

A prospective reassessment of the utility of the Wells score in identifying pulmonary embolism

Kenneth SK Yap, Victor Kalff, Alla Turlakow and Michael J Kelly

Pulmonary embolism (PE) occurs in less than 35% of patients investigated for suspicion of this diagnosis.¹ As indiscriminate investigation of all such patients could result in a major health care burden, there is a need to stratify patients to ensure that they are investigated with appropriate detail and urgency. To help quantify the pre-test probability of PE, clinical scoring systems such as the Wells score² (Box 1) were introduced.

Since the completion of the multidetector computed tomography (MDCT)-based Prospective Investigation of Pulmonary Embolism Diagnosis (PIOPED) II study³ in 2006, diagnostic strategies for PE have changed. The Chair of the PIOPED II study nuclear medicine working group has recommended that the interpretation of all PE imaging results be performed in conjunction with the Wells score and that ventilation/perfusion (V/Q) results be classified as “positive”, “negative” or “inconclusive”.⁴

As the Wells score may now be used both as a pre-imaging tool to help select patients and as a post-imaging tool in conjunction with the interpretation of results, it is uncertain whether such a combined role would change the robustness of the score, given that it was originally developed only as a clinical pre-test probability tool for diagnosing PE. Furthermore, PE imaging techniques have changed with time.

The Wells score, published in 2000, was based on a Canadian multicentre study that used a significantly different V/Q ventilation technique, with xenon (Xe)-133 as the radioactive ventilation tracer and with venous ultrasound as the most common additional imaging investigation. Currently, V/Q ventilation scintigraphy is widely performed using technetium (Tc)-99m Technegas,⁵ with MDCT as the most common additional imaging investigation. The aim of our study was to evaluate the Wells score in these current clinical circumstances, and specifically, to determine its utility and robustness as a clinical pre-test probability tool for identifying or excluding PE.

METHODS

Our study was carried out at the Alfred Hospital, Melbourne, a major metropolitan

ABSTRACT

Objective: To assess whether a cumulative clinical risk score (Wells score) can optimise imaging strategies in patients with suspected pulmonary embolism (PE).

Design, setting and participants: Prospective, consecutive series of 633 studies on 595 patients referred to a major teaching hospital for ventilation/perfusion (V/Q) scanning for suspected acute PE between September 2004 and November 2005. Ventilation scintigraphy was performed using technetium-99m Technegas, and V/Q results were interpreted in conjunction with Wells scores.

Main outcome measures: Likelihood of PE for each Wells score interval; overall prevalence of PE.

Results: The likelihood of PE for a given Wells score in our study was not significantly different from the likelihood in the original study by Wells et al. Scores of < 2 in our study were associated with a 4% risk of PE, scores between 2 and 6 with a 13% risk, and scores > 6 with a 67% risk. The overall prevalence of PE in our study was significantly less than that in the original study (9% v 16%; $P < 0.01$), attributable to a significantly larger proportion of our patients having scores of < 2 (66% v 40%; $P < 0.0001$).

Conclusion: The Wells score remains a robust clinical tool for stratifying the likelihood of PE. Patients with Wells scores of > 2 warrant imaging assessment for PE, but for those with scores < 2, further imaging may be problematic.

MJA 2007; 187: 333–336

For editorial comment, see page 325

teaching hospital. The study protocol was approved by the Alfred Hospital Ethics Committee.

All consecutive adult inpatients and outpatients referred to the Department of Nuclear Medicine between 16 September 2004 and 7 November 2005 for V/Q scintigraphy for evaluation of acute PE were enrolled. Patients referred for other indications were excluded. Patients who re-presented at a later time with renewed suspicion of acute PE were enrolled separately.

The age, sex and referral source of each enrolled patient, and presence of any chest pain, were noted. This was followed by clinical assessment to complete the Wells score,² which was done prospectively for all enrolled patients by the reporting nuclear medicine physician just before scintigraphy. Components of the Wells score are shown in Box 1. A detailed explanation of how these components should be scored has been previously described,^{2,6} especially with respect to completing the subjective component entitled “An alternative diagnosis is less likely than PE”.⁶ This latter component was scored “0” if chest pain was the only symptom, as this symptom correlated negatively

with PE in the study by Wells et al.² For each patient, the total point score of the individual components in Box 1 constitutes the Wells score. We divided Wells scores into three intervals: < 2, 2–6 and > 6.

Every patient then underwent V/Q scintigraphy. A similar technique has been described elsewhere.⁷ In brief, the ventilation phase involved inhalation of Tc-99m Technegas (Vita Medical, Sydney, NSW). The perfusion phase involved intravenous administration of Tc-99m-macroaggregated albumin (CIS-US, Bedford, Mass, USA). Planar static images of ventilation then perfusion of the lungs were acquired in six identical projections.

V/Q scans were reported as positive or negative according to the criteria shown in Box 2. The remainder of the scans were defined as inconclusive, with recommendation to the treating unit for further evaluation using MDCT scanning. All patients were followed up at least 1 week after V/Q scintigraphy to determine whether MDCT had been performed and, if so, what the result was.

Final classification of PE for each presentation was “positive” if either the V/Q or MDCT result was positive; “negative” if the

1 Components of the simplified Wells score*²

Component	Points
Clinical signs and symptoms of DVT (minimum of leg-swelling and pain with palpation of the deep veins)	3.0
An alternative diagnosis is less likely than PE	3.0
Heart rate greater than 100 beats/min	1.5
Immobilisation (complete bedrest for ≥ 3 days in the 4 weeks before presentation) or surgery in the previous 4 weeks	1.5
Previous objectively diagnosed DVT or PE	1.5
Haemoptysis	1.0
Malignancy (receiving treatment, treated in the last 6 months or receiving palliative care)	1.0

DVT = deep vein thrombosis. PE = pulmonary embolism. *A total score of < 2 implies a low clinical pre-test probability of PE; a score of 2–6 represents a moderate clinical pre-test probability of PE; a score of > 6 represents a high clinical pre-test probability of PE.²

V/Q result was negative or, if initially inconclusive, the follow-up MDCT result was negative; or “inconclusive” if an initially inconclusive V/Q result was not followed up by an MDCT evaluation.

To allow comparison with the original study by Wells et al, we combined and retrospectively analysed the published results from the study's derivation and validation sets.² The rate of PE in our study was determined and compared with the original study results in total and according to the score intervals (< 2 , 2–6 and > 6). The negative predictive value (NPV) for scores < 2 and positive predictive value (PPV) for scores > 6 were calculated. Subanalysis of the proportion of patients diagnosed with PE who also had chest pain was compared with those without chest pain. Comparisons between PE rates were made using the χ^2 test, with $P < 0.05$ being considered significant.

RESULTS

During the period of the study, 1044 V/Q scan referrals were received, of which 411 were excluded because the indication was not for evaluation of acute PE. The remaining 633 studies included in the analysis represented 595 patients. The patients had a mean age of 60 years (range, 16–95 years), with a male : female scan ratio of 319 : 314. The number of scans of patients referred as inpatients, outpatients and from the emergency department was 349, 86 and 198, respectively.

Investigation results and final PE diagnoses of the 633 initial referrals are shown in Box 3. After exclusion of the eight inconclusive cases, there were 625 conclusive results. Fifty-four of the 625 patients (9%)

were diagnosed with PE, compared with 201/1219 (16%) in Wells and colleagues' original study ($P < 0.01$). A significantly larger proportion of patients scored < 2 in our study (415/625 [66%]) than in the original study (491/1219 [40%]) ($P < 0.0001$).²

In addition to the 12 patients with inconclusive results who had both V/Q and MDCT evaluation (Box 3), a further 25 patients with *conclusive* V/Q results also underwent MDCT: of the 15 patients with negative V/Q results, 14 had negative and one had inconclusive results on MDCT; of the 10 patients with positive V/Q results, eight had positive, one had inconclusive and one had negative results on MDCT.

In our study, patients with scores of < 2 were at low risk of PE, as only 4% of these patients had PE (Box 4), with an associated NPV of 96%. Patients with scores of 2–6 had a PE rate of 13% and were thus at moderate

risk of PE. Patients with scores > 6 were at high risk, as 67% of these patients were diagnosed with PE, implying that the PPV of scores > 6 was also 67%. PE rates among patients with low (< 2) and high (> 6) Wells scores in our study were not significantly different from rates in Wells and colleagues' original study (Box 4). The difference in PE rates among patients with moderate Wells scores (2–6) between our study (13%, 26/195) and the original study (20%, 129/639) achieved borderline statistical significance ($P < 0.05$).

Subanalysis of patients with and without chest pain showed PE rates of 7.1% and 9.6%, respectively. For patients with Wells scores of < 2 , the rates were 3.4% and 5.0%, respectively. Neither of these differences were statistically significant.

DISCUSSION

Our principal finding was that a given Wells score in our study predicted an equivalent rate of PE to that observed in Wells and colleagues' original study. These findings add confirmation to the robust nature of the Wells score. In particular, a Wells score of > 2 predicted a $\geq 13\%$ likelihood of PE, while a Wells score of < 2 was associated with a $< 5\%$ risk of PE and a high NPV.

A second finding is that a significantly higher proportion of patients referred for imaging investigations in our study scored < 2 (66%) than in the original study (40%) ($P < 0.0001$).

The strengths of our study were that it was a prospective evaluation in a large population of consecutively enrolled patients with suspected PE. Our results reaffirmed the value of the Wells scoring system in the current imaging and clinical environment.

2 Criteria for interpreting positive and negative ventilation/perfusion (V/Q) scans

Positive V/Q result

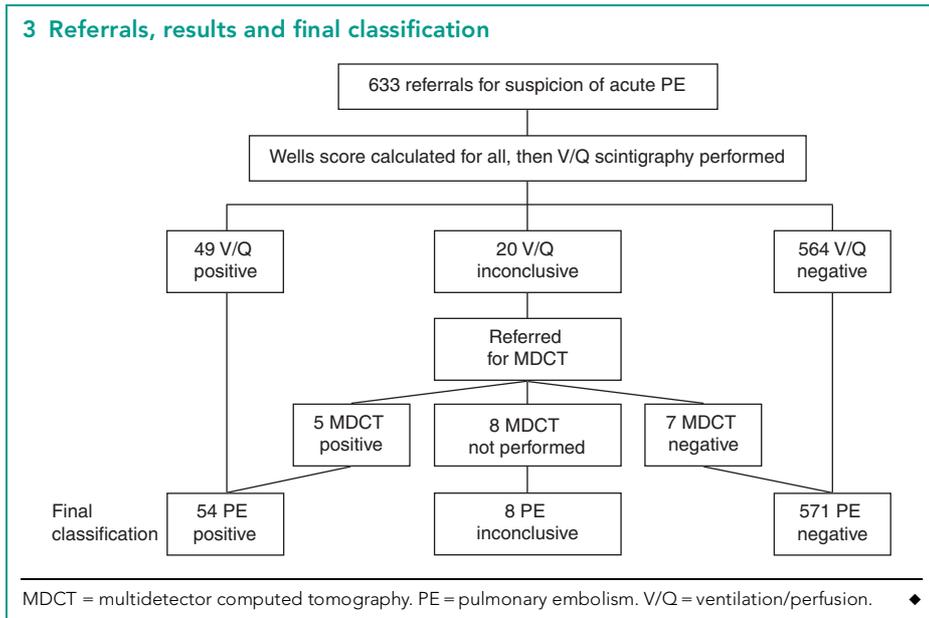
Multiple or new unmatched segmental perfusion defects

Negative V/Q result

Any one of the following:

- Normal perfusion
- Features of a very low-probability V/Q scan⁸
 - Nonsegmental perfusion abnormality
 - Perfusion defect smaller than the corresponding radiographic defect
 - Matched V/Q defects in two or three zones of a single lung (normal radiograph)
 - One to three small segmental perfusion defects
 - Tripled matched defect in the upper or middle lung zone
 - Stripe sign⁹
- “Reverse mismatch” (ie, a perfusion defect associated with a correspondingly larger ventilation defect)¹⁰

3 Referrals, results and final classification



Furthermore, in contrast to the original study, which used Xe-133 as the tracer for the ventilation scan, our study used Tc-99m Technegas, which markedly reduces the proportion of non-diagnostic results.^{5,7} In line with current recommendations,⁴ we interpreted scans in conjunction with the Wells score and used MDCT as the most frequent follow-up modality, rather than venous ultrasound, which was used in the original study.²

Weaknesses of our study included the absence of prolonged follow-up of all negative imaging results to identify any false negatives. In particular, we did not determine whether all patients with negative results were investigated for deep vein thrombosis, which is often used as a proxy for classifying patients as PE positive. However, other studies that have followed patients for up to a year after exclusion of venous thromboembolism have shown only about a further 1% incidence of positive cases.^{11,12} Furthermore, among the 15 patients with negative V/Q results who were also assessed by MDCT, no further PE was identified. These considerations suggest that the number of missed diagnoses is likely to represent only a small proportion of the entire study and would be unlikely to introduce a major bias in results.

A potential source of bias is the fact that the reporting nuclear medicine physician was responsible for generating both the Wells score and the V/Q report — raising the possibility that knowledge of the Wells scores may have influenced reporting. Our results suggest that, if such an effect was

present, it was not enough to significantly change the relationship between given Wells scores and their associated PE rate, as our results were similar to those obtained by Wells et al (Box 4).

In the clinical presentation spectrum of PE, chest pain remains a common presenting symptom. The component of the Wells score entitled “An alternative diagnosis is less likely than PE” is subjective, yet represents a significant part of the total score. It may sometimes be inappropriately allocated positively when chest pain is the only presenting symptom. In our study, as in the studies by Wells et al² and Wicki et al,¹³ chest pain was not found to be significantly associated with PE. Hence, when a Wells score is applied, chest pain alone should be scored “0” for its subjective component.

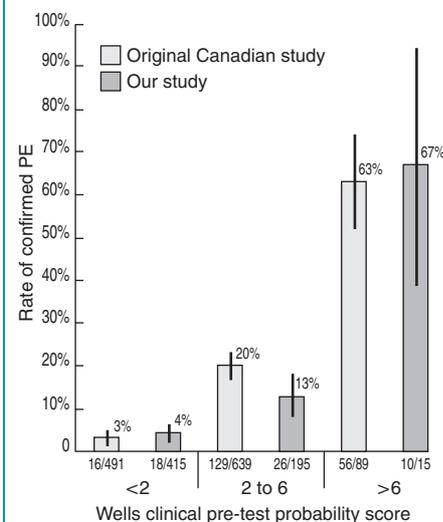
The proportion of inconclusive V/Q scans reported in our study (20/633 [3.2%]) is lower than that reported in earlier studies that used Xe-133.¹⁴ However, our findings are consistent with those of other studies performed using Tc-99m Technegas.⁷ The latter agent allows improved matching of V/Q abnormalities, resulting in fewer inconclusive results. It is also to be expected that the lower prevalence of PE in our study population, as reflected by the lower total proportion of positive results, would reduce the number of inconclusive results. Another important factor is likely to be our application of the currently recommended interpretation principles.⁴ In our study, the ratio of inconclusive to positive V/Q results (20 : 49 [0.41]) was not statistically different from

the same ratio for MDCT in the PLOPED II population (51 : 175 [0.29]).³

Our data suggesting that a Wells score of >2 predicts a high enough likelihood of PE to justify urgent imaging are consistent with the findings of the original study by Wells et al. They are also consistent in showing that a Wells score of <2 has an NPV for PE of over 95%.

It is a notable and somewhat surprising feature of our study and other recent studies that the majority of patients being investigated are actually at low risk for PE (ie, have a Wells score of <2).^{15,16} In 1998, Wells and colleagues commented that, because the prevalence of PE in patients with a low clinical probability of having the disease is not dissimilar to the rate of PE in patients with normal and near-normal V/Q scans, it may not be worthwhile performing further imaging investigations in such patients.⁶ Sijens et al,¹⁷ commenting on a subsequent study by Wells et al,¹⁵ concluded that patients with Wells scores of <2 needed no further diagnostic imaging or D-dimer testing. The absolute risk of PE for patients with a Wells score of <2 in our study (4%) was similar to the risk of PE in patients who have a V/Q scan reported as being very low-probability for PE (2.5%)¹⁴ and to the 2.6% rate found incidentally in cancer patients

4 Rates of PE confirmed by imaging for three clinical pre-test probability intervals (Wells scores): rates in the original study by Wells et al² compared with rates in our study*



without PE symptoms.¹⁸ It is also less than the 9% PE rate found in the total population of patients with a negative MDCT scan in the PIOPED II study.³ Thus, one could argue that, in patients with a Wells score of <2, further imaging may be unwarranted based on their low prevalence of PE alone. It is important to emphasise this point, as the frequency of further investigation in such low-risk patients has substantially increased since the original study done by Wells et al.^{15,16}

While diagnosis of PE in low-risk situations may still be considered desirable, there are also potential problems with false positive imaging results and the need to balance the low but fixed imaging investigation risks against the low risk of PE in this situation. In the PIOPED II data, the PPV of MDCT for PE in patients with Wells scores of <2 was only 58%,³ indicating that a large proportion of positive results in these patients are false positives. This suggests that overly aggressive investigation of these patients may have not only low economic value but also problematic diagnostic yield. False positives may lead to complications from unnecessary anticoagulation and risks from imaging investigations that may not be negligible in certain patient groups. These include toxicity related to the contrast medium, especially in patients with renal impairment, and breast radiation exposure,¹⁹ in the case of women of reproductive age. The absorbed dose to the breast with CT angiography has been calculated as 10–50 mGy.^{19–21} This compares with the absorbed dose to the breast with perfusion scintigraphy (which represents the large majority of the radiation dose from the entire V/Q scan) of 0.28 mGy,²⁰ standard two-view mammography of 3 mGy,¹⁹ and the normal annual background radiation exposure rate of 1–2 mSv. Breast radiation exposure of >20 mGy has been reported to be associated with detectable excess risk of breast cancer in women of reproductive age.²²

The use of D-dimer testing in combination with Wells scores has been proposed as a way of helping to select patients for whom imaging may be warranted.^{2,15} While it has been shown that negative D-dimer results reduce the likelihood of PE for any Wells score,² the PE rate in patients with scores of <2, even with a positive D-dimer result, was 7.0% in the study by Wells et al.² This is comparable to the PE rate for patients with a negative MDCT scan.³

In conclusion, the results of our study are relevant to Australian clinicians, as they

confirm the Wells score as a robust clinical tool for stratifying PE risk, despite changes in imaging and interpretation techniques and in the population prevalence of PE since the original study by Wells et al. Our results confirm that patients with Wells scores of >2 warrant imaging to assess whether PE is present. However, further imaging investigations in patients with Wells scores of <2 are associated with problematic diagnostic yields and are not without risk.

In low-risk patients, further evaluation based on outcome studies is desirable to provide a clearer understanding of the optimal strategy for balancing diagnostic certainty against the risks of investigation and economic costs.

COMPETING INTERESTS

None identified.

AUTHOR DETAILS

Kenneth SK Yap, MB BS, FRACP, Nuclear Medicine Physician

Victor Kalff, MB BS, FRACP, FACC, Deputy Director

Alla Turlakow, MB BS, FRACP, Nuclear Medicine Physician

Michael J Kelly, MB BS, FRACP, Director Department of Nuclear Medicine, The Alfred Hospital, Melbourne, VIC.

Correspondence: kskyap@yahoo.com

REFERENCES

- Value of the ventilation/perfusion scan in acute pulmonary embolism. Results of the prospective investigation of pulmonary embolism diagnosis (PIOPED). The PIOPED Investigators. *JAMA* 1990; 263: 2753–2759.
- 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: 416–420.
- Stein PD, Fowler SE, Goodman LR, et al. Multi-detector computed tomography for acute pulmonary embolism. *N Engl J Med* 2006; 354: 2317–2327.
- Gottschalk recognized with Cassen prize. *J Nucl Med* 2006; 47: 25N.
- Burch WM, Sullivan PJ, McLaren CJ. Technegas — a new ventilation agent for lung scanning. *Nucl Med Commun* 1986; 7: 865–871.
- Wells PS, Ginsberg JS, Anderson DR, et al. Use of a clinical model for safe management of patients with suspected pulmonary embolism. *Ann Intern Med* 1998; 129: 997–1005.
- Howarth DM, Lan L, Thomas PA, et al. 99mTc technegas ventilation and perfusion lung scintigraphy for the diagnosis of pulmonary embolus. *J Nucl Med* 1999; 40: 579–584.

- Stein PD, Gottschalk A. Review of criteria appropriate for a very low probability of pulmonary embolism on ventilation-perfusion lung scans: a position paper. *Radiographics* 2000; 20: 99–105.
- Sostman HD, Gottschalk A. The stripe sign: a new sign for diagnosis of nonembolic defects on pulmonary perfusion scintigraphy. *Radiology* 1982; 142: 737–741.
- Li DJ, Stewart I, Miles KA, et al. Scintigraphic appearances in patients with pulmonary infection and lung scintigrams of intermediate or low probability for pulmonary embolism. *Clin Nucl Med* 1994; 19: 1091–1093.
- Miniati M, Monti S, Bauleo C, et al. A diagnostic strategy for pulmonary embolism based on standardised pretest probability and perfusion lung scanning: a management study. *Eur J Nucl Med Mol Imaging* 2003; 30: 1450–1456.
- Le Gal G, Righini M, Roy PM, et al. Prediction of pulmonary embolism in the emergency department: the revised Geneva score. *Ann Intern Med* 2006; 144: 165–171.
- Wicki J, Perneger TV, Junod AF, et al. Assessing clinical probability of pulmonary embolism in the emergency ward: a simple score. *Arch Intern Med* 2001; 161: 92–97.
- Worsley DF, Alavi A. Comprehensive analysis of the results of the PIOPED Study. Prospective Investigation of Pulmonary Embolism Diagnosis Study. *J Nucl Med* 1995; 36: 2380–2387.
- Wells PS, Anderson DR, Rodger M, et al. Excluding pulmonary embolism at the bedside without diagnostic imaging: management of patients with suspected pulmonary embolism presenting to the emergency department by using a simple clinical model and D-dimer. *Ann Intern Med* 2001; 135: 98–107.
- Runyon MS, Webb WB, Jones AE, et al. Comparison of the unstructured clinician estimate of pretest probability for pulmonary embolism to the Canadian score and the Charlotte rule: a prospective observational study. *Acad Emerg Med* 2005; 12: 587–593.
- Sijens PE, Oudkerk M. Should patients be managed for suspected pulmonary embolism on the basis of pretest clinical probability and D-dimer results? *Ann Intern Med* 2002; 136: 781.
- Sebastian AJ and Paddon AJ. Clinically unsuspected pulmonary embolism — an important secondary finding in oncology CT. *Clin Radiol* 2006; 61: 81–85.
- Parker MS, Hui FK, Camacho MA, et al. Female breast radiation exposure during CT pulmonary angiography. *AJR Am J Roentgenol* 2005; 185: 1228–1233.
- Cook JV, Kyriou J. Radiation from CT and perfusion scanning in pregnancy [letter]. *BMJ* 2005; 331: 350.
- Managing patient dose in computed tomography. A report of the International Commission on Radiological Protection. *Ann ICRP* 2000; 30: 7–45.
- Land CE. Studies of cancer and radiation dose among atomic bomb survivors. The example of breast cancer. *JAMA* 1995; 274: 402–407.

(Received 20 Nov 2006, accepted 19 Apr 2007) □