4</th>
<th>Modified Wells ≤2</th>
<th>Simplified Wells ≤1</th>
<th>Original revised Geneva ≤5*</th>
<th>Simplified revised Geneva ≤2*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensitivity</td>
<td>95% (87% to 98%)</td>
<td>95% (87% to 98%)</td>
<td>96% (88% to 99%)</td>
<td>90% (81% to 96%)</td>
<td>88% (78% to 94%)</td>
</tr>
<tr>
<td>Specificity</td>
<td>51% (47% to 55%)</td>
<td>50% (46% to 55%)</td>
<td>49% (45% to 53%)</td>
<td>48% (44% to 53%)</td>
<td>53% (49% to 57%)</td>
</tr>
<tr>
<td>Positive predictive value</td>
<td>21% (17% to 26%)</td>
<td>21% (17% to 26%)</td>
<td>21% (17% to 25%)</td>
<td>20% (15% to 24%)</td>
<td>21% (16% to 26%)</td>
</tr>
<tr>
<td>Negative predictive value</td>
<td>99% (96% to 100%)</td>
<td>99% (96% to 100%)</td>
<td>99% (97% to 100%)</td>
<td>97% (95% to 99%)</td>
<td>97% (94% to 99%)</td>
</tr>
<tr>
<td>Efficiency</td>
<td>46% (41% to 50%)</td>
<td>45% (41% to 49%)</td>
<td>43% (39% to 48%)</td>
<td>44% (40% to 48%)</td>
<td>48% (44% to 52%)</td>
</tr>
<tr>
<td>Failure rate</td>
<td>1.5 (0.4% to 3.7%)</td>
<td>1.5 (0.4% to 3.8%)</td>
<td>1.2 (0.2% to 3.3%)</td>
<td>2.7 (1.1% to 5.4%)</td>
<td>3.1 (1.4% to 5.9%)</td>
</tr>
</tbody>
</table>

*Main analysis used composite variable “pain and swelling,” in which missing values for “pain on deep vein (DV) palpation” were assigned as “pain on DV palpation” absent; results for analysis in which all missing values were assigned as “pain on DV palpation” present are identical and are not presented separately.

Table 4 | Differences in failure rates between different diagnostic prediction models in AMUSE-2 cohort (with 95% confidence intervals based on 200 bootstrap samples)

<table>
<thead>
<tr>
<th>Comparison (model 1 vs model 2)</th>
<th>Difference between failure rates (model 1 vs model 2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original Wells vs simplified Wells</td>
<td>0.32% (~0.08% to 1.04%)</td>
</tr>
<tr>
<td>Original Wells vs modified Wells</td>
<td>−0.02% (~0.04% to 0.00%)</td>
</tr>
<tr>
<td>Original Wells vs original revised Geneva</td>
<td>−1.21% (~2.49% to 0.06%)</td>
</tr>
<tr>
<td>Original Wells vs simplified revised Geneva</td>
<td>−1.67% (~3.33% to 0.21%)</td>
</tr>
<tr>
<td>Simplified Wells vs modified Wells</td>
<td>−0.33% (~1.06% to 0.06%)</td>
</tr>
<tr>
<td>Simplified Wells vs original revised Geneva</td>
<td>−1.53% (~2.80% to 0.36%)</td>
</tr>
<tr>
<td>Simplified Wells vs simplified revised Geneva</td>
<td>−1.98% (~3.33% to 0.74%)</td>
</tr>
<tr>
<td>Modified Wells vs original revised Geneva</td>
<td>−1.20% (~2.47% to 0.18%)</td>
</tr>
<tr>
<td>Modified Wells vs simplified revised Geneva</td>
<td>−1.65% (~3.21% to 0.25%)</td>
</tr>
<tr>
<td>Original revised Geneva vs simplified revised Geneva</td>
<td>−0.45% (~1.23% to 0.21%)</td>
</tr>
</tbody>
</table>

Characteristics of three cases missed by all five diagnostic prediction models combined with D-dimer testing when validated in AMUSE-2 cohort, consisting of 598 primary care patients with suspected pulmonary embolism

Case 1—75 year old man
Venous thromboembolism in history; heart rate 59 bpm
Simply D-dimer negative
Scores of models:
- Original Wells—1.5 (previous PE)
- Modified Wells—1.0 (previous PE)
- Simplified Wells—1.0 (previous PE)
- Original revised Geneva—4.0 (age ≥65, previous PE)
- Simplified revised Geneva—2.0 (age ≥65, previous PE)

Case 2—30 year old woman
Use of oral contraceptives; heart rate 92 bpm
Simply D-dimer negative
Scores of models:
- Original Wells—3.0 (PE most likely diagnosis)
- Modified Wells—2.0 (PE most likely diagnosis)
- Simplified Wells—1.0 (PE most likely diagnosis)
- Original revised Geneva—3.0 (heart rate 75-94 bpm)
- Simplified revised Geneva—1.0 (heart rate 75-94 bpm)

Case 3—25 year old woman
Use of oral contraceptives; heart rate 80 bpm
Simply D-dimer negative
Scores of models:
- Original Wells—0.0
- Modified Wells—0.0
- Simplified Wells—0.0
- Original revised Geneva—3.0 (heart rate 75-94 bpm)
- Simplified revised Geneva—1.0 (heart rate 75-94 bpm)

PE=pulmonary embolism; bpm=beats per minute.
study in a secondary (or tertiary) care setting. The failure rates observed in our validation study in primary care are largely in line with the previous studies regarding the performance of the Wells rules in combination with D-dimer testing, as can be appreciated from figure 3: most secondary care studies yielded a failure rate of the diagnostic approach of combining the Wells rule with D-dimer testing around 1% (range 0.5-2.2%), whereas in our study we observed a failure rate close to 1.5% for the Wells rules combined with D-dimer testing.

Contrastingly, for the (simplified) revised Geneva scores, the failure rates observed in this primary care cohort are slightly higher than those in the previous studies (fig 3). Reasons for this are not entirely clear, but one can think of the differences in cohort characteristics that might have influenced the diagnostic performance. Importantly, the revised Geneva scores have been developed in a tertiary care setting, with a pre-selected and thus more severe disease presentation. Conversely, an unselected patient population and a lower prevalence of pulmonary embolism are typical features of a primary care setting, and such differences in case mix of patients can influence the diagnostic performance of the rule.

Clinical implications and conclusions

On the basis of our findings of a (slightly) higher failure rate for the Geneva rules, we propose that these rules are less suitable for use in this particular setting, specifically when compared with the Wells rules. We would suggest that general practitioners use the (simplified version of the) Wells rule, combined with a (point of care) D-dimer test. Pulmonary embolism can be excluded in about four in every 10 patients with suspected pulmonary embolism, with an acceptably low failure rate of below 2%.

We thank AMUSE-2 project members R Oudega, H ten Cate, and M H Prins for their contribution to the design and initiation of the AMUSE-2 cohort. We thank Joris de Groot for his support with using the DeLong method and Peter Zuthoff and Karlijn Groenewegen for their statistical input.

Contributors. HEJHS, HCPMwV, HRB, ANW, and KGMM had the original idea for the study and were involved in writing the original study protocol. GJG, PMGE, and WAML were involved in data collection. JMTH and GJG drafted the first version of the manuscript, which was subsequently revised by the other authors. All authors participated in the final approval of the manuscript. JH, GJG, PMGE, and WAML had full access to all of the data in the study. JMTH and GJG are guarantors.

Funding: KGMM received a grant from The Netherlands Organization for Scientific Research (ZONMW 918.10.615 and 912080004). GJG is supported by a VENI grant from The Netherlands Organization for Scientific Research (ZONMW 016.166.030). All funding sources had no role in the design, conduct, analyses, or reporting of the study or in the decision to submit the manuscript for publication.

Competing interests: All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf (available on request from the corresponding author) and declare: no support from any organisation that might have an interest in the submitted work.

Ethical approval: Not required.

Transparency: The lead authors (the manuscript’s guarantors) affirm that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned have been explained.

Data sharing: Additional data are available on request from the corresponding author.


© BMJ Publishing Group Ltd 2015