Chapter 3: Results

Literature Search and Abstract Review

Systematic reviews

The literature search process identified 463 unique citations potentially relevant to key questions for which the EPC team evaluated *systematic reviews.* During the review of abstracts, 86 percent (399 articles) were found not to meet the criteria for inclusion. Articles were excluded for the following reasons: the article was not in English (62); the article did not include human data (3); the article was a review but did not include a systematic review, meta-analysis, or cost-effectiveness analysis (84); the article was not a review (49); reports primary data only (49); the article focused on prevention only (86); and the article did not apply to a key question designated to be addressed by systematic reviews (153). The total number of exclusions exceeded the number of articles reviewed because some articles were excluded for more than one reason.

Primary Literature

The literature search process identified 1786 unique citations potentially relevant to key questions for which the EPC team evaluated *primary literature*. During the review of abstracts, 92 percent (1638 articles) were found not to meet the criteria for inclusion. Abstracts were excluded for the following reasons: the article was not in English (99); the article did not include human data (18); the citation was a meeting abstract only (3); the study was limited to prevention of VTE (126); the article was a case report (26); the article contained no original data (354); the article did not apply to a key question designated to be addressed by review of primary literature (956) or all data in the article were presented elsewhere (2). For articles relating only to key questions 3 or 4, the EPC team excluded 18 studies that did not involve a comparison group or did not include a cost-effectiveness analysis. For articles relating only to key question 5, the team excluded studies that did not include a clinical prediction rule (i.e., at least two of history, physical exam, and/or laboratory testing, used together) (11) or did not specify a reference standard (1). For articles relating only to key question 7, the team excluded studies that did not report test characteristics of CT or MRI for the diagnosis of PE (3) or did not have an appropriate reference standard (21). The total exclusions exceeded the number of articles reviewed because some articles were excluded for more than one reason.

Articles Eligible for Review

Following the abstract review process, 63 reviews and 146 primary studies remained eligible. Of these, 31 reviews were tagged for key question 1 or 2 (LMWH for treatment of DVT or PE), 33 primary studies addressed key question 3 (efficacy and cost-effectiveness of outpatient treatment for DVT), 22 primary studies addressed key question 4 (duration of therapy), 61 primary studies pertained to key question 5 (use of clinical prediction rules), 16 reviews addressed key question 6 (ultrasonography for DVT diagnosis), 9 reviews and 30 primary studies pertained to key question 7 (helical CT or MRI/MRA for PE diagnosis), and 15 reviews addressed key question 8 (D-dimer for thromboembolism diagnosis). Added together, the total number of articles identified as pertaining to key questions exceeded the actual number of articles reviewed because some articles were identified as relevant for more than one key question.

Results of the Key Questions

Q1. What are the efficacy and safety of LMWH compared with UFH for the treatment of DVT?
Q2. What are the efficacy and safety of LMWH compared with UFH for treatment of PE?

Introduction

Because DVT and PE have similar underlying pathophysiology and often occur together, most of the published clinical trials evaluated the use of LMWH in patients with DVT with or without concomitant PE. Also, several systematic reviews of clinical trials have already been published about the efficacy and safety of LMWH for VTE. Therefore, for the purposes of this report, we combined questions 1 and 2 and searched the literature for systematic reviews that have evaluated the efficacy and safety of LMWH versus UFH in patients with VTE, emphasizing the quality and content of these reviews.

Results of Literature Search

Thirty-one articles were identified at article review for possible relevance to key questions 1 or 2. Of these, 17 were excluded: nine did not include a systematic review, one focused on prevention of VTE, three did not apply to any key question, three duplicates were found with different citations, and two did not discuss any relevant outcomes. The number of exclusions exceeded the number of articles reviewed as reviewers could indicate more than one reason for exclusion. After article review, 14 systematic reviews remained eligible for the review on key questions 1 and 2.[32-45]

Characteristics of Reviews

In Evidence Table 1 we have summarized the study aims, number of trials included, and quality scores for the 14 systematic reviews of clinical trials for Questions 1 and 2. The reviews were published between 1994 and 2000; nine included trials that enrolled patients with DVT or PE,[33,36-39,42-45] while five limited their review to trials of patients with DVT only.[32,34,35,40,41] No systematic review published to date has focused exclusively on patients with PE with or without concomitant DVT. The number of randomized controlled trials (RCTs) reviewed in each article varied substantially (mean 13, range 6 to 21) and was not related to either year of publication or whether the review included patients with VTE or those limited to DVT. There was little overlap among the trials included in the systematic reviews of this topic. The most recent reviews. those by van den Belt, et al., and van der Heijeden, et al., included many of the trials that were included in earlier reviews. Most of the systematic reviews included RCTs evaluating the efficacy of many different LMWHs, with the exception of one that focused solely on dalteparin.[36]

Quality of Reviews

The overall quality scores varied substantially (mean 58 percent, range 22 to 92 percent), with more recent studies tending to have higher scores (see Evidence Table 1). Most reviews adequately described the study aims, search strategy, and study inclusion criteria, and provided conclusions consistent with the results of their analyses. Fewer reviews adequately described their methods to pool data across the RCTs. Only four reviews included a formal assessment of the quality of the included RCTs.[33,41,42,44]

Results of Reviews

Evidence Table 2 describes patient populations and outcomes of trials included in the systematic reviews. A few articles limited their review to specific subpopulations of VTE (e.g., first episode of VTE[33] or first episode of DVT[40]). Several reviews analyzed data for all participants in the RCTs combined and then separately for patients with cancer.[33,34,38,41,44] The clinical outcomes most commonly compared between treatment groups were recurrence of VTE, major bleeding, and all-cause mortality. Most reviews reported recurrence of VTE and mortality data at three or six months after VTE diagnosis, although some also examined differences in outcomes at several earlier times (e.g., days 1 to 15, 16 to 90, 1 to 90[33] or during the period of heparin use[44]). Bleeding, however, was generally assessed during the initial period of heparin treatment (LMWH or UFH). A few reviews evaluated other outcomes as well, particularly thrombus extension,[32,34,35,40,43,44] minor bleeding,[33,41,42] and thrombocytopenia.[37,41,42] Four systematic reviews published in 1997[36] and 1998[37,39,40] were only descriptive and did not quantitatively pool results. The remaining 10 systematic reviews provided a summary measure of treatment effect based on a quantitative pooling of data from the RCTs.[32-35,38,41-45]

During the three or six months of followup in the RCTs, the rate of recurrence of VTE among RCT participants was approximately five percent. The systematic reviews relied on the definition of VTE recurrence used in the various RCTs. Of the 10 reviews that quantitatively examined the results of the various RCTs, four reported that LMWH significantly reduced the risk of recurrent thrombosis,[32-34,45] and six indicated a trend toward a protective effect with LMWH.[35,38,41-44] A review published in 1995[33] found that the benefit of LMWH in preventing recurrence of VTE occurred primarily during days 1 to 15 ;[33] a later review reported a similar magnitude of benefits extending up to six months after initiation of therapy.[44] Results of the descriptive reviews were discordant, indicating that LMWH was more effective,[39] that there was no difference between LMWH and UFH,[37,40] or that data were insufficient to answer the question.[36]

Of the six reviews that compared rates of thrombus extension in LMWH and UFH groups,[32,34,35,40,43,44] five reported that LMWH was superior to UFH,[32,34,35,43,44] and one (a descriptive review) suggested no difference.[40]

All reviews compared rates of major bleeding during the initial treatment period with heparin. Authors of the systematic reviews generally relied on the definition of major bleeding used in the various RCTs. The overall rate of major bleeding reported in the systematic reviews was approximately two percent. In eight of the 10 reviews that reported results from the quantitative pooling of the data, patients treated with LMWH had fewer episodes of major bleeding than those treated with UFH.[32-35,38,43-45] Gould et al. reported a significant benefit when using a fixed-effects model, but only a trend toward benefit when using a random-effects model;[41] the remaining review indicated a trend toward less bleeding with LMWH.[42] As with recurrence of VTE, the descriptive reviews either indicated that LMWH was more effective,[39] that there was a lack of difference between LMWH and UFH,[37,40] or that there were insufficient data.[36]

Eleven of the fourteen systematic reviews examined differences in rates of all-cause mortality in patients according to treatment assignment.[33-35,37,38,40-45] The systematic reviews reported a mortality rate of approximately five percent across the RCTs. All nine reviews that employed quantitative pooling for this outcome indicated that LMWH significantly reduced mortality during the three or six months of followup compared to UFH,[33-35,38,41-45] with one review indicating a similar benefit of LMWH in days 1 to 15 and days 16 to 90 after VTE diagnosis.[33] Two descriptive reviews suggested that mortality was no lower with LMWH than with UFH.[37,40] Five reviews[33,34,38,41,44] examined mortality in patients with cancer according to their treatment assignment. Two of these reviews[33,44] concluded that LMWH reduced mortality in patients with cancer, but not in patients without cancer.

In general, published clinical trials evaluating the efficacy of LMWH for VTE enrolled patients with DVT with or without concomitant PE. Only three published trials have been specifically designed to compare LMWH with UFH for patients with PE. These three trials include two smaller pilot studies (fraxiprine versus UFH, 101 patients;[46] (fragmin versus UFH, 60 patients[47]) and a large unblinded multicenter trial (tinzaparin versus UFH, 612 patients[48]) of patients without "massive" PE (i.e., were not in shock, did not receive thrombolytic therapy or embolectomy). One systematic review presented in this report included all three trials of patients with PE,[39] with five systematic reviews only including the tinzaparin versus UFH trial.[37,38,42,44,45]

Only three systematic reviews reported summary results for patients with PE, concluding that LMWH was as effective as UFH in this population.[36,38,44]

Since publication of these systematic reviews, data from a previously published double-blind double-placebo clinical trial of 432 patients with proximal DVT[49] were presented as part of re-analyses comparing LMWH (tinzaprin) versus UFH to patients who also had PE.[50] Perfusion lung scanning was performed on 97 percent of participants with proximal DVT at study entry. Investigators found evidence of PE in about 50 percent of participants (defined as high probability perfusion scans); about half of these patients were asymptomatic for PE. In this population with DVT and concomitant PE, patients assigned LMWH (N=97) were less likely than patients assigned UFH (N=103) to have a recurrence of VTE (0 versus 6.8 percent; 95 percent confidence interval (CI) for difference 1.9 to 11.7 percent) but had similar rates of major bleeding during heparin therapy (1.0 versus 1.9 percent; 95 percent CI for difference was -2.4 to 4.3 percent).[50]

Summary of Reviews

Compared to the five reviews published between 1994 and 1997, the nine reviews published more recently, from 1998 to 2000, tended to report smaller magnitudes of risk reduction from use of LMWH (recurrence of VTE: relative risk (RR) 0.7 to 0.8 versus 0.4 to 0.7; major bleeding: RR 0.6 to 0.7 versus 0.3 to 0.5; mortality: RR 0.7 to 0.8 versus 0.6 to 0.7). These differences could be due to variations in methodological quality, types of LMWH examined, and populations of included patients with VTE.

Overall, these data provided evidence that the efficacy (reduced rate of VTE recurrence, thrombus extension, and mortality) and safety (lower rates of major bleeding) of LMWH are superior to that of UFH for DVT (Evidence Grade: A). The evidence for treatment of submassive PE (with or without DVT) is more limited, but suggests that LMWH is likely to be as effective and safe as UFH (Evidence Grade: B).

Q3a. What are the efficacy and safety of outpatient versus inpatient treatment of DVT with LMWH or UFH?

Q3b. What is the cost-effectiveness of outpatient versus inpatient treatment of DVT with LMWH or UFH?

Introduction

In the first part of this document, we reviewed all published systematic reviews that evaluated the efficacy and safety of LMWH compared with UFH for the treatment of acute DVT. The evidence demonstrated that LMWH is at least as efficacious as UFH for the treatment of DVT, without an increase in major hemorrhagic complications. As with any new medication or technology, the costs associated with its use must be evaluated before it can be recommended for widespread use in a population.

Most of the trials described in these systematic reviews tested LMWH compared to UFH in an inpatient setting. As LMWH does not require intravenous administration, it may be used in an

outpatient setting or at home. If hospital stays are eliminated or shortened by the use of LMWH in place of UFH, the total costs of treatment can be expected to be less, despite higher medication costs. Furthermore, as partial thromboplastin times do not need to be monitored with the use of LMWH, the reduction in laboratory costs can be expected to reduce the total costs.

To better understand the efficacy and safety associated with use of LMWH in an outpatient setting and to address the cost implications of this practice, we reviewed the literature addressing the two study questions noted above.

Results of Literature Search

At article review, 14 articles were excluded from the 33 articles originally identified for possible relevance to key question 3. Of these, two contained no original data, six had no comparison group, one compared only two groups of outpatients, one presented data that were reported elsewhere, and four did not apply to any key question. After article review, 19 primary studies remained eligible for the review on key question 3 including ten on key question 3a and nine on key question 3b.[41,51-68]

Characteristics of Studies

Eight of the identified studies on key question 3a reported on the outcomes of patients with DVT treated with LMWH administered at home compared with outcomes of patients treated with UFH in the hospital[51-58] (see Evidence Table 3). Three of these were randomized trials,[51-53] while the others were cohort studies. An additional two studies compared clinical outcomes and costs for patients receiving LMWH at home to patients receiving LMWH administered in the hospital.[59,60] One of these studies enrolled only patients with PE.[60] We identified nine studies on key question 3b that were cost-effectiveness or cost-minimization studies.[41,61-68]

Outpatient versus Inpatient Therapy

The ten studies on key question 3a were published between 1996 and 2002 (see Evidence Tables 4, 5, and 6). Four of these were randomized controlled trials.[51-53,59] The smallest study enrolled 28 patients in each arm[57] and the largest was a retrospective cohort study with 1850 patients (164 of whom had received LMWH).[58] All of the trials used enoxaparin, nadroparin, or dalteparin during the intervention, and then an oral anticoagulant during the followup period. Enoxaparin was always used at a dosage of 1 mg/kg twice daily, but the dosage of nadroparin varied across studies.

In all of the studies, UFH was given in the hospital, except for one trial in which one group at home used UFH given subcutaneously.[51] In all studies, LMWH was administered at home or was completed at home after a brief in-patient admission. In two studies, however, outpatient LMWH was compared with LMWH administered as an inpatient treatment.[59,60] Among randomized trials, only one study required a visiting nurse to administer the medication.[59] In the trial by Koopman et al., only 15 percent of participants received help at home with drug administration. In the study by Levine et al., the patients administered the drug themselves,[53] and in the trial by Belcaro et al., patients received one home visit by a nurse for instruction and then self-

administered the drug.[51]

All studies excluded patients with PE except for the study by Kovacs et al. that exclusively enrolled patients with PE.[60] The exclusion criteria were fairly extensive; most studies excluded patients with known thrombophilic conditions, including prior VTE and patients unlikely to comply with outpatient therapy (see Evidence Table 3). Only three of the studies used scheduled, radiological surveillance procedures to detect recurrences.[51,53,59]

Quality of Studies on Outpatient versus Inpatient Therapy

Generally, the quality of the studies was not high. The studies were mostly complete in their description of the patient populations, but weaker in the description of the interventions (particularly regarding the UFH interventions) with little description of the adequacy of anticoagulation during the acute intervention or the followup period. Few studies adequately described whether other therapies, such as aspirin, were allowed or prohibited during the followup period (see Evidence Table 4).

Results of Studies on Outpatient versus Inpatient Therapy

The studies reported few differences in outcomes between study groups (see Evidence Table 6). Across studies, the percentages of recurrent DVT ranged from zero to nine percent. Only one study reported a significant difference between groups in the percentages of patients with recurrences.[58] The single study that enrolled patients with PE also found no difference in adverse event rates; unfortunately, it was a small study and underpowered for seeing a difference in these rates.[60]

The occurrence of PE was rare and not different between arms in any study. Similarly the incidence of major bleeding was very low (from zero to four percent) and not different between arms. The percentage of patients dying during followup ranged from zero to 11 percent, again with no difference between study arms.

The number of inpatient days was fewer in the study arms that used LMWH either entirely at home or after a brief inpatient stay than in the arms that used UFH in the hospital. Few studies reported the statistical significance of these differences. The duration of the hospitalization depended strongly on how the study was designed.

Five of these 10 studies reported on costs[51,54,57-59] (see Evidence Table 6). Although only two studies reported on the statistical significance of the difference in costs between the study arms,[54,59] it seems likely that this difference was also statistically significant in other studies. Huse et al. showed that outpatient costs with LMWH were higher, but stated that the anticipated savings of 2.5 hospital days in this group would save 1,911 U.S. dollars per patient.[58]

Cost-Effectiveness or Cost-Minimization Studies

Nine cost-effectiveness or cost-minimization studies were published between 1997 and 2000 (see Evidence Table 7). Four were designed as cost-effectiveness studies,[41,62,64,66] four were cost-minimization studies,[61,63,65,67] and one used a decision-model but could not be classified as either of the above.[68] A societal perspective was used in quantifying costs in two studies,[41,65] while the other seven took the perspective of a payer.

The modeled comparisons fell into two categories. Four of the studies modeled the use of LMWH compared with UFH, with all drugs administered in the hospital.[41,61,62,67] The other studies modeled the use of LMWH at home compared with UFH in the hospital.[63,65,66,68] Two of these modeled the use of LMWH in patients at home if they were medically eligible to be treated as outpatients, and in the hospital if they were not.[64,66]

The source of the estimates for costs used in the models varied (see Evidence Table 8). Half

of the studies used actual costs measured in the setting of a clinical trial. The others used costs obtained from databases of costs maintained by the government or payer, or used costs abstracted from review of the literature. Similarly, the rates of events included in the models came from actual data observed in trials or from the literature. For the models, two of the studies assumed, on the basis of earlier work, that the rates of recurrent thromboses and adverse events were equivalent for LMWH and UFH.[61,63]

Quality of Studies on Cost-Effectiveness or Cost-Minimization

The overall quality of the studies was good (see Evidence Table 7). According to the quality assessment instrument that we designed, the study quality score ranged from 67 percent to 100 percent. The two questions on which the studies performed worst concerned the adequacy of the sensitivity analysis and the description of the population to whom the results could be expected to apply. Thus, readers of these studies may have some difficulty generalizing the results.

Results of Studies on Cost-Effectiveness or Cost-Minimization

Of the four studies that compared inpatient LMWH treatment to inpatient UFH treatment, two were cost-minimization studies. One projected a 57 percent cost savings with use of nadroparin instead of UFH.[61] The other study found no difference in costs between enoxaparin and UFH. It concluded that, since these costs were accrued in the setting of a clinical trial, some of the laboratory tests were protocol-driven, thus raising the costs in the enoxaparin arm above what would be seen in usual practice[67] (see Evidence Table 9).

One of the cost-effectiveness studies addressing this comparison found that inpatient tinzaparin dominated the UFH arm, i.e. tinzaparin was less costly and more efficacious.[62] This study predicted an 11 percent cost savings with the use of tinzaparin in the hospital in place of UFH. The high-quality cost-effectiveness study by Gould et al. modeled the use of enoxaparin and UFH in the hospital and found that while enoxaparin treatment is more expensive, it can be considered cost-effective compared with UFH because of the gain in quality-adjusted life-years, i.e. gain in years of life adjusted for the quality of those years.[41] In a secondary analysis in which the outcomes modeled that some of the patients on enoxaparin were treated as outpatients, they found that if only eight percent were treated as outpatients, this treatment would be cost-saving.

Of the studies investigating outpatient LMWH treatment compared with inpatient UFH treatment, all found that use of LMWH in outpatients is less costly than hospitalization for UFH. The cost-effectiveness study by Estrada et al. found that use of LMWH at home for clinically stable patients and in the hospital for unstable patients, yields a 10 percent cost savings over use of UFH in the hospital for all patients.[66] The authors noted that the cost savings were largely due to savings on inpatient costs. Rodger et al. similarly found a cost savings of 23 percent when this same comparison was made.[64] The two cost-minimization studies found outpatient LMWH to yield a cost-savings of 57 percent[65] and 64 percent[63] compared with inpatient UFH. The final study by Tillman et al. provided little data on event rates in the UFH arm so that the results were harder to interpret.[68] However, the authors stated that there was a 60 percent cost savings with enoxaparin at home compared with UFH in the hospital, and indicated that this treatment would be cost-saving even if hospitalization costs were to decrease by 77 percent.

Summary of Studies

The randomized trials that compared treatment with LMWH, in outpatients or in inpatients with early discharge, to inpatient treatment with UFH did not demonstrate a difference in adverse outcomes between groups, and showed a major reduction in duration of hospitalization and associated costs. Similarly, the comparison between LMWH in the hospital or at home revealed no difference in outcomes, but did demonstrate a major savings in hospitalization costs. However, no study alone was adequately powered to detect small differences in rates of adverse events between groups. For example, the largest trial had only 12 percent power to detect a difference in the observed rates of recurrent DVT between groups.[53] The frequency of adverse events in all studies was small; a difference in outcomes between groups was not be demonstrated, however equivalency cannot be definitively claimed. Still, the direction of the results suggested that it is unlikely that LMWH at home will be found to be substantially less safe than UFH. The results also suggest a substantial savings in duration of hospitalization and a savings in costs. Overall, we concluded that outpatient treatment of DVT with LMWH is likely to be efficacious and safe (Evidence Grade: B). These studies primarily enrolled patients who were selected as being appropriate for outpatient therapy and the results may not be applicable to all patients presenting with VTE.

The cost-effectiveness studies were consistent in suggesting that LMWH is either cost-saving or cost-effective compared with UFH (Evidence Grade: B). This is the conclusion regardless of whether this drug is administered in the hospital or at home, although the cost savings should be greater if hospitalization can be avoided. Given the different units of benefit and years of the studies, it was difficult to compare the studies directly with one another, but the direction of the benefit was uniform across studies.

Q4. What is the optimal duration of treatment for DVT and PE in patients without known thrombophilic disorders and in patients with thrombophilic disorders?

Introduction

Immediate therapy of symptomatic VTE employs UFH, LMWH or thrombolytic therapy (in severe cases) followed by heparin to inhibit coagulation and promote initial clot lysis. Once therapeutic heparin anticoagulation is achieved, a vitamin K antagonist (warfarin, acenocoumarol, fluindione, etc.) is initiated with the goal of attaining a target INR of at least 2.0 with concomitant use of heparin for an additional four to five days. Longer periods of heparin therapy (ten days) may be appropriate for massive pulmonary emboli or iliofemoral thrombosis.[69] Initial therapy of symptomatic VTE with a vitamin K antagonist alone is associated with a significantly higher incidence of recurrent VTE within three months.[70]

Continuation of warfarin therapy beyond the initial period of heparin anticoagulation permits continued thrombus resolution and reduces the risk of recurrent thrombotic episodes. The

benefits of warfarin therapy must be weighed against the risk of hemorrhagic morbidity and mortality associated with anticoagulation. The risk to benefit ratio is influenced by variables such as the acuity and location of the clot, the intensity, stability and duration of anticoagulation, patient age, comorbidities, and both intrinsic and extrinsic predispositions to thrombus formation. Intrinsic predispositions include inherited and acquired thrombophilic disorders such as Factor V leiden and antiphospholipid antibodies. Extrinsic predispositions include surgery, trauma, and immobility. Since excessive or inadequate anticoagulation can each lead to adverse outcomes, it is important to evaluate of the evidence on the optimal duration of oral anticoagulation therapy for patients with VTE. To this end, we conducted a systematic review of the English language literature that assessed the duration of anticoagulation for VTE. For the purposes of this review, idiopathic VTE is considered to be thrombosis that occurs in the absence of an obvious intrinsic or extrinsic risk factor. Secondary VTE refers to thrombotic events that occur in association with one or more temporary or permanent risk factors.

Results of Literature Search

At article review, 10 articles were excluded from the 23 articles originally identified for possible relevance to key question 4. Of these, seven were not relevant to any key question, three contained no original data, and one had no comparison group. After article review, 13 primary studies remained eligible for the review on key question 4.

Characteristics of Studies

The 13 studies, published between 1972 and 2001, included a total of 4137 patients (range of patients per study: 80 to 897)[71-83] (see Evidence Table 10). Twelve were RCTs;[71-78,80-83] one was a retrospective cohort study.[79] Inclusion criteria varied considerably with recent studies more precisely specifying eligible study subjects.[71,75,80,82,83] Most of these studies excluded subjects at high risk for recurrent thrombosis (known thrombophilia or malignancy) or bleeding (malignancy, recent surgery or trauma).[71-73,75-77,79,80,82,83] Differences in exclusion criteria were common even among more recent studies.

Five studies focused exclusively on patients being treated for a first episode of thrombosis,[71,74,75,80,82] while one evaluated the treatment of patients following a second episode of VTE.[83] Three included patients with isolated calf vein thrombosis,[74,76,80] one of which focused exclusively on this population.[76]

Quality of Studies

Evidence Table 11 summarizes the quality assessment of these studies, with the earlier trials providing less information about the setting and participants' characteristics.[72-74,78,79,81] Recently designed studies were less likely to be at risk of having results affected by confounding and biases. In this regard, studies by Levine et al.[77] and Kearon et al.,[75] which employed placebo-controlled triple-blind designs, were particularly strong. Among older studies, the one by Petitti et al. may be especially vulnerable to bias because of the retrospective cohort design.[79] More complete and precise assessments of patient outcomes characterized the recently published

literature.[71,75,80,82,83]

Unlike the earlier trials, five recent studies used independently-adjudicated, well-defined radiological criteria for the diagnosis of VTE.[71,75,77,82,83] Older studies used several different coagulation assays to monitor the intensity of oral anticoagulation and failed to provide data on the time within the therapeutic range,[72,73,78] whereas more recent studies routinely used the INR and reported data on therapeutic intensity over time.[71,75,77,80,82,83] Statistical analyses were also of higher quality in later reports.[71,75,77,80,82,83] Precise characterizations of the study populations, therapeutic intensity and outcome definitions, as well as randomization, blinded outcome assessment, and appropriate statistical analysis distinguished the highest-quality studies.[71,75,77,82,83]

Results of Studies

The twelve randomized trials enrolled 3767 patients (range of patients per study: 80 to 897) with a mean age of 61.5 years (range of mean ages from 56 to 67.7 years); a mean of 56 percent of participants were men (range of mean percentages from 40 to 75 percent) (see Evidence Table 12).

As shown in Evidence Table 13, most early studies found no evidence of increased benefit with a longer duration of anticoagulation for VTE. This finding, however, was weakened by methodological limitations including small study populations, unblinded assessment of outcomes, and the absence of radiological confirmation of VTE.[72-74,78,79,81]

Recent studies clearly demonstrated that oral anticoagulation effectively prevents recurrent thromboembolism as long as patients remain on treatment.[71,75,77,82,83] Prolonged anticoagulation for patients with a first idiopathic VTE[75] or a second VTE[83] was associated with fewer VTE recurrences but at the expense of a trend toward more bleeding and no difference in survival. Consequently, since the incidence of recurrent VTE decreased as time elapsed from a thrombotic event (recurrence rate 2.1 percent per month between six weeks and six months[82] and 0.45 percent per month between six months and indefinite treatment[83]) while bleeding risk remained constant (two percent per year), the therapeutic benefit of continued anticoagulation may decline over time.

For patients with a first episode of idiopathic DVT, the rate of recurrent VTE after discontinuation of anticoagulation was similar for patients treated for three months (5.1 percent per patient-year) or 12 months (5.0 percent per patient-year).[71] In contrast, six weeks of oral anticoagulation for patients with a first episode of VTE in the absence of malignancy, pregnancy or known thrombophilia was associated with an initially increased rate of recurrence (2.1 percent per month during months 1.5 to 6) compared with patients treated for six months (0.1 percent per month during months 1.5 to 6). After six months, the VTE recurrence rates over the next 18 months were equivalent between treatment groups (0.4 percent per month in the 6 week group versus 0.5 percent per month in the 6 month group).[82]

Agnelli et al. found that the incidence of recurrent VTE within two years of stopping anticoagulation was similar among patients who received three months compared with 12 months of treatment for idiopathic DVT.[71] These studies suggest that at least 3 months of anticoagulation is required for patients with idiopathic DVT.[71,82]

For calf vein thrombosis, three months of oral anticoagulant therapy in addition to five days

of heparin was superior to five days of heparin alone,[76] but, in another study, six weeks was equivalent to three months of oral anticoagulation.[80]

Subgroup analysis among the more methodologically sound trials demonstrated that the presence of permanent risk factors for VTE increased the risk of recurrence[75,77,80,82] Patients with permanent risk factors for VTE may benefit from longer therapy.[75,82] Specific permanent risk factors identified in subgroup analyses included antiphospholipid antibody syndrome[75] and malignancy.[80] In contrast, the presence of Factor V Leiden and the prothrombin mutation did not increase the risk of recurrence.[75] However, a small number of patients in the latter study reduced the certainty of these subgroup analyses and larger prospective clinical trials are needed to validate the findings. Increasing the duration of anticoagulation from six weeks to six months significantly reduced the two-year incidence of recurrence among patients with: a) permanent risk factors, b) a proximal DVT or c) inadequate anticoagulation (INR adequately elevated less than 75 percent of the time).[82] Among patients with these risk factors, the incidence of recurrent VTE was very high during the first 10 weeks after discontinuation of anticoagulation in the six week group.[82]

Conversely, there was no evidence that patients with temporary risk factors benefitted from a longer duration of treatment. Schulman, et al. and Pinede, et al. found no difference in recurrence among VTE patients with temporary risk factors treated for shorter versus longer durations.[80,82] VTE patients with temporary risk factors are significantly less likely to have a recurrence than those with permanent risk factors.[77]

Summary of Studies

For a first episode of idiopathic DVT, the evidence demonstrated that at least three months of oral anticoagulation is optimal, meaning that this duration of therapy reduces the risk of recurrent VTE without an excessive increase in episodes of major bleeding[71,77] (Evidence Grade: B). For symptomatic calf vein thrombosis, six weeks appeared to be sufficient.[76,80] Although no randomized studies focused exclusively on patients with PE, the outcomes of patients with first VTE, including PE, indicated that six months of therapy is superior to six weeks.[82] Although one study suggested that three months may be sufficient,[80] the more persuasive data supported a longer treatment duration.[75] For patients with a first episode of VTE associated with a temporary risk factor, three months of therapy is probably sufficient.[77,80,82]

For patients with an objectively documented second episode of VTE, the evidence suggested that indefinite anticoagulation is highly efficacious, albeit associated with a steady 2 percent per year incidence of major bleeding.[83] Subgroups of patients at exceptionally high risk of recurrent VTE such as those with the antiphospholipid antibody syndrome are particularly likely to benefit from prolonged anticoagulation.[75] However, since the incidence of recurrent VTE appeared to decline over time while the incidence of major bleeding remained constant, indefinite anticoagulation may not benefit all subgroups of patients with a second episode of VTE (Evidence Grade: C).

Q5. How accurate are clinical prediction rules used for the diagnosis of DVT or PE?

Introduction

Optimal use of diagnostic tests requires an appreciation of the pretest probability of disease in a patient. The results of a diagnostic test are best interpreted with knowledge of this pretest probability to yield a posttest probability that the patient actually has the disease. A number of clinical prediction rules have been created to help clinicians estimate accurately the pretest likelihood of disease.

Some of the scoring systems used to generate pretest probabilities of DVT or PE may be accurate enough to serve as diagnostic tests by themselves. If this is so, this approach could eliminate more invasive or expensive testing. Examples are the use of the Ottawa ankle rules,[84] which have markedly reduced the use of radiography of injured ankles, and the use of "strep throat" prediction rules, which have safely reduced the use of throat culture and antibiotics.[85,86]

Thus, we evaluated clinical prediction rules that are used in the diagnosis of DVT or PE.

Results of Literature Search

At article review, 44 articles were excluded from the 63 articles originally identified for possible relevance to key question 5. Of these, 30 did not report on clinical prediction rules as defined by the EPC team (i.e., two of the three from history, physical exam, or laboratory testing), seven were retrospective studies, four contained no original data, two did not address any key question, and one focused on prevention of VTE. After article review, 19 primary studies remained eligible for the review on key question 5 (Evidence Tables 14 to 17).

Characteristics of Studies

The articles were stratified according to the event that the clinical prediction rule was predicting (Evidence Table 14). We identified 14 studies that prospectively evaluated clinical prediction rules for the diagnosis of DVT,[87-103] and five studies evaluating prediction rules for diagnosis of PE.[100,101,104-107] Of the 14 studies using clinical prediction rules for the diagnosis of DVT, 12 were studies in which the Wells prediction model was evaluated.[88] Of these 12 studies, only one included a comparison of the Wells model to other proposed models.[95]

The clinical prediction rules for the diagnosis of DVT were evaluated in a total of 5411 patients. Most of the studies were done in Canada and Europe with only two studies having been done in the United States. Fifty-eight percent of the studies reported that the patients had idiopathic DVT, and most of them excluded patients for whom there was a suspicion of a concomitant PE. Among studies, the mean age for the patients evaluated was between 54 and 68 years. Men accounted for 25 to 62 percent of the subjects in the studies. The most commonly reported risk factors for the development of DVT were surgery and immobilization; only a few patients in each study had a malignancy (5 to 17 percent).

The clinical prediction rules for the diagnosis of PE were evaluated in a total of 3284 patients.[101,104-107] All of the studies were done in Canada or Europe. Among studies, the reported mean age ranged from 51 to 64 years. The risk factors for the development of PE were not consistently reported.

Quality of Studies

We report on the quality of these studies in Evidence Table 15. The population was well described in most of the studies. The low scores in the bias and confounding sections were due to most of the studies not having two independent observers applying the clinical prediction rules to the study subjects, to an absence of blinding in interpretation of the reference test, or to an absence of independent observers interpreting the reference test.

The overall quality of the studies was fairly high and there were no major differences in quality between the studies evaluating clinical prediction rules for the diagnosis of DVT and for PE.

Results of Studies

The Wells model is a scoring system that allocates pretest probability as high, moderate, and low based on a score derived from risk factors and physical findings of DVT (see Table 1).[108] In the 12 studies in which the model was tested, patients who had a high pretest probability based on this model had a prevalence of DVT that ranged between 17 and 81 percent (Evidence Table 17). Those found to be at a moderate pretest probability had a prevalence of DVT between zero and 28 percent; the group with a low pretest probability had a prevalence of DVT between zero and 13 percent.

The negative predictive value is a useful summary statistic in this setting because it indicates what proportion of patients who have a *low* score will truly *not* have thrombosis. These patients may be able to forego further testing or, alternatively, the results of their subsequent radiological tests can be interpreted with this knowledge.

The negative predictive values across the studies evaluating DVT were high. If patients with either moderate or high scores were classified as having DVT, the median negative predictive value was 96 percent with a range from 81 percent to 100 percent. If only patients with the highest category of prediction scores were classified as having DVT, the median negative predictive value was slightly lower, 87 percent, with a range from 75 percent to 100 percent. With a higher cutoff score, a greater number of patients can potentially be spared further testing although there is more misclassification of patients as being free of DVT when they are not.

The positive predictive values were not high indicating that these rules were not as useful for definitively identifying patients who do have thrombosis. Even with a high cutoff score, the positive predictive values rarely exceeded 75 percent.

The Wells model for the prediction of DVT, across all studies, had an area under the ROC curve (AUC) that ranged from 0.74 to 0.90. This indicates that the model has a probability of 0.74 to 0.90 of correctly discriminating a random pair of patients in which one has DVT and one does not. An AUC of 0.50 means that a test has no discriminating ability.[109] For detection of proximal DVT, the AUCs ranged from 0.79 to 0.92, whereas for distal DVT, the AUCs ranged only from 0.65 to 0.79, thereby suggesting that the Wells model is more accurate for the diagnosis of proximal DVT than for distal DVT.

A number of studies tested the addition of a D-dimer assay to the Wells model for improving the performance of the model.[91,92,94,96-99,102] In the majority of these studies the area under the

ROC curve increased with addition of the D-dimer assay indicating better discrimination between patients with and without thrombosis. The predominant conclusion was that a D-dimer assay that is normal (low), in the setting of a low clinical probability of VTE, even further lowers the likelihood of thrombosis.

In the studies evaluating the clinical prediction rules for diagnosis of PE, the percentages of patients that had a PE in the high pretest probability group ranged from 38 to 78 percent, the percentages for the moderate pretest probability group ranged from 16 to 39 percent, and for low pretest probability, percentages ranged from 3 to 28 percent. The Wells model for the prediction of PE had negative predictive values ranging from 72 percent to 98 percent when a lower cutoff was used for classifying patients as having PE, and from 64 percent to 89 percent when a high score cutoff was used.[104-106] By comparison, the Prospective Investigation of Pulmonary Embolism Diagnosis (PIOPED) model had a negative predictive value of 81 percent when a lower cutoff was used, and 73 percent when a high cutoff was used.[105]

Other clinical prediction rules, besides the Wells model, had AUCs that ranged from 0.51 to 0.87; however, the models were each tested in only a single patient population.[87,92,95,105,107] The only direct comparison between the Wells model and any other prediction rule found that the Sant-Andre Hospital rule performed similarly to the Wells model, with negative predictive values of 89 percent for Sant-Andre and 90 percent for Wells when a low score cutoff was used for classifying patients having DVT, and 79 percent and 84 percent, respectively, when a higher cutoff was used.[95,105]

Summary of Studies

Studies were relatively consistent in showing that the Wells clinical prediction rule for diagnosing DVT is useful for generating an estimate of the probability that a patient has a DVT, identifying patients who have no more than a ten percent chance of having a DVT, and identifying patients with a high enough risk of DVT to warrant additional testing (Evidence Grade: B). The evidence indicated that the model is not sufficiently specific for ruling *in* the diagnosis of DVT without further radiological testing. The model performed best if the DVT was proximal, and addition of the D-dimer assay to the model improved the diagnostic performance. Other models performed similarly to the Wells model, but there were not enough data to make conclusive comparisons. The evidence also indicated that the Wells model for PE has less predictive value than the DVT model (Evidence Grade: C).

Q6a. What are the test characteristics of ultrasonography for diagnosis of DVT?
Q6b. Are calf vein thromboses adequately identified with ultrasound?

Introduction

Contrast venography is the test that serves as the reference standard for the diagnosis of DVT. It is, however, a procedure that is avoided when possible because of its invasiveness and the risk of complications including thrombosis, phlebitis, bleeding, and allergic reaction to the

contrast dye. A noninvasive and safe diagnostic test is ultrasonography. Many studies have been done to determine the sensitivity and specificity of ultrasonography for the diagnosis of DVT. In these studies, patients received both ultrasonography and the reference standard, and the resulting diagnoses were compared. We describe here the systematic reviews that have qualitatively and quantitatively summarized this primary literature.

Results of Literature Search

At article review, nine articles were excluded from the 16 articles originally identified for possible relevance to key question 6. Of these, six did not contain a systematic review, and three did not address any key question. After article review, seven systematic reviews remained eligible for the review on key question 6.

Characteristics of Studies

The reviews were published between 1989 and 2002 (see Evidence Table 18). All of the reviews included only studies that compared ultrasonography to venography.

Four of the reviews summarized studies aimed specifically at diagnosing proximal DVT[75,110-112] (see Evidence Table 19). One review included studies of calf vein thrombosis exclusively,[113] and one included studies of upper-extremity DVT diagnosis only.[114] Most reviews specified that the studies must have had a prospective design and enrolled consecutive patients meeting the study entry criteria.

Five reviews included only trials of symptomatic patients,[110-113,115] while the review by Wells et al. focused on studies of asymptomatic, post-operative patients.[112] One review included trials of asymptomatic and symptomatic patients and stratified the results.[18] Two studies stratified the studies into two levels based on study quality.[110,112] Level one studies were prospective and employed blinded interpretation of both diagnostic tests. Level 2 studies failed to meet all criteria for a level 1 designation. Another review carefully assessed study quality but did not stratify on that basis.[115]

Quality of Studies

The description of the search methods used to identify studies for inclusion were reasonably strong although no review contacted experts in the field to identify other studies for inclusion (see Evidence Table 18). Most reviews provided little detail about the included study populations, although it is possible that many of the primary studies provided little clinical information. Two of the reviews made no assessment of the quality of the included studies.[111,113] It was difficult to assess the quality of the methods of combination of the studies as there is no consensus about the ideal way to pool results from diagnostic testing studies. Several studies appropriately avoided a quantitative summary of the data (i.e., did not pool the sensitivities and specificities). Others pooled the data, but stratified it in some way to minimize heterogeneity between studies.

Results of Studies

As the reviews had different criteria for inclusion of trials, the included studies overlapped less than anticipated. The reviews with the most overlap were those by Kearon et al., Cogo et al., and White et al., reviews that focused on studies enrolling patients with symptoms of lower-extremity DVT.[75,110,111] The review by Becker included studies lacking prospective designs and many of these were not included in the later reviews.[115]

All of the reviews used a simple weighted average of the individual sensitivities and specificities to yield aggregate results (see Evidence Table 19). One review incorporated the heterogeneity between the studies in calculating the CI surrounding the estimates of sensitivity and specificity.[112] These authors also included a summary ROC curve for the included studies, which is a useful way to present these data. There is no consensus on the best methodology for combining results of diagnostic tests, and aggregate sensitivities and specificities may not adequately capture the heterogeneity of the included studies.

The reviews that focused on studies of patients with symptoms of lower-extremity DVT reported uniformly high sensitivity and specificity for ultrasonography. The level of ultrasound technology (i.e., use of compression, duplex or Doppler) did not influence the results greatly. In these included studies, the prevalence of DVT was high, roughly 40 to 60 percent, a finding that suggests the positive predictive value of an abnormal ultrasound will be very high. This suggests that the test is useful in a population of patients selected to have a high prevalence of disease (such as with suggestive clinical criteria).

Upper-extremity DVT, even if symptomatic, was often missed with ultrasound alone, although the highest quality study included in the review had a sensitivity of 100 percent and a specificity of 93 percent.[114] The studies included in this review had an extremely high prevalence of upper extremity DVT, thus making the positive predictive value of this test fairly high despite a low sensitivity and specificity.

For diagnosing VTE in asymptomatic patients, ultrasonography retained its high specificity, but its sensitivity was markedly reduced, as shown in two reviews.[18,112]

For diagnosing calf vein thrombosis, three reviews found that ultrasound had low sensitivity in both asymptomatic and symptomatic patients.[18,111,112] One review found fairly high sensitivity for diagnosing calf vein thrombosis among the studies that were included,[113] although the authors noted many indeterminate test results throughout the included studies. The uncertain clinical significance of calf vein thrombosis was not addressed in these systematic reviews.

Looking only at the primary literature as defined by the reviews' authors, ultrasonography for diagnosing proximal DVT in symptomatic patients was sensitive and very specific. In these studies, doppler and color doppler capability offered no important advantage over compression ultrasound alone in diagnosing proximal DVT. In trials of asymptomatic patients, the performance characteristics of ultrasonography were fairly low in the high quality primary studies.

Summary of Studies

We conclude that the evidence was consistent in showing that ultrasonography has relatively high sensitivity and specificity for diagnosis of proximal lower extremity DVT in symptomatic patients (Evidence Grade A). However, with a false negative rate ranging from 0 to 6 percent, a negative ultrasound cannot absolutely exclude disease. The evidence indicated that ultrasound has considerably less utility for diagnosing DVT in asymptomatic patients, such as in a post-operative screening setting. The studies in which screening asymptomatic patients seemed promising were mostly of lower quality than those in which it was less useful.

The evidence was somewhat inconsistent, but suggested that ultrasound had relatively low sensitivity and specificity for diagnosing upper-extremity DVT (Evidence Grade: C). The identification of one successful high quality study suggests that this topic needs further study. Additionally, a high quality primary study was recently published. This recent study suggested that upper extremity DVT can be diagnosed with ultrasound with acceptable accuracy if the ultrasound examination shows venous incompressibility.[116]

The evidence suggested that ultrasound has poor sensitivity for the diagnosis of calf vein thrombosis. The need for diagnosis of calf vein thrombosis was not addressed by these reviews and is a separate issue (Evidence Grade: B).

Q7a. What are the test characteristics of helical CT for diagnosis of PE relative to V/Q scanning and/or standard angiography? Q7b. What are the test characteristics of MRI and MRA for diagnosis of PE relative to V/Q scanning and/or standard angiography?

Introduction

Imaging is an important component in the diagnostic evaluation of patients who are suspected of having PE (see Evidence Table 20). V/Q scintigraphy is widely used in the initial evaluation for PE, but the usefulness of this test is limited by a substantial proportion of indeterminate exams and the possibility that PE may be present despite a low probability scan. By contrast, pulmonary arteriography is highly accurate in the diagnosis of PE, but it is accompanied by the risks and discomfort associated with an angiographic procedure.

Examination of the pulmonary arteries with contrast-enhanced CT was made possible by the introduction of high-speed helical CT scanners in the early 1990s.[117] The advantages of helical CT include rapid exam times, high availability in emergent clinical settings, non-invasiveness, and relatively low cost. Helical CT scanners have since become widely available, and examination of the pulmonary arteries by helical CT has become a routine practice.[118] Given the high reported accuracy, it is reasonable to consider whether helical CT can replace traditional imaging modalities for detecting PE, namely, V/Q scan and pulmonary arteriography by catheterization. More recently MRI/MRA has been studied for diagnosis of PE. Its benefits include the ability to avoid the use of iodinated contrast material, and faster scanning sequences that have enabled imaging to be done more quickly than older techniques (see Table 2).

This key question was addressed in two parts. In part one, we examined all published systematic reviews of the use of helical CT or MRI/MRA for the diagnosis of PE. In part two, we examined original studies reporting the sensitivity and specificity of helical CT for the diagnosis

of PE compared to pulmonary arteriography, and the sensitivity and specificity of MRI/MRA for the diagnosis of PE.

Results of Literature Search

At article review, four reviews and 15 primary studies were excluded from the ten reviews and 30 primary studies originally identified for possible relevance to key question 7. The reviews were excluded for not being systematic reviews. For the primary studies, seven did not use a diagnostic testing study design, five did not address any key question, two contained no original data, and two did not use an appropriate reference standard. The total number of reasons for exclusion may exceed the number reviewed as reviewers may indicate more than one reason for exclusion. After article review, six systematic reviews and 15 primary studies remained eligible for the review on key question 7 (eight primary studies for key question 7a and seven for key question 7b).

Part One: Examination of Systematic ReviewsCharacteristics of Studies.

Six systematic reviews have examined the use of helical CT for the diagnosis of PE (see Evidence Table 20).[93,119-123] The most recent systematic review included the literature published before December 2000.[123] A major difference in these systematic reviews was the reference standard against which CT was compared. Two of the reviews[120] examined only studies in which the reference standard was pulmonary arteriography.[119,120] Two reviews defined the reference standard as either pulmonary arteriography or V/Q scan.[122,123] The remaining two reviews did not limit the reference standard to specific imaging modalities.[93,121] Two of the reviews included an article evaluating contrast-enhanced electron beam CT.[119,120] No systematic review addressed the use of MRI/MRA for diagnosis of PE.

Quality of Studies. Evidence Table 21 summarizes our assessment of the quality of the systematic reviews. Except for one review,[122] the quality scores for the reviews had a range from 72 to 78 percent. The articles with the lowest quality evaluation scored lowest in all categories, indicating no single area of weakness.[122,123] Among these systematic reviews, description of search methods received the lowest quality scores, whereas statements of study aims and conclusions received the highest quality scores.

Results of Studies. The findings of the systematic reviews are shown in Evidence Table 22. All of the reviews reported the sensitivity and specificity of helical CT for diagnosing PE as a main index of test performance. In five of the reviews, the sensitivities and specificities of each reviewed study were averaged, weighted according to each study's sample size. The combined sensitivities of CT across reviews ranged from 66 percent to 93 percent, and the combined specificities of CT ranged from 89 percent to 97 percent. In one of the reviews, combined sensitivity and specificity were not reported because the authors felt that the heterogeneity of included studies did not allow mathematical combination.[93] In that review, sensitivity was reported as a range from 53 percent to 100 percent, and specificity was reported as a range from 81 percent to 100 percent.

Part Two: Examination of Primary Studies

Our examination of the published systematic reviews was supplemented by a review of the

primary literature. Our initial aim was to update our analysis of the systematic reviews with pertinent studies published after completion of the systematic reviews. However, because of the wide variation in sensitivities reported by the systematic reviews, we felt a more meaningful approach would be to focus on the strongest evidence, instead of focusing only on the most recent. Therefore, we completed our primary literature review on all prospective studies evaluating helical CT for the diagnosis of PE in which all participants received the optimal reference test to confirm the diagnosis. We excluded studies evaluating electron beam CT because this technology is not routinely available. Our review of the primary studies on MRI/MRA also included all prospective studies that evaluated this modality against an acceptable reference test (pulmonary angiography or V/Q scan).

Characteristics of Studies. Evidence Table 23 summarizes key aspects of the eight eligible studies of CT, which were published between 1994 and 2001.[117,124-130] All studies were diagnostic test evaluations in which all participants received the diagnostic test and the reference test. None were multi-center studies, and none of the reports stated the specific dates of participant recruitment. Although some of the studies were included in the systematic reviews in Part One, none of the systematic reviews reviewed all of the studies selected for our primary literature review.

One study employed dual-detector helical CT, a faster form of helical CT.[128] All of the other studies employed conventional single-detector helical CT, and all studies used pulmonary arteriography as the reference standard. Only one study used explicit clinical findings to define the suspicion of PE.[130] In six of the studies, clinical suspicion of PE was implied as all participants in these studies were referred for imaging.[117,124,126-129] In one study, it was unclear if patients were enrolled because of referral for imaging or because of symptomatology.[125]

We identified seven studies of MRI /MRA for diagnosis of PE; the earliest was published in 1993. Five of these studies used MRA,[131-135] while the other two used perfusion MRI techniques.[136,137] The five MRA studies enrolled consecutive patients with suspicion of PE and required pulmonary angiography as the reference test. One MRI study enrolled nonconsecutive patients with suspected PE referred for either V/Q or angiography.[137] Finally, one study of MRI evaluated two groups of patients for perfusion defects due to either PE or severe emphysema.[136]

Quality of Studies. The study quality scores are given in Evidence Table 24. For the eight studies of CT, the scores ranged from 44 percent to 84 percent. The CT study with the lowest quality score was a brief report describing a study of 10 patients in whom massive PE was clinically suspected.[124] The study with the second lowest quality score was similarly a brief report, and the low scores may be related to the brief format.[125] The two categories with the lowest average quality scores across the eight studies of CT were for the descriptions of the included patients, and for the potential for bias and confounding in the study.

The five MRA studies were of similar and reasonably high quality. Their weakness as a group was incomplete description of the study population and key patient characteristics. The MRI perfusion studies were of lower quality than the MRA studies. Berthezene et al. described two series of patients with suspected perfusion defects, but did not describe the patient populations very well.[136] Erdman et al. enrolled nonconsecutive patients and allowed different reference tests.[137] All MRA studies used some form of blinding during the interpretation of the MRA examinations.

Results of Studies. The eight studies of CT reported data on a total of 443 individuals with the prevalence of PE ranging from 27 percent to 70 percent. The basic population characteristics for each of the studies are given in Evidence Table 23. The results of each study are summarized in Evidence Table 24. The reported sensitivity of CT ranged from 45 percent to 100 percent, and the reported specificity ranged from 78 percent to 100 percent. The only study reporting a sensitivity of 100 percent was the one that enrolled patients with clinically suspected massive PE, which was also the study with the highest prevalence of PE.[124]

The variability in sensitivity was greater than the variability in specificity, a fact we also noted in the prior systematic reviews. This variability in sensitivity was present in our primary literature review even though it had more stringent study inclusion criteria than did the earlier systematic reviews (i.e., we required that all patients in a study undergo both the diagnostic test and the reference test). This observation suggests that study design may not be an important contributor to the variations in sensitivity and specificity.

To summarize the CT studies graphically, a representative sensitivity and specificity for each study is plotted in Figure 2. We specified that the sensitivity/specificity pair be calculated using data from all the participants in the study and using the cutoff that yielded the best test performance (if several cutoffs were studied). The greater variability in sensitivity relative to the variability in specificity is also apparent in Figure 2. In Figure 3 we examined the relationship between prevalence of PE and the reported sensitivity and specificity. There is no apparent relation between prevalence and test performance. Therefore, the variability in reported sensitivities and specificities did not appear to be related to disease prevalence. However, the variability in disease prevalence is expected to strongly influence the reported positive and negative predictive values.

When the representative sensitivity/specificity pairs from the eight studies were pooled using simple addition, the sensitivity of CT was 86 percent (95 percent CI 80 to 90 percent) and the specificity was 92 percent (95 percent CI 88 to 95 percent). However, such pooling assumes that the studies were similar enough to be pooled, (i.e., each study is assumed to have the same underlying sensitivity and specificity so that random variation is the only source of variance between the results of different investigations). Figure 2 suggests that two of the studies are outliers having sources of variance outside of random variation.[126,130] The study by Velmahos et al. reported the lowest sensitivity and specificity, but theirs was also the only study in which all participants came from a specific clinical setting (a surgical intensive care unit).[130] Therefore, interpretation of the pooled sensitivity and specificity for the reviewed studies must be done with caution because of potential underlying heterogeneity.

Two of the studies suggested that the relatively low sensitivity may be related to whether CT interpretation included the finding of subsegmental clots that were seen on the reference tests. Velmahos et al. included interpretation of subsegmental clot, and their study was associated with the lowest sensitivity of all of the studies reviewed.[130] In the study by Goodman et al., inclusion of subsegmental clot lowered the sensitivity from 86 percent to 64 percent.[127] However, the study by Qanadli et al. differed from this pattern because it reported relatively high sensitivity and specificity despite the inclusion of subsegmental clot.[128] Therefore, in the studies reviewed, there did not appear to be a definite relation between test accuracy and vessel level interpreted.

The sensitivity of helical CT found in our examination of both the primary literature and systematic reviews is generally higher than was found in a recent large study of outpatients,

which reported a sensitivity of 70 percent and a specificity of 91 percent.[138] The latter study incorporated other imaging modalities as well as clinical followup to establish the diagnosis of PE rather than pulmonary arteriography alone, and this difference in study design may at least partially explain the lower sensitivity compared to the literature we reviewed.

The MRA studies demonstrated fairly consistent specificities. Sensitivities ranged across studies from 77 percent to 100 percent. The prevalence of PE across studies ranged from 27 percent to 55 percent. Berthezene et al., who presented aggregate data from two populations of patients (those with suspected PE and those with emphysema), found that sensitivity for picking up perfusion defects was low.[136] Erdman et al. found fairly high sensitivity and specificity and included an analysis of a subgroup of patients with pulmonary angiography as the reference test.[137] In this subpopulation, sensitivity was similar to that observed in other MRA studies; specificity, however, was lower.

Interpretation of our examination of the primary literature should be made with the knowledge of some important limitations in the evidence. First, participants in all but one of the studies [130] were enrolled because of suspicion of PE that led to referral for imaging. This introduced a potential selection bias in the study populations because nothing is known about individuals in whom PE was suspected but who were not referred for imaging. The real effect of this potential selection bias was difficult to determine from the data, however. Individuals referred for imaging may have been selected because of clinically obvious (rather than occult) disease and perhaps have a form of disease that is easier to detect by imaging than the typical case (inflating sensitivity and specificity), as exemplified by the one study in our review that included only patients suspected of having massive PE.[124] On the other hand, referring physicians may have referred only clinically difficult cases which could have more subtle imaging findings than clinically obvious cases.

There is also obvious heterogeneity in the prevalence of PE in the published studies. While disease prevalence strongly influences the positive and negative predictive values of a test, it classically should not affect the sensitivity and specificity of a test. However, if the variation in prevalence is indicative of a variation in disease spectrum or severity, then sensitivity and specficity may be affected. This principle is exemplified by the study of patients suspected of having massive PE.[124]

Summary of Studies

In our examination of both systematic reviews and primary studies, we found a moderate amount of variation in reported sensitivity of helical CT for the diagnosis of PE, ranging from 45 to 100 percent; reported specificity was generally greater than 90 percent with less variability (Evidence Grade: B). Pooled estimates of sensitivity and specificity of helical CT reported by systematic literature reviews should be interpreted with caution due to potential selection bias and heterogeneity in the reviewed studies. The source of the variability in sensitivity was unclear and was not completely explained by differences in study design, prevalence of PE, or smallest arterial level (segmental or subsegmental) interpreted by the radiologists. Potential sources of variability that could not be systematically evaluated from the literature included variations in scanning protocols, timing of contrast injection, scanner technology, and experience of radiologists.

Our review of the evidence also indicated that MRA is sensitive and specific in detecting acute PE of the lobar and segmental branches of pulmonary arteries in patients whose clinical presentation suggests PE (Evidence Grade: B). The accuracy of detecting smaller emboli was reduced substantially as one moves distal to the lobar segment of the arteries.

Q8. What are the test characteristics of D-dimer for diagnosis of VTE?

Introduction

The diagnosis of VTE employs clinical assessment followed by objective testing. Most of the available non-invasive diagnostic tests are radiological procedures that require expensive equipment, technicians, and radiologists for their performance and interpretation. These tests, are costly, time-consuming, and burdensome to patients.

A blood test that is both highly sensitive and specific for the diagnosis of VTE would be ideal. The test that has been most studied for this purpose is the D-dimer assay. D-dimers are fragments of cross-linked fibrin that are generated by fibrinolysis. Thus, elevated D-dimer levels indicate that clot formation and lysis have occurred. Many qualitative and quantitative D-dimer assays are available. Qualitative assays generally rely on the agglutination of latex particles or red cells coated with monoclonal antibodies to detect D-dimers in patient samples. Quantitative assays typically employ enzyme-linked immunosorbent assay (ELISA) to measure precisely the amount of D-dimer present in plasma.[80,139,140]

Over 70 articles in the primary literature have evaluated the characteristics of different D-dimer assays in various patient populations using different criteria for positivity. We sought to determine the usefulness of these assays in the diagnosis of VTE by reviewing systematic reviews of this primary data.

Results of Literature Search

At article review, 13 articles were excluded from the 15 articles originally identified for possible relevance to key question 8. Of these, 11 were not systematic reviews, and two did not apply to any key question. After article review, two systematic reviews remained eligible for the review on key question 8.

Characteristics of Studies

Of the eligible two reviews, the study by Kraaijenhagen et al. addressed multiple questions regarding the diagnosis of VTE, one of which was the role of D-dimer in patients with normal ultrasound exams.[141] The study by Becker et al. evaluated 29 published primary studies and presented detailed characteristics of the various D-dimer assays and their accuracies.[142] There was no overlap in the primary literature included in the two reviews.

Quality of Studies

Both reviews cleared stated the purpose of their study.[141,142] Pertinent English-language literature was identified by electronic and hand searches in both reviews. In the Kraaijenhagen et al. review, this search was supplemented by a query of experts in the field.[141] Inclusion criteria were reported in sufficient detail to allow replication in that review.[141] A validated instrument to assess study quality was used in the Becker review;[142] no instrument was reported in the other.[141] Reproducibility of quality assessments was not reported. Kraaijenhagen et al. pooled their selected studies and found no evidence of significant heterogeneity. Becker et al. found that the heterogeneity among the selected studies precluded pooling. The conclusions of both reviews were supported by the reported analysis. Based on these criteria for assessing the quality of systematic reviews, we assigned a quality score of 71 percent to the review by Kraaijenhagen et al. and 38 percent to the review by Becker et al.

Results of Studies

The two systematic reviews that we evaluated were methodologically very different. As part of a more extensive review, the authors of the review by Kraaijenhagen et al. focused upon two specific clinical questions; the utility of the D-dimer assay in patients with suspected DVT and a normal initial compression ultrasound result, and the utility of the D-dimer assay in patients evaluated with impedance plethysmography (IPG) and a clinical prediction rule.[141] The assays used and the thresholds for defining abnormal results were not reported. Of a total of 1128 patients with normal ultrasounds pooled from two of the primary studies identified by Kraaijenhagen et al., 250 had an abnormal D-dimer result and underwent a second ultrasound at one week. Two-hundred thirty-four patients had normal serial ultrasounds, but 4 (1.7 percent) of these patients developed non-fatal VTE during three months of followup. Only one fatal PE occurred (0.4 percent). Of the 878 patients with a normal initial ultrasound and normal D-dimer result only two (0.2 percent) went on to develop VTE during the three-month followup period. The overall VTE complication rate for this strategy was only 0.6 percent. Only patients with abnormal D-dimer assays had the followup ultrasonography mandated, introducing the likelihood of ascertainment bias, which could make the D-dimer test appear to be more predictive than it really is.

To further discuss the content of the Kraaijenhagen et al. review, we describe the included studies briefly. One of the primary studies, included in the review by Kraaijenhagen et al., evaluated the utility of D-dimer assays in patients evaluated with IPG after application of a clinical prediction rule.[143] Of 401 patients with clinically suspected DVT, 352 had a normal IPG. Seventy-six of these 352 had an abnormal D-dimer and venography confirmed a DVT in one-third of these patients. Of the remaining 276 patients with normal D-dimer levels, 177 patients with low clinical likelihood of DVT were followed without treatment for three months. Only one of these patients developed a VTE. Another patient, with a normal IPG and D-dimer result but a high clinical likelihood of thrombosis developed a DVT during followup. Therefore, the total VTE complication rate for this strategy was low. Again, ascertainment bias was possible because not all patients had clinical followup.

The systematic review by Becker et al., included 29 studies evaluating the test characteristics of D-dimer measurements (12 for diagnosis of DVT, 13 for diagnosis of PE, and four for either).[142] Thirteen of these studies were identified by the review's authors as being of high

quality. These studies employed a reference test, described the patient selection process, and studied test subjects representative of patents with suspected VTE. Marked heterogeneity was present among the studies and, appropriately, the results were not pooled. The authors plotted the studies' true positive and false positive rates on a summary ROC curve, a useful way to summarize this information. The authors identified, on the plot, the cutoffs used to define an abnormal test for each study. They identified at least 10 different cutoffs in these 29 studies.

As expected, the plots showed clearly that the ELISA studies that used very high D-dimer cutoffs (1000 ng/mL or 2000 ng/mL) had low sensitivity (five percent to 90 percent) and higher specificity (50 percent to 99 percent) for identifying patients with VTE. Studies using very low cutoffs (100 ng/mL or 200 ng/ml) had much higher sensitivity (75 percent to 100 percent) and lower specificity (one percent to 70 percent). A similar pattern was seen with the latex agglutination studies, with the summary ROC curve having a similar shape to that generated from the ELISA quantitative studies.

The authors noted that the major determinants of the specificity of D-dimer tests were the type of assay, the cutoff value, and the spectrum of clinical characteristics of enrolled patients free of thromboembolic disease. Overall, specificities were higher for outpatients than for inpatients, and for patients without co-morbidities, for both ELISA and agglutination assays. The authors concluded that D-dimer assays could not yet be used as a diagnostic test for VTE and recommended that further research be done with attention to the clinical spectrum of the patients, the duration of symptoms, the clinical setting, the age, and the comorbidities of the patients.

Summary of Studies

The systematic reviews reported widely varying estimates for sensitivity and specificity for D-dimer in the diagnosis of DVT. The specificities were generally higher than the sensitivities, particularly for outpatients and patients without comorbid diseases. This being so, D-dimer may eventually prove to have a role in risk stratification of patients, particularly when used with clinical prediction rules. However the evidence to date was not strong enough to allow us to draw definitive conclusions (Evidence Grade: C).

Figure 2: Plot of the representative sensitivity of helical computerized tomography for the diagnosis of pulmonary embolism versus one hundred minus the representative specificity reported in the eight studies in the primary literature review.

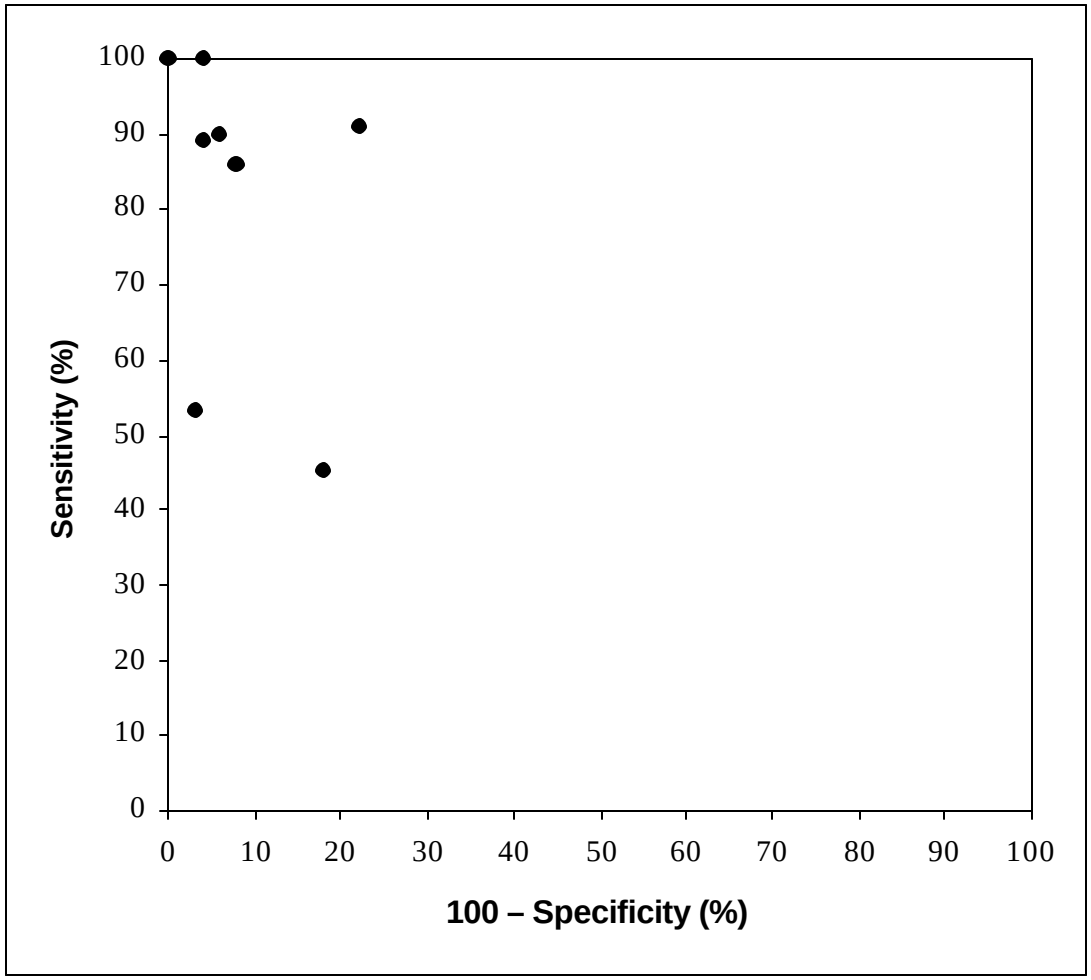

Figure 3: Plot of the

representative sensitivity and specificity of helical computerized tomography for the diagnosis of pulmonary embolism versus the prevalence of pulmonary embolism in the eight studies in the primary literature review.

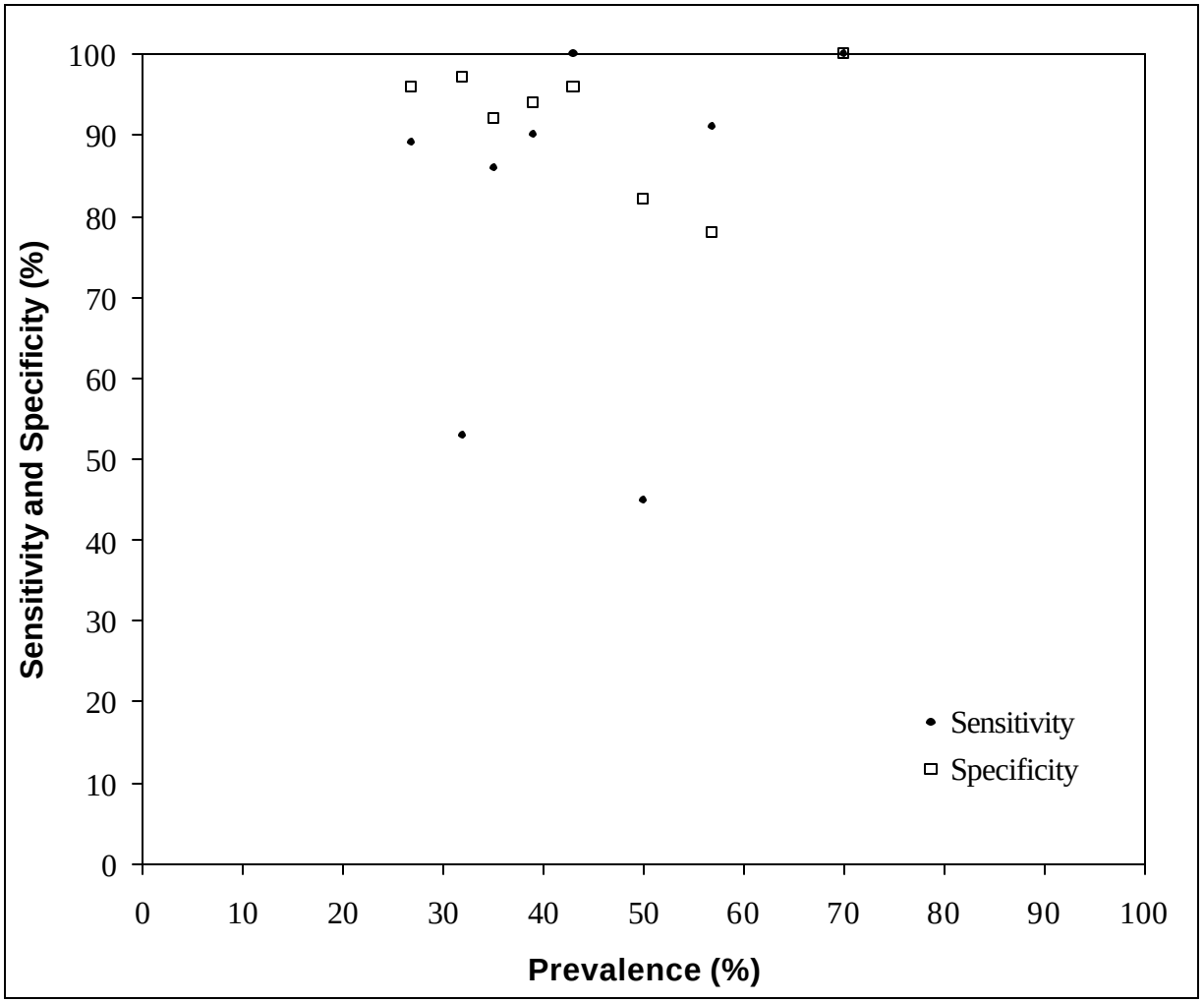

Table 1: Clinical model for predicting pretest probability for deep-vein thrombosis

Checklist
Major Points
Active cancer (treatment ongoing or within previous 6 months or palliative)
Paralysis, paresis, or recent plaster immobilization of the lower extremities
Recently bedridden >3 days and/or major surgery within 4 weeks
Localized tenderness along the distribution of the deep venous system
Thigh and calf swollen (should be measured)
Calf swelling 3 cm >symptomless side (measured 10 cm below tibial tuberosity)
Strong family history of DVT ($\geq$ 2 first degree relatives with history of DVT)
Minor Points

History of recent trauma (≥60 days) to the symptomatic leg
Pitting oedema; symptomatic leg only
Dilated superficial veins (non-varicose) in symptomatic leg only
Hospitalization within previous 6 months
Erythema
Clinical Probability
<u>High</u>
≥3 major points and no alternative diagnosis
≥2 major points and ≥2 minor points + no alternative diagnosis
<u>Low</u>
1 major point + ≥2 minor points + has an alternate diagnosis

1 major point + ≥ 1 minor point + no alternative diagnosis
0 major points + ≥3 minor points + has an alternative diagnosis
0 major points + ≥2 minor points + no alternative diagnosis
Moderate
All other combinations
Active cancer did not include non-melanomatous skin cancer; deep-vein tenderness had to be elicited either in the calf or thigh in the anatomical distribution of the deep venous system.

Table 2: Comparison of imaging modalities used in the diagnosis of PE

Characteristic	V/Q Scintigraphy	Pulmonary Arteriography	Helical CT	MRI
Noninvasive?	Yes	No	Yes	Yes
Does not require iodinated contrast?	Yes	No	No	Yes
Available in many emergency departments?	No	No	Yes	No

Quick examination (<15 minutes)?	No	No	Yes	No
Minimal patient discomfort?	Yes	No	Yes	No
Relatively inexpensive (<500 USD)?	Yes	No	Yes	No