Title	Diagnostic Performance of Clinical Prediction Models for Pulmonary Embolism
Agency	AETSA, Andalusian Agency for Health Technology Assessment Av. Innovación s/n. Edificio Renta Sevilla, 2ª planta, 41020 Seville, Spain; Tel: +34 955 407 233, Fax: +34 955 407 238; leda.ojeda.ext@juntadeandalucia.es, www.juntadeandalucia.es/salud/aetsa
Reference	Report no. 3/2008. ISBN 978-84-691-4638-3. www.juntadeandalucia.es/salud/servicios/contenidos/aetsa/pdf/2008_3_Embolismo_PulM.pdf

Aim

To assess the diagnostic performance of scores or clinical models used to determine pre-test probability in diagnosing pulmonary embolism (PE). To describe other outcomes from the scores or models that differ from the diagnostic performance (DP).

Conclusions and results

The search resulted in 428 references, and 11 original observational studies with no control group were finally included in the review. Most assessed more than one score. The accuracy of the scores as opposed to empiricism was studied in 3 good-quality works. The studies did not find any relevant difference in DP of the 2 options when estimating pre-test probability of PE. When comparing the accuracy of scores, the results varied widely.

The 2-category Wells score was more specific than sensitive in 2 good-quality studies; sensitivity varied in both, from 60% to 62%. Nine studies assessed the 3-category Wells score. Two studies assessed it in groups of patients with low frequency of PE, providing different results on sensitivity (S) and specificity (Sp) (92% of S and moderate Sp in one study, and 54% of S in the other), although both studies coincided in high negative predictive value of the score (>96%). In patients with moderate and high frequency of PE, the studies matched in presenting the 3-category Wells score as more sensitive than specific: S was over 90% in patients with moderate frequency (MF) of PE (two studies), and it ranged from 66% to 91% (5 studies) in patients with high frequency (HF) of PE. The Geneva score, assessed in 5 studies, was more sensitive than specific. Its results matched those of Wells scores favorably in the only study that had applied the score prospectively. Both results achieved S >70%. The DP of Wells algorithm, Charlotte rules, and Rodger and Pisa models were each assessed in one study only. Only the Pisa model proved, in a statistically significant way, a superior DP than Wells and Geneva scores (94% of the area under the ROC curve for this model vs 54% for the Geneva score and 75% for Wells score). Despite being a good-quality study, problems related to external validity jeopardize its applicability. These studies reported no results on clinical effectiveness in applying the scores.

Methods

MEDLINE, ECRI, Cochrane Library, CRD, INAHTA, NGC, and EMBASE. Inclusion criteria: adults with suspected PE, intervention (any tool structured to estimate pre-test probability of PE), gold standard (usual and supplementary tests to diagnose PE and/or 3 months clinical follow-up), results (DP and/or clinical outcomes from applying the scores) and design (CT and observational studies with or without control group). We used the QUADAS tool to critically assess the original studies. The data were summarized qualitatively.

Written by Reyes López Márquez and Soledad Márquez Calderón, AETSA, Spain