Measurement of the clinical and cost-effectiveness of non-invasive diagnostic testing strategies for deep vein thrombosis

Goodacre S, Sampson F, Stevenson M, Wailoo A, Sutton A, Thomas S, Locker T, Ryan A

CRD summary
This very large review assessed the clinical and cost-effectiveness of non-invasive testing options for the diagnosis of deep vein thrombosis. It concluded that diagnostic algorithms based on Wells score, D-dimer and ultrasound are likely to be the most feasible and cost-effective option for UK hospitals. These conclusions represent a reasonable interpretation of the data and are likely to be reliable.

Authors' objectives
To estimate the diagnostic accuracy, clinical effectiveness and cost-effectiveness of non-invasive tests and diagnostic algorithms for deep vein thrombosis in patients with clinically suspected deep vein thrombosis or high risk asymptomatic patients.

Searching
MEDLINE, EMBASE, CINAHL, Web of Science, BIOSIS Previews, the Cochrane Library, DARE, NHS EED, HTA database and the American College of Physicians Journal Club were searched to April 2004 for articles in English, Spanish, French, or Italian. Search strategies were reported. The bibliographies of included articles were searched for additional studies and manufacturers were contacted for unpublished studies.

Study selection
Diagnostic cohort studies, with ten or more participants, which assessed the diagnostic accuracy of non-invasive tests for deep vein thrombosis, were eligible for inclusion. Diagnostic tests eligible for inclusion were: individual clinical characteristics or clinical scores; D-dimer assays (latex, enzyme-linked immunosorbent assay/ELISA, whole-blood agglutination); plethysmography and rheography (impedance, strain-gauge and air plethysmography, phleborheography and light-reflex rheography); ultrasound; magnetic resonance imaging (MRI); or computed tomography (CT).

Different reference standards were accepted for different tests: ultrasound studies were required to use venography as the reference standard; plethysmography, rheography, CT and MRI studies were required to use venography or ultrasound as the reference standard; D-dimer and clinical characteristics studies could use any other non-invasive test, or venography as the reference standard. Studies of patients with suspected pulmonary embolism, except for studies of CT, were excluded.

Cohorts were recruited from various settings (inpatient, outpatient, emergency, mixed) and the mean age of participants, in the majority of studies, ranged from 39 to 70 years.

Studies were independently screened for inclusion by two reviewers, a kappa score was calculated, and disagreements were resolved by discussion.

Assessment of study quality
The methodological quality of included studies was assessed using three criteria: whether the reference standard was applied independently of the results of the non-invasive test; whether those performing and interpreting the non-invasive test were blind to the results of the reference standard; and whether those performing and interpreting the reference standard were blind to the results of the index test.

The authors did not state how quality assessment was performed, or how many reviewers were involved in this process.

Data extraction
Data were extracted on: prevalence of deep vein thrombosis (separately for proximal and distal, if reported); reported
sensitivity and specificity of the non-invasive test and numbers of true positives, false negatives, false positives and true negatives; inter-observer variation for the non-invasive test. Separate estimates of sensitivity for all deep vein thrombosis, proximal deep vein thrombosis and distal deep vein thrombosis, were made where possible. Specificity was always defined as the ability of a test to correctly identify cases without any deep vein thrombosis (i.e. for all deep vein thrombosis).

The authors did not state how data were extracted, or how many reviewers performed the data extraction.

Methods of synthesis

For each non-invasive test, pooled estimates of sensitivity and specificity, with 95% confidence intervals, were calculated separately for patients with clinically suspected deep vein thrombosis, asymptomatic cohorts and mixed cohorts, using a random-effects model. Where zero counts occurred, a continuity correction of 0.5 was added to all values from that study. Between study heterogeneity was assessed using a \( \chi^2 \) test.

For individual clinical characteristics, pooled estimates of likelihood ratios and their 95% confidence intervals were calculated in place of sensitivity and specificity.

For clinical scores, ordinal logistic regression was used to estimate the probability of a classification of high risk, intermediate risk, and low risk. These were based upon the score, using separate models for participants with any deep vein thrombosis, proximal deep vein thrombosis, distal deep vein thrombosis, and no deep vein thrombosis. Sensitivity, specificity and likelihood ratios were then estimated for the two thresholds, high versus intermediate and low risk, and high and intermediate versus low risk.

Where numbers of studies were sufficient, meta-regression was used to investigate potential sources of heterogeneity in the diagnostic performance of individual tests; subgroups of studies with significant co-variates were analysed separately. For clinical scores, the ordinal regression model was extended to fit a summary receiver operating characteristic curve. The influence on the curve of adding covariates to the model was explored.

Results of the review

Clinical characteristics (29 studies; 13 individual clinical characteristics): The prevalence of deep vein thrombosis ranged from 12 to 70% and, where reported, proximal deep vein thrombosis ranged from 18 to 92%. The reference standard always included ultrasound or venography and was applied to 28 studies independently of clinical characteristics. Blinding was reported in a minority of studies. All individual clinical characteristics had little diagnostic value; both positive and negative likelihood ratios were close to one in all cases.

Clinical probability scores (22 studies): The majority of studies assessed the performance of the Wells score and used ultrasound as the reference standard. The prevalence of deep vein thrombosis ranged from 10 to 47%. Proximal deep vein thrombosis ranged between 46 and 92%, where reported. The reference standard was applied independently of clinical probability in 17 studies. Clinical probability was estimated blind to the reference standard result in eight studies. The reference standard was interpreted blind to the clinical probability estimate in three studies. The Wells probability score (21 studies) had a positive likelihood ratio of 5.4 (95% confidence interval (CI): 4.1 to 7.2) for high risk versus intermediate and low risk, and a negative likelihood ratio of 0.25 (95% CI: 0.21 to 0.29) for low risk vs. intermediate and high risk. Significant between study heterogeneity was observed. Results of meta-regression were reported in the review. Empirical assessment of clinical probability produced similar likelihood ratio estimates (four studies).

D-dimer (110 articles reported 213 data sets; 198 clinically suspected deep vein thrombosis; 15 asymptomatic participants; one to 13 assays tested per study): In cohorts with clinically suspected deep vein thrombosis, the prevalence of deep vein thrombosis ranged from 2 to 78% and proximal deep vein thrombosis ranged from 27 to 100%, where reported. The reference standard was applied independently of the results of D-dimer testing in 86 cohorts, D-dimer was measured blind to the reference standard result in 43 cohorts; the reference standard was interpreted blind to the D-dimer test result in 50 cohorts. For all deep vein thrombosis and all D-dimer assays, sensitivity was 90% (95% confidence interval (CI): 89 to 91) and specificity was 55% (95% CI: 54 to 56). For enzyme-linked immunosorbent assay (ELISA, 58 cohorts, 91data sets), sensitivity was 94% (95% CI: 93 to 95) and specificity was 45% (95% CI: 44 to
46). For latex agglutination assays (52 cohorts, 74 data sets), sensitivity was 89% (95% CI: 88 to 90) and specificity was 55% (95% CI: 54 to 56). For whole-blood agglutination assays (29 cohorts, 29 data sets), sensitivity was 87% (95% CI: 85 to 88) and specificity was 68% (95% CI: 67 to 69). All analyses reporting separate data for proximal and distal deep vein thrombosis showed higher sensitivity for proximal deep vein thrombosis. Specificity appeared to be dependent on pre-test clinical probability, as defined by the Wells score, being higher in low-risk participants. Full results of meta-regression and data for asymptomatic cohorts were reported.

**Plethysmography and rheography** (88 articles reported 98 data sets; 82 clinically suspected deep vein thrombosis; two mixed cohorts; 14 asymptomatic participants): In cohorts with clinically suspected deep vein thrombosis, the prevalence of deep vein thrombosis was 15 to 83% and distal deep vein thrombosis ranged from 0 to 58%, where reported. For all deep vein thrombosis: impedance plethysmography (42 cohorts) sensitivity was 75% (95% CI: 73 to 77) and specificity was 90% (95% CI: 89 to 91); strain-gauge plethysmography (20 cohorts) sensitivity of was 83% (95% CI: 81 to 85) and specificity was 81% (95% CI: 79 to 82); air plethysmography (4 cohorts) sensitivity was 85% (95% CI: 79 to 90) and specificity was 91% (95% CI: 81 to 95); light-reflex rheography (9 cohorts) sensitivity was 91% (95% CI: 87 to 94) and specificity was 71% (95% CI: 66 to 75); phleborheography (7 cohorts) sensitivity was 86% (95% CI: 83 to 89) and specificity was 93% (95% CI: 91 to 95). All analyses reporting separate data for proximal and distal deep vein thrombosis showed higher sensitivity for proximal deep vein thrombosis, with the exception of light reflex rheography where values were similar. Full results of meta-regression and data for asymptomatic cohorts were reported.

**Ultrasound** (142 articles reported 150 data sets; 100 clinically suspected deep vein thrombosis; five mixed cohorts; 45 asymptomatic participants): In cohorts with clinically suspected deep vein thrombosis, the prevalence of deep vein thrombosis 20 to 94% and proximal deep vein thrombosis ranged from 48 to 100%, where reported. The reference standard was applied independently of the results of ultrasound in all studies; ultrasound was interpreted blind to the venography in 62 cohorts; venography was interpreted blind to ultrasound results in 56 cohorts. For all deep vein thrombosis: compression ultrasound (22 cohorts) sensitivity was 90% (95% CI: 88 to 92) and specificity was 98% (95% CI: 97 to 98); colour Doppler ultrasound (five cohorts) sensitivity was 82% (95% CI: 77 to 86) and specificity was 93% (95% CI: 90 to 95); continuous wave Doppler ultrasound (16 cohorts) sensitivity was 81% (95% CI: 78 to 84) and specificity was 84% (95% CI: 81 to 86); triplex ultrasound (25 cohorts) sensitivity was 91% (95% CI: 89 to 93) and specificity was 94% (95% CI: 93 to 96); duplex ultrasound (28 cohorts) sensitivity was 92% (95% CI: 91 to 94) and specificity was 94% (95% CI: 93 to 95). All analyses reporting separate data for proximal and distal deep vein thrombosis showed higher sensitivity for proximal deep vein thrombosis. Full results of meta-regression and data for asymptomatic and mixed cohorts were reported.

**Computed tomography** (CT, nine studies): Most studies compared CT with ultrasound in patients with suspected pulmonary embolism; only three studies included patients with clinically suspected deep vein thrombosis. The reference standard was applied independently of the results of CT in six studies; CT was interpreted blind to the reference standard in five studies; the reference standard was interpreted blind to CT results in four studies. The pooled estimate of sensitivity was 95% (95% CI: 91 to 97) and the pooled estimate of specificity was 97% (95% CI: 95 to 98). Significant between study heterogeneity was noted for both parameters.

**Magnetic resonance imaging** (MRI, 14 studies): Most studies compared two-dimensional time-of-flight MRI with contrast venography in patients with suspected deep vein thrombosis. The reference standard was applied independently of the results of MRI in all studies; MRI was interpreted blind to the reference standard in 11 studies; the reference standard was interpreted blind to MRI results in 11 studies. The pooled estimate of sensitivity was 92% (95% CI: 88 to 95) and the pooled estimate of specificity was 95% (95% CI: 93 to 97). Significant between study heterogeneity was noted for both parameters.

**Diagnostic algorithms** (18 studies; 11 diagnostic algorithms): All studies used follow-up to determine whether the strategy was safe. The rate of thromboembolic events in untreated patients ranged from 0 to 2.6%.

**Cost information**

Two algorithms, that would be feasible in most hospitals, offered high benefit. Both algorithms used a combination of Wells score, D-dimer and above-knee ultrasound. For willingness to pay thresholds of £10,000 or £20,000 per quality-adjusted life-year, the optimal strategy was to discharge patients with a low or intermediate Wells score and negative D-dimer, and to use ultrasound to further investigate those with a high Wells score or positive D-dimer, with repeat
scanning for those with a negative initial scan. For a threshold of £30,000, the optimal strategy was similar, but involved repeat ultrasound for all patients with a negative initial scan.

**Authors' conclusions**

Diagnostic algorithms based on Wells score, D-dimer and ultrasound in patients with a high Wells score or positive D-dimer (with repeat scanning if negative) were likely to be the most feasible and cost-effective option for UK hospitals. Further investigation of patients with a low Wells score and negative D-dimer was unlikely to be cost-effective.

**CRD commentary**

This was a very large review covering a number of different non-invasive testing options for the diagnosis of deep vein thrombosis. The inclusion criteria were clearly stated for each review question and the search strategy was comprehensive. Although included studies were restricted by language, a large number of data sets were included, so it seems unlikely that the conclusions of the review would have been significantly changed by the inclusion of studies in other languages. The authors attempted to identify unpublished studies, reducing the potential for publication bias. Detail of the review process was limited, which made it difficult to assess the potential for error or bias. Limited, but appropriate, criteria were used to assess the methodological quality of the included studies. The results of this assessment were included in the text and analyses, where appropriate. Although improved methods are now available, the meta-analyses presented the best option at the time the review was conducted; the value of pooled estimates from heterogeneous data sets may be limited, but the authors also presented ranges and explored potential sources of heterogeneity. Overall, the authors’ conclusions represent a reasonable interpretation of the data presented and are likely to be reliable.

**Implications of the review for practice and research**

**Practice**: The authors stated that diagnostic algorithms based on Wells score, D-dimer and ultrasound in patients with a high Wells score or positive D-dimer (with repeat scanning if negative) are likely to be the most feasible and cost-effective option for UK hospitals.

**Research**: The authors recommended the following: the evaluation of clinical and cost-effectiveness of using optimal algorithms in routine practice; the development of subgroup specific algorithms (e.g. pregnancy, previous deep vein thrombosis); the evaluation of the role of plethysmography and its interaction with other tests. They also stated that future studies should be of higher quality in terms of reporting.

**Funding**

NHS Research and Development Health Technology Assessment Programme, project number 02/03/01.

**Bibliographic details**


**PubMedID**

16707072

**Original Paper URL**

http://www.hta.ac.uk/fullmono/mon1015.pdf

**Other publications of related interest**


**Indexing Status**

Subject indexing assigned by NLM

**MeSH**
Record Status
This is a critical abstract of a systematic review that meets the criteria for inclusion on DARE. Each critical abstract contains a brief summary of the review methods, results and conclusions followed by a detailed critical assessment on the reliability of the review and the conclusions drawn.