

RESEARCH METHODOLOGY SERIES

The challenges of determining noninferiority margins: a case study of noninferiority randomized controlled trials of novel oral anticoagulants

Grace Wangge MD MSc, Kit C.B. Roes PhD, Anthonius de Boer MD PhD, Arno W. Hoes MD PhD, Mirjam J. Knol PhD

A randomized controlled trial (RCT) can have either a superiority design or a noninferiority design. A superiority design aims to show that a new drug is better than placebo or an active comparator, whereas a noninferiority design aims to show that a new drug is not worse than its comparator, which is typically an active drug. Noninferiority trials can be used when a new drug is anticipated to have an efficacy profile similar to its comparator and could offer advantages over the existing drug, such as a novel method of administration.

We have seen a large increase in publications of noninferiority trials since 2000. A search in PubMed for the term “non-inferior*” in titles and abstracts found 9 publications in 2000 and 260 publications in 2010. These results show the growing importance for readers and clinicians of understanding the concept of this sort of trial.

The crucial but difficult step in designing such a trial is prespecifying a noninferiority margin: a threshold below which it can be established that the new drug is not worse than its comparator. This margin should be chosen such that the new drug can be considered to be effective relative to placebo (even when a placebo group is not included) and needs to account for the uncertainty in the effect size of the active control versus placebo. Previously, we found that only 106 of 232 noninferiority trials (46%) reported the method they used to determine the noninferiority margin, and these methods varied considerably.¹ In 22% of the trials, the margin was determined based solely on the investigator’s own assumption (without providing a rationale for the choice); in 8.6% of the trials, the margin was stated as an acceptable clinical difference according to the literature.² These observations are worrisome, as the choice of the noninferiority margin determines the conclusion of the trial and, thus, clinical decision-making.

Here, we explain one method for determining a noninferiority margin, as outlined in the draft US Food and Drug Administration (FDA) guideline on noninferiority trials.³ In addition, we present a case study on the noninferiority margins used in trials of novel anticoagulant drugs. The case study shows substantial variability in the noninferiority margins applied in the selected trials.

Determining a noninferiority margin

Most of the guidelines on noninferiority trials⁴⁻⁶ state that a margin should account for both clinical and statistical considerations. However, details on how such a margin should be determined are not clearly specified, with the exception of the recently drafted guideline on noninferiority trials issued by the FDA.³ The guideline was composed based on previous guidelines⁴⁻⁶ and methodological publications on noninferiority trials⁷⁻¹⁰ published since the 1980s. The guideline is only one example of

Competing interests:

Grace Wangge, Anthonius de Boer and Mirjam Knol have received grant funding from Top Institute Pharma. No other competing interests were declared.

This article has been peer reviewed.

Correspondence to:

Anthonius de Boer,
a.deboer@uu.nl

CMAJ 2012. DOI:10.1503/
cmaj.120142

KEY POINTS

- The aim of a noninferiority trial is to show that a new drug is not worse than its comparator.
- How a noninferiority margin is chosen for a trial is often not explained; methods can be highly variable, resulting in inconsistent conclusions of noninferiority.
- A noninferiority margin should be based on both statistical and clinical considerations.
- The constancy assumption — that the effect of the active comparator versus placebo is present in the current trial — should be discussed.

This article is part of an occasional series that examines controversial aspects of research methods and reporting.

determining a noninferiority margin, and it reflects regulatory interest; thus, its focus is on showing indirect efficacy of the test drug compared with placebo.

The guideline recommends the fixed-margin method, or 95%–95% method, which is considered the most straightforward and readily understood approach. The method starts by identifying M1 and M2. M1 is the effect of the active control compared with placebo, which is assumed to be present in the noninferiority trial. M1 is chosen as a conservative estimate (smallest effect size possible) of the effect of the active comparator, which is the upper bound of the 95% confidence interval (CI) of the pooled effect size, rather than the point estimate. M2 reflects the clinical judgement about how much of M1 should be preserved and represents the largest clinically acceptable difference (degree of inferiority) of the test drug compared with the active control. For example, if it is necessary that a test drug preserve 75% of a mortality effect, M2 would be 25% of M1, the loss of effect that must be ruled out. Determining M2 assures that the test drug will be superior to placebo.

Determining M1, as the first step in defining a noninferiority margin, can be based on one or more placebo-controlled trials of the active comparator that have a design similar to the current noninferiority trial. A meta-analysis of several placebo-controlled trials is preferable, because it will result in a pooled, more precise effect estimate of the active comparator.

The second step is to calculate M2 from M1 by choosing a certain amount of the effect to be preserved. The draft FDA guideline implicitly recommends using a preserved-effect of 50% to determine M2. Choosing a higher percentage to be preserved (e.g., 67%, where M2 is 33% of M1) results in a stricter or more conservative noninferiority margin, meaning it is more difficult to conclude noninferiority. The formula to calculate M2 for a risk difference (RD) is:

$$(1 - \text{preserved effects}) \times -M1$$

For the relative risk (RR), and other ratio measures, the guideline discusses 3 methods for calculating M2. The preferred method calculates the margin using the natural logarithm:

$$e^{\ln(1/M1) \times (1 - \text{preserved effects})}$$

$$\text{or } (1/M1)^{(1 - \text{preserved effects})}$$

The results of the noninferiority trial are compared with the prespecified noninferiority margin (M2) as follows: if the upper bound of the 95% CI for the effect estimate is smaller than the

noninferiority margin, noninferiority is concluded. For example, if a noninferiority trial shows that the RR of the new drug compared with the active comparator is 0.90 (95% CI 0.68 to 1.20), and the noninferiority margin is 1.25, it can be concluded that the new drug is noninferior to the active comparator.

Determining M2 is also related to how much of the treatment effect is judged necessary to be preserved, a consideration that may reflect the seriousness of the outcome, the benefit of the active comparator and the relative safety profiles of the test drug and the comparator. This factor has considerable practical implications. For example, in large cardiovascular studies, it is unusual to seek retention of more than 50% of the effect of the control drug, even if this might be clinically reasonable, because doing so will usually cause the size of the study to become infeasible.

Case study

Noninferiority trials

Recently, new classes of anticoagulant medications, direct thrombin inhibitors (DTIs) and direct inhibitors of factor Xa (DXAIs), were claimed to be as effective as conventional therapies, such as heparin or low-molecular weight heparins (LMWHs), but with a more convenient route of administration and no requirement for monitoring after discharge from hospital. DTIs and DXAIs were first registered for the prevention of venous thromboembolism in patients undergoing elective hip- or knee-replacement surgery. Many of the trials were noninferiority trials. We found 12 such trials in PubMed and the Cochrane central register of controlled trials in May 2012 (Appendices 1 and 2, available at www.cmaj.ca/lookup/suppl/doi:10.1503/cmaj.120142/-/DC1).

All of the trials used enoxaparin as the active comparator (40 mg once daily, or 30 mg twice daily). Most of the trials used RD to define the noninferiority margin, which ranged from 2.0% to 9.2%. Three trials used RR, setting the noninferiority margin at 1.25. Only 4 of the 12 trials stated how they determined the noninferiority margin. One trial stated that an independent expert committee determined the margin, which was the same noninferiority margin that had been used in a previous active-controlled trial of enoxaparin versus tinzaparin.¹¹ Three trials used 67% preserved-effect of the (pooled) effect of 1 or 3 placebo-controlled trials.^{12–14}

Reference noninferiority margin

We determined a reference noninferiority margin using the fixed-margin method recommended in the draft guideline.

First, we performed a meta-analysis of placebo-controlled trials with enoxaparin for prophylaxis of venous thromboembolism after elective hip- or knee-replacement surgery. We found 6 trials in PubMed and the Cochrane register in May 2012 (Appendix 3, available at www.cmaj.ca/lookup/suppl/doi:10.1503/cmaj.120142/-/DC1). The placebo-controlled trials were quite similar to the noninferiority trials with respect to enoxaparin's dosage and duration, patients' ages and sex distribution. However, death was not included as an outcome in the placebo-controlled trials, whereas most noninferiority trials included death by all causes in their composite outcome. Because the noninferiority trials in our case study started recruiting patients after 2000, we only included the 4 placebo-controlled trials¹⁵⁻¹⁸ published before 2000 in the meta-analysis. We calculated the pooled RD and RR with 95% CIs using a fixed- and random-effects model (Appendix 4, available

at www.cmaj.ca/lookup/suppl/doi:10.1503/cmaj.120142/-/DC1). We considered the upper bound of the pooled CI to be M1. The fixed- and random-effects model for RD resulted in different CIs, and therefore resulted in different values for M1.

Second, we calculated values for M2 using a 50% and 67% preserved-effect of M1 (Appendix 4). For example, calculating M2 with 50% preserved-effect for RD based on the fixed-effects model resulted in the following calculation: $(1 - 0.5) \times -(-0.26) = 0.130$.

In addition, we used a 67% preserved-effect because 3 of the noninferiority trials included in our meta-analysis used this value.

Comparison between the reference and published noninferiority margins

We plotted the point estimates and 95% CIs of the noninferiority trials, their noninferiority margins and the reference noninferiority mar-

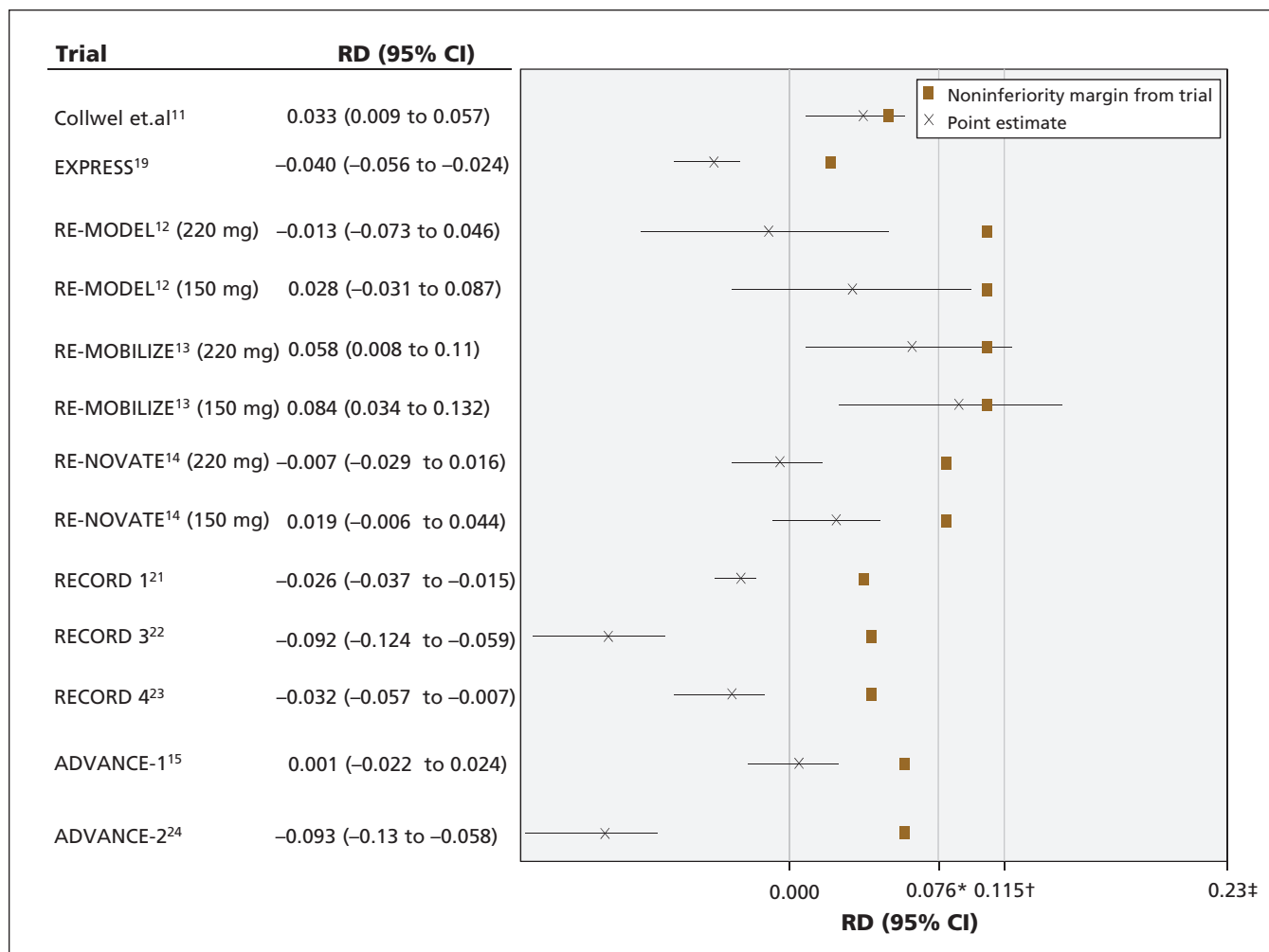


Figure 1: Results of noninferiority trials, noninferiority margins of the trials, and preserved-effects reference noninferiority margins using risk difference. The confidence intervals resemble the effects of test drug – active comparator (negative treatment effect is desirable). Point-of-no-difference between test drug and enoxaparin is 0.000. CI = confidence interval, M1 = effect of active control versus placebo, M2 = largest clinically acceptable difference between the test drug and the active control, RD = risk difference. *M2 for 67% preserved effect. †M2 for 50% preserved effect. ‡M1.

gins to assess whether the conclusion of the trials would have been different had the reference noninferiority margin been used (Figures 1 and 2). We did not include one of the trials in our figures,¹⁹ because it was stopped early for safety concerns and therefore lacked data on efficacy.

Figure 1 shows that the noninferiority margins for the RDs from the trials were stricter than the 50% preserved-effects reference noninferiority margin (0.02–0.092 v. 0.115); thus, the conclusion of noninferiority in these trials does not change when using the reference noninferiority margin, with the exception of the trial by Colwell and colleagues.¹¹ The noninferiority margins in the RE-MODEL,¹² RE-MOBILIZE¹³ and RE-NOVATE¹⁴ trials were larger (i.e., less conservative) than the 67% preserved-effect reference noninferiority margin (0.092 and 0.077 v. 0.076). In the RE-MODEL trial,¹² dabigatran (150 mg) would not have been found noninferior to enoxaparin if the 67% preserved-effect reference noninferiority margin had been used. Moreover, if the most conservative noninferiority margin from the EXPRESS^{w2} trial was used (0.02), the RE-MODEL¹² and RE-NOVATE¹⁴ trials would not have concluded noninferiority to enoxaparin.

Figure 2 shows that the noninferiority margins in these trials were smaller (i.e., more conservative) than the 50% and 67% preserved-effect references (1.25 v. 1.46 and 1.28). In the ADVANCE 1 trial,¹⁵ noninferiority of apixaban was not concluded by the authors owing to inconsistency between results for the RD and RR. If the 50% preserved-effect reference noninferiority margin was used for both the RD and

RR, apixaban would have been found noninferior to enoxaparin.

Lessons learned

We found substantial variation in noninferiority margins used in noninferiority trials of oral anticoagulant medications compared with enoxaparin for prophylaxis of venous thromboembolism after orthopedic surgery. Such variation could lead to inconsistent conclusions on noninferiority and the efficacy of the studied drugs compared with placebo. Furthermore, when determining a noninferiority margin using the method from the draft FDA guideline, we noted some issues that are not explicitly described in the guidelines, including the amount of effect that should be preserved, how similar the characteristics of the placebo-controlled trials and noninferiority trials need to be, and whether the RD or RR should or could be used to calculate the margin.

The different values for preserved effect used in the trials could be the reason for this variability in noninferiority margins. The draft FDA guideline suggests using a preserved-effects value of 50% to assure that the active control is better than placebo. However, there may be other specific considerations related to the test drug or the trial itself for choosing a higher preserved-effect value. These considerations include the seriousness of the outcomes (e.g., a stricter margin for irreversible outcomes, such as death), the treatment effect of the active comparator versus placebo (e.g., using larger mar-

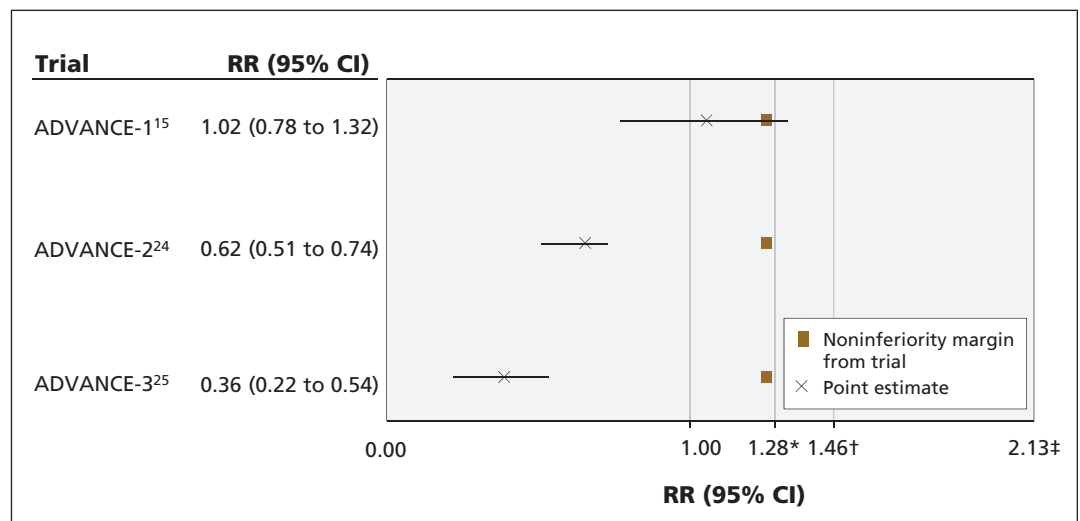


Figure 2: Results of noninferiority trials, noninferiority margins of the trials, and preserved-effects reference noninferiority margins using relative risk. The confidence intervals resemble the effects of test drug – active comparator (negative treatment effect is desirable). Point-of-no-difference between test drug and enoxaparin is 1.000. CI = confidence interval, M1 = effect of active control versus placebo, M2 = largest clinically acceptable difference between the test drug and the active control, RR = relative risk. *M2 for 67% preserved effect. †M2 for 50% preserved effect. ‡M1.

gins for larger effects), adverse effects of the test drug (e.g., using a larger margin if the test drug has fewer serious adverse effects than available therapies), the availability of other drugs (e.g., using a stricter margin if other efficacious and safe drugs are available) and overall cost and benefit–risk assessment.^{3,16} Although all of the noninferiority trials in our case study were similar in terms of these considerations, substantial variation in the noninferiority margin existed between the trials, suggesting that the different clinical judgments and perceptions of the investigators played a role.

Furthermore, for valid inference of a noninferiority trial, one must assume that the treatment effect between the active comparator and the placebo remains accurate during the current trial. This is known as the “constancy assumption” and cannot be assessed with total objectivity. However, it can be supported by a proper meta-analysis and by showing similarity between the current trial and the trials used for setting the margin in terms of the characteristics of patients, the intensity of treatment and the definition of outcomes.¹⁷ In our case study, although the placebo-controlled trials were quite similar to the noninferiority trials, they did differ in their definition of outcomes. The question, therefore, remains as to whether the noninferiority trials and placebo-controlled trials were similar enough. This is another subjective judgement inherent to noninferiority trials. In addition to the similarity in the characteristics of trials, the constancy assumption relies on the absence of any influence from several other factors that are not easily verifiable, such as changes in the standards of care. Uncertainty of the validity of the constancy assumption in a noninferiority trial can raise concerns over the conclusion of noninferiority.

Another challenge related to the use of meta-analysis is the risk of publication bias. It is possible that the result of our pooled analysis would have been different if unpublished results of placebo-controlled trials on enoxaparin had been included. However, accessing such data might be difficult. Only recently have pharmaceutical companies been obliged to publish all results of clinical trials done to get market authorization, either in a peer-reviewed publication or on an independent website (e.g., www.clinicaltrials.gov).¹⁸ Such disclosure of data will certainly help improve the quality of future trials.

The draft FDA guideline does not explicitly state whether the noninferiority margin should be based on an absolute measure, such as the RD, or a relative measure, such as the RR. For clinicians, the RD is more relevant to treatment

decisions for individual patients. Furthermore, the RD is particularly useful when considering trade-offs between the benefits and harms of an intervention, which is crucial in noninferiority trials. The RR, however, is less dependent on the baseline risk, less likely to show heterogeneity between trials and is mathematically more convenient. It is worth noting that, in the context of noninferiority trials, the RDs and RRs can yield opposite conclusions regarding noninferiority if the rate of events seen in the active comparator group differs from the assumed rate that was used to define the noninferiority margin. In a superiority trial, this cannot occur.

Substantial variation in noninferiority margins exists among noninferiority trials of anticoagulant medications for prophylaxis of venous thromboembolism after orthopaedic surgery, which could lead to inconsistent conclusions of a drug’s noninferiority to an active comparator and its efficacy compared with placebo. This inconsistency is undesirable both from a clinical and regulatory perspective. Further research is needed to provide clearer guidance on how to deal with certain crucial aspects of determining a noninferiority margin.

References

1. Wangge G, Klungel OH, Roes KC, et al. Interpretation and inference in noninferiority randomized controlled trials in drug research. *Clin Pharmacol Ther* 2010 ;88:420-3.
2. Wangge G, Klungel OH, Roes KCB, et al. Room for improvement in conducting and reporting non-inferiority randomized controlled trials on drugs: a systematic review. *PLoS ONE* 2010;5:e13550.
3. *Guidance for industry non-inferiority clinical trials*. Silver Spring (MD): Center for Drug Evaluation and Research and Rockville (MD): Center for Biologics Evaluation and Research, US Food and Drug Administration; 2010. Available: www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM202140.pdf (accessed 2012 June 15).
4. ICH Expert Working Group. *ICH Harmonised Tripartite Guideline: Statistical principles for clinical trials. E9*. Geneva: ICH; 1998. Available: www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E9/Step4/E9_Guideline.pdf (accessed 2012 June 15).
5. ICH Expert Working Group. *ICH Harmonised Tripartite Guideline: Choice of control group and related issues in clinical trials. E-10*. Geneva: ICH; 2000. Available: www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E10/Step4/E10_Guideline.pdf (accessed 2012 June 15).
6. Committee for Medicinal Products for Human Use. *Guideline on the choice of the non-inferiority margin*. London (UK): European Medicines Agency; 2005. Available: www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003636.pdf (accessed 2012 June 15).
7. D’Agostino RB S, Massaro JM, Sullivan LM. Non-inferiority trials: design concepts and issues — the encounters of academic consultants in statistics. *Stat Med* 2003;22:169-86.
8. Fleming TR. Design and interpretation of equivalence trials. *Am Heart J* 2000;139:S171-6.
9. Hung HM, Wang SJ, O’Neill R. A regulatory perspective on choice of margin and statistical inference issue in non-inferiority trials. *Biom J* 2005;47:28,36; discussion 99-107.
10. Lange S, Freitag G. Choice of delta: requirements and reality — results of a systematic review. *Biom J* 2005;47:12,27; discussion 99-107.
11. Colwell CW, Berkowitz SD, Davidson BL, et al. Comparison of ximelagatran, an oral direct thrombin inhibitor, with enoxaparin for the prevention of venous thromboembolism following total hip replacement. A randomized, double-blind study. *J Thromb Haemost* 2003;1:2119-30.

12. Eriksson BI, Dahl OE, Rosencher N, et al. Oral dabigatran etexilate vs. subcutaneous enoxaparin for the prevention of venous thromboembolism after total knee replacement: The RE-MODEL randomized trial. *J Thromb Haemost* 2007;5:2178-85.
13. Ginsberg JS, Davidson BL, Comp PC; Re-Mobilize Writing Committee, et al. Oral thrombin inhibitor dabigatran etexilate vs north american enoxaparin regimen for prevention of venous thromboembolism after knee arthroplasty surgery. *J Arthroplasty* 2009;24:1-9.
14. Eriksson BI, Dahl OE, Rosencher N, et al. Dabigatran etexilate versus enoxaparin for prevention of venous thromboembolism after total hip replacement: a randomised, double-blind, non-inferiority trial. *Lancet* 2007;370:949-56.
15. Lassen MR, Raskob GE, Gallus A, et al. Apixaban or enoxaparin for thromboprophylaxis after knee replacement. *N Engl J Med* 2009;361:594-604.
16. Kaul S, Diamond GA. Making sense of noninferiority: clinical and statistical perspective on its application to cardiovascular clinical trials. *Prog Cardiovasc Dis* 2007;49:284-99.
17. Fleming TR, Emerson SS. Evaluating rivaroxaban for nonvalvular atrial fibrillation — regulatory considerations. *N Engl J Med* 2011;365:1557-9.
18. New joint industry clinical trials transparency position requires companies to disclose all clinical trials in patients [news release]. The Pharma Letter [London (UK)] 11 Nov 2009. Available: www.thepharmaletter.com/file/33724/new-joint-industry-transparency-position-requires-companies-to-disclose-all-clinical-trials-in-patients.html (accessed 2012 June 15).
19. Eriksson BI, Agnelli G, Cohen AT, et al. The direct thrombin inhibitor melagatran followed by oral ximelagatran compared with enoxaparin for the prevention of venous thromboembolism after total hip or knee replacement: the Express study. *J Thromb Haemost* 2003;1:2490-6.
20. Eriksson BI, Borris LC, Friedman RJ, et al. Rivaroxaban versus enoxaparin for thromboprophylaxis after hip arthroplasty. *N Engl J Med* 2008;358:2765-75.
21. Lassen MR, Ageno W, Borris LC, et al. Rivaroxaban versus enoxaparin for thromboprophylaxis after total knee arthroplasty. *N Engl J Med* 2008;358:2776-86.
22. Turpie AGG, Lassen MR, Davidson BL, et al. Rivaroxaban versus enoxaparin for thromboprophylaxis after total knee arthroplasty (RECORD 4): a randomised trial. *Lancet* 2009;373:1673-80.
23. Lassen MR, Raskob GE, Gallus A, et al. Apixaban versus enoxaparin for thromboprophylaxis after knee replacement (ADVANCE-2): a randomised double-blind trial. *Lancet* 2010;375:807-15.
24. Lassen MR, Gallus A, Raskob GE, et al. Apixaban versus enoxaparin for thromboprophylaxis after hip replacement. *N Engl J Med* 2010;363:2487-98.

Affiliations: From the Division of Pharmacoepidemiology and Clinical Pharmacology (Wangge, de Boer, Knol), Utrecht Institute for Pharmaceutical Sciences, Utrecht University; and Julius Centre for Health Sciences and Primary Care (Roes, Hoes, Knol), University Medical Centre Utrecht, Utrecht, the Netherlands

Contributors: All of the authors were involved in the conception and design of the study. Grace Wangge and Mirjam Knol acquired the data and conducted the analysis. Grace Wangge, Mirjam Knol and Kit Roes interpreted the data. Grace Wangge wrote the first draft of the manuscript, and all of the authors contributed to subsequent revisions of the manuscript and approved the final version submitted for publication. Anthonius de Boer is the guarantor of this article.

Funding: This study was performed in the context of the Escher project (T6-202), a project of the Dutch Top Institute Pharma. The Escher project brings together university and pharmaceutical partners with the aim to energize pharmaceutical research and development by identifying, evaluating and removing regulatory and methodological barriers to bring efficacious and safe medicines to patients in an efficient and timely fashion. The project focuses on delivering evidence and credibility for regulatory reform and policy recommendations. The funders had no role in the study's design, data collection and analysis or preparation of the manuscript.