Challenges for decision-makers when assessing within-class comparative effectiveness: the case of anticoagulation therapy for atrial fibrillation

Andrew Briggs*,1,2 [ID], Amanda Howarth2, Shawn Davies2, John Schneider2, Georgios Spentzouris3, Farhan Mughal4, Ahmet Fuat5,6 & Matthew Fay7,8

1London School of Hygiene & Tropical Medicine, London, UK
2Avalon Health Economics, Morristown, NJ 07960, USA
3Daiichi Sankyo Europe Gmbh, Munich, Germany
4Daiichi Sankyo UK Ltd, Buckinghamshire, UK
5Primary Care Physician, Carmel Medical Practice, Darlington, Durham, UK
6Honorary Professor of Primary Care Cardiology, School for Health, Durham University, Durham, UK
7The Willows Medical Practice, Osbourne Drive, Queensbury, Bradford, BD13 2GD, UK
8Department of Science, Engineering, and Medicine, University of Warwick Medical School, The University of Warwick, Coventry, CV4 7AL, UK
*Author for correspondence: andrew.briggs@lshtm.ac.uk

> "The aim of this manuscript is to elucidate the methodological challenges involved in any decision made by a reimbursement authority, such as NICE in the UK, and ask an open question regarding the strength of evidence of comparative effectiveness be required in order to differentiate similarly priced drugs within a single drug class."

In September 2020, the UK National Institute of HealthCare Excellence (NICE) published a draft report for consultation on anticoagulation therapy for stroke prevention in people with atrial fibrillation (AF) [1]. This intervention evidence review is part of the process of updating the NICE clinical guideline on management of AF [2] and addressed the question of which nonvitamin K oral anticoagulation (NOAC) therapy is most clinically and cost-effective for stroke prevention in people with AF. Four NOACs were included in the evidence review, each with a low and high dose formulation: apixaban, dabigatran, rivaroxaban and edoxaban, and all of which had been previously evaluated and recommended in NICE single technology appraisals [3–6].

On the basis of a systematic literature review to identify relevant evidence, network meta-analysis (NMA) to address comparative effectiveness, and subsequent cost–effectiveness analysis (CEA), the committee concluded that apixaban and dabigatran had the most favorable cost–effectiveness (CE) results of the four NOACs, at the NHS list prices and this led to a draft recommendation in the report that apixaban and dabigatran should be used as first-line options [1]. A further recommendation was made that patients who are stable on one of the other anticoagulants (rivaroxaban, edoxaban or warfarin) should discuss switching with their physician [1].

This draft decision by the committee, to recommend two NOACs within a class of four, represents an unusually strong recommendation to differentiate drugs within a particular class. This is especially so, given that the committee themselves state that the acquisition costs based on NHS list prices of the four NOACs are similar (page 74, line 28) [1]. This suggests that the opinion reached by the committee was grounded in the results of the NMA produced comparative effectiveness estimates that then formed the basis of the CEA.

This manuscript reports on the process of examining the committee's draft decision to differentiate drugs within the NOAC class on the basis of clinical and CE in light of the evidence considered by the committee. This

Future Medicine

manuscript was sponsored by the manufacturer of one of the NOACs that was not recommended as a first-line option by the committee (edoxaban, Daiichi Sankyo) in order to help them respond to the consultation process NICE initiated and which closed in November 2020. The aim of this manuscript is to elucidate the methodological challenges involved in any decision made by a reimbursement authority, such as NICE in the UK, and ask an open question regarding the strength of evidence of comparative effectiveness be required in order to differentiate similarly priced drugs within a single drug class.

The manuscript is structured around three further sections. The next section looks at the evidence on comparative effectiveness of seven key clinical outcomes, the need for NMA to generate those estimates of comparative effectiveness and the representation of uncertainty in the results of NMAs. The third section looks specifically at how the results of the NMA are then turned into a CEA that uses quality-adjusted life-years (QALYs) as the metric to synthesize the seven clinical outcomes into a QALY comparison, before considering the inclusion of costs and the representation of uncertainty in the CE results. A final section then discusses the challenges of making recommendations in the context of the uncertainty in clinical and CE both between the NOAC and vitamin-K antagonist (VKA) drug classes and within the NOAC class.

Comparative effectiveness & NMA

On the basis of a systematic literature review of the evidence, the committee concluded that a recently published NMA by López-López and colleagues [7] had included much of the same studies and was an appropriate foundation from which to produce an evidence-based guideline relevant to the UK NHS (page 9, lines 19–21) [1]. For the purposes of this paper, we consider only five treatments: each of the four NOACs at licensed high dose and the VKA warfarin. The original NMA and NICE considered additional treatments, such as antiplatelet therapy and usual care in the absence of treatment, but these were shown to be inferior to VKA treatment. We focus therefore on the key decision of how the newer NOAC treatments compare to the VKA standard of care.

Synthesizing multiple trial outcomes using NMA

In considering how the NOACs compare to warfarin and to each other, we focus on the seven main clinical outcomes of interest which have been subjected to a NMA as part of NICE's review and which are subsequently used in the CE model.
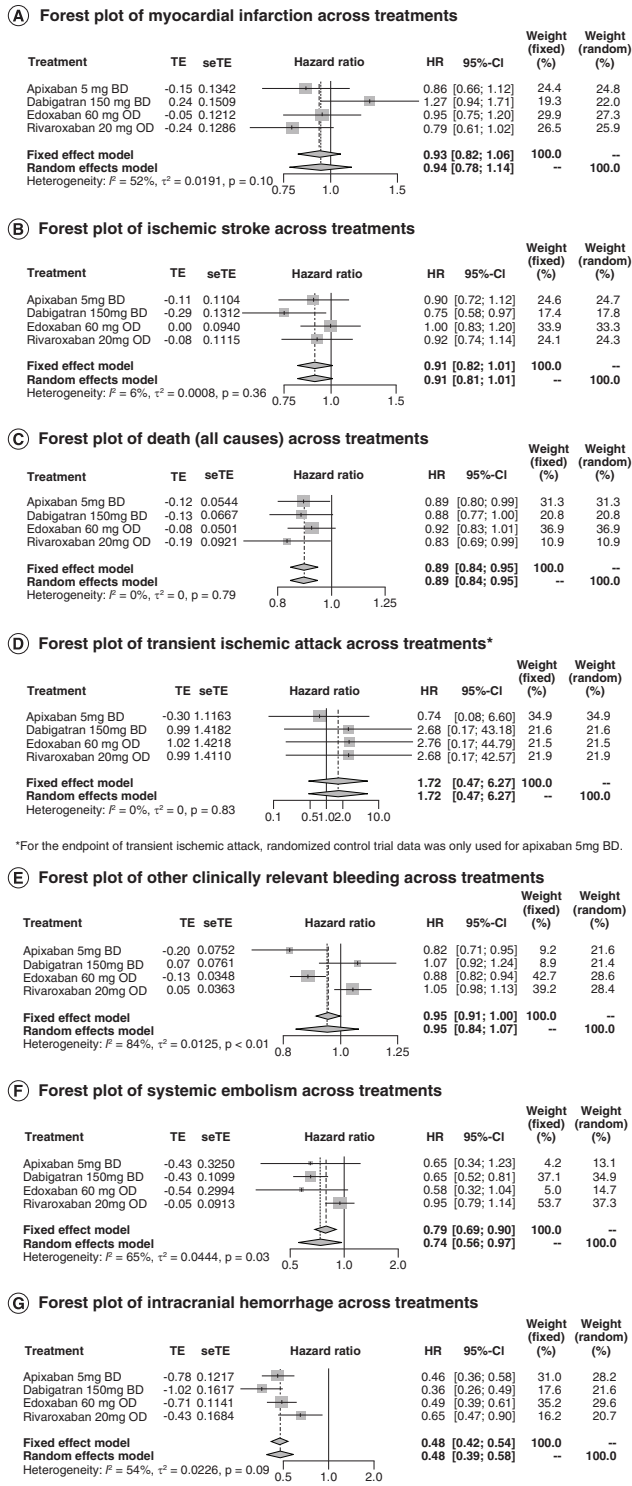
Looking specifically at Supplementary Tables 18–25 of the draft guidance, we find the odds ratios of each treatment (apixaban 5 mg bd, dabigatran 150 mg bd, edoxaban 60 mg od and rivaroxaban 20 mg od) for each outcome [1]. The point estimates of these odds ratios suggest that apixaban does not outperform the other NOACs across all clinical outcomes. The same conclusion can be made from interpretation of Figure 1A–G, which displays forest plots of the hazard ratio and associated 95% CI for seven main events based on our reproduction of the results of the NMA (Supplementary Table 52, page 172) (Box 1) [8].

Heterogeneity across treatments for the seven events of interest in the competing risk NMA is reported at the bottom of Figures 1A–G (Q-statistic p-values and I^2 values). The I^2 values indicate moderate heterogeneity across treatments for myocardial infarction ($I^2 = 52\%$) and intracranial hemorrhage ($I^2 = 54\%$). There was substantial or considerable heterogeneity across treatments for systemic embolism ($I^2 = 65\%$) and other clinically relevant bleeding ($I^2 = 84\%$).

Heterogeneity in trial design

There are distinct differences between the four pivotal Phase III randomized controlled trials of NOACs for stroke prevention in patients with nonvalvular atrial fibrillation that formed the basis of the NMA. These differences have the potential to make direct comparison of the results misleading even within the context of a carefully conducted NMA [12]. Methodological variations, differences in outcome definitions and other variations between studies contribute to the heterogeneity between studies.

The heterogeneity present between these studies further contributes to the uncertainty in the between treatment comparisons of the results of any NMA. We understand that the use of a NMA is necessary but it is important to highlight that any uncertainty intervals calculated as a result of a NMA assume that the data from the trials included are exchangeable. As described by Camm and colleagues [12], this is unlikely to be the case. Heterogeneity across comparator arm populations in each trial should be considered when evaluating and comparing trial outcomes for within-class treatments. In the case of NOACs, patients taking warfarin across trials may differ in characteristics

(A) **Forest plot of myocardial infarction across treatments**

Treatment	TE	seTE	Hazard ratio	HR	95%-CI	Weight (fixed) (%)	Weight (random) (%)
Apixaban 5 mg BD	-0.15	0.1342		0.86	[0.66; 1.12]	24.4	24.8
Dabigatran 150 mg BD	0.24	0.1509		1.27	[0.94; 1.71]	19.3	22.0
Edoxaban 60 mg OD	-0.05	0.1212		0.95	[0.75; 1.20]	29.9	27.3
Rivaroxaban 20 mg OD	-0.24	0.1286		0.79	[0.61; 1.02]	26.5	25.9
Fixed effect model				**0.93**	**[0.82; 1.06]**	**100.0**	--
Random effects model				**0.94**	**[0.78; 1.14]**	--	**100.0**
Heterogeneity: $I^2 = 52\%$, $\tau^2 = 0.0191$, $p = 0.10$			0.75 1.0 1.5				

(B) **Forest plot of ischemic stroke across treatments**

Treatment	TE	seTE	Hazard ratio	HR	95%-CI	Weight (fixed) (%)	Weight (random) (%)
Apixaban 5mg BD	-0.11	0.1104		0.90	[0.72; 1.12]	24.6	24.7
Dabigatran 150mg BD	-0.29	0.1312		0.75	[0.58; 0.97]	17.4	17.8
Edoxaban 60 mg OD	0.00	0.0940		1.00	[0.83; 1.20]	33.9	33.3
Rivaroxaban 20mg OD	-0.08	0.1115		0.92	[0.74; 1.14]	24.1	24.3
Fixed effect model				**0.91**	**[0.82; 1.01]**	**100.0**	--
Random effects model				**0.91**	**[0.81; 1.01]**	--	**100.0**
Heterogeneity: $I^2 = 6\%$, $\tau^2 = 0.0008$, $p = 0.36$			0.75 1.0 1.5				

(C) **Forest plot of death (all causes) across treatments**

Treatment	TE	seTE	Hazard ratio	HR	95%-CI	Weight (fixed) (%)	Weight (random) (%)
Apixaban 5mg BD	-0.12	0.0544		0.89	[0.80; 0.99]	31.3	31.3
Dabigatran 150mg BD	-0.13	0.0667		0.88	[0.77; 1.00]	20.8	20.8
Edoxaban 60 mg OD	-0.08	0.0501		0.92	[0.83; 1.01]	36.9	36.9
Rivaroxaban 20mg OD	-0.19	0.0921		0.83	[0.69; 0.99]	10.9	10.9
Fixed effect model				**0.89**	**[0.84; 0.95]**	**100.0**	--
Random effects model				**0.89**	**[0.84; 0.95]**	--	**100.0**
Heterogeneity: $I^2 = 0\%$, $\tau^2 = 0$, $p = 0.79$			0.8 1.0 1.25				

(D) **Forest plot of transient ischemic attack across treatments***

Treatment	TE	seTE	Hazard ratio	HR	95%-CI	Weight (fixed) (%)	Weight (random) (%)
Apixaban 5mg BD	-0.30	1.1163		0.74	[0.08; 6.60]	34.9	34.9
Dabigatran 150mg BD	0.99	1.4182		2.68	[0.17; 43.18]	21.6	21.6
Edoxaban 60 mg OD	1.02	1.4218		2.76	[0.17; 44.79]	21.5	21.5
Rivaroxaban 20mg OD	0.99	1.4110		2.68	[0.17; 42.57]	21.9	21.9
Fixed effect model				**1.72**	**[0.47; 6.27]**	**100.0**	--
Random effects model				**1.72**	**[0.47; 6.27]**	--	**100.0**
Heterogeneity: $I^2 = 0\%$, $\tau^2 = 0$, $p = 0.83$			0.1 0.51.02.0 10.0				

*For the endpoint of transient ischemic attack, randomized control trial data was only used for apixaban 5mg BD.

(E) **Forest plot of other clinically relevant bleeding across treatments**

Treatment	TE	seTE	Hazard ratio	HR	95%-CI	Weight (fixed) (%)	Weight (random) (%)
Apixaban 5mg BD	-0.20	0.0752		0.82	[0.71; 0.95]	9.2	21.6
Dabigatran 150mg BD	0.07	0.0761		1.07	[0.92; 1.24]	8.9	21.4
Edoxaban 60 mg OD	-0.13	0.0348		0.88	[0.82; 0.94]	42.7	28.6
Rivaroxaban 20mg OD	0.05	0.0363		1.05	[0.98; 1.13]	39.2	28.4
Fixed effect model				**0.95**	**[0.91; 1.00]**	**100.0**	--
Random effects model				**0.95**	**[0.84; 1.07]**	--	**100.0**
Heterogeneity: $I^2 = 84\%$, $\tau^2 = 0.0125$, $p < 0.01$			0.8 1.0 1.25				

(F) **Forest plot of systemic embolism across treatments**

Treatment	TE	seTE	Hazard ratio	HR	95%-CI	Weight (fixed) (%)	Weight (random) (%)
Apixaban 5mg BD	-0.43	0.3250		0.65	[0.34; 1.23]	4.2	13.1
Dabigatran 150mg BD	-0.43	0.1099		0.65	[0.52; 0.81]	37.1	34.9
Edoxaban 60 mg OD	-0.54	0.2994		0.58	[0.32; 1.04]	5.0	14.7
Rivaroxaban 20mg OD	-0.05	0.0913		0.95	[0.79; 1.14]	53.7	37.3
Fixed effect model				**0.79**	**[0.69; 0.90]**	**100.0**	--
Random effects model				**0.74**	**[0.56; 0.97]**	--	**100.0**
Heterogeneity: $I^2 = 65\%$, $\tau^2 = 0.0444$, $p = 0.03$			0.5 1.0 2.0				

(G) **Forest plot of intracranial hemorrhage across treatments**

Treatment	TE	seTE	Hazard ratio	HR	95%-CI	Weight (fixed) (%)	Weight (random) (%)
Apixaban 5mg BD	-0.78	0.1217		0.46	[0.36; 0.58]	31.0	28.2
Dabigatran 150mg BD	-1.02	0.1617		0.36	[0.26; 0.49]	17.6	21.6
Edoxaban 60 mg OD	-0.71	0.1141		0.49	[0.39; 0.61]	35.2	29.6
Rivaroxaban 20mg OD	-0.43	0.1684		0.65	[0.47; 0.90]	16.2	20.7
Fixed effect model				**0.48**	**[0.42; 0.54]**	**100.0**	--
Random effects model				**0.48**	**[0.39; 0.58]**	--	**100.0**
Heterogeneity: $I^2 = 54\%$, $\tau^2 = 0.0226$, $p = 0.09$			0.5 1.0 2.0				

Figure 1.   Network meta-analysis of seven key clinical events. (A) Forest plot of myocardial infarction across treatments. **(B)** Forest plot of ischemic stroke across treatments. **(C)** Forest plot of death (all causes) across treatments. **(D)** Forest plot of transient ischemic attack across treatments*. **(E)** Forest plot of other clinically relevant bleeding across treatments. **(F)** Forest plot of systemic embolism across treatments. **(G)** Forest plot of intracranial hemorrhage across treatments.
*For the end point of transient ischemic attack, randomized control trial data was only used for apixaban 5 mg BD.
BD: Twice daily; HR: Hazard ratio; OD: Once daily; seTE: Standard error of the treatment estimate; TE: Estimate of treatment effect.

> **Box 1. Reproduction of the results of the NICE network meta-analysis.**
>
> - We reproduced the results from the reported network meta-analysis in order to be able to assess visualizations of the results as well as consider standard tests for heterogeneity. The 'metagen' function was used from the 'meta' package in R to perform a fixed effect and random effects meta-analysis of the event hazard ratios. To obtain the standard error as an input for the 'metagen' function in Figures 1A–G, as well as in subsequent forest plot figures, the lower CI estimate was subtracted from the upper CI estimate and divided by $2 \times 1.96 = 3.92$ (i.e., assuming a normal distribution) [9]. For all forest plot figures, the inverse variance method was used to obtain summary measures and the DerSimonian-Laird estimator was used for tau2. The treatment hazard ratio and 95% CI approximate the values presented in Table 52 of the NICE evidence review [8], with the exception of the CIs for transient ischemic attack, which are much wider in Figure 1D. While our estimates report slight differences in values of the CIs compared with those in the review, we do not believe these differences affect our conclusions. A limitation of the results in Figures 1A–G, as well as results presented in subsequent forest plots, is that the meta package makes the assumption that each effect size estimate is independent of each other, an assumption that is unlikely to hold in our analyses. Additionally, the fixed effect approach assumes that all studies and effect sizes come from one, homogenous population, which is also not met in our analyses [10]. Thus, we added the pooled estimate from a random effects approach, producing similar results for the fixed effect pooled estimate values. The CIs are wider because a random effects approach accounts for additional uncertainty (both within study variance and between study variance) [11].

including age, comorbidities or geographic location – all factors that may contribute to increasing or decreasing costs as well as factors that affect the rates of clinical outcomes to be used in comparative and CEA.

Furthermore, in subgroup analyses of the four pivotal trials some evidence indicative of effect modification was found. For example, effect modification due to age was found in RE-LY [13] and of renal function was found in ARISTOTLE [14]. This means that the transitivity assumption is unlikely to have been met. An additional potential issue is that of ecological bias according to which, the differences seen in subgroup analyses at the individual trial level are not evident at the summary level. Without individual patient data from the individual trials, it will be difficult to detect this effect and adjust for it in NMAs using meta-regression techniques [15]. In conventional (pairwise) meta-analysis, the Cochrane methods guide recommends a minimum of ten studies is necessary to perform a meta-regression [16]. López-López *et al.* conducted a meta-regression to evaluate potential effect modifiers [7]. While López-López *et al.* analyzed 23 studies [7], the number of studies providing head-to-head contrasts within the NMA is likely to be too low to form the basis of a robust meta-regression: only 12 trials explored the efficacy of a NOAC against warfarin. The remaining trials explored warfarin compared with aspirin (n = 10) and apixaban compared with aspirin (n = 1). Across NOACs, no treatment had greater than a total of four studies reporting efficacy outcomes compared with warfarin (edoxaban [n = 4], apixaban [n = 3], dabigatran [n = 3] and rivaroxaban [n = 2]). This may be particularly misleading because, for example, of the three studies comparing apixaban to warfarin, only one study compares the low dose of apixaban to warfarin. Similarly, of the two rivaroxaban studies, one study includes only outcomes for the high dose and the other study includes only outcomes for the low dose. It is likely; therefore, that the meta-regression conducted was not powered to adjust adequately for the differences between the treatment comparisons in terms of potential effect modifiers.

The validity of any technique is conditional on the assumptions. In the absence of effect modification, the fixed effect model may still obtain unbiased point estimates, but given the heterogeneity in the trial design the uncertainties are likely to be underestimated and create a false sense of certainty. In the case of effect modification and given differences in trial design, the random effect analyses as suggested by Ren *et al.* to obtain informative priors of the between trial variance using expert elicitation would be more appropriate [17].

Another limitation resulting from the small number of studies for each comparison was that a NMA using a fixed effect logistic regression approach was conducted [7]. López-López *et al.* indicated that a study limitation of the NMA was that random effects models could not be fitted due to the limited number of comparisons that had been replicated in two or more trials [7].

In the face of these limitations, we consider that the level of uncertainty reported for the NMA that is the basis of the NICE report should be considered the most optimistic. In reality, the limitations described above would suggest greater overall uncertainty than is represented by the presented CIs.
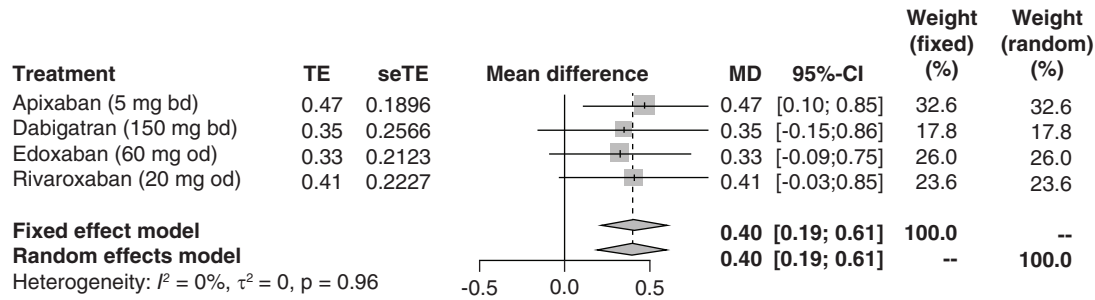
Treatment	TE	seTE	Mean difference	MD	95%-CI	Weight (fixed) (%)	Weight (random) (%)
Apixaban (5 mg bd)	0.47	0.1896		0.47	[0.10; 0.85]	32.6	32.6
Dabigatran (150 mg bd)	0.35	0.2566		0.35	[-0.15;0.86]	17.8	17.8
Edoxaban (60 mg od)	0.33	0.2123		0.33	[-0.09;0.75]	26.0	26.0
Rivaroxaban (20 mg od)	0.41	0.2227		0.41	[-0.03;0.85]	23.6	23.6
Fixed effect model				**0.40**	**[0.19; 0.61]**	**100.0**	**--**
Random effects model				**0.40**	**[0.19; 0.61]**	**--**	**100.0**
Heterogeneity: $I^2 = 0\%$, $\tau^2 = 0$, p = 0.96			-0.5 0.0 0.5				

Figure 2.   Forest plot of mean quality-adjusted life-years differences across treatments compared with warfarin.

Treatment	TE	seTE	Net monetary benefit	Net monetary benefit	95%-CI	Weight (fixed) (%)	Weight (random) (%)
Apixaban 5 mg BD	15264.00	5152.5510		15264.00	[5165.19; 25362.81]	28.5	28.5
Dabigatran 150 mg BD	13450.00	5991.3265		13450.00	[1707.22; 25192.78]	21.1	21.1
Edoxaban 60 mg OD	10088.00	4988.9541		10088.00	[309.83; 19866.17]	30.4	30.4
Rivaroxaban 20 mg OD	10647.00	6154.1071		10647.00	[-1414.83; 22708.83]	20.0	20.0
Fixed effect model				**12384.50**	**[6992.04; 17776.95]**	**100.0**	**--**
Random effects model				**12384.50**	**[6992.04; 17776.95]**	**--**	**100.0**
Heterogeneity: $I^2 = 0\%$, $\tau^2 = 0$, p = 0.89			-30,000 -10,000 0 10,000 30,000				

Figure 3.   Net-benefit forest plot (NOACs compared with warfarin)*.
*NHS List prices were used for the treatment costs.

CE & representation of uncertainty

Synthesizing across multiple outcomes

One of the purposes of a CE model is to synthesize disparate clinical outcomes into a common outcome measure. The standard outcome in CEA is the QALY. The seven key clinical outcomes reported in Figure 1A–G are combined with estimates of life-expectancy, event impacts on quality of life as well as background quality of life to generate an estimated QALY for each treatment in the NICE CEA. Based on the expected QALY estimates and 95% CIs reported in Supplementary Table 65 (page 192) [8], mean QALY differences compared with warfarin were calculated across NOACs. From the mean QALY differences across treatments compared with warfarin in the forest plot presented in Figure 2, one can see that the mean differences are nearly the same for all four NOACs. Only apixaban shows a significant QALY benefit to warfarin; however, the differences between the NOACs are slight and the magnitude of QALY benefits are similar. The conventional tests for heterogeneity do not support differentiation between NOACs on the QALY outcome.

CE & net (monetary) benefit

Supplementary Figure 24 from the evidence review (page 193) combines the simulated results of the QALY differences of the NOACs with the estimated additional costs on the incremental CE plane [8]. Each set of simulated points on the CE plane represents the joint density of incremental costs and effects of each of the four NOACs and no treatment relative to warfarin. That it is difficult to distinguish the 'clouds' of points from each other, ignoring the values for 'no treatment', emphasizes that the NOACs have similar CE.

This result is further emphasized in Figure 3 which shows the net-monetary-benefit forest plot, created using values from the results of the base case analyses in Supplementary Table 65 of the evidence review (page 192) [8]. Net-monetary-benefit is a measure that combines the net cost and net-QALY outcomes together using the decision threshold as an exchange rate.

$$NMB_k = \Delta Q_k \cdot \lambda - \Delta C_k$$

Where ΔQ_k and ΔQ_k are the additional QALYs and costs for NOAC k compared with warfarin and λ is the cost per QALY used for decision making. For the results presented in Figure 3, a cost per QALY of £30,000 was

employed. Note that in constructing this plot, we identified some differences between the values of the CIs in Figure 3 and Supplementary Table 65 of the review likely due to our lack of information on the covariance between costs and effects in the data. Nevertheless, we believe the differences are not large enough to affect our general conclusions on how the evidence should be interpreted.

CE acceptability curves

Supplementary Figure 25 of the evidence review (page 194) displays a cost-effectiveness acceptability curve (CEAC) plotting the probability that each NOAC is cost-effective against the willingness to pay for all four NOAC's, warfarin and no treatment [8]. While it is common to use a CEAC for health economic assessment, CEACs may give a misleading view of the strength of evidence for each treatment because the CEAC uses ordinal measuring, and because 100% of the simulations are split between the total number of treatment options which means that viable options may nevertheless have an apparently 'low' number of times that they are considered optimal. Consider for example that four completely identical treatments in all respects were compared and were all considered cost-effective compared with no treatment. In this case, we would expect the CEACs to show all options as having 25% probability of being cost-effective. Even small apparent differences between treatments would have a profound impact on the rank-order probabilities, which we contend is the issue with the presentation in Supplementary Figure 25 of the evidence review [8].

In the recent NICE consultation process on updating their methods (and in the supporting Task and Finish Report that accompanied the consultation document), the potential confusions arising from relying on CEACs in the presence of many competing options has been highlighted as an area of concern for concealing uncertainty in the ranking process [18,19]. The proposed solution to the problem is to focus on the net-monetary-benefit results of the treatments.

Discussion

Following a NICE draft clinical guideline on stroke prevention in patients with AF which recommended distinguishing drugs within the NOAC class, we reviewed the evidence available to NICE when making such a recommendation. We argued that the evidence supporting such a distinction was weak and unlikely to be robust to the many assumptions underlying the comparative effectiveness and CE calculations using NHS list prices. We believe that a clear and consistent trend across all clinical outcomes of interest that can be translated in a clear gain of QALYs should be a prerequisite to differentiate therapeutic alternatives of the same class with similar costs. We have shown that this was not established, and we have shown that apixaban did not outperform the other NOACs across all of the clinical outcomes included in the NMA and CEA. However, our analyses did not adjust for baseline risk, and thus, we acknowledge that the comparison is limited.

Attention should be given to the totality of the evidence and not focused on a single way of presenting the results. We suspect that in making the initial recommendation, members of the committee may have been influenced by the rank ordering of CE rather than on an appropriate consideration of overall uncertainty in those estimates. However, the overlapping CIs of net benefit and QALY differences help to give a more balanced picture of the strength of evidence to differentiate the NOACs. Graphical methods, such as forest plots, can also be used for QALY and CE outcomes and help to highlight the challenge of distinguishing a treatment option that clearly dominates and outperforms within the class of NOACs.

In the assessment of within-class comparative effectiveness, where differences are small and there is a lack of head-to-head data, it will be generally important to include a textual overall assessment of uncertainty. The objective being, to describe the different types of uncertainty and their potential impact on CE in a simplified and easy to understand manner for readers without technical background. We believe this is aligned with the intended audience of the clinical guidelines document since the majority of physicians do not have a background in health economic methodologies. In addition, we believe that given the confusion and the risk of overinterpretation of CEACs that are showing multiple options, it will be useful to give more weight to the incremental net-benefits of treatments when seeking to understand the implications of uncertainty for relative CE.

While we have focused on within-class comparative effectiveness and CE in the setting of randomized controlled trials, we acknowledge that there are different and distinct methodological variations and biases intrinsic to real-world studies that can make cross-study comparisons and meta-analyses that use real-world data problematic. In the case of NOACs, additional methodological challenges in real-world studies may include appropriately accounting

for varying treatment adherence rates and understanding how improper adherence may affect clinical outcomes, as well as accounting for patients switching treatments.

We believe that patient choice and shared decision making should be fundamental elements of NICE's guideline development process and recommendations should be made with these principles in mind. By recommending the use of only two NOACs, the committee halves patients' options for tailored oral anticoagulation therapy, removes access to once daily regimens and puts pressure on prescribers to switch treatments to approximately 50% of their patients. Switching OAC treatment to patients without any clinical rationale introduces unnecessary risks without any proven advantages. Additionally, commercial deals may exist for treatments at subnational levels, which would change the evidence base for CE.

NICE is a vanguard of health technology assessment methods and its decision-making process can have an impact and influence decisions regarding the management of AF patients outside of its immediate jurisdictional boundary of England and Wales. Although NICE bases its clinical recommendation on CE, in the case of NOACs where the four treatment alternatives have similar acquisition costs using NHS list prices, NICE's clinical guidance recommendations should be aligned with guidelines by worldwide clinical societies such as European Society of Cardiology. We are reassured that, following the consultation process, the final guidance issued reversed the recommendation within the original draft to distinguish within the class of NOACs on the basis of comparative and CE. NICE appears to acknowledge the totality of the evidence is more in line with a general parity between NOACs.

Conclusion

In an earlier draft of their updated guidance, NICE appeared to favor distinguishing the four currently recommended NOACs in terms of effectiveness and therefore CE. Following consultation, involving views from patient advisory groups and clinical experts, this recommendation has been reversed and the guidance no longer seeks to distinguish the NOACs within class. Given the evidentiary challenges of showing differences within drug classes are robust to the underlying analysis, this appears to be the appropriate resolution and continues to offer patients and their treating physicians, the greatest choice of clinically effective and cost-effective treatments taking into account practical benefits of treatment to prevent stroke in atrial fibrillation.

Author contributions

A Briggs drafted the manuscript with support from A Howarth, S Davies, J Schneider, G Spentzouris, F Mughal, A Fuat and M Fay. A Howarth conducted statistical analysis.

Financial & competing interests disclosure

Open access

References

Papers of special note have been highlighted as: ● of interest; ●● of considerable interest

1. National Institute for Health and Care Excellence (NICE); National Guideline Centre (NGC). Atrial fibrillation. Anticoagulant therapy for stroke prevention in people with atrial fibrillation: NICE guideline intervention evidence review, September 2020. Draft for consultation (2020). https://www.nice.org.uk/guidance/ng196/documents/evidence-review-5

●● **This is a draft report for consultation that assesses the most clinically and cost-effective anticoagulation therapy for stroke prevention in people with atrial fibrillation. The draft report serves as the main reference in our manuscript by providing the example case we use throughout the paper. Reading this draft report will provide readers better context and insight into a recent example of how decision makers evaluated within-class comparative effectiveness.**

2. National Institute for Health and Care Excellence (NICE). Atrial fibrillation: the management of atrial fibrillation. (Full Guideline CG180) (2014). https://www.nice.org.uk/guidance/cg180

3. National Institute for Health and Care Excellence (NICE). Dabigatran etexilate for the prevention of stroke and systemic embolism in atrial fibrillation. (TA249) (2012). https://www.nice.org.uk/guidance/TA249

4. National Institute for Health and Care Excellence (NICE). Apixaban for preventing stroke and systemic embolism in people with nonvalvular atrial fibrillation. (TA275) (2013). https://www.nice.org.uk/guidance/TA275

5. National Institute for Health and Care Excellence (NICE). Edoxaban for preventing stroke and systemic embolism in people with nonvalvular atrial fibrillation. (TA355) (2015). https://www.nice.org.uk/guidance/ta355

6. National Institute for Health and Care Excellence (NICE). Rivaroxaban for the prevention of stroke and systemic embolism in people with atrial fibrillation. (TA256) (2012). https://www.nice.org.uk/guidance/ta256

7. López-López JA, Sterne JAC, Thom HHZ *et al.* Oral anticoagulants for prevention of stroke in atrial fibrillation: systematic review, network meta-analysis, and cost effectiveness analysis. *BMJ* 359, j5058 (2017).

•• **The NICE committee decided that the published network meta-analysis by López-López and colleagues was an appropriate foundation from which to produce an evidence-based guideline relevant to the UK NHS. Thus, the results of the publication contributed to National Institute of HealthCare Excellence's decision making process and were used as evidence in assessing NOACs.**

8. Sterne JA, Bodalia PN, Bryden PA *et al.* Oral anticoagulants for primary prevention, treatment and secondary prevention of venous thromboembolic disease, and for prevention of stroke in atrial fibrillation: systematic review, network meta-analysis and cost-effectiveness analysis. *Health Technol. Assess.* 21(9), 1–386 (2017).

•• **The evidence review provides the data we used as inputs for Figures 1A–G, Figure 2 and 3.**

9. Harrer M, Cuijpers P, Furukawa TA, Ebert DD. Chapter 4.3 Binary outcomes. In: *Doing Meta-Analysis in R: A Hands-on Guide.* CRC Press, FL, USA (2019).

10. Harrer M, Cuijpers P, Furukawa TA, Ebert DD. Chapter 4.1 Fixed-effects-models. In: *Doing Meta-Analysis in R: A Hands-on Guide.* CRC Press, FL, USA (2019).

11. Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. Chapter 13: Fixed effect versus random-effects models. In: *Introduction to Meta-Analysis.* Sharples K (Ed.). John Wiley & Sons, Ltd, NJ, USA (2009).

12. Camm AJ, Fox KAA, Peterson E. Challenges in comparing the non-vitamin K antagonist oral anticoagulants for atrial fibrillation-related stroke prevention. *Europace* 20(1), 1–11 (2018).

•   **Camm and colleagues review and describe the methodological differences in nonvitamin K antagonist oral anticoagulant stroke prevention studies. Although specific to challenges in comparing NOAC-focused studies, the challenges they describe, such as difference in study design, patient characteristics and end points, can extend beyond NOACs and apply to the evaluation of other studies and treatments as well.**

13. Lauw MN, Eikelboom JW, Coppens M *et al.* Effects of dabigatran according to age in atrial fibrillation. *Heart* 103(13), 1015–1023 (2017).

14. Hohnloser SH, Hijazi Z, Thomas L *et al.* Efficacy of apixaban when compared with warfarin in relation to renal function in patients with atrial fibrillation: insights from the ARISTOTLE trial. *Eur. Heart J.* 33(22), 2821–2830 (2012).

15. Thompson SG, Higgins JP. Treating individuals 4: can meta-analysis help target interventions at individuals most likely to benefit? *Lancet* 365(9456), 341–346 (2005).

16. Higgins JPT, Green S. *Cochrane Handbook for Systematic Reviews of Interventions.* The Cochrane Collaboration (2011). https://handbook-5-1.cochrane.org/chapter_9/9_6_4_meta_regression.htm

17. Ren S, Oakley JE, Stevens JW. Incorporating genuine prior information about between-study heterogeneity in random effects pairwise and network meta-analyses. *Med. Decis. Making* 38(4), 531–542 (2018).

18. National Institute for Health and Care Excellence (NICE). The NICE methods of health technology evaluation: the case for change (2020). https://www.nice.org.uk/about/what-we-do/our-programmes/nice-guidance/chte-methods-consultation

19. National Institute for Health and Care Excellence (NICE). CHTE methods review. Exploring uncertainty. Task and finish group report (2020). https://www.nice.org.uk/Media/Default/About/what-we-do/our-programmes/nice-guidance/chte-methods-consultation/Exploring-uncertainity-task-and-finish-group-report.docx