The protein C system and venous thromboembolism in reproduction

Pathophysiology, epidemiology and diagnosis

van Mens, T.E.

CHAPTER 8

Diagnostic tests for pregnancy-related deep vein thrombosis (Protocol)

Bistervels IM
Scheres LJJ
Spijker R
Middeldorp S
van Mens TE

Cochrane Diagnostic Test Accuracy Review Protocol
Submitted

Background

Target condition being diagnosed

Venous thromboembolism (VTE) is a disease that can manifest as both deep vein thrombosis (DVT) and pulmonary embolism (PE). Pregnancy is a major risk factor for VTE in women of fertile age. Pregnant women have a five-fold increased VTE risk compared to non-pregnant age-matched controls [1–3]: in absolute terms, the incidence of pregnancy-related VTE, meaning VTE either during pregnancy or postpartum, is one to two in a 1000 pregnancies [4–7]. This is caused by mostly physiological changes during pregnancy and the puerperium: i) immediately following conception, procoagulant factors rise while natural anticoagulant and fibrinolytic protein activities decrease, ii) obstetric interventions and delivery lead to vascular damage, and iii) compression of the gravid uterus on the pelvic venous system, venous distension and increased intravascular volume causing venous stasis [8].

Episodes of VTE are equally distributed over pregnancy and the postpartum period [9]. The risk of VTE remains elevated until three months after delivery, but 80% of postpartum VTE occur in the first three weeks of the postpartum period and the risk reduces over time [10,11].

Certain aspects of the manifestation of DVT are specific to pregnancy. Due to compression by the gravid uterus, the proximal veins, including the femoral, iliac and caval veins, are more frequently affected than the distal veins (72% in the pregnant population versus 9% in the non-pregnant population) [12]. Because of the asymmetrical vascular anatomy, the gravid uterus exerts more pressure on the left iliac vein then on the right, causing a predominance of left-sided DVT in pregnancy (over 80%) [13].

DVT requires immediate treatment with anticoagulants, the main reason being the capacity of the thrombus to embolize to the pulmonary circulation causing potentially fatal PE. Moreover, pregnancy-related DVT is associated with a high risk of post-thrombotic syndrome, a debilitating cluster of chronic leg symptoms including skin discolorations, swelling, pain and ulcers, associated with a decreased quality of life [14]. According to data presented by Wik et al. [15], over 40% of pregnant women develop post-thrombotic syndrome, while the incidence of post-thrombotic syndrome in the (older) general population ranges from 20% to 50% [16–19].

Heparins are the anticoagulants of first choice for treatment of VTE during pregnancy as they do not cross the placenta [20]. In the postpartum period, vitamin K antagonist

can be used safely when lactating, while direct oral anticoagulants can be detected in breast milk and are therefore contraindicated when breastfeeding [20]. Anticoagulant therapy is effective but it entails a risk of haemorrhage, challenges for management of the delivery and neuraxial anaesthesia, and importantly significant discomfort to the patient who is required to subcutaneously self-inject on a daily basis. A diagnosis of pregnancy-related DVT has consequences for decisions on anticoagulant therapy in subsequent pregnancies and contraception counselling as well [20].

This review protocol is complementary to existing Cochrane review protocols on the use of duplex ultrasound for the diagnosis of DVT in the lower limb [21] and on D-dimer tests in combination with a clinical prediction rule score for the diagnosis of DVT in symptomatic hospital outpatients [22], since the current review protocol focusses specifically on the pregnancy-related DVT. Non-Cochrane reviews on this topic include a systematic review by Nijkeuter et al from 2006 [23].

Index tests

The following index tests will be reviewed: clinical decision rules (CDR) and D-dimer testing performed as triage tests, and ultrasound examination (US) and magnetic resonance imaging (MRI), performed either as first-line tests or following other diagnostic testing.

CDRs are decision tools that are used to predict the likelihood of a DVT and identify patients that require further testing. CDRs sometimes incorporate D-dimer testing. Outside pregnancy, the Wells score is often used [24,25]. A low Wells score combined with a normal D-dimer level rules out DVT without the need for further imaging. Most studies evaluating CDRs have excluded pregnant patients.

D-dimer testing is a well-established test used outside pregnancy to rule out DVT in patients with a low pre-test probability. D-dimer is a degradation product of fibrin, the final clot-forming protein in the coagulation cascade. Elevated D-dimer plasma levels indicate activated coagulation and fibrinolysis and can have several causes. A concentration below 500 μg/L is considered normal [25]. Various assays from different manufacturers are available, including enzyme-linked immunofluorescent immunoassays, microplate and membrane enzyme-linked immunosorbent assays (ELISA), and latex and whole blood agglutination assays [26]. Different tests have different advantages. Some are rapid and easy to perform, such as agglutination assays. In terms of accuracy, the enzyme-linked immunofluorescent immunoassays, microplate ELISA and latex quantitative assays have the highest sensitivity (96%, 94% and 93%, respectively) [26].

US is the main imaging test for DVT outside pregnancy. It is a widely available,

inexpensive, non-invasive test that reveals the two-dimensional structure of, and the blood flow in, the vessels [27]. The technique is based on recording and displaying of an ultrasound signal that is reflected to a different degree by different tissues. Compression US can be performed as proximal compression US (compression of only femoral and popliteal vein) or as whole leg US (compression of proximal veins and calf veins). Veins are normally fully compressible with the ultrasound probe and a non-compressible vein is diagnostic of a (residual) clot in the vein, i.e. DVT. Compression US cannot be applied to veins proximal of the groin, but modern US machines combine ultrasound with Doppler flow measurements. Flow is visualized in different ways, including with pulsed wave and colour mode. The Valsalva maneuver can be performed during the examination to assess flow changes indicative of proximal vein patency. The accuracy of US is partly dependent on the experience and skills of the examiner. A systematic review estimated sensitivity of 89.7% (95% confidence interval (CI) 88.8% to 90.5%) and specificity of 93.8% (95% CI 93.1% to 94.4%) [28]. US is not harmful for the fetus and can therefore be used safely.

A less commonly used imaging technique for DVT is MRI. The technique captures anatomical images by applying magnetic fields that are perturbed by radiofrequency pulses. Advantages include avoidance of ionising radiation, high resolution and good visualization of the pelvic veins [29]. MRI is however relatively expensive, intravenous contrast is sometimes required and availability is not universal. Magnetic resonance direct thrombus imaging (MRDTI) is a special application of MRI. It utilizes the endogenous contrast agent of methaemoglobin in a thrombus and therefore requires no contrast injection [30–33].

Computer tomography angiography is another, highly accurate, imaging modality for DVT [34]. However since it applies ionising radiation to the lower body, the technique is unsafe in pregnancy. It will hence not be evaluated as index test for pregnancy-related DVT in this review.

Clinical pathway

Patients with pregnancy-related DVT typically present with pain, swelling, redness and warmth of the affected leg. Proximal thrombosis can present with pain in the pelvic, lower back or buttock region, especially when the iliac and caval veins are affected. Inferior caval vein thrombosis can cause bilateral leg symptoms.

For the diagnosis of DVT outside pregnancy, CDRs - generally the Wells score [25] - combined with D-dimer testing is used as a diagnostic strategy. A low clinical suspicion based on such a rule combined with normal D-dimer levels is sufficient to rule out DVT without further imaging. Elevated D-dimer levels or high clinical

suspicion demand additional imaging testing, which will often be US. If a proximal compression US is negative but pretest probability is high, the test is repeated after 5-7 days (serial testing). Alternatively, a whole leg US can be performed. A negative whole leg US rules out DVT [35].

The diagnostic pathway in the case of suspicion of pregnancy-related DVT may differ among physicians, centers and patients. CDRs have not been validated for diagnosis of pregnancy-related DVT. In general, ultrasound of the leg veins will be performed as a first line test. The leading American College of Chest Physicians guideline recommends proximal compression US as a first-line test in pregnant patients suspected of DVT [35]. In case of a negative US, additional testing with either serial compression US or D-dimer testing at the time of presentation is advised to rule out DVT. In patients suspected of having an isolated iliac thrombosis and who have a negative compression US, either Doppler US, venography or MRI are recommended. MRI does not have a routine place in pregnancy-related DVT diagnosis otherwise, but it might potentially prove to be a valuable ionising radiation-free modality with good visualization of pelvic veins.

Rationale

Pregnant and postpartum patients are a distinct clinical subgroup when it comes to diagnosing DVT. Physiological changes during pregnancy often result in mild leg edema and pelvic and lower back pain, and hence signs and symptoms of DVT may be masked. Conventional CDRs used to triage patients with a suspicion of DVT have not been validated in pregnancy or postpartum patients. D-dimer levels increase progressively throughout healthy pregnancy, often above the diagnostic cut-off value, and remain elevated after delivery [36]. The gravid uterus makes imaging of the proximal veins challenging, while at the same time the most proximal veins are more likely to be affected in isolation from popliteal veins during pregnancy than in non-pregnant patients. Approximately 10% of ultrasounds performed in pregnant women confirm the clinical suspicion of a DVT [37].

An accurate diagnosis of pregnancy-related DVT is crucial. A venous thrombus embolizing to the lung can potentially become fatal for mother and child. On the other hand, false positive diagnoses with subsequent anticoagulant therapy must be avoided as delivery and neuraxial anaesthesia may be complicated by excessive bleeding. A (false) positive diagnosis also has implications for future thromboprophylaxis, both during and outside pregnancy, and the counselling on hormonal contraceptive use.

Approaches to the diagnosis of pregnancy-related DVT vary in clinical practice and guidelines are based on low-level evidence [35,38]. Recommendations are partly

extrapolated from findings in the non-pregnant population [35]. The research field is hampered by the relative contra-indication of the reference test of venography because of the radiation exposure to the fetus.

Objectives

To determine the diagnostic accuracy of CDRs, D-dimer, US and MRI for detecting DVT during, respectively, pregnancy and the postpartum period. Effects of varying D-dimer cut-off values on test accuracy will be investigated. CDR and D-dimer are analysed as triage tests and will not be directly compared to each other since they are not alternatives but potentially complimentary, analogous to the application in non-pregnant patients. US and MRI are alternatives for the diagnosis of lower extremity DVT in pregnancy and the postpartum period and will therefore be compared for diagnostic superiority if sufficient data are available.

Secondary objectives

Secondary objectives shall be the following:

- To determine the number of inconclusive test results for US and MRI, defined as neither showing nor excluding DVT.
- To determine the prevalence of DVT in the included study populations.
- To investigate the effects of the following clinical factors on the accuracy outcomes of index tests:
    1. Prior testing during the same pregnancy (this will apply to US and MRI)
    2. Trimester
    3. History of VTE
- To evaluate accuracy according to different technological aspects of D-dimer assays, US and MRI. These will include D-dimer assay types; use of Doppler flow measurement in US; whole leg versus proximal versus US; single versus serial US; MRI scanner type and scanning technique.
- To explore accuracy for different thrombus locations for US and MRI: distal (distal from popliteal vein), proximal (including and proximal from popliteal vein, but below the groin), and pelvic (above the groin).

Methods

Criteria for considering studies for this review

Types of studies

We will include diagnostic cohort studies with a prospective or retrospective design. We will consider studies that report at least three DVT diagnoses, counting pregnant or post-partum women separately. Cohorts with lower numbers are deemed to not contribute to accurate summary point estimates. We will exclude case-control studies as these can overestimate the clinical relevant estimates of sensitivity and specificity [39]. We will only include studies that present the following data in extractable format: index test results and reference standard results for pregnant and post-partum patients separately.

Participants

We will include data on patients who presented with symptoms and signs of DVT of the lower extremities during pregnancy or in the three months postpartum in the hospital in- or outpatient setting. We will not include data on D-dimer performance in patients that received therapeutic dose anticoagulants. If this is not reported at a patient level, we will only include in the analyses on D-dimer data from studies in which < 20% of patients received therapeutic dose anticoagulants for > 24 hours at the time of the blood draw. Prophylactic doses of anticoagulants will be allowed. Clinical suspicion of DVT and confirmation of pregnancy will be left to the discretion of the authors of the included studies. We will not include data on patients suspected of PE who underwent diagnostic testing to detect asymptomatic DVT.

Index tests

The index tests for this review comprise of all diagnostic tests with a potential role in the diagnostic work-up of pregnancy-related DVT: CDRs, D-dimer levels, US and MRI. Because of the variety in technical aspects of the index tests and the anticipated low yield of the literature search, we will not exclude studies based on index test characteristics. We will carefully report details of the technique evaluated and investigate the effect on test accuracy where possible. D-dimer cut-off levels can be either conventional (500 μg/L) or adjusted. All US and MRI modalities will be eligible for inclusion, as well as all CDRs including physician's clinical judgement, i.e. gestalt, if this can be evaluated against the reference standard. If a patient has undergone the same index test more than once, we will include data on the first test performed. If multiple types of index tests were used, we will include all results. Data on original assessments of the index test result in the clinical setting will be preferred over re-

assessment for study purposes.

Target conditions

Symptomatic DVT, either primary or recurrent, of the lower extremities, including the pelvic veins and the inferior vena cava.

Reference standards

For US and MRI as index tests, the reference standard will be clinical follow-up or venography. When venography is used, it overrules the results of clinical follow-up. However, it will likely not be performed in any study due to radiation exposure to the fetus.

The reference standard for CDR and D-dimer level testing will be clinical follow-up, which may include US, MRI and/or venography that is performed as index test in the primary study. US and MRI are thus part of the reference standard for CDR and D-dimer, even though they are also investigated as index tests in this review. US and MRI are evidently not gold standards. However, they will have a substantially superior accuracy compared to the screening tests of CDR and D-dimer. Given the place of CDR and D-dimer in the clinical pathway as triage tests, and the anticipated absence of studies applying more rigorous reference standards, we deem the imaging modalities in combination with clinical follow-up acceptable reference standards.

Another issue is that the patients with a negative CDR or D-dimer will likely only receive follow-up, without imaging testing, while patients with a positive CDR or D-dimer will have follow-up that includes imaging, leading to a form of differential verification. We might alternatively not include US and/or MRI results into the reference standard for any patient in the analysis of CDR and D-dimer, which would eliminate differential verification. However, ignoring the data from the imaging tests, which have high accuracy compared to the triage tests, in patients marked as high probability of DVT by D-dimer or CDR would lead to less reliable accuracy estimates. We therefore will accept all available follow-up information including imaging to obtain reliable estimates, at the cost of the possibility of differential verification bias.

Clinical follow-up will be considered positive in case of occurrence of any DVT or PE within three months after the index test has taken place, confirmed by objective testing. We recognize that a new episode of VTE might occur during the follow-up period, especially in the presence of additional risk factors (e.g. caesarian section, immobility) [20]. However, the absolute risk is relatively low. A positive test during follow-up therefore likely indicates a previously missed DVT. In case of mortality during clinical follow-up, we will assume PE to be the cause of death when there is no likely alternative diagnosis and no autopsy has been performed. Under these circumstances, we will consider the reference standard as positive, giving rise to a

conservative estimate of the sensitivity of the index test.

Search methods for identification of studies

Electronic searches

We will search Medline (Ovid), Embase (Ovid), BIOSIS Previews (Ovid), Science Citation Index Expanded (ISI Web of Knowledge), and HTA section of CRD database. See Appendix 1 for proposed MEDLINE search strategy. We will design similarly-structured search strategies using search terms appropriate to each database. We will use MeSH terms and other controlled vocabulary where appropriate. In short, we defined five concepts and combined these to achieve an optimal balance between precision and recall of the search.

1. Patients: pregnant or postpartum patients
2. Target condition: DVT
3. Index test: D-dimer, ultrasound, and MRI
4. DTA filter [40]
5. Prediction filter [41]

The final combination will be 1 AND 2 AND (3 OR 4 OR 5).

For details and results from unpublished and ongoing studies, we will additionally search:

- World Health Organization International Clinical Trial Platform (http://apps.who.int/trialsearch/)
- ClinicalTrials.gov (http://clinicaltrials.gov/)
- Current Controlled Trials (http://www.controlledtrials.com/)

The search will be performed by an official Cochrane Group information specialist.

Searching other resources

Grey literature and proceedings: chosen electronic databases include assessments of conference proceedings.

Handsearching: We will not perform handsearching as there is little published evidence of the benefits of handsearching for diagnostic studies [42].

Reference lists: We will check the reference lists of all relevant studies and reviews in the field for further possible titles and will repeat the process until no new titles are found [43].

Correspondence: We will contact research groups who have published or are conducting work on DVT diagnosis, informed by results of initial search.

We will use relevant studies in PubMed to search for additional studies using the 'Related article' feature. We will examine key studies in citation databases such as Science Citation Index to ascertain any further relevant studies.

Data collection and analysis

Selection of studies

After performing the electronic search, two review authors (IB and LS) will independently screen the titles and abstracts of all the retrieved articles for eligibility. The same authors will then independently read the full text of the selected studies and apply in- and exclusion criteria. At all stages of the selection and data extraction process we will resolve any disagreement through discussion with the other review authors (SM and TvM). We will document the reason for exclusion of studies that undergo full-text review. We will present our selection process using a PRISMA flow diagram [44].

Data extraction and management

Two review authors (IB and LS) will independently extract data and assess the quality of the studies. Data extraction will be done using a data extraction form designed for the current review which we will pilot on two studies. We will fill two by two contingency tables containing true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN) with the extracted data from the studies. We allow ourselves to back-calculate the data from the test accuracy estimates (sensitivity, specificity, positive and negative predictive values), when the absolute numbers are not provided in the study. We will also collect data on baseline characteristics, mortality, adverse events, inconclusive tests, previous VTEs, the use of thromboprophylaxis, technical aspects if the index and reference tests, and the presence of additional risk factors (e.g. caesarean section, immobility) at baseline and during the interval between the index and the reference test.

Assessment of methodological quality

We will use the Quality Assessment of Diagnostics Test Accuracy Studies-2 tool (QUADAS-2) to score included studies on risk of bias and applicability to the review's research question [45]. We have adjusted the standard tool to our review question (Appendix 2). We will apply the QUADAS-2 tool for each index test separately for a

single study that applies multiple tests. Clinical follow-up is accepted as a reference standard in this review as the authors anticipate gold standard imaging will not be performed in the pregnant population. Clinical follow-up has severe limitations however, which we have included in the review-tailored signalling questions in the QUADAS-2 tool and which we will deal with methodologically as described in the statistical analysis section below. We do not have a signalling question on case-control studies, since these studies will be excluded. A signalling question about use of anticoagulant treatment was added. Two review authors (IB and LS) will independently complete the QUADAS-2 assessments. Completed QUADAS-2 forms will be compared and disagreements will be solved by discussion with the other review authors (SM and TvM).

Statistical analysis and data synthesis

Criteria for index test positivity will be left to the discretion of the authors of the primary studies. Serial US will be regarded as a single index test in the main analysis.

The reference standard for US and MRI will consist of clinical follow-up. Clinical follow-up of a patient who has had a DVT according to the index imaging test will be inadequate to revert that diagnosis, since the follow-up is aimed at identifying and not excluding DVT. The most common alternative explanations for symptoms suggestive of DVT will be self-limiting. Patients will be treated for DVT and symptoms will likely regress. A false positive will thus mimic a true positive in these studies. For US and MRI, we therefore regard the false positive rate as invalid. Consequently, the only reliable test accuracy statistics are sensitivity and negative predictive value (NPV), which will be the reported outcome measures for US and MRI. Although this does not take into account the inherent trade-off between sensitivity and specificity resulting from implicit threshold variations [46], calculating test accuracy using false positive rates as determined by clinical follow-up will produce misleading estimates of specificity. We will use random-effects models to take into account, at least to some extent, the between-study variation resulting from implicit threshold variation. Inconclusive US or MRI results will be regarded as negative as this approach gives the most conservative estimate of the sensitivity [47].

The reference standard for CDR and D-dimer will be clinical follow-up which may include imaging at presentation. This reference standard is able to confirm or reject a diagnosis of DVT after an initial risk stratification by D-dimer or CDR . For these index tests sensitivity and specificity will therefore be the main outcomes. If accuracy results are only reported as CDR in combination with D-dimer, then the two will be analysed as one index test.

All outcomes will be analysed and presented for pregnant patients and postpartum

patients separately. Data for each index test will be presented in a forest plot combined with a two by two contingency table. Where possible we will regard patients rather than lower extremities as unit of analysis. Only accuracy estimates of D-dimer and CDR will be plotted in an ROC space, as we will not estimate specificity for the imaging modalities. Different CDRs will be reported separately.

The authors anticipate that the search will yield only a limited number of eligible studies. Meta-analysis will be performed under certain conditions. Exact criteria cannot be pre-specified because of the multitude of factors involved and because we plan to perform multiple primary analyses, for each index test in the pregnant and the postpartum population. Factors that will be taken into account in the decision whether to perform meta-analysis include the following: the quality of the studies as rated with the QUADAS-2 tool, since pooling low quality studies can lead to misleading conclusions [48]; the heterogeneity in study design, study execution, technical aspects of the index test, the reference test and study populations; the number of studies; and the convergence and fit of the statistical model.

In case meta-analysis is appropriate, we will perform the following analyses. For CDR we will use a bivariate random-effects model to estimate a summary sensitivity and specificity with 95% confidence intervals (CI) and prediction regions. Different CDRs will not be pooled. If included studies use distinct D-dimer cut-off levels, we will use hierarchical summary receiver operator characteristics if more than two different cut-offs are used. If one or two cut-offs are used we will use a bivariate model for each cut-off. For US and MRI we cannot apply the optimal meta-analytical techniques for diagnostic studies as these require false positive rates. We will perform a meta-analysis according to the DerSimonian and Laird method [49] to obtain a summary estimate of the sensitivity of US and MRI. US and MRI will be compared by adding index test as a covariate to the model. If more than four studies performed both US and MRI in the same population, we will restrict this analysis to those studies.

In all cases we will use two univariate models, for sensitivity and specificity, instead of a bivariate model, if the number of studies is below five or if the bivariate model shows noncovergence or other signs of poor model fit [50]. We will not meta-analyse NPV as this parameter is heavily dependent on the prevalence, which will vary across studies [48]. We will follow guidance of the Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Chapter 10 on issues with model convergence [48].

We define prevalence as the sum of true positives and false negatives divided by the total number of patients with a conclusive index test result. The inconclusive rate will be the number of inconclusive index test results divided by the total number of index tests performed. Confidence intervals for both proportions will be based on

the Wilson score interval. If meta-analyses are deemed inappropriate we will present the parametric or non-parametric descriptive summary statistics for the main accuracy measures, as well as for prevalence and inconclusive rates, depending on the distribution of the data.

All analyses will be performed in R version 3.3.3 (cran.r-project.org) or Review Manager (RevMan) Version 5.3. Copenhagen: The Nordic Cochrane Centre, The Cochrane Collaboration, 2014.

Investigations of heterogeneity

All planned investigations of heterogeneity and sensitivity analyses are subject to the availability of adequate data. We will investigate heterogeneity informally by visually inspecting forest plots stratified by anticipated sources of heterogeneity. If test accuracy data are presented separately for each trimester, we will perform separate meta-analyses for each trimester for each index test. If only mean gestational age is available at study level we will use this as a covariate in the above described models. For US and MRI we will investigate prior testing as a source of heterogeneity. Any form of diagnostic testing before imaging will be modelled as a dichotomous variable at study level, except if only a proportion of the primary study population underwent screening. In this case we will use this proportion as a covariate. Technical aspects of the index tests will also be investigated using meta-regression.

Sensitivity analyses

We shall investigate the effect of quality of the primary studies on the outcomes by re-performing the main meta-analyses omitting studies scoring as unclear or high risk of bias. If accuracy data of US and MRI are extractable for different thrombus locations, we will determine summary estimates as described above for each separate location.

Acknowledgements

We thank Dr BA Hutten and Prof Dr AH Zwinderman for critically reading the analysis section.

References

1. James AH, Tapson VF, Goldhaber SZ. Thrombosis during pregnancy and the postpartum period. Am J Obs Gynecol 2005; 193: 216–9.

2. James AH, Jamison MG, Brancazio LR, et al. Venous thromboembolism during pregnancy and the postpartum period: Incidence, risk factors, and mortality. Am J Obstet Gynecol 2006; 194: 1311–5.

3. Pomp ER, Lenselink AM, Rosendaal FR, et al. Pregnancy, the postpartum period and prothrombotic defects: risk of venous thrombosis in the MEGA study. J Thromb Haemost 2008; 6: 632–7.

4. Heit JA. Trends in the Incidence of Venous Thromboembolism during Pregnancy or Postpartum: A 30-Year Population-Based Study. Ann Intern Med 2005; 143: 697.

5. Kane E V, Calderwood C, Dobbie R, et al. A population-based study of venous thrombosis in pregnancy in Scotland 1980-2005. Eur J Obstet Gynecol Reprod Biol 2013; 169: 223–9.

6. Kourlaba G, Relakis J, Kontodimas S, et al. A systematic review and meta-analysis of the epidemiology and burden of venous thromboembolism among pregnant women. Int J Gynaecol Obs 2016; 132: 4–10.

7. McColl MD, Ramsay JE, Tait RC, et al. Risk factors for pregnancy associated venous thromboembolism. Thromb Haemost 1997; 78: 1183–8.

8. Van Mens TE, Middeldorp S. Inherited thrombophilia and early pregnancy. In: Farquharson RG, Stephenson M., editors. Early pregnancy. 2nd ed. 2017. p. 234–44.

9. Meng K, Hu X, Peng X, et al. Incidence of venous thromboembolism during pregnancy and the puerperium: a systematic review and meta-analysis. J Matern Neonatal Med 2015; 28: 245–53.

10. Ficheur G, Caron A, Beuscart JB, et al. Case-crossover study to examine the change in postpartum risk of pulmonary embolism over time. BMC Pregnancy Childbirth 2017; 17: 119.

11. Jacobsen AF, Skjeldestad FE, Sandset PM. Ante- and postnatal risk factors of venous thrombosis: a hospital-based case-control study. J Thromb Haemost 2008; 6: 905–12.

12. Greer IA. Thrombosis in pregnancy: updates in diagnosis and management. Hematol Am Soc Hematol Educ Progr 2012; 2012: 203–7.

13. Chan W-S, Spencer FA, Ginsberg JS. Anatomic distribution of deep vein thrombosis in pregnancy. CMAJ 2010; 182: 657–60.

14. Wik HS, Jacobsen AF, Sandvik L, et al. Long-term impact of pregnancy-related venous thrombosis on quality-of-life, general health and functioning: results of a cross-sectional, case-control study. BMJ Open 2012; 2.

15. Wik HS, Jacobsen AF, Sandvik L, et al. Prevalence and predictors for post-thrombotic syndrome 3 to 16 years after pregnancy-related venous thrombosis: a population-based, cross-sectional, case-control study. J Thromb Haemost

2012; 10: 840–7.

16. Brandjes DP, Buller HR, Heijboer H, et al. Randomised trial of effect of compression stockings in patients with symptomatic proximal-vein thrombosis. Lancet 1997; 349: 759–62.

17. Kahn SR, Shrier I, Julian JA, et al. Determinants and time course of the postthrombotic syndrome after acute deep venous thrombosis. Ann Intern Med 2008; 149: 698–707.

18. Prandoni P, Villalta S, Bagatella P, et al. The clinical course of deep-vein thrombosis. Prospective long-term follow-up of 528 symptomatic patients. Haematologica 1997; 82: 423–8.

19. Prandoni P, Kahn SR. Post-thrombotic syndrome: prevalence, prognostication and need for progress. Br J Haematol 2009; 145: 286–95.

20. Bates SM, Greer IA, Middeldorp S, et al. VTE, thrombophilia, antithrombotic therapy, and pregnancy: Antithrombotic Therapy and Prevention of Thrombosis, 9th ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. Chest 2012; 141: e691S–e736S.

21. Chappell FM, Crawford F, Andras A, et al. Duplex ultrasound for the diagnosis of symptomatic deep vein thrombosis in the lower limb. [Protocol]. Crawford F, editor. Cochrane Database Syst Rev 2014; : CD010930.

22. Chappell FM, Andras A, Welch K, et al. D-Dimer tests for the diagnosis of deep venous thrombosis in symptomatic hospital outpatients with a clinical prediction rule [Protocol]. In: Crawford F, editor. Cochrane Database of Systematic Reviews. 2016. p. CD012356.

23. Nijkeuter M, Ginsberg JS, Huisman M V. Diagnosis of deep vein thrombosis and pulmonary embolism in pregnancy: a systematic review. J Thromb Haemost 2006; 4: 496–500.

24. Wells PS, Anderson DR, Bormanis J, et al. Value of assessment of pretest probability of deep-vein thrombosis in clinical management. Lancet 1997; 350: 1795–8.

25. Wells PS. The role of qualitative D-dimer assays, clinical probability, and noninvasive imaging tests for the diagnosis of deep vein thrombosis and pulmonary embolism. Semin Vasc Med 2005; 5: 340–50.

26. Di Nisio M, Squizzato A, Rutjes AWS, et al. Diagnostic accuracy of D-dimer test for exclusion of venous thromboembolism: a systematic review. J Thromb Haemost 2007; 5: 296–304.

27. Sampson FC, Goodacre S, Kelly AM, et al. How is deep vein thrombosis diagnosed and managed in UK and Australian emergency departments? Emerg Med J 2005; 22: 780–2.

28. Goodacre S, Sampson F, Thomas S, et al. Systematic review and meta-analysis of the diagnostic accuracy of ultrasonography for deep vein thrombosis. BMC Med Imaging 2005; 5: 6.

29. Ruehm SG, Wiesner W, Debatin JF. Pelvic and lower extremity veins: contrast-enhanced three-dimensional MR venography with a dedicated vascular coil-initial experience. Radiology 2000; 215: 421–7.

30. Fraser DG, Moody AR, Morgan PS, et al. Diagnosis of lower-limb deep venous thrombosis: a prospective blinded study of magnetic resonance direct thrombus imaging. Ann Intern Med 2002; 136: 89–98.

31. Kelly J, Hunt BJ, Moody A. Magnetic resonance direct thrombus imaging: a novel technique for imaging venous thromboemboli. Thromb Haemost 2003; 89: 773–82.

32. Moody AR. Magnetic resonance direct thrombus imaging. J Thromb Haemost 2003; 1: 1403–9.

33. Tan M, Mol GC, van Rooden CJ, et al. Magnetic resonance direct thrombus imaging differentiates acute recurrent ipsilateral deep vein thrombosis from residual thrombosis. Blood 2014; 124: 623–7.

34. Thomas SM, Goodacre SW, Sampson FC, et al. Diagnostic value of CT for deep vein thrombosis: results of a systematic review and meta-analysis. Clin Radiol 2008; 63: 299–304.

35. Bates SM, Jaeschke R, Stevens SM, et al. Diagnosis of DVT: Antithrombotic Therapy and Prevention of Thrombosis, 9th ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. Chest 2012; 141: e351S–e418S.

36. Khalafallah AA, Morse M, Al-Barzan A-M, et al. D-Dimer levels at different stages of pregnancy in Australian women: a single centre study using two different immunoturbidimetric assays. Thromb Res 2012; 130: e171-7.

37. Le Gal G, Kercret G, Ben Yahmed K, et al. Diagnostic value of single complete compression ultrasonography in pregnant and postpartum women with suspected deep vein thrombosis: prospective study. BMJ 2012; 344: e2635–e2635.

38. Khan F, Vaillancourt C, Bourjeily G. Diagnosis and management of deep vein thrombosis in pregnancy. BMJ 2017; 357: j2344.

39. Rutjes AW, Reitsma JB, Vandenbroucke JP, et al. Case-control and two-gate designs in diagnostic accuracy studies. Clin Chem 2005; 51: 1335–41.

40. Devillé WLJM, Bossuyt PMM, de Vet HCW, et al. [Systematic reviews in practice. X. Searching, selecting and the methodological assessment of diagnostic evaluation research]. Ned Tijdschr Geneeskd 2002; 146: 2281–4.

41. Geersing G-J, Bouwmeester W, Zuithoff P, et al. Search filters for finding prognostic and diagnostic prediction studies in Medline to enhance systematic reviews. PLoS One 2012; 7: e32844.

42. Beynon R, Leeflang MMG, McDonald S, et al. Search strategies to identify diagnostic accuracy studies in MEDLINE and EMBASE. Cochrane database Syst Rev 2013; : MR000022.

43. Greenhalgh T, Peacock R. Effectiveness and efficiency of search methods in systematic reviews of complex evidence: audit of primary sources. BMJ 2005; 331: 1064–5.

44. Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009; 6: e1000097.

45. Whiting PF, Rutjes AWS, Westwood ME, et al. QUADAS-2:

a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 2011; 155: 529–36.

46. Reitsma JB, Glas AS, Rutjes AW, et al. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. J Clin Epidemiol 2005; 58: 982–90.

47. van Mens TE, Scheres LJ, de Jong PG, et al. Imaging for the exclusion of pulmonary embolism in pregnancy. Cochrane Database Syst Rev 2017; 1: CD011053.

48. Macaskill P, Gatsonis C, Deeks JJ, Harbord RM TY. Chapter 10: Analysing and Presenting Results. In: Deeks J, Bossuyt P, Gatsonis C, editors. Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 1.0. 2010.

49. DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials 1986; 7: 177–88.

50. Takwoingi Y, Guo B, Riley RD, et al. Performance of methods for meta-analysis of diagnostic test accuracy with few studies or sparse data. Stat Methods Med Res 2015; .

Appendices

	Appendix 1. MEDLINE Search
1	exp Pregnancy/ or exp Pregnancy Trimesters/ or Pregnancy Complications, Cardiovascular/ or Postpartum Period/
2	(pregnan$ or prepartum or antepartum or trimester? or puerperium).ti,ab.
3	1 or 2
4	((Thrombosis/ or (thrombus or thrombo$ or thrombilic or thrombotic).ti,ab,kf.) and (exp Lower Extremity/ or (leg? or (lower adj extremit*)).ti,ab,kf.)) or Venous Thrombosis/ or (DVT or (deep adj (vein* or venous) adj thromb*)).ti,ab,kf.
5	3 and 4
6	exp ultrasonography, doppler/ or ultrasonography/ or exp ultrasonography, prenatal/ or ultrasonics/ or (ultrasound or ultrasonogra$ or ultrasonic$ or echograph$ or (doppler or duplex) or sonograph$ or sonogram$ or (contrast adj4 US) or echotomograp*).ti,ab,kf.
7	magnetic resonance imaging/ or magnetic resonance angiography/ or Diffusion Magnetic Resonance Imaging/ or ((magn* adj2 (resonance or imag*)) or MR or MRI or NMR or MRA or MR-venograph* or MRDTI).ti,ab,kf.
8	Fibrin Fibrinogen Degradation Products/ or Enzyme-Linked Immunosorbent Assay/ or "Nephelometry and Turbidimetry"/
9	(d-dimer or (fibrin adj2 d) or dimeri?ed plasmin or elisa? or elfa? or enzyme linked or latex agglutination or (latex adj3 assay?) or blood agglutination or Immunoturbidimetr$ or turbidimetr$ or SimpliRed or Minutex or NycoCard or "Instant I.A" or Vidas or LIATEST or ("IL test" or IL-DD) or Turbiquant or Asserachrom or Enzygnost or Fibrinostika or "BC DD" or (Tinaquant or Tina-quant) or TriniLIZE or biopool or TintElize or HemosIL or Innovance-DD or stratus or FDP or Dimertest or (LPIA or EIA)).ti,ab,kf.
10	exp "sensitivity and specificity"/ or exp "mass screening"/ or "reference values"/ or "false positive reactions"/ or "false negative reactions"/ or specificit$.tw. or screening.tw. or false positive$.tw. or false negative$.tw. or accuracy.tw. or predictive value$.tw. or reference value$.tw. or roc$.tw. or likelihood ratio$.tw.
11	Validat$.mp. or Predict$.ti. or Rule$.mp. or (Predict$ and (Outcome$ or Risk$ or Model$)).tw. or ((History or Variable$ or Criteria or Scor$ or Characteristic$ or Finding$ or Factor$) and (Predict$ or Model$ or Decision$ or Identif$ or Prognos$)).tw. or (Decision$.tw. and ((Model$ or Clinical$).tw. or logistic models/)) or (Prognostic and (History or Variable$ or Criteria or Scor$ or Characteristic$ or Finding$ or Factor$ or Model$)).tw. or ("Stratification" or "Discrimination" or "Discriminate" or "c-statistic" or "c statistic" or "Area under the curve" or "AUC" or "Calibration" or "Indices" or "Algorithm" or "Multivariable").tw.
12	6 or 7 or 8 or 9 or 10 or 11
13	5 and 12

Appendix 2. QUADAS-2.

Domain 1: Patient selection	
A. Risk of bias	
Describe methods of patient selection	
· Signalling Question 1: Was a consecutive or random sample of patients enrolled? · Rating criteria: v Yes: We will score 'yes' when it is stated that at least 95% of eligible patients were consecutively enrolled or that the patients are a random sample. v No: We will score 'no' if less than 95% of eligible patients were consecutively enrolled and if it the patients were not a random sample. We will also score 'no' if patients received one of multiple possible index tests where the choice of index test was not based on randomization. v Unclear: We will score 'unclear' in case this is not mentioned in the full text or appendix.	Yes/No/Unclear
· Signalling Question 2: Did the study avoid inappropriate exclusions? · Rating criteria: v Yes: We will score 'yes' when studies included symptomatic, pregnant or postpartum patients that were suspected of a DVT and underwent one of the selected index test(s), with or without prior testing, in a prospective fashion. v No: We will score 'no' when studies excluded pregnant patients in a specific trimester, in a retrospective fashion. v Unclear: We will score 'unclear' of the inclusion process was inadequately described.	Yes/No/Unclear
· Signalling Question 3: Did the study avoid inappropriate inclusions? · Rating criteria: v Yes: We will score 'yes' when studies included less than 10% of patients who have received therapeutic dose anticoagulants for longer than 24 hours prior to the index test. v No: We will score 'no' when studies included more than 10% of patients who have received therapeutic dose anticoagulants for longer than 24 hours prior to the index test, or when studies included more than 25% of patients with prior DVT. v Unclear: We will score 'unclear' of the inclusion process was inadequately described.	Yes/No/Unclear
Risk of bias question: Could the selection of patients have introduced bias? · Rating criteria: v Low: We will score 'low' when all signalling questions above are answered 'yes'. v High: We will score 'high' when at least one signalling question above are answered 'no'. v Unclear: We will score 'unclear' when in- or exclusion criteria and patient selection are not clearly mentioned.	RISK: LOW/HIGH/ UNCLEAR
B. Concerns regarding applicability	
Describe included patients (prior testing, intended use of index test, setting, gestational age, presentation of complaints)	
Applicability question: Is there concern that the included patients do not match the review question? · Rating criteria: v Low: We will score 'low' when the studied population meets the criteria of the participants, target condition and the intended use of index test(s) as described in this protocol. v High: When the study's population differs from the criteria of the participants, target conditions and the intended use of index test(s) as described in this protocol, we will score 'high'. We will score this item 'high' if more than 25% of the patients underwent prior testing. Specifically for D-dimer, we will score this item as 'high' when more than 10% of patients were on therapeutically dosed anticoagulants during D-dimer testing. v Unclear: We will score 'unclear' when the study population is not clearly described in the full text or appendix.	CONCERN: LOW/ HIGH/UNCLEAR
Domain 2: Index test(s) (if more than 1 index test was used, please complete for each test)	
A. Risk of bias	
Describe the index test(s) and how it was conducted and interpreted:	
· Signalling Question 1: Were the index test results interpreted without knowledge of the results of the reference standard? · Rating criteria: v Yes: We will score 'yes' when the index test interpreter was blind to the reference test result. In case clinical follow-up was the reference test, we will score 'yes'. v No: We will score 'no' when the index test interpreter was not blind to the reference test result. v Unclear: We will score 'unclear' in case this is not mentioned in the full text or appendix.	Yes/No/Unclear

· Signalling Question 2: If a threshold was used, was it pre-specified? · Rating criteria: v Yes: We will score 'yes' when a pre-specified threshold was used. v No: We will score 'no' if no threshold was used or when it was based on the results of the study. v Unclear: We will score 'unclear' in case it is not mentioned in the full text or appendix whether a prespecified threshold was used. v Not applicable: We will score 'not applicable' when studies using US or MRI as index test are evaluated.	Yes/No/Unclear/Not applicable
Risk of bias question: Could the conduct or interpretation of the index test have introduced bias? · Rating criteria : v Low: We will score 'low' when all signalling questions above are answered 'yes'. v High: We will score 'high' when at least one signalling question above are answered 'no'. v Unclear: We will score 'unclear' when the study does not provide sufficient information to conclude whether index tests were interpreted blind or whether pre-specified thresholds were used.	RISK: LOW/HIGH/ UNCLEAR
B. Concerns regarding applicability	
Applicability question: Is there concern that the index test, its conduct, or interpretation differ from the review question? · Rating criteria: v Low: We will score 'low' when the conduct and the interpretation of the index test(s) meet this review's description of the index test(s). v High: When the conduct and interpretation of the index test(s) differs from this review's description of the index test(s), we will score 'high'. We will also score 'high' if risk stratification by unstructured clinical judgement is investigated as an index test. v Unclear: We will score 'unclear' when the conduct and the interpretation of the index test(s) is not clearly described in the full text or appendix.	CONCERN: LOW/ HIGH/UNCLEAR
Domain 3: Reference standard	
A. Risk of bias	
Describe the reference standard and how it was conducted and interpreted:	
· Signalling Question 1: Is the reference standard likely to correctly classify the target condition? · Rating criteria: v Yes: We will score this item 'yes' if the clinical follow-up included at least a telephone interview, patient visit or admission at the end of the follow-up period and all patients with symptoms of either DVT or PE received objective diagnostic testing. In case of death, autopsy had to be performed. In case of a negative index test, the reference test could only interpreted correctly when no anticoagulant therapy was administered in the meantime. v No: We will score 'no' if clinical follow-up was based on hospital records or when objective testing or autopsy was not performed. v Unclear: We will score 'unclear' in case it is not mentioned in the full text or appendix how the reference test was assessed.	Yes/No/Unclear
· Signalling Question 2: Were the reference standard results interpreted without knowledge of the results of the index test? · Rating criteria: v Yes: We will score 'yes' when the reference test interpreter was blind to the index test result. v No: We will score 'no' when the reference test interpreter was not blind to the index test result. v Unclear: We will score 'unclear' in case this is not mentioned in the full text or appendix whether the reference test interpreter was blinded.	Yes/No/Unclear
Risk of bias question: Could the reference standard, its conduct, or its interpretation have introduced bias? · Rating criteria: v Low: We will score 'low' when all signalling questions above can be answered with 'yes'. v High: We will score 'high' when at least one signalling question above should be answered with 'no'. v Unclear: We will score 'unclear' when the conduct and the interpretation of the reference test(s) is not clearly described in the full text or appendix.	RISK: LOW/HIGH/ UNCLEAR
B. Concerns regarding applicability	
Applicability question: Is there concern that the target condition as defined by the reference standard does not match the review question? · Rating criteria: v Low: We will score 'low' when venography was performed as reference standard. v High: We will score 'high' when clinical follow-up was used as reference standard. v Unclear: We will score 'unclear' when the conduct and the interpretation of the reference test(s) is not clearly described in the full text or appendix.	CONCERN: LOW/ HIGH/UNCLEAR

Domain 4: Flow and timing	
A. Risk of bias	
Describe any patients who did not receive the index test(s) and/or reference standard or who were excluded from the 2x2 table (refer to flow diagram):	
Describe the time interval and any interventions between index test(s) and reference standard:	
· Signalling Question 1: Was there an appropriate interval between index test(s) and reference standard? Rating criteria: v Yes: We will score 'yes' if imaging was part of the reference test in all patients and it was performed within 24 hours after the index test. v No: We will score 'no' if clinical follow-up was used as the only reference test as it necessarily bares the risk of disease either occurring or resolving during the follow-up period. v Unclear: We will score 'unclear' in case it is not mentioned in the full text or appendix what the interval between index and reference test was.	Yes/No/Unclear
· Signalling Question 2: Did all patients receive a reference standard? · Rating criteria: v Yes: We will score 'yes' if more than 99% of patients underwent a reference test. v No: We will score 'no' if less than 99% of patients underwent a reference test. v Unclear: We will score 'unclear' in case it is not mentioned in the full text or appendix what percentage of patients received a reference test.	Yes/No/Unclear
· Signalling Question 3: Did patients receive the same reference standard? · Rating criteria: v Yes: We will score 'yes' if all patients received the same reference test, where clinical follow-up without imaging is not considered the same test as clinical follow-up with imaging. v No: We will score 'no' if not all patients received the same reference test. v Unclear: We will score 'unclear' in case it is not mentioned in the full text or appendix which patients received which reference test.	Yes/No/Unclear
· Signalling Question 4: Were all patients included in the analysis? · Rating criteria: v Yes: We will score 'yes' if more than 99% of all patients contribute to the 2x2 table. v No: We will score 'no' if less than 99% of all patients contribute to the 2x2 table. v Unclear: We will score 'unclear' in case it is not mentioned whether there were patients included that did not contribute to the 2x2 table.	Yes/No/Unclear
· Signalling Question 5 (only for index tests for which imaging was part of the reference test): Did patients not receive anticoagulant treatment between the index test and the reference standard? · Rating criteria: v Yes: We will score 'yes' if lessmore than 10% of patients received over 72 hours of therapeutically dosed anticoagulants before the imaging test. v No: We will score 'no' if moreless than 10% of patients received over 72 hours of therapeutically dosed anticoagulants before the imaging test. v Unclear: We will score 'unclear' in case it is not mentioned in the full text or appendix whether and for how long patients received anticoagulants.	Yes/No/Unclear
Risk of bias question: Could the patient flow have introduced bias? · Rating criteria: v Low: We will score 'low' when all signalling questions above are answered 'yes'. v High: We will score 'high' when at least one signalling question above are answered 'no'. v Unclear: We will score 'unclear' when concerns above are not clearly described in the full text or appendix.	RISK: LOW/HIGH/ UNCLEAR