INNOVATIONS in pharmacy

Volume 7 | Number 3 Article 7

9-9-2016

Imaginary Worlds: The Status of Modeled Quality Adjusted Life Year Claims for New Oral Anticoagulants in Atrial Fibrillation Published Between January 2012 and February 2016

Paul C. Langley
University of Minnesota, langley@maimonresearch.com

Taeho Greg Rhee
University of Minnesota, Twin Cities, rhee0041@umn.edu

Imaginary Worlds: The Status of Modeled Quality Adjusted Life Year Claims for New Oral Anticoagulants in Atrial Fibrillation Published Between January 2012 and February 2016

Cover Page Footnote
Acknowledgement: The authors would like to thank Professor Jon Schommer for his support in preparing this commentary

Imaginary Worlds: The Status of Modeled Quality Adjusted Life Year Claims for New Oral Anticoagulants in Atrial Fibrillation Published Between January 2012 and February 2016

Paul C. Langley, PhD ; Taeho Greg Rhee, PhD(c)
College of Pharmacy University of Minnesota

Abstract

The purpose of this commentary is to evaluate modeled quality adjusted life year claims (QALYs) for new oral anticoagulants (NOACs) published in the period from January 2012 to February 2016. The focus of this commentary is to assess whether or not the modeled claims meet the standards of normal science in supporting falsification and replication. A systematic and consensus review by the authors identified a total of 23 cost-utility NOACs evaluations along with four single technology appraisals undertaken by the National Institute for Health and Care Excellence (NICE) in the UK. Each study was evaluated in terms of four criteria: (i) did the study generate evaluable claims (ii) id the authors attempt to generate evaluable claims (iii) did the authors suggest how the claims might be evaluated and (iv) did the authors caution readers as to the implications of generating non-evaluable projections or claims for credibility in health system decision making? None of the 23 studies assessed or the four NICE single technology appraisals met any of the four assessment criteria. None of the studies presented projections or claims in a form suitable for empirical evaluation. None could support falsification or replication. They failed the standards associated with the scientific method. Failure to meet the standards of normal science meant that the studies, from a formulary assessment perspective, are not credible. The claims made were either impossible to verify, or if potentially verifiable, were not presented in a testable form. There was no basis for assessing whether the claims were right or even if they were wrong. This lack of scientific credibility is a major concern. In particular, the choice of a lifetime cost-utility framework for assessing the NOACs against warfarin and against each other effectively precludes any experimental assessment. If medical economics is to advance through the formulation and testing of hypotheses, then editors of journals should consider whether or not to set standards for the acceptance of publications to include the requirement for testable claims and the results of claims assessment. If this is not acceptable, then it should be made clear that published modeled claims and simulations are simply imaginary worlds or thought experiments. Editors cannot sit back and assume that at some time in the future non-testable projections will possibly be evaluated.

Keywords: new oral anticoagulant (NOAC), quality-adjusted life years (QALYs), modeling, credibility, imaginary worlds, pseudoscience, adherence

Introduction

The entry of new oral anticoagulants (NOACs) into treatment practice for non-valvular atrial fibrillation has led to a substantial investment in a number of modeled and simulated claims for product cost-effectiveness. These technology assessments have included comprehensive simulated comparisons between warfarin and all competing NOACs, as well as individual NOAC comparisons against warfarin and each other. A common feature of these models or simulations has been their attempt to extrapolate from pivotal phase 3 randomized controlled trials (RCTs) to capture constructed clinical events and direct medical costs attributable to the comparator OACs over the lifetime of the patient cohort. The respective pivotal trials are: edoxaban ENGAGE-AF; dabigatran RE-LY; rivaroxaban ROCKET-AF; and apixaban ARISTOTLE [1 2 3 4]. The common element has been to demonstrate that the entry of the NOACs into clinical practice and the switching from warfarin in non-valvular atrial fibrillation can be justified, not only on clinical grounds, but in terms of projected cost-utility outcomes and willingness-to-pay from a health system perspective. Needless to say, from manufacturing and marketing perspectives this is a highly competitive market.

The purpose of this commentary is to consider how credible the claims for NOACs are in non-valvular atrial fibrillation. The review covers the period from January 2012 to February 2016. Credibility in this research program is determined by the standards of normal science: Does the projection or claim for a product support falsification and replication? [5 6] If, in applying the recognized standards that support the scientific method, the judgment is that the claim lacks credibility, then the recommendation is that the claim should be rejected [7 8]. In focusing on the claims, irrespective of the perceived merits of the model or simulation generating those claims, if they

Corresponding author: Paul C. Langley, PhD
Adjunct Professor
College of Pharmacy University of Minnesota
Tel: 520-577-0436; Email: langley@maimonresearch.com

fail the standards of falsification and replication, then they are not credible inputs for health care decision making [9 10 11]. The standards of normal science are absolute. A constructed model or simulation does not test hypotheses.

This commentary is part of an ongoing evaluation of formulary evidentiary standards supported by the Social and Administrative Pharmacy Program at the University of Minnesota College of Pharmacy. A number of commentaries have already been published. These have focused on (i) the proposed Minnesota guidelines for formulary evaluations [12]; (ii) standards for health technology assessments published in three leading journals: *Pharmacoeconomics*, *Value in Health* and the *Journal of Medical Economics* [13 14 15]; (iii) guidelines for formulary submissions by the Academy of Managed Care Pharmacy (AMCP) in the US and by the National Institute for Health and Care Excellence (NICE) in the UK and the Pharmaceutical Management Agency (PHARMAC) in New Zealand [16 17 18]; and (iv) the inappropriateness of quality adjusted years (QALY) as a recommended 'gold standard' outcome measure in value claims for competing pharmaceuticals [19 20].

Methods

A systematic review, using the PRISMA-P checklist, of PubMed with MeSH terms 'atrial fibrillation AND cost AND effectiveness AND utility' (as of March 24, 2016) yielded 52 studies [21]. Supplementary searches of the bibliography of each study were also undertaken and abstracts extracted. All abstracts were then reviewed independently by both authors with an agreed final selection based on three criteria. These criteria were:

- Did the study compare one or more NOACs against the standard of care warfarin?
- Did the study utilize a decision model or simulation?
- Did the study present claims for competing therapies?

A total of 23 studies were identified. These were classified as:

- Comparative cost-effectiveness reviews: all NOACs: 9 studies
- Comparative cost-effectiveness reviews: single NOAC: 1 study
- Comparative cost effectiveness: dabigatran vs. warfarin: 3 studies
- Comparative cost effectiveness: rivaroxaban vs. warfarin: 4 studies
- Comparative cost effectiveness: apixaban vs. warfarin/other NOACs: 4 studies
- Comparative cost effectiveness: edoxaban vs. warfarin: 2 studies

To evaluate the credibility of a study, each author reviewed the study in terms of four criteria. These are:

- Did the study generate evaluable claims?
- Did the author(s) attempt to generate evaluable claims?
- Did the authors suggest how the claims might be evaluated?
- Did the author(s) caution readers as to the implications of generating non-evaluable claims for the credibility of the analysis in health system decision making?

There was no restriction on the type of claim. A claim could be expressed in cost-effectiveness terms, it could be expressed as quality adjusted life years (QALYs) gained, it could be expressed as direct medical costs or it could be expressed as adverse clinical events avoided. Irrespective of the claim made, this was assessed in the context of providing feedback to a formulary committee to support ongoing disease area and therapeutic reviews.

If these credibility standards are accepted, then the responsibility is on the authors of a modeled claim or simulation to structure their analysis to generate evaluable claims. If claims are put in cost-utility terms, then it has to be shown how those claims might be evaluated. If claims are expressed in adverse events avoided or if the claims were disaggregated by a base-line risk of stroke, again it should be shown how those claims are to be assessed. If the claims rest on assumptions of product discontinuation or adherence/persistence behavior, then once again it should be indicated how these data are to be captured.

As well as reviewing the selected papers, the analysis also considered four single technology assessments by NICE for dabigatran, rivaroxaban, apixaban and edoxaban [22 23 24 25]. These were chosen because they all subscribe to a common model format, the NICE reference case [26]. The reference case is important in health technology assessment because it has been adopted as the preferred format for formulary submissions in a number of single payer health systems [27 28]. It has also influenced the Academy of Managed Care Pharmacy (AMCP) *Format for Formulary Submissions* in the US [29]. Through standards established by professional groups such as the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) and the adoption of quality check lists such as CHEERS, the reference case standard has had a significant effect on the widespread adoption of constructed, lifetime cost-per-QALY models to generate non-evaluable claims for product performance [30 31 32]. In none of these guidelines is there any requirement for projections or claims for cost-effectiveness to be put in evaluable terms.

Results

The results are presented in Table 1 under four heads:

- Study and country
- Sponsor (if stated)
- Type of modeled or simulated claim
- QALYs gained by OAC

As all studies evaluated utilized a lifetime cost-per-QALY framework, QALYs gained were chosen as the representative outcome claim. This avoided attempting to rationalize country specific estimates of direct medical cost and resultant incremental cost-effectiveness ratios (ICERs). QALYs are also of interest because while QALY gains for the OAC interventions are potentially substantial in years of life gained, the results of the pivotal phase 3 RCTs would suggest that QALY gains between the individual OACs are less substantial. Even so, costs and QALYs are not independent of each other; they are both constructed from the underlying clinical event structure.

The results of the systematic review are presented in Table1. The key findings are as follows:

- All studies presented claims in lifetime cost-utility terms (the term lifetime could vary from a minimum time horizon of 20 years to death);
- All of the studies extrapolated from the pivotal phase 3 trials for the respective NOACs utilizing a Markov framework;
- None of the studies presented claims in evaluable terms;
- None of the authors attempted to generate or suggest possible evaluable claims;
- None of the authors suggest how the claims might be evaluated; and
- None of the authors cautioned the reader as to the constructed nature of non-evaluable claims.

Constructed QALYs Gained

In virtually all models or simulations projected QALY gains are similar. Comparisons between the various OACs point to the comparative gains being measured in days or months rather than years. While these should not be taken at face value it is important to put them in the perspective of the survival rates of the constructed atrial fibrillation cohorts. For example, the Coyle et al study, which covered all OACs, yielded estimates of QALYs gained, for a lifetime cohort with average age of 72 years, ranging from 6.480 years to 6.543 years (or a range of 23 days; 0.063 of 365 days) [35]. The Harrington et al study which compared warfarin, rivaroxaban, dabigatran (150mg) and apixaban yielded, for a hypothetical cohort of 70-year-old patients, QALYs gained ranging from 7.97 for warfarin to 8.47 for apixaban (or a range of 6 months) [40]. Within the NOACs the incremental gain was only 0.21 QALYS (2.5 months). The Krejczy et al. modeling for Germany yielded for a follow-up period of 20 years extrapolated from the principal RCTs a gain of 0.04 QALYS comparing dabigatran 110mg with warfarin (15 days), 0.07 QALYS comparing dabigatran 150 mg with warfarin (26 days), 0.08 QALYs comparing rivaroxaban with warfarin (29 days) and 0.21 QALYs for apixaban comparing apixaban with warfarin (77 days) [38]. The Lantis et al study[29], which again utilized a lifetime Markov 6-week cycle model (adapted from models developed by Dorian et al and Lip et al. [55 57]) generated discounted QALY estimates of 6.099 for warfarin, 6,289 for apixaban, a range of 6.186 to 6.221 for dabigatran and 6.242 for rivaroxaban [39]. Although the study claimed that apixaban dominated the other NOACs, the gain in terms of QALYs is relatively small. In respect of warfarin, the gain was 69 days and for rivaroxaban 17 days.

The Harrington et al assessment of competing NOACs and warfarin found for a cohort of 70-year old patients a difference between dabigatran and apixaban of 0.06 QALYs and a superiority of dabigatran over rivaroxaban of 0.15 QALYs [40]. Similar comparisons for models in Portugal for a 73-year old cohort yielded an increment of 0.02 QALYs comparing rivaroxaban to warfarin, while a similar comparison for Belgium yielded an increment of 0.11 QALYs [50 52].

Table 1

Imaginary Worlds: Status of Comparative Incremental QALY Claims for NOACs January 2012 to March 2016

Study [country]	Sponsor (if any)	Modeled or Simulated Claims	QALYs Gained (base case)
Comparative cost-effectiveness reviews: all NOACs			
Jarungsuccess and Taerakun [Thailand][33]	None stated	Markov 30-year time horizon 1-year cycle	Warfarin: 2.29 Dabigatran 110mg: 2.29 Dabigatran 150mg: 2.34 Apixaban: 2.33 Rivaroxaban: 2.31
Coyle et al [Canada][34]	Canadian Institutes of Health Research	Markov 20-year, 10-year and 2-year time horizon.	Warfarin: 6.480 Dabigatran 110mg: 6.543 Dabigatran 150mg: 6.617 Apixaban: 6.617 Rivaroxaban: 6.541 Edoxaban: 6.543
Vestegaard & Ehlers [Denmark][35]	None stated	Markov 10-year time horizon3-month cycle length with starting age of 70 years (with extension to lifetime)	Model to assess cost-effectiveness of implementing a national treatment strategy of strict adherence to 2012 ESC guidelines compared to previous strategy. Base-case estimated QALY gain of 5.316 compared to 4.942.
Wu et al [China][36]	Program of Shanghai Chief Science	Markov state transition lifetime cost-per-QALY model comparing rivaroxaban, warfarin, aspirin plus clopidogrel, aspirin and no intervention. Cycle length 1-month.	No prevention: 3.33 Aspirin: 3.6 Aspirin plus clopidogrel:3.81 Warfarin: 3.93 Rivaroxaban: 4.08
Krejczy et al [Germany][37]	None stated	Lifetime cost-per-QALY Markov model for a hypothetical cohort of patients at 65 years of age at increased risk of stroke	Warfarin: 6.480 Dabigatran 110mg: 6.543 Dabigatran 150mg: 6.617 Apixaban: 6.617 Rivaroxaban: 6.541 Edoxaban: 6.543
Lantis et al [France][38]	Pfizer and Bristol-Myers Squibb	Markov lifetime 6-week cycle cost-per- QALY model comparing apixaban, dabigatran ,rivaroxaban, warfarin and aspirin: warfarin and apixaban optimal treatment choices, with apixaban cost-effective alternative to warfarin	Warfarin: 6.099 Dabigatran 110mg: 6.186 Dabigatran 150/110 mg switch: 6.220 Dabigatran 150mg: 6.221 Apixaban: 6.289 Rivaroxaban: 6.242
Harrington et al [US][39]	None stated	Lifetime cost-per-QALY Markov model for a hypothetical cohort of 70-year old patients at increased risk of stroke	Warfarin: 7.97 Dabigatran 150mg: 8.41 Apixaban: 8.47 Rivaroxaban: 8.26
Rognoni et al [Italy][40]	None stated	Decision tree combined with a lifetime cost-per-QALY Markov model with a 3-month cycle for a hypothetical cohort of 71-year olds differentiated by $CHADS_2$ score	$CHADS_2 \leq 1$ Warfarin: 11.133 Apixaban: 11.890 Dabigatran: 12.223 $CHADS_2 = 2$ Warfarin :8.764 Apixaban: 9.402 Dabigatran: 9.587 Rivaroxaban: 9.122 $CHADS_2 \geq 3$ Warfarin: 7.127 Apixaban:7.997 Dabigatran: 7.518 Rivaroxaban: 7.581

Kongnakorn et al [Belgium] [41]	Pfizer and Bristol-Myers Squibb	Lifetime Markov model with 6-week cycle	Incremental results vs. warfarin Dabigatran 110mg: 0.078 Dabigatran 110/150mg:0.118 Rivaroxaban: 0.132 Apixaban: 0.193
Comparative cost-effectiveness reviews : single NOACs			
Sorensen et al [42]	Boehringer Ingelheim International GmbH	Systematic review of dabigatran lifetime Markov cost-utility models to evaluate key model features to account for cost and QALY differences vs. warfarin	UK setting: incremental QALY Base case Kansal et al: 0.242 Base case Pink et al: 0.146 [43] US setting: incremental QALY Base case: Sorensen et al [44]: adapted to Freeman et al [45]/Shah and Gage [46]: 0.183 Base case: Freeman et al: 0.56 Base case: Shah and Gage: 0.25
Comparative cost effectiveness: dabigatran vs. warfarin			
You et al [US] [47]	None stated	Markov cost-utility model with 25 year horizon and a monthly cycle. Starting age 65 years	Genotype guided AC: 9.554 Usual AC: 9.444 Dabigatran 150mg:10.065 Dabigatran 110mg:10.026
Miguel et al [Portugal] [48]	Boehringer Ingelheim Lda	Lifetime Markov cost-utility model with 3-month cycle	Dabigatran: 0.439 gain (therapy before the age of 80 years) Dabigatran: 0.166 gain (therapy begins 80 years and over)
Kansal et al [Canada] [49]	Boehringer Ingelheim International GmbH	Lifetime Markov model with 3 month cycle and simulated age of 71 years (primary analysis – dabigatran vs. rivaroxaban; secondary analysis vs. warfarin.	Dabigatran: 6.167 Rivaroxaban: 6.015 Warfarin: 5.940
Comparative cost effectiveness: rivaroxaban vs. warfarin			
Morais et al [Portugal] [50]	None stated	Lifetime (20 year) Markov cost-utility model with a cycle length of 3 months. Mean starting age 73 years	Warfarin: 3.81 Rivaroxaban: 3.83
Mensch et al [Germany] [51]	None stated	Markov cost-utility model over 35 years with a hypothetical cohort of 65 year olds with at moderate to high risk of stroke with a 30-day cycle	Warfarin: 10.35 Rivaroxaban: 11.06
Kleintjens et al [Belgium] [52]	None stated	Markov lifetime cost-utility model with a hypothetical cohort mean age of 73 years	Warfarin: 10.51 Rivaroxaban: 10.62
Kourlaba et al [Greece] [53]	Bayer Hellas	Lifetime Markov model with 3-month cycle with starting age of 75 years.	Warfarin: 6.28 Rivaroxaban: 6.50
Comparative cost effectiveness: apixaban vs. warfarin/other NOACs			
Kamae et al [Japan] [54]	Bristol Myers KK and Pfizer Japan Inc.	Lifetime Markov model with 6-week cycle length (adapted from Dorian et al) comprising 17 health states. Starting age 70 years	Warfarin: 7.23 Apixaban: 7.47
Dorian et al. [UK] [55]	Pfizer and Bristol Myers Squibb	Lifetime Markov cost-utility model with 11 health states	Warfarin: 6.08 Apixaban: 6.26
Ademi et al. [Australia] [56]	Pfizer Australia	State transition Markov model with 1-year cycle comparing apixaban and warfarin. The model comprised 5 health states and seven transition states. Time horizon 20 years (effectively lifetime). Starting age 70 years	Warfarin: 5.84 Apixaban: 6.15
Lip et al [UK] [57]	Pfizer and Bristol-Myers Squibb	Lifetime Markov model with cycle of 6 weeks and 11 health states. Starting age 70 years	Apixaban: 6.26 Dabigatran 110mg: 6.16 Dabigatran 150mg: 6.19 Rivaroxaban: 6.21

Comparative cost effectiveness: edoxaban vs. warfarin			
Rognoni et al [Italy][58]	None	Lifetime Markov model with 3-month cycle	Warfarin: 8.425 Edoxaban: 9.022
Krejczy et al [Germany][59]	None	A 20-year Markov model with a cycle length of 1-year and a base-case hypothetical population with starting age of 65 years who were at increased risk for stroke.	Warfarin :7.48 Edoxaban 30mg: 7.65 Edoxaban 60mg 7.69 Warfarin: 7.64 Dabigatran 110mg 7.68 Dabigatran 150mg: 7.71 Warfarin 7.59 Rivaroxaban 20mg: 7.67 Warfarin 7.56 Apixaban 5mg: 7.75

The NICE Single Technology Appraisals

In each of the four NICE single technology appraisals, manufactures were required to meet the standards of the reference case. This mandates a framework consistent with the natural course of a chronic disease, which, in practical terms, means a lifetime horizon. Model endpoints are to be expressed in cost-per-QALY terms. Costs are from the perspective of the National Health Service (NHS) and Personal Social Services (PPS). QALYs are to be based upon community preference utilities with the EQ-5D the preferred instrument. Standard techniques are to be applied in the modeling to include Markov processes and the application of sensitivity and probabilistic sensitivity analysis. Costs and benefits are to be discounted at 3.5%. As noted above, there is no requirement for the simulated cost-outcomes projections to be presented in evaluable terms

.

Consistency with the reference case ensures the evidence presented to NICE is constructed. As such the projected claims are both untestable and immune to failure. The only basis for challenging the manufacturer's modeled or simulated claims is through an assessment of the 'reasonableness' or otherwise of the simulation given the evidence base for the modeled inputs and assumptions. This is achieved through an evidence group (ERG) review of the submission, focusing on the quality and scope of the clinical inputs and then a review of the structure, assumptions and inputs to the cost-utility model. Typically, there is a base-case model with supplementary sensitivity and sub-group evaluations to assess robustness. The ERG's review is then subject to a further review by the NICE appraisal committee. A draft guidance is then prepared, reviewed and a final guidance published.

The four manufacturer's submissions follow a similar model structure for all NOACs. In the case of apixaban, the manufacturer submitted a Markov cohort model comprising 18 health states including the absorbing state of death. Patients transitioned between health states in cycles of 6 weeks with only one clinical event permitted per cycle. It was noted that, although previous NOAC submissions had utilized a similar process, given the influence of individual patient characteristics on outcomes in atrial fibrillation, a discreet event simulation rather than a Markov cohort modeling approach may have been more appropriate. The Markov cohort, however, was accepted. The model projected the risk of events and treatment discontinuation from the within-trial into the post-trial period with the impact of treatment on the risk of events assumed to remain constant for the full model time horizon. The same approach is taken for the other NOAC models. Utility scores for the various health states, as they were not collected in the trials, were estimated from a systematic literature review. Unit direct medical costs were assumed to be unchanged over the time horizon of the model. Together with the constructed utility values they were discounted at 3.5%.

The internal validity of the model was evaluated through an extreme value analysis to identify any flawed algorithms or irregularities. Face validity was assessed by comparison of the model assumptions against published results. In this case, the projected results were validated against the constructed simulated results reported for dabigatran and rivaroxaban. The base-case model was presented as a fully incremental analysis between all considered interventions. The review group assessed the manufacturer's deterministic results against those estimated by the manufacturer's probabilistic sensitivity analysis. A summary of the results is presented in Table 2.

The results reported for the deterministic as opposed to the probabilistic projections are quite different. The former simulation has higher projected costs, claims for life years gained and QALYs gained, yet lower ICERs. In the deterministic model lifetime discounted costs range from £7,188 for warfarin to £8,983 for apixaban; in the

probabilistic model the range is from £5,331 for warfarin to £7,228 for apixaban. Corresponding life years gained are 7.469 to 7.614 and 6.869 to 7.002 for the deterministic as opposed to probabilistic model. Projected days gained over warfarin are slightly higher for the deterministic simulation. A similar pattern is reported for QALYs gained. In the deterministic model they are consistently (yet marginally) higher ranging from 5.696 to 5.860. Days gained over warfarin in the deterministic model range from 22.6 for dabigatran 110mg to 59.9 days gained for apixaban. Differences within the four NOACs identified, life days and QALYs gained are minimal. In the deterministic model life days gained over warfarin range from12.24 to 52.20 and for QALY days the range is 22.6 to 59.9.

In both the deterministic and probabilistic incremental results, dabigatran 110 mg is strictly dominated by dabigatran blend (150 mg/110 mg) with apixaban extendedly dominated rivaroxaban and dabigatran blend. Apixaban had an ICER versus warfarin of £11,008 and £16,852 in the deterministic and probabilistic incremental analyses, respectively.

Table 2

Lifetime Base-Case Modeled QALY Claims for NOACs: Nice Single Technology Appraisal for Apixaban

NOAC	Costs (£)	Life years gained	Life years gained over warfarin	Life days gained over warfarin	QALYs gained	QALYs gained over warfarin	QALY days gained over warfarin	Incremental QALYs gained**	ICER vs. warfarin (£)
Deterministic*									
Warfarin	7,188	7.469	-	-	5.696	-	-	-	-
Dabigatran (150/110mg)	8,437	7.537	0.068	24.48	5.788	0.092	33.6	0.091	13,648
Dabigatran (110 mg)	8,634	7.503	0.034	12.24	5.756	0.060	22.6	-0.032	25,308
Rivaroxaban	8,778	7.553	0.084	30.24	5.809	0.113	41.2	0.054	14,071
Apixaban	8,983	7.614	0.145	52.20	5.860	0.164	59.9	0.05	11,008
Probabilistic*									
Warfarin	5,331	6.869	-	-	5.303	-	-	-	-
Dabigatran (150/110mg)	6,737	6.921	0.052	18.72	5.342	0.039	14.24	0.04	36,4505
Dabigatran (110 mg)	6,832	6.889	0.030	10.80	5.321	0.018	6.6	-0.02	83,628
Rivaroxaban	7,070	6.943	0.074	26.60	5.366	0.063	23.0	0.05	27,565
Apixaban	7,228	7.002	0.133	47.90	5.416	0.113	41.2	0.05	16,852

Note: in the deterministic simulation mean values are used for each parameter; in the probabilistic simulation distributions are used
**versus the next least costly technology. Source: Ref 25

Discussion

All of the studies considered in this review modeled the comparative claims for the four NOACs and warfarin in a lifetime (or long-term) cost per quality adjusted life year (QALY) framework. Presumably, the authors of the various studies accepted the cost-utility NICE reference case as the appropriate standard. If we accept the scientific method and the evidentiary standards of normal science, this is an unfortunate choice of evaluation framework as it is impossible for the projected claims to be evaluated [17 19] . The choice of time horizon precludes testable claims. In this context, the claims are immune to failure. None of authors recognized this limitation and none attempted to generate testable claims or to suggest how the framework of the model might be utilized to develop and evaluate claims for these competing interventions. None of the authors cautioned as to the implications of generating non-testable claims for the credibility of the analysis. There was no consideration given as to how, if this was the intent of the analysis, these lifetime cost-per-QALY claims in respect of QALYs gained and incremental-cost-per-QALY estimates might be factored into formulary decisions.

Clearly, considerable effort was put into developing these modeled claims. Studies ranged from single comparator claims contrasting an individual NOAC against warfarin as well as those studies that attempted a comparison across competing NOACs with warfarin as the baseline comparison. Considerable attention was given to justifying the Markov structure, the cycle length and the number of health states captured in the Markov process. A number of studies extended the analysis from a deterministic to a probabilistic framework, presenting results for a deterministic base-case analysis to a range of probabilistic sensitivity analyses. The clinical input event data for both the individual NOACs and warfarin were typically referenced back to the pivotal RCTs and converted to rates that corresponded to the cycle length. Background mortality was based on country-specific life tables and mortality attributable to the specific clinical events was both included in most models. In all models considerable

detail was provided on the drug and event cost assumptions. These typically include drug acquisition costs and assumptions regarding the frequency and cost of physician visits. Given the focus on generating QALY estimates as the gold standard endpoint, considerable effort was also directed to justifying the utility weights to be attached to the various Markov health states. None of the pivotal trials included quality of life as a secondary efficacy outcome. Utility weights were generated, therefore, from literature reviews. There was considerable variation in the scope of the literature review.

A key point to note is that, in presenting modeled or simulated projections, irrespective of the inherent complexity and attractiveness of the model, the evidence presented (the claims for comparative advantage) is constructed. From the perspective of the agreed standards of normal science, the puzzling feature of these attempts to assess (and justify) the entrance and formulary acceptance of the NOACs is why they are presented in a framework that generates non-testable claims? And why, over some 6 or 7 years, have analysts persisted in replicating this approach? [13 14 15] It might also be reasonable to ask why journal editors have accepted these projections at face value? Projections based on the lifetime cost-per-QALY standards of the reference case have no possibility of ever being evaluated and challenged. It seems unreasonable that an editor would accept the projection as provisional, subject to the hope that it would be put in an evaluable form and the results of a future evaluation reported. If this is the case then it may go some way to providing an explanation for the backlog of cost-effectiveness projections that have been published over the past 20 years but never evaluated.

In the case of the NICE reference case it is apparent that there is no intention of evaluating the simulated claims [17 60]. There is no requirement in the reference case for the projections to be put in an evaluable form, supported by a protocol that details how the projected claims might be evaluated and reported back for further appraisal by NICE. Rather, the standards support, for chronic diseases, a lifetime horizon. This guarantees that the projections have to be taken at face value.

Justifying Lifetime Cost-Utility Models

If the NICE reference case is accepted as the gold standard for modeled claims, then we are, in effect, adopting a relativist position [61]. For the relativist, attempting to assess these studies from the perspective of normal science is to miss the point. There is no intention to generate claims that have the potential to be empirically evaluated. There is no intention to subscribe to the standards of normal science. The NICE reference case is more appropriately viewed as a construct, validated by the ERG process and the final adjudication by the NICE advisory committee to support threshold pricing decisions. The matching of projected ICERs against national notional willingness-to-pay thresholds in a number of the studies reviewed here confirms the acceptance of the NICE paradigm.

The relativist believes that all perspectives are equally valid. In the relativist's advocacy of the equivalence or symmetry principle, health care decisions are to be understood sociologically. No one body of evidence is superior to another. Results of a simulation are on an equal basis with those of a RCT. For the relativist the success of a scientific research program, in this case one of non-testable projections built on models and simulations, rests not on its ability to generate new knowledge but on its ability to mobilize the support of the technology assessment community. Basing decisions on models and simulations underpins the consensus view that evidence is constructed, never discovered. Instead of coming to grips with reality science is about rhetoric, persuasion and authority. Truth is consensus [62].

If we accept the relativist opinion and argue that decisions for formulary listing are most appropriately based on non-testable claims generated by models and simulations then we have to address the possibility that simulations can fail. Simulations or models are accepted because in the consensus view, the view of the authorities in the discipline, the ability to capture what they see as the critical or corresponding features of the reality of a decision is all that is required. If the simulated input conditions and the simulated core mechanism correspond to reality, the sufficient condition character of the simulation assures us that the output is necessarily entailed and predictions must corresponded to reality [62]. In the evaluation of apixaban described above, choice of a probabilistic framework as opposed to the manufacturer's choice of a deterministic model produces quite different projections [4]. Indeed, if we consider the models presented in Table 1 it is reasonable to infer that there is considerable latitude in the choice of model to support a product specific comparative QALY and cost-effectiveness claims based on willingness-to-pay thresholds. This is illustrated in the Sorensen et al paper which, in the case of dabigatran incremental QALY and cost-effectiveness claims [42]. While Sorensen et al concluded that the various models 'reached generally similar conclusions' in respect of the cost-effectiveness of dabigatran, there were some substantive differences in model design and in the inability to replicate the claimed QALY results.

It is unlikely ever to be agreement on correspondence, sufficiency and necessary entailment in lifetime cost-per-QALY models. Practitioners can agree that a Markov process is appropriate to capture the natural course of a disease, yet

disagree on the cycle length, the number of health states and transition probabilities that the model accommodates. Indeed, it is always possible to reverse engineer any simulation to generate competing claims. Apart from a probably fruitless debate over competing 'core' models, the assumptions driving the model and the validity of data that is trawled from the literature to populate the model there is no basis, apart from an appeal to experimentation that could distinguish one modeled or simulated claims from others. Rather than capturing the essence of a decision problem, the simulation captures the perception of the essence of the problem held by the authors of the simulation guided by a reference case or similar standards. We are asked to believe that it is possible capture the both present clinical reality and that reality will continue to unfold and present itself over succeeding decades. From the perspective of standards that are commonly accepted in evaluation and replication of evaluable claims, these models are best characterized as pseudoscience; intelligent design as opposed to natural selection [63].

Persistence, Adherence and Discontinuation

Three features of the pivotal trials for the four NOACs should be noted. First, the sheer numbers enrolled and the number of observations supporting the intention to treat analysis; second, the length of the respective trials; and third, the rates of discontinuation or persistence. In the case of rivaroxaban, 14,264 patients underwent randomization in the period December 2006 through June 2009; with study termination in May 2010 [3]. A total of 23.7% in the rivaroxaban arm and 22.2% in the warfarin arm permanently stopped their assigned therapy before an end-point event and before the study termination date, with 14.3% discontinuing in the first 12 months. The median duration of treatment exposure was 20 months.

In the case of dabigatran, 18,113 patients were enrolled and randomized between December 2005 and December 2007, with final follow-up visits between December 2008 and March 2009 [2]. At year 1, rates of discontinuation were 14.5% for dabigatran 110mg, 15.5% for dabigatran 150mg and 10.2% for warfarin (at 2 years the corresponding figures were 20.7%, 21.2% and 16.6%). Median duration of follow-up was 2 years. For apixaban 18,140 patients were enrolled between December 2006 and April 2010 [1]. Discontinuation before the end of the study was 25.3% for apixaban (3.6% due to death) and 27.5% for warfarin (3.8% due to death). Finally, in the case of edoxaban 21,026 patients were enrolled and randomized between November 2008 and November 2010. Median duration of treatment exposure was 907 days. In the warfarin arm 34.4% of patients permanently discontinued therapy prematurely and 33.6% for the two edoxaban arms (34.3% and 32.8%).

A recent assessment of rates of premature discontinuation in long-term NOAC clinical trials also points to a substantial proportion of patients discontinuing therapy prematurely [64]. Across ten long-term trials unscheduled cessation within the NOAC arm ranged from 11.3% to 34.4%, with discontinuation due to adverse events ranging from 1.7% to 17.2%. The duration of follow-up to assess these premature discontinuations ranged from 3 months to just under 3 years. These patterns of discontinuation are not, however, out of line with those for other vitamin K antagonists with randomized clinical trials. Chatterjee et al, from a meta-analysis of 18 randomized trials, found a risk ratio for discontinuation of NOACs versus other vitamin K antagonists for all causes in atrial fibrillation 1.01(95% CI 0.87 – 1.17) [65].

In the studies reviewed here, the practice was to apply discontinuation and average rates of adherence and persistence (where they are applied) generated by the long-term clinical trials supporting the individual NOACs. Apart from differences in the protocols supporting these trials, trial based estimates of discontinuation are unlikely to be a satisfactory basis for extrapolating modeled claims and constructing simulation models. While treatment switching and discontinuation of treatment is a feature of virtually all of the lifetime models, none of the models factor in observed patterns of adherence and persistence over the study time horizon. The Dorian et al model, for example, assumes that only patients experiencing stroke or MI were assumed to discontinue treatment permanently [55]. In a number of the models survivors of major bleeding were assumed to stop anticoagulation therapy and switch to aspirin. The Wu et al study for China made no mention of the need to accommodate adherence and persistence within the model, apart from mentioning that treatment administered in clinical practice might not be effective as one administered in clinical trials given lower levels of adherence [36]. The Miguel et al model of dabigatran versus warfarin similarly makes no mention of adherence of persistence with therapy [48]. The two Rognoni et al papers which focus on claim from the perspective of the Italian health system similarly make no mention of adherence or persistence in their simulations [40 58]. Models to explicitly consider all cause persistence and discontinuation are found in two warfarin versus rivaroxaban papers: the Morais et al model for Portugal and the Kourlaba et al model for Greece [50 53]. For the initial 3-month cycle, a discontinuation rate from the ROCKET trial of 8.9% was assumed for rivaroxaban and 8.0% for warfarin; in subsequent cycles it was 4.4% and 4.5% respectively.

A recent study by Yao et al compares adherence patterns for warfarin with those for rivaroxaban, dabigatran and apixaban and their impact on risk of stroke and major bleeding [66]. During a median follow up of 1.1 years, only 47.5% of NOAC patients were adherent, defined as a medication possession

ratio (MPR) of ≥ 80%. Adherence to warfarin was 40.2%. Apixaban had the highest unadjusted adherence (61.9%) and dabigatran the lowest (38.5%). The rivaroxaban rate was 58.4%. Applying a multivariate logistic regression, adjusted adherence rates were 38.7% for warfarin and 47.5% for all NOACs. Higher rates of adherence were found across all treatments for those at higher risk. For those with a $CHAa_2DS_2$-VASc ≥ 4 the warfarin adherence was 53.4% and the average for the NOACs 59.8%.

Persistence with NOACs has been reported in three recent observational studies. Forslund et al utilizing data from the administrative health register of the Stockholm region evaluated crude and adjusted persistence from the index OAC prescription. In the period April 2011 to December 2014 [67]. At the end of the first year crude overall persistence was 88.2% and 82.9% at the end of the second year. Persistence with warfarin at the end of the first year was 85.0%, apixaban 85.9%, dabigatran 74.4% and rivaroxaban 77.4%. In the UK, Martinez et al reported on persistence with longitudinal data from the Primary Care Clinical Practice Research Datalink between January 2011 and May 2014 [68]. Persistence with warfarin at the end of the first year was 63.6$ and 79.2% for all NOACs.

In Germany, Beyer-Westendorf et al reported persistence from primary care patients at 180 days of 66.0% for rivaroxaban, 60.3% for dabigatran and 58.1% for VKA. At 1 year corresponding persistence estimates were 53.1%, 47.3% and 25.45% respectively. An MPR ≥ 0.8 was found for 61.4% of rivaroxaban and 49.5% of dabigatran patients [69].

Longer term studies suggest that by 3 years from index prescriptions no more than 30% of patients met the standard of ≥ 80% days covered. There are limited data for longer periods. Experience in Australia, for example, in the period November-December 2013 to March 2015 with records from the Pharmaceutical Benefits Scheme reported by Simons et al found that for index prescriptions in a sample of 1,471 atrial fibrillation patients with a mean age of 76 years on NOACs and 74 years on warfarin found that 62% discontinued within 12 months [70]. The corresponding figure for NOACs was 30%. Only 9% of those on NOACs failed to pick up the first repeat prescription compared to 14% of those on warfarin.

Overall, these estimates suggest that by the end of one year after the index prescription persistence with warfarin is in the range 60 to 70% with a corresponding NOAC rate of 70 to 80%. By the end of year 2, persistence is likely to be 15 to 20% lower. Beyond two years is sheer speculation, although it would not be unreasonable, given evidence for persistence in other chronic disease states that the overwhelming majority of patients have discontinued within 3 to 4 years. Given the age at which treatment is usually initiated for atrial fibrillation, deaths to patients need to be factored in to persistence estimates. In the edoxaban pivotal trial, for example, 10-8% of patients died before the end of the trial. Under reasons for discontinuation death was given in 3.1% of warfarin patients and 2.8% of edoxaban patients.

Adherence patterns add a further dimension. Although not captured in any of the pivotal trials, other than reporting time in therapeutic range for warfarin, adherence is a potentially important offset to claimed therapy gains. While it is often difficult in observational studies to distinguish adhere from discontinuation, their additive effect on projected model end-points in therapy could be significant.

If the majority of patients initiated to an OAC have discontinued therapy, for event related reasons, non-event related reasons and death within 3 to 4 years of their initial prescription, then it seems rather odd to focus on creating projections for discounted direct medical costs and utilities over a lifetime horizon. Instead of modeling switching and discontinuation behavior over the lifetime of a treatment cohort, a more practical and useful approach would be to recognize the likelihood of early discontinuation and generate comparative predictions for events and costs for a meaningful timeframe that can be captured from existing data sources as feedback to formulary committees.

Quality Adjusted Life Years

While application and acceptance of the reference case model is understandable in the case of NICE, it is puzzling as to why authors continued to 'replicate' cost-per-QALY models when the simulated NOAC gains were minimal and the only points of reference were other non-evaluable claims simulations that adopted lifetime Markov processes. The QALY standard for constructing endpoints was never questioned. There was no consideration of alternative endpoints that might be more clinically appropriate as a basis for differentiating the OACs. There is no reason why QALYs should be accepted as the outcomes gold standard. The argument that QALYs allow us to compare interventions across disease states and therapeutic areas loses much of it impact when it is pointed out that QALYs are seldom recorded in electronic medical records (EMRs) and never in administrative claims and other 'big' data sets. The effect of expressing claims in cost-per-QALY terms may simply be to erect one more barrier to evaluating claims. It we accept that lifetime cost-per-QALY models (or in these evaluations costs-per- quality adjusted life day) are the gold standard, then there is no chance that such modeling exercises will generate anything other than an accumulation of constructed and non-testable projections.

In the absence of any commitment to falsification and replication of experimental claims, there is no basis for validating QALY projections other than through re-evaluating

simulation input assumptions, model structure and matching projections against those from other simulations. The latter comparisons illustrate the sensitivity of constructed cost-utility claims to individual model structures and assumptions. In the case of dabigatran, as noted above, the assessment by Sorensen et al undertakes a quantitative comparison between a reference model and two US and two UK simulations. Assumptions that were compared included the lifetime modelling approach, the defined patient population and their baseline characteristics, treatment effectiveness, projected future risk of stroke and bleeding events, cost and utility inputs, drug switching and discontinuation. In the UK setting differences in the constructed cost-effectiveness results were attributed primarily to the cost and utility model inputs (e.g., assumed cost and quality of life following intracranial haemorrhage), while in the US setting the differences went beyond unit cost and utility values to adjustments in the risk for intracranial haemorrhage and ischaemic stroke as the modelled cohorts aged and accommodation of discontinuation due to non-adherence.

If QALY claims are impossible to evaluate experimentally, then there is the option of falling back on clinical endpoints: non-fatal ischaemic strokes, fatal strokes, non-fatal hemorrhagic strokes and deaths. These endpoints are currently accessible from claims data and from electronic medical records [71]. A recent presentation, for example, to the American College of Cardiology in April 2016 reports on a comparison of differences between rivaroxaban, dabigatran and apixaban in preventing stroke or systemic embolism in patients with atrial fibrillation [72]. Utilizing data from the Optum Labs Data Warehouse in the US the study captured data from 160,328 patients who had in initiated therapy between October 2010 and February 2015. Three propensity score matched cohorts based on 19 clinical variables were constructed: rivaroxaban vs dabigatran n = 31,574; apixaban vs. dabigatran n=12,084; and apixaban vs. rivaroxaban n=13,130. The results of the assessment using Cox proportional hazard models for an intention to treat analysis found that there was no statistically significant differences in any of the pairwise comparisons. The only advantage observed was a reduced risk of bleeding and major bleeding events with apixaban compared to dabigatran and rivaroxaban. Rivaroxaban increased major bleeding.

It is not clear, therefore, what is to be gained from quantitative, non-experimental assessments of competing constructed cost-utility claims in atrial fibrillation. The fact that the cost-utility claims generated by these various simulations claims were incapable of ever being evaluated suggests that the exercise is probably pointless. Rather than focusing on lifetime cost-utility simulations, the construction and comparison of imaginary NOAC worlds, reviews of simulated models might be more productive if they addressed the issue of testable claims for the competing OACs and the modeling or simulation frameworks that might support such claims.

The fundamental objection to simulations driving decision making is that there is no opportunity to generate feedback. Health systems that have relied on reference case modeling and thresholds to support formulary decisions have no idea, in the absence of experimental data, whether those claims are right or even if they are wrong. Indeed, as the pipeline for new entrants to the atrial fibrillation marketplace is crowded, what are the implications for the simulation of new compounds, consequent patient switching and the likely continuing place on formulary of the NOACs? A simulation that took an adherence profile into account might conclude that the QALY benefits accrue primarily in the first 3 to 4 years of therapy and to extend to model a further 10 to 20 years adds little to overall claims for comparative benefits and costs when it is likely that patients who persist on therapy will switch to new compounds.

Conclusions

If the arguments presented here are accepted then, from the perspective of experimental evidence supporting formulary decisions, the lifetime cost-per-QALY reference case is a redundant standard. The reference case supports the development of simulation models where the projections are immune to failure. The only challenge is from competing simulations. Rather than accepting the standards of falsification and replication, these standards are put to one side. The process of the discovery of new facts is not only put to one side, it is irrelevant.

This does not mean that models or simulated claims are incapable of generating or suggesting testable hypotheses that conform to the standards of normal science. Given the evidence for anticipated adherence and persistence patterns, attempts to generate claims for the competing merits of the NOACs and potential benefits from switching from warfarin could focus on the more immediate short term. Rather than attempting to develop lifetime models, time horizons of, at most, 3 to 4 years should be considered for competing cost-effectiveness claims. These would capture the baseline risk of stroke to capture relative treatment effects, generate specific claims for the primary efficacy outcome of reducing stroke and other potential adverse events, evaluate claims for decreasing efficacy of NOACs with increasing creatinine clearance, potentially capturing QALYs and identify key cost components. Protocols could be developed by manufacturers to detail how the clams would be assessed utilizing observational data and the time horizon for reporting (and publishing) the results. At the same time, manufacturers might suggest alternative intervention strategies to support adherence behavior. If the objective is to reduce the risk of

stroke and bleeding, the clinical benefits of adherence strategies may far outweigh attempts to demonstrate fairly trivial and imaginary QALY benefits between the competing NOACs and warfarin. The key point is that claims can be evaluated and the outcomes reported back to formulary committees in a timely manner.

Acknowledgement: The authors would like to thank Professor Jon Schommer for his support in preparing this commentary

References

[1] Giugliano RP Ruff CT, Braunwald E et al. Edoxaban versus warfarin in patients with atrial fibrillation. *New Eng J Med. 2013*;369:2093-2104

[2] Connolly SJ, Ezekowitz MD, Yusuf S et al. Dabigatran versus warfarin in patients with atrial fibrillation. *New Eng J Med*. 2009;361:1139-1151

[3] Patel MR, Mahaffey KW, Garg J et al. Rivaroxaban versus warfarin in nonvalvular atrial fibrillation. *New Eng J Med*. 2011;365:883-891

[4] Granger CB, Alexander JH, McMurray JJ et al. Apixaban versus warfarin in patients with atrial fibrillation. *New Eng J Med*. 2011;365:981-992

[5] Popper KR. The logic of scientific discovery. New York: Harper, 1959

[6] Popper KR. Objective knowledge (Rev. Ed.). Oxford: Clarendon Press, 1979

[7] Langley PC. The status of modeled claims. *J Med Ec*on. 2015;18(12):991-992

[8] Langley PC. Validation of modeled pharmacoeconomic claims in formulary submissions. *J Med Econ*. 2015;18(12):993-99

[9] Editorial. *Nature*. 2016;529:256

[10] Centre for Evidence Based Medicine. COMPare Project. Oxford, 2016. http://compare-trials.org/results

[11] Langley PC. He ao pohewa: The PHARMAC Prescription for Pharmacoeconomic Analysis in New Zealand and the standards of normal science. *INNOVATIONS in Pharmacy*. 2016;7(2): No. 13. http://pubs.lib.umn.edu/innovations/vol7/iss2/13

[12] Langley PC. Nullius in verba: the University of Minnesota Social and Administrative Pharmacy program proposed Guidelines for Formulary Evaluations. *INNOVATIONS in Pharmacy*. 2016;7(2): Article 10 http://pubs.lib.umn.edu/innovations/vol7/iss2/10

[13] Langley PC**.** Imaginary worlds: Modeled claims for cost-effectiveness published in *PharmacoEconomics* January 2015 to December 2015. *INNOVATIONS in Pharmacy*. 2016;7(2): Article 9 http://pubs.lib.umn.edu/innovations/vol7/iss2/9

[14] Langley PC, Rhee TG. Imaginary worlds: A systematic review of the status of modeled cost-effectiveness claims published in the Journal of Medical Economics from January 2015 to December 2015. *INNOVATIONS in Pharmacy*. 2016;7(2): Article 16 http://pubs.lib.umn.edu/innovations/vol7/iss2/16

[15] Langley PC, Rhee TG. Imaginary worlds: The status of modeled economic evaluation claims published in *Value in Health* January 2015 to December 2015. *INNOVATIONS in Pharmacy*. 2016;7(2): Article 18 http://pubs.lib.umn.edu/innovations/vol7/iss2/18

[16] Langley PC. Modeling imaginary worlds: Version 4 of the AMCP Format for Formulary Submissions. *INNOVATIONS in Pharmacy*. 2016;7(2): Article 11 http://pubs.lib.umn.edu/innovations/vol7/iss2/11

[17] Langley PC**.** Sunlit uplands: the genius of the NICE reference case. *INNOVATIONS in Pharmacy*. 2016;7(2): Article 12 http://pubs.lib.umn.edu/innovations/vol7/iss2/12

[18] Langley PC. He ao pohewa: The PHARMAC Prescription for Pharmacoeconomic Analysis in New Zealand and the standards of normal science. *INNOVATIONS in Pharmacy*. 2016;7(2): Article 13 http://pubs.lib.umn.edu/innovations/vol7/iss2/13

[19] Langley PC. Great expectations: Cost-utility models as decision criteria. *INNOVATIONS in Pharmacy*. 2016;7(2): Article 15 http://pubs.lib.umn.edu/innovations/vol7/iss2/14

[20] Langley PC. Supporting formulary decisions: The discovery of new facts or constructed evidence? *INNOVATIONS in Pharmacy*. 2016;7(2): Article 15 http://pubs.lib.umn.edu/innovations/vol7/iss2/15

[21] Moher D, Shamseer L, Clarke E et al. Preferred reporting items for systematic reviews and met-analyses protocol (PRISMA-P) 2015 statement. *Systematic Reviews*. 2015;4:1 DOI: 10.1186/2046-4053-4-1

[22] National Institute for Health and Care Excellence. Dabigatran etexilate for the prevention of stroke and systemic embolism in atrial fibrillation. (NICE technology appraisal guidance 249). March 2012.

[23] National Institute for Health and Care Excellence. Rivaroxaban for the prevention of stroke and systemic embolism in people with atrial fibrillation. (NICE technology appraisal guidance 256). May 2012.

[24] National Institute for Health and Care Excellence. Apixaban for preventing stroke and systemic embolism in people with non-valvular atrial fibrillation. (NICE technology appraisal guidance 275). February 2013.

[25] National Institute for Health and Care Excellence. Edoxaban for preventing stroke and systemic embolism in people with non-valvular atrial fibrillation. (NICE technology appraisal guidance 355). SEptember 2015.

[26] National Institute for Health and Care Excellence. Guide to the methods of technology appraisal. London, NICE April 2013

[27] New Zealand. Pharmaceutical Management Agency. Prescription for Pharmacoeconomic Analysis: Methods for cost-utility analysis (v 2.1). PHARMAC, 2012. https://www.pharmac.govt.nz/assets/pfpa-final.pdf

[28] Ireland. Health Information and Quality Authority. Guidelines for the Economic Evaluation of Health Technologies in Ireland. Dublin: HIQA; 2014

[29] Academy of Managed Care Pharmacy. Format for Formulary Submissions. AMCP: April 2016

[30] Caro JJ, Briggs A, Siebert U et al. Modeling good research practices – Overview: A report of the ISPOR-SMDM Modeling good practices task force – 1. *Value Health*. 2012 (15:796-803

[31] Eddy DM, Hollingworth W, Caro JJ et al. Model transparency and validation: A report of the ISPOR-SMDM modeling good research practices Task Force – 7. *Value Health*. 2012;15:843-850

[32] Husereau D, Drummond M, Petrou S et al. Consolidated health economics evaluation reporting standards (CHEERS) statement. *J Med Econ*. 2013;16(6):713-19

[33] Jarungsuccess S, Taerakun S. Cost-utility analysis of oral anticoagulants for nonvalvular atrial fibrillation patients at the Police General Hospital, Bangkok, Thailand. *Clin Ther*. 2014;36(10):139-1394

[34] Coyle D, Coyle K, Cameron C et al. Cost-effectiveness of new oral anticoagulants compared with warfarin in preventing stroke and other cardiovascular events in patients with atrial fibrillation. *Value Health*. 2013;16:498-506

[35] Vestergaard AS, Ehlers LH. A health economic evaluation of stroke prevention in atrial fibrillation: Guideline adherence versus the observed treatment strategy prior to 2012 in Denmark. *PharmacoEconomics*. 2015;33:967-979

[36] Wu B, Kun L, Liu X et al. Cost-effectiveness of different strategies for stroke prevention in patients with atrial fibrillation in a health resource-limited setting. *Cardiovasc Drugs Ther*. 2014;28:87-98

[37] Krejczy M, Harenberg J, Marx S et al. Comparison of anticoagulation with dabigatran, rivaroxaban and apixaban in patients with non-valvular atrial fibrillation across countries. *J Thromb Thrombolysis*,. 2014;37:507-23

[38] Lantis T, Cotté FE, Gaudin AF et al. Stroke prevention in patients with atrial fibrillation in France: comparative cost-effectiveness of new oral anticoagulants (apixaban, dabigatran, and rivaroxaban), warfarin and aspirin. *J Med Econ*. 17(8):587-98

[39] Harrington AR, Armstrong EP, Nolan PE et al. Cost-effectiveness of apixaban, dabigatran, rivaroxaban, and warfarin for stroke prevention in atrial fibrillation. *Stroke*. 2013 DOI:10.1161/STROKEAGHA.111

[40] Rognoni C, Marchetti M, Quaglini S et al. Apixaban, dabigatran, and rivaroxaban versus warfarin for stroke prevention in non-valvular atrial fibrillation: a cost-effectiveness analysis. *Clin Drug Investig*. 2014;34:9-17

[41] Kongnakorn T, Lantis T, Annemans L et al. Stroke and systemic embolism prevention in patients with atrial fibrillation in Belgium: comparative cost-effectiveness of new oral anticoagulants and warfarin. *Clin Drug Investig*. 2015;35:109-119

[42] Sorensen SV, Peng S, Monz BU et al. A comparative analysis of models used to evaluate the cost-effectiveness of dabigatran versus warfarin for the prevention of stroke in atrial fibrillation. *PharmacoEconomics*. 2013;31:589-604

[43] Pink J, Lane S, Pirmohamed M et al. Dabigatran etexilate versus warfarin in management of non-valvular atrial fibrillation in Uk context: quantitative benefit-harm and economic analyses. *BMJ*. 2011;343:d6333

[44] Sorensen SV, Kansal AR, Connolly S et al. Cost-effectiveness of dabigatran etexilate for the prevention of stroke and systemic embolism in atrial fibrillation: a Canadian payer perspective. *Thromb Haemost*. 2011;105(5):908-19

[45] Freeman JV, Zhu Rp, Owens DK. Cost-effectiveness of dabigatran compared with warfarin for stroke prevention in atrial fibrillation. *Ann Intern Med*. 2011;154(1):1-11

[46] Shah SV, Gage BF. Cost-effectiveness of dabigatran for stroke prophylaxis in atrial fibrillation. *Circulation.* 2011;123(22):2562-70

[47] You J, Tsui K, Wong R et al. Cost-effectiveness of dabigatran versus genotype-guided management of warfarin therapy for stroke prevention in patients with atrial fibrillation, *Plos One*. 2012;7(6):e39640

[48] Miguel LS, Rocha E, Ferreira J. Economic evaluation of dabigatran for stroke prevention in patients with non-valvular atrial fibrillation. *Rev Port Cardiol*. 2013;32(7-8):557-565

[49] Kansal AR, Sharma M, Bradley-Kennedy C et al. Dabigatran versus rivaroxaban for the prevention of stroke and systemic embolism in atrial fibrillation in Canada. Comparative efficacy and cost-effectiveness. *Thromb Haemost*. 2012;108(4):672-82

[50] Morais JM Aguilar C, McLeod E et al. Cost-effectiveness of rivaroxaban for stroke prevention in atrial fibrillation in the Portuguese setting. *Rev Port Cardiol*. 2014;33(9):535-544

[51] Mensch A, Stock S, Stollenwerk B et al. Cost effectiveness of rivaroxaban for stroke prevention in German patients with atrial fibrillation. *PharmacoEconomics*. 2015;33:271-283

[52] Kleintjens J, Li X, Simoens S et al. Cost-effectiveness of rivaroxaban versus warfarin for stroke prevention in atrial fibrillation in the Belgian healthcare setting. *Pharmacoeconomics*. 2013;31:909-918

[53] Kourlaba G, Maniadakis N, Andrikopoulos et al. Economic evaluation of rivaroxaban in stroke prevention for patients with atriual fibrillation in Greece. *Cost Effectiveness and Resource Allocation*. 2014;12:5

[54] Kamae I, Hashimoto Y, Koretsune Y et al. Cost-effectiveness analysis of apixaban anainst warfarin for stroke prevention in patients with nonvalvular atrial fibrillation in Japan. *Clin Thera*. 2015;37(12):2837-2851

[55] Dorian P, Kongnakorn T, Phatak H et al. Cost-effectiveness of apixaban vs. current standard of care for stroke prevention in patients with atrial fibrillation. *Eur Heart J*. 2014;35:1897-1906

[56] Ademi Z, Pasupathi K, Liew D. Cost-effectiveness of apixaban compared to warfarin in the management of atrial fibrillation in Australia. *Eur J Prev Cardiol*. 2015;22(3):344-53

[57] Lip G, Kongnakorn T, Phatak H et al. Cost-effectiveness of apixaban versus other new oral anticoagulants for stroke prevention in atrial fibrillation. *Clin Ther*. 36(2):192-210

[58] Rognoni C, Marchetti M, Quaglini S et al. Edoxaban versus warfarin for stroke prevention in non-valvular atrial fibrillation: a cost-effectiveness analysis. *J Thromb Thrombolysis*. 2015;39:149-154

[59] Krejczy M, Harenberg J, Wehling M et al. Cost-effectiveness of anticoagulation in patients with nonvalvular atrial fibrillation with edoxaban compared to warfarin in Germany. *Biomed Res Int*. 2015: ID 876923

[60] Belsey J. Predictive validation of modeled health technology assessment claims: lessons from NICE. *J Med Econ*. 2015;18(12):1007-12

[61] Wootton D. The Invention of Science. New York: Harper, 2015

[62] Grim P, Rosenberger R, Eason RE et al. How simulations fail. *Synthese*. 2011 DOI 10 1007/s11229-011-9976-7

[63] Pagliucci M. Nonsense on Stilts: How to tell science from bunk. Chicago: University of Chicago Press, 2010

[64] Suryanarayan D, Schulman S. When the rubber meets the road: Adherence and persistence with non-vitamin K antagonist oral anticoagulants and old oral anticoagulants in the real world – a problem or a myth. Semin Thromb Hemost. 2014;40:852-59

[65] Chatterjee S, Sardar P, Giri JS et al. Tratment disctinuations with new oral agents for long-term anticoagulation: Insights from a meta-analysis of 18 randomized trials including 101, 801 patients. Mayo Clin Proc. 2014;89(7):896-907

[66] Yao X, Abraham NS, Alexander GC et al. Effect of adherence to oral anticoagulants on risk of stroke and major bleeding among patients with atrial fibrillation. *J Am Heart Assoc*. 2016;5:e003074 doi: 10.1161/JAHA.115.003074

[67] Forslund T, Wettermark B, Hjemdahl PO. Comparison of treatment persistence with different oral anticoagulants in patienbts with atrial fibrillation. *Eur J Clin Pharmacol*. 2016;72(3):329-38

[68] Martinez C, Katholing A, Wallenhorst C et al. Therapy persistence in newly diagnosed non-valvular atrial fibrillation treated with warfarin or NOAC. A cohort study. *Thromb Haemost*. 2015;115(1):31-9

[69] Beyer-Westendorf J, Enliken B, Evers T. Real-world persistence and adherence to oral coagulation for stroke risk reduction in patients with atrial fibrillation. *Europace*. 2016 Jan 31. Pii:euv421

[70] Simons LA, Ortiz M, Freedman SB et al. Improved persistence with non-vitamin K oral anticoagulants compared with warfarin in patients with atrial fibrillation: recent Australian experience. *Curr Med Res Opin*. 2016; Jul 27:1-14 (epub]

[71] Wasser T, Haynes K, Barron J, Cziraky M. Using 'big data' to validate claims made in the pharmaceutical approval process. *J Med Econ.* 2015;18(12):1013-1019

[72] MedPage Today (5 April 2016). NOACs similarly effective for stroke prevention. http://www.medpagetoday.com/clinical-context/Strokes/57142