9.

Ultrasonography for clinically suspected upper extremity deep vein thrombosis

Marcello Di Nisio
Ettore Porreca
Ankie Kleinjan
Patrick MM Bossuyt
Harry R Büller
Anne WS Rutjes

In preparation; to be published
in the Cochrane Database of Systematic Reviews
ABSTRACT

Background and objectives
While ultrasonography (US) is broadly used as the sole diagnostic test to confirm or exclude the presence of upper extremity deep vein thrombosis (UEDVT), the accuracy of this test remains unclear. Therefore, we aimed to determine the sensitivity and specificity of US for the diagnosis of UEDVT, verified by venography, in patients with clinically suspected disease.

Methods
A comprehensive search for studies reporting on the diagnostic accuracy of US for the diagnosis of clinically suspected UEDVT was conducted in MEDLINE and EMBASE databases up to March 2012. Prospective and retrospective cohorts and nested case-control studies on the diagnostic accuracy of US among inpatients and outpatients with clinically suspected acute UEDVT were eligible, as long as the index test results were verified by venography. US included Doppler US, compression US and Doppler with compression US. The QUADAS tool was applied to assess methodological quality.

Results
Of 1053 studies identified by the search, 9 were included (687 patients). Overall, the methodological quality was considered to be low, sample sizes were small, and large between-study differences were observed in spectrum and design. The summary estimates of sensitivity and specificity (95% confidence interval) of US were respectively 89.7 (80.2 to 94.9) and 94.9 (84.4 to 98.4). Doppler US was evaluated in 6 studies (n=258), compression US in 2 (n=157), and Doppler with compression US in 5 (n=272). The summary estimates of sensitivity and specificity were 86.8 (74.2 to 93.8) and 93.8 (78.7 to 98.4) for Doppler US, 87.0 (71.2 to 94.8) and 98.3 (91.7 to 99.7) for compression US, and 93.4 (84.7 to 97.3) and 95.6 (83.8 to 98.9) for Doppler with compression US. No statistically significant difference was found when comparing the summary estimates of Doppler US to either compression US or Doppler with compression US. The exploration of between study heterogeneity showed that the summary estimate of specificity was higher in studies with a low (below 35%) prevalence of UEDVT (p<0.001), while the sensitivity was significantly higher in studies that did not use venography to confirm the presence of UEDVT in all patients (p=0.039).

Authors’ conclusions
Although Doppler US with compression is currently widely used as a substitute for venography, the data from this review suggest caution since confidence intervals around the summary estimates were wide, and the quality of the majority of the studies was low. While awaiting properly sized and designed studies clarifying the diagnostic accuracy of US, the non-invasiveness and ease of the test make US a reasonable option in the work-up of clinically suspected DVT. Repeat US testing or venography may be warranted in doubtful cases, although we currently lack evidence supporting this strategy.
INTRODUCTION

Upper extremity deep vein thrombosis (UEDVT) represents an increasingly important clinical disease with the potential for considerable morbidity due to recurrent venous thromboembolism and the development of post-thrombotic syndrome (1;2). UEDVT involves one or more venous territories between the brachial, axillary, subclavian, and jugular veins. It is estimated that UEDVT currently accounts for up to 10% of all cases of deep vein thrombosis (DVT), with an incidence in the general population of approximately three per 100,000 persons per year (1). The incidence of UEDVT has risen over the past several decades as an effect of the increasing use of central venous catheters for chemotherapy, bone marrow transplantation, dialysis, and parenteral nutrition (3). Nevertheless, in about one third of patients UEDVT occurs in the absence of evident predisposing factors (i.e. primary UEDVT).

As for DVT of the lower extremity, several tests including clinical rules, D-dimer and imaging tests are available for the diagnostic work-up of UEDVT (4-6). Magnetic resonance imaging, rheography, and plethysmography have been poorly evaluated and are not commonly used in routine clinical practice (5). Venography is still considered as the reference standard for the diagnosis of UEDVT although it is an invasive test that is often difficult to perform, not always or not broadly available, and requires the use of ionising radiation which may induce allergic reactions. These limitations have prompted many centres around the world to adopt ultrasonography (US) as the sole diagnostic test to confirm or exclude the presence of UEDVT. US can nonetheless be difficult to perform because of the clavicle, which hinders and limits the possibility to image and to compress the middle part of the subclavian vein, potentially resulting in false negative results. In general, the accuracy of US for clinically suspected UEDVT is still unclear (1;7-9). To be considered as an adequate alternative, US would need to have a sensitivity and specificity of around 95% as for symptomatic lower limb DVT (10).

US for the diagnosis of UEDVT embraces several methods including compression US and Doppler US, used alone or in combination (11). B-mode (grey scale) imaging with or without the addition of colour permits the visualization of the venous vessels, while their patency is verified by means of compression or pulsed-wave Doppler, or both. Compression US assesses the presence or absence of vein compressibility; Doppler US evaluates the characteristics of venous flow including the phasicity, pulsatility, and variation with physiologic manoeuvres. In clinical practice these tests are often used in combination.

The diagnostic accuracy of US for suspected UEDVT has been summarized in a number of narrative reviews (1;7;8) and it was recently evaluated by Mustafa and colleagues who attempted a more quantitative evaluation (9). Although well performed, the systematic review by Mustafa and colleagues was based on a search strategy limited to the MEDLINE electronic database. In our recent systematic review summarising the evidence on all available diagnostic tests for UEDVT, space constraints prohibited a comprehensive evaluation of sources of heterogeneity (5).
Therefore, the objective of the present Cochrane review is to determine the sensitivity and specificity of US for the diagnosis of UEDVT, verified by venography, in patients with clinically suspected UEDVT. Furthermore, we investigated how test accuracy varies with patient group, prevalence, and methodological criteria related to the verification of disease status.

METHODS

Criteria for considering studies for this review
Studies evaluating the diagnostic accuracy of US in patients with clinically suspected UEDVT were eligible as long as the index test results were verified by venography. Prospective and retrospective cohorts and nested case-control designs were eligible. We excluded reports of studies if, after contacting authors for more information, we could not extract the data to calculate a 2x2 table for symptomatic UEDVT. Case reports and other types of case-control studies were not eligible.

Participants
Studies including inpatients and outpatients with clinically suspected acute UEDVT were eligible. Signs and symptoms of UEDVT include swelling, pain and functional impairment, erythema, a heavy and hot limb, and new visible veins at the shoulder girdle. If a study included both clinically symptomatic and asymptomatic participants, we excluded the report if data could not be obtained from the report or the authors to calculate 2x2 tables for symptomatic UEDVT.

Ultrasonography techniques
Index tests included for the present analyses include: compression US (B-mode imaging and compression technique), Doppler US (B-mode imaging and Doppler technique) and Doppler US with compression. We recorded whether colour was added to the B-mode imaging and, where available, information about the variability between operators.

The preferred reference standard was venography, which had to be applied in all or some of the patients. Studies using alternative reference standards, such as magnetic resonance imaging and (spiral) computed tomography or clinical follow up, in a fraction of the patients, were only considered if at least a sample of the patients was verified by venography. The influence of the differential verification on heterogeneity was evaluated.

Search methods and data extraction
We searched MEDLINE and EMBASE databases through the Ovid platform (from inception up to March 2012) to identify studies reporting on the diagnostic accuracy of US for the diagnosis of clinically suspected UEDVT. No language restrictions were applied. The search terms are presented in the Appendix.

Furthermore, Science Citation Index was used to retrieve reports citing the relevant articles identified from our MEDLINE and EMBASE searches, and relevant studies were
entered into PubMed. Subsequently, the Related Articles feature was used as suggested by Sampson and colleagues (12). In addition, we manually screened reference lists of all included studies and related reviews to identify additional potentially eligible studies.

Two review authors independently assessed eligible articles for inclusion from the titles and abstracts obtained in the initial search. Any disagreement was resolved through discussion or by involvement of a third review author. If multiple reports described the same study, we considered all reports to derive suitable diagnostic data for the analyses.

Two review authors independently extracted study characteristics using predefined, piloted data-extraction forms. Any disagreement was resolved by consensus and, if necessary, by involving a third review author. No attempts were made to mask for authorship, journal name, or institution. We extracted information on: author, year of publication, and journal; study design (cohort or nested case-control); timing of data collection (prospective, retrospective); setting (inpatients, outpatients); study population (age, gender, presence of oncologic conditions or central venous lines); type of reference standard; US method; QUADAS-items; and data for the 2x2 table. If 2x2 tables could not be constructed, we contacted the authors for additional data.

Study quality
Two review authors independently assessed study quality using a shortened version of the QUADAS-list, with each item scored as ‘Yes’, ‘No’, or ‘Unclear’ (13). We refer to the original QUADAS publication and the STARD statement for detailed explanations of the potentially biasing effects of suboptimal design characteristics (13;14). We omitted three reporting items from the QUADAS list; addressing the description of the index test, reference standard and selection criteria, as recommended by the Cochrane Handbook for Diagnostic Test Accuracy Reviews (http://srdta.cochrane.org/). These three items were assessed, however, and embedded in the ‘Characteristics of included studies’ table. Results were presented in the text, in a quality graph, and in an ‘assessment of methodological quality’ table. Some of the QUADAS items were used in the exploration of heterogeneity, as described in the section ‘Investigations of heterogeneity’. The items of the shortened QUADAS tool and how they were dealt with, are presented below.

We considered the participant spectrum to be representative when patients were consecutively or randomly selected and if the percentages of patients with cancer or central venous lines were between 35% and 55% (7). We scored ‘No’ if patient selection was not consecutive and not random, or if the proportion was out of the specified range. We classified this item as ‘Unclear’ if any of these characteristics were not reported.

We considered the reference standard acceptable if venography was applied in all patients and the test technology, execution, and interpretation matched current standards. We scored ‘No’ if venography was applied to only a proportion or if the test technology, execution, and interpretation were outdated. We scored ‘Unclear’ if the verification scheme was not reported in sufficient detail to allow a judgement.
The reliability of the diagnostic accuracy estimates of US is higher when the time delay between the US and the final diagnosis is kept reasonably short, to avoid changes in disease status either induced by natural progression or by the start of anticoagulant treatment. We judged a time interval of maximally 24 hours between US and venography as acceptable.

Partial verification occurs when some participants receive no reference standard at all. Partial verification and withdrawals are difficult to disentangle in some diagnostic test accuracy studies. Some argue that partial verification is design based, driven by the risk profile or other test results of patients. We used a broader definition where incomplete verification could be a consequence of the study design, the patient’s or physician’s choice, or other events such as unavailability of imaging equipment. Any ‘withdrawal’ after the application of the index test (US) was counted here, regardless of the reasons for withdrawal. We scored ‘Yes’ if the reference standard was applied in at least 90% of the patients. Although arbitrary, a cut-off of 10% is typically used in reviews or meta-epidemiologic designs of diagnostic test accuracy studies (15-17). We scored ‘Unclear’ if the percentage was not reported and could not be derived from the report.

Differential verification occurs when the diagnosis is made in all participants but by using different reference standards. We scored ‘Yes’ if venography was applied in at least 90% of the patients receiving US. A cut-off of 10% to classify the potential biasing effect of differential verification is typically used in reviews of diagnostic test accuracy studies (15-17). We scored the item as ‘Unclear’ if the percentage was not reported and could not be derived from the report.

Concerning the item ‘incorporation avoided’, we scored ‘Yes’ if ultrasonographic testing was not a prerequisite for diagnosing UEDVT. We scored ‘Unclear’ if it was not clearly described and ‘No’ if authors explicitly stated that US results were used to make the final diagnosis of UEDVT.

Blinding of the index test for reference test results and vice versa was scored ‘Yes’ if the authors explicitly stated that the assessment of US was blinded for the reference test results, and vice-versa. If authors unambiguously described that one test was always undertaken before the other, we scored ‘Yes’ for that test. We scored ‘Unclear’ if blinding was not reported and the sequence of testing was either mixed, random, or not reported. We scored ‘No’ if authors explicitly described an open verification procedure.

Clinical information provided to the researchers should be scored ‘Yes’ if the same clinical data are available when test results are interpreted. Reports typically lack detailed descriptions of the clinical information provided while interpreting test results. We have therefore chosen a pragmatic solution and we scored ‘Yes’ if the authors explicitly mentioned that clinical information was given or if they state that the study was non-blinded for clinical information.

Uninterpretable or indeterminate results are those where the authors report that the index test result does not allow them to conclude with certainty about the presence or absence of the target condition, results are neither positive nor negative. We scored ‘Yes’
if authors reported any US result that was uninterpretable or indeterminate, or if it was clear that there were no uninterpretable or indeterminate results. Although the QUADAS background document states that this item refers to the results of the index test only, we also assessed if any uninterpretable or indeterminate results of the reference standard occurred. We added this information, when present, to the methodological quality table, but it did not influence our classification of ‘Yes’, ‘No’, or ‘Unclear’.

We defined withdrawal as any patient formally included in the study who, for any reason, did not receive US. In addition, patients who received both US and reference standard, but were withdrawn from the 2x2 table, were considered as withdrawals. We scored ‘Yes’ if withdrawals were clearly described or if the report explicitly mentioned that no withdrawals occurred. We scored ‘No’ if the number of patients recruited did not match with the number of patients contributing to the 2 x 2 table and no explanation was given. We scored ‘Unclear’ if the number of patients initially recruited at the start of the study was insufficiently described.

**Statistical analysis and data synthesis**

We included only studies reporting sufficient data for the construction of a 2x2 table. The data in the 2x2 tables were used to calculate sensitivity and specificity for each study. We present individual study results graphically by plotting the estimates of sensitivity and specificity (and their 95% confidence intervals (CI)) in forest plots and, if thought illustrative, in the receiver operating characteristic (ROC) space (RevMan 5 software; Nordic Cochrane Centre, The Cochrane Collaboration, Copenhagen). RevMan 5 software was also used to construct methodological quality summary graphs. We used a bivariate random-effects approach to obtain summary estimates of the pairs of sensitivity and specificity and to construct a summary ROC curve (18). Our main analysis was a bivariate model with a covariate indicating the type of ultrasound (Doppler US versus Doppler with compression US or versus compression US). ‘Bivariate models’ was performed with the xtmelogit module in STATA statistical software, version 12.1 (StataCorp, College Station, Texas). P-values below 0.05 were considered to indicate statistical significance.

Heterogeneity was investigated stratified per type of US method. We explored several other possible sources of heterogeneity related to spectrum and design characteristics. The investigation of heterogeneity was performed through visual examination of both the ROC plot of raw data and the forest plots of sensitivities and specificities. We formally explored sources of variation in bivariate models, by adding covariates indicating patient or design features. This enabled us to explore whether, on average, studies that differ with respect to these features resulted in different estimates of diagnostic accuracy. The main sources of heterogeneity are likely to be related to differences in verification tests and the existence of clinical subgroups (spectrum effects). We anticipated inclusion of a low number of studies, thus we restricted our exploration of heterogeneity to these potential sources of variation. Our hypothesis was that studies using venography as the
sole reference standard result in different estimates of sensitivity and specificity compared with studies using venography in a (selected) subset of patients and other tests such as magnetic resonance imaging, (spiral) computed tomography or clinical follow up in the remainder (differential verification). Similarly, we expected that studies using complete verification versus incomplete (partial) verification resulted in different estimates of accuracy. The evaluation of possible effects of clinical subgroups was analysed according to risk groups and the prevalence of UEDVT. We first classified studies into those reporting predominant inclusion of patients at high risk of UEDVT (that is patients with cancer, central venous catheters, or coagulation defects) versus studies reporting other sampling methods. Second, we classified the studies into high prevalence versus typical or low prevalence studies. The prevalence typically seen in an unselected cohort of symptomatic patients suspected of UEDVT is as high as 50% (7). We defined a prevalence of up to 40% as low, between 40% and 60% as typical, and above 60% as high.

We planned one sensitivity analysis, which was to restrict the analysis to studies that used venography as the reference standard in all patients.

Deeks and colleagues have shown that the tools available for evaluating publication bias in intervention studies do not work as good for test accuracy studies (19) and the common tests, based on the standard error of the estimated odds, give misleading results. Deeks 2005 suggested the use of an effective sample size plot to detect publication bias, but acknowledged that the corresponding tests of asymmetry lack power in situations where, for example, sample variability is present. In addition, Leeflang and colleagues concluded that tests for publication bias are typically not useful (15). Therefore we did not plan to use funnel plots to evaluate the impact of publication bias or other biases associated with small studies.

RESULTS

Search strategy
The search strategy yielded 1053 studies, 25 were eligible based on the title/abstract. Nine studies were included (20-28) and sixteen excluded for the following reasons: 2x2 table could not be reproduced (29-31), not a diagnostic accuracy design (32-34), reviews, case-reports, or editorial (35-39); more than one of the above (40;41).

Methodological quality
The spectrum was considered to be representative in 1 study (11%, Figure 1; Figure 2). One article reported inclusion to be consecutive but omitted the percentage of patients with cancer or central venous lines (27). Venography was applied in 7 studies (78%) as the sole reference standard, avoiding the risk of differential verification bias, and in 5 (55%) it was applied in all patients, avoiding the risk of partial verification bias. The existence of uninterpretable or intermediate results and withdrawals were explained in only 2
Accuracy of ultrasonography in suspected UEDVT

studies (22%). The proportion of differential and partial verification, withdrawals and uninterpretable results are given in the appendix. Only two studies fulfilled all quality criteria and were considered to be at low risk of bias (20;26).

**Findings**

Nine studies (687 patients) evaluated in total 13 test comparisons, including Doppler US in 6 studies (258 patients) (20-22;25-27), compression US in 2 studies (157 patients) (20;26), and Doppler with compression US in 5 studies (272 patients) (20;23;24;26;28). UEDVT was diagnosed in 301 patients for an overall prevalence of 44%.

The bi-variate random effects approach resulted in summary estimates of sensitivity and specificity of 86.8 (95%CI 74.2 to 93.8) and 93.8 (95%CI 78.7 to 98.4) for Doppler US, 87.0 (95%CI 71.2 to 94.8) and 98.3 (95%CI 91.7 to 99.7) for compression US, 93.4 (95%CI 84.7 to 97.3) and 95.6 (95%CI 83.8 to 98.9) for Doppler with compression US. The overall sensitivity and specificity of US were respectively 89.7 (80.2 to 94.9) and 94.9 (84.4 to 98.4). Investigation of heterogeneity showed that these estimates were significantly affected by the type of US method (p=0.021) as well as the prevalence of UEDVT (p= 0.0037), but not by differential and partial verification (Table 1). Although the model including a covariate for the type of US fitted the data better than a model without (P-value from Likelihood-ratio test = 0.021), no statistically significant differences between estimates of sensitivity and specificity of the

<table>
<thead>
<tr>
<th>Table 1. Summary of results</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Variable</strong></td>
</tr>
<tr>
<td>All studies</td>
</tr>
<tr>
<td>Type of Ultrasound</td>
</tr>
<tr>
<td>Doppler US</td>
</tr>
<tr>
<td>Compression US</td>
</tr>
<tr>
<td>Doppler US with compression</td>
</tr>
<tr>
<td>UEDVT Prevalence</td>
</tr>
<tr>
<td>Typical prevalence</td>
</tr>
<tr>
<td>Low prevalence</td>
</tr>
<tr>
<td>High prevalence</td>
</tr>
<tr>
<td>Reference standard</td>
</tr>
<tr>
<td>All venography</td>
</tr>
<tr>
<td>Differential verification</td>
</tr>
<tr>
<td>Type of verification</td>
</tr>
<tr>
<td>Complete</td>
</tr>
<tr>
<td>Incomplete (partial)</td>
</tr>
<tr>
<td>Unclear</td>
</tr>
</tbody>
</table>

* P-values have been derived by comparing a model with and without the respective design feature, assuming equal variances (LR test)
Three US techniques could be detected (Table 2; summary relative sensitivity and specificity of compression US versus Doppler US \( p=0.965 \) and \( p=0.171 \); summary relative sensitivity and specificity for Doppler with compression US versus Doppler US \( p=0.118 \) and \( p=0.447 \), respectively). Two studies directly compared the three US methods in the same set of
patients (20;26). While the relative small size of the studies precluded a significant statistical comparison of the tests within each study, visual inspection of the forest plots suggested that no US method performed consistently better than the others. The estimates of both sensitivity and specificity had broad confidence intervals and varied largely between the two studies. As an example, Doppler with compression US had a sensitivity of 100% in the study of Prandoni 1997, and only 82% in the study of Baarslag 2002 with the lower limit of the 95%CI as low as 67%. The corresponding estimates for specificity were 93% and 82% with lower limits of the 95%CI below 70% in both studies.

The model including a covariate for the prevalence of UEDVT fitted the data better than a model without (Table 1; p-value from Likelihood-ratio test: 0.0037). The summary relative specificity in studies with a low UEDVT prevalence (<35%) were significantly higher (logit relative specificity 0.13; 0.07 to 0.19) compared to studies with typical prevalence (between 40% and 60%). The sensitivity did not differ between studies with low or high prevalence versus those with typical prevalence (Table 2). Partial verification had no significant effect on test accuracy estimates. Studies using different reference standards to verify the diagnosis of UEDVT (differential verification) appeared to have a higher US sensitivity compared to studies using venography as the sole reference standard (logit relative sensitivity 0.12; 0.01 to 0.24), with no significant influence on specificity. In sensitivity analyses where only the studies that used venography as the reference standard in all patients were included, the sensitivity and specificity of US were respectively 85.4 (74.2 to 92.2) and 93.2 (77.6 to 98.2).

Figure 3. Summary receiver operating curve plots. Doppler with compression (squares), Doppler ultrasonography (triangles) and compression ultrasonography (circles).
Table 2. The optimal design feature is set as the reference category (*), to which the other features are compared. For the type of ultrasound, Doppler US was set as reference category, to enable the most relevant comparisons. P-values refer to the relative sensitivities and specificities (logit scale), where a significant P-value indicates that the summary estimate of studies with a suboptimal design feature is significantly different from the reference category.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Logit relative sensitivity (95%CI)</th>
<th>P-value</th>
<th>Logit relative specificity</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type of Ultrasound</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Doppler US</td>
<td>Ref Cat*</td>
<td>-</td>
<td>Ref Cat</td>
<td>-</td>
</tr>
<tr>
<td>Compression US</td>
<td>0.00 (-0.10 to 0.10)</td>
<td>0.965</td>
<td>0.05 (-0.02 to 0.11)</td>
<td>0.171</td>
</tr>
<tr>
<td>Doppler US with compression</td>
<td>0.07 (-0.02 to 0.16)</td>
<td>0.118</td>
<td>0.02 (-0.03 to 0.07)</td>
<td>0.447</td>
</tr>
<tr>
<td>UEDVT Prevalence</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Typical prevalence</td>
<td>Ref Cat</td>
<td>-</td>
<td>Ref Cat</td>
<td>-</td>
</tr>
<tr>
<td>Low prevalence</td>
<td>0.04 (-0.12 to 0.19)</td>
<td>0.619</td>
<td>0.13 (0.07 to 0.19)</td>
<td>0.000</td>
</tr>
<tr>
<td>High prevalence</td>
<td>-0.03 (-0.26 to 0.21)</td>
<td>0.825</td>
<td>-0.27 (-0.84 to 0.30)</td>
<td>0.356</td>
</tr>
<tr>
<td>Reference standard</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All venography</td>
<td>Ref Cat</td>
<td>-</td>
<td>Ref Cat</td>
<td>-</td>
</tr>
<tr>
<td>Differential verification</td>
<td>0.12 (0.01 to 0.24)</td>
<td>0.039</td>
<td>0.04 (-0.07 to 0.15)</td>
<td>0.448</td>
</tr>
<tr>
<td>Type of verification</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Complete</td>
<td>Ref Cat</td>
<td>-</td>
<td>Ref Cat</td>
<td>-</td>
</tr>
<tr>
<td>Incomplete (partial)</td>
<td>0.11 (-0.02 to 0.25)</td>
<td>0.111</td>
<td>0.12 (-0.39 to 0.15)</td>
<td>0.384</td>
</tr>
<tr>
<td>Unclear</td>
<td>0.11 (-0.03 to 0.26)</td>
<td>0.114</td>
<td>0.06 (-0.03 to 0.16)</td>
<td>0.205</td>
</tr>
</tbody>
</table>

**DISCUSSION**

**Summary of main results**

While the accuracy of US has been largely established for clinically suspected DVT of the lower limbs, the accuracy of the method relative to venography in clinically suspected UEDVT remains inconclusive. In the current review, the overall sensitivity of US was 89.7% and the specificity 94.9%, but confidence intervals were wide, and between study heterogeneity was large. No statistically different estimates were observed between Doppler US, compression US, and Doppler with compression US.

The lower limit of the confidence interval for sensitivity was as low as 80.2% implying that up to 20% of UEDVT could be missed by US, potentially posing patients at risk of developing (fatal) pulmonary embolism. Similarly, the percentage of false positive results was as high as 15%, which would expose these patients to unnecessary anticoagulant treatment. In the exploration of between study differences, we found that the specificity of the test varied with the prevalence of the disease. Specificity appeared significantly worse in studies including patients with a prevalence of UEDVT above 35%. In the pre-specified subgroup with a high prevalence of UEDVT, which could include patients with cancer carrying central venous catheters, the specificity was 66.6% with a lower bound as low as 26.7%, which would translate into an unacceptably high rate of false positive results. Although these data would suggest that positive US findings prompt further testing with
either repeat US or venography, any such interpretation is hampered by the uncertainty around the estimate and the broad confidence intervals around it.

In general, our summary estimates appear comparable to those on US for lower limb DVT where the specificity is around 94% whereas sensitivity is 89%, 97%, 73% for overall, proximal, and distal lower DVT, respectively (10;20), but the evidence they rely upon is not as strong and convincing. Several studies have demonstrated the accuracy of US in the diagnosis of DVT of the lower extremities which has led many centres worldwide to replace venography with US (10). By extension of these findings, physician often prescribe US to confirm or exclude UEDVT. However, the interpretation of US may be more challenging in the upper extremities where anatomical constrains do not allow to image and compress adequately few segments of the vein as the middle part of the subclavian vein hindered by the clavicle, potentially resulting in false negative results. The diagnostic accuracy of US for suspected UEDVT has been the object of a number of reviews which, however, lacked a systematic assessment of the literature, did not systematically check the effect of sources of bias on the diagnostic accuracy indexes (1;7-9), and did not attempt a more quantitative evaluation (1;7;8). In a recent systematic review, we summarised the available evidence on the accuracy of the US for clinically suspected UEDVT (5), but we did not explore the influence of study design characteristics on the accuracy estimates due to space constrains.

Strengths and weaknesses of the review
The main weakness of this review is the uncertainty around the estimates and the size of the estimates which was due to the small number of studies with relatively few patients. For instance, visual inspection of the summary ROC suggested that Doppler with compression US was superior to Doppler US in terms of sensitivity or specificity, although these differences were not statistically significant. Latter finding may thus be related to the relative low power of the analysis or due to the absence of a true difference. Another major limitation in the interpretation of the findings was that most of the studies presented significant methodological shortcomings which further reduced the confidence in the estimates. An attempt was done to assess the influence of possible sources of bias such as partial or differential verification on the summary estimates, although this was again hampered by the low number of studies. For some studies, 2x2 tables were not based on patient level, but on the number of arms examined. We did not account for such correlation in the data nor corrected for the number of patients analysed, which may have artificially inflated the power of the analysis.

Applicability of findings to clinical practice and policy
When evaluating a patient with suspected UEDVT, physicians should be aware of the relative scarce evidence regarding the accuracy of US in this setting, and consider the possibility of further testing by venography in cases that remain doubtful after US.
Conclusions

Doppler with compression US may be an acceptable alternative to venography, but this needs to be confirmed in adequately designed studies. The potential adverse events and the non feasibility of venography in approximately 20% of patients due to renal dysfunction, poor vascular access, or contrast allergies, should be taken into consideration. In the absence of large accuracy or management studies which could clarify whether the benefits gained from the exclusive use of venography would justify the associated risks, it seems reasonable not to suggest venography for all patients with clinically suspected UEDVT, but to limit its use to the clinical situations where there is a concern for false positive or false negative ultrasound results (S).

Implications for research

Future studies should focus on direct comparisons between the US methods in adequately powered studies, or should evaluate the effective implementation of US in the clinical routine in terms of patient relevant outcomes. In addition, questions remain regarding the use of a single test versus serial US, the combination of US with pre-test clinically probability and/or D-dimer test within diagnostic algorithms, and the accuracy of US in patient groups at different prevalence of UEDVT. A large prospective study evaluating the safety and feasibility of a diagnostic algorithm including pre-test clinical probability, D-dimer test and (serial) US in patients with clinically suspected UEDVT is completing follow-up and the results are eagerly awaited by 2013 (ClinicalTrials.gov Identifier: NCT01324037).

REFERENCE LIST

Accuracy of ultrasonography in suspected UEDVT

**SUPPLEMENTARY INFORMATION: SEARCH STRATEGY**

*Medline*

**Search terms for upper extremity DVT**
1. venous thrombosis/
2. vein$ thrombo$.tw.
3. (deep adj2 thrombo$).tw
4. venous thrombo$.tw.
5. upper extremity/
6. arm/
7. arm$1.tw
8. forearm/
9. upper limb$.tw.
10. upper extremit$.tw.
11. axilla/
12. axillary vein/
13. subclavian vein/
14. (axillary or subclavian).tw.

**Search terms for ultrasonography**
15. ultrasonography/
16. ultrasonography, Doppler/
17. ultrasonography, Doppler, color/
18. ultrasonography, Doppler, duplex/
19. ultrasonography, Doppler, pulsed/
20. (ultrasonography or ultrasound).tw
21. ultrasonic imaging.tw

**Combining terms**
22. animal/
23. animal/ and human/
24. 22 not 23
25. or/1-14
26. or/15-21
27. and/25-26
28. 27 not 24
29. remove duplicates from 28

The “/” refers to MeSH, medical subject headings, and (tw) to text word in the title or abstract; the $ is a truncation character which allows all possible suffix variations of the root word.
Embase

Search terms for upper extremity DVT
1. vein thrombosis/
2. vein$ thrombo$.tw
3. (deep adj2 thrombo$).tw
4. venous thrombo$.tw
5. Deep vein thrombosis/
6. arm/
7. arm$1.tw
8. upper limb$.tw
9. upper extremit$.tw
10. forearm/
11. axilla/
12. axillary vein/
13. subclavian vein/
14. (axillary or subclavian).tw

Search terms for ultrasonography
15. ecography/
16. Doppler ecography/
17. gray scale ecography/
18. real time ecography/
19. Doppler flowmetry/
20. color ultrasound flowmetry/
21. ultrasonogr$.tw
22. ultrasonic imaging.tw
23. ultrasound.tw

Combining terms
24. animal/
25. animal/ and human/
26. 24 not 25
27. or/1-14
28. or/15-23
29. and/27-28
30. 29 not 26
31. remove duplicates from 30

The "/" refers to EMTREE, medical subject headings, and (tw) to text word in the title or abstract; the $ is a truncation character which allows all possible suffix variations of the root word.