Variation in the diagnostic performance of D-dimer for suspected deep vein thrombosis

S. GOODACRE¹, F.C. SAMPSON¹, A.J. SUTTON², S. MASON¹ and F. MORRIS³

From the ¹Medical Care Research Unit, University of Sheffield, Sheffield, ²Centre for Biostatistics and Genetic Epidemiology, Department of Health Sciences, University of Leicester, Leicester, and ³Department of Emergency Care, Sheffield Teaching Hospitals, Sheffield, UK

Received 12 January 2005 and in revised form 24 March 2005

Summary

Background: Numerous studies have evaluated the accuracy of D-dimer in diagnosing suspected deep vein thrombosis (DVT), but results are conflicting.

Aim: To overview estimates of the diagnostic accuracy of D-dimer and identify causes of variation.

Design: Systematic review, meta-analysis and meta-regression.

Methods: We searched Medline, EMBASE, CINAHL, Web of Science, Cochrane Database of Systematic Reviews, Cochrane Controlled Trials Register, Database of Reviews of Effectiveness, the ACP Journal Club, citation lists, and contacted manufacturers. We selected studies that compared D-dimer to a reference standard in patients with suspected DVT. Data were analysed by random effects meta-analysis and meta-regression.

Results: We included 97 studies reporting 198 assays in 99 different patient groups. Overall estimated sensitivity and specificity of D-dimer were 90.5% and 54.7%, but both estimates were subject to significant heterogeneity (p < 0.001). Meta-regression identified that some heterogeneity was explained by study setting, exclusion criteria, whether recruitment was consecutive or the study prospective, whether D-dimer and the reference standard were measured blind, and whether the D-dimer threshold was determined a priori. Sensitivity and specificity also varied between ELISA (94% and 45% respectively), latex (89% and 55%) and whole blood agglutination assays (87% and 68%). Sensitivity was higher for proximal than distal DVT. Specificity was dependent upon whether clinical probability of DVT was high (specificity 51%), intermediate (67%) or low (78%).

Discussion: D-dimer has good sensitivity, but poor specificity, for DVT. Estimates are subject to substantial heterogeneity from various sources. D-dimer specificity appears to be strongly dependent upon the pre-test clinical probability of DVT.

Introduction

Clinical suspicion of deep vein thrombosis (DVT) is a frequent cause for emergency diagnostic assessment.¹ The D-dimer assay is a simple, cheap test that may be useful in DVT diagnosis. Numerous research studies have been published evaluating the diagnostic accuracy of D-dimer for DVT, and attempts have been made to systematically review these studies and draw broad conclusions.²–⁴ The most recently published review⁴ examined the use of D-dimer for both DVT and pulmonary embolus (PE), and concluded that an enzyme-linked immunoassay (ELISA) D-dimer had a negative predictive

Address correspondence to Dr S. Goodacre, Medical Care Research Unit, Regent Court, 30 Regent Street, Sheffield S1 4DA. email: s.goodacre@sheffield.ac.uk
© The Author 2005. Published by Oxford University Press on behalf of the Association of Physicians. All rights reserved. For Permissions, please email: journals.permissions@oupjournals.org
value for DVT that was as diagnostically useful as negative duplex ultrasonography findings. However, this analysis also reported that the ELISA assay has poor specificity, a finding that may limit its usefulness in practice.

Several issues were not explored in this review, or others, and merit detailed analysis. Although there is substantial evidence of variation in the results of different studies, only a limited search for causes of heterogeneity was undertaken. Identifying sources of heterogeneity by using techniques such as meta-regression may produce valuable information with which to interpret results, extrapolate findings to different settings, and guide future research. Two specific issues warrant particular attention. Firstly, D-dimer may have greater sensitivity for proximal (rather than distal) DVT. Proximal DVT carries a higher risk of propagation than distal DVT, so missed distal DVT is not as potentially serious. Ultrasonography has become a widely accepted diagnostic test for DVT despite limited sensitivity for distal DVT, because its sensitivity for proximal DVT is high. Secondly, the diagnostic parameters of D-dimer may vary with the pre-test clinical probability of DVT. This has important implications if D-dimer is used in a Bayesian framework (i.e. by applying the likelihood ratio for a D-dimer result to an estimate of pre-test probability, to derive an estimate of post-test probability), because Bayes’ Theorem assumes that diagnostic test performance is independent of pre-test probability.

We aimed to: (i) undertake a systematic review of the literature and meta-analysis to estimate the sensitivity and specificity of D-dimer in the diagnosis of patients presenting with clinically suspected DVT; (ii) undertake meta-regression to identify causes of heterogeneity between studies; and (iii) undertake meta-analysis limited to studies that report proximal and distal DVT separately, or report results stratified by pre-test clinical probability.

Methods

Search strategy

We sought to identify all diagnostic cohort studies of patients with clinically suspected DVT who underwent testing with D-dimer followed by a reference standard investigation. We searched Medline, EMBASE, CINAHL, Web of Science, Cochrane Database of Systematic Reviews, Cochrane Controlled Trials Register, Database of Reviews of Effectiveness, and ACP Journal Club using the search strategy outlined in the Appendix. The bibliographies of all articles selected for the review were scanned for potentially relevant articles that were not identified by the original search. Manufacturers of D-dimer assays were contacted and asked to supply any unpublished studies.

Selection of articles

Two reviewers (FS and SG) screened the titles and abstracts of all articles to independently identify potentially relevant articles. Full copies of all selected articles were retrieved. These were then reviewed by the same two reviewers, who independently selected articles that measured the diagnostic performance of D-dimer, compared to a reference standard of venography, ultrasound or plethysmography, in patients with suspected DVT. At both stages of selection, a kappa score was calculated and disagreements resolved by discussion. Studies published in English, French, Spanish, Italian or German were included. Studies published in other languages were excluded. Abstracts and letters were included if they reported data in sufficient detail to allow inclusion in the analysis. If not, the authors were contacted and asked to provide details of the data or any full publications.

We specifically excluded: studies that measured the risk of DVT developing after D-dimer testing, rather than the probability of DVT being present at the time of testing; case-control studies, in which D-dimer measurements in a group of patients with DVT were compared to a control group of patients without DVT; management studies that ruled out DVT on the basis of a negative D-dimer alone; studies that used a reference standard other than venography, ultrasound or impedance plethysmography; studies with <10 patients; and studies of patients with suspected pulmonary embolus.

Data extraction

Two independent reviewers (SG and FM or SM) extracted the following data from the selected studies onto a standardized pro forma: the setting for patient recruitment, any exclusion criteria that might influence D-dimer performance (malignancy, post-operative patients, pregnant, anticoagulated, trauma, sepsis, or prolonged history), population demographics, whether recruitment was consecutive and/or data collection prospective, which D-dimer assay(s) were used, the threshold value(s) used, the reference standard used, and the number of true positives (proximal and distal), true negatives, false positives and false negatives (proximal and distal), either as reported or calculated from the reported data. Discrepancies were checked and resolved by an independent reviewer (FS). If it was
not possible to extract the necessary data from the published report, we contacted the authors for clarification. We reviewed the data reported by each study, and removed studies that contained duplicated data. Although we collected data from groups of asymptomatic patients receiving D-dimer testing, we have only reported data here from patients with clinically suspected DVT.

**Quality assessment**

The same two reviewers independently assessed the quality of each study according to the following criteria, with disagreements being resolved by an independent reviewer (FS): (i) was application of the reference standard independent of the D-dimer result? (ii) was D-dimer measured by operators who were blind to the reference standard result? (iii) was the reference standard performed by operators who were blind to the D-dimer result? (iv) was a pre-determined threshold value for D-dimer used (a priori threshold value), or was the threshold derived from the data (post hoc threshold)?

**Statistical analysis**

Analyses used MetaDiSc statistical software. Each analysis used a random effects model to estimate overall sensitivity and specificity, and a $\chi^2$ test for heterogeneity between studies. Where zero counts occurred for study data, a continuity correction of 0.5 was added to the raw data for that study in order to make the calculation of sensitivity and specificity defined. Initially all studies were analysed together and random effects meta-regression undertaken to identify potential causes of heterogeneity for sensitivity and specificity separately. Any covariate that showed an association with sensitivity or specificity ($p < 0.1$) was selected, and subgroups of studies identified by such covariates were meta-analysed separately. We also undertook the following separate meta-analyses: studies of ELISA, latex and whole-blood agglutination assays; proximal and distal DVT (where studies reported them separately); studies of each individual D-dimer assay; studies reporting results by Wells clinical risk stratification; and groups of patients with malignancy.

Funnel plots for sensitivity and specificity were constructed to assess the possibility of publication bias. Such plots are constructed by plotting the log odds of sensitivity and specificity against their corresponding standard errors. This transformed scale should produce a funnel that is approximately symmetrical if no publication bias is present. Fixed-effect pooled estimates, together with guidelines in which 95% of studies should lie if no heterogeneity is present, are overlaid on the plots to aid interpretation of heterogeneity and bias.

**Results**

The flow of articles considered for the review is outlined in Figure 1. We scanned the titles or abstracts of 1140 articles and selected 247 potentially relevant articles for retrieval (kappa 0.87). Review of the full articles identified 131 that met the inclusion criteria (kappa 0.94). Six of these contained duplicated data and were excluded. We were unable to extract or analyse appropriate data from a further 15 articles. Thirteen articles reported groups of asymptomatic patients and are not reported here. Thus 97 articles were included in the meta-analysis: 90 were published in English, three in French, two in Spanish, and one each in Italian and German. Two articles reported two groups of patients, so a total of 99 groups were analysed. The number of assays tested on each groups varied from one to 13. Overall, 198 analyses of D-dimer assays were reported.

The patients were recruited from the following settings: emergency department 16, out-patient referrals 31, in-patients 9, mixed setting 29, not reported 14. Exclusions were reported for 49 groups: 10 excluded post-operative patients, 19 excluded pregnant patients, 33 excluded anticoagulated patients, 23 excluded those with previous thromboembolism, three excluded those with recent trauma, four excluded those with sepsis, and 18 excluded patients with a prolonged history.

The mean or median age of the sample varied from 51 to 69 years (median 59), with the exception of one study that recruited exclusively over 70s. The proportion of males ranged from 17% to 62% (median 42%). Overall prevalence of DVT varied from 2% to 78% (median 36%). Prevalence of proximal and distal DVT was reported for 51 groups. The proportion of proximal DVT (of all DVT detected) varied from 27% to 100% (median 77%).

The reference standard consisted of venography (34 studies), ultrasound alone (28), ultrasound with clinical follow-up (10), serial ultrasound (6), either ultrasound or venography (13), or a combination of ultrasound and plethysmography (8). The reference standard was definitely independent of the results of D-dimer testing in 86, D-dimer measurement was definitely blind to the reference standard in 43, the reference standard measurement was definitely blind to D-dimer in 50, and the threshold value for D-dimer was defined before analysis in 82.
Figure 2 shows the results for all studies plotted on an ROC plane. The true positive rate (sensitivity) is plotted against the false positive rate (1-specificity) for each study included in the meta-analysis. A study of a perfect test (100% sensitivity and specificity) would be plotted in the top left hand corner. The ROC curve thus demonstrates the overall diagnostic performance of D-dimer assays and the variation between the results of individual studies. The pooled estimates (95%CI) for sensitivity and specificity were 90.5% (90.0–91.1) and 54.7% (54.0–55.4) respectively, but the results show substantial heterogeneity, particularly for specificity. Sensitivity ranged from 48% to 100% ($\chi^2$ test for
heterogeneity \(p < 0.001\). Specificity ranged from 5\% to 100\% (\(\chi^2\) test for heterogeneity \(p < 0.001\)). Specificity ranged from 5\% to 100\% (\(\chi^2\) test for heterogeneity \(p < 0.001\)).

Table 1 shows the results of meta-regression. A number of covariates were associated with systematic variation in sensitivity and specificity. However, the significant associations for age and D-dimer threshold were due to one group in each analysis having an extreme value for the covariate that exerted substantial leverage upon the regression line: Le Blanche et al. selected patients aged over 70 to recruit a group with a mean age of 86 years,\(^{78}\) while Crippa et al. used a threshold of 2300 ng/ml for positivity.\(^{128}\) Neither association was significant when the group with an extreme value was removed.

Random effects meta-analysis was repeated, stratified by each significant covariate, to estimate sensitivity and/or specificity for studies with, or without, the covariate. The results are shown in Table 2. More selective groups (i.e. recruited from out-patient or ED only, or reporting exclusion criteria) tended to have higher sensitivity and specificity. Higher quality studies (prospective studies, those recruiting consecutive patients, venographic reference standard, D-dimer and reference standard measured blind) tended to have higher specificity. Studies that determined the D-dimer threshold after data analysis had higher sensitivity. However, there was still substantial heterogeneity among results, even when stratified by these criteria.

ELISA assays were reported by 91 analyses in 58 cohorts (35 analyses reported proximal and distal DVT separately). Pooled sensitivity was 94\% (95\%CI 93–95) and specificity 45\% (95\%CI 44–46). Latex assays were reported by 74 analyses in 52 cohorts (22 analyses reported proximal and distal DVT separately). Pooled sensitivity was 89\% (95\%CI 88–90) and specificity 55\% (95\%CI 54–56).

Whole-blood agglutination assays were reported by 29 analyses in 29 cohorts (10 analyses reported proximal and distal DVT separately). Pooled sensitivity was 87\% (95\%CI 85–88) and specificity 68\% (95\%CI 67–69). Analysis of studies that reported proximal and distal DVT separately showed that all assays had higher sensitivity for proximal than distal DVT: 98\% (95\%CI 97–99) vs. 86\% (95\%CI 84–88) for ELISA, 94\% (95\%CI 92–95) vs. 79\% (95\%CI 75–83) for latex, and 84\% (95\%CI 80–88) vs. 64\% (95\%CI 55–73) for whole-blood agglutination.

Table 3 shows the pooled estimates of sensitivity and specificity for individual assays that were investigated by three or more eligible studies, and are thus suitable for meta-analysis. Although

<table>
<thead>
<tr>
<th>Variable</th>
<th>Sensitivity (\rho)</th>
<th>Specificity (\rho)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patients recruited</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mixed</td>
<td>0.030</td>
<td>0.010</td>
</tr>
<tr>
<td>ED only</td>
<td>0.32</td>
<td>0.065</td>
</tr>
<tr>
<td>Out-patient only</td>
<td>0.010</td>
<td>0.022</td>
</tr>
<tr>
<td>In-patient only</td>
<td>0.94</td>
<td>0.24</td>
</tr>
<tr>
<td>Patients excluded</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Post-operative</td>
<td>0.85</td>
<td>0.92</td>
</tr>
<tr>
<td>Pregnant</td>
<td>0.049</td>
<td>0.079</td>
</tr>
<tr>
<td>Anticoagulated</td>
<td>0.011</td>
<td>0.007</td>
</tr>
<tr>
<td>Past history of DVT</td>
<td>0.38</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Trauma</td>
<td>0.47</td>
<td>0.21</td>
</tr>
<tr>
<td>Sepsis</td>
<td>0.76</td>
<td>0.89</td>
</tr>
<tr>
<td>Long history</td>
<td>0.061</td>
<td>0.82</td>
</tr>
<tr>
<td>Mean age</td>
<td>0.046 (0.11)*</td>
<td>0.01 (0.21)*</td>
</tr>
<tr>
<td>% Males</td>
<td>0.30</td>
<td>0.13</td>
</tr>
<tr>
<td>Prevalence of DVT</td>
<td>0.27</td>
<td>0.38</td>
</tr>
<tr>
<td>Consecutive patients recruited</td>
<td>0.47</td>
<td>0.016</td>
</tr>
<tr>
<td>Prospective study</td>
<td>0.026</td>
<td>0.047</td>
</tr>
<tr>
<td>D-dimer threshold used</td>
<td>0.090 (0.97)**</td>
<td>0.68 (0.45)**</td>
</tr>
<tr>
<td>Reference standard used (venography vs. other)</td>
<td>0.37</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Reference standard independent of D-dimer result</td>
<td>0.11</td>
<td>0.21</td>
</tr>
<tr>
<td>D-dimer measured blind to reference standard</td>
<td>0.40</td>
<td>0.054</td>
</tr>
<tr>
<td>Reference standard measured blind to D-dimer</td>
<td>0.98</td>
<td>0.035</td>
</tr>
<tr>
<td>D-dimer threshold defined before the study</td>
<td>0.002</td>
<td>0.34</td>
</tr>
</tbody>
</table>

\(\rho\) values when included in random effect weighted meta-regression. *The revised value excluded one study with an extreme value for mean age. **The revised value excluded one study that used an extreme threshold value.
differences are apparent between the individual assays, the confidence intervals for estimates are relatively wide, suggesting that differences may not be significant. Furthermore, even these estimates for individual assays are subject to significant heterogeneity, suggesting that heterogeneity in estimates of D-dimer sensitivity and specificity cannot be explained simply differences between assays.

Five studies,\textsuperscript{51,64,80,85,121} using six assays (five whole blood agglutination and one latex), reported results stratified by Wells clinical risk score. Meta-analysis of these studies showed that D-dimer had a sensitivity (95\% CI, $p$ for heterogeneity) of 96\% (92–99, $p = 0.047$) and specificity of 51\% (38–53, $p = 0.001$) in patients with a high Wells score; a sensitivity of 88\% (79–94, $p = 0.18$) and specificity of 67\% (62–72, $p = 0.003$) in patients with an intermediate score; and a sensitivity of 91\% (72–99, $p = 0.69$) and specificity of 78\% (74–81, $p < 0.001$) in patients with a low Wells score. Thus the specificity of D-dimer appears to be dependent upon the clinical probability of DVT. Five studies reported patients with malignancy separately.\textsuperscript{49,50,52,80,127} Meta-analysis of these studies showed that D-dimer had a sensitivity of 95\% (90–97, $p = 0.026$) and a specificity of 46\% (39–52, $p = 0.0026$) in patients with malignancy.

Funnel plots of the studies are shown by sensitivity in Figure 3 and specificity in Figure 4. The funnel plot for specificity appears to be symmetrical, whereas the funnel plot for sensitivity is markedly asymmetrical. Although there are many small studies reporting high sensitivity, there appear to be very few small studies reporting low sensitivity. One possible explanation for this is publication bias: small studies have only been published if they reported high sensitivity.

**Discussion**

Our meta-analysis confirms that D-dimer has good sensitivity for DVT but poor specificity. ELISA assays are more sensitive, but less specific, than latex and whole blood agglutination assays. However, sensitivity and specificity vary between individual assays, so that some latex assays may have better sensitivity than some ELISA assays. Since D-dimer is principally used to rule out DVT, clinicians will usually be more concerned that the test should have optimal sensitivity, although the need for rapid results may favour latex, whole blood agglutination and rapid ELISA assays over the classical ELISA assays. Table 3 may provide some guidance in comparing assay types, but care should be taken to consider confidence intervals, heterogeneity among individual study results, and the effect of varying the threshold for positivity upon sensitivity and specificity.

There was substantial heterogeneity among results of almost all analyses. We found that patient selection and (to a lesser extent) quality criteria may explain some heterogeneity. Studies that recruited from a single setting or reported using exclusion criteria had higher sensitivity and specificity. This is probably because factors that may lead to exclusion from the studies may be associated with false positive D-dimer results (e.g. pregnancy) or false negative results (e.g. prolonged history). Studies that recruited consecutively, prospective studies, and those with blinded measurement of D-dimer and the reference standard had slightly higher specificity. This was an unexpected finding, as failure to adhere to quality criteria is usually associated with higher estimates of sensitivity and specificity.\textsuperscript{140} It is possible that this finding is confounded by another

<table>
<thead>
<tr>
<th>Variable</th>
<th>Sensitivity</th>
<th>Heterogeneity $p$</th>
<th>Specificity</th>
<th>Heterogeneity $p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mixed patient selection</td>
<td>87% (86–88)</td>
<td>$&lt;0.001$</td>
<td>50% (49–51)</td>
<td>$&lt;0.001$</td>
</tr>
<tr>
<td>ED patients only</td>
<td>89% (87–91)</td>
<td>$&lt;0.001$</td>
<td>62% (60–64)</td>
<td>$&lt;0.001$</td>
</tr>
<tr>
<td>Out-patients only</td>
<td>94% (93–95)</td>
<td>$&lt;0.001$</td>
<td>59% (58–60)</td>
<td>$&lt;0.001$</td>
</tr>
<tr>
<td>Studies excluding pregnant patients</td>
<td>95% (93–97)</td>
<td>$&lt;0.001$</td>
<td>57% (56–58)</td>
<td>$&lt;0.001$</td>
</tr>
<tr>
<td>Studies excluding anticoagulated patients</td>
<td>93% (92–94)</td>
<td>$&lt;0.001$</td>
<td>61% (60–62)</td>
<td>$&lt;0.001$</td>
</tr>
<tr>
<td>Studies excluding patient with past history of thromboembolism</td>
<td>91% (90–93)</td>
<td>$&lt;0.001$</td>
<td>65% (63–66)</td>
<td>$&lt;0.001$</td>
</tr>
<tr>
<td>Studies excluding patients with a prolonged history</td>
<td>93% (92–94)</td>
<td>$&lt;0.001$</td>
<td>54% (52–56)</td>
<td>$&lt;0.001$</td>
</tr>
<tr>
<td>Consecutive patients recruited</td>
<td>91% (90–92)</td>
<td>$&lt;0.001$</td>
<td>57% (56–58)</td>
<td>$&lt;0.001$</td>
</tr>
<tr>
<td>Prospective study</td>
<td>90% (89–91)</td>
<td>$&lt;0.001$</td>
<td>58% (57–59)</td>
<td>$&lt;0.001$</td>
</tr>
<tr>
<td>Venographic reference standard</td>
<td>91% (90–92)</td>
<td>$&lt;0.001$</td>
<td>62% (61–64)</td>
<td>$&lt;0.001$</td>
</tr>
<tr>
<td>D-dimer measured blind to reference standard</td>
<td>90% (89–91)</td>
<td>$&lt;0.001$</td>
<td>56% (55–57)</td>
<td>$&lt;0.001$</td>
</tr>
<tr>
<td>Reference standard measured blind</td>
<td>90% (89–91)</td>
<td>$&lt;0.001$</td>
<td>57% (56–58)</td>
<td>$&lt;0.001$</td>
</tr>
<tr>
<td>D-dimer threshold derived from data</td>
<td>95% (94–96)</td>
<td>0.245</td>
<td>41% (39–44)</td>
<td>$&lt;0.001$</td>
</tr>
</tbody>
</table>
Table 3  Sensitivity and specificity (95% CI and p for heterogeneity) for each D-dimer assay

<table>
<thead>
<tr>
<th>Test</th>
<th>Assay type</th>
<th>Manufacturer</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asserachrom (n=19)</td>
<td>Classical ELISA</td>
<td>Diagnostica Stago</td>
<td>95% (93–96)</td>
<td>36% (34–38) p &lt; 0.001</td>
</tr>
<tr>
<td>Enzygnost (n=5)</td>
<td>Classical ELISA</td>
<td>Dade Behring</td>
<td>96% (92–98)</td>
<td>62% (57–67) p &lt; 0.001</td>
</tr>
<tr>
<td>Fibrinostika (n=3)</td>
<td>Classical ELISA</td>
<td>Organon Teknika</td>
<td>95% (90–98)</td>
<td>37% (30–45) p = 0.761</td>
</tr>
<tr>
<td>Dimertest EIA Gold (n=6)</td>
<td>Classical ELISA</td>
<td>Agen</td>
<td>90% (85–93)</td>
<td>65% (62–71) p &lt; 0.001</td>
</tr>
<tr>
<td>VIDAS (n=17)</td>
<td>Classical ELISA</td>
<td>Biomerieux</td>
<td>96% (95–98)</td>
<td>39% (36–42) p &lt; 0.001</td>
</tr>
<tr>
<td>Nyco card (n=14)</td>
<td>Rapid ELISA</td>
<td>Nycomed</td>
<td>89% (87–91)</td>
<td>50% (47–54) p &lt; 0.001</td>
</tr>
<tr>
<td>Instant IA (n=8)</td>
<td>Rapid ELISA</td>
<td>Diagnostica Stago</td>
<td>92% (90–94)</td>
<td>54% (49–58) p &lt; 0.001</td>
</tr>
<tr>
<td>Minutex (n=8)</td>
<td>Latex</td>
<td>Biopool</td>
<td>92% (90–94)</td>
<td>45% (42–49) p &lt; 0.001</td>
</tr>
<tr>
<td>D-Di latex (n=4)</td>
<td>Latex</td>
<td>Diagnostica Stago</td>
<td>80% (72–86)</td>
<td>58% (51–64) p &lt; 0.001</td>
</tr>
<tr>
<td>LPIA (n=3)</td>
<td>Latex</td>
<td>Mitsubishi Kasei</td>
<td>98% (94–99)</td>
<td>58% (51–65) p = 0.244</td>
</tr>
<tr>
<td>BC (n=4)</td>
<td>Latex</td>
<td>Dade Behring</td>
<td>90% (84–94)</td>
<td>53% (47–60) p &lt; 0.001</td>
</tr>
<tr>
<td>IL test (n=6)</td>
<td>Latex</td>
<td>Instrumentation Labs</td>
<td>94% (91–97)</td>
<td>52% (47–56) p &lt; 0.001</td>
</tr>
<tr>
<td>Miniquant (n=4)</td>
<td>Latex</td>
<td>Biopool</td>
<td>96% (91–99)</td>
<td>36% (31–40) p &lt; 0.001</td>
</tr>
<tr>
<td>STA-latest (n=8)</td>
<td>Latex</td>
<td>Diagnostica Stago</td>
<td>94% (92–96)</td>
<td>46% (43–50) p = 0.002</td>
</tr>
<tr>
<td>Tinaquant (n=7)</td>
<td>Latex</td>
<td>Roche</td>
<td>96% (93–97)</td>
<td>52% (47–56) p = 0.14</td>
</tr>
<tr>
<td>Turbiquant (n=3)</td>
<td>Latex</td>
<td>Dade Behring</td>
<td>88% (81–93)</td>
<td>46% (41–52) p &lt; 0.002</td>
</tr>
<tr>
<td>SimpliRED (n=29)</td>
<td>Whole blood agglutination</td>
<td>Agen</td>
<td>87% (85–88)</td>
<td>68% (67–69) p &lt; 0.001</td>
</tr>
</tbody>
</table>
unreported factor. For example, high-quality studies may also have more rigorous procedures for measuring D-dimer or selecting patients for testing that result in improved diagnostic performance. Blinded measurement of D-dimer in particular would be unlikely to influence results, unless a qualitative assay was used. However, reporting that D-dimer was measured blind may reflect a more rigorous approach to measurement that is in turn associated with fewer false positive results. Finally, we found that using a threshold value that was derived from the study data led to higher estimates of sensitivity. This suggests that when a threshold is derived \textit{post hoc}, it tends to be chosen at a level that maximizes sensitivity.

Although some covariates were highly statistically significant predictors, analysis stratified by these covariates did not dramatically affect estimates of sensitivity or specificity, and estimates remained subject to substantial heterogeneity. Much of the observed heterogeneity therefore remains unexplained and presumably due to unreported factors. Our analysis provides a clue to the cause of this unexplained heterogeneity. We found that specificity was strongly dependent upon clinical probability of DVT. This suggests that differences in clinical probability can have a powerful effect upon specificity. It also has important practical consequences. Firstly, using D-dimer in patients with a low clinical probability of DVT is likely to generate fewer false positives than one might expect from estimates of specificity derived from an unselected population. Secondly, if D-dimer specificity, and thus likelihood ratios, are dependent upon pre-test probability then a fundamental assumption of Bayes theorem (that test performance is independent of pre-test probability) is not fulfilled. This means that if we use D-dimer in a Bayesian framework (i.e. applying the likelihood ratios of D-dimer results to the pre-test odds of DVT to determine the post-test odds of DVT in an individual patient) then our estimates of post-test odds, or probability, may
be misleading. Finally, it may be appropriate to use different thresholds for positivity according to pre-test probability. For example, since D-dimer is principally used to rule out DVT in low risk patients, it may be appropriate to use a lower threshold for positivity to minimize the risk of inadvertent discharge with DVT. Our data suggest that, although using a lower threshold may reduce specificity, it may still be acceptable in patients with a low clinical probability.

Analysis of studies reporting proximal and distal DVT separately showed that all three assays had higher sensitivity for proximal DVT. The pooled sensitivities of ELISA and latex assays were comparable to the estimated sensitivity of ultrasound for proximal DVT. This suggests that if we are not concerned about detecting and treating distal DVT then D-dimer may have sufficient sensitivity to rule out DVT. However, the clinical importance of distal DVT is controversial, and many clinicians would want to detect and treat this condition.141,142

**Limitations**

This is the most comprehensive review yet undertaken of the diagnostic performance of D-dimer for suspected DVT, but it has some limitations that must be appreciated. We undertook only limited efforts to identify unpublished data. Although publication bias is a potentially important factor in meta-analysis of diagnostic tests, there is little empirical data to determine how this might influence findings. Most published meta-analyses of diagnostic tests have only searched electronic databases and reference lists, and have not addressed the possibility of publication bias.143 This is perhaps not surprising. Registries of diagnostic test studies have not yet been established and, because diagnostic test studies are relatively easy to undertake, unpublished data could reside in a wide variety of potential locations. Tracing such data would be arduous and potentially unproductive.

The funnel plots showed that small studies reporting high sensitivity are more numerous than small studies reporting low sensitivity. It is highly plausible that this could be due to publication bias. D-dimer is likely to be more clinically useful as a rule-out test for patients at low risk of DVT. Authors may therefore be more eager to submit their work, and journals to publish it, if it reports high sensitivity. Specificity would not be expected to influence publication to a similar degree: since D-dimer is not likely to be used to definitively diagnose DVT, poor specificity would not be a barrier to publication. If this meta-analysis has been subject to publication bias, then our estimate of sensitivity is likely to represent an over-estimate.

Reporting of the studies was frequently poor and many of the details we sought for meta-regression were missing. We wanted to explore a number of other covariates, such as the proportion of the study population with co-morbidities, but were unable to because these variables were rarely reported. Even where potential sources of heterogeneity were identified it is not clear what the ‘true’ values for diagnostic parameters are. Selective groups may produce higher estimates, but these may not be accurate if they do not reflect the population in which the test will be used. Readers need to consider how the groups reported here compare to their patient population.

**Implications for future research**

Future studies of D-dimer should collect and report accurate data on patient selection, application of exclusion criteria, and the presence of co-morbidities. We recommend that the proportion of the study population with the following characteristics should be reported: malignancy, anticoagulant treatment, past history of DVT, pregnancy, sepsis, recent trauma or surgery. Most importantly, patients should be classified according to their clinical probability of DVT, and results reported separately for different risk groups. The Wells score is a widely used, validated score that is ideal for this purpose. Finally, in view of the evidence of publication bias, there is perhaps a need for registers of diagnostic studies, similar to those recently developed for clinical trials.

**Conclusions**

D-dimer has good sensitivity, but poor specificity, for diagnosing suspected DVT. All analyses showed substantial heterogeneity, which was partially explained by differences in patient selection and methodology in the individual studies. D-dimer sensitivity is higher for proximal than distal DVT, while specificity appears to be dependent upon the pre-test clinical probability of DVT, being higher in patients with a low clinical probability of DVT.

**Acknowledgements**

The United Kingdom Health Technology Assessment R&D Programme funded this project (reference number 02/03/01). The views and opinions expressed therein are those of the authors, and do not necessarily reflect those of the UK Department of Health. We thank Kathryn Paulucy for clerical...
assistance in preparing this article, Angie Ryan for assistance with the literature searches and Eddie Hampton for constructive comments during development of this project.

References
13. D-dimer testing and impedance plethysmography were effective for the exclusion of deep vein thrombosis. ACP J Club 1997; 127:74.


117. Mossaz A, Grandrille S, Vitoux JF, Abdouchel-Baudot N, Aiach M, Fiessinger JN. The value of D-dimer assays in...


Appendix: Search strategy and years searched on each database

Strategy

Key terms and mesh headings may vary between databases.

1. exp ‘Sensitivity and Specificity’/
2. exp Diagnostic Errors/
3. Reference Values/
4. Reproducibility of Results/
5. likelihood functions/
6. 1 or 2 or 3 or 4 or 5
7. specificity.mp. [mp=title, abstract, cas registry/ ec number word, mesh subject heading]
8 sensitivity.mp.
9 false negative$.mp.
10 false positive$.mp.
11 true negative$.mp.
12 true positive$.mp.
13 predictive value$.mp.
14 reproducibility.mp.
15 ROC curve.mp.
16 Diagnos$.mp.
17 reference value$.mp.
18 likelihood function$.mp.
19 likelihood ratio$.mp.
20 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or
21 15 or 16 or 17 or 18 or 19
22 exp Venous Thrombosis/
23 deep vein thrombosis.tw.
24 phlebothrombosis.tw.
25 venous thrombosis.tw.
26 venous thromboembolism.tw.
27 DVT.tw.
28 21 or 22 or 23 or 24 or 25 or 26 or 27
29 20 and 28
30 Fibrin Fibrinogen Degradation Products/ or
   Enzyme-Linked Immunosorbent Assay/
31 Reagent Kits, Diagnostic/
32 d-dimer.tw.
33 simplired.tw.
34 30 or 31 or 32 or 33
35 29 and 34

Years searched
Medline 1996–2004
CINAHL 1982–2004
Embase 1980–2004
Web of science 1985–2004
CCTR 1898–2004
CDSR 1990–2004
DARE 1995–2004