To the Editors,

We thank the reviewers for their thoughtful review of our science. We have addressed concerns comprehensively in a point-by-point fashion, making changes and edits where appropriate.

Reviewer 1

1. **Consider using an alternative term to NOAC, since these agents are no longer "novel". Perhaps something like "target-specific oral anticoagulants (TSOACs)". Also, either mention all of the available agents in the 1st paragraph (vs only dabig & riva), or just mention the drug classes as a whole.**

   We appreciate the reviewer’s concern regarding the common reference to this class of drugs as novel oral anticoagulants (NOAC). However, we do not believe it is appropriate to adopt a new vernacular as suggested by the reviewer given NOAC is the current industry standard accepted by the pharmaceutical industry, the US Food and Drug Administration, the European Medicines Agency and cardiology communities. To introduce new terminology when these products have only been available for less than 3 years would be unnecessarily confusing to the wider audience who will be interested in this scientific work.

   We have adjusted our text in the first paragraph to accommodate the request that we mention the drug class as a whole, instead of naming specific agents. If the reviewers and editors feel strongly about the nomenclature, we would be happy to change it to ‘direct oral anticoagulants’ as the mechanism for each of the new agents is direct while that of warfarin is indirect.

2. **Why were inducers/inhibitors of warfarin specifically mentioned, but not agents interacting with either dabigatran or rivaroxaban?**

   There is very little data regarding important clinical drug-drug interactions involving dabigatran or rivaroxaban. The current pharmacodynamic data suggests a potential interaction with inhibitors of the CYP3A4 and P-gp enzyme systems¹. Drugs such as azole antifungals or systemic HIV protease inhibitors may pose a problem. However, there have been few reported cases to date and data from the RE-LY trial showed potential interactions did not meet the specified criteria², therefore, no labeled interactions were required by the FDA. In fact, it has been suggested that drug interactions are far less common with NOAC than seen with warfarin³. In comparison, there is a large body of literature demonstrating the clinical relevance of drug-drug interactions with warfarin. Thus, to exclude the warfarin interactions would lead to potential systematic misclassification errors in exposure-outcome assessment. In contrast, given the paucity of real-world knowledge regarding the clinical impact of theoretic drug-drug interactions associated with NOAC, the inclusion of these drugs as a covariate in models would be speculative, at best. The far more important modifier of NOAC toxicity is renal
dysfunction, which has been shown to impact adverse events caused by prolonged effects of the anticoagulants. We have appropriately adjusted our analyses for renal dysfunction.

3. **Also, consider using the more universally recommended CHADS2-VASc stroke risk score rather than the outdated CHADS2.**

We believe it is appropriate to use the CHADS2 score in this manuscript as all the existing NOAC trials and observational studies use the CHADS2 score, thus, our use of the CHADS2 score allows for greater comparability with the existing literature. Furthermore, a study comparing several stroke risk stratification tools for atrial fibrillation, CHADS2-VASC did not outperform CHADS2. The authors, Gregory Y.H. Lip et al. found that the CHADS2 and CHADS2-VASC schema demonstrated broadly similar predictive ability for thromboembolism events (0.64 and 0.65, respectively). However, we did re-run all the atrial fibrillation analyses using CHADS2-VASC. Overall, the results were similar and are presented below. While we believe that it would be better to include the results with CHADS2 due to the comparability to other studies, we would be happy to add the CHAD2-VASC results to either the main manuscript or to the appendix if the reviewers or editors would prefer.

<table>
<thead>
<tr>
<th></th>
<th>Events per 100 Patient-Yrs.</th>
<th>Dabigatran vs Warfarin Bleeding</th>
<th>Events per 100 Patient-Yrs.</th>
<th>Rivaroxaban vs Warfarin Bleeding</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Hazard Ratio* (95% CI)</td>
<td></td>
<td>Hazard Ratio* (95% CI)</td>
</tr>
<tr>
<td><strong>Atrial Fibrillation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total Bleeding Events</td>
<td>Dabigatran</td>
<td>2.29</td>
<td>Warfarin</td>
<td>3.10</td>
</tr>
<tr>
<td>Upper GI Bleeding Events</td>
<td>Dabigatran</td>
<td>1.43</td>
<td>Warfarin</td>
<td>1.74</td>
</tr>
<tr>
<td>Lower GI Bleeding Events</td>
<td>Dabigatran</td>
<td>0.86</td>
<td>Warfarin</td>
<td>1.36</td>
</tr>
</tbody>
</table>

4. **I am not clear on why so many differences remained different between groups (particularly in the rivaroxaban non-AF patients) following propensity-score matching. Did the model that was used for matching not fit the data appropriately? Please provide the model diagnostics to support/refute this. Were too few variables used to match?**

We examined the standardized differences for each covariate in the PS model and few were >0.10 (10%); these were controlled for in the multivariable model. The standardized differences compare the difference in means in units of the pooled standard deviation. Furthermore, it is not influenced by sample size and allows for the comparison of the relative balance of variables measured in different units. See supplemental figures A1-A4 in Appendix.
5. **How do you explain why the hazard of GI bleeds was higher in the non-AF group with dabigatran vs. warfarin? While not statistically significant, the results are trending opposite of the AF cohort. Why might this be?**

Both the underlying rates of bleeding as well as the hazard ratio for the non-AF group was different compared to the AF group. During the period of our study, the use of dabigatran for indication other than atrial fibrillation was off-label and resulted in the relatively small sample size. It is likely that the use during this time frame was driven by physician preferences and the patients are likely a very different group who would now be prescribed the drug now that Dabigatran has broader clinical indications. To further comment on the reasons for these results in text would be speculative at best.

6. **When discussing the results of the rivaroxaban analyses, please keep wording consistent. Stating that there were numerically, albeit not statistically significantly, fewer events with riva vs. warf in the AF-cohort, while saying "similar rates of GIB when compared to warfarin" in the non-AF cohort is inconsistent. The non-AF cohort had confidence intervals much closer to statistical significance than the AF cohort.**

We agree with the reviewer and have changed the language so that it suggests similar rates of GIB for both AF and non-AF analyses.

7. **In the first paragraph, why was only the AF findings mentioned, and the non-AF findings ignored?**

As requested, we have summarized the non-AF findings in the first paragraph of the Discussion.

8. **When discussing the differences in age between your cohort & the clinical trials, you seem to suggest that the difference of 4 years in mean age could explain the differences in GIB rates. Please substantiate how this magnitude of age difference relates to significant GIB rates.**

From a GI bleeding perspective, our real-world population should have had an inherently lower risk of GI bleeding when compared with RCT. They differ from trial populations in two important GI bleeding risk factors—mean age (median 67 years in our study vs. 71 years in RCTs) and CHADS2 score (the majority of patients in our study had a CHADS2 score of 0-1 vs. mean CHADS2 scores of 3.5 in ROCKET-AF and 2.1 in RE-LY and ARISTOTLE RCTs.)
Current GI bleeding literature supports the notion that the risk of GI bleeding increases with age and with associated co-morbidity. These two risk factors exceed that of concomitantly prescribed medications when assessing overall risk of GI bleeding. Our data is consistent with these data. Patients with a lower cardiovascular co-morbidity score for stroke (i.e., lower CHADS2 score) and at lower age were at less risk of GIB. However, with increasing age, GIB increased with both NOAC.

9. **Is there a mechanistic rationale for why one might expect differences in upper vs. lower GIB rates with these agents?**

The epidemiology of GI bleeding is shifting with the ageing population and the widespread use of proton pump inhibitors (PPIs) and *Helicobacter pylori* eradication. PPIs and *H. pylori* eradication have decreased the overall burden of peptic ulcer disease-related upper GI bleeding, however, there has been an increase in lower GIB events associated with the aging population. This is due to an increased prevalence of age-related GI mucosal defects which predominantly affect the lower GI tract (i.e., diverticulosis, angiodysplasia and other vascular abnormalities) and the increasing rates of cardiac co-morbidity in the ageing population which has contributed to increased use of anticoagulants either alone or in combination with aspirin or other thienopyridine agents. These pharmacologic agents are known to increase the risk of GI bleeding associated with mucosal defects or vascular lesions.

10. **Figures 2 & 3: Consider adding the p-value for each comparison to allow for easier interpretation of the data within the figures (same for those in the appendix).**

As requested, we have labeled the p-value for each comparison in Figure 2 & 3 and Figure A5 & A6. However, we believe that instead of improving the clarity of the data, the figures have become more difficult to interpret.

Reviewer 2

1. **The authors never state whether they included GI Hemorrhage on inpatient claims only, or inpatient and outpatient (and whether they included all GI hemorrhage diagnoses, or only those appearing in the primary position on claims).**

We agree, we have modified the manuscript to reflect the below statement.

The primary outcome was the occurrence of a gastrointestinal bleed (GIB) as defined using the Food and Drug Administration’s Mini-Sentinel post-marketing surveillance system diagnostic codes (ICD-9-CM codes used to define these outcome are listed in supplemental Table A1 in the Appendix). Gastrointestinal (GI) bleeding was
identified by using inpatient hospital claims for relevant primary and secondary discharge diagnoses indicative of bleeding in the GI tract.

2. How were different dosages handled (e.g., 150 mg and 75 mg for dabigatran). Did the study include 75 mg dabigatran as well as 150 mg? Given the relatively infrequent use of 75 mg dabigatran, presenting results for 150 mg only is perhaps more relevant, or at least presenting results for 150 and 75 mg separately.

The majority of our patients were prescribed 150 mg of dabigatran, which is the more common dose in the United States. Despite this increased dose, our results still demonstrate a more favorable risk-benefit profile of dabigatran, until patients reach an advanced age. The Graham Medicare study also demonstrated no protective dose-effect associated with dabigatran 75 mg\textsuperscript{11}. We have presented our data in a stratified fashion by dose (below) to demonstrate the comparative results. Overall, the sample size for the 75mg group was relatively small. For the atrial fibrillation group, the rates of GIBs are lower for the 150mg dose. However, we have chosen not to present the data separately in the manuscript as the 150 mg capsule was the predominant prescription strategy in our dataset and we agree with the reviewer, it is the more relevant dose and the primary driver of the overall results. We would be happy to add these results to the appendix if the reviewers and/or editors would prefer.

Events of GI Bleeding Events by dose

<table>
<thead>
<tr>
<th></th>
<th>Dabigatran with 75mg Capsule</th>
<th>Warfarin N = 594</th>
<th>P-value</th>
<th>Dabigatran N = 7,155</th>
<th>Warfarin N = 7,155</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atrial Fibrillation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total Bleeding Events</td>
<td>13 (2.19%)</td>
<td>12 (2.02%)</td>
<td>0.84</td>
<td>85 (1.19%)</td>
<td>115 (1.61%)</td>
<td>0.03</td>
</tr>
<tr>
<td>Upper GI Bleeding Events</td>
<td>10 (1.68%)</td>
<td>7 (1.18%)</td>
<td>0.46</td>
<td>51 (0.71%)</td>
<td>73 (1.02%)</td>
<td>0.05</td>
</tr>
<tr>
<td>Lower GI Bleeding Events</td>
<td>3 (0.51%)</td>
<td>5 (0.84%)</td>
<td>0.48</td>
<td>34 (0.48%)</td>
<td>42 (0.59%)</td>
<td>0.36</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Non-Atrial Fibrillation</th>
<th>N = 76</th>
<th>N = 76</th>
<th>P-value</th>
<th>N = 656</th>
<th>N = 656</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Bleeding Events</td>
<td>2 (2.63%)</td>
<td>2 (2.63%)</td>
<td>1.00</td>
<td>13 (1.98%)</td>
<td>11 (1.68%)</td>
<td>0.68</td>
</tr>
<tr>
<td>Upper GI Bleeding Events</td>
<td>1 (1.32%)</td>
<td>1 (1.32%)</td>
<td>1.00</td>
<td>9 (1.37%)</td>
<td>8 (1.22%)</td>
<td>0.81</td>
</tr>
<tr>
<td>Lower GI Bleeding Events</td>
<td>1 (1.32%)</td>
<td>1 (1.32%)</td>
<td>1.00</td>
<td>4 (0.61%)</td>
<td>3 (0.46%)</td>
<td>0.70</td>
</tr>
</tbody>
</table>

3. How did the analysis deal with death? This is never discussed in the description of analysis, but if death occurs at different rates in patients receiving a novel oral anticoagulant versus warfarin (as was found in the Re-LY trial), the likelihood of observing a GI bleeding event will be impacted. One option is
to censor patients at the time of death. Of course, this can be problematic because ideally, with Cox regression models, the reason for censoring should be unrelated to the patient’s health status.

Ascertainment of death in this dataset is limited to the category of “disenrollment” from the plan. Reasons for disenrollment include: change in health plan provider, loss of plan coverage or death. In all models, we censored patients’ data from the analysis on the date of disenrollment. While it would be ideal to consider death separately from disenrollment from health plan, privately insured claims typically do not allow for the ability to link to death data, and thus would be a limitation of any studies based on privately insured claims.

The main advantage of looking at the population studied here is that it is broader and includes a younger group compared to the studies based on Medicare claims where death information is available. However, we did test to see if there were differences in disenrollment rates between the two groups considering that a subset of that group might have died. We did not observe any differences in disenrollment rates between the groups compared in any of the analyses.

4. If I understand correctly, use of the index anticoagulant was considered continuous until one of the following events occurred: a gap of 30 days, a prescription was provided for another anticoagulant, disenrollment from the healthcare plan, or the absence of a new prescription 45 days after the last medication prescription fill. It is never stated clearly in the text, but I believe these events were treated as censoring events in the analysis. If this is correct please clarify in the text. If not correct, describe how these events were handled in the analysis.

We agree the wording of our text was confusing. We have modified our description of exposures and primary outcome for clarity in the methods section (page 5) to the following:

“We considered patients as being continuously exposed from index date (t0) for the duration of their prescription until a gap of 30 days occurred or until occurrence of GIB, disenrollment from the healthcare plan or treatment termination as defined by the absence of a new prescription by the end of the 45 day period from the last identified index medication fill”.

5. It would be very useful to also include a sensitivity analysis based on ‘intent to treat’ analysis, in which patients are assumed to continue the index medication until end of follow-up or GI bleed. Given that several publications from the RE-LY and ROCKET-AF data used ‘intent to treat’, this facilitates direct comparison to published results of clinical trials

See Table A3 in the appendix and below which demonstrates similar Hazard Ratios when an intention treat analysis is done. These analyses have greater effects on the non-AF group as this group would be typically
anticoagulated for a limited time period (e.g. for VTE prophylaxis), but the intention to treat assumes a much longer follow-up.

‘Intent to Treat’ Events and Adjusted Hazards of GI Bleeding Events (Censoring on End of Enrollment)

<table>
<thead>
<tr>
<th>Event Type</th>
<th>Dabigatran vs Warfarin</th>
<th>Rivaroxaban vs Warfarin</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atrial Fibrillation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total Bleeding Events</td>
<td>1.21 (0.60, 0.94)</td>
<td>2.45 (0.72, 1.23)</td>
</tr>
<tr>
<td>Upper GI Bleeding Events</td>
<td>0.71 (0.55, 0.94)</td>
<td>1.57 (0.73, 1.44)</td>
</tr>
<tr>
<td>Lower GI Bleeding Events</td>
<td>0.84 (0.58, 1.22)</td>
<td>0.82 (0.53, 1.27)</td>
</tr>
<tr>
<td>Non-Atrial Fibrillation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total Bleeding Events</td>
<td>1.02 (0.51, 2.04)</td>
<td>0.74 (0.47, 0.86)</td>
</tr>
<tr>
<td>Upper GI Bleeding Events</td>
<td>1.02 (0.44, 2.35)</td>
<td>0.48 (0.46, 1.00)</td>
</tr>
<tr>
<td>Lower GI Bleeding Events</td>
<td>1.01 (0.29, 3.50)</td>
<td>0.26 (0.34, 0.97)</td>
</tr>
</tbody>
</table>

*Hazard Ratio Adjusted for those covariates significant at the <0.05 level.

6. **It would be nice to see a more complete description of these patients (i.e., a description of other indications for which dabigatran, rivaroxaban, and warfarin are approved and/or used).**

As requested, we have examined the number of non-AF patients who had other approved indications for anticoagulant use including pulmonary embolism, deep venous thromboembolism and total hip and knee replacement. As we were identifying the indications for use, we identified a small group of patients in the non-AF group that had a diagnosis of AF within the first 7-days after the medication fill (most during 1-3 days). We had initially limited the identification of diagnoses up to the date of the medication fill. To be sure that we did not miss-categorize any patients, we went out up to 6-months after the first medication fill date to look at indications/diagnoses and were only able to identify a few additional individuals that would be in the atrial fibrillation cohort in the first 7 days. We re-ran all the analyses and have updated all the tables and figures. None of the results changed substantively.

We explored most common conditions and procedures for which anticoagulants were prescribed in the non-AF cohorts. Identification of indication for anticoagulant use in non-AF patients are not consistently easy to identify. The most common off-label uses of dabigatran were for pulmonary embolism (PE), deep venous thromboembolism (DVT), total knee and hip replacement. Similarly, majority of the non-AF use among rivaroxaban users could be attributed to post knee and hip replacement, DVT, and PE.
7. There is a lack of clarity about how medication cessation was treated in the analysis. Were patients who switched anticoagulants excluded altogether (and if so, why?), or was termination of the index anticoagulant treated as a censoring event (whether or not a different anticoagulant was prescribed)?

We agree, the wording was confusing and we have edited the “Exposures and Primary Outcome” section (see previous response to Reviewer 2, comment 4). No, we did not exclude all switchers, we excluded only those that switched within the 45 day period from their last identified index medication fill. If patients switched outside of the 45 day period we would have censored them at the 45 day period.

8. The authors should demonstrate the success of their matching algorithm using two graphs of standardized differences before and after matching (one graph for Afib patients taking rivaroxaban, dabigatran, or warfarin, and the other for non-Afib patients).

This has been done, as requested. Please see Figures A1-A4 in the appendix.

9. I would prefer to see more detail provided regarding results by age. Rather than presenting overall results for all ages in the tables (e.g., Table 3), I would like to see the tables show bleeding rates and hazard ratios for patients <75, and patients >=75.

This has been done as requested. Additional detail regarding the bleeding rates, stratified by age, can be found in Figures A7 to A10 in the appendix.

10. I would like to see more comparison in the discussion section to previous investigations of differences in GI bleeding rates by age, and how the current study adds to that literature. (for example, Eikelboom et al investigated rates of major bleeding by age using RE-LY data [Eikelboom et al, Circulation 2011]).

Below is a table comparing the rates of GIBs by age. While the measures and sub-groups presented in each of the studies vary, overall results are similar to the ones observed in our study. More importantly, we did not find any studies that have previously published rates of bleeds for those under the age of 65, for rivaroxaban (outside of randomized trials), and for those using anticoagulants for non-AF indications. We have also added the comparisons to these studies in the discussion section.
Editors’ comments:

1. **We would like the authors to clarify if there are other databases that capture similar information, whether there is any overlap between the databases, and why this particular one was chosen.**

Data for this study were obtained from the Optum Labs Data Warehouse (OLDW). OLDW contains longitudinal health information on over 149 million, geographically diverse individuals in the United States. The advantage of using the OLDW data is that it includes both the privately insured and Medicare Advantage enrollees. Thus, our dataset is one of the most comprehensive population datasets available in the United States and permits a more comprehensive comparative effectiveness evaluation of the NOACs in a varied population. While there are other datasets of privately insured individuals in the US available such as Marketscan and IMS Lifelink, they would provide no advantage over this data as they include comparable amounts of covered lives, but often do not have the same rates of older enrollees. In fact, prior to starting this study we looked at sample sizes from two other datasets and they were both smaller than what we have in this study.

2. **Several authors noted that the doses of medications used in this study are not available in Europe. The authors might want to comment on how this affects generalizability.**

Dabigatran is approved for 75mg BID and 150mg BID in the US and 110mg BID and 150mg BID in Europe. Since majority of the use in our population was for 150mg BID, we would not be able to draw any inferences on rates of GIBs for individuals treated with the 110mg BID dose in Europe. We have added this as a limitation to the manuscript.

3. **The propensity scoring seems to have been well done, although there seem to have been some residual differences in the non-atrial fibrillation group.**
Please see our response to Reviewer 1, comment 4 above and the supplemental figures A1-A4 in Appendix.

4. **There are differences between the results of this paper and prior RCTs. We would be interested to see if the results of this paper change after the reviewers’ concerns are taken into account. These include questions about how data were collected, and which individuals were included in the datasets, how exposure was defined.**

Our results do not change after significant re-analysis to accommodate reviewer’s concerns. We have clarified issues regarding assessment of exposure and outcome; re-calculated the outcomes using an intention to treat analysis; provided the model diagnostics supporting the robustness of our propensity score matching algorithm; and, provided additional figures in the appendix demonstrating differences in the outcomes by specific ages.

We hope that these changes and comments satisfy the Editors and reviewers. On behalf of the authors, please allow me to thank you for your consideration of our work.

¹ (www.mini-sentinel.org)