The D-Dimer Test for Deep Venous Thrombosis: Gold Standards and Bias in Negative Predictive Value

John T. Philbrick and Steven Heim

Background: Because venous ultrasound (US) fails to fully image the calf veins, there is the potential for US gold standard studies to classify patients with calf deep venous thrombosis (DVT) in the nondiseased category, causing bias in test index calculations. A false increase in negative predictive value (NPV) is especially likely because calf DVT false-negative tests will be counted in the numerator along with the true-negative tests in NPV calculations. We verified the presence and magnitude of this bias for the d-dimer test.

Methods: We abstracted data on overall (calf and thigh) and thigh-only test sensitivity, specificity, and NPV from the six English language studies published between March 1995 and October 2001 that compared d-dimer to a gold standard (GS) capable of imaging both thigh and calf veins and that also stratified results by thigh and calf location. Thigh specificity and NPV were calculated classifying calf DVT patients as free of disease.

Results: The six studies included 81–214 participants and provided 26 comparisons of 16 different d-dimer assays to the GS. Thigh sensitivity was higher than overall sensitivity in 22 of 26 comparisons (range, −0.3 to 8.6); thigh specificity was lower than overall specificity in all comparisons (range, −0.7 to −7.8); and thigh NPV was higher than overall NPV in 22 of 26 comparisons and unchanged in 4 comparisons (range, 0.0–9.2). NPV was >95% in 20 of the thigh results but >95% in only 8 of the overall results.

Conclusions: Different GS can produce clinically significant differences in test indices. Care must be taken in interpreting DVT studies that evaluate d-dimer as a rule-out test and that use US as a GS, because missed calf DVT can falsely increase the NPV.

The measurement of d-dimer, a degradation product of circulating cross-linked fibrin, has attracted much attention as a rule-out test for venous thromboembolism. In the past 20 years, more than 40 studies have been published comparing d-dimer concentrations with a gold standard for lower extremity deep venous thrombosis (DVT) (1–29). Until 1995, d-dimer studies usually used lower extremity venography as the gold standard for DVT diagnosis (1). More recently, paralleling clinical practice, duplex ultrasonography has assumed a prominent role as the gold standard. In 28 studies published between March 1995 and October 2001, duplex ultrasound alone was used as the gold standard in 9 studies (2–5, 7, 8, 11, 14, 17), ultrasound and/or venography in 9 studies (6, 10, 12, 18, 25, 26, 28, 29), and venography alone in 10 studies (13, 15, 16, 19–22, 24, 27, 30).

Although clinicians now use duplex ultrasound as the test of choice for the initial evaluation of patients with suspected lower extremity DVT (31), ultrasound has disadvantages as a gold standard for lower extremity DVT research because of its limited ability to image calf DVT (32). In clinical practice, this limitation has been circumvented through the use of serial testing strategies designed to detect propagation of calf DVT to the thigh (31). However, in d-dimer research, the use of a single ultrasound as the gold standard creates the potential for bias. A gold standard that images both thigh and calf DVT, such as venography, will enable all patients with lower extremity DVT to be properly counted as having disease (Table 1A). However, to the extent that a gold standard such as duplex ultrasound fails to fully image the calf veins, patients with disease will be counted in the nondiseased category (Table 1B). This may lead to biased indexes of test efficacy. For example, if the true sensitivity of a d-dimer test for thigh DVT were higher than the sensitivity for calf DVT, the measured sensitivity would...
Table 1. Effect of different gold standards on content of
two-by-two table cells and indices of test efficacy.

<table>
<thead>
<tr>
<th>D-Dimer test (Tinaquant)</th>
<th>Present</th>
<th>Absent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>TP&lt;sub&gt;Thigh&lt;/sub&gt; + TP&lt;sub&gt;Calf&lt;/sub&gt;</td>
<td>FP</td>
</tr>
<tr>
<td></td>
<td>(52)</td>
<td>(6)</td>
</tr>
<tr>
<td>Negative</td>
<td>FN&lt;sub&gt;Thigh&lt;/sub&gt; + FN&lt;sub&gt;Calf&lt;/sub&gt;</td>
<td>TN</td>
</tr>
<tr>
<td></td>
<td>(0)</td>
<td>(1)</td>
</tr>
</tbody>
</table>

Sensitivity<sub>Overall</sub> = (TP<sub>Thigh</sub> + TP<sub>Calf</sub>) / All patients with DVT present = 58 / 58 + 1 = 98.3%

Specificity<sub>Overall</sub> = (TN) / All patients with DVT absent = 24 / 24 = 100.0%

NPV<sub>Overall</sub> = (TN) / All negative d-dimer tests = 24 / 24 = 100.0%

A. Overall gold standard that images thigh and calf veins.

B. Gold standard that images thigh veins only.

Our findings are summarized in Table 2. Overall prevalence ranged from 25% to 52% and thigh prevalence from 20% to 46%. Thigh sensitivity was lower than overall sensitivity in 2 of the 26 comparisons, unchanged in 2, and higher in 22 (range, −0.3 to 8.6). The two instances in which thigh sensitivity was lower than overall sensitivity occurred when there were fewer false-negative tests in patients with isolated calf DVT than in patients with thigh DVT. Thigh specificity was lower than overall specificity in all studies (range, −0.0 to −7.8). Thigh NPV was unchanged in 4 of the 26 comparisons and higher in the remaining 22 comparisons (range, 0.0 to 9.2). The NPV was >95% in 20 of the 26 thigh comparisons but in only 8 of 26 overall NPV calculations.

Discussion

The measurement of d-dimer is a simple test that has potential advantages over current testing strategies for lower extremity DVT. However, because d-dimer assays are often positive in a variety of common conditions, they are not appropriate for use as a “rule-in” test. Although studies of the test as a rule-out test have shown promise, our review shows that care must be taken in interpreting
results of studies that use ultrasound as the gold standard. To the degree that ultrasound misses isolated calf DVT, the crucial index of NPV will be higher than the true value. We believe that the magnitude of this bias, as much as 9.2% (a thigh NPV of 90.7 compared with an overall NPV 81.5), is of clinical significance.

Thigh sensitivities were generally higher than overall sensitivities, confirming our prediction that the sensitivities of the d-dimer tests for thigh DVT would be higher than the sensitivities for isolated calf DVT. Thigh specificities were generally lower than overall specificities, indicating that the sensitivities of the d-dimer tests for isolated calf DVT were higher than the overall specificities. We did not pool the results of the different studies.
because we did not want larger studies and those that evaluated multiple d-dimer tests to have undue influence. We also judged that the patient groups were too heterogeneous to allow combining of data.

In calculating the potential for bias, we compared thigh-only DVT with overall lower extremity DVT. This assumption is likely to maximize the differences in test indices that we have reported because ultrasonographers in clinical practice often extend their evaluation to the trifurcation of the popliteal vein and report the presence of calf DVT when identified. However, it is also possible that the true differences are even larger than we have shown. Venous thromboembolism frequently is not limited to the part of the body identified by signs and symptoms. Silent pulmonary embolism occurs in 40–50% of patients with proximal DVT (33–36). It is possible that there are patients with undetected pulmonary emboli counted in the nondiseased column along with patients with undetected calf DVT. If so, the bias caused by lack of an appropriate gold standard would be even greater.

Venous thromboembolism research presents challenges because the best gold standards are invasive, expensive, and somewhat risky. Research on new imaging tests has been simplified by the opportunity to make direct anatomic correlations. Areas of abnormality on duplex ultrasound and spiral computed tomography can be compared with the same areas on lower extremity venography and pulmonary angiography, respectively. For example, research on duplex ultrasound is not invalidated by the limitations of ultrasound in imaging the calf and iliac veins because correlations between ultrasound and venographic imaging of the thigh veins can still be made. Research on spiral computed tomography is not invalidated by its limitations in imaging subsegmental pulmonary arteries because correlations between the larger pulmonary arteries and the same areas on pulmonary angiography can still be made. In addition, these analyses are unaffected by whether patients also have another thrombosis elsewhere.

The d-dimer test has no corresponding capability of showing location of venous thromboemboli but has the potentially important advantage of being used to rule out venous thromboembolism in all locations, including the lungs, upper extremities, pelvis, thigh, and calf. However, to establish this role for d-dimer, we have shown that researchers must carefully choose the appropriate gold standard. It clearly is not enough to compare d-dimer concentrations to lower extremity duplex ultrasound. Is it enough to compare d-dimer concentrations only to venography in patients with suspected lower extremity DVT? Is it enough to compare d-dimer concentrations only to pulmonary angiography in patients with suspected pulmonary embolism? A gold standard that underdiagnoses disease creates a bias that erroneously raises the key index of NPV. For d-dimer research, the definitive diagnostic testing study would be one that uses a comprehensive gold standard for venous thromboembolism, one that includes complete testing for both DVT (e.g., venography) and pulmonary embolism (e.g., lung scan, spiral computed tomography, or pulmonary arteriography).

References


