Cancer and thrombosis
van Es, N.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
CHAPTER 9

THE ORIGINAL AND SIMPLIFIED WELLS RULES AND AGE-ADJUSTED D-DIMER TESTING TO RULE OUT PULMONARY EMBOLISM: AN INDIVIDUAL PATIENT DATA META-ANALYSIS

Nick van Es, Noémie Kraaijpoel, Frederikus A. Klok, Menno V. Huisman, Paul L. den Exter, Inge C. Mos, Javier Galipienzo, Harry R. Büller, and Patrick M. Bossuyt

Journal of Thrombosis and Haemostasis, 2017
Supplemented with data from: Thrombosis Research, 2016 Apr; 140 Suppl 1:S179
ABSTRACT

BACKGROUND The Wells score and D-dimer testing can safely rule out pulmonary embolism (PE). A simplification of the Wells score has been proposed to improve clinical applicability, but evidence on its performance is scarce.

OBJECTIVES To compare the performance of the original and simplified Wells scores alone and in combination with age-adjusted D-dimer testing.

METHODS Individual patient data from 7,268 patients with suspected PE enrolled in 6 management studies were used to evaluate the discriminatory performance of the original and simplified Wells scores. The efficiency and failure rate of the dichotomized original and simplified scores combined with age-adjusted D-dimer testing were compared using a one-stage random effects meta-analysis. Efficiency was defined as the proportion of patients in whom PE could be considered excluded based on a ‘PE unlikely’ Wells score and a negative age-adjusted D-dimer. Failure rate was defined as the proportion of patients with symptomatic venous thromboembolism during 3-month follow-up.

RESULTS The discriminatory performance of the original and simplified Wells scores was comparable (c-statistic 0.73 [95% CI 0.72-0.75] vs. 0.72 [95% CI 0.70-0.73]). When combined with age-adjusted D-dimer testing, the original and simplified Wells rules had comparable efficiency (33% [95% CI 25-42%] vs 30% [95% CI 21-40%]) and failure rates (0.9% [95% CI 0.6-1.5%] vs. 0.8% [95% CI 0.5-1.3%]).

CONCLUSION The original and simplified Wells rules combined with age-adjusted D-dimer testing have similar performance in ruling out pulmonary embolism. Given its ease of use in clinical practice, the simplified Wells rule is to be preferred.
INTRODUCTION

In patients with clinically suspected pulmonary embolism (PE), objective diagnosis is mandatory to identify those who require anticoagulant treatment. The clinical signs and symptoms of PE are, however, non-specific. Computed tomography pulmonary angiography or ventilation-perfusion scanning are therefore warranted to confirm the diagnosis, but associated costs, exposure to ionizing radiation, risk of contrast-induced nephropathy, and potential for overdiagnosis make them less desirable as a first-line test. This has led to the recommendation to start the diagnostic management of suspected acute PE with assessment of clinical pretest probability followed by D-dimer testing to reduce the number of imaging tests.¹ For example, imaging can be safely withheld in approximately 28% of patients who are classified as ‘PE unlikely’ by the Wells rule and have a D-dimer result below 500 μg/L.² Imaging can be safely withheld in an additional 5% of patients by applying an age-adjusted D-dimer positivity threshold, defined as a patient’s age times 10 μg/L for those older than 50 years. This age-adjustment increases the specificity of D-dimer testing in elderly patients.²³

Despite the recommendation to use a well-validated algorithm in the diagnostic work-up of PE, the threshold to perform imaging appears to have been lowered over the years, probably due to the fear of missing a potentially fatal diagnosis, facilitated by the readily and widespread availability of computed tomography at emergency departments. It is also possible that using a clinical decision score, such as the Wells score, may be perceived as too complicated for use at the bedside, as seven clinical items have to be remembered, assessed, assigned different weights, and then summed.

To partly overcome the practical limitations of the original Wells score, a simplification has previously been proposed, in which only 1 point is assigned to each item.⁴ As a consequence, when using the score at the recommended positivity threshold of 2 points, physicians do not have to sum points for the individual items anymore, but rather evaluate whether two or more features are present or not; if so, the patient has to be referred for imaging. This simplified Wells rule appeared to be as safe and efficient as the original score when combined with fixed D-dimer testing in a prospective management study of 807 patients.⁵ Yet, despite these encouraging results and apparent clinical applicability, the simplified Wells score has received little attention; it is infrequently used in clinical practice.

The objective of the present study was to compare the performance of the simplified and original Wells scores alone as well as in combination with age-adjusted D-dimer testing using individual patient data of more than seven thousand patients with suspected PE, enrolled in six prospective management studies.
METHODS

We used data from 7,268 patients with clinically suspected PE. These had been enrolled in 6 prospective, diagnostic management studies between November 2002 and January 2013. These studies were previously identified in a systematic review; the data collected has been used for an individual patient data meta-analysis of the performance of the original Wells rule and D-dimer testing.

Study selection

Briefly, MEDLINE and EMBASE were searched from January 1st, 1998 to February 13th, 2016 for PE diagnostic management studies in secondary care that had enrolled consecutive, hemodynamically stable adults with signs and symptoms suggestive of acute PE. All patients had to be assessed by the original, dichotomous Wells rule, followed by quantitative D-dimer testing in those classified as ‘PE unlikely’ by the Wells score (4 points or less). Patients with a ‘PE unlikely’ Wells score and a negative D-dimer, i.e. a D-dimer below the protocol-defined threshold, were to be managed without imaging, not given anticoagulants, and followed for 3 months for the occurrence of symptomatic VTE.

Study description

Five studies evaluated the Wells rule in combination with a conventional, fixed D-dimer threshold of 500 μg/L, whereas the remaining study evaluated an age-adjusted D-dimer threshold, defined as a patient’s age times 10 μg/L in those older than 50 years.

D-dimer testing was performed using the locally available method, which could be an enzyme-linked immunosorbent or quantitative latex-based assay. Five different quantitative D-dimer assays were used in the studies: VIDAS (bioMérieux), Tina-quant (Roche), STA-Liatest (Stago), INNOVANCE (Siemens), or D-dimer HS (Instrumentation Laboratory). The study protocols did not mandate D-dimer testing in patients with a ‘PE likely’ Wells score.

Patients with a ‘PE likely’ Wells score or a ‘PE unlikely’ Wells score and a positive D-dimer underwent diagnostic imaging for PE, which was CTPA in 5 studies and CTPA or V/Q-scanning in 1 study. In all studies, patients with a ‘PE unlikely’ Wells score and a negative D-dimer remained untreated and were followed for 3 months by clinic visits or by telephone for the occurrence of symptomatic VTE. Suspected outcomes were adjudicated in 4 of 6 studies.

Risk of bias as assessed by the QUADAS-2 tool was judged to be ‘low’ for patient selection and index test in all studies and ‘unclear’ for the reference standard and flow and timing in 3
Simplified vs original Wells rule for ruling out pulmonary embolism

Studies. No applicability concerns were identified (see previous report for full QUADAS-2 results).

**Simplified vs original Wells score**

The performance of the original and simplified Wells scores was compared (Table 1). First, in each of the 6 studies, a multivariable logistic regression model was fitted with the Wells score items as independent variables and a diagnosis of VTE at baseline or during follow-up as dependent variable. Odds ratios with 95% confidence intervals (CI) across the studies were compared with the odds ratios reported in the original study by Wells and colleagues, to evaluate whether the derivation model may have been overfitted, which could have resulted in inflated weights assigned to the respective variables. The discriminative performance of the original and simplified Wells scores was assessed by calculating the area under the receiver operating characteristic (ROC) curve (i.e. the c-statistic), with 95% CIs calculated according to DeLong’s method. The proportion of patients with confirmed PE was estimated in the groups of ‘PE unlikely’ patients (i.e. one minus negative predictive value) and ‘PE likely’ patients (i.e. positive predictive value).

Table 1. Original and simplified Wells scores

<table>
<thead>
<tr>
<th>Items</th>
<th>Original Wells score (points)</th>
<th>Simplified Wells score (points)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinical signs of lower extremity DVT</td>
<td>+3</td>
<td>+1</td>
</tr>
<tr>
<td>Alternative diagnosis less likely than PE</td>
<td>+3</td>
<td>+1</td>
</tr>
<tr>
<td>Heart rate &gt;100 beats per minute</td>
<td>+1.5</td>
<td>+1</td>
</tr>
<tr>
<td>Surgery of immobilization within 4 weeks</td>
<td>+1.5</td>
<td>+1</td>
</tr>
<tr>
<td>Previously objectively diagnosed PE or DVT</td>
<td>+1.5</td>
<td>+1</td>
</tr>
<tr>
<td>Hemoptysis</td>
<td>+1</td>
<td>+1</td>
</tr>
<tr>
<td>Active cancer</td>
<td>+1</td>
<td>+1</td>
</tr>
<tr>
<td>Classification</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pulmonary embolism unlikely</td>
<td>≤4</td>
<td>≤1</td>
</tr>
<tr>
<td>Pulmonary embolism likely</td>
<td>&gt;4</td>
<td>&gt;1</td>
</tr>
</tbody>
</table>

Abbreviations: DVT indicates deep vein thrombosis; PE, pulmonary embolism.

**Original and simplified Wells rules and age-adjusted D-dimer testing**

Since the Wells score alone cannot safely rule out PE, it has to be used in conjunction with D-dimer testing. Therefore, we evaluated the efficiency and failure rate of the original and simplified Wells rules combined with quantitative D-dimer testing using an age-adjusted D-dimer positivity threshold, defined as a patient’s age times 10 μg/L in those aged 51 years or
older. We focused on age-adjusted D-dimer testing, since this approach has now replaced the conventional, fixed threshold in many centers following the results of the ADJUST-PE study. A comparison between the fixed and age-adjusted D-dimer thresholds was beyond the scope of the present analysis.

Efficiency was defined as the proportion of patients in whom PE could be considered excluded based on a ‘PE unlikely’ Wells score and a D-dimer below the age-adjusted threshold. The failure rate was defined as the number of these patients who were diagnosed with PE at baseline or with symptomatic VTE during 3-month follow-up, relative to all patients with a ‘PE unlikely’ Wells score and a D-dimer below the age-adjusted threshold. Patients who received anticoagulation for other indications than VTE or who were lost to follow-up were excluded from the failure rate analysis.

Since by definition the age-adjusted D-dimer threshold is only different from the conventional threshold in patients older than 50 years, we performed an analysis restricted to this group of patients. An analysis restricted to cancer patients was also performed, since it is well-known that clinical decision rules and D-dimer testing may perform differently in this group.

**Statistical analysis**

It is well-known that missing data can lead to biased estimates. Therefore, multiple imputation was used to replace missing values 10 times within each study separately, assuming a missing at random pattern wherein the missingness mechanism depends on the observed data. Results across the imputed datasets were combined using Rubin’s rule.

In all analysis, summary estimates with 95% confidence intervals were obtained by a one-stage meta-analysis using generalized linear mixed-effects models. A random intercept was specified for each study to account for the clustering of observations within studies. To assess between-study heterogeneity, a 90% prediction interval was calculated around the estimates of the mean. This prediction interval reflects both the statistical uncertainty around the estimate and the between-study variability. The 95% confidence interval for the absolute difference in efficiency was obtained by repeating the efficiency analyses in 500 bootstrap samples.

R, version 3.3.1, (R Foundation for Statistical Computing, Vienna, Austria, www.R-project.org) was used for all analyses, in particular using the mice package, version 2.21, for multiple imputation, and the lme4 package, version 1.1-12, for the generalized linear mixed-effects models.
RESULTS

The study group comprised 7,268 patients with clinically suspected PE previously enrolled in 6 diagnostic management studies. Baseline characteristics are summarized in Table 2. The mean age was 56 years and 58% were female. Multiple imputation was used to replace missing quantitative D-dimer in 104 patients with a ‘PE unlikely’ Wells score (2%) and to calculate the simplified Wells score in 128 patients (1.8%). The proportion of patients with confirmed PE ranged from 13% to 42% across the studies (mean: 23%; 95% CI, 17 to 30).

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Overall (N=7,268)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean age (SD), y</td>
<td>56 (18)</td>
</tr>
<tr>
<td>Age &gt;50 years, n (%)</td>
<td>4,598 (63)</td>
</tr>
<tr>
<td>Female, n (%)</td>
<td>4,239 (58)</td>
</tr>
<tr>
<td>Inpatient, n (%)</td>
<td>804 (11)</td>
</tr>
<tr>
<td>Active cancer, n (%)</td>
<td>938 (13)</td>
</tr>
<tr>
<td>Previous venous thromboembolism, n (%)</td>
<td>1,116 (15)</td>
</tr>
<tr>
<td>Clinical signs of lower extremity DVT, n (%)</td>
<td>297 (4)</td>
</tr>
<tr>
<td>Alternative diagnosis less likely than PE, n (%)</td>
<td>3,812 (52)</td>
</tr>
<tr>
<td>Heart rate &gt; 100 beats per minute, n (%)</td>
<td>1,716 (24)</td>
</tr>
<tr>
<td>Surgery of immobilization within 4 weeks, n (%)</td>
<td>1,259 (17)</td>
</tr>
<tr>
<td>Previously objectively diagnosed PE or DVT, n (%)</td>
<td>1,116 (15)</td>
</tr>
<tr>
<td>Haemoptysis, n (%)</td>
<td>359 (5)</td>
</tr>
<tr>
<td>Active cancer, n (%)</td>
<td>938 (13)</td>
</tr>
<tr>
<td>Original Wells score classification</td>
<td></td>
</tr>
<tr>
<td>PE unlikely (≤4 points)</td>
<td>5223 (72)</td>
</tr>
<tr>
<td>PE likely (&gt;4 points)</td>
<td>2045 (28)</td>
</tr>
<tr>
<td>Simplified Wells score classification</td>
<td></td>
</tr>
<tr>
<td>PE unlikely (≤1 point)</td>
<td>293 (4)</td>
</tr>
<tr>
<td>PE likely (&gt;1 point)</td>
<td>4514 (62)</td>
</tr>
<tr>
<td>Missing, n (%)</td>
<td>128 (1.8)</td>
</tr>
<tr>
<td>D-dimer</td>
<td></td>
</tr>
<tr>
<td>Median in µg/L (interquartile range)</td>
<td>780 (350-1850)</td>
</tr>
<tr>
<td>Missing, n (%)</td>
<td>995 (14)</td>
</tr>
<tr>
<td>Missing in ‘PE unlikely’ patients, n (%)</td>
<td>104 (2)</td>
</tr>
</tbody>
</table>

Abbreviations: DVT, deep vein thrombosis; PE, pulmonary embolism; SD, standard deviation.

Comparison of the original and simplified Wells scores

The most frequently observed Wells score element was the subjective item ‘alternative diagnosis less likely than PE’ (52%), whereas the items ‘clinical signs of lower extremity DVT’ (4%) and ‘hemoptysis’ (5%) were infrequently seen. The results of the multivariate analysis demonstrated that the odds ratios for the Wells score items were often lower in the present
analysis than in the derivation study by Wells and colleagues (Figure 1). The area under the ROC-curve of the original Wells score (0.73; 95% CI, 0.72 to 0.75) was comparable to that of the simplified Wells score (0.72; 95% CI, 0.70 to 0.73; Figure 2). The proportion of confirmed PE among patients classified as ‘PE unlikely’ and ‘PE likely’ by the original Wells rule was 15% (95% CI, 12 to 19) and 42% (95% CI, 38 to 46), respectively. For the simplified Wells rule, the proportion of ‘PE unlikely’ patients with confirmed PE was 12% (95% CI, 10 to 15) and the proportion of ‘PE likely’ patients with confirmed PE was 37% (95% CI, 33 to 42).

Figure 2. Receiver operating characteristics curves of the original and simplified Wells scores
Figure 1. Adjusted odds ratios with 95% confidence intervals for the Wells score items

The y-axis was capped at 15, but the upper limit of the 95% confidence interval of the odds ratio for ‘signs and symptoms of deep vein thrombosis’ in the study by Goekoop et al. was 37. The solid horizontal line indicates an odds ratio of 1.

Abbreviations: CI, confidence interval; DVT, signs of deep vein thrombosis; ALT, alternative diagnosis less likely than pulmonary embolism; HR, heart rate >100 beats per minute; IMMO, recent surgery or immobilization; VTE, previous venous thromboembolism; HPT, hemoptysis; MAL, active malignancy.
Performance of the original and simplified Wells rules and D-dimer testing

In the pooled analysis, 70% of patients (95% CI, 57 to 80; 90% PI, 32 to 92) were classified as ‘PE unlikely’ by the original, dichotomous Wells rule compared to 58% (95% CI, 43 to 72; 90% PI, 20 to 89) by the simplified, dichotomous Wells rule. Overall, 9.4% of patients were reclassified from 'PE unlikely' to 'PE likely' when applying the simplified Wells rule.

The proportion of patients in whom imaging could have been withheld based on a ‘PE unlikely’ score and a D-dimer below the age-adjusted threshold was comparable with the original Wells rule and with the simplified Wells rule. The mean efficiency was 33% (95% CI, 25 to 42) with the original Wells rule, while the 90% prediction interval (PI) extended from 18% to 52%. With the simplified Wells rule, the mean efficiency was 30% (95% CI, 21 to 40; 90% PI, 14 to 51). As is evident from the width of the prediction intervals, considerable heterogeneity was observed in both analyses. The absolute difference in efficiency between the original and simplified Wells rules was 2.8% (95% CI, 2.1 to 3.5).

The failure rates of the original and simplified Wells rules combined with age-adjusted D-dimer testing were also comparable: mean failure rate for the original Wells 0.9% (95% CI, 0.6 to 1.5), with a 90% PI from 0.6% to 1.4%, versus a mean of 0.8% for the simplified rule (95% CI, 0.5 to 1.3), and a 90% PI from 0.6% to 1.2%.

The efficiency of both rules was analyzed in the subgroup of 4,598 patients older than 50 years and in the subgroup of 938 cancer patients. The efficiency of the original and simplified Wells rules in combination with age-adjusted D-dimer testing was comparable in patients older than 50 years, with a mean of 25% (95% CI, 19 to 33) for the original, and a 90% PI from 10% to 49%, versus a mean of 22% (95% CI, 15 to 31) with the simplified, and a 90% PI from 7% to 51%. The respective failure rates were comparable as well, with a mean of 1.4% for the original (95% CI, 0.6 to 2.9), and a 90% PI from 0.4% to 4.9%, versus a mean of 1.3% (95% CI, 0.6 to 2.6) for the simplified Wells rule, with a 90% PI from 0.4% to 3.7%. Among cancer patients, the efficiency of the original Wells rule combined with age-adjusted D-dimer testing was 13% (95% CI, 11% to 15%) compared to 3.9% (95% CI, 2.0% to 7.6%) of the simplified Wells rule. The failure rates when applying the original and simplified Wells rules were 1.5% (95% CI, 0.13 to 15) and 2.4% (95% CI, 0.34 to 15), respectively.
DISCUSSION

The findings of the present analysis indicate that the original and simplified Wells rule have similar performance in the diagnostic management of patients with clinically suspected acute PE. When combined with age-adjusted D-dimer testing, both rules can rule out the disease without additional imaging in approximately one-third of patients, while less than 1% will be subsequently diagnosed with VTE.

A strong point of this analysis is that we were able to combine individual patient data from more than 7,000 patients with clinically suspected PE who were enrolled in 6 prospective outcome studies, which were homogeneous in design. All patients were managed by the Wells rule and quantitative D-dimer testing, currently one of the recommended initial diagnostic approaches for PE. Modern statistical methods were used to impute missing data and obtain estimates in a one-stage random-effects meta-analysis.

Some limitations of our study require consideration. The simplified Wells score and age-adjusted D-dimer threshold were calculated post-hoc for most of the patients, while management decisions had been based on the original Wells rule and conventional D-dimer testing in the original studies. However, since the Wells score items were systematically assessed in all studies prior to D-dimer testing, we feel that the post-hoc calculated simplified Wells scores are likely to be the same as if they were calculated prospectively. By applying the age-adjusted D-dimer threshold post-hoc, those patients with a ‘PE unlikely’ Wells score and a D-dimer between the fixed and age-adjusted threshold were considered to have a negative D-dimer result, while in fact they had received imaging at baseline. As a consequence, the failure rate may have been overestimated due to the detection of pulmonary emboli with sometimes uncertain clinical significance.

As illustrated by the relatively wide prediction intervals around the estimates, considerable between-study heterogeneity was observed in the performance of the diagnostic strategies, most likely due to differences in patient population and in the proportion of patients with PE across the studies. Caution is warranted when extrapolating these findings to other populations or healthcare settings with a lower proportion of patients with PE, where both the absolute and relative performance of these algorithms may be different.

We acknowledge that a number of different quantitative D-dimer assays were used in the studies. We were unable to stratify the results by assay, since patient-level information on the assay used was not available. However, all were modern, well-validated quantitative D-dimer assays with similar fixed cut-off levels.
We are also aware that the combination of the simplified Wells rule and age-adjusted D-dimer testing itself was not formally evaluated in a prospective management study and would require validation before application in clinical practice. The Prometheus study, included in the present analysis, is the only study to have prospectively evaluated the simplified Wells rule. Among 807 patients with clinically suspected PE, the simplified Wells rule was as efficient and safe as the original Wells score in combination with conventional D-dimer testing. A subsequent post-hoc analysis of this study demonstrated that the original and simplified Wells rules had comparable efficiency and failure rates when combined with age-adjusted D-dimer testing, which is line with the findings of the present study. Similarly, two other retrospective studies found no significant differences in efficiency and failure rate between the original and simplified Wells scores in combination with conventional D-dimer testing. Taken together, although prospective confirmation is formally needed, it is unlikely that the results will be different from the present analysis.

The similar performance of the simplified Wells rule compared to the original Wells rule was consistent in patients older than 50 years, but not in cancer patients. In the latter group, both the efficiency and safety of the simplified Wells rule were poorer than that of the original Wells rule. A possible explanation is that cancer patients frequently scored two or more items on the Wells score (data not shown). This automatically classifies patients as ‘PE likely’ when using the simplified Wells rule, hence resulting in a lower efficiency. This finding suggests that the original Wells rule may still be preferred in cancer patients.

It is well-known that overfitting due to model and parameter uncertainty can be a serious problem in the derivation of clinical prediction models. Although the number of events in Wells’ derivation study was adequate for valid prediction modelling, still, the applied method of variable selection and lack of use of shrinkage methods may have resulted in overestimated points in the final score. Indeed, odds ratios tended to be lower in the present analysis than in the derivation study. Nevertheless, it appeared that only little predictive performance was lost when all items were assigned unit points. In the setting of VTE, this phenomenon has, for example, also been observed for simplifications of the Pulmonary Embolism Severity Index and revised Geneva score. Selecting the right predictors seems to be more important than assigning the right points to each item.

The practical advantage of the simplified Wells score over the original score is its ease of use in deciding whether a patient should undergo D-dimer testing or can proceed to imaging directly. Rather than assessing seven clinical items at the bedside, assigning different points to each item, and calculating a sum score, the simplified Wells score only requires an
assessment of the clinical items. If two or more of the Wells items are present, the physician can immediately refer the patient for diagnostic imaging.

Some argue that we should not aim for simplification of clinical decision scores now that we have entered an era wherein physicians increasingly use digital devices during their clinical work; their argument is that we should develop even more complex scores to be calculated on a website or in an app. Some discriminatory performance may be lost with an oversimplification of a score. Future research aimed at optimizing the diagnostic management of PE could focus on the development of new decision models with state-of-the-art statistical techniques in which more information, such as quantitative D-dimer results and additional results from physical examination, is incorporated. Until then, given its similar performance, the simplified Wells rule in combination with age-adjusted D-dimer testing presents an attractive alternative to the original score in the diagnostic management of acute PE.
REFERENCES


