Comparative Effectiveness Research Review Disposition of Comments Report

Research Review Title: Stroke Prevention in Atrial Fibrillation

Draft review available for public comment from August 31, 2012, to September 28, 2012.

Research Review Citation: Lopes RD, Crowley MJ, Shah BR, Melloni C, Wood KA, Chatterjee R, Povsic TJ, Dupre ME, Kong DF, Barros e Silva PGM, Santos MHH, Armaganijan LV, Katz M, Kosinski A, McBroom AJ, Chobot MM, Gray R, Sanders GD. Stroke Prevention in Atrial Fibrillation. Comparative Effectiveness Review No. 123. (Prepared by the Duke Evidence-based Practice Center under Contract No. 290-2007-10066-I.) AHRQ Publication No. 13-EHC113-EF. Rockville, MD: Agency for Healthcare Research and Quality; August 2013. www.effectivehealthcare.ahrq.gov/reports/final.cfm.

Comments to Research Review

The Effective Health Care (EHC) Program encourages the public to participate in the development of its research projects. Each comparative effectiveness research review is posted to the EHC Program Web site in draft form for public comment for a 4-week period. Comments can be submitted via the EHC Program Web site, mail or email. At the conclusion of the public comment period, authors use the commentators' submissions and comments to revise the draft comparative effectiveness research review.

Comments on draft reviews and the authors' responses to the comments are posted for public viewing on the EHC Program Web site approximately 3 months after the final research review is published. Comments are not edited for spelling, grammar, or other content errors. Each comment is listed with the name and affiliation of the commentator, if this information is provided. Commentators are not required to provide their names or affiliations in order to submit suggestions or comments.

The tables below include the responses by the authors of the review to each comment that was submitted for this draft review. The responses to comments in this disposition report are those of the authors, who are responsible for its contents, and do not necessarily represent the views of the Agency for Healthcare Research and Quality.

Source: http://effectivehealthcare.ahrq.gov/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productID=1668Insert URL here
Published Online: August 23, 2013

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #1	General: Quality of the Report	Superior	Thank you
Peer Reviewer #2	General: Quality of the Report	Fair	Acknowledged
Peer Reviewer #3	General: Quality of the Report	Good	Thank you
Peer Reviewer #4	General: Quality of the Report	Good	Thank you
Peer Reviewer #5	General: Quality of the Report	Superior	Thank you
Peer Reviewer #6	General: Quality of the Report	Fair	Acknowledged
Peer Reviewer #7	General: Quality of the Report	Good	Thank you
TEP #1	General: Quality of the Report	Superior	Thank you
TEP #2	General: Quality of the Report	Good	Thank you
TEP #3	General: Quality of the Report	Fair	Acknowledged
TEP #4	General: Quality of the Report	Good	Thank you
TEP #5	General: Quality of the Report	Superior	Thank you
TEP #6	General: Quality of the Report	Good	Thank you
TEP #7	General: Quality of the Report	Superior	Thank you

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #2	General: Clarity and usability	The review concludes that CHADS2 and CHADSVASC are similar in their predictive ability for stroke. However, they do not provide the user with any information about the thresholds (in each scoring system) where the risks of anticoagulant therapy may outweigh the benefits. In their future research section, the authors should provide a clear statement about the need for clinical decision-making tools that will allow care-givers to easily weight the risks and benefits of anticoagulant therapy in a particular patient with AF.	Thank you. We have added a statement in the Discussion conclusion regarding the need for tools where the bleeding may outweigh the benefits and systems that could be used to deploy these tools. We have also added the thresholds for which oral anticoagulants are recommended for CHADS2 and CHA_2DS_2-VASc scores.
Peer Reviewer #3	General: Clarity and usability	Other than being somewhat verbose and repetitious, as noted above, the report was well and logically structured. Not a fault of the study, but there were few conclusions that will change practice. From a policy perspective, beyond a solid summary of the literature, the report was most helpful in identifying current knowledge gaps. Hopefully this will inform research priorities with funding agencies.	We thank the reviewer for their comments.
Peer Reviewer #4	General: Clarity and usability	All yes. this should be very helpful	Thank you
Peer Reviewer #5	General: Clarity and usability	The report is well-structured and organized but is very long. It can definitely serve as reference resource for policy and or practice decisions, but the succint take aways or key points for practicing clinicans are more limited. Some more succinct key statements or conclusions for each of the KQ could be helpful.	Comparative Effectiveness Reviews (CERs) of this sort are typically long and technical in nature. We hope the Executive Summary and the Key Points in the revised CER provide more succinct summaries for practicing clinicians.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #6	General: Clarity and usability	The authors should be congratulated on this exhaustive review. However, I would hesitate to make any strong recommendations based on the fair/poor methodological rigor of the many observational studies included. How does one inform policy from studies that did not include independent adjudication of events, but instead relied on administrative databases? or studies without validation of critical covariates, without assurance that the same definitions were used across studies, without verification that key covariates were uniformly available and ascertained (aspirin, fall risk, prior TIA, blood pressure measurements, etc.) in each of the studies included. This review constitutes a very loud alarm for robust, methodologically rigorous studies to truly inform patient care.	We agree that there is a need for additional rigorous studies in this field, and we have attempted to moderate our conclusions accordingly, particularly with respect to the KQs that were informed by observational evidence (e.g. KQ 1 and KQ 2). We have included discussion of the overall limitations of the evidence based in the future research needs section.
Peer Reviewer #7	General: Clarity and usability	I think the authors have been reasonably clear within the constraints of an AHRQ report, which as I've said above, is simply very poor format indeed in requiring as much as it does.	We thank the reviewer for their comment.
TEP #1	General: Clarity and usability	Yes.	Thank you
TEP #2	General: Clarity and usability	As with many of these reviews, a major challenge in terms of clarity and usability is the sheer number of comparisons and outcomes evaluated. e.g. KQ 3 is essentially one comparison and outcome after another. I think from the reader's perspective it would probably be more helpful to have the results summarized in a table and the actual text focus on important findings.	We thank the reviewer for their suggestion and feel that the strength of evidence tables which are included in each key question and then summarized in a more concise form in the Executive Summary already serve to synthesize the expansive evidence. Although extensive, we feel that the included detail about comparisons and outcomes reported in the sections is needed to serve the needs of diverse stakeholders and users of the report.

Commentator & Affiliation	Section	Comment	Response
TEP #3	General: Clarity and usability	Some sections are just confusing, for example, mixing stroke scores into bleeding risk assessment. Also, more text on identification of low risk patients patients with AF	Thank you for the comments. We have tried to make these sections clearer. As part of the review, we found numerous studies that examined the utility of stroke risk scores, such as $CHADS_2$ and CHA_2DS_2-VASc, to also predict bleeding risk. We initially felt it was important to evaluate these studies in KQ 2. However, given that these stroke risk scores are not used clinically for bleeding prediction, we have decided to remove these from the KQ 2 section to assure clarity for readers. Additionally, we provided data of these risk scores with regard to identification of low-risk patients. While we agree that these are important patients to highlight, the critical gaps and treatment dilemmas occur in the higher risk populations; thus, the proportionally higher attention in this review to the identification and treatment regimens in this group.
TEP #3	General: Clarity and usability	Suggest you need to read the 2012 ESC focused update AF guideline which is currently state of the art	Thank you for this suggestion. We now refer to the European (ESC) guidelines in the Discussion section for KQ 3 and discuss their incorporation of novel antithrombotics and what they currently recommend, particularly for patients at lower risk for thromboembolic events.
TEP #4	General: Clarity and usability	Yes.	Thank you
TEP #5	General: Clarity and usability	The report is well structured and Key Questions are detailed in a logical sequence. Conclusions are valid and applicable to policy decision making.	Thank you
TEP #5	General: Clarity and usability	For KQ 1 continuous CHADS2 and CHA2DS2-VASc were most predictive of stroke events as compared to the Framingham Categorical Score. In a meta analysis of the reviewed literature CHADS2 Continuous score and CHA2DS2-VASc Continuous score had comparable predictive ability for stroke risk (0.71) and greater predictive ability than other scores. For KQ 2 HAS-BLED tool has the highest predictive accuracy for bleeding events in AF patients on warfarin, off antithrombotic therapy, and on aspirin.	Acknowledged

Commentator & Affiliation	Section	Comment	Response
TEP #5	General: Clarity and usability	For KQ3 for interventions for preventing thromboembolic events, triple therapy increases fatal and non-fatal bleeding compared to warfarin. Dabigatran at 150mg dose was superior to warfarin, and non inferior at 110mg, in reducing the incidence of composite stroke or systemic embolism. For KQ4 the literature is too small to draw any conclusion regarding a clinically important issue of anticoagulation strategies in patients undergoing invasive procedures. Similarly there was lack of literature regarding KQ 5 and KQ 6.	Acknowledged
TEP #6	General: Clarity and usability	As mentioned previously, this report is very well written and nicely structured. On those key questions which have been addressed explicitly in this report will provide valuable information health policy makers. The authors should be commended for taking such a difficult task in such an important patient population.	Thank you
TEP #7	General: Clarity and usability	The report is very clearly structured and clinically useful. I strongly believe that this document will be very useful both for clinical practice and for formulation of policy related to stroke prevention in atrial fibrillation	Thank you
Peer Reviewer #1	General	The report is clinically meaningful. The target population and audience are defined. The key questions are appropriate and explicitly stated.	Thank you
Peer Reviewer #2	General	Throughout the report, the authors use the terms “Factor II” and “thrombin” interchangeably. To be correct, the authors should know that Factor II is PROthrombin and Factor IIa is THROMBIN. Dabigatran is a direct inhibitor of thrombin, FIIa.	Thank you. We have clarified these instances in the text to ensure we refer to inhibition of thrombin or inhibition of Factor IIa.
Peer Reviewer #2	General	Overall, this report strikes me as being somewhat biased in favor the new oral anticoagulants and it is otherwise of limited utility for the practicing clinician.	The protocol and processes used by the EPC were designed to reduce potential bias in our report. We feel that the review accurately conveys the state of the available evidence both in terms of its strengths and limitations.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #3	General	The most striking observation was the extreme level of duplication in material presented across the entire document. Most notable was the almost verbatim repetition of material from the executive summary in the body of the review. If possible, I would suggest trimming the executive summary down to a more concise and digestible review.	The duplication of the material between the main report and Executive Summary is intentional. The Executive Summary is published independently of the main report and is the only document read by many users; therefore, it is important for it to include all the main methods and findings of the systematic review. The full report provides additional detail for those users who need such information.
Peer Reviewer #4	General	The listing of quality ranking for individual studies is inconsistent. Sometimes a studies quality is listed, other times it is not.	We have included a quality rating breakdown of the relevant studies in the Description of Studies section of each KQ and provided additional descriptions of quality ratings in the key points sections. Quality (or Risk of Bias) ratings may also be found in several of the text tables, and are provided for all studies in the appendixes.
Peer Reviewer #5	General	This is a comprehensive review of recent literature on the predictors of stroke and bleeding in AF patients as well as an assessment of the newest agents to prevent stroke with AF. The KQs are explicitly stated and appropriate. although KQ 4, 5, and 6 had very little to no data to evaluate. The target population is clearly stated. The review of the data and results for KQs 1-3 are comprehensive and clear. The large number of comparisons is exhaustive and it was sometimes difficult to prioritize the main or primary outcome and then evaluate the secondary outcomes. In order to be more useful to the practicing MD it would help to have more focused summary statements or conclusions for each KQ.	We have revised the key points throughout the report and feel that the revised report is more focused and useful to the practicing clinician. Note also that as part of the EPC Program process, the Eisenberg Center may prepare dissemination documents specifically targeted at practicing clinicians and patients, which distill down the findings even further.
Peer Reviewer #6	General (KQ1)	1. Sole reliance on administrative data and ICD-9 codes for covariates is fraught with problems. Do any of the included studies include statistically robust samples to assess agreement between ICD-9 codes and medical record reviews?	Thank you for raising this important point. Many of the studies included sufficient samples to assess agreement with ICD-9 coding and medical record reviews. We have discussed this potential limitation in the Discussion section.
Peer Reviewer #6	General (KQ1)	2. Differences in definitions of key covariates, i.e., congestive heart failure, hypertension, is not mentioned. This statement applies to the risk score itself and the manner in which these key covariates were ascertained in the validation populations.	We agree and have included a statement regarding the lack of standard definitions in the Discussion.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #6	General (KQ1)	3. Framingham utilized blood pressure measurements and not a dichotomous yes/no variable. How was this operationalized in retrospective and prospective observational studies that would not routinely have this level of clinical detail?	All but one of the studies utilizing Framingham risk explicitly mentioned that they had actual blood pressure measurements that were used in their analyses.
Peer Reviewer #6	General (KQ1)	4. Similarly in studies that did not have access to patient medical charts or patient interviews, how was "prior TIA" or "prior stroke" determined? This important covariate is often missing from active problem lists. Patients often move, switch insurance, or switch health systems, so reliance on retrospective hospital billing data is problematic. This also applies to "prior bleed" for the bleeding scores.	We agree that common and standard definitions ideally should be used in evaluating scores and comparing analyses. However, given that many of the other studies used billing or ICD9/10 data to identify these events, we agree that they may have underestimated the number of events. We have acknowledged this limitation in the Discussion and highlighted those studies with clinical definitions of events or a formal adjudication process in the KQ 1 section.
Peer Reviewer #6	General (KQ1)	5. NONE of these scores includes chronic kidney disease which is an established risk factor for stroke in AF.	We agree and have expanded upon text in the Discussion to comment on the need to continue to refine risk prediction tools, including testing and validating the use of newer clinical markers (such as renal failure) either with or in addition to current risk tools.
Peer Reviewer #6	General (KQ1 – Stroke Outcomes)	1. How many of the included studies validated the strokes with review of primary data? Reliance solely on primary diagnosis from administrative data is fraught with error. It is unclear why any study would be included in this analysis that did not demand this level of rigor.	We thank the reviewer for the suggestion and now include in the text discussion of the reliance on administrative data. Unfortunately as suspected very few studies (5) described a clinical adjudication process to validate these events. The scores and outcomes reported in these studies did not allow us to perform a sensitivity analysis of this subset.
Peer Reviewer #6	General (KQ1 – Stroke Outcomes)	2. Hospitalized patients are different from outpatients in many respects: stability of covariates, concomitant medications, stroke prone procedures, etc. Is it valid to include studies of inpatients to assess the value of these stroke prediction tools?	We agree with the assessment of the heterogeneity of patients that are admitted versus those in an outpatient setting exclusively. We now discuss this issue in the Discussion section.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #6	General (KQ2)	1. It is unclear how one goes about comparing prediction tools across studies/populations that would have been missing many of the key covariates included in the score. For example, HEMORRHAGES includes ethanol abuse, genetic factors, and "excessive fall risk". Were any of these key risk factors uniformly available? Aspirin is a well-established risk factor for hemorrhage but it would be unavailable in population studies that rely on pharmacy databases as it is available without a prescription. Aspirin is a key exposure that was not included in the ATRIA bleeding risk score because it was not available in administrative data. Poorly controlled hypertension is included in HAS-BLED, but blood pressure measurements are not routinely available in population databases that include thousands of patients. Instead presence of "hypertension" in a problem list of ICD-9 codes is often substituted. Do large population based studies report on the availability of hemoglobin values, platelet counts, liver function tests, or creatinine values/GFRs? What is the temporal relationship to the bleeding event in regard to baseline covariate assignment in these studies? What weight was used to calculate GFR? What equation was used to calculate GFR? To assess the validity of a risk tool in an independent population, at a minimum, the variables used to derive the score should be available. The summary utility of the stroke and bleeding scores is being based on studies with serious methodological limitations.	The reviewer raises a key point—thank you for this feedback. An issue with the included studies is that different studies used different approaches to calculating the risk score, particularly for the HEMORR$_2$HAGES and HAS-BLED scores. For example, in HEMORR$_2$HAGES, due to unavailability of information on genetic factors, multiple database studies left out the "genetic factors" component of the score and so were, in effect, evaluating a modified HEMORR$_2$HAGES. To further complicate this issue, not all studies described in detail whether certain factors were omitted from their calculations of these scores. We decided that inter-study differences in approach to calculating some of the bleeding risk scores of interest limits comparison of bleeding risk scores across populations and precludes meta-analysis. We have added information regarding these issues to the "Detailed Synthesis" sections of KQ 2.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #6	General (KQ2)	2. Definitions of major hemorrhage vary across studies and this needs to be addressed. The ISTH definition for major hemorrhage is difficult to apply if investigators do not have access to hemoglobin or hematocrit values or blood banking data for number of units of packed red cells transfused. Investigators may substitute "bleeding requiring hospitalization" as the default definition for major hemorrhage which biases toward more bleeding events in the elderly population and those individuals who live alone. The decision to admit someone for "observation" is subjective.	Thank you for this comment. Our review is limited in that we are required to use the definition of bleeding provided by the studies' authors in the published literature. It is true that bleeding definitions differed somewhat from study to study. We now discuss this limitation in the Major Bleeding section overview.
Peer Reviewer #6	General (KQ2)	3. The point often made about CHADS scores tracking with risk of bleeding exemplifies the clinical dilemma regarding discrimination of stroke risk among medically complicated individuals in whom these same factors correlate with hemorrhage. CHADS and CHADSVASC schemes were derived and evaluated as stroke prediction tools. It is nonsensical that they would appear as bleeding prediction tools and should be removed from the KQ2 tables.	We thank the reviewer for their comment. We have removed discussion of stroke risk scores throughout the bleeding risk score section.
Peer Reviewer #6	General (KQ2)	4. The authors oversimplify decisions regarding antithrombotic therapy. Bleeding scores are aggregate scores that lump nosebleeds with 2 unit transfusions with intracranial hemorrhage. Not all bleeding outcomes are associated with the morbidity of an ischemic stroke. One cannot simply equate stroke and bleeding risk scores. This is one of the fundamental reasons that bleeding scores are not widely adopted in clinical practice. Bleeding scores should not be used to withhold anticoagulant therapy. They should be used to help minimize the risk associated with anticoagulant therapy.	We agree that how to balance the risks of stroke and bleeding against one another is important – and often unclear. We discuss the need for clinical decision tools that balance this risk although the development and use of such a decision analytic framework was outside of the scope of our project.
Peer Reviewer #6	General (KQ3)	1. ? INR range used in studies included from Asia.	Acknowledged. An INR range of 2.0 to 3.0 is clinically accepted regardless of patient age or race.
Peer Reviewer #6	General (KQ3)	2. 13559 patients in ATRIA were not all taking warfarin. Please check this statement.	The reviewer is correct. However, we were unable to locate where in the document we stated that all ATRIA patients were taking warfarin and so it has not been edited.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #6	General (KQ3)	3. Rivaroxaban was non-inferior to warfarin for stroke prevention in AF. Rivaroxaban reduced stroke compared to warfarin in the on-treatment group.	We agree. Both the intention-to-treat results as well as the per-protocol results, a prespecified secondary analysis, are presented within the Discussion section. In the results section for KQ 3, we have also presented ITT results first for this and other studies.
Peer Reviewer #6	General (KQ3)	4. It is a bit far-reaching to consider LAA occluding devices as a viable alternative to antithrombotic therapy in the vast majority of individuals with AF. The individuals with the highest bleeding risk on anticoagulant therapy (older frail medically complex) were not included in the trials of these devices and their safety in this population has not been determined.	The reviewer is correct that LAA occluding devices are still investigative devices pending FDA approval; we now make this clear in the text. However, because the technical expert panel, key informants, and investigative team felt that LAA may be a viable alternative to antithrombotic therapy it was included in our review. We also note in the Research Gaps section of the Discussion for KQ 3 that studies are needed to address the comparative safety and effectiveness of: (a) new oral anticoagulants and LAA closure for stroke prevention in nonvalvular AF patients, and (b) medication and procedural interventions in specific subpopulations (e.g., patients with advanced renal failure or on dialysis, and elderly patients)
Peer Reviewer #6	General (KQ3)	5. The BAFTA trial randomized individuals 75 years of age and older to ASA vs warfarin and contributed significantly to our understanding of the efficacy and safety of ASA in this age group.	We agree . This specific substudy is included in the KQ 3 discussion of the “Elderly patients with AF” subgroup section.
Peer Reviewer #7	General	The report is encyclopedic in length and detail, which is the usual style of an AHRQ report, and doubtlessly is style required. But it’s mind numbing to have all of that material recited and tabulated. The only part I read in any detail was the Executive summary, as the rest was just too much. AHRQ should reconsider the format for these reports, which guarantees that only a handful of people will ever read them. The summaries prepared for peer-review journals are much better. The key questions are good. The report is generally well done, and gathers material into a single place. That being said, it doesn’t break any new ground, or add too much to what has been written already on these topics. There were some missed opportunities to do more.	We thank the reviewer for their comments and have passed along suggestions/concerns about the EPC process to AHRQ for consideration.
TEP #1	General	Yes.	Thank you

Commentator & Affiliation	Section	Comment	Response
TEP #2	General	This was obviously a lot of work and the authors have spent a lot of time trying to put the evidence together. Some general suggestions on things that might be considered in revising the review:	Thank you
TEP #2	General	1. I am not sure that there is any strong justification to limiting the searches to studies from 2000 on. This resulted in all except two observational studies of warfarin vs. aspirin from being excluded, leaving a very incomplete view of the evidence on this important question. At the least the authors should have included systematic reviews of older trials of warfarin and aspirin foro prevention of stroke, to at least summarize the available evidence and so the reader isn't left with the mistaken impression that the only studies are those included in the report.	Thank you for this comment. We agree that information regarding the earlier evidence would be helpful to users of this report. To provide data and context of the results from earlier studies (pre-2000), we now include a summary in KQ 3 of the findings from two meta-analyses by Hart (published in 1999 and 2007) that evaluated the effect of warfarin, aspirin, and other antiplatelet therapies for stroke prevention in patients with non-valvular atrial fibrillation.
TEP #2	General	2. I have some questions/reservations about the methods used to analyze the studies of risk prediction (see methods below for more specifics).	Acknowledged
TEP #2	General	3. Some of the conclusions seem a little strong given the evidence provided, e.g. reporting of one risk prediction instrument for bleeding as superior over others even though the CI's overlapped in 2 of the 3 studies, and the estimates were very close in at least one study (see below for more detail).	We thank the reviewer for their comment. We have softened the conclusions accordingly.
TEP #2	General	4. It seems like in the conclusions harms of newer antithrombotics were not really discussed, in particular studies suggesting higher MI risk (or trend towards higher risk) with dabigatran.	Acknowledged. We have now included the higher risk of MI and GI side effects associated with use of dabigatran as compared to warfarin in the Discussion section.
TEP #3	General	Overall, a little mixed up in some places	We thank the reviewer for their comment and hope they feel the revised report addresses their concern.
TEP #4	General	Report represents a comprehensive review of the literature of population studies underlying the use of anticoagulation in patients with atrial fibrillation.	Thank you
TEP #4	General	What appears to be missing are the areas of concern to practitioners for which data from large population studies is lacking. These include: (1) Pros and cons of different methods of improving the time in therapeutic range (TTR) for patients on warfarin; ie., home INR testing, frequency of INR testing.	We acknowledge that safety concerns are very important to practitioners, particularly for the newer agents that are found to be as efficacious and possibly safer than warfarin. We have included safety concerns, including monitoring and reversal of anticoagulant effects of the newer agents as a research gap in the Discussion section.

Commentator & Affiliation	Section	Comment	Response
TEP #4	General	(2) Of concern I was unable to find any mention of the lack of an antidote to reverse anticoagulation with the newer oral thrombin inhibitors. This is a major concern of patients and physicians for in choosing warfarin vs. the direct thrombin inhibitors especially in the elderly at risk for falls. Head injury in the elderly due to falls is a growing public health problem. Similarly I did not find mention of current issues surrounding reversal of warfarin. On the other hand there is some attention to the question of whether to stop anticoagulation, or bridge from long acting agents to shorter acting agents for patients with AF undergoing common procedures.	Thank you for this comment. We have included the issue of reversal of these agents as a research gap in the Discussion section of this report.
TEP #4	General	(3) There was no mention of the quandry of how to manage patients with atrial fibrillation found to have microbleeds on MRI. These are now found commonly in persons with longstanding hypertension or amyloid angiopathy who undergo a brain MRI for another reason.	We thank the reviewer for this comment; however, we feel that the discussion of this patient subgroup is outside the scope of this review.
TEP #4	General	(4) Though there is mention of the fact that only a fraction of the patients with AF are currently treated with anticoagulation in the real world, there is little discussion of the risks of warfarin that underly conscious witholding of anticoagulation. What is the evidence based for the common contraindications for anticoagulation, ie. risk of falls, concomitant alcohol abuse, medical non compliance, dementia, etc.	Although a more detailed discussion of these risks is outside the scope of our review, we have introduced some additional details concerning these risks in the Introduction section.

Commentator & Affiliation	Section	Comment	Response
TEP #4	General	(5)There is no mention of the hypercoaguable state that infrequently occurs with institution of warfarin therapy thought due to time dependence of blockade of Protein C vs. other clotting factors. When it occurs it is mostly associated with skin necrosis. There is less written about reports of a transient hypercoaguable state upon acute warfarin discontinuation but a few reports of greater than expected incidence of stroke within first few days after warfarin discontinuation. Not clear if this is due to a hypercoaguable state or due to unmasking of a thromboembolic condition in exceptionally high risk vascular disease patients.	We thank the reviewer for this comment but feel that the discussion of this patient subgroup is outside the scope of this review.
TEP #5	General	This is a state of the art, exhaustive review of the available English language literature on the topic. The review fully explains the scope of the review in a logical and systematic fashion, details the current knowledge gap and the need for the systematic review of the literature. The review is strengthened by the accompanying appendices detailing the methodology of the literature search and the literature included for this review. Inclusion/exclusion criteria are well defined. The executive summary in 31 pages including 35 references is a superb summary of the entire report further enhanced by tables and figures. The Key Questions are fully explained and the target patients and subgroups of interest are well defined.	Thank you
TEP #5	General	Few typos were noted as follows: Executive Summary Page 8, line 15: Change the word ³and² to ³as² ES, Page 19, line 38 and line 41 delete the repeated word ³other showed² Page 4, line 18: add the word ³inhibitor² after Factor II Page 51, line 13: delete the word ³was² after therefore. Page 101, line 24: delete the repeated word ³agent²	Thank you for identifying these typos. We have corrected them.

Commentator & Affiliation	Section	Comment	Response
TEP #6	General	The authors have selected a very important cohort of patients population for prevention of stroke in non-valvular AF. They have reviewed close to 100 articles related to 74 unique studies. The report is nicely written and summarized alone with some shortcomings. The target population is well defined and the key questions are clearly stated and addressed in this report. The report is heavily reviewed 3 currently completed large clinical trials, mainly RE-LY, ROCKET-AF and Aristotle with over 50,000 patients in studied in these 3 trials. All were designed as a non-inferiority trial against warfarin, but the authors should have pointed out that these were a double-blind randomized trial, except for RE-LY which had an open-label dose-adjusted warfarin arm. In addition, the risk-profile of patient populations in these 3 trials somehow were different. Results from these studies were report as ITT population and some were reported as treated population, which is a point of controversy in non-inferiority trials.	Thank you. We now clarify when we are reporting ITT versus as-treated findings and which trials were designed as noninferiority trials.
TEP #7	General	In general, the report is a very comprehensive survey of the literature related to the topic. The key questions are clear, relevant, and appropriately stated.	Thank you

Commentator & Affiliation	Section	Comment	Response
Public Reviewer #1 – Sabine Luik, Boehringer Ingelheim Pharmaceuticals, Inc.	General	**Clearly delineate if a drug has or has not received FDA-approval for the indication under review** BIPI appreciates AHRQ's thorough review of currently available therapeutic options ("available strategies for anticoagulation" as indicated in the report) for stroke prevention in atrial fibrillation (SPAF). BIPI agrees that examining studies on a wide range of available products provides a comprehensive overview for providers, patients, and others who may use the findings to inform healthcare decisions. However, we are concerned that the draft report does not explicitly indicate whether the drugs being compared for the given indication are FDA-approved at the time of the report's publication. In order to provide the most accurate representation and up-todate information of reviewed drugs in this and future reports, BIPI strongly recommends that AHRQ clearly and explicitly indicate the FDA-approval status of all identified products for the particular indication under review in both the executive summary and throughout the report.	We have added explicit clarification regarding the FDA approval status of the new anticoagulant drugs in the Introduction, KQ 3 Results section, and Discussion. We also differentiate between the approval/marketing status of the 150 mg and 110 mg dosages of dabigatran in the KQ 3 Results and the Discussion.
Public Reviewer #1 – Sabine Luik, Boehringer Ingelheim Pharmaceuticals, Inc.	General	Of note, apixaban, edoxaban, and dabigatran 110mg are not currently FDA-approved as options for treating SPAF or preventing thromboembolic events, however, these statuses are not currently cited in the report. Already delayed on two separate occasions, apixaban is still pending FDA-approval while edoxaban is currently considered experimental in the U.S. While the 150mg dose of dabigatran is FDA-approved and currently on the market, the 110mg dose is not approved or marketed in the US. The report attempts to briefly address these points by noting on page 4 that "New devices and systemic therapies have been developed for stroke prophylaxis and are in testing or have been approved for use," however, it does not explicitly delineate which drugs are or are not currently FDA-approved.	As described above, we now clarify in the Introduction the FDA status of the different novel anticoagulants.

Commentator & Affiliation	Section	Comment	Response
Public Reviewer #1 – Sabine Luik, Boehringer Ingelheim Pharmaceuticals, Inc.	General	Similarly, BIPI encourages AHRQ to closely review the report and revise language as necessary to ensure the accuracy of the general context used to describe these non-approved drugs. In particular, on page ES-24 in the Key Findings section, and on page 100 in the Findings in Relationship to What is Already Known section, the language references the 2006 U.S. guidelines for the management of AF and then states, "Since that time, newer anticoagulants have entered the marketplace (namely dabigatran, rivaroxaban, and apixaban)." However, apixaban is not currently marketed in the U.S. despite what the statement suggests. As such, BIPI requests AHRQ to remove apixaban from this statement and provide clarification throughout the document around the approval status of any drug that is not FDA-approved.	Thank you; we agree it is important to explicitly note the status of drugs that have not received FDA approval. Since apixaban has now received FDA approval, we have not altered the sentence specifically mentioned in this comment; however, we have added additional clarification throughout the report to clearly note drugs that are not FDA approved.
Public Reviewer #1 – Sabine Luik, Boehringer Ingelheim Pharmaceuticals, Inc.	General	While these findings may offer some insight to physicians and patients on the relative effectiveness and safety of these products, BIPI believes that it is especially important for AHRQ to clearly identify each product's FDA-approval status to ensure that stakeholders are being provided with the most accurate and up-to-date information when making choices about their care. Products lacking FDA-approval have not undergone the rigorous regulatory process by which FDA evaluates all safety issues associated with a drug and thus, cannot be appropriately compared to FDA-approved drugs. By clearly identifying each comparator's FDA-approval status in this and future reports, AHRQ will make stakeholders aware of current safety and efficacy information, ultimately aiding them in healthcare decision-making, as well as preventing any misinterpretation of the studies' conclusions.	We agree and their status has been clarified.

Commentator & Affiliation	Section	Comment	Response
Public Reviewer #1 – Sabine Luik, Boehringer Ingelheim Pharmaceuticals, Inc.	General	**Address conclusions drawn from studies with "as treated" models vs. "intention to treat" models** BIPI commends AHRQ for its thorough and detailed assessment of the quality of individual studies included in the review. A core criterion for all included studies, among others, was placing a primary focus on intention-to-treat (ITT) clinical trial populations. We agree with AHRQ's emphasis on ITT populations, as analyzing ITT data helps to avoid the potential bias associated with systematic patient dropout. Therefore, we do have concerns with reporting related to the "all stroke or systemic embolism" outcome in the ROCKET AF trial (rivaroxaban vs. warfarin). The trial results for this outcome were based on the "as treated" population, which demonstrated a statistical superiority for rivaroxaban vs. warfarin (HR = 0.79 [0.65 – 0.95]; page ES-22). However, when the ITT population is examined, the statistical significance of rivaroxaban's superiority is eliminated (HR = 0.88 [0.74 – 1.03]).1 Therefore, the choice of which clinical trial population to analyze can directly impact the statistical model results and conclusions. ITT analyses (1) allow analysts to retain balance in patient characteristics originating from the original random treatment allocation, (2) are more likely to yield an unbiased treatment effect estimate when compared to "as treated" analyses, and (3) document non-compliance and departures from trial protocol (e.g., non-adherence), which better reflects actual clinical practice.2 As such, BIPI strongly requests that AHRQ continue its commitment to placing primary focus on ITT populations, thereby evading potential bias associated with analyzing "as treated" populations in clinical trial data.	Thank you for this comment. We have gone through the trials included for these analyses and have consistently presented the ITT analyses first when available. Most trials reported primary efficacy endpoints in an ITT population. In trials reporting safety and bleeding data, often the population was either the "as-treated" population or population that had received at least one dose of the study drug.

Commentator & Affiliation	Section	Comment	Response
Public Reviewer #1 – Sabine Luik, Boehringer Ingelheim Pharmaceuticals, Inc.	General	**Address conclusions drawn from studies with poorly-controlled warfarin comparison arms** Additionally, we interpret results from the ROCKET AF study cautiously. The rates of INR control in patients taking warfarin in ROCKET AF were lower than in other OAC clinical trials (e.g., RE-LY, ARISTOTLE), which may have impacted the study results. An INR or ≥2.0 has been shown to reduce the risk of ischemic stroke, as well as severity and stroke-related mortality.3 In ROCKET AF, patients on warfarin had therapeutic INR levels (2.0-3.0) of 55% of the time, while patients on warfarin in the RE-LY and in ARISTOTLE trials had therapeutic INR levels of 64% and (only according to the publication) 62.2% of the time, respectively. This disparity in INR control likely affects the overall relative efficacy of an anticoagulant when compared to warfarin. Thus, BIPI recommends AHRQ to note and account for this difference in INR control when evaluating research.	Thank you for this comment. We have added a description of the different TTR for those on warfarin in the ARISTOTLE vs ROCKET trial in the section of KQ 3 headed "Xa Inhibitors Versus Warfarin." While those on warfarin in the ROCKET trial did have a lower average TTR compared to the other trials, their TTRs were closer to real-world settings. Therefore, results of ROCKET are applicable to clinical practice. But, as pointed out, the different TTRs of the participants on warfarin in the different trials should prevent any direct or indirect comparisons of these results.

Commentator & Affiliation	Section	Comment	Response
Public Reviewer #1 – Sabine Luik, Boehringer Ingelheim Pharmaceuticals, Inc.	General	**Revise language to avoid making inadvertent indirect comparisons** BIPI agrees with and appreciates AHRQ's recognition on page ES-24 that direct comparisons of oral anticoagulant (OAC) medications, at this time, are not possible given the lack of comparative trials. As AHRQ notes in the report, the pre-market trials of OACs analyzed AF patients at differing risks of stroke and used different dosing strategies and event definitions, making any cross-trial conclusions unreliable and potentially misleading. In the same regard, AHRQ should also note that, for medications not approved by the FDA, the authors' report relies solely on published data generated by the manufacturer whereas all data on approved drugs have been vetted by the FDA and undergone scrutiny to ensure their claims on safety and efficacy are accurate. Despite acknowledgments, BIPI is concerned by language in the report used to describe apixaban's safety and efficacy in patients not suitable for warfarin. BIPI recommends AHRQ review and revise language throughout the report that may inadvertently highlight apixaban's safety and efficacy as compared to other OACs, thereby avoiding inappropriate indirect comparisons with other OACs.	We agree that due to different methodologies of the trials, direct or indirect comparisons of these novel antithrombotic agents based on data from these trials should not be done. Head-to-head comparisons are needed to compare therapeutic effects and safety of these agents. This point has been clarified in the Discussion section.

Commentator & Affiliation	Section	Comment	Response
Public Reviewer #1 – Sabine Luik, Boehringer Ingelheim Pharmaceuticals, Inc.	General	One example of such language includes a statement in the Conclusion section of page V as well as page ES-28, “Apixaban in particular shows safety and efficacy in patients who are not candidates for warfarin.” By referencing apixaban, specifically, and excluding mention of other OACs, the report seems to implicitly suggest apixaban’s superiority over the other OACs in patients contraindicated to warfarin. While the recently published AVERROES clinical trial (Connolly et al, 2011) did find apixaban to be superior compared to use of aspirin in patients who are eligible for anticoagulation but “unsuitable” for warfarin, this does not support the implied claim that apixaban is the universally superior therapeutic option for this population.4 Additionally, the AVERROES trial presents several separate reasons for unsuitability of warfarin that do not preclude treatment with new FDA-approved OACs.5 It is important to be clear that few absolute contraindications to any anticoagulation therapy exist (usually related to elevated bleeding risk). Ultimately, this implied superiority of apixaban over other OACs is not supported in the existing evidence-base. Therefore, BIPI recommends AHRQ to remove the phrase, “in particular” or further clarify this statement to avoid potential misinterpretations about the drug’s effectiveness and safety compared to other OACs for patients who are not candidates for warfarin.	We have removed the phrase “in particular” from this statement as suggested.

Commentator & Affiliation	Section	Comment	Response
Public Reviewer #1 – Sabine Luik, Boehringer Ingelheim Pharmaceuticals, Inc.	General	Separate from the above concerns regarding inadvertent indirect comparisons, another section of the report on page ES-25 states, “recent evidence showed that for the first time a new oral anticoagulant agent (apixaban) reduced all-cause mortality in patients with AF.” Again, BIPI is concerned that stakeholders would interpret this statement as apixaban outperforming all other treatment modalities with respect to all-cause mortality when in fact, data from ARISTOTLE only indicates that apixaban is superior to warfarin regarding all-cause mortality (HR=0.89 [CI: 0.80 – 0.998]; page ES-21). Dabigatran also showed a reduction in all-cause mortality compared to warfarin (HR=0.88 [CI: 0.77 – 1.00]; page ES-20) and these data have been subject to FDA scrutiny unlike the proposed apixaban data. As such, BIPI requests further accuracy surrounding the use of wording when interpreting study results in order to avert the aforementioned unintended consequence. Revising or removing such language is necessary to provide stakeholders with an unbiased and accurate frame of reference for interpreting the findings.	We have revised this statement to clarify that the mortality effects observed for the 3 new anticoagulant agents (dabigatran, apixaban, rivaroxaban) are not specific to apixaban.
Public Reviewer #1 – Sabine Luik, Boehringer Ingelheim Pharmaceuticals, Inc.	General	**Provide clarity surrounding references to observational data on bleeding rates** On page 103 of the draft report, a statement in the Implications for Clinical and Policy Decisionmaking section claims “early observational research suggest that the bleeding risks of OACs may have been underestimated in clinical trials.” From the report, the specific citations used to support this potentially misleading claim are lacking. Further, the term “early observational data” suggests there is very limited observational data to support this claim. BIPI recommends that AHRQ remove this statement from the report or, at minimum, revise it to cite the specific data to which the statement is referencing.	We have removed this statement as suggested.

Commentator & Affiliation	Section	Comment	Response
Public Reviewer #1 – Sabine Luik, Boehringer Ingelheim Pharmaceuticals, Inc.	General	In general, causal links between medication administration and safety outcomes via observational data are difficult to establish due to a number of potential confounding factors. Properly-powered randomized, controlled clinical trials (RCT) far exceed observational study designs and alternative data collection methods, such as spontaneously-reported adverse events, in their ability to establish safety and efficacy. Spontaneous adverse event reporting cannot establish causality and does not provide rates of events to compare to determine if the event rates are higher or lower than expected. Relying on spontaneous adverse event reporting for recently-approved drugs is inherently biased because numerous factors can increase spontaneous adverse event reporting, including level of utilization, degree of innovation, litigation and advertising.	Acknowledged
Public Reviewer #1 – Sabine Luik, Boehringer Ingelheim Pharmaceuticals, Inc.	General	Without directing readers to the data source to support or clarify this assertion, the statement could impact patient, provider, and other stakeholder perceptions that real-world observational research has found higher bleeding rates in patients using OACs than those found in the seminal RCTs. Current evidence of fatal bleeding rates of OACs is consistent with our expectations, based on findings from the RE-LY trial, and are aligned with the U.S. Prescribing Information. Similarly, a recent FDA Drug Safety Communication that addressed bleeding events in patients taking dabigatran supported the continued use of dabigatran for SPAF, highlighting that it provided an "important health benefit when used as directed," but recommended adherence to the approved drug label.6	Acknowledged
Public Reviewer #1 – Sabine Luik, Boehringer Ingelheim Pharmaceuticals, Inc.	General	Providing publicly available and peer-reviewed information would clarify AHRQ's use and reliance on observational research for drawing conclusions about the best treatment options for specific therapeutic areas, in addition to ensuring providers and AF patients have accurate information about the risks of various AF treatment options.	Acknowledged

Commentator & Affiliation	Section	Comment	Response
Public Reviewer #2 – Freda Lewis-Hall, Pfizer	General	**There is an opportunity to apply a more consistent approach in the evaluation and discussion of the technical elements of the studies included in the report.** **Please provide further explanation/context for the use of specific study concepts when comparing interventions, particularly in summary sections.** In many of the report's summaries, the review refers to technical elements of included studies or discusses the effectiveness of evaluated therapies, but does not always provide full context or definition. For example, in the discussion of rivaroxaban, superiority versus non-inferiority determinations are presented, but additional caveats or explanations around the comparisons could help the reader better contextualize the results. The potential value of additional context can also be noted in the discussion of intent-to-treat and on-treatment results for the analyses. For example, in evaluating Factor Xa inhibitors versus warfarin, the report states that: "In a second study, in the per-protocol population, rivaroxaban was shown to be noninferior to warfarin in preventing stroke and systemic embolism (1.7% per year vs. 2.2% per year for rivaroxaban and warfarin, respectively; HR 0.79; 95% CI, 0.66 to 0.96; p<0.001 for non-inferiority; 1.7% per year vs. 2.2% per year for rivaroxaban and warfarin, respectively; HR 0.79; 95% CI, 0.65 to 0.95; p=0.01 for superiority)." As safety analyses are not designed to evaluate efficacy outcomes, their use in these types of comparisons may lead the reader to potentially incorrect conclusions. We recommend that evidence from intent-to-treat populations should be evaluated in order to correctly draw conclusions about product efficacy. We also suggest that the p-value used to identify superiority should be 0.02, rather than the 0.01 per publication value used in the draft report.	Thank you for these comments. We have gone through all of the trials included in KQ 3 and have presented the ITT analyses first and as the primary results for the efficacy outcomes. For trials reporting on safety endpoints, most often the study populations described were not the ITT populations. We cannot discard these data, but we have tried to define the population better for the descriptions of safety outcomes.

Commentator & Affiliation	Section	Comment	Response
Public Reviewer #2 – Freda Lewis-Hall, Pfizer	General	**Please clarify summary language on the comparative effectiveness of different therapies for atrial fibrillation stroke prevention.** In the structured abstract conclusions, the authors note that: "[n]ewer anticoagulants show initial early promise of reducing stroke and bleeding events when compared to warfarin, and apixaban in particular shows safety and efficacy in patients who are not candidates for warfarin." It may be the case that this could be misinterpreted by readers to imply that apixaban has only been evaluated in patients who are not candidates for warfarin therapy. To clarify, we recommend that this be edited as: "Newer anticoagulants show initial early promise of reducing stroke and bleeding events when compared to warfarin. In addition, apixaban in particular showed safety and efficacy when compared to aspirin in patients who are not candidates for warfarin."	We have made the suggested edit.

Commentator & Affiliation	Section	Comment	Response
Public Reviewer #2 – Freda Lewis-Hall, Pfizer	General	**Please consider a more uniform application of strength of evidence assemssnents to all studies in the report.** At points, the report appears to employ an variable approach to evaluating the strength of evidence presented in captured trials, particularly as it relates to the precision metric. For example, the Apixaban for the Prevention of Stroke in Subjects with Atrial Fibrillation (ARISTOTLE) trial presents nearly identical hazard ratios and confidence intervals for the endpoints on ischemic stroke and myocardial infarction. However, the strength of evidence from ARISTOTLE for ischemic stroke is deemed to be 'high', whereas the strength of evidence for myocardial infarction from this same trial is deemed to be 'moderate'. Given that both endpoints are derived from the same trial, it would seem to follow that the evaluation should rate the strength of the evidence for both endpoints similarly. We recommend that a more uniform application of strength of evidence assessment standards across trial results would produce more consistent findings for the reader.	We have globally reviewed the SOE ratings within the report and modified where needed for consistency.
Public Reviewer #2 – Freda Lewis-Hall, Pfizer	General	**Please consider incorporating major gastrointestinal bleeding as an endpoint in Key Question #3 on "Interventions for Preventing Thromboembolic Events."** Gastrointestinal bleeding is a central issue for atrial fibrillation therapies. Although gastrointestinal bleeding is a prevalent and important endpoint captured in studies of therapies for stroke prevention, Key Question #3 of the report currently does not include this important and clinically meaningful endpoint. We recommend that this endpoint be included as part of thus question.	As described in the Methods section, we discuss in the systematic review bleeding categorized into major and minor bleeding. Within the results we highlight when these bleeds were gastrointesinal (or otherwise) when reported by the authors although categorization and reportng of bleed sites varied.

Commentator & Affiliation	Section	Comment	Response
Public Reviewer #2 – Freda Lewis-Hall, Pfizer	General	**Please consider revising the conclusions on the usefulness of CHADS2 versus CHA2DS2-VASc scores, and please consider the increased benefit of using CHA2DS2-VASc to identify stroke risk when compared to the CHADS2 tool.** In the summary section for Key Question #1, "Predicting Thromboembolic Risk," the report states that: "CHADS2 and CHA2DS2-VASc continuous risk scores appear to be similar and the most predictive of stroke events when compared with the CHADS2 categorical score, the CHA2DS2-VASc categorical score, and the Framingham categorical score. This finding was, however, statistically significant only when compared with the Framingham categorical score." The CHA2DS2-VASc (Congestive heart failure, Hypertension, A2 for age greater than or equal to 75, Diabetes mellitus, S2 to recognize prior stroke or transient ischemic attack or thromboembolism, Vascular disease, Age 65-74 years old, Sex category) was developed after the CHADS2 (Congestive heart failure, Hypertension, Age greater than or equal to 75, Diabetes mellitus) clinical prediction rule. As a results, the CHA2DS2-VASc measure and has been evaluated in fewer studies than the CHADS2 measure, thereby limiting the ability to fully draw conclusions about the usefulness of CHA2DS2-VASc when compared to other risk evaluation tools.	We thank the reviewer for their perspective. Even though the CHA_2DS_2-VASc has more elements for stroke prediction but fewer studies reviewed for our analysis, our review did not show that its performance was superior to the $CHADS_2$ score to warrant the statement and therefore we have not made the modifications as suggested.

Commentator & Affiliation	Section	Comment	Response
Public Reviewer #2 – Freda Lewis-Hall, Pfizer	General	**Please consider capturing additional evidence on therapies used to prevent atrial fibrillation-related stroke.** **Please consider including data from the supplementary appendix to the rivaroxaban versus warfarin in nonvalvular atrial fibrillation (ROCKET-AF) trial primary publication.** The section of the report evaluating outcomes associated with Factor Xa Inhibitors and warfarin does not currently reference a supplement to the 2011 New England Journal of Medicine article titled "Rivaroxaban versus warfarin in nonvalvular atrial fibrillation." 1 This supplement provides valuable additional data on various endpoints used to evaluate warfarin and rivaroxaban, including rates of ischemic or uncertain stroke, hemorrhagic stroke, systemic embolism, major gastrointestinal bleeding, vascular death, and adverse events. We recommend that this supplement and its associated evidence tables be incorporated into the draft report.	Thank you for these suggestions. We have now added these secondary efficacy outcomes from ROCKET and have clarified that these are from the on-treatment population
Public Reviewer #2 – Freda Lewis-Hall, Pfizer	General	**Please consider incorporating recently published indirect comparison analyses of novel oral anticoagulants (NOACs).** The report notes that heterogeneity of NOAC trials prevents indirect comparisons across studies. However, we note that numerous indirect comparison analyses of NOACs have been published in 2012. This suggests that the scientific community does view indirect comparison analysis within this class of products appropriate, given the limitations. As such, we recommend that the report capture available meta-analyses using indirect comparisons of NOACs, and that the authors seek to perform similar assessments of the quality and level of evidence so that stakeholders can be better informed as to their relative merits.	Although we stand by our decision to not combine the novel oral anticoagulant data with indirect comparison meta-analyses given the heterogeneity of study designs, therapies, populations, and concomitant therapies in the included studies, we now include within the discussion a brief summary of the main analyses published in the existing literature.

Commentator & Affiliation	Section	Comment	Response
Public Reviewer #2 – Freda Lewis-Hall, Pfizer	General	**Please consider capturing data on subanalyses from high profile trials to support better decisionmaking on interventions for preventing thromboembolic events.** Key Question #3 of the report focuses on "interventions for preventing thromboembolic events," but does not currently include important data relevant to clinical decision making. Specifically, notable trials such as Randomized Evaluation of Long-term anticoagulant therapy (RE-LY) and ARISTOTLE offer subanalyses that seem to be relevant to include in the report. We recommend including subanalysis data on renal impairment from ARISTOTLE and subanalyses on international normalized ratios from the RE-LY study.	We agree with the reviewer; recently published data from these subanalyses were identified during the search update and pertinent information been added to the report.

Commentator & Affiliation	Section	Comment	Response
Public Reviewer #3 – Cynthia Bens, Afib Optimal Treatment Task Force	General	We agree with AHRQ's finding that stroke prevention in patients with AFib in clinical practice is complex and challenging but critically important given the morbidity and mortality associated with stroke events. AFib is associated with an approximate doubling of mortality risk, in large part due to the increased risk of stroke1. There are currently conflicting clinical guidelines and educational efforts, which leads to confusion about how healthcare providers should determine stroke risk and bleeding risk in patients with AFib, what tools should be used, and how best to incorporate scores from these tools into treatment decision making with their patients. The result is commonly the underutilization of anticoagulants, particularly in older patients who are often at a perceived higher risk of bleeding. This is a major obstacle to effective care that reduces morbidity and mortality for this condition. We were pleased that AHRQ included a key question in this review on the strength of evidence to support the validity of current bleeding risk assessment tools in addition to tools available to determine a patient's risk of stroke. Bleeding risk assessment is not only important because it can uncover risk factors for serious bleeding such as intracranial hemorrhage, but also because it can identify modifiable risk factors that can be addressed before a patient receives anticoagulation therapy.	We thank the reviewer for these comments.

Commentator & Affiliation	Section	Comment	Response
Public Reviewer #3 – Cynthia Bens, Afib Optimal Treatment Task Force	General	The AFib Optimal Treatment Task Force convened a roundtable of experts in cardiology, neurology and other related fields on January 18, 2012 to forge consensus on the best practices for assessing stroke and bleeding risk in anticoagulation decision-making using available risk assessment tools. These experts subsequently developed a consensus statement (see Appendix A) to provide guidance to health care providers on evaluating patients with AFib. This consensus statement aligns with and supports the AHRQ recommendations. In the consensus statement, the experts recommend a three-step approach: •First, a patient's stroke risk should be assessed and recorded no less than annually using an established scoring tool. Those identified as intermediate or high risk should be put on an anticoagulant—warfarin or a direct thrombin inhibitor or a factor Xa inhibitor. Aspirin is not recommended for stroke prophylaxis in AFib. •Second, if the patient is at high enough risk to require anticoagulation therapy, the patient's bleeding risk should then be evaluated to estimate the net clinical benefit of an anticoagulant Risk factors for intracranial hemorrhage should be considered although routine screening for these risk factors is not currently indicated. For the majority of patients, the net benefit of stroke prophylaxis supersedes the "net harm" of serious bleeding events—even in older patients. •Third, the decision to undergo anticoagulation therapy must reflect patient preferences and values. The patient must also understand the relative benefits and risks involved in the discussion and decision surrounding the clinical net benefit of anticoagulation therapy.	Thank you for pointing out this relevant reference, which provides important context to KQ 1 and KQ 2. We agree with the reviewer's comments and have added this reference to the "Description of Included Studies" section of KQ 2 and KQ 1 in order to contextualize our approach to these sections. We have also added discussion of the recommended three-step approach in the overall report Discussion.

Commentator & Affiliation	Section	Comment	Response
Public Reviewer #3 – Cynthia Bens, Afib Optimal Treatment Task Force	General	In addition to this recommended approach, our roundtable experts agreed that priority should be given to collecting and analyzing real-world data on new anticoagulants to identify which patients are best suited for specific agents. The experts identified needed health care professional and patient education materials and tools to support both risk assessment and implementation of new anticoagulation therapies. They also highlighted areas requiring additional research. Supporting AHRQ's identified research gaps in the areas of risk stratification for thromboembolic and bleeding risk, our roundtable experts concluded that more research is needed into specific risks associated with intracranial hemorrhage, including the biological (versus chronological) age, frailty and specific findings on brain imaging. The roundtable experts also suggest that further refinement of stroke risk stratification could result in identifying patients who are truly at low risk for a stroke and who presumably have a low net clinical benefit from anticoagulation. Such refinement could improve the predictive value of risk stratification tools, however there will still be the need for additional healthcare provider education on the risks and benefits of new and existing treatments for stroke prevention in AFib and facilitating a dialogue with patients about their individual risk of stroke and the benefit of treatment.	We agree with the reviewer's perspective and have added to our Discussion that real-world data, particular in examining VKA and the use and uptake of new anticoagulants, must be examined to understand their true effectiveness and risks. We also agree that with the advent of more digitized clinical data, continuous refinement of risk tools needs to occur for better discrimination of risk and benefit in atrial fibrillation patients. Thank you for raising these issues. Our comprehensive literature review is consistent with the findings from the expert roundtable described by the reviewer. We agree that: (1) real-world data on new anticoagulants; (2) biological age, frailty, and specific brain imaging findings as RFs for ICH; and (3) refinement of risk scores to better identify a truly "low risk" population are all important issues and represent gaps in the current evidence base. We have added detail to our Discussion's "Research Gaps" section.
TEP #2	Abstract	Specific comments regarding abstract: a. There is little indication of the number of studies or quality in the various results; trying to work some of that in would be helpful for readers. b. The c-statistic is a measure of "discrimination" not really "prediction" so I don't think "predictive power" is the correct term (abstract and elsewhere).	We appreciate the reviewer's comments. We include detail in the abstract regarding the overall number of included studies, the number of studies reporting data relevant to each KQ, and the strength of evidence for the conclusions. We appreciate the reviewer's interest in the additional detail regarding study quality; however, we feel that the strength of evidence (of which study quality is one element) provides the most informative, yet concise, format for a reader in the context of the brief abstract. Regarding c-statistics, the reviewer's point is correct. We have replaced "predictive power" terminology with "risk discrimination ability" terminology in the abstract and elsewhere in the report.

Commentator & Affiliation	Section	Comment	Response
TEP #3	Abstract	Abstract - seems to be very focussed on the NOACs - each only has one Phase 3 trial vs wafarin, apart from apixaban vs aspirin. In contrast there are numerous trials for warfarin. Overall, the abstract needs rewritten to ensure the main message is not over-whelmed by the NOAC sentences.	Thank you for this comment. Prior meta-analyses done through 2007 do focus on warfarin and aspirin. Our review updates these results and does tend to focus on the newer treatment options and risk stratification tools, which are the focus of most of the recently published literature. We have tried to make this clearer in the abstract. We have also emphasized that warfarin continues to be a viable treatment option.
Peer Reviewer #2	ES (Results)	The final bullet on page ES-17 states: "Patients with renal impairment, with different INR control, and with prior stroke seem to benefit equally from the new anticoagulant agents when compared with warfarin (low strength of evidence). This finding is based on one study of patients with renal impairment, two studies of patients with different INR control, and seven studies of patients with prior stroke." I do not agree that there is low-quality evidence to support this set of conclusions. The conclusion is being drawn on sub-group analyses (highly subject to bias) and, in the case of the statement about "different INR" control, the conclusion is based on analyses of CENTER-based (not patient-based) INR control...I do not think we know, based on the available evidence, whether the point estimates for the relative effect of the NOACs (vs. warfarin) would apply similarly to patients with different INR control.	Thank you for this comment; we agree that the INR control discussed in this statement refers to center-based INR control rather than patient-based INR control and we have indicated that difference in the key points and main results.. Although based on subgroup analyses of the RCTs, thes findings for the different subgroups were consistent and the analysis prespecified. We feel that the low-quality evidence rating is appropriate
Peer Reviewer #2	ES (Results)	In line 19 or page ES-21, the authors state in the table (and previously in the text) that dabigatran increases the risk of MI. However, the lower bound of the 95% CI for the HR is 0.98. How can this be moderate quality evidence of an increase in risk? Are the authors also considering the meta-analysis by Uchino et al when reaching this conclusion? They should state explicitly...	We have modified the SOE table, key points, and text to clarify that the strength of evidence related to MI risk differs between the two different dabigatran doses with moderate strenght of evidence supporting an increase in MI risk with 150mg dabigratran but only low strength of evidence for the 110mg dose of such an increased risk. We reviewed the references of the Uchino meta analysis during the study selection phase of our review and included individual studies from that analysis which met our inclusion criteria.
Peer Reviewer #2	ES (Results)	In lines 15-17 of ES-22, the authors show data indicating rivaroxaban is superior to warfarin for the prevention of stroke + systemic embolism. While this was true in the 'on treatment' analysis of the ROCKET-AF study, it was not true in the intention-to-treat analysis. Thus, I do not believe a superiority claim can be stated.	Thank you for this comment. We have now reported the ITT results for the primary efficacy outcome of the ROCKET-AF trial before the on-treatment results.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #2	ES (Discussion/ Conclusion)	In lines 13-14 of ES-25, the authors state that "All the new oral anticoagulants were better tolerated than warfarin, and rates of study drug discontinuation were lower with the new agents when compared with warfarin." This is not true - patients in the RE-LY trial (open-label dabigatran vs. warfarin) discontinued dabigatran more frequently than warfarin. The other statements about the NOACs in this section are accurate but they leave out certain negative findings: more major GI bleeding with dabi 150 and riva, trend toward more MI with dabi, etc.	We have added in clarification that this statement refers to "all the new oral anticoagulants *tested in a blinded fashion.*" As noted, myocardial infarction and gastrointestinal bleeds were outcomes of interest and are reported within the main report and also highlighted in the key points and discussion section as appropriate.
Peer Reviewer #4	ES (Methods)	ES-10 line 35: first I have heard of starting at moderate for observational studies. Usually you start at low and upgrade as merited. And, how can the results of one study be listed as ³consistent²	We now clarify that studies of risk prediction started with moderate strength of evidence and cite the appropriate reference. We have clarified that studies needed to include more than 1,000 patients to be considered consistent when there was only one study.
TEP #1	ES (Results)	In the Executive Summary, for KQ1 and KQ2 please indicate how many studies directly compared the risk scores in a head to head comparison, and what the results of these studies were.	We now clarify in the tables within the main report which studies included head-to-head comparisons of the different scores. This information has also been highlighted in the Executive Summary.
TEP #1	ES (Introduction)	ES-5: bridging is associated with increased risk of bleeding—please provide references for this statement.	We have provided references for this statement in the Executive Summary and main report as suggested.
TEP #2	ES (Methods)	ES-10: The sentence that "For outcomes were confounding was not believed to be an issue..." is talking about risk prediction studies and doesn't seem quite on target. The point isn't that confounding isn't an issue (there are plenty of other sources of bias in such studies, and confounding can certainly play an important role, e.g. if patients with high risk assessment scores are treated more aggressively); it's that by their nature studies of risk prediction and diagnostic accuracy are observational studies. I don't think you need a justification, just say that studies of risk prediction started with moderate strength of evidence (can refer to the GRADE method for diagnostic reviews here, or AHRQ methods if they are out).	We have modified this text as suggested.

Commentator & Affiliation	Section	Comment	Response
Public Reviewer #4 – Judy Wagner	ES (Introduction)	Page:ES-4, Para: 3, Line(s): 1 Existing Text: New anticoagulants are challenging the predominance of VKAs for stroke prophylaxis in AF. Since 2007, three large trials comparing novel anticoagulants with VKAs have been completed, with a combined sample size of ~50,000 subjects. Recommendation: Include a statement that rivaroxaban and dabigatran have been approved in the US for stroke prevention in nonvalvular atrial fibrillation. Apixaban has not been approved.	Since apixaban has now received FDA approval, we have not added the specific statement recommended by the reviewer; however, we have added statements throughout the report to note the status of drugs that are not yet FDA approved.
Peer Reviewer #1	Introduction	The Introduction is appropriate and excellent.	Thank you
Peer Reviewer #2	Introduction	OK	Thank you
Peer Reviewer #3	Introduction	Well structured with a good summary of background and justification for key questions.	Thank you
Peer Reviewer #4	Introduction	good. no comments	Thank you
Peer Reviewer #5	Introduction	The background info on the importance of AF is well stated and clear. On page 5, the authors mention addressing health care costs, but the latter may be beyond the scope of this report since no cost-utility analyses were addressed. No discussion fo the cost differentials for the community use of the newer agenst compared to the older agenst for AF was provided. The model in the analytical framework is helpful but the outcomes listed there are not all equally importnat. Some ordered prioritization as to the most clinically meaningful or primary vs secondary could be helpful.	The reviewer is correct that health care costs were considered outside of the scope of this review. We have reorganized the outcomes listed in the analytic framework to highlight the importance of the outcomes of interest.
Peer Reviewer #6	Introduction	Intro is fine.	Thank you
Peer Reviewer #7	Introduction	Generally reasonable, but the motivation for the risk scores was not as good as it could be.	We have added in additional text to provide the motivations for KQ 1 and KQ 2.

Commentator & Affiliation	Section	Comment	Response
TEP #1	Introduction	Yes. Minor point--the definition of atrial fibrillation in the first sentence (Background, ES-1 and in the body of the report) is unclear. It is easy to mix up the definition of supraventricular tachycardia with the definition of afib. If you feel you need to define "supraventricular tachycardia" please do so in a separate sentence.	Thank you for this comment. We have made the suggested edits.
TEP #2	Introduction	The intro seems to presuppose to some degree that CHADS2 and HAS-BLED are the best tools, which they should be careful about not doing.	The $CHADS_2$ and HAS-BLED tools are certainly the most studied tools and this is indicated in the Introduction. We do not, however, feel that the Introduction presupposes that these are the best tools but instead demonstrates the need to determine the comparative effectiveness of all available tools.
TEP #2	Introduction	The intro doesn't really mention other risk assessment tools. Even though many have fallen out of use since the widespread endorsement of CHADS2 etc they should at least be mentioned. Perhaps evidence was excluded b/c many of these were developed/studies before 2000.	We have inserted an additional sentence as suggested.
TEP #2	Introduction	A number of places in the intro talk about how a comparative review is needed to better understand the clinical value/efficacy etc and will help clinical practice. This is repetitive and really not needed in the first place (p ES-3, line 28-37; ES-4, lines 47-49; ES-5, lines 23-26).	Thank you for this comment. We have addressed this issue by reducing the repetition.
TEP #2	Introduction	There is also repetition with regard to underuse of thromboprophylaxis.	Thank you for this comment. We have addressed this issue by reducing the repetition.
TEP #3	Introduction	Page 11 1st para - A lot of detail about CHADS2. Overlaps with section on risk stratification. Should be rewritten to discuss stroke risk factors identified in Stroke in AF Working Group systematic review. Also, large systematic review on stroke risk factors just published - Pisters R et al Circ J Sept 2012	Thank you for this comment. We have eliminated the redundancy as mentioned and have provided a reference to the Stroke In AF working group. We considered the 2012 Pisters article in the batch of articles screened for potential inclusion in the final report. Since this article is a review, it did not individually meet our protocol-specified criteria for direct inclusion in the CER. We manually reviewed its reference list to identify other studies for screening and potential inclusion.
TEP #3	Introduction	Risk stratification - please see focussed update ESC guidelines (Camm et al EHJ 2012). All the stroke risk schemes have modest predictive value for 'high risk', and past focus on identifying 'high risk' patients was reasonable when we had an inconvenient drug, warfarin.	We agree and have addressed these limitations and the ESC focused update in the Discussion.

Commentator & Affiliation	Section	Comment	Response
TEP #3	Introduction	We are in a new era now - thus, we need a major practice shift (as recommended in focussed ESC guideline 2012) to identify 'truly low risk' patients who do not need any antithrombotic therapy. All other AF patients (with ≥1 stroke risk factors) can be offered effective stroke prevention, which is oral anticoagulation.	As above, we agree that we are in a new paradigm, but these new therapies still carry an increased bleeding risk. We therefore still need tools to separate low risk patients from the rest, which we have attempted to communicate. We have included the ESC focused update guidelines in the Discussion as we summarize the performance of the risk tools and their use clinically there.
TEP #3	Introduction	Page 12, line 18 - HASBLED recommended also by Canadian Guidelines. HASBLED is superior to other scores for assessing bleeding risk, including HEMORRHAGES, ATRIA etc - see Apostolakis et al JACC 2012, Roldan et al Chest 2012, Lip et al Circ Arrhyth Electrophysiol 2012	We have verified that all of the suggested studies have been included in our revised report.
TEP #3	Introduction	Net clinical benefit balancing stroke risk (CHADS2, CHA2DS2-VASc) vs bleeding risk (HAS-BLED) is in favour of OAC rather than not (Friberg et al Circulation 2012, with editorial by Fuster Circulation 2012). This aspect has been totally ignored in this review	Thank you for this feedback. We now contextualize these points in the "Description of Included Studies" sections for KQ 1 and KQ 2.
TEP #3	Introduction	Page 14 - I am not sure why bridging is here. Should be in the Discussion. HAS-BLED is predictive of bleeds during bridging (Omran et al Thromb Haemostat 2012)	We thank the reviewer for their comment but feel that discussing bridging at this spot provides important context.
TEP #3	Introduction	Introduction makes no discussion about aspirin - and how weak the data are, esp in the elderly	Thank you for this comment. We have added a sentence in the Introduction referring to the use of aspirin as an alternative strategy for stroke prevention, despite limited evidence.
TEP #4	Introduction	Comprehensive review of the problems is given in the Intro with the caveats above.	Thank you
TEP #5	Introduction	This section covers the available information on warfarin and the non warfarin oral anticoagulants and new devices and procedures for prevention of stroke in patients with atrial fibrillation. It details the scope of the review and the Key Questions rather well.	Thank you
TEP #6	Introduction	Background and data extraction steps are well defined and stated.	Thank you
TEP #7	Introduction	The introduction is appropriate for the document.	Thank you

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #1	Methods	The inclusion and exclusion criteria are justifiable. The search strategies are explicitly stated and logical. The definitions or diagnostic criteria for the outcome measures are appropriate. The statistical methods used are appropriate.	Thank you
Peer Reviewer #2	Methods	OK	Thank you
Peer Reviewer #3	Methods	Inclusion/exclusion criteria were very conservative, but this is appropriate and helped to highlight the key information gaps. I found the level of description of methodological detail a bit dense and more verbose than I have seen in other evidence summary documents from AHRQ. I suppose readers can simply skip some of that detail, but it contributes to making the read a bit arduous.	We thank the reviewer for these comments.
Peer Reviewer #4	Methods	All items listed are done well.	Thank you
Peer Reviewer #5	Methods	This is a real strength of the report which is very explicit in terms of the inclusion criteria, definitions, and search strategies. The PICOTS criteria are excellent and clear. Definitions and diagnostic criteria and outcomes are well outlined. As mentioined above some sorting as to the clinical importance of some of the outcomes could have been helpful. For example addressing the pre-specified primary outcomes of the new anticoagulant trials and the major secondary outcomes and then major bleeding in some ordered fashion would be helpful rather than an exhaustive list of all of the outcomes. The statistical methods seem appropriate. The summary statistics of the C-stats are helpful in comaprisons for KQ1 and 2.	Thank you. We have ordered the outcomes in a consistent fashion in the revised report.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #7	Methods	The systematic review search strategy appears to be standard, but these searches are ridiculously inefficient and ought to be revisited entirely. I sincerely doubt that the extensive search and reading of abstracts, papers, etc turned up more than a couple articles that were not known to experts in the field. Again, this is not so much a criticism of the team, but the way this whole program is designed and run. So much effort and money are wasted on the search that would be better applied to better analysis.	We thank the reviewer for their comment and share their concern regarding making sure we are as efficient as possible with limited resources when performing systematic reviews. We have passed along this comment to AHRQ, which is always looking to improve the systematic review process and its efficiency.
Peer Reviewer #7	Methods	The whole section on risk scores could be greatly improved. I did not see any consideration of what would be "good practice" in deriving and testing risk scores. I didn't see a summary of how the scores were derived and validated, and how much independent replication in additional populations was done, and how much validation was done by independent teams (ie not the score developers).	Thank you for this comment. While we feel that a discussion of "good practice" in deriving and testing risk scores is beyond the scope of this review, we agree that noting which studies were "derivation studies" for the different risk scores is important. We have added detail to KQ1 (in the Description of Included Studies section) and in the tables and text of KQ 2 such that it should be clear which study was the derivation study for each score, and which were independent cohorts.
Peer Reviewer #7	Methods	There is a framework for evaluating novel risk markers set out by the AHA (Circulation 2009; 119: 2408-2416) that should have been considered and applied to KQ1 and KQ2. One key criterion is incremental value of the score over and above easier and simpler approaches. It certainly would be applicable to assessing these scores. Particularly the use of reclassification measures would have been some "value added". The whole idea of "categorical" scores is really bad here - it discards information, and is a low quality way of assessing risk.	We agree that highlighting an established framework for assessing risk scores will add to our framing of KQ 1 and KQ 2. In order to add context to the KQ 1/KQ 2 sections, we have added discussion of a specific framework for the assessment of stroke and bleeding risk laid out by the AF Optimal Treatment Task Force. We highlight the expert panel's recommended approach to assessing net clinical benefit of anticoagulation in the "Description of Included Studies" section for KQ1 and KQ2. Regarding our reporting of categorical scores, the review is limited by being able to report what is in the published literature. We have however now added in additional information when available regarding the predictive accuracy for categorical vs. continuous scores and our finding of the benefit of continuous scores compared with categorical scores.
Peer Reviewer #7	Methods	On KQ3, the authors say there is "strong evidence" from single large RCTs of an agent. I don't agree that a single trial, even a large one, provides this level of evidence. Multiple RCTs should be required. There is not enough evidence yet about the newer agents What about a network meta-analysis for KQ3?	Although the evidence is based on one large RCT, these studies were assessed to be high quality, direct, precise, and have a low risk of bias. We did not feel that an additional study would change our confidence in the direction or size of the effect. As described, we felt that the included studies differed in important and significant ways making a meta analysis inappropriate.

Commentator & Affiliation	Section	Comment	Response
TEP #1	Methods	The inclusion and exclusion criteria were not appropriate to address the question of whether warfarin was superior to aspirin for prevention of stroke in atrial fibrillation. This is because studies prior to 2000 were not included, including several important high quality RCTs of warfarin vs. aspirin therapy. The paper should be revised to either exclude this question of whether warfarin is superior to aspirin, or the criteria should be revised for this question only, so that the best evidence can be included.	Thank you for this comment. We appreciate the reviewer's concerns about the need to put the findings of our review in relation to warfarin and aspirin into the larger context of investigations conducted prior to 2000. We have now included an introduction to KQ 3 that summarizes previous meta-analyses that focus on the use of aspirin and warfarin for stroke prevention, and comment on this point in the Methods.
TEP #2	Methods	The analytic framework is rather messy and it's hard to understand the logic. Some of the boxes with rounded corners represent interventions, some seem to represent patient populations, others seem to represent potentialy modifying or subgroup characteristics. I don't understand why there are arrows going in two directions from the risk assessment boxes, and they aren't placed correctly in the AF since they don't directly affect patient outcomes; they affect them through use of therapies etc. Also, the individual characteristics don't directly lead to the outcomes so it isn't clear why it has an arrow going straight there. In my opinion the AF can be simplified substantially (I would just take out the individual charactersitics box as wel as the three boxes for the populations for KQ's 4-6 and just include a risk assessment step in between "adults with afib" and the subsequent interventions), and also should use consistent convesntions so that it is clear what the different arrows and boxes mean.	We have revised the Analytic Framework in response to these comments.
TEP #2	Methods	Consider including systematic reviews for studies published prior to 2000, particularly for the risk assessment instruments and warfarin/aspirin.	Thank you for this comment. We appreciate the reviewer's concerns about the need to put the findings of our review in relation to warfarin and aspirin into the larger context of investigations conducted prior to 2000. We have now included an introduction to KQ 3 that summarizes previous meta-analyses that focus on the use of aspirin and warfarin for stroke prevention, and comment on this point in the Methods.
TEP #2	Methods	Should describe which Random Effects model was used in meta-analysis.	We have revised the Methods text to clarify that we used the random-effects model analysis option in Comprehensive Meta-Analysis software (Version 2.2.057; Biostat, Englewood, NJ) which uses the Dersimonian and Laird method.

Commentator & Affiliation	Section	Comment	Response
TEP #2	Methods	There needs to be more description of the methods used to perform meta-analysis for the risk assessment tools. Most methods guidance (e.g., Cochrane, AHRQ) suggest methods to develop summary ROC's. I am not familiar with the method that seeemed to be used in this review (should be described more explicitly) of simply pooling the c-statistics across studies. The authors should cite references if they choose to stick with this method. Leeflang et al. Ann Intern Med 2008;149:889-897; see also the Cochrane Handbook and AHRQ guidance on reviews of diagnostic tests	We agree that the summary ROC approach may be preferred in instances where it can be applied. We have previously utilized that approach in another AHRQ report summarizing performance of diagnostic tests. However, here the available data across studies was limited to the estimated c-statistic with its confidence interval (CI). We agree that the c-statistics is not the best summary of a predictive model performance. If the published literature had consistently reported the raw data we indeed could have considered the Net Reclassification Index (NRI) or Integrated Discrimination Index (IDI) summarizing incremental benefit of a score (biomarker) when added to a model with other covariates. Unfortunately this information was not available from the included studies. We now include this reporting suggestion in the future research section. The c-statistics were pooled by considering their estimated values (point estimates) and confidence intervals, and the "Generic point estimates" effect specification option in the Comprehensive Meta-Analysis software. We have added text to the Methods to clarify these points.
TEP #2	Methods	As mentioned above, the authors should be clear about describing the c-statistic as a measure of discrimination. Also, it is not clear to me why they did not evaluate other measures used to assess risk assessment instruments such as measures of calibration, or strengths of association (e.g., RR's/OR's) or measures of diagnostic accuracy (sensitivity, specificity, LR's). All of these provide different information about the usefulness of risk prediction instruments; the shortcomings of the c-statistic as the sole measure of a risk prediction instrument has been well described. e.g., the c-statistic doesn't tell us whether higher scores on the CHADS2 are associated with progressively higher risk, or how much risk a score of 5 confers compared to a score of 2, etc. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N. et al., Assessing the performance of prediction models: a framework for traditional and novel measures.. Epidemiology. 2010;21128-38 Cook NR. Circulation 2007;115:928-935.	We appreciate the reviewer's comment. One of the biggest challenges for KQ 1 and KQ 2 was the inconsistency in how the results of included studies were reported. Not all studies included measures of strength of association, and very few included measures of diagnostic accuracy. When such measures were included, they were often based on different iterations of the risk scores in question (e.g., categorical scores for some, continuous for others). For this reason, we included event rate data for all studies, as this was most consistently reported, and included c-statistics, as well, because these also tended to be frequently reported. When available, we also reported NRI (though this was available for few studies). We have highlighted the limitations of using the c-statistic to evaluate risk prediction tools throughout the document – that it is a measure of discrimination. Also, we have highlighted consistent reporting of measures for calibration, strength of association, diagnostic accuracy, and discrimination as an important feature for future studies of stroke and bleeding risk prediction tools.

Commentator & Affiliation	Section	Comment	Response
TEP #2	Methods	In comparing risk assessment instruments the report doesn't attempt to evaluate risk reclassification rates (e.g., what proportion of patients are correctly re-classified into different risk categories), which are considered more clnically informative than the c-statistic alone (or measures of calibration or risk estimates) for assessing clinical usefulness. Cook NR and Ridker P. Ann Intern Med 2009;115:928-935	We appreciate the reviewer's comment. Measures of risk reclassification, such as net reclassification improvement, were infrequently or inconsistently reported in the included studies, as were other measures such as hazard ratios and likelihood ratios. In order to address this concern, we have fleshed out the discussion of the inconsistent reporting of relevant statistics by the included studies, highlighted the limitations of c-statistics, and added information on NRI where available for KQ 2. Of note, for KQ 1, none of the included studies conducted NRI.
TEP #3	Methods	Generally OK Page 18, last para - limitations of c-statistics should be discussed, and when comparing different schemes, IDI and NRI should be used. Also, comparing c-statistic from one study with that from another is like comparing apples and oranges, given the inter-study heterogeneity	We appreciate this comment. We have removed inter-study comparisons of c-statistics and have highlighted the limitations of c-statistics in assessing risk prediction models. We have added information on NRI where available for KQ 1/2, but measures of risk reclassification, such as net reclassification improvement, were infrequently or inconsistently reported in the included studies, as were other measures such as hazard ratios and likelihood ratios.
TEP #4	Methods	Inclusion exclusion criteria are very good for population studies. Less good for generating the universe of controversial issues in managing patients on anticoagulation.	We thank the reviewer for their comments and hope they feel that the revised report addresses their concerns.
TEP #5	Methods	Methodology covers the steps in data extraction and assessment of the studies. Importantly, the strength of the body of evidence is systematically rated using the latest accepted tools in the field. When appropriate, meta analysis techniques have been applied to the available literature.	Thank you
TEP #6	Methods	statistical methods of using meta-analysis with random-effects models are appropriate. However, inference from meta-analysis should be used with caution due to very large heterogeneity between studies. The report on page 12 line 45 stated a measure of "% reduction" and it's not clear if this point estimate (95% CI) is calculated on relative risk reduction scale or not? (ie., 1-RR or 1-HR). Table C, page 28 line 40, the sentence need correction. page 29 line 31, the point estimates (95% CIs) are reported as HRs, however, Connolly et al 2009 reported RE-LY efficacy endpoints as relative risk (RR). Table C may need a footnote for major bleeding to state the criteria used such as ISTH or TIMI major?	We thank the reviewer for the careful review of the report. We now clarify that the % reduction was a relative risk reduction. We have corrected the typo in Table C and changed the HR to RR as appropriate Although not listed in the summary tables, information about how major and minor bleeding was defined in the studies is listed in the text when available/appropriate.

Commentator & Affiliation	Section	Comment	Response
TEP #7	Methods	The methods are appropriate and well described.	Thank you
Peer Reviewer #1	Results	The amount of detail presented in the Results section is appropriate. The characteristics of the studies are clearly described. The key messages are explicit and applicable. The figures, tables, and appendices are adequate and descriptive. The investigators did not overlook any studies that ought to have been included or included any study that ought to have been excluded.	Thank you
Peer Reviewer #3	Results	Very inclusive and detailed. Excellent description of results.	Thank you
Peer Reviewer #4	Results	All items listed are done well.	Thank you
Peer Reviewer #5	Results	The details are very extensive and could be summarized in tables and figures more succintly with reduction in some of the text. Studies are well described and appropriate. They have included the appropriate studies with no clear gaps. The key messages could be made more explicit. In the interest of being complete the authors make it easy for the reader to get lost in the mountain of details.	We have revised the key points throughout the report and hope the reviewer feels the revised report is clearer.
Peer Reviewer #7	Results	I am not convinced that the CHA2DS2-VASc score is as good as the authors suggest. The c-index values given look almost exactly the same in those studies that tested them both in the same population. This is exactly where risk reclassification would be the most useful. In the case of rough equivalence, certainly the preference should be to use the simpler score (CHADS2) that has been more widely used and validated by multiple independent groups. I don't think the data presented back up the conclusion of the report here.	Thank you for this comment and we agree that CHA_2DS_2-VASc does not appear to be much better at risk stratification than $CHADS_2$. We have removed the statement of superiority. Additionally, we have noted that within our Discussion section that CHA_2DS_2-VASc is used more commonly in Europe and has been incorporated into their atrial fibrillation practice guidelines. The score is not routinely applied in the US and is not in the current AHA/ACC guidelines.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #7	Results	Similarly, the added value of the HASBLED score vis a vis the others is not well laid out here. What about the incremental value over simple things like a prior history of bleeding? application of a better framework for evaluation should have been considered here.	Thank you for this comment. We attempted to follow a framework for assessing risk scores, which was established by the AF Optimal Treatment Task Force. We highlight the expert panel's recommended approach to assessing net clinical benefit of anticoagulation in the "Description of Included Studies" section for KQ 2. This group recommends using established bleeding risk scores to determine bleeding risk in patients whose stroke risk justifies anticoagulation. With respect to comparing established bleeding risk scores to individual risk factors, due to the large number of individual factors and bleeding risk tools, this was felt to be beyond the scope of this already large project. Although the 95% confidence intervals on the c-statistics overlap between scores, many of the point estimates when given direct comparison of scores are better for HAS-BLED than for the other scores. In addition the net reclassification improvement data is promising for the HAS-BLED score. These led us to suggest a potential benefit of the HAS-BLED score albeit it with low strength of evidence/limited confidence.
Peer Reviewer #7	Results	Table B - The claim that HASBLED outperforms the others is not well supported here. There is no overall c-index, and no formal meta-analysis (or head to head comparisons) cited. I would not fully trust developers of score A comparing their product with score B developed by someone else.	We agree with the reviewer's concern and have moderated our discussion of HAS-BLED accordingly. Although we still feel that the evidence supports a suggested discrimination benefit of HAS-BLED based on the direct comparison studies, and those reporting the net classification improvement scores – we have rated this finding to have low strength of evidence. Note that our low strenght of evidence rating indicates that we have "Low confidence that the evidence reflects the true effect. Further research is likely to change the confidence in the estimate of effect and is likely to change the estimate."
Peer Reviewer #7	Results	The data on ³tolerability² of the new agents vs warfarin should be summarized in Table C and the text of the Executive Summary. The whole rationale for the newer agents is that they are supposed to be easier to use and better tolerated, but what are the data that prove this?	Thank you for this comment. We have highlighted the risks associated with use of dabigatran re: risk of MI as well as dyspepsia and GI side effects in the Discussion section and the SOE tables.
Peer Reviewer #7	Results	Should distinguish procedures with short-term only problems from PCI w stents, where long-term antiplatelet tx is needed	We now indicate in KQ 4 which studies focused on short-term only problems as suggested
TEP #1	Results	A recent analysis of RCT data relevant to KQ4 was not included (Circulation 2012 Jul 17;126(3):343-8).	This study by Healey and colleagues was captured in our updated search and is included in the revised final report.

Commentator & Affiliation	Section	Comment	Response
TEP #2	Results	for KQ I think there is a typo in Table A and throughout other parts of theh report, at least according to the Forest plot the CI For Framingham should be 0.61 to 0.64, not 0.61 to 0.74. The correct CI of course will have an impact on the conclusions. It will be important to be clear that conclusions are based on indirect comparisons of c-statistics across studies, unless there are studies that directly compared the instruments. See above for my concerns about relying solely on the c-statstic and doing a simple pooling.	Thank you; the summary forest plot and CI for the Framingham meta-analysis has been updated throughout the report to reflect the inclusion of a new study identified during the literature search update. We have also added language to be clearer about which studies are providing data from direct vs. indirect comparisons.
TEP #2	Results	For KQ 2 it's hard to see that the CI's for the c-statistics really differ at all for the different risk insruments (they really seem to overlap). Also Table 15 showing direct comparisons really shows little difference between HAS-BLED and HEMORR2HAGES, esp Pisters and Olesen, where the confidence intervals overlap and there is no real difference in the point estimates (esp for Olesen). So the conclusion that HAS-BLED has the highest discrimative accuracy seems quite overstated, both for patients on and not on anticoagulation.	We agree with the reviewer's concern and have moderated our discussion of HAS-BLED accordingly. We now rate the strenght of evidence as low indicating the uncertainty in the finding but feel that the direct comparison data (Tables 15-17) and the net reclassification improvement data from Table 18 support this limited confidence in the HAS-BLED score.
TEP #2	Results	See previous comment about needing to include at least a summary of older trials on warfarin and aspirin. Otherwise it looks like the only evidence on warfarin vs. aspirin are two observational studies, which is simply not true.	Thank you for this comment. We have now included an introduction to KQ 3 that summarizes previous meta-analyses that focus on the use of aspirin and warfarin for stroke prevention, and comment on this point in the Methods.
TEP #2	Results	I didn't really see any exploration of heterogeneity, which was extremely high for a number of the risk assessment meta-analyses. Again I'm not sure that simple meta-analysis is the right way to go here, but if you do stick with it, it is critical to investigate heterogeneity through subgroup/sensitivity analyses etc.	We agree that the heterogeneity in the KQ2 meta analyses was high and those analyses are no longer performed. For KQ1 we discuss the heterogeneity of the findings and how our SOE ratings were reduced because of it. In addition, we chose a random-effects approach due to substantial variability between the studies.
TEP #3	Results	KQ1 - see comments above in Intro box ... we need to move away from the obsession to 'identify' high risk, when all risk schemes based on clinical factors only have modest priorityt	We have modified the text to highlight the focus on identifying low risk patients. However, identifying the patients still requires understanding those factors that would place patients in a 'high risk' category, so making sure those risks are identified and quantified is important.

Commentator & Affiliation	Section	Comment	Response
TEP #3	Results	KQ2 - need to add in the new papers by Apostolakis et al JACC 2012, Roldan et al Chest 2012, Lip et al Circ Arrhyth Electrophysiol 2012. Line 34, page 23 - no studies comparing HAS-BLED and ATRIA' is WRONG! Why would you use stroke scores CHADS2 and CHA2DS2-VASc for major bleeding events? Does not make sense! Table 10 is just wrong! See Gallego et al Circ Arrhyth Electrophysiol 2012 - HASBLED predicts bleeding as good as a multivariate analysis. Has less good predictive value for CV events (and less good compared to a multivariate analysis) HAS_BLED has a good predictive value for ICH - see Apostolakis et al JACC 2012	We appreciate this input, and have incorporated all of the suggested references (which were published after our initial inclusion window). We have also removed discussion of stroke risk scores throughout the bleeding risk score section. As part of the review, we found numerous studies that examined the utility of stroke risk scores, such as $CHADS_2$ and CHA_2DS_2-VASc, to also predict bleeding risk. We initially felt it was important to evaluate these studies in KQ 2. However, given that these stroke risk scores are not used clinically for bleeding prediction, we have removed these from the KQ 2 section to assure clarity for readers.
TEP #3	Results	KQ3 - very confusing, mixing up trials, cohorts and retrospective analyses. I think this needs to be re-presented with RCTs, then large cohorts etc Layout of Table C is confusing, and does not allow comparisons between different interventions Comment probably needed on the numerous indirect comparisons between the different NOACs that have recently been published	Thank you for this comment. We have tried to present results by comparison of drugs and by outcome which we felt would be a meaningful way to present current findings regarding stroke prevention for clinicians. Unfortunately, because of the different trial designs within these categories, we have had to describe results with different study types in the same sections. We did not include studies with indirect comparisons of NOACs given the heterogeneity of the studies. Rather, we feel that head-to-head comparisons between NOACs are needed.
TEP #3	Results	KQ4 - large registries eg AFCAS have been published. Also, consider systematic review in the consensus document from the ESC Working Group on Thrombosis - see Thromb Haemost. 2010 Jan;103(1):13-28. Also, North American consensus document - Thromb Haemost 2011 Oct;106(4):572-84.	We appreciate these suggestions. The article by Lahtela et al. reporting data from the AFCAS registry was identified in our search update, screened, and included for abstraction. We now present data from this study in KQ 4. We also manually reviewed the component references of the two consensus documents to identify additional articles for screening and potential inclusion in the report.
TEP #3	Results	KQ6 - Friberg et al EHJ 2012 shows that prior ICH increases risk of subsequent stroke by 49% Also, see Arch Intern Med 2012; DOI:10.1001/archinternmed.2012.4261.	Although the Friberg study listed did not meet the specific inclusion criteria for our KQ 6 studies, it is included in our updated draft for KQ 1 and KQ 2. Also note that the Archives of Internal Medicine study listed was screened during our review but was excluded because it did not allow atrial fibrillation patients to be evaluated as a specific subgroup of interest.
TEP #4	Results	Data tables are extremely comprehensive.	Thank you

Commentator & Affiliation	Section	Comment	Response
TEP #5	Results	The authors have provided a succinct review of each major study along with conclusions for studies related to each KQ followed by a summary later on in the discussion section. The quality of the studies, region where the studies were conducted, and source of funding when available, have been adequately outlined. Tables provide additional information regarding the strength of evidence. Exclusion of non English language papers is appropriate.	Thank you
TEP #6	Results	The results are summarized very well for the intended purposes. Given large patients contribution from countries other than US, the report could have summarized some geographical difference especially with regard to time in therapeutic range (TTR) for patients treated with warfarin. Suggest not to use “high strength of evidence” when the 95% CI includes unity since the evidence is inconclusive. Figures and the tables were very useful and adequate.	Given the focus of the report on the US population, we did not include additional details about the varying geographical differences in TTR. We do however include information about the location of the included studies for each key question. Those cases with “high strength of evidence” when the 95% CI includes unity correspond to outcomes where the evidence suggests that there is *no difference* between the two treatments and therefore it is appropriate to include unity.
TEP #7	Results	The results are well written and appropriate. the level of detail is appropriate for the document. The key messages are clearly stated. The figures and tables are clinically useful and clear. I am not aware of any relevant studies not included by the authors.	Thank you

Commentator & Affiliation	Section	Comment	Response
Public Reviewer #4 – Judy Wagner	Results	Page: 60, Para: 1, Line: 1 Existing Text: For the section entitled “Ischemic or Uncertain Stroke” only data for apixaban vs. warfarin was presented. Recommendation: Include data from ROCKET AF on rivaroxaban vs. warfarin on ischemic stroke. Rates were: 1.34% per year for rivaroxaban vs. 1.42% per year for warfarin; HR 0.94;95% CI,0.75, 1.17; p=0.581. (safety on treatment population). Page(s): 60, Para: 2, Line: 1 Existing Text: For the section entitled “Hemorrhagic Stroke” only data for apixaban/idraparinux vs. warfarin was presented. Recommendation: Include data from ROCKET AF on rivaroxaban vs. warfarin on hemorrhagic stroke. Rates were: 0.26% per year for rivaroxaban vs. 0.44% per year for warfarin. HR 0.59;95% CI,0.37, 0.93, p=0.024. (safety on treatment population). Page(s): 60, Para:4, Line: 1 Existing Text: For the section entitled “Systemic Embolism” only data for apixaban/idraparinux/edoxaban vs. warfarin was presented. Recommendation: Include data from ROCKET AF on rivaroxaban vs. warfarin on systemic embolism. Rates were: 0.04% per year for rivaroxaban vs. 0.19% per year for warfarin. HR 0.23;95% CI,0.09, 0.61, p=0.003 (safety on treatment population). Page: 61, Para: 3, Line(s): 1 Existing Text: For the section entitled “Death from Cardiovascular Causes” only data for apixaban/edoxaban vs. warfarin was presented. Recommendation: Include data from ROCKET AF on rivaroxaban vs. warfarin on vascular death. Rates were: 1.53% per year for rivaroxaban vs. 1.71% per year for warfarin. HR 0.89;95% CI,0.73, 1.10, p=0.289. (safety on treatment population). Reference for all comments above: Supplement to: Patel MR, et al. NEJM 2011;365:883-91. Available online at: www.nejm.org/action/showSupplements?doi=10.1056%2FNEJMoa1009638&viewType=Popup&viewClass=Supplpg 17	Thank you for these suggestions. We have now added these secondary efficacy outcomes from ROCKET and have clarified that these are from the on-treatment population.

Commentator & Affiliation	Section	Comment	Response
Public Reviewer #4 – Judy Wagner	Results	Page: 64, Para 2 Existing Text: In the section with evidence regarding “patients with AF and renal impairment”, the report cited data from a substudy of ROCKET-AF, the pivotal study for rivaroxaban. Yet in the last Key Point (page 49), the report implied that all new anticoagulant agents are comparable to warfarin, despite a lack of evidence for the other new agents. Recommendation: Include a separate key point for patients with renal impairment, and specify that these patients benefit equally from rivaroxaban when compared with warfarin.	As suggested, we now include a separate key point focusing on patients with renal impairment and clarify which new anticoagulants have been studied in this population.
		Page 72,Para: 2, Line 1 Existing Text: Section entitled “Elderly Patients with AF” Recommendation: Include results from the ROCKET AF substudy in elderly patients with nonvalvular AF. Results demonstrated that the overall relative effects of rivaroxaban versus warfarin were consistent among the elderly (age ?75) and younger patients for both efficacy and safety. Data were presented as an oral presentation at the International Stroke Conference February 1-3, 2012 (see full reference citation below). Reference: Halperin JL, et al. Efficacy and safety of rivaroxaban compared with warfarin among elderly patients with nonvalvular atrial fibrillation in the ROCKET AF trial. Oral presentation and abstract 148. Presented at International Stroke Conference Feb 1-3, 2012; New Orleans, LA.	Thank you for this suggestion. As part of our review criteria, we excluded information published only in abstract form since that data has not yet passed a full course of peer review. Since this ROCKET AF subgroup data is not yet available in a full-length peer-reviewed format, we have not incorporated it into the report.
		Page: 84 Table Existing Text: Table for Xa inhibitor (rivaroxaban) versus warfarin Recommendation: Include the following data points for rivaroxaban to ensure consistency with data presented for apixaban and dabigatran in the tables: • Ischemic stroke • Hemorrhagic stroke • Systemic embolism • Death from vascular causes Reference: Supplement to: Patel MR, et al. NEJM 2011;365:883-91. Available online at: www.nejm.org/action/showSupplements?doi=10.1056%2FNEJMoa1009638&viewType=Popup&viewClass=Supplpg 17	We have incorporated findings from the suggested citation both into the text of KQ 3 and in to the summary strength of evidence table.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #1	Summary/ Discussion/ Conclusion	The implications of the major findings are clearly stated. The limitations of the review/sudies are described adequately. No important literature was omitted in the Discussion section. The future research section is clear and easily translated into new research. Research gaps were identified for all 6 key questions including research gaps for risk stratification for thromboembolic risk and bleeding risk, comparative effectiveness and safety of different anticoagulation strategies, and comparative effectiveness and safety of changing anticoagulation therapies for different reasons.	Thank you
Peer Reviewer #2	Summary/ Discussion/ Conclusion	OK	Thank you
Peer Reviewer #3	Summary/ Discussion/ Conclusion	Unfortunately, the conclusion in a number of the key areas was that there were not sufficient, well performed studies to draw firm conclusions. This is not a reflection of the review, but a comment on the field and level of knowledge. Research gaps were well identified as a result.	Acknowledged
Peer Reviewer #4	Summary/ Discussion/ Conclusion	This section was very well done. I especially appreciated the barriers to implementation section	Thank you
Peer Reviewer #5	Summary/ Discussion/ Conclusion	The discussion re-states some of the results and seems repetitious in some areas. The Findings in Relationship to What is Already Known is very well-done and the most clinically meaningful. Some of the remaining issues and questions regarding these new agents are mentioned here. The lack of FDA approval of apixaban should be mentioned somewhere sicne it is not yet publicaly available. The gaps and future research questions are excellent and can be translated into new research studies.	Acknowledged. Since apixaban has now received FDA approval, we have not added the specific statement recommended by the reviewer; however, we have added statements throughout the report to note the status of other interventions that are not yet FDA approved.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #7	Summary/ Discussion/ Conclusion	Discussion is generally reasonable. But there are really only 3 large trials of the new agents, the conclusisons can't be very strong. Might the shorter half-lives (pharmacodynamic) of these drugs be a disadvantage in poorly compliant patients? No known reversibility agents for the new drugs.	Thank you for this comment. The population size for each of these trials was very large, allowing for robust results and conclusions regarding the comparisons of the NOACs with warfarin, which is what each of these large trials examined. However, we do feel that conclusions cannot be made between the NOACs and that head-to-head comparisons are needed. Regarding the reversibility of the NOACs, none of the studies we examined addressed this issue, and we have included this as an important research gap. We have added a sentence about the potential negative of the shorter-half life in poorly compliant patients to the discussion.
Peer Reviewer #7	Summary/ Discussion/ Conclusion	Surely the lack of cost-effectiveness studies, or at least documenting economic outcomes, is a research gap for new agents. I'm surprised this wasn't covered in the report.	Although we agree that the cost effectiveness of the different agents would be of interest to many readers, this outcome was considered outside of the scope of this project and therefore is not covered in this report.
TEP #1	Summary/ Discussion/ Conclusion	Yes.	Thank you
TEP #2	Summary/ Discussion/ Conclusion	As mentioned above I think the possible association between dabigatran and increased risk of MI deserves more discussion (there may have been few or small events and the confidence intervals are close to 1 or just cross it; it's worth stating these things).	Thank you for this comment. We have highlighted the increased risk of MI with dabigatran in the discussion section of KQ 3.
TEP #2	Summary/ Discussion/ Conclusion	Might bring up the issue of adherence and how it might affect results of efficacy trials vs. real-world practice (might make the trials look better for warfarin than they are in practice).	Thank you for this comment. We have added additional text to the Discussion to expand on the need for real-word data on the new anticoagulants, including the opportunity offered by the growing prevalence of electronic health records to monitor and evaluate the real world uptake of these therapies.
TEP #2	Summary/ Discussion/ Conclusion	An important applicability issue is that few trials reported CHADS2 (or presumably other) risk assessment scores or bleeding risk scores. Future studies need to provide this infromation.	Thank you for this comment. We have addressed this concern within our future research needs section.
TEP #3	Summary/ Discussion/ Conclusion	Key findings - disproportionate text on NOACs. Much of the KQs were on stroke and bleeding risk, so why limit the discussion?	The key findings list the findings across each key question but the reviewer is correct that much of the discussion focuses on the newer treatment options and risk stratification tools which are the focus of most of the recently published literature. We feel that the areas highlighted in the discussion are those where there is either the newest evidence needing synthesis or where there are identified future research needs.

Commentator & Affiliation	Section	Comment	Response
TEP #3	Summary/ Discussion/ Conclusion	ESC guideline 2012 also recommends CHA2DS2-VASc as the preferred stroke risk assessment score, given the focus on identification of low risk patients. This should be discussed	We have noted that CHA_2DS_2-VASc is used more commonly in Europe and has been incorporated into their atrial fibrillation practice guidelines. The score is not routinely applied in the US and is not in the current AHA/ACC guidelines. We have not emphasized CHA_2DS_2-VASc in our report since the audiences for this document are primarily in the US.
TEP #3	Summary/ Discussion/ Conclusion	Research Gaps - KQ1 - need to discuss what is recommended in the ESC guideline, to focus on identification of low risk (rather than being obsessed with 'high risk') See also Potpara et al Circ Arrhyth Electrophysiol 2012	As mentioned above, we have modified the text to highlight the focus on identifying low risk patients. However, identifying the patients still requires understanding those factors that would place patients in a 'high risk' category so making sure those risks are identified and quantified are important.
TEP #3	Summary/ Discussion/ Conclusion	Page 37, line 14 - CHA2DS2-VASc score should be mentioned	Acknowledged
TEP #4	Summary/ Discussion/ Conclusion	Major missing part is around the reversal of anticoagulation in patients who have emergency bleeding conditions with warfarin vs. the direct thrombin inhibitors. This is of major concern for patients and their physicians, especially in the large elderly population.	Thank you for this comment. This has been included as an important research gap and we have added some additional discussion of this concept in the "findings in relation to what is already known" section
TEP #5	Summary/ Discussion/ Conclusion	The discussion section provides a brief summary for each KQ and is supplemented by summary tables outlining the strength of evidence (extensive four pages table for KQ3). The Discussion outlines the findings of this Comparative Effectiveness Review (CER) in relation to what is already known and its implications for clinical and policy decision making, as well as, potential issues with applicability of the included studies to the usual clinical practice. The limitations of this CER process have been outlined by the authors. The research gaps for all Key Questions have been described.	Thank you

Commentator & Affiliation	Section	Comment	Response
TEP #6	Summary/ Discussion/ Conclusion	Conclusions were well summarized and addressed the key questions. Further research in this area needed with respect to bridge patients who undergo invasive procedures as well as switching patients on warfarin to one of the new anticoagulant therapy. The report could have mentioned an important AF study currently ongoing and funded by NHLBI "Catheter Ablation Versus Antiarrhythmic Drug Therapy for Atrial Fibrillation Trial—CABANA", once completed this study will provide valuable information to health policy makers on strategies in treatment of AF.	Thank you for this comment. We do include issues surrounding bridging for procedures as an important research gap (KQ 4). CABANA was not included in this report as its primary focus is not on stroke prevention, although we agree that it will provide important information regarding the treatment of atrial fibrillation.
TEP #7	Summary/ Discussion/ Conclusion	The major findings are clearly stated, and the discussion is generally appropriate. I would have liked to see more discussion about important limitations of the novel anticoagulants (dabigatran, apixaban, and rivaroxaban) that while the primary randomized trials were large, we have very little in the way of real-world safety data. Furthermore, the lack of reversal agents for these drugs make the potential issues of bleeding, especially around surgical procedures potential issues that should have been more discussed in the document.	Thank you for this comment. We have included these important issues regarding reversibility and bridging for invasive procedures (KQ 4) as research gaps.