VALUE IN HEALTH 17 (2014) 416–423

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/jval

Comparative Effectiveness Research/HTA

Searching for Indirect Evidence and Extending the Network of Studies for Network Meta-Analysis: Case Study in Venous Thromboembolic Events Prevention Following Elective Total Knee Replacement Surgery

Pascale Dequen, MSc[1,*], *Alex J. Sutton, PhD*[1], *David A. Scott, MA, MSc*[2], *Keith R. Abrams, PhD*[1]

[1]*Department of Health Sciences, University of Leicester, Leicester, UK;* [2]*Oxford Outcomes Ltd., Oxford, UK*

ABSTRACT

Objective: To evaluate the effect of study identification methods and network size on the relative effectiveness and cost-effectiveness of recommended pharmacological venous thromboembolic events (VTEs) prophylaxis for adult patients undergoing elective total knee replacement surgery in the United Kingdom. **Methods:** A stepwise literature search specifically designed to identify indirect evidence was conducted to extend the original clinical review from the latest National Institute for Health and Care Excellence (NICE) VTE technology appraisal. Different network sizes or network orders, based on the successive searches, informed three network meta-analyses (NMAs), which were compared with a replicated base case. The resulting comparative estimates were inputted in an economic model to investigate the effect of network size on cost-effectiveness probabilities. **Results:** Searches increased the number of indirect comparisons between VTE interventions, progressively widening the relevant network of studies for NMA. Precision around mean relative treatment effects was increased as the network was extended from the base case to first-order NMA, but further extensions had limited effect. Cost-effectiveness analysis results were largely insensitive to variation in clinical inputs from the different NMA orders. **Conclusions:** No standard methodology is currently recommended by NICE to identify the most relevant network of studies for NMA. Our study showed that optimizing the identification of studies for NMA can extend the evidence base for analysis and reduce the uncertainty in relative effectiveness estimates. Although in our example network extensions did not affect the acceptability of available treatments in VTE prevention based on cost-effectiveness results, it may in other applications.

Keywords: evidence synthesis, indirect treatment comparison, network meta-analysis, relative effectiveness, venous thromboembolism.

Introduction

The quantitative synthesis of clinical data is a key and often necessary step to the relative effectiveness assessment of medical interventions both premarket and postmarket launch. Meta-analysis is widely used to combine results from multiple clinical studies and considered best practice by many regulatory and health technology assessment bodies in Europe and worldwide [1]. The potential advantages, as well as standard methodology for conducting meta-analysis, are well established in the scientific community with acknowledged guidelines by the Cochrane Collaboration and the Centre for Reviews and Dissemination [2,3]. Recent statistical developments are extending this analytical approach to networks of studies, synthesizing evidence from both direct and indirect treatment comparisons [4–6].

When no head-to-head trial is available, studies evaluating A versus B and B versus C can be used to compare A and C indirectly using network meta-analysis (NMA). Indirect comparisons must be connected by at least one common comparator, that is, treatment B. Additional intermediate links may be required to connect two treatments of interest, thereby increasing the degree of "removal" or "separation" between comparisons and decreasing the degree of influence on the analysis [7]. A number of methodological concerns have been raised when extending an evidence base to include indirect comparisons within a network of studies such as how to best identify indirect evidence. The ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices published guidance on how to conduct NMA and recommended Hawkins et al.'s iterative search strategy to identify indirect evidence [7,8]. Although this search methodology can maximize the NMA

* *Address correspondence to*: Pascale Dequen, Department of Health Sciences, University of Leicester, Adrian Building, University Road, Leicester LE1 7RH, UK.

E-mail: pascale.dequen@le.ac.uk.

1098-3015$36.00 – see front matter

http://dx.doi.org/10.1016/j.jval.2014.02.013

network by efficiently identifying indirect evidence, authors warn that if more than a few links separate treatments (e.g., A and C), results may be unreliable. Additional links can provide useful information but may also increase between-study heterogeneity, uncertainty around estimates, and inconsistency between direct and indirect comparisons [7–9]. We carried out a case study to evaluate the effect of study identification methods and network size on indirect treatment comparisons for the prevention of venous thromboembolic events (VTEs) after total knee replacement (TKR) surgery.

The use of pharmacological, as well as mechanical, prophylaxis for VTE—deep vein thrombosis (DVT) and/or pulmonary embolism—after elective orthopaedic surgery is common practice in the United Kingdom. In 2010, the National Institute for Health and Care Excellence (NICE) published a clinical guideline on reducing the risk of VTE in patients admitted to hospital; at that time, five drugs were recommended: dabigatran etexilate, fondaparinux sodium, low molecular weight heparins, rivaroxaban, and unfractionated heparin for patients with renal failure [10]. Based on relative effectiveness estimates compared with these existing medicines, apixaban was also recommended in 2012 by NICE for use in adult patients scheduled for elective total hip or knee replacement [11]. These drugs were evaluated over time in single technology appraisals and all shown to be cost-effective for their given indication [11–13].

Objectives

We built on the latest NICE VTE technology appraisal TA245 for apixaban [11] to reanalyze the relative effectiveness and cost-effectiveness of recommended pharmacological VTE prophylaxis for adult patients undergoing elective TKR surgery in the United Kingdom using NMA. We sought to evaluate the effect of different network sizes on decision making for VTE prevention.

Methods

Literature Review

A stepwise systematic literature review was conducted in MEDLINE, Medline-in-Process, OLD Medline, EMBASE, and the Cochrane Library in October 2012 to identify relevant studies. The searches were replicated using the reported search strategies for the apixaban appraisal clinical review and adapted using Hawkins et al.'s [7] breadth-first search methodology presented in Table 1 [11,14,15].

Table 1 – Breadth-first search strategy.

Search order	Search iteration	Search comparators
1	i	All first-order comparators *except one*
	ii	First-order comparator previously omitted
2	iii	All second-order comparators *except one*
	iv	Second-order comparator previously omitted
3	v	All third-order comparators *except one*
	vi	Third-order comparator previously omitted

Note. Adapted from Table 1 of Hawkins et al. [7].

Breadth-first searching is based on graph theory; it is an uninformed or "naive" search process that aims to exhaustively search a sequence or a combination of sequences from a "root" node on a graph to all "neighboring" nodes without considering a final limit until it is reached. A parallel can be drawn between nodes on a graph to interventions on a network map and the need to identify all common comparators within a network without knowing the final size or shape of the network. Hawkins et al. [7] refer to search "orders" and associated search comparators to describe each sequential step in the breadth-first search. Treatments directly compared with first-order comparators following first-order searches become second-order comparators, and so on. The sequence of searches in Table 1 progressively include first-, second-, and third-order comparators, allowing us to identify all trials contributing to a network of evidence, until no further comparators are identified. From the set of identifiable trials, all relevant indirect comparisons are also identified at any given order.

In accordance with Hawkins et al. [7], searches were divided further for each order. In Table 1, search orders are numbered 1 to 3 and searches within each order i to vi. For example, in the first-order searches, all but one first-order comparator are included in the search terms (cf. search (1i) in Table 1). The omitted comparator is searched separately in a subsequent search iteration to ensure that all trials including one or more first-order comparators are captured and all possible second-order comparators identified (cf. search (1ii) in Table 1). Search (1i) will identify all trials comparing more than one of the first-order treatments, thus identifying any direct head-to-head evidence, albeit one of the treatments is not included in the search syntax. If the objective is to capture only first-order (i.e. direct) comparisons, the subsequent search (1ii) of the omitted comparator is not required. In this instance, dividing the search into two steps has the potential to reduce the search burden if a particular comparator is associated with a large number of hits. Hawkins et al. [7] thus recommend omitting a widely used comparator such as placebo or best supportive care; however, this is arbitrary. If further search orders are conducted and abstracts reviewed, search (1ii) is redundant and each order comparators could be searched at once. First-order comparators can be arbitrarily selected within or outside the original scope of searches and include treatments not of interest for appraisal. Moreover, study selection is intentionally broadened to include all clinical trials evaluating a first-order comparator without a restriction on comparator criteria, allowing for treatments that may not fall within the scope for appraisal, such as unlicensed drugs, nonrelevant treatments for decision making, or nonpharmacological interventions, to contribute to the network of evidence.

Studies were selected at the abstract and publication level on the basis of the indicated population for TKR and restricted to prospective, phases II to IV randomized controlled trials. To replicate the search conditions and provide comparable model results to the original technology appraisal, abstracts were further restricted by date to studies published before September 2011 and to English language. Date restrictions were included in the search strategy and exclusion of non-English abstracts and publications took place during the screening phase.

Network Meta-Analysis

Network sizes were based on the studies selected following each search order, thereafter referred to as first-, second-, and third-network orders. The base case was defined a priori in the apixaban appraisal from three pivotal phase III clinical trials

comparing apixaban 2.5 mg/bd, dabigatran etexilate 220 mg/qd, and rivaroxaban 10 mg/qd to enoxaparin 40 mg/qd, respectively [16–18]. In accordance with the submitted apixaban economic model [14], these interventions form the decision space for VTE prevention after TKR and are routinely used in clinical practice in the United Kingdom. A comparison with fondaparinux was not considered relevant by manufacturers or the evidence review group because of its low market share in the United Kingdom and was therefore excluded from the analysis. The evidence network used in the original technology appraisal is referred to as the base case and shown in Figure 2A.

A Bayesian NMA was conducted for each network order for the composite outcome of total VTE and all-cause death, as well as for total DVT, and any bleeds. Multiple outcomes were analyzed for economic modeling purposes and to curb potential outcome reporting bias for the composite measure of all VTE/all-cause death used in more recent trials as primary outcome measure but not frequently calculated in older studies [16–18]. Fixed- and random-effects NMA models adjusted for multiarm trials were used in WinBUGS version 1.4.3 to estimate odds ratios (ORs), using Ades et al.'s codes available online [19,20].

The first 20,000 simulations were discarded as a burn-in and achieved reasonable convergence according to visual inspection of trace and history plots. Main analyses were based on a further 50,000 iterations to ensure robustness of results. Model fit was evaluated using the total residual deviance and the deviance information criterion (DIC) for each network size [21]. Between-study heterogeneity was compared using the standard deviation (SD) across random-effects models [22]. Inconsistency was assessed by plotting the residual deviances against the number of intervention arms in each included study, and looking at the proportion of mixed *P* values under 5% and 10% significance [23,24]. We expect that if there was no inconsistency, the residual deviance would equal the number of arms in each trial because it should be equal to 1 for each data point. Mixed *P* values provide an approximation to cross-validation *P* values, which can be calculated in a single model run. According to Welton et al. [25], mixed *P* values calculated from the same data set should follow a uniform distribution on the interval (0,1). We plotted the ordered *P* values for each study and each network order against uniform order statistics to evaluate inconsistency looking at unusually small or large *P* values [25].

Economic Model

A combined decision tree and Markov chain was built in Excel to model the initial prophylaxis/90-day postsurgery phase and the following 35-year time horizon, respectively. The economic model was rebuilt using the input data provided in the apixaban manufacturer submission and evidence review group report. The modeling approach and assumptions were externally validated against the original model [14,15]. Figure A in the Appendix in Supplemental Materials found at http://dx.doi.org/10.1016/j.jval.2014.02.013 illustrates the two-phase model diagram.

Treatment effect was demonstrated only during the first 90 days of the clinical pathway. We applied the ORs for all VTE/all-cause death and any bleeds from the NMA to adjust the baseline risk and inform transition probabilities in the decision tree. Baseline risks were taken from the Apixaban Dose Orally vs. Anticoagulation with Enoxaparin-2 trial for enoxaparin 40 mg/qd as in the original technology appraisal [16]. The parameterization of the Markov model was identical for all treatments compared. Uncertainty around parameters was expressed in distributions; a probabilistic sensitivity analysis was performed using 1000 model runs sampling from these distributions. ORs for all VTE/all-cause death and any bleeds were sampled from 10,000 Markov chain Monte-Carlo simulations extracted from WinBUGS. Quality-adjusted life-years were used to estimate incremental cost-effectiveness ratios (ICERs) compared with enoxaparin; the 2.5th and 97.5th percentiles were also extracted to demonstrate the variation in uncertainty around mean ICER estimates at each given order.

Results

Literature Review

We considered the list of comparators included in the original apixaban submission search strategy as first-order comparators. More than 25 product names and drug classes of interest for VTE prevention in both total hip and knee replacement were included as first-order comparators. Different dosages were considered as individual treatments in the analysis. A full search strategy and the complete list of comparators included in each search order are included in the Appendix (cf. Table A1-3 and Table B) in Supplemental Materials found at http://dx.doi.org/10.1016/j.jval.2014.02.013.

Fifty-three clinical trials met the inclusion criteria over the three network orders. Figure 1 shows the study selection flow diagram broken down by search and network order. The numbers of studies included and excluded for each search iteration are also presented and totaled by network order. Figure B in the Appendix in Supplemental Materials found at http://dx.doi.org/10.1016/j.jval.2014.02.013 illustrates the network map representing all treatment comparisons identified by successive search orders. The number of randomized controlled trials included in the NMA was limited to focus solely on treatment comparisons that would inform the relative effectiveness estimates for apixaban 2.5 mg/bd versus relevant comparators for decision making (i.e., dabigatran etexilate 220 mg/qd, enoxaparin 40 mg/qd, rivaroxaban 10 mg/qd). Graphically, these comparisons are referred to as "closed loops" within the network of studies. Focusing on these loops allowed us to reduce the size of the evidence base and make data sets more manageable without biasing results, because excluded studies did not contribute to indirect comparisons relevant to the decision space. Figure 2 illustrates the network diagrams for each search order including only the closed loops with the interventions of interest shaded in gray, as well as the base-case Indirect Treatment Comparison (ITC) network for reference. Asterisks in Figure 2 indicate that multiple drug dosages were represented by one node; although different dosages were considered as individual treatments in the analyses, these were not illustrated in the networks for readability. Note that we included interventions from three-arm trials even if only one treatment comparison from the trial was of interest, such as in Wang et al. [26] comparing placebo, fraxiparine (nadroparin calcium) 0.2 to 0.4 ml/qd, and indomethacin 25 mg/bd. Lastly, not all studies reported the outcomes of interest and were *de facto* excluded from the NMA. The final numbers of studies in each NMA order for TKR are included in Figure 1 and presented in tabular format in the Appendix (cf. Table C) in Supplemental Materials found at http://dx.doi.org/10.1016/j.jval.2014.02.013, including studies reporting separate results for total hip and knee replacement in the same publication.

Network Meta-Analysis

Goodness-of-fit statistics for the fixed- and random-effects NMA models are presented in Table 2. Fixed-effects models for all network orders were used because they provided the best fit to the data according to the DIC. Forest plots in Figure 3 summarize

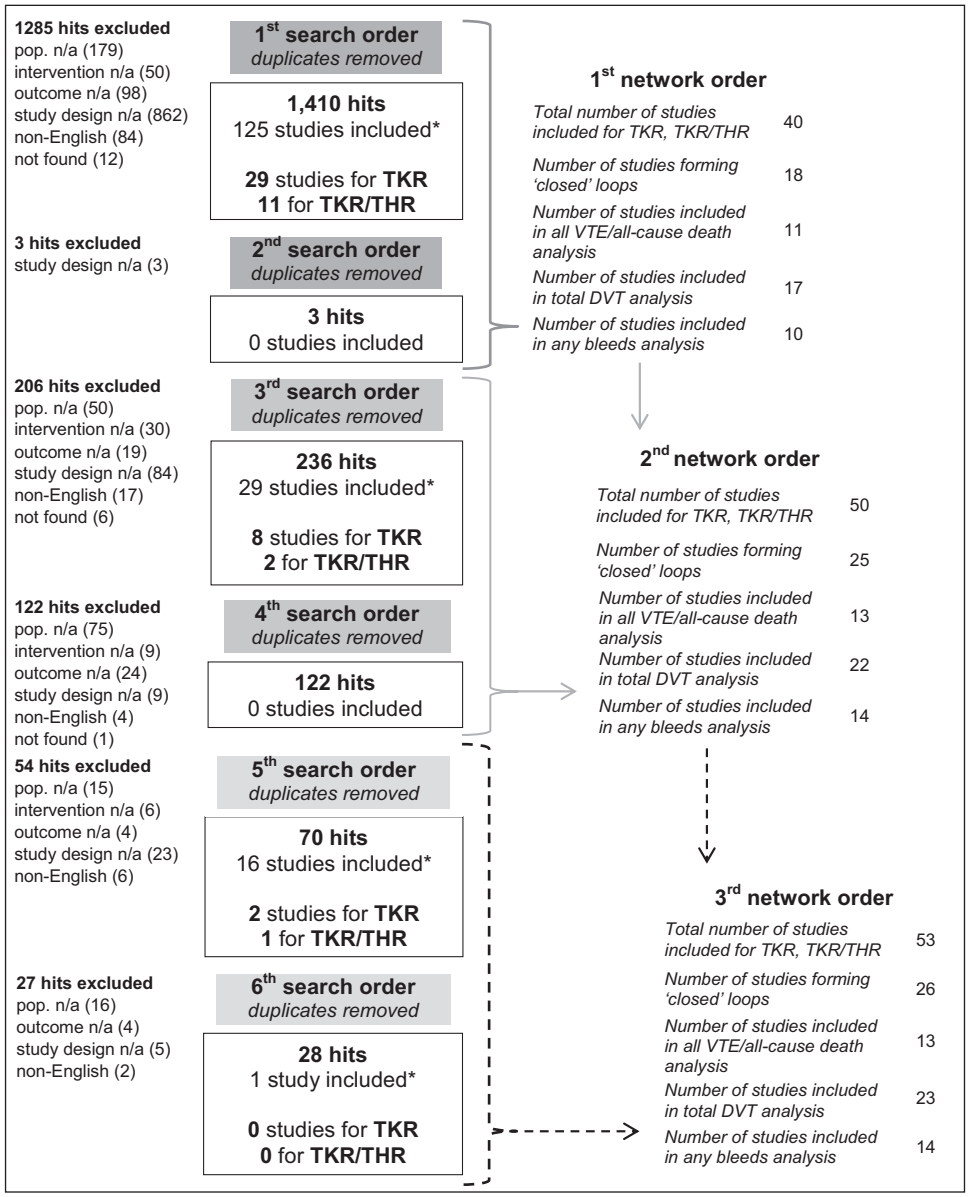

Fig. 1 – Study selection flow diagram. Asterisk indicates that the remainder of the included studies were THR only. DVT, deep vein thrombosis; n/a, not applicable; pop., population; THR, total hip replacement; TKR, total knee replacement; VTE, venous thromboembolic event.

the mean ORs and 95% credible intervals (CrI) for all VTE/all-cause death, total DVT, and any bleeds obtained for the base case and three network sizes. Given the number of studies included (cf. Fig. 1), second- and third-order NMAs for all VTE/all-cause death and any bleeds were the same and results in Figure 3 are presented only for completeness. The growing evidence base from the base case to first-network order marginally increased precision around the mean ORs for all outcomes. For example, the all VTE/all-cause death mean OR for dabigatran versus enoxaparin decreased from 0.95 (95% CrI 0.74–1.22) to 0.90 (0.73–1.10) between the base-case and first-order analysis; similarly, the uncertainty in any bleeds mean OR for apixaban versus enoxaparin was reduced from 0.78 (0.51–1.26) to 0.72 (0.55–0.97). Apixaban and rivaroxaban were superior to enoxaparin for both efficacy outcomes; however, ORs for dabigatran versus enoxaparin were inconclusive. Results favored apixaban over dabigatran for all VTE/all-cause death for all network orders, with a mean OR of 0.65 (0.51–0.85) for first- and second-order analyses. The NMA also estimated that patients are less likely to experience a VTE event/death with rivaroxaban than with apixaban at higher network orders, although the base-case ITC did not support the statistical superiority of rivaroxaban and this was not demonstrated for total DVT. Apixaban showed the most favorable safety profile versus enoxaparin and versus rivaroxaban for first- and second-order NMA.

Although the fixed effects provided the best model fit for all outcomes and all network orders, we considered the random-effects models to assess between-study heterogeneity and the consistency of the evidence. Results for the random-effects models are included in the Appendix (cf. Figure C) in

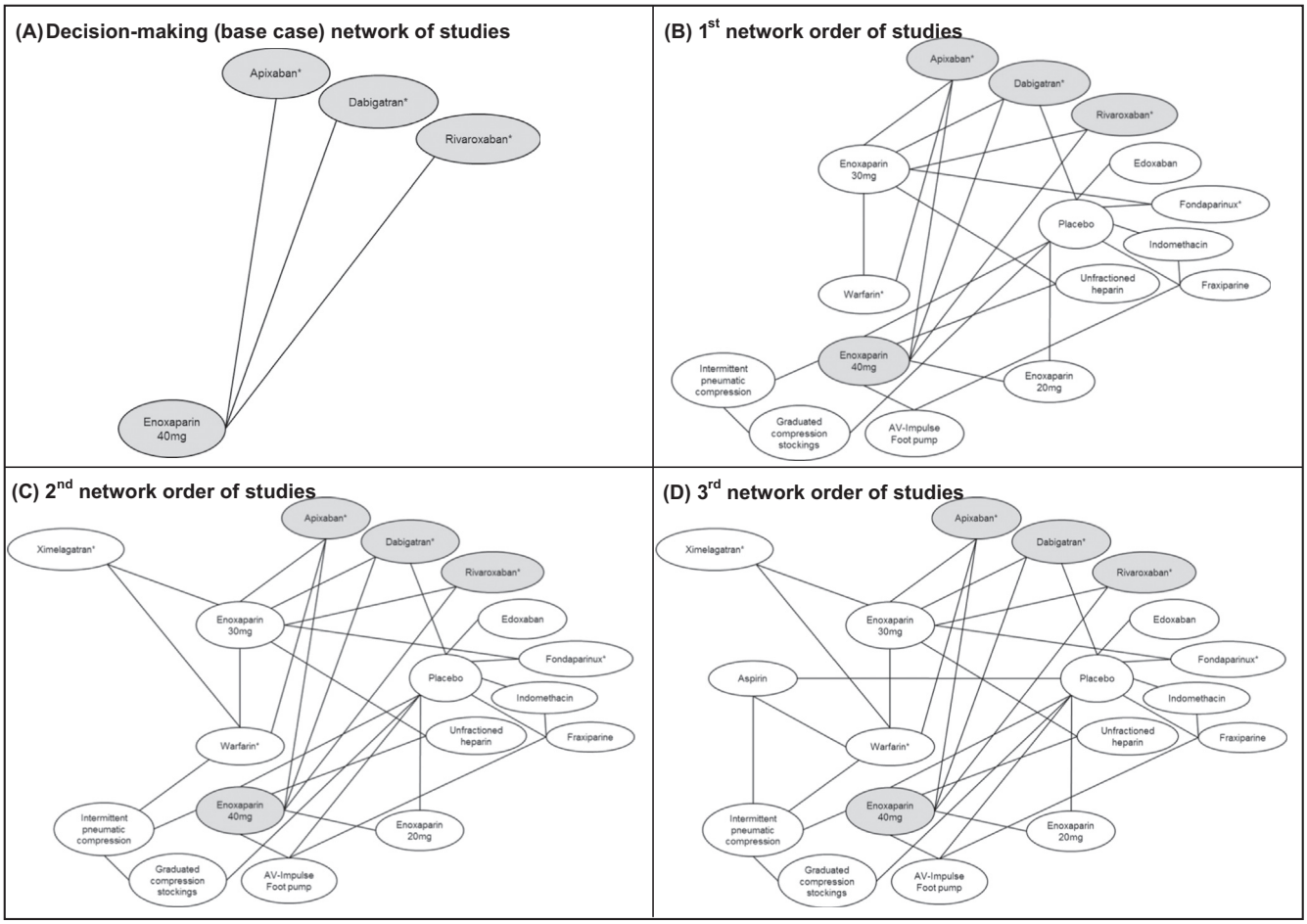

Fig. 2 – Network of studies including only "closed loops" based on search orders. Asterisk indicates multiple dosages included.

Supplemental Materials found at http://dx.doi.org/10.1016/j.jval.2014.02.013 Overall, results were consistent across all network orders for both fixed- and random-effects models with little variation between respective point estimates and CrI. The between-study heterogeneity estimates and CrI were reduced for all VTE/all-cause death from 0.156 (0.005–0.588) to 0.108 (0.004–0.379)
and from 0.115 (0.003–0.569) to 0.108 (0.004–0.350) for any bleeds from first- to second-order NMA. The SDs increased, but not considerably, from 0.092 (0.002–0.307) to 0.112 (0.006–0.341) and

Table 2 – Goodness-of-fit statistics for fixed- and random-effects NMA models for all network orders.

Network order	Fixed effects		Random effects		
	DIC	Total residual deviance	DIC	Total residual deviance	SDs (95% CrI)
Total VTE/all-cause death					
First-order	260.97	39.92	262.27	39.33	0.156 (0.005– 0.588)
Second-order	303.14	44.23	304.45	43.95	0.108 (0.004– 0.379)
Third-order	NA	NA	NA	NA	NA
All DVT					
First order	366.15	52.48	369.14	52.96	0.092 (0.002– 0.307)
Second order	468.65	70.59	471.05	69.45	0.112 (0.006–0.341)
Third order	490.00	80.1	492.11	77.98	0.138 (0.015–0.391)
Any bleeds					
First order	237.87	33.46	239.46	34.12	0.115 (0.003–0.569)
Second order	303.89	42.29	305.36	42.86	0.108 (0.004–0.350)
Third order	NA	NA	NA	NA	NA

CrI, credible interval; DIC, deviance information criterion; DVT, deep vein thrombosis; NA, not available; NMA, network meta-analysis; VTE, venous thromboembolic event.

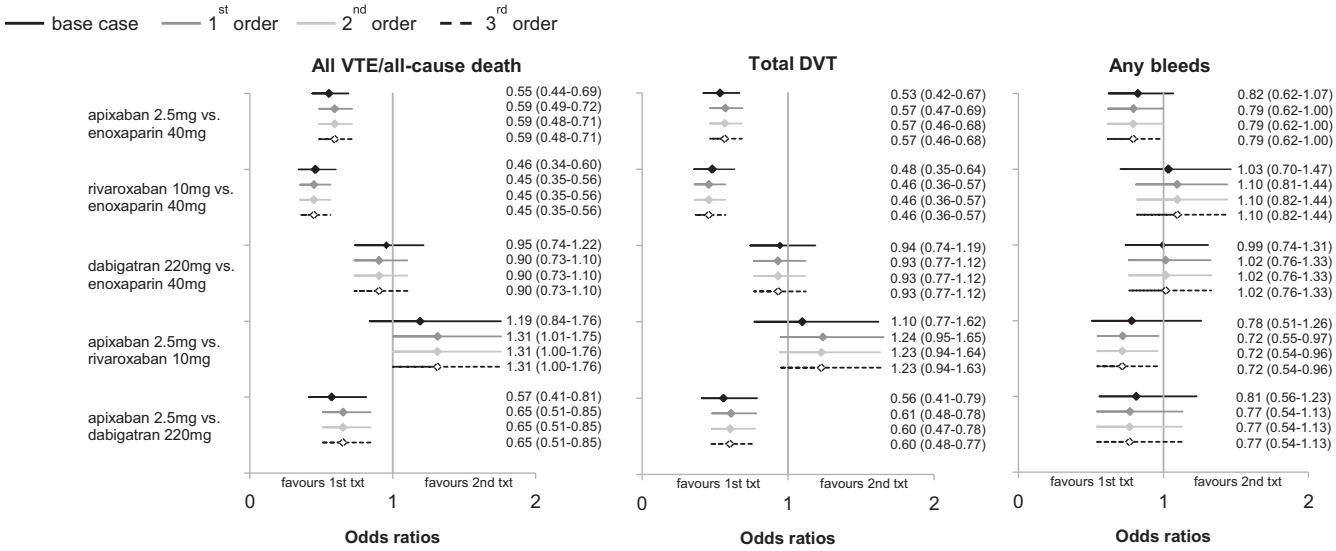

Fig. 3 – Odds ratios for all VTE/all-cause death, total DVT, and any bleeds from fixed-effects NMA models. DVT, deep vein thrombosis; NMA, network meta-analysis; VTE, venous thromboembolic event.

0.138 (0.015–0.391) as the network of studies grew across all three total DVT networks. Spiegelhalter et al. [22] provide a possible interpretation of the random-effects SD by describing a "range" of ORs. This range is in fact the ratio of the 97.5% to the 2.5% point of the distribution of ORs for any given relative treatment effect. They state that SDs on the OR scale of 0.1 or 0.2 will only ever correspond to a range of ORs of 1.48 or 2.19, respectively [22]. Therefore, the SDs reported in Table 2, all smaller than 0.2, showed little evidence of between-study heterogeneity.

Investigatory plots of residual deviances against the number of intervention arms for each trial, outcome, and network order, as shown in Figure D in the Appendix in Supplemental Materials found at http://dx.doi.org/10.1016/j.jval.2014.02.013, do not suggest any inconsistencies between direct and indirect evidence across all models. We also plotted the ordered mixed predicted *P* values against uniform order statistics and found the evidence to be consistent across the three outcomes and network orders (cf. Figure E in the Appendix in Supplemental Materials found at http://dx.doi.org/10.1016/j.jval.2014.02.013). Although the plotted mixed *P* values appear to deviate from a uniform distribution, no individual *P* value was significant at 5% or, more appropriately, at 10% due to the estimates being conservative by nature [25].

Lastly, analysis of both efficacy outcomes—that is, all VTE/all-cause death and total DVT—showed little variation largely due to the relatively low risks of pulmonary embolism (fatal and non-fatal) and death among surgical patients, suggesting no outcome reporting bias for composite measures in the VTE literature.

Economic Model

Apixaban, dabigatran etexilate, and rivaroxaban were found to be cost-effective versus enoxaparin for all network orders. These results were in line with findings from NICE appraisals that recommended these treatments on the basis of their dominance over enoxaparin. Table 3 presents the probabilistic sensitivity analysis means for total costs, total quality-adjusted life-years, and ICERs for the base case and first- and second-network orders. As previously stated, second- and third-order NMA results for all VTE/all-cause death and any bleeds were the same, because these were the clinical inputs to our model, and comparative effectiveness analysis results for the third-network order were redundant and not included in Table 3.

The mean ICERs for rivaroxaban, apixaban, and dabigatran etexilate were negative across all models, suggesting that treatments were on average both more effective and less costly than enoxaparin. The cost-utility analysis results showed little variation in outcomes despite the growing evidence base for the NMA parameterizing the economic model. Figure F in the Appendix in Supplemental Materials found at http://dx.doi.org/10.1016/j.jval.2014.02.013 shows the cost-effectiveness planes based on the probabilistic sensitivity analysis results. At face value, these plots appear uninformative with regard to the effect of network size on the economic evaluation of compared pharmacological treatments for VTE. The percentages in Table 3, however, indicate a reduction in the uncertainty for which treatment is most cost-effective at a £20,000 willingness-to-pay threshold from the base case to first-network order, with rivaroxaban's predicted percentages increasing from 83.2% to 97.1% cost-effective. In addition, although the dominance of dabigatran versus enoxaparin is asserted by all network orders and the mean outcomes do not reflect any significant change, the 2.5th and 97.5th percentiles presented show the widest uncertainty in the ICERs.

Discussion

Using a breadth-first search strategy specifically designed to optimize the identification of indirect evidence allowed us to extend the network of relevant studies for analysis. Extensions of the network maximized the number of indirect comparisons between existing VTE interventions, and precision was increased from the base case to first-network order because additional studies reduced the uncertainty around mean ORs for all VTE/all-cause death, total DVT, and any bleeds. Estimates, however, became more stable as fewer studies were included in the evidence networks with each subsequent search order. Authors believe that additional information provided by trials comparing existing treatments to a lower dose of enoxaparin (30 mg/bd) identified in first-order searches contributed in large part to the increased precision across all outcome estimates. Overall, results from the NMA were consistent across network orders and extending the networks did not increase heterogeneity or inconsistency between studies. The cost-utility analysis was insensitive to NMA results; variation in the clinical input data according

Table 3 – Cost-effectiveness results of apixaban, dabigatranetexilate, and rivaroxaban vs. enoxaparin for the base case, first-order, and second-order networks.

Interventions	Total costs (£)	Total QALYs	ICERs (£)	2.5th percentile uncertainty (£)	97.5th percentile uncertainty (£)	% cost-effective at 20K	% cost-effective at 30K
Base case (ITC)							
Rivaroxaban	703	9.32	− 3,412	−4,171	−2,957	83.2	83.3
Apixaban	810	9.27	−3,703	−4,627	−3,109	16.8	16.7
Dabigatran etexilate	1,377	9.04	− 17,920	−76,636	75,111	0	0
Enoxaparin 40 mg	1,746	9.02				0	0
First-network order							
Rivaroxaban	688	9.34	−3,387	−4,044	−2,956	97.1	97.1
Apixaban	860	9.26	−3,851	−4,807	−3,225	2.9	2.9
Dabigatran etexilate	1,275	9.09	−7,907	−48,454	25,412	0	0
Enoxaparin 40 mg	1,748	9.03				0	0
Second-network order							
Rivaroxaban	695	9.33	−3,380	−4,043	−2,920	96.3	96.3
Apixaban	868	9.25	−3,841	−4,771	−3,181	3.7	3.7
Dabigatran etexilate	1,293	9.08	−8,197	−55,296	47,541	0	0
Enoxaparin 40 mg	1,754	9.02				0	0

Note. Third-network orders for included model inputs are the same as second-network orders so model results not presented above.
ICER, incremental cost-effectiveness ratio; ITC, indirect treatment comparison; QALY, quality-adjusted life-year.

to network order did not affect mean ICERs but reduced the uncertainty in outcomes without influencing the acceptability of interventions.

A number of limitations and methodological challenges should be addressed. Authors did not find an in-depth exploration of heterogeneity and inconsistency (e.g., node-splitting) across order NMAs was warranted given the results and therefore it was not performed. The selection of first-order comparators was arbitrary because no clear definition of how to optimally choose these search terms currently exists. Hawkins et al. start the iterative searches in their practical example looking at all currently licensed treatments for non–small cell lung cancer across regulatory jurisdictions [9]; our first-order search considered indicated pharmaceutical interventions for VTE prophylaxis in the United Kingdom. Although the NICE scoping process can provide some grounds for defining first-order comparators, depending on the therapeutic area, these can include four interventions, that is, for second-line stage III/IV non–small cell lung cancer, or 30 in our case study. This should not make a difference but could affect how many search iterations are needed in the breadth-first strategy. In our case study, no particular gains were achieved from further dividing search orders because the additional burden of including all comparators, even placebo, rather than all but one comparator was marginal. Ultimately, all relevant comparators will be identified in the sequence of searches; however, the incremental value of higher search and network orders for NMA should be weighed against the associated additional search and computational burden. For example, the authors found that initially splitting each search order as recommended by Hawkins et al. to minimize the search burden, that is, searching for "all except one" comparators and subsequently searching the omitted comparator separately, proved inefficient. We agree with Hawkins et al. [7] that such omission is redundant if the next search order is conducted and abstracts reviewed, as was the case in our example. In practice, searches conducted as part of a clinical evidence review could inform first-network order searches, even if distinct study selection criteria may be required, and this could help alleviate the search burden.

Efforts to widen an evidence base for analysis are highly dependent not only on the literature available but also what outcomes are reported in trial publications. Across all networks, between 3 and 13 studies were excluded from our analyses because they did not report outcomes of interest. Recent work in multiple outcomes analysis could help maximize the evidence base and improve NMA methods [27–29]. Moreover, König et al. [30] propose a new method to characterize the flow of evidence in an NMA using linear coefficients to interpret the "parallelism" and "indirectness" of networks to gauge the risk of bias, heterogeneity, and inconsistency within an indirect treatment comparison. Such methodological extensions to understand an evidence base, including how searching and identifying indirect evidence could be examined quantitatively to optimize network shape and size, are desirable.

Our application of Hawkins et al. [7] search methods to evaluate the cost-effectiveness of VTE prophylaxis suggests that more exhaustive searches to identify indirect evidence can provide valuable additional information for NMA. As we extend the breadth of searches, we can draw on more treatment comparisons to inform the network of studies for analysis. However, we are also more likely to include small sample size and older studies, which may contribute to greater between-study heterogeneity and increase the potential for time bias. Given the contradictory results found by Hawkins et al. in their similar study evaluating relative effectiveness estimates for non–small cell lung cancer treatments across multiple network sizes, the effect of extending the network size on uncertainty remains case-specific [9]. Taken together with our findings, however, this highlights the case for examining a wider network of evidence and in the absence of guidelines, we tentatively recommend Hawkins et al.'s search strategy to both future researchers and

reviewers. This awareness should prevent, or at least discourage, "gaming" when undertaking and reporting NMAs. To ensure transparency, health technology assessment bodies should consider wider networks for clinical review and evidence synthesis, as well as to justify the use of narrower networks for economic modeling and decision making. A simulation study to evaluate the effect of network sizes and shapes for NMA would provide generalizable findings and help formalize guidance on the added value of indirect searching and network extensions.

Source of financial support: This study was funded by Pfizer Ltd. via an unrestricted educational grant to the University of Leicester. K. Abrams is partially supported by the National Institute for Health Research (NIHR) (grant no. NF-SI-0508-10061) as a senior investigator. K. Abrams is a member of the National Institute for Health and Care Excellence (NICE) Technology Appraisal Committee and is also a member of the NICE Decision Support Unit (DSU). K. Abrams, A.J. Sutton, and D.A. Scott have provided methodological consultancy services to Pfizer Ltd. The publication of study results was not contingent on the sponsor's approval or censorship of the manuscript.

Supplemental Materials

Supplemental material accompanying this article can be found in the online version as a hyperlink at http://dx.doi.org/10.1016/j.jval.2014.02.013 or, if a hard copy of article, at www.valueinhealthjournal.com/issues (select volume, issue, and article).

REFERENCES

[1] Kleijnen S, George E, Goulden S, et al. Relative effectiveness assessment of pharmaceuticals: similarities and differences in 29 Jurisdictions. Value Health 2012;15:954–60.
[2] Higgins JPT, Green S, eds. Cochrane handbook for systematic reviews of interventions Version 5.1.0 [updated March 2011]. 2011. Available from: www.cochrane-handbook.org. [Accessed October 15, 2012].
[3] Centre for Reviews and Dissemination. Systematic Reviews CRD's Guidance for Undertaking Reviews in Health Care. York: Centre for Reviews and Dissemination, University of York. Available from: www.york.ac.uk. [Accessed August 28, 2012].
[4] Lumley T. Network meta-analysis for indirect treatment comparisons. Stat Med 2002;21:2313–24.
[5] Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. Stat Med 2004;23:3105–24.
[6] Cooper NJ, Peters J, Lai MCW, et al. How valuable are multiple treatment comparison methods in evidence-based health-care evaluation? Value Health 2011;14:371–80.
[7] Hawkins N, Scott DA, Woods B. How far do you go? Efficient searching for indirect evidence. Med Decis Mak 2009;29:273–81.
[8] Hoaglin DC, Hawkins N, Jansen JP, et al. Conducting indirect-treatment-comparison and network-meta-analysis studies: report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: part 2. Value Health 2011;14:429–37.
[9] Hawkins N, Scott DA, Woods BS, Thatcher N. No study left behind: a network meta-analysis in non-small-cell lung cancer demonstrating the importance of considering all relevant data. Value Health 2009;12:996–1003.
[10] National Institute for Health and Care Excellence. Venous Thromboembolism—Reducing the Risk. Care guidelines CG92. London, UK: National Institute for Health and Care Excellence, 2010. Available from: guidance.nice.org.uk/CG92. [Accessed September 1, 2012].
[11] National Institute for Health and Care Excellence. 2012. Apixaban for the Prevention of Venous Thromboembolism After Total Hip or Knee Replacement in Adults. Technology Appraisal TA245. London, UK: National Institute for Health and Care Excellence. Available from: www.nice.org.uk. [Accessed September 1, 2012].
[12] National Institute for Health and Care Excellence. Dabigatran Etexilate for the Prevention of Venous Thromboembolism After Hip or Knee Replacement Surgery in Adults. Technology Appraisal TA157. London, UK: National Institute for Health and Care Excellence, 2008. Available from: www.nice.org.uk. [Accessed September 1, 2012].
[13] National Institute for Health and Care Excellence. Rivaroxaban for the Prevention of Venous Thromboembolism After Total Hip or Total Knee Replacement in Adults. Technology Appraisal TA170. London, UK: National Institute for Health and Care Excellence, 2009. Available from: www.nice.org.uk. [Accessed September 1, 2012].
[14] Bristol-Myers Squibb Pharmaceuticals Ltd. and Pfizer Ltd. Apixaban (Eliquis®) for the Prevention of Venous Thromboembolic Events in Adult Patients Who Have Undergone Elective Hip or Knee Replacement Surgery. Venous Thromboembolism—Apixaban (Hip and Knee Surgery): Manufacturer Submission. London, UK: National Institute for Health and Care Excellence, 2011.
[15] Kleijnen Systematic Reviews Ltd. Apixaban for the Prevention of Venous Thromboembolism in People Undergoing Elective Knee and Hip Replacement Surgery. Venous Thromboembolism—Apixaban (Hip and Knee Surgery): Evidence Review Group Report. London, UK: National Institute for Health and Care Excellence, 2011.
[16] Lassen MR, Raskob GE, Gallus A, et al. Apixaban versus enoxaparin for thromboprophylaxis after knee replacement (ADVANCE-2): a randomised double-blind trial. Lancet 2012;375:807–15.
[17] Eriksson BI, Dahl OE, Rosencher N, et al. Oral dabigatran etexilate vs. subcutaneous enoxaparin for the prevention of venous thromboembolism after total knee replacement: the RE-MODEL randomized trial. J Thrombosis Haemostasis 2007;5:2178–85.
[18] Lassen MR, Ageno W, Borris LC, et al. (RECORD3 Investigators). Rivaroxaban versus enoxaparin for thromboprophylaxis after total knee arthroplasty. N Engl J Med 2008;358:2776–86.
[19] Lunn DJ, Thomas A, Best N, Spiegelhalter D. WinBUGS—a Bayesian modelling framework: concepts, structure, and extensibility. Stat Comput 2000;10:325–37.
[20] Ades AE, Welton N, Lu G. Introduction to Mixed Treatment Comparisons. , Bristol, UK: University of Bristol. Available from: www.bristol.ac.uk. [Accessed October 23, 2012].
[21] Spiegelhalter DJ, Best NG, Carlin BP, et al. Bayesian measures of model complexity and fit. J Royal Stat Soc Ser B (Stat Methodol) 2002;64:583–616.
[22] Spiegelhalter DJ, Abrams KR, Myles JP. Bayesian Approaches to Clinical Trials and Health-Care Evaluation. Chichester, UK: John Wiley & Sons Ltd., 2003.
[23] Caldwell DM, Welton NJ, Ades AE. Mixed treatment comparison analysis provides internally coherent treatment effect estimates based on overviews of reviews and can reveal inconsistency. J Clin Epidemiol 2010;63:875–82.
[24] Marshall EC, Spiegelhalter DJ. Approximate cross-validatory predictive checks in disease mapping models. Stat Med 2003;22:1649–60.
[25] Welton NJ, Sutton AJ, Copper NJ. Evidence Synthesis for Decision Making in Healthcare. Chichester, UK: John Wiley & Sons Ltd, 2012.
[26] Wang CJ, Wang JW, Weng LH, et al. Prevention of deep-vein thrombosis after total knee arthroplasty in Asian patients: comparison of low-molecular-weight heparin and indomethacin. J Bone Joint Surg 2004(86-A):136–40.
[27] Ades AE, Mavranezouli I, Dias S, et al. Network meta-analysis with competing risk outcomes. Value Health 2004;13:976–83.
[28] Welton NJ, Cooper NJ, Ades AE, et al. Mixed treatment comparison with multiple outcomes reported inconsistently across trials: evaluation of antivirals for treatment of influenza A and B. Stat Med 2008;27:5620–39.
[29] Lu G, Ades AE, Sutton AJ, et al. Meta-analysis of mixed treatment comparisons at multiple follow-up times. Stat Med 2007;26:3681–99.
[30] König J, Krahn U, Binder H. Visualizing the flow of evidence in network meta-analysis and characterizing mixed treatment comparisons. Stat Med 2013;32:5414–29.