Testing the therapeutic equivalence of novel oral anticoagulants for thromboprophylaxis in orthopedic surgery and for prevention of stroke in atrial fibrillation

Andrea Messori, Valeria Fadda, Dario Maratea, Sabrina Trippoli, and Claudio Marinai

HTA Unit, ESTAV Toscana Centro, Regional Health Service, Firenze, Italy

Abstract. Background: In studying the comparative effectiveness of novel oral anticoagulants (NOACs) in orthopedic surgery and in non-valvular atrial fibrillation, previous meta-analyses have found no proof of difference in head-to-head indirect comparisons between individual agents. However, the question of their therapeutic equivalence remains unanswered. Objectives: The objective of this analysis was to test the equivalence of three NOACs (dabigatran, rivaroxaban, apixaban) in orthopedic surgery and four NOACs (dabigatran, rivaroxaban, apixaban, and edoxaban) in non-valvular atrial fibrillation. Methods: Standard pairwise meta-analysis and network meta-analysis for indirect comparisons were combined with equivalence testing. The endpoint was venous thromboembolism in orthopedic surgery and a composite of stroke or systemic embolism in atrial fibrillation. Comparisons were expressed as risk difference (RD). Margins for equivalence testing were derived from the original trials. Results: Our results indicate that rivaroxaban and apixaban (but not dabigatran) are equivalent for thromboprophylaxis in orthopedic surgery. In atrial fibrillation, all the four NOACs we tested were found to meet the criterion of therapeutic equivalence. Some concern, however, is raised by some findings focused on adverse events of these agents, in which the equivalence was not proven in all analyses. Conclusions: Regardless of clinical implications, our results can be the basis to develop local acquisition tenderings on NOACS. In Italy, a new law has been issued according to which equivalence analyses have become a mandatory prerequisite for local tenderings.

Introduction

In studying the therapeutic evidence of innovative drug treatments, a growing role is attributed to differentiating between results that indicate no significant differences among treatments (“no proof of difference”) and results that demonstrate therapeutic equivalence (“proof of no difference”). It is well known that this latter conclusion is more informative than the former because “proof of no difference” is a conclusive result whereas “no proof of difference” is inconclusive. Several analyses of this type have already examined quite a large number of pharmacological treatments (e.g., anti-TNF agents in ulcerative colitis [1] and Crohn’s disease [2], DPP-4 inhibitors in type 2 diabetes [3], anti-VEGF agents in age-related macular degeneration [4], and new oral anticoagulants (NOACs) in acute coronary syndrome [5]). Demonstrations of therapeutic equivalence have obvious clinical implications. However, these findings also have important practical consequences because hospital acquisition tenderings can be conducted more selectively in cases where equivalence has been demonstrated across different agents of the same pharmacological class [6, 7].

The two main approved indications of NOACs include thromboprophylaxis in orthopedic surgery and prevention of stroke in non-valvular atrial fibrillation. A few preliminary studies on NOACs have already examined some issues related to the presence or absence of equivalence [5, 8, 10]. One of these [8] has simply emphasized the need to conduct these analyses; the other two have been focused on a controversial indication of NOACs (i.e., acute coronary syndrome [5]) and on a specific side effect of these agents (i.e., intracranial hemorrhage [10]). No equivalence analysis is available on the two main indications of NOACs.

The present study was conducted to investigate the therapeutic equivalence of NOACs...
Messori, Fadda, Maratea, et al.

for thromboprophylaxis in orthopedic surgery
and for prevention of stroke in non-valvular
atrial fibrillation. The clinical material was
obtained from the published literature. Tests
of equivalence were carried out according to
standard evidence-based methods.

Methods

Study design

The present study was carried out as an
evidence-based analysis aimed at assessing:
[a] the therapeutic equivalence of dabigatran,
rivaroxaban, and apixaban when employed
for thromboprophylaxis in orthopedic sur-
gery; [b] the therapeutic equivalence of dabi-
gatran, rivaroxaban, apixaban, and edoxaban
when employed for stroke prevention in non-
valvular atrial fibrillation. Testing equiva-
ience according to evidence-based methods
is particularly straightforward when a “com-
bined approach” is adopted that integrates
information on margins with standard Forest
plots [10, 11, 12]. In particular, margins are
assumed to represent a threshold between
clinically relevant incremental benefits and
irrelevant ones, and can be retrieved from
statistical power calculations of the original
trials [13, 14, 15]. In this framework, com-
parisons of effectiveness between individ-
ual agents can be designed and carried out
according to standard meta-analysis tech-
niques.

Literature search

The aim of our study was to carry out
equivalence tests based on published mate-
rial and not to carry out an original literature
search of primary studies. In addition, we
knew that a large number of meta-analyses
had previously been focused on these two in-
dications of NOACs (e.g., 13 meta-analyses
on stroke prevention in atrial fibrillation [8]
and 20 meta-analyses on thromboprophyl-
axis in orthopedic surgery [9]). For these
reasons, we did not look for original clinical
studies, but we examined only the clinical
material already reported in the (numerous)
meta-analyses published thus far. Accord-
ingly, our literature search was simply aimed
at examining reviews and/or meta-analyses
already focused on these topics and at select-
ing the best paper(s) suitable for our pur-
poses.

In more detail, our literature search (based
on the PubMed version of MEDLINE) cov-
ered the last 5 years; only reviews, systemat-
ic reviews, and meta-analyses (according to
PubMed definitions of these terms and its re-
spective “filters”) were eligible for being the
source of our clinical material. The terms for
our PubMed searches were intentionally kept
very generic (i.e., “dabigatran OR rivaroxa-
ban OR apixaban OR edoxaban”) because
the presence of filters restricted the eligible
articles to a number of citations suitable for
being examined individually.

The “best” paper(s) for the purposes of
our analyses were identified through a sim-
plicated procedure that we had already ad-
opted in a previous paper [7]. Firstly, we
identified all reviews/meta-analyses that
were considered eligible for our purposes
(namely: a first subset of reviews suitable for
the indication of thromboprophylaxis and a
second subset of reviews suitable for atrial
fibrillation), and we retrieved their full text.
Finally, to identify the best article(s) from
these subsets of select publications, we ad-
opted an evaluation form in which, for each
article, four items were rated on a 0 to 5 scale
((a) number of clinical trials suitable for our
analysis; (b) impact factor of the journal in
which the article was published; (c) degree
of literature update; (d) description of clini-
cal trials not already reported in articles as-
signed a better rating for items (a) through
(c) than the present one). Four authors (AM,
VF, DM, ST) were involved in this process
of literature selection and evaluation. This
procedure, which selected the best paper(s)
for our analysis, was separately applied for
thromboprophylaxis in surgical patients and
for stroke prevention in non-valvular atrial
fibrillation.

Equivalence testing based on
standard Forest plots
incorporating margins from
randomized trials

Previous studies have shown that Forest
plots can easily be integrated with equiva-
lence derived from the statistical power section of randomized trials. Although there has been some debate on the role of these margins, a widely accepted concept is that these parameters should mainly be intended to represent a threshold separating clinically relevant differences from irrelevant ones [11, 12, 13, 14, 15, 16]; in this context, no difference should exist, whether margins were intended for non-inferiority comparisons or for equivalence testings. The α-level adopted in equivalence testing is commonly set at either 5% or 2.5%; in our analyses, we chose the more conservative value of α 2.5%.

**Analysis on thromboprophylaxis in orthopedic surgery**

The following agents were assessed: rivaroxaban 10 mg/d; apixaban 5 mg/d; dabigatran 220 mg/d; dabigatran 150 mg/d. The primary endpoint for our analysis was a composite of symptomatic or asymptomatic deep vein thrombosis, non-fatal pulmonary embolism, and/or all-cause mortality. Our analysis adopted risk difference (RD) as outcome measure. Consequently, we planned to recompute all outcomes expressed differently (e.g., as relative risk) and to re-determine all meta-analytical results as well. Meta-analytical values of RD were estimated by random-effect model according to the Open Meta-Analyst Software (Brown University, city?, USA). Information on margins (expressed as RDs) was obtained from the trials included in the “best” meta-analysis identified from our literature search.

**Presentation of the results**

We designed a two-fold presentation of our results, which included a preliminary analysis of the effectiveness data based on direct comparisons and a final analysis based on indirect head-to-head comparisons. Direct comparisons were characterized by the presence of a common reference treatment, which was low-molecular weight heparins in the analysis on thromboprophylaxis and adjusted-dose warfarin in the analysis on stroke prevention in atrial fibrillation. The values of RD for the indirect comparisons were estimated according to standard network meta-analysis techniques [17]. All outcome measures were accompanied by their respective 95% confidence intervals (CIs).

Our preliminary analysis (which of course differed between the two clinical indications) mainly served the purpose of providing an overall picture of the therapeutic homogeneity that can be expected in terms of effectiveness within this class of different agents. The final analysis, based on indirect comparisons, instead explored, case by case, the degree of therapeutic similarity (i.e., equivalence) that can be attributed to each of the combinations of pair-wise comparisons between individual NOACs.

**Results**

**Thromboprophylaxis in orthopedic surgery**

Our initial PubMed query (run on 1 May, 2014) extracted 857 eligible articles. Among these articles, 32 were potentially suitable for providing the material for our analysis (see Reference [9] for individual citations). We examined the full text of these 32 articles, and then we selected the meta-analysis by Maratea et al. [18] as our source of effectiveness (8 trials comparing NOACs vs. low-molecular weight heparins). Since the selected study employed relative risk as outcome measure, we re-analyzed the results to express them as RDs. Margins were obtained from the trial by Eriksson et al. [19], where their value was set at RD = ± 7.7%.

The results of our preliminary analysis on effectiveness are presented in Figure 1, Panel
The equivalence interval was centered around the pooled RD across all studies (i.e., –3.1%) and therefore ranged from –10.8% to +4.6%. First of all, this Forest plot showed that, in comparison with the average effectiveness of all NOACs, both rivaroxaban and apixaban were numerically more effective, whereas both dosages of dabigatran were numerically less effective. Equivalence testings from direct comparison showed that rivaroxaban and dabigatran (but not apixaban) met the equivalence criterion (even though both apixaban and dabigatran 150 mg/d were very close to “touching” the equivalence margins). It should be noted that, in the case of apixaban 5 mg/d, the lower extreme of the 95% CI (at RD = –12.4%) clearly crossed the lower margin of the equivalence interval (at –10.8%). This is probably because the two studies available for apixaban were dishomogeneous in their results, and so the meta-analytic RD for this regimen showed a wide 95% CI (pooled RD = –5.7%; 95% CI: –12.4% to –1%; random-effect model).

More interestingly, our final analysis based on indirect comparison (Figure 1, Panel (b)) provided a much clearer message. In fact, the criterion of equivalence was met only for two indirect comparisons (apixaban vs. rivaroxaban and dabigatran 150 mg/d vs. dabigatran 220 mg/d), whereas the other 4 indirect comparisons were unable to demonstrate equivalence. Of the two comparisons showing equivalence, the one regarding the two dosages of dabigatran has quite little interest, whereas the demonstrated equivalence between apixaban and rivaroxaban shows that the two agents that are characterized by the highest numerical effectiveness in the class of NOACs have also an important homogeneity in their degree of expected effectiveness.

Prevention of stroke in non-valvular atrial fibrillation

Among the 857 eligible articles extracted through our PubMed search, 37 were potentially suitable for providing the material for this analysis. We examined the full text of these 37 articles, and then we selected the meta-analysis by Ruff et al. [20] as our source of effectiveness (4 trials comparing
Equivalence of new oral anticoagulants

NOACs vs. warfarin. The results of Ruff et al. [20] were re-expressed as risk difference (RD) instead of relative risk (RR). As described by Fadda [21], the margin was set at the same value (RD = 2.5%) adopted in pivotal trials comparing NOACs vs. warfarin.

Our results (Figure 2) included, on the one hand, the preliminary analysis based on direct comparisons (Panel (a)) and, on the other hand, the final analysis based on indirect head-to-head comparisons (Panel (b)). In this case, the interpretation of these results was very straightforward. In fact, both analyses indicated that these NOACs are equivalent in their degree of expected effectiveness (under our pre-specified assumptions about margins and α-level).

Discussion

According to our results (in particular, the two analyses presented in Figures 1 and 2), rivaroxaban and apixaban proved to be equivalent in their effectiveness for thromboprophylaxis in orthopedic surgery, while dabigatran, rivaroxaban, apixaban, and edoxaban proved to be equivalent in their effectiveness for stroke prevention in atrial fibrillation. These demonstrations of equivalence suffer from the same limitations already pointed out in previous studies [1, 2, 3, 4, 5, 6, 7]. For example, margins continue to be a point of controversy in the overall area of equivalence studies [11, 12, 13, 14, 15, 16]; furthermore, in the case of our analysis on NOACs, one might not agree from a clinical perspective on the magnitude of the margins (± 7.7% and ± 2.5% for the endpoints of thromboprophylaxis and stroke prevention, respectively).

In Italy, there has been a renewed interest in defining the criteria for therapeutic equivalence because, at the end of 2012, a national regulation was issued (“decreto Balduzzi” [7, 8]) concerning the acquisition tenderings run by our NHS. According to this regulation, when these tenderings are aimed at drugs that belong to the same pharmacological class, a preventive authorization must be obtained from our National Medicines Agency to certify that the agents under examination are therapeutically equivalent. Otherwise, these tenderings...
can no longer be launched. Hence, defining equivalence is, in Italy, not only of scientific interest but also has important implications in terms of regulatory decisions and pharmaceutical governance.

In the field of NOACs, our study indicates that the magnitude of the clinical improvement cannot represent the main criterion for selecting a specific agent in a given patient, since the expected improvement is essentially the same across the different agents (at least those demonstrated to be equivalent). Other criteria should, therefore, prevail, including the dosing schedule, the profile of adverse effects, and, last but not least, the cost.

As pointed out by Kirchhof et al. [22], NOACs are going to replace a substantial part of warfarin prescriptions in patients with atrial fibrillation, but the extent to which this replacement takes places varies markedly across different countries. Many countries have shown a high degree of switch from warfarin to NOACs, and this raises important economic issues. In comparison with warfarin, NOACs tend to have a favorable cost-effectiveness profile [23], but this pharmacoeconomic advantage cannot be the only basis upon which this decision should be made.

In the prevention of stroke and systemic embolism in patients with non-valvular atrial fibrillation, the efficacy of individual NOACs is influenced by the occurrence of bleeding. Very recently, Liew et al. [24] have published an updated review in this area and have provided meta-analytic estimates of the bleeding risk with individual NOACs. Given that the occurrence of fatal bleeding is – of course – the strongest negative end-point in these patients, we carried out a re-analysis of the findings published by Liew et al. [24] regarding this specific end-point (see Appendix B). Our results confirm that this negative outcome occurs with the same frequency across dabigatran, rivaroxaban, apixaban, and edoxaban (see Figure 3 in Appendix B); in particular, we have been able to provide the proof of no difference among these four agents (besides the already known demonstration of no proof of difference). Therefore, fatal bleedings negatively affect the efficacy of NOACs, but this occurs to an equivalent extent across these four agents.

In general, the safety of NOACs is, of course, another important factor for defining their respective role in comparative terms. The evidence on this point, however, is more difficult to interpret because of the diversity of the safety end-points and their relatively low frequency of occurrence [10, 25, 26, 27, 28, 29]. It should be noted that a few reports on NOACs safety [10, 20, 27, 28, 29] seem to indicate that some differences might exist.
In conclusion, although our analyses have been entirely based on the same clinical material already published by Maratea et al. [18] and by Ruff et al. [20], our results convey original information to better interpret the effectiveness of NOACs in terms of equivalence and possibly to better design the acquisition tenders run by our NHS.

**Conflict of interest**

None declared.

**References**


*Authors, please clarify:
Should the Appendix be published?
If yes, print and online publication? Or only online?

Appendix A

The values of RD shown in Figures 1 and 2 are as follows:

**Figure 1, direct comparisons (top to bottom) in Panel (a):** [1] rivaroxaban 10 mg/d vs. LMWHs (3 trials), RD = –6.2% (95% CI: –10.5% to –2%); [2] apixaban 5 mg/d vs. LMWHs (2 trials), RD = –5.7% (95% CI: –12.4% to +1%); [3] dabigatran 220 mg/d vs. LMWHs (3 trials), RD = –0.9% (95% CI: –2.5% to +0.8%); [4] dabigatran 150 mg/d vs. LMWHs (2 trials), RD = +2.0% (95% CI: –0.3% to +4.3%).

**Figure 1, indirect comparisons (top to bottom) in Panel (b):** [1] R (10 mg/d) vs. D (220 mg/d), RD = –5.3% (95% CI: –9.9% to –0.7%); [2] A (5 mg/d) vs. D (220 mg/d), RD = –4.8% (95% CI: –11.7% to 2.1%); [3] D (150 mg/d) vs. D (220 mg/d), RD = +2.9% (95% CI: +0.07% to +5.7%); [4] A (5 mg/d) vs. R (10 mg/d), RD = +0.5% (95% CI: –6.6% to +7.6%); [5] R (10 mg/d) vs. D (150 mg/d), RD = –8.2% (95% CI: –13% to –3.4%); [6] A (5 mg/d) vs. D (150 mg/d), RD = –7.7% (95% CI: –13.8% to –1.6%).

**Figure 2, direct comparisons (top to bottom) in Panel (a):** [1] D vs. W, RD = –1.10% (95% CI: –1.68% to –0.52%); [2] R vs. W, RD = –0.52% (95% CI: –1.17% to 0.13%); [3] A vs. W, RD = –0.59% (95% CI: –1.06% to –0.13%); [4] E vs. W, RD = –0.58% (95% CI: –1.27% to 0.1%).

**Figure 2, indirect comparisons (top to bottom) in Panel (b):** [1] R (10 mg/d) vs. D (220 mg/d), RD = 0.58% (95% CI: –0.7% to 1.86%); [2] A vs. D, RD = 0.51% (95% CI: –0.68% to 1.70%); [3] E vs. D, RD = 0.52% (95% CI: –0.78% to 1.82%); [4] A vs. R, RD = –0.07% (95% CI: –0.95% to 0.81%); [5] E vs. R, RD = –0.06% (95% CI: –1% to 0.88%); [6] E vs. A, RD = –0.01% (95% CI: –0.84% to 0.82%).

Appendix B

In studying the occurrence of fatal bleeding across the four NOACs, Liew et al. [24] have reported in detail the event frequencies observed in the pivotal trials that compared individual NOACs versus warfarin in patients with non-valvular atrial fibrillation (see Figure 1c of Reference 24). From these primary data, we have computed the values of RD (with 95% CI) by using a random-effect meta-analysis model (OMA Software, Open Meta-Analyst, version 4.16.12, Tufts University, url http://tuftscaes.org/open_meta/).

These values of RD (direct comparisons) are as follows:

[1] dabigatran vs. warfarin, RD = –0.23% (95% CI: –0.46% to +0.01%);
[2] rivaroxaban vs. warfarin, RD = –0.39% (95% CI: –0.64% to –0.14%);
[3] apixaban vs. warfarin, RD = –0.23% (95% CI: –0.44% to –0.03%);
[4] edoxaban vs. warfarin, RD = –0.46% (95% CI: –0.70% to –0.23%).
Then, the above values of RD for direct comparisons were incorporated into a network meta-analysis (Bucher’s method) that estimated the following values of RD for the various indirect comparisons:

[1] rivaroxaban vs. dabigatran, RD = –0.16% (95% CI: –0.50% to +0.18%);
[2] dabigatran vs. apixaban, RD = 0.0% (95% CI: –0.31% to 0.31%);
[3] rivaroxaban vs. apixaban, RD = –0.16% (95% CI: –0.48% to +0.16%);
[4] edoxaban vs. rivaroxaban, RD = –0.07% (95% CI: –0.41% to +0.27%);
[5] edoxaban vs. apixaban, RD = –0.23% (95% CI: –0.54% to +0.08%);
[6] edoxaban vs. dabigatran, RD = –0.23% (95% CI: –0.56% to +0.10%).

Figure 3 shows the Forest plot with the RD values for direct comparisons (Panel (a)) and for indirect comparisons (Panel (b)). These results indicate that, in regards to indirect comparisons, the post-hoc margin of equivalence is at ± 0.56%.