Non-inferiority trials:

methodological and regulatory challenges

Grace Wangge

Non-inferiority trials: methodological and regulatory challenges

Thesis, Utrecht University, with summary in Bahasa Indonesia and in Dutch

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 R&D 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.

ISBN	978-94-6182-163-8
Author	Grace Wangge
Layout	Masadepan@tech
Cover design	Grace Wangge (inspired by the works of MC Escher)
Printed by	Offpage.nl

Non-inferiority trials:
methodological and regulatory challenges

Uji klinis Non-inferior:
tantangan dalam metodologi dan regulasi
(dengan ringkasan dalam bahasa Indonesia)

'Non-inferiority' studies:
methodologische en regelatoire uitdagingen
(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht op gezag van de rector magnificus, prof. dr. G.J. van der Zwaan, ingevolge het besluit van het college voor promoties in het openbaar te verdedigen op maandag 5 november 2012 des middags te 16.15 uur

door

Grace Wangge
geboren op 16 oktober 1977 te Jakarta, Indonesië

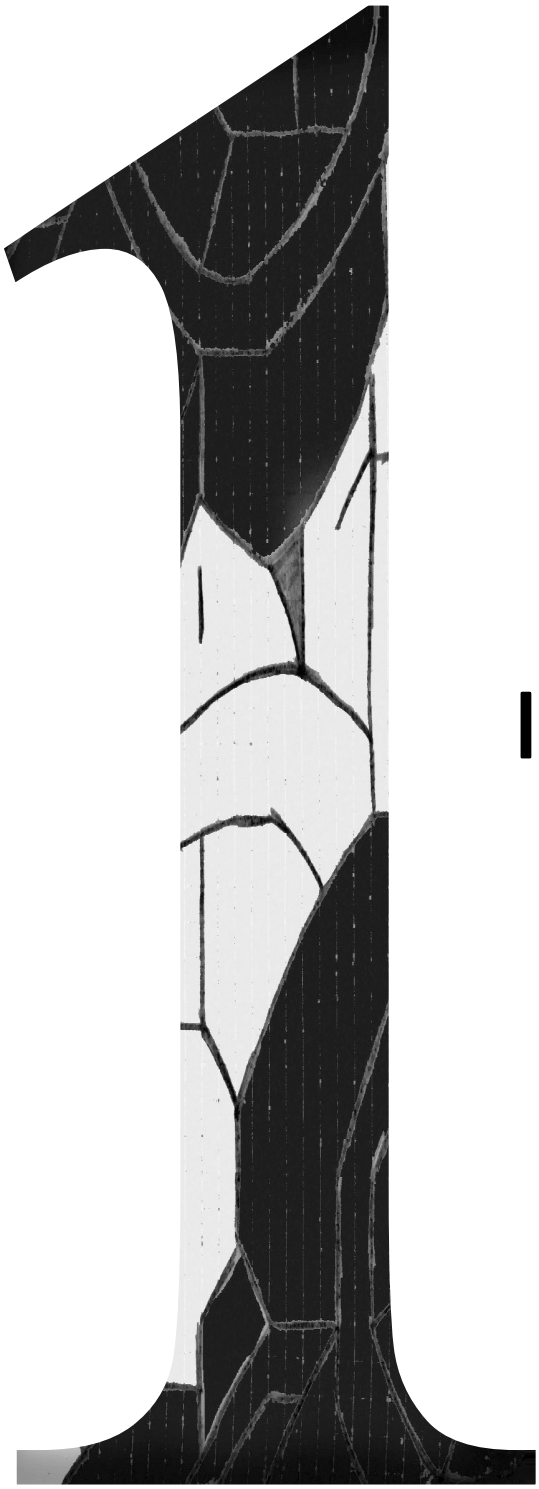
Promotoren :	Prof. dr. A. de Boer
	Prof. dr. A. W. Hoes
Co–promotoren :	Dr. M. J. Knol
	Dr. O.H. Klungel

It is human nature to want to exchange ideas

MC Escher, Dutch graphic artist (1898– 1972)

Dedicated to all scientists, artists and story tellers

Table of contents



Introduction

Introduction

Randomized controlled trials (RCT) are considered as the gold standard to confirm a drug's efficacy.[1] Nowadays, active-controlled trials are often requested instead of, or in addition to, placebo-controlled trials as the basis for marketing authorization and reimbursement decisions. A study showed that for 48 % of new medicines approved between 1999 and 2005 in the European Union at least one active-controlled trial was performed in their development phase.[2]

An active-controlled trial may have a non-inferiority (NI) design, a design used to show that a new drug is not worse than an active comparator. NI trials can be used in a situation when a new drug considered has a similar efficacy profile as its comparator but may offer other advantages over the existing drug such as a novel method of administration or a better safety profile. In a regulatory setting, NI trials can be used to provide primary, but indirect, evidence of efficacy of a novel drug in cases where a placebo control treatment is not ethically justified. [3,4]

The concept of NI trials has been developed in the 1970s and was inspired by the methodology of (bio) equivalence trials. [4-6] During that time, the term non-inferiority and therapeutic equivalence were used interchangeably. NI trials have become popular in the 1990s, especially after the introduction of several regulatory guidelines that regulate the use of active-controlled trials. This is shown by a major increase of publications on NI trials since the first guideline, the ICH E9, was published in 1998. A search in Pubmed revealed only 5 publications in the year 1998 and the number increased to more than 200 publications per year since 2008 (Figure 1). This illustrates the growing interest in NI trials as well as the increased need for readers and clinicians to understand the concept of this methodology.

The ICH E9[7], the ICH E10[8], CHMP guidelines[9], and FDA draft guideline on NI trials[10] are the currently available guidelines for the appropriate conduct of NI trials. Furthermore, for reporting NI trials, the Consolidated Standards of Reporting Trials (CONSORT) organization has released the extension of the CONSORT statement on NI trials [11] that recommends how to report an NI trial.

From a methodological perspective, compared to superiority trials, NI trials have challenges in design and analysis that can influence proper inference. First, there are different methods to determine the NI margin and there are debates on whether the NI margin should be determined based on statistical or clinical considerations or both. Second, a difficulty in interpreting NI trials is their lack of ability to distinguish an effective drug from an ineffective drug i.e. assay sensitivity [7, 8], without relying on evidence outside the trial. A drug is considered effective if it shows a significant treatment effect compared with placebo. An additional placebo arm is recommended to confirm assay sensitivity [2,6,9]. However, this is often impossible due to ethical reasons. Last, the validity of the historical data that was used as the reference for the current trial, i.e. constancy assumption, is a critical point in the interpretation of NI trials.

The Escher project: science-driven drug regulation and innovative research throughout phased drug development

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 R&D by identifying,

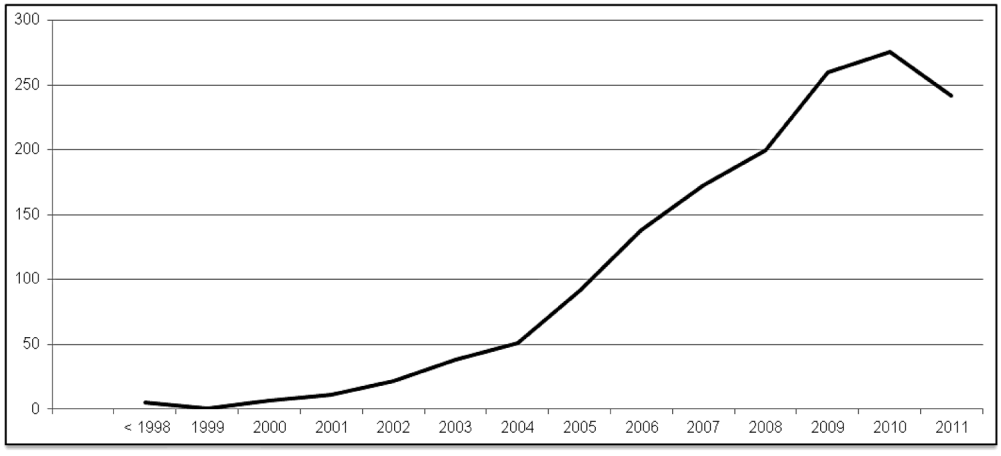


Figure 1. Number of publications of NI trials per year found in Pubmed

Note: search was done on 12th March 2012 with keywords : non-inferior*[All Fields] OR noninferior*[All Fields] OR ("active control"[All Fields] AND "equivalence"[All Fields]) NOT "bioequivalence"[All Fields] AND "humans"[MeSH Terms] AND English[lang] AND Randomized Controlled Trial[Publication Type]

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.

Objective and outline of the thesis

In the context of the Escher project, the objective of this thesis is to look deeper into the challenges in the methodology of NI trials, and the role of regulatory guidelines in it.

Challenges in methodology of NI trials

The first part of this thesis (Chapter 2) focuses on the methodology of NI trials and its challenges. In **chapter 2.1** and **2.2**, we analysed publications of 232 NI trials to identify how NI trials are currently designed, analyzed, and reported. Furthermore, we explain and address the complications in the interpretation of NI trials that arise from the indirect comparison with placebo.

In **chapter 2.3** we explain the method to determine an NI margin according to the draft FDA guideline on NI trials [10] and present a case study on the NI margins used in trials on novel anticoagulants, drugs for which many NI trials are and were performed.

We gathered opinions of clinical experts, regulators and researchers on a clinically relevant NI margin in **chapter 2.4** using an online questionnaire. A case scenario of a prospective NI trials in oral anticoagulants for prophylaxis of venous thromboembolic events post orthopaedic surgery was developed. We described the experts' choices of NI margin and their reasoning.

Regulatory challenges on NI trials methodology

The second part of this thesis focuses on the regulatory challenges of NI trials. In **chapter 3.1** we reported the content analysis of 156 final-advice letters from 94 different applicants of scientific advice to the European Medicines Agency (EMA).

In **chapter 3.2** we looked at 41 published post-authorization NI trials. We determined whether these trials reported benefit claims beyond clinical efficacy and how these additional claims were supported or proven in the trials.

General discussion

This thesis is ends with a discussion (chapter 4), where we discussed whether the pledge made by Garattini and Bertele[12] five years ago is still valid. In their article published in 2007, Garattini and Bertele have condemned NI trials as unethical because it disregards interest of the patients by exposing patients to a drug without the intention to show a new drug is better than the comparator. The new drug might even be worse than the standard drug. They argued that the scientific community should ban NI (and equivalence) trials, whatever measures are taken to prevent their methodological pitfalls.

We will focus our discussion on the ethical, methodological and regulatory arguments for and against banning of NI trials.

References

(1) Hadorn DC, Baker D, Hodges JS, Hicks N. Rating the quality of evidence for clinical practice guidelines J Clin Epidemiol 1996;49(7):749-54.

(2) van Luijn JC, van Loenen AC, Gribnau FW, Leufkens HG. Choice of Comparator in Active Control Trials of New Drugs Ann Pharmacother 2008;42(11):1605-12.

(3) 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 Feb;47(1):28-36; discussion 99-107.

(4) 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 Jan 30;22 (2):169-86.

(5) Blackwelder W. “Proving the null hypothesis” in clinical trials Control Clin Trials 1982;3 (4):345 -53.

(6) Ebbutt AF, Frith L. ,Practical issues in equivalence trials. Stat Med. 1998 Aug 15-30;17 (15-16):1691-701.

(7) ICH Expert Working Group. (1998) Stastical principles for clinical trials. E9 . Available at : http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E9/Step4/E9_Guideline.pdf

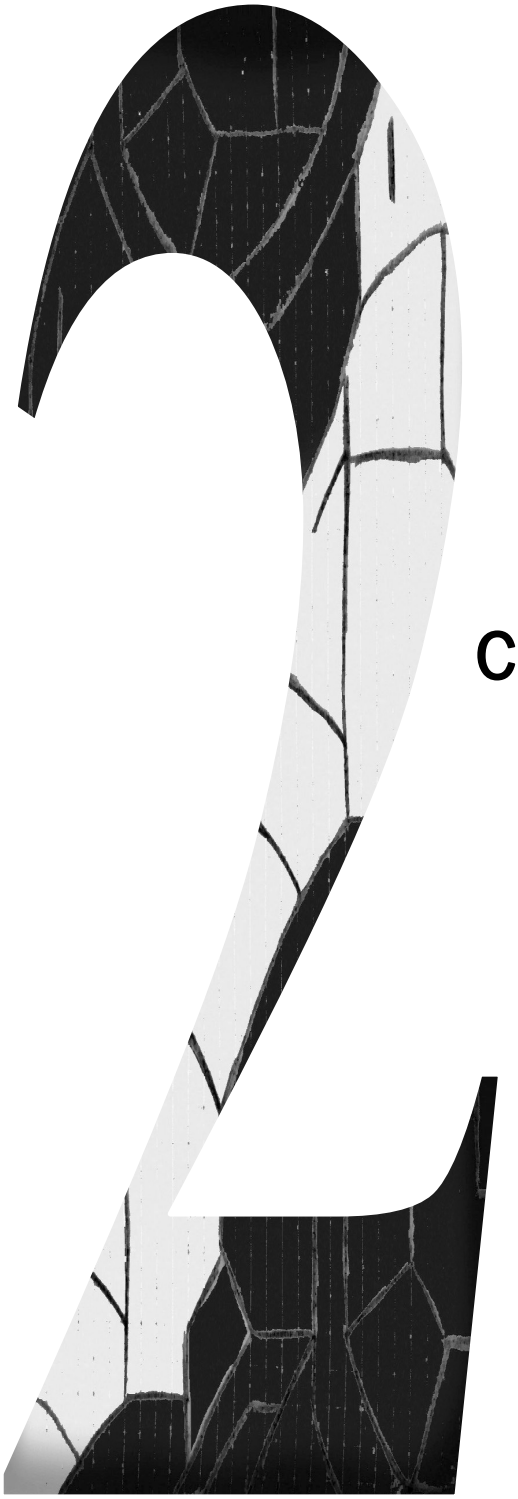
(8) ICH Expert Working Group. (2001) Choice of control group and related issues in clinical trials. E-10 .Available at : http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E10/Step4/E10_Guideline.pdf

(9) Committee for Medicinal Products for Human Use (CHMP). (2005) Guideline on the choice of the non-inferiority margin. EMEA/CPMP/EWP/2158/99 . Available at : http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003636.pdf

(10) Center for Drug Evaluation and Research (CDER) Center for Biologics Evaluation and Research (CBER). (2010) DRAFT GUIDANCE : Guidance for industry non-inferiority clinical trials . Available at : http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM202140.pdf

(11) Piaggio G. Reporting of Noninferiority and Equivalence Randomized Trials: An Extension of the CONSORT Statement JAMA: The Journal of the American Medical Association 2006;295(10):1152-60.

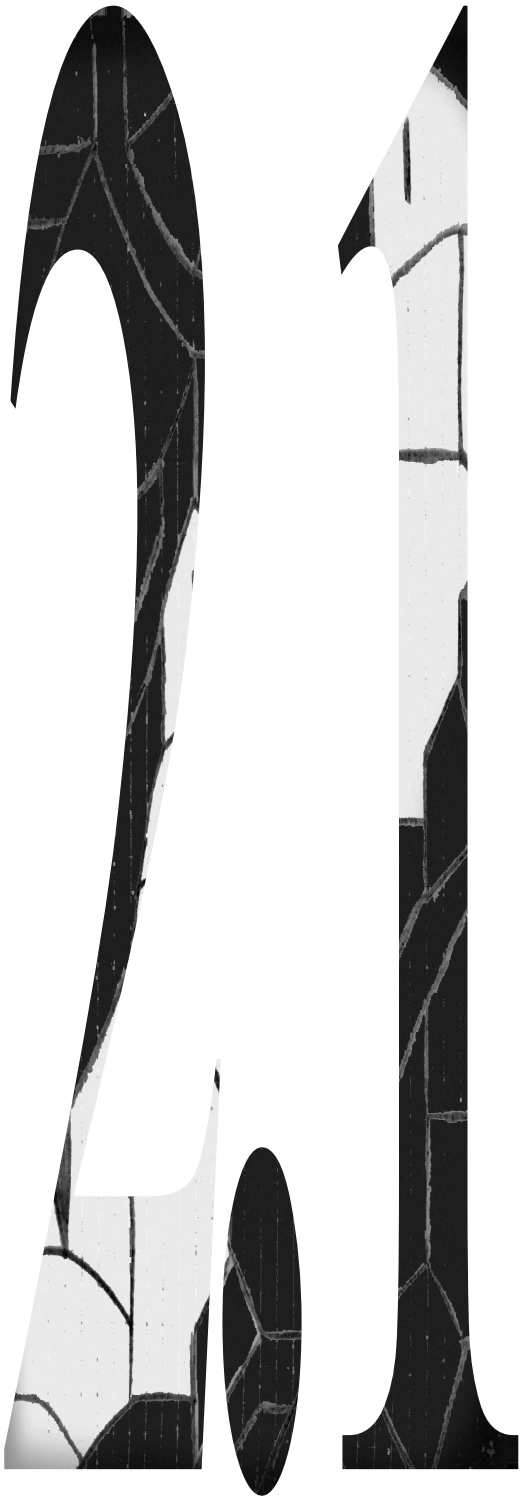
(12) Garattini S, Bertele V. Non-inferiority trials are unethical because they disregard patients' interests. The Lancet 2007;370(9602):1875-77.



Challenges in the methodology of non-inferiority trials

Are you really sure that a floor can't also be a ceiling ?

MC Escher, Dutch graphic artist (1898– 1972)



Interpretation and inference in non-inferiority randomized controlled trials in drug research

Grace Wangge
Olaf H. Klungel
Kit CB Roes
Anthonius de Boer
Arno W. Hoes
Mirjam J. Knol

Clin Pharmacol Ther. 2010
Sep;88(3):420-3.

Non-inferiority (NI) trials in drug research are used for the purpose of demonstrating that a new treatment is not worse than a proven active comparator, thereby indirectly showing that the treatment is effective. This article explains and addresses the complications in the interpretation of NI trials that arise from the indirect comparison. On the basis of our review of 232 trials, we conclude that the interpretation and inference of NI trials are complicated, partly because of the incompleteness of the information.

Although often criticized as incapable of distinguishing between "method effectiveness" (efficacy) and "use effectiveness" (effectiveness), the randomized controlled trial (RCT) is still a preferred design for determining the efficacy profile of a new drug. In order to determine the profile, an RCT may have either a superiority design or an NI design. In a superiority design, the objective is to demonstrate that the new drug is effective by showing that it is significantly better than its comparator, typically placebo treatment. In this context, an NI trial is intended to demonstrate that the new drug is not worse than its comparator, thereby indirectly showing that the new treatment is effective (i.e., more effective than placebo).[1,2] Typically, an NI trial is employed if use of placebo is not ethically justified. NI trials can also be used if a new drug is expected to have an efficacy profile similar to that of its comparator but may offer other advantages, including novel methods of administration or a better safety profile. Critics, however, have pointed out that NI trials seem to allow the pharmaceutical industry to gather additional data for their marketing activities by permitting drugs without additional efficacy to enter the market and that it is problematic to conclude, from the results of NI trials, whether the candidate drug would have a more beneficial treatment effect than a placebo.[3]

The interpretation of NI trials can be difficult because of the use of an NI margin and also because of related issues including assay sensitivity and the constancy assumption. Our aim in this article is to provide an improved understanding of NI trials, by describing data related to 232 NI trials. We performed a search in PubMed on 5 February

2009 using the search terms " non-inferior*" and "noninferior*" or "active control" and "equivalence", in combination with the MeSH term "humans" and "randomized controlled trial" as publication type. This search resulted in 669 articles, and, based on pragmatic consideration rather than formal sample-size calculations, we randomly selected 300 for our review. Subsequently, we excluded studies on bioequivalence, phase I studies, non–drug trials, and articles that did not have full text in English, reducing the number to 227 articles reporting 232 NI trials.

NI Margin

The underlying objective of many NI trials is to indirectly demonstrate efficacy against placebo; this is done by showing that the new drug is not worse than the active comparator. The main step of designing an NI trial is pre-specification of a margin, or boundary, at which it can still be established that the new drug is similar or not worse than its comparator. The NI margin should be chosen such that the new drug can be shown to be effective relative to placebo and needs to account for the uncertainty in the effect size of the active control vs. placebo. In many cases, it is sensible to adopt stricter margins so as to ensure that the effect is clinically relevant, in the sense that it preserves a substantial part of the efficacy of the active control. This concept of margin determination is emphasized and discussed in detail in the US Food and Drug Administration's new draft guideline for NI trials.[4]

These essential features of an NI margin lead to the following basic approach for setting the margin (assuming larger outcomes are better). The statistical margin for the effect between the new drug and active control is set at the upper boundary of the 95% confidence interval (CI) of difference between placebo and active control. This limiting value is obtained from relevant previous placebo-controlled trial(s) of the active comparator[5,6] Subsequently, a clinical relevance margin is set to preserve a fraction of the active comparator effect, e.g., 50%, relative to placebo. This approach is the one that was recommended by the US Food and Drug Administration in 1999[5] included in a more elaborate form in its new draft guidance.[4] Note that this is a substantial simplification,

and the margin setting depends on the type of outcome and the effect measure (e.g., difference, ratio, or hazard ratio). The basis for determining the clinical relevance margin is "clinical judgment." This includes weighting factors such as the severity of the disease, the available treatment options, and the safety profile.[6] This step is mainly subjective but is considered to be a key step in determining the margin[7] because it helps prevent "biocreep," i.e., moving gradually to less effective treatments.

To illustrate how the NI margin is determined in practice, we describe a trial by Cooper et al.[8] In that trial, the efficacy and safety of ibandronate administered orally once a week were compared with those of the same drug administered orally once a day. The primary outcome was the relative change in lumbar spine bone mineral density after 48 weeks. The statistical margin was based on the observed superiority of a daily ibandronate regimen over a placebo regimen in increasing lumbar spine bone mineral density in a similar population in a previous trial. The statistical margin was set at 3.3%,

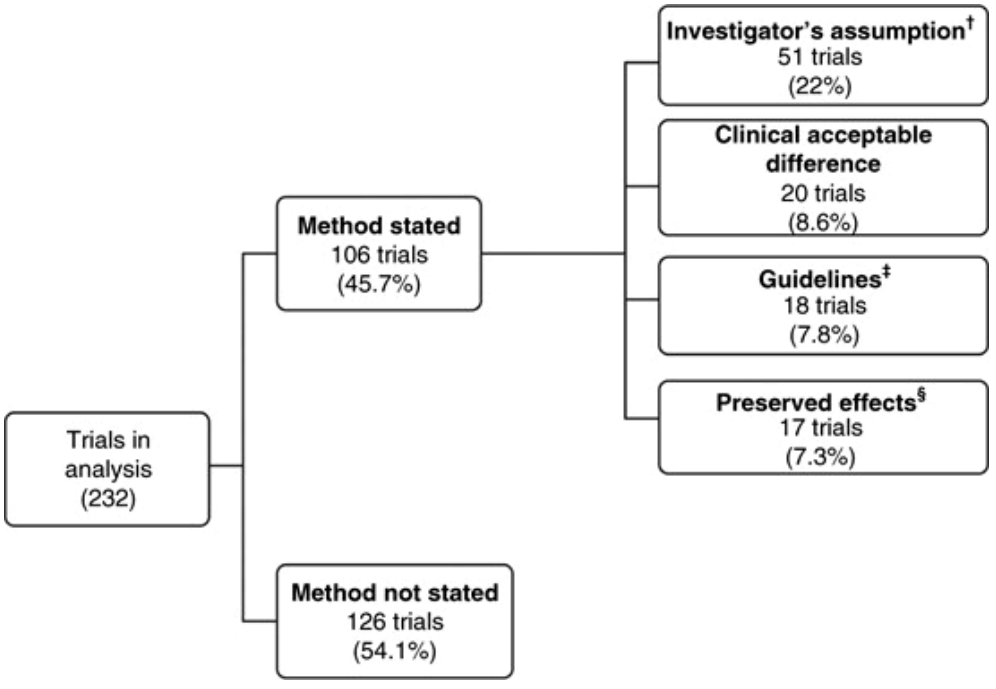


Figure 1. Methods to determine the margin

†In these trials, the source of data to determine the margin was not clear, and a statement about whether the margin is a clinically acceptable margin was missing. ‡The margin was determined on the basis of a guideline or recommendation from regulatory bodies on NI margins of specific drugs. §In 15 trials, NI margin was based on ≥50% preserved effect of comparator vs. placebo; in two trials, the margin was based on <50% preserved effect.

representing the effect of daily ibandronate administration vs. placebo. Next, the investigator performed discounting and preserved a 50% fraction, resulting in an NI margin of 1.65%. In the final step, the investigator determined whether this margin was clinically acceptable. However, the health authorities subsequently asked for a tighter margin of 1.10% to be set, based on a 33% preserved fraction.

In our review, 227 (97.8%) trials reported the NI margins (Figure 1). However, only 106 (45.7%) trials described how they had determined the margin. The International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use, the European Medicines Agency, and the most recent draft of the US Food and Drug Administration guidelines emphasize that determination of the NI margin should be based on both clinical and statistical margins. Although it is hard to determine from the literature whether the trials actually followed the guidelines, we found 20 (8.6%) trial reports stating that the margin was an acceptable clinical difference. In 51 (22%) of the trials, the margin was determined merely on the basis of the investigator's own assumption. Of 17 (7.3%) trials that stated that the NI margin had been determined after taking into account preserved effects, 15 used a preserved fraction of ≥ 50%.

Interpretation of NI trials

The inference from the result of an NI trial is based on the CI of the treatment difference between the new drug and its comparator. NI is inferred when the CI is at the correct side and excludes the NI margin.[9]

To illustrate this, we categorized the possible CIs in NI trials into six types as presented in Figure 2 (assuming that larger outcomes are better).

The basic interpretation of the CIs in terms of NI of the new drug[10] in Figure 2 is as follows: NI can be shown from type A, B and C since their CI excludes the NI margin. While in type D, E and F NI of the new drug is not shown.[10] Of course, given that the CIs also quantify the treatment difference directly, type C, which lies completely beyond the point-of-no difference line, would potentially demonstrate that the new drug is superior to its comparator. This represents a switch from NI to superiority, which is regulated, for

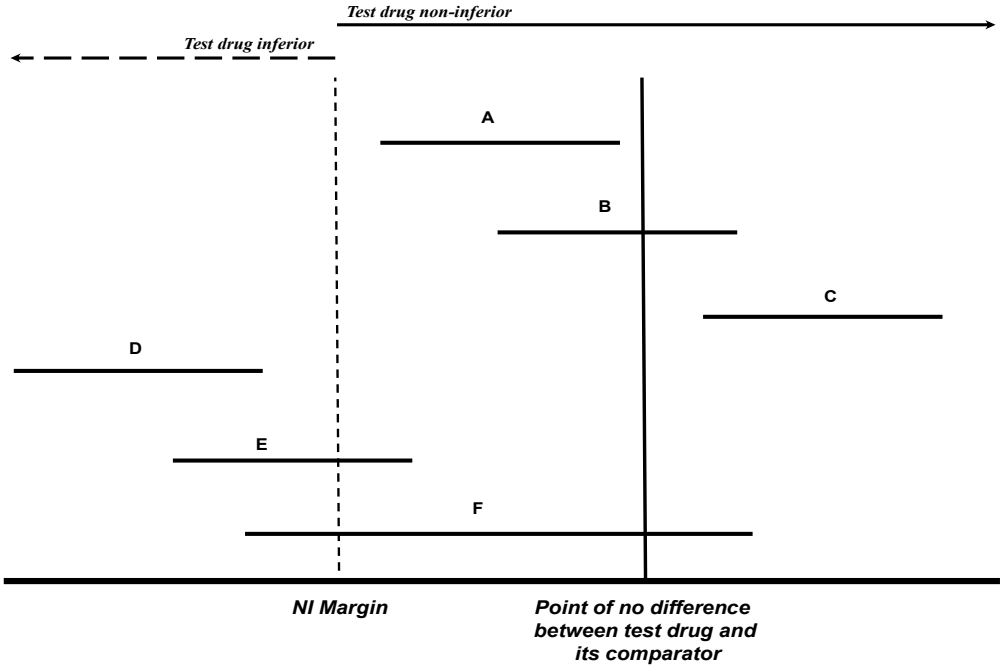


Figure 2. The confidence interval categories and non-inferiority interpretation.

The dashed vertical line represents the NI margin, the solid vertical line is the point-of-no-difference line, and the horizontal line represents CI. The point-of-no-difference is the point at which the estimated treatment difference between the new drug and comparator is neutral: zero for a difference in outcome or one for a ratio.

example, by the Committee for Proprietary Medicinal Products guidelines.[11] If there is a possibility of switching to superiority, this should be indicated at the planning stage as part of a stepwise approach. Also, the results from the intention-to-treat and per-protocol analyses should be consistent because the results of these analyses may have a different interpretation in an NI trial as compared with a superiority trial. In superiority trials, intention-to-treat analysis generally gives more conservative results than per-protocol analysis, whereas in NI trials, per-protocol analysis may give more conservative results.[11]

Type A also requires cautious interpretation. Although the lower limit lies above the NI margin, thereby showing NI, the upper limit lies below the point-of-no-difference, indicating that the new drug is actually statistically inferior to its comparator. However, the new drug can still be claimed to be clinically noninferior if the NI margin was determined on the basis of clinical relevance.

In our review, we extracted the CIs of the differences between the two active

Table 1. The comparison of guidelines requirements, the author's conclusion, and the CI in 198 trials with single type of CI

Confidence interval		Interpretation based on guidelines	Author's conclusion, N (%)			
Type	N (%)		Non-inferiority	Superiority	Inferiority	Other*
A	10 (5.0)	NI	9 (4.5)		1 (0.5)	
B	138 (69.7)	NI	134 (67.7)	1 (0.5)		3 (1.5)
C	19 (9.6)	NI	15 (7.6)	4 (2.0)		
D	3 (1.5)	NI not shown			1 (0.5)	2 (1.0)
E	9 (4.5)	NI not shown	2 (1.0)#		6 (3.0)	1 (0.5)
F	19 (9.7)	NI not shown	14 (7.2)#		5 (2.5)	
Total	198 (100)		174 (88.0)	5 (2.5)	13 (6.5)	6 (3.0)

Note : * : 'Other' conclusion defined as a very general conclusion stated by the investigator on the basis of trial results and not related to the main hypothesis. # : Trials with incorrect interpretation relative to the guideline. NI : Non-inferiority

treatments on the primary end points as presented in each trial report. Of 232 trials, 198 (85.3%) had a single type of CI, whereas 13 (5.6%) trials had multiple CIs, and 21 (9.0%) did not report their CI. We included only the trials with a single type of CI. Of these 198 trials, 138 (69.7%) had a type B CI (Table 1).

As can be seen from Table 1, the majority of the trials were interpreted correctly with regard to NI, although there were 17 (8.7%) trials with incorrect interpretation. Sixteen of these trials concluded NI of the new drug, although, according to the guideline, NI was not shown. In one trial with a type A CI, the conclusion was that the new drug was inferior to the comparator, whereas, according to strict interpretation of the guideline, the new drug was noninferior. Of 19 (9.6%) trials with CI of type C, we found that 15 authors claimed NI for the respective drugs, and four authors claimed superiority. Three of these latter four trials did not state that they had preplanned a switch from an NI trial to a superiority trial.

This observation shows that the interpretation becomes complicated when the main objective of the trial is altered post hoc, with the possibility of switching from NI to superiority being explored at a later stage of the trial instead of being preplanned.

Assay sensitivity and constancy assumption

The approach to designing and interpreting results of NI trials as described here depends greatly on two key aspects that cannot be verified within the trial: assay sensitivity and the constancy assumption[10,12] Assay sensitivity is defined as the ability of an RCT to distinguish an effective treatment from an ineffective treatment.[10,13] A drug is considered effective if it shows a significant treatment effect as compared with placebo.

In a superiority RCT, a significant difference between two treatments directly confirms assay sensitivity. In contrast, an NI trial does not directly show the efficacy of both drugs as compared with placebo. An NI could mean that both drugs were effective, but it could also mean that both drugs were ineffective. One possible solution is to include a placebo arm to confirm that both the new drug and the comparator drug are better than placebo.[2,14]However, the very rationale for conducting an NI trial is usually that the use of a placebo is impossible because of ethical considerations. We observed only 14 (6.0%) trials that included placebo arms. Most of these trials were safety trials.

When designing NI trials, other options should be considered before making the decision to omit the placebo arm. The options include unequal random allocation of treatment (fewer in the placebo group), shorter duration of treatment in the placebo group, adaptive trial designs in which placebo non-responders can be reallocated, or a Bayesian approach[4,15,16] In the absence of a placebo arm, interpretation of results from an NI trial relies on the strong assumption that the conditions of the trial, the manner of its execution, and the population included were such that the data pertaining to the active control would have separated from those of the placebo had a placebo arm been included.[6] This renders NI trials aimed at demonstrating efficacy virtually useless for applications in which the data pertaining to treatments of known efficacy frequently fail to separate sufficiently from those of placebo, even in well-controlled trials (e.g., in psychiatry).

Another, partially related, assumption in NI trials that cannot be verified within the trial is the constancy assumption. The determination of the NI margin relies directly on the size of the estimated treatment effect between the active comparator and the placebo.

Hence, for the inference to be valid, it has to be assumed that this estimate is (still) accurate for the trial at hand. This cannot be assessed with total objectivity. However, it can be supported by a proper meta-analysis and by a demonstration of similarity between the current trial (and the patients enrolled in it) and the trials used for setting the margin. [2,6,17]But the constancy assumption also relies on the absence of any influence from a number of factors, e.g., changes in standard of care, which are not easily verifiable. We found only nine (3.9%) trials that discussed the constancy assumption.

If a placebo arm cannot be included, the authors should discuss how they have arrived at the conclusion that the trial had assay sensitivity and provide data-driven as well as clinical reasons for assuming that the constancy assumption holds true. Without these assessments, the reader cannot reliably judge whether the conclusions from the trial are valid and relevant for treatment decisions.

Conclusion

In this article, we explain what NI trials are, how they should be designed, and which aspects need to be included in the interpretation of such trials. Our review showed that <50%of the NI trials reported the method used to determine the NI margin, and <10% of the trials stated that the NI margin was a priori justified on the basis of clinical margin. Importantly, the recently released Food and Drug Administration draft guideline for NI trials also emphasizes the need for prior justification of NI margin based on a clinical margin and on statistical grounds.

Furthermore, we found that >8% of the trials were interpreted incorrectly, and <10% of the trials included placebo arms to ensure assay sensitivity or even discussed assay sensitivity or the validity of the constancy assumption. These findings provide evidence that the interpretation of and inference from NI trials are complicated, and publications do not routinely contain the information needed.

Improvement in design, interpretation, and publication of NI trials is necessary, and this overview paper may help authors and readers of NI trials to achieve that.

References

(1) 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 Feb;47(1):28-36; 99-107.

(2) 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 Jan 30;22 (2):169-86.

(3) Garattini S, Bertele V. Non-inferiority trials are unethical because they disregard patients' interests. The Lancet 2007;370(9602):1875-77.

(4) Center for Drug Evaluation and Research (CDER) Center for Biologics Evaluation and Research (CBER). (2010) DRAFT GUIDANCE : Guidance for industry non-inferiority clinical trials . Available at : http://www.fda.gov/downloads/Drugs/ GuidanceComplianceRegulatoryInformation/Guidances/UCM202140.pdf

(5) Snapinn SM. Alternatives for discounting in the analysis of noninferiority trials. J Biopharm Stat 2004 May;14(2):263-73.

(6) Hung HMJ, Wang S, O'Neill R. Challenges and regulatory experiences with non-inferiority trial design without placebo arm. Biometrical Journal 2009;51(2):324-34.

(7) Committee for Medicinal Products for Human Use (CHMP). (2005) Guideline on the choice of the non-inferiority margin. EMEA/CPMP/EWP/2158/99 . Available at : http:// www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/ WC500003636.pdf

(8) Cooper C, Emkey RD, McDonald RH, Hawker G, Bianchi G, Wilson K, et al. Efficacy and Safety of Oral Weekly Ibandronate in the Treatment of Postmenopausal Osteoporosis. Journal of Clinical Endocrinology & Metabolism 2003 October 01;88(10):4609-15.

(9) ICH Expert Working Group. (1998) Stastical principles for clinical trials. E9 . Available at : http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/ E9/Step4/E9_Guideline.pdf

(10) ICH Expert Working Group. (2001) Choice of control group and related issues in clinical trials. E-10 .Available at : http://www.ich.org/fileadmin/Public_Web_Site/ ICH_Products/Guidelines/Efficacy/E10/Step4/E10_Guideline.pdf

(11) Committee for Proprietary Medicinal Products (CPMP). (2000) Points to consider on switching between superiority and non-inferiority. Available: http://www.emea.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003658.pdf

(12) Snapinn SM. Noninferiority trials. Curr Control Trials Cardiovasc Med 2000;1(1):19-21.

(13) Kay R. Equivalence and non-inferiority. In: Kay R, editor. Statistical thinking for non-statisticians in drug regulation: Wiley, John & Sons, Incorporated; 2007. p. 173-91.

(14) Senn S. Active control equivalence studies. Statistical Issues in Drug Development, 2nd Edition Glasgow, UK: Wiley; 2007. p. 235-47.

(15) McMahon RP, Kelly DL, Boggs DL, Li L, Hu Q, Davis JM, et al. Feasibility of Reducing the Duration of Placebo-Controlled Trials in Schizophrenia Research. Schizophrenia Bulletin 2008 March 01;34(2):292-301.

(16) Britton M. Bayesian Approach to Three-Arm Non Inferiority Trials Mathematics Theses 2007 -05-03(27).

(17) Piaggio G. Reporting of Noninferiority and Equivalence Randomized Trials: An Extension of the CONSORT Statement JAMA: The Journal of the American Medical Association 2006;295(10):1152-60.

Room for improvement in conducting and reporting non-inferiority randomized controlled trials on drugs

A systematic review

Grace Wangge
Olaf H. Klungel
Kit CB Roes
Anthonius de Boer
Arno W. Hoes
Mirjam J. Knol

PLoS One. 2010 Oct 27;5 (10):e13550.

ABSTRACT

Background: A non-inferiority (NI) trial is intended to show that the effect of a new treatment is not worse than the comparator. We conducted a review to identify how NI trials were conducted and reported, and whether the standard requirements from the guidelines were followed.

Methodology and Principal Findings: From 300 randomly selected articles on NI trials registered in PubMed at 5 February 2009, we included 227 NI articles that referred to 232 trials. We excluded studies on bioequivalence, trials on healthy volunteers, non-drug trials, and articles of which the full-text version could not be retrieved. A large proportion of trials (34.0 %) did not use blinding. The NI margin was reported in 97.8 % of the trials, but only 45.7 % of the trials reported the method to determine the margin. Most of the trials used either intention to treat (ITT) (34.9 %) or per-protocol (PP) analysis (19.4 %), while 41.8 % of the trials used both methods. Less than 10% of the trials included a placebo arm to confirm the efficacy of the new drug and active comparator against placebo, and less than 5.0 % were reporting the similarity of the current trial with the previous comparator's trials. In general, no difference was seen in the quality of reporting before and after the release of the CONSORT statement extension 2006 or between the high-impact and low-impact journals.

Conclusion: The conduct and reporting of NI trials can be improved, particularly in terms of maximizing the use of blinding, the use of both ITT and PP analysis, reporting the similarity with the previous comparator's trials to guarantee a valid constancy assumption, and most importantly reporting the method to determine the NI margin.

Introduction

In the drug development process, the randomized controlled trial (RCT) can have a superiority, equivalence or a non-inferiority design. A superiority trial aims to demonstrate the superiority of a new therapy compared to an active comparator or a placebo, while an equivalence trial aims to demonstrate that a new therapy is equivalent (within margins) to its active comparator. In non-inferiority (NI) trials, the aim is to show that the new treatment is not worse than the comparator, which typically is an active drug.

NI trials can be used in a situation when a new drug considered has a similar efficacy profile as its comparator but may offer other advantages over the existing drug such as a novel method of administration or a better safety profile. In a regulatory setting, NI trials can be used to provide primary, but indirect, evidence of efficacy of a novel drug in cases where a placebo control treatment is not ethically justified.[1,2]

Critics have pointed at various drawbacks of NI trials, questioning whether they are really useful. Some argue that NI trials only benefit pharmaceutical industry as they allow drugs without additional clinical efficacy to enter the market.[3,4] However, as argued by Jones et.al, in some cases the new treatment may have no direct advantage but may present an alternative or second line therapy**.**[5]

From a methodological perspective, compared to superiority trials, NI trials have methodological issues in design and analysis that can influence proper inference. First, the value of blinding in NI trial is under debate, especially if the endpoints are subjective.[6] In a superiority trial, a blinded investigator who has a preliminary belief in superiority of the test drug cannot manipulate the results to support his belief. On the contrary, in an NI trial, the blinded investigator with a preliminary belief in non-inferiority of the test drug can bias the result by assigning similar ratings to the treatment responses of all patients. Others argued that blinding is still important to show the differences between drugs in NI trials.[7] Second, there are different methods to determine the NI margin and there are debates on whether the margin should be determined based on statistical or clinical considerations. Third, although there is a degree of consensus that non-inferiority should be shown for both the intention-to-treat (ITT) and per-protocol (PP) analysis sets, it is not clear whether

this will be conservative or anti-conservative in a particular situation.[6,7] Fourth, a difficulty in interpreting NI trials is their lack of ability to distinguish an effective drug from an ineffective drug i.e. assay sensitivity[7,8], without relying on evidence outside the trial. A drug is considered effective if it shows a significant treatment effect compared with placebo. An additional placebo arm is recommended to confirm assay sensitivity[2,6,9]. However, this is often impossible due to ethical reasons. Last, the validity of the historical data that was used as the reference for the current trial, i.e. constancy assumption, is a critical point in the interpretation of NI trials. Related to the last issue, the CONSORT statement has recommended authors to mention whether the eligibility criteria, interventions and outcomes are identical or very similar to any trial that established efficacy of the reference treatment.[10] The effort is encouraged to support the validity of the constancy assumption.

The ICH E9 [11], the ICH E10 [8], CHMP guidelines [12] and the extension of the CONSORT statement on NI trials[10] are the currently available guidelines for the appropriate conduct and report of NI trials. We summarized the guidelines' recommendations on the five methodological issues described above in Table 1. Furthermore, we included the FDA draft guideline on NI trials[13] in Table 1 for consideration. The draft FDA guideline is not in effect yet and still open for changes (as per 18th March 2010).

In this review, we described how published NI trials were conducted and reported, and whether the standard requirements from the guidelines were followed.

Methods

Search strategy and publication selection

We performed searches for NI trials in PubMed on 5 February 2009 and retrieved 669 articles as described in Figure 1. Subsequently, based on pragmatic consideration rather than formal sample size calculation, we used SPSS 16 to select a random sample of 300 articles. We subsequently excluded study design papers, reviews, trials using healthy volunteers, non-drug trials, non-RCTs, and articles of which the full-text version could not

be retrieved. If one article reported multiple trials, we analyzed the trials separately. If multiple articles reported the result of one trial, we considered them as one subject, and included only the first publication.

Data extraction

To extract relevant data, we created a standardized data extraction form, accompanied by an operational definition of each extracted variable. GW extracted all articles and MK extracted a randomly chosen 10 % of the articles. GW and MK then compared the extraction results from the 10 % articles. Disagreements occurred in seven articles and in three of 38 variables. The cause of the disagreements was the interpretation on vague information listed in the articles. We then decided that only a literal extraction was allowed, thus disallowing interpretation during extraction. For example for the degree of blinding, if only the description on how the investigator did the blinding but no clear terms e.g. double blind were written in the articles, we categorized it as 'ambiguously stated'. We then updated the operational definition accordingly and GW rechecked the

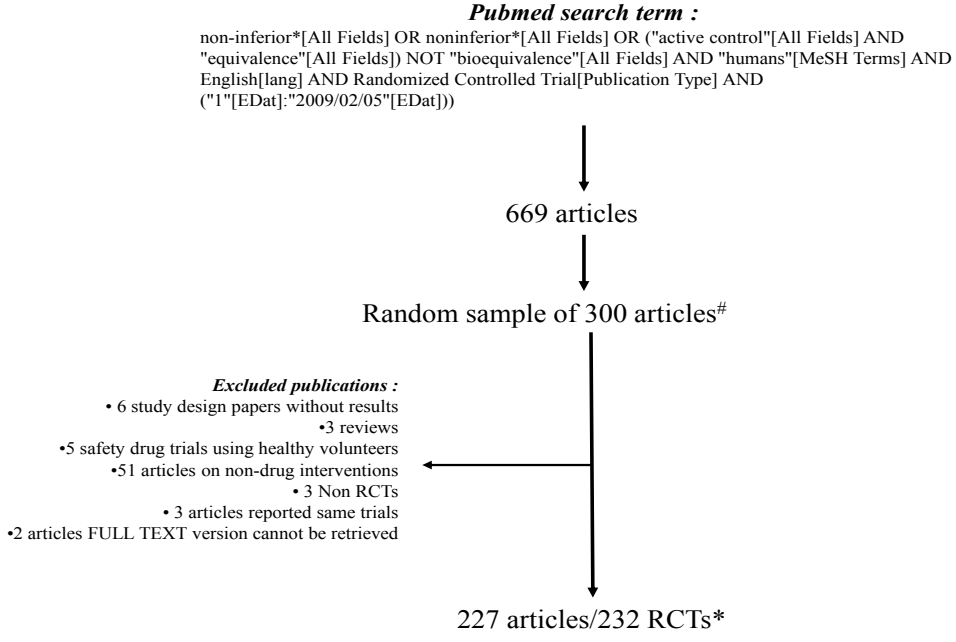


Figure 1. Flow diagram of trial selection process

Sample size was chosen based on pragmatic consideration rather than formal calculation

* 1 publication may contain more than 1 RCT

extraction results of those three variables in all the articles again and if necessary revised them.

For any missing information, if the articles referred to a registration database or previous paper for full description of the methods, information from these sources was retrieved.

Characteristics of the trials

From each article, we extracted information on the journals' impact factor, type of drug, phase of the trial, trial's sponsor (independent investigator, pharmaceutical industry, or government), trial's design, primary endpoints, sample size, and the trial's conclusion of the new drug.

In addition, we extracted specific information whether the authors mentioned any additional benefit of the new drug and whether the additional benefit was addressed in the trial. For example, if the author mentioned that the additional benefit of the new drug was its better safety profile, we evaluated whether any formal safety profile comparison was included in the results section of the article. We classified the journals based on their impact factor listed in the Journal Citation Reports® (JCR) 2008 edition. We arbitrarily chose a cut-off point of ten to classify the journal as high or low-impact.

We extracted the phase of the trial according to the statement in the publications or the referred clinical-trial's database e.g. clinicaltrials.gov. The classification was Phase I, II, III and IV. Phase II and III might be divided into 2 parts, A and B. Phase IIA's primary aims are assessment and exploration of efficacy and pharmacodynamic aspects of the drug in patients with the target disease. In phase IIB, the main objectives are to confirm efficacy in a relatively large group of patients and determine optimal dose and dosing regimen to be implemented in phase III trials. In phase III trials, the main objectives are to confirm and to gather the additional information about the effectiveness and safety of the drug that are needed to evaluate the overall benefit-risk profile of the drug. Phase IIIA is conducted prior to application for marketing authorization, and phase IIIB is conducted after application.[14, 15]

Table 1. The requirements in the guidelines for conducting and reporting NI trials

Issues in NI trials	Requirements in the guidelines
Blinding method	. Blinding is necessary to minimize bias (ICH E9 and E10)
	. It is critical to provide reassurance and procedures that ensure maintenance of blinding (draft FDA guideline on NI trial 2010)
NI margin	. An acceptable non-inferiority margin should be defined (ICH E10, CPMP/EMEA 2000)
	. Should be pre-specified, and can be no larger than the presumed entire effect of the active control in the NI trial (draft FDA guideline on NI trial 2010)
	. Should be specified in publication (CONSORT statement extension, 2006)
Method to determine NI margin	. The determination of the margin in a non-inferiority trial is based on both statistical reasoning and clinical judgment (ICH E10)
	. Margin is chosen by defining the largest difference that is clinically acceptable, so that a difference bigger than this would matter in practice (CPMP/EMEA 2000)
	. The NI margin should be generally identified based on previous experience in placebo-controlled trials of adequate design under conditions similar to those planned for the current trial, but could also be supported by dose response or active control superiority studies.(ICH E10, CHMP/EMEA 2005)
	. Fixed margin method (two CIs method) is recommended. It is referred to as fixed because the past studies comparing the drug with placebo are used to derive a single fixed value for statistical margin, even though this value is based on results of placebo-controlled trials (one or multiple trials versus placebo) that have a point estimate and confidence interval for the comparison with placebo. This approach is relatively conservative, as it keeps separate the variability of estimates of the treatment effect in the historical studies and the variability observed in the NI trial, and uses a fixed value for the estimate of the control effect based on historical data (the 90% or 95% CI lower bound), a relatively conservative estimate of the control drug effect. (draft FDA guideline on NI trial 2010)
	. should be specified in publication (CONSORT statement extension,
Similarity with trial of reference treatment	. The report should contain whether the eligibility criteria, interventions and outcomes are identical (or very similar) to that of any trial that established efficacy of the reference treatment (CONSORT statement extension, 2006)
Type of statistical analysis	. Use of the full analysis set is generally not conservative and its role should be considered very careful (ICH E9)
	. Both ITT and PP have equal importance (CPMP/EMEA 2000)
	. Important to conduct both ITT and as-treated analyses. Differences in results using the two analyses will need close examination. (draft FDA guideline on NI trial 2010)
Assay sensitivity	. A trial should have the ability to distinguish an effective from an ineffective drug (ICH 10, draft FDA guideline on NI trial 2010)
	. A three-armed trial with test, reference and placebo allows some within-trial validation of the choice of non-inferiority margin and should be used wherever possible.(CHMP/EMEA 2005, draft FDA guideline on NI trial 2010)
Constancy assumption	. The similarity of the new trial to the historical trial should be sufficient (CHMP/EMEA 2005, draft FDA guideline on NI trial 2010)

We classified the type of primary endpoints as hard endpoints, intermediate endpoints and subjective endpoints. Hard endpoints were direct clinical events, such as mortality or stroke; intermediate endpoints were indirect outcome measurements that might not necessarily have a direct relationship with the clinical event such as laboratory data or biomarkers; and subjective endpoints are endpoints based on subjective perspectives of investigator or patient, such as quality-of-life questionnaires.

We extracted from the article specific characteristics of NI trials: degree of blinding, the method to determine the NI margin, the type of analysis, the use of a placebo arm to confirm assay sensitivity, and whether the authors discussed the constancy assumption. Furthermore, we extracted reasons for not including a placebo arm.

In terms of blinding, we extracted the literal term reported by the authors in the manuscript and classified the blinding into open-label, single, double, triple and "ambiguously stated" blinding.

Since there are guidelines on the NI margin for anti-infective drugs, we assessed within these trials whether their NI margin was consistent with them. Based on a guideline of the FDA (1992) and CPMP (1997), the recommended NI margin for anti-infective drugs is percentage difference of 10 - 20 %.

We analyzed the quality of conducting NI trials by comparing the design and analysis characteristics of the trials reported in the high-impact vs. low-impact journals; and between the trials that were sponsored by industry and non-industry.

Quality of reporting

To evaluate the quality of reporting, we compared the requirements from the extension of the CONSORT statement for NI and equivalence trials[10] between articles published before and after 2006 to evaluate the impact of the CONSORT statement extension on the reporting of NI trials. According to the extension of the CONSORT statement for NI trials, the method section should include additional information on how identical the inclusion and exclusion criteria, type of interventions and outcomes to previous efficacy trial of the active comparator were. The additional information should also include the NI margin and the method to determine it, sample size calculation, and

whether a one-sided or two-sided confidence interval (CI) was used. The side of the CI is important in an NI trial as its inference of non-inferiority is based on the CI of the treatment difference between the test drug and its comparator. NI is concluded when the CI excludes and lays beyond the NI margin.[11] Furthermore, we compared the compliance to the CONSORT statement extension's requirements between trials reported in the high-impact and low-impact journals; and between the trials that were sponsored by industry and non-industry.

Data analysis

Data were entered into a database using Epidata 3.1 (EpiData Association, Odense, Denmark; www.epidata.dk) and all statistical analyses were performed using SPSS 16 (SPSS Inc, USA; www.spss.com). The p-values for the differences were calculated using the Chi-square or Fischer's Exact test.

Results

The selection process of the NI trials is outlined in Figure 1. After filtering the articles based on the exclusion criteria, we included 227 articles in the analysis, which referred to 32 trials. One hundred eleven (47.8 %) trials were published after 2006, the year in which the extension of the CONSORT statement on NI trials was published.

The missing data we retrieved from the registry were mostly data on the trial's phases and sponsorship. We only referred to the database as suggested by the author, so we believe the data in the register were reliable. We retrieved data of 34 trial's phases from clinicaltrial.gov; data of one trial's phase and data of one trial's sponsor from ISCRTN; data of one trial's sponsor from WHO international clinical trial registry; and data of one trial's phase from a sponsor clinical-trial registry website.

Table 2. The general characteristics of trials

	N (%) unless stated otherwise
Published in high-impact factor journals	46 (19.8)
Type of drug	
Anti infective drugs	53 (22.9)
Cardiovascular and thrombolytic drugs	40 (17.2)
Drugs for endocrine disorders	26 (11.2)
Vaccines	24 (10.4)
Anti inflammatory and anti rheumatics drugs	17 (7.3)
Respiratory drugs	16 (6.9)
Neurological and psychiatric drugs	14 (6.0)
Anticancer drugs	11 (4.7)
Others	31 (13.4)
Phase	
Phase II	7 (3.0)
Phase III	69 (29.7)
Phase IV	12 (5.2)
Phase IIIB and IV	3 (1.3)
Not stated	141 (60.8)
Sponsor	
Independent investigator	39 (16.8)
Pharmaceutical industry	171 (73.7)
Government	6 (2.6)
Combination of any above	2 (0.9)
Not clear	14 (6.0)
Design	
Parallel	216 (93.1)
Cross-over	13 (5.6)
Factorial	2 (0.9)
Cluster-randomized	1 (0.4)
Primary endpoints	
Hard endpoints	97 (41.8)
Intermediate endpoints	102 (44.0)
Subjective endpoints	33 (14.2)
Sample size (median, interquartiles range)	
Number of planned subjects	388(242 – 673)
Number of subjects in ITT analysis divided by number of subjects planned	1.1 (1 - 1.2)
Number of subjects in PP analysis divided by number of subjects planned	1.0 (0.8 – 1.1)
Conclusion	
Non-inferiority was shown	209 (90.1)
Non-inferiority was not shown	17 (7.3)
Not clear	6 (2.6)

The general characteristics of the trials

The general characteristics of the trials are described in Table 2. Most of the trials were published in low-impact journals (84.5 %). Anti-infective drugs were the most studied drugs (22.9 %).

Almost one-third (29.7 %) of the studies were phase III studies and the majority had pharmaceutical industry involvement in their trial process (73.7 %).

Almost all studies had a parallel design (93.1 %), and both hard and intermediate endpoints were often investigated. Variability between studies in the ratio of number of subjects in the analysis population versus the planned number of subjects was considerable. Most of the trials concluded that the new drug was shown to be non-inferior compared with its comparator (209 trials – 90.1 %).

In 124 trials (53.4 %), the authors mentioned additional advantages of the new drug. Most of the additional benefits mentioned and addressed were in terms of the safety profile of the drug, as shown in Table 3.

The quality of conducting NI trials

The design and analysis characteristics of the trials are described in Table 4, while stratification according to journal impact factors is shown in Table 5. Six journals did not have their impact factor listed in the JCR 2008 edition and were not included in the analysis. We found no significant difference in terms of trials' characteristics between trials that were sponsored by pharmaceutical industry or not (data not shown).

More than half of the trials were stated as double blinded, while a substantial number (79, 34.0 %) was open label. We found no difference in terms of blinding method between trials that were published in high-impact or low-impact journals.

We observed that 227 (97.8 %) trials reported their NI margin in the articles. Nevertheless, only 106 (45.7 %) trials reported the method by which the NI margin was determined. In 51 (22 %) trials, the margin was determined merely based on investigator's assumption. In 20 (8.7 %) trials, the NI margins were obtained from other publications or reviews. In 18 (7.7 %) trials, the NI margins were obtained from guidelines and in 17 (7.3 %) trials the NI margins were calculated by the investigators based on data from previous

Table 3. Design and analysis characteristics

	N (%)
Blinding method	
Open-label	79 (34.0)
Single	18 (7.8)
Double	125 (53.9)
Triple	1 (0.4)
Ambiguously stated	9 (3.9)
Method to determine NI margin	
Based on investigator's assumption	51 (22.0)
Based on other publications or reviews	20 (8.7)
Based on guidelines	18 (7.7)
Calculated by the investigator based on previous trial's result	17 (7.3)
Not clear	126 (54.3)
Type of statistical analysis	
Both ITT and PP	97 (41.8)
Only ITT	81 (34.9)
Only PP	46 (19.8)
Not clear	8 (3.5)
Including placebo-arm to confirm assay sensitivity	14 (6.0)
Discuss constancy assumption	9 (3.9)

trials. Among the last, 15 of them used a preserved fraction of 50% or greater. We also found in 95 (40.9 %) trials, the authors mentioned that the NI margin was a clinical acceptable margin. Among them three trials mentioned that the decision to use the margin was validated by a panel of clinical experts. We found no difference in terms of method to determine the NI margin between trials that were published in high-impact or low-impact factor journals.

Within 53 anti-infective trials, most of the trials (42, 77.8 % of all anti-infective trials) used an NI margin of percentage difference between 10 to 20 %. Only four trials used a NI margin less than 10 % or more than 20 %. In the rest of seven trials, six trials did not use percentage difference as an NI margin, and in one trial, the NI margin was not clear.

In terms of statistical analysis, most of the trials (127, 54.7%) used either ITT or PP, while 97 (41.8 %) trials used both ITT and PP analysis. We found among the trials that used both ITT and PP analysis, 94 of them concluded that the new drug was non-inferior to its comparator. In 53 trials of the latest, the conclusions were deducted from similar results of

both ITT and PP analysis. In the rest of the trials: 22 trials concluded non-inferiority based on the results of their PP analysis; 18 trials were based only on the results of their ITT analysis; while in three trials, it was not clear on which analysis their conclusion was based. We found a significant difference in terms of type of statistical analysis between the trials published in high-impact and low-impact factor journals. Trials published in the high-impact journals mostly used only ITT analysis (54.3 % of 46 trials), while in the low-impact journals, both analysis methods were most frequently used (44.4 % of 180 trials).

In our review, we observed that 210 trials (90.5 %) did not include a placebo arm to confirm assay sensitivity. Only 19 trials mentioned the reason why a placebo arm was not included in trials, and almost half of them were due to ethical reasons. We observed that the inclusion of a placebo is quite common (28.6 %) in trials with neurology/ psychiatric drugs. This is probably because in this type of drugs, the constancy assumption will often not hold, as the placebo effect in previous placebo-controlled trials is difficult to rule out**.** In addition, we found no difference in terms of using a placebo arm to confirm the

Table 4. Additional benefit of the new drug mentioned in the publication

	N, % *	Addressed¶ (% from N)
Better safety profile	45 (36.3)	43 (95.6)
Better method of administration	19 (15.4)	6 (31.6)
Better safety profile and method of administration	12 (9.7)	10 (83.3) †
Better method of administration and induce higher patient's compliance rate	12 (9.7)	5 (41.7) ‡
Better safety profile and induce higher patient's compliance rate	7 (5.6)	5 (71.4) §
Induce higher patient's compliance rate	6 (4.8)	3 (50.0)
Better method of administration and low cost	5 (4.0)	2 (40.0) ¥
Others than above	18 (14.5)	10 (55.6)

***Note**:*

* Percentage is based on 124 trials that mentioned any additional benefit of the new drugs irrespective of whether or not data were shown to support the claim

¶ The authors show any analysis or argument of the additional benefit

† Three trials addressed both the safety profile and the method of administration, six trials only addressed the safety profile, and one trial only addressed the method of administration

‡ One trial addressed both the method of administration and patient's compliance rate, and four trials addressed only the patient's compliance rate

§ Four trials addressed both safety profile and patient's compliance and one trial only addressed the safety profile

¥ One trial addressed both better method of administration and cost and one trial only addressed cost.

Table 5. Stratification of the articles according to their journal impact factors

Trials' design and analysis issues	High-impact (N =46)	Low-impact (N = 180)	p value
		N (%)	
Blinding method			0.11
Open-label	20 (43.5)	56 (31.1)	
Single	5 (10.9)	12 (6.7)	
Double	18 (39.1)	105 (58.3)	
Triple	0 (0)	1 (0.6)	
Ambiguously stated	3 (6.5)	6 (3.3)	
Method to determine NI margin			0.34
Based on investigator's assumption	16 (34.8)	35 (19.4)	
Based on other publications or reviews	5 (10.9)	14 (7.8)	
Based on guidelines	2 (4.3)	16 (8.9)	
Calculated by the investigator based on previous trial's result	3 (6.5)	14 (7.8)	
Not clear	20 (43.5)	101(56.1)	
Type of statistical analysis			*0.01*
Both ITT and PP	14 (30.4)	80 (44.4)	
Only ITT	25 (54.3)	56 (31.1)	
Only PP	6 (13.2)	39 (21.7)	
Not clear	1 (2.1)	5 (2.8)	
Including placebo-arm to confirm assay sensitivity	2 (4.3)	12 (6.7)	0.74
Discuss constancy assumption	1 (4.3)	7 (8.9)	0.65

assay sensitivity between trials that were published in high-impact or low-impact factor journals.

Additionally, we observed only nine (3.9 %) authors discussed the constancy assumption and there was no difference in this respect between trials that were published in high-impact or low-impact journals.

Compliance in reporting NI trials

Only 3.0 % of the trials reported the similarity of the inclusion and exclusion criteria with previous trials studying the effect of the active comparator, 5.6 % of the trials reported the similarity of the type of intervention with previous trials, and 3.4 % of the trials reported the similarity of the outcomes. Seventy-seven (33.2 %) trials did not report whether they were going to present the data using one-sided or two-sided CI in the methods section as

Table 6. Comparison of reporting of essential information in NI trials

Reported in the method section of a NI article	Between High-impact journals and low-impact journals			Before and after CONSORT statement in 2006		
	Percentage of trials			Percentage of trials		
	High-impact (N = 46)	Low-impact (N=180)	p-value	Before and until 2006 (n=121)	After 2006 (n=111)	p-value
Eligibility similarity	2.2 %	2.8 %	1.00	2.5 %	3.6 %	1.00
Type of Intervention similarity	10.9 %	3.9 %	0.07	5.8 %	5.4 %	0.90
Outcomes similarity	4.3 %	2.8 %	0.63	2.5 %	4.5 %	1.00
NI margin	97.8 %	98.3 %	1.00	97.5 %	98.2 %	1.00
Method to determine NI margin	56.5 %	43.9 %	0.12	50.4 %	40.5 %	0.15
Side of CI	50.0 %	71.7 %	< 0.01	64.5 %	69.4 %	0.26

required by the CONSORT statement. Furthermore, we found that the papers in low-impact journals reported the side of the CI more frequently than those in the high-impact factor journals, and the difference was significant (Table 5).

The compliance in reporting the items required by the extension of the CONSORT statement before and after 2006 is described in Table 6. We did not observe improvement of reporting after the release of the CONSORT statement extension for NI trials. The method of determination of the NI margin was even reported less frequently in trials published after 2006 than in trials published before and in 2006.

Discussion

In this review, we found five main issues in the design, analysis and reporting of NI trials. First, many of the trials were open label trials. Second, reporting the method to determine the NI margin was infrequent and limited. Third, most of the trials analyzed their data with one statistical analysis method; ITT or PP. Fourth, we observed that only few trials included placebo-arm to confirm assay sensitivity and that only few trials discussed the constancy assumption. Lastly, we did not observe any difference in terms of

reporting in NI trials published before or after the release of the extension of the CONSORT statement for NI trials in 2006.

In our review, about a third of the trials were open label trials. This surprising finding was not consistent with the guidelines[8,11] that suggest to use blinding whenever possible to minimize the risk of bias. This leads to discussion on the importance of blinding in an NI trial. Snappin believes that blinding only gives minor protection in NI trials, since a blinded investigator with a preliminary belief in non-inferiority of the test drug can bias the result by assigning similar ratings to the treatment responses of all patients.[16] There is no doubt, however, that blinding does offer protection against information bias. In addition, there will usually be endpoints (e.g. safety) for which differences are expected and for which blinding will ensure stronger evidence. We therefore conclude that blinding is still important in NI trials to avoid bias**.** If blinding is not possible, subjective endpoints need to be avoided and more stringent monitoring should be conducted.

The method to determine the NI margin was not reported in more than half of the trials. This finding is consistent with previous reviews in 2005 to 2006, where the methods were presented in 46 % or less of the trials.[17,18,19] Apparently, the extension of the CONSORT statement in 2006 has not brought any significant impact yet. Furthermore, the statement has suggested that the NI margin should be preferably justified on clinical grounds and its relation to the effect of the reference treatment relative to placebo in any previous trials should be noted.[10] We found that most of the authors included a statement that the NI margin was a clinically acceptable difference, but only three trials mentioned that the margin was validated by a panel of clinical experts. This finding was consistent with other reviews[17,18,19], where many trials claimed that their margin was clinically relevant without any clear details how the clinically acceptable NI margin was chosen. Putting merely a statement that the margin was determined based on clinically acceptable difference is not sufficient for any subsequent trial replications. Thus, more details are needed in the description on how the NI margin was determined. Furthermore, a detailed description on how the margin was determined can help the reader to decide whether the NI margin and the rationale for the margin's choice influenced the validity of the results.

We observed in anti-infective drug trials, that most of them used a constant difference of 10-20 % in treatment difference as their NI margin. Regulators recommend an NI margin of 10 % for vaccines and anti-bacterials.[20,21] This margin of 10 % is acceptable as long as the primary outcome of interest has a high incidence rate. The implication of using a 10 % constant margin in vaccines and anti-infective drugs should be further explored and any improvement on the guidelines to determine NI margin should cover this issue.

We observed that most of the trials reported the result only from ITT analysis or PP analysis. Our results were consistent with a previous review that observed that more NI trials used ITT rather than PP.[18] We also observed that ITT analysis was more reported in high-impact journals. The CPMP guidelines and the new draft FDA guidelines for NI trials already stated that both analyses have equal importance in NI trials. For superiority trials, ITT analysis is the preferred analysis as it adheres to randomization[11] and might best reflect clinical practice. PP analysis might violate randomization and not reflect clinical practice very well. Several reviews with RCT simulation showed that both ITT and PP could be problematic in NI trials, especially if the trial had large number of non-compliance. [22,23,24] In addition, in our data, we did not observe any evidence that ITT will lead to more NI conclusions than PP. We conclude that both analyses are equally important, as each approach brings a different interpretation for the drug in daily practice.

We observed that only a small number of trials included placebo arms to support assay sensitivity. Although our data did not provide sufficient evidence whether the use of placebo was appropriate or not in the trials, we believe that the use of a placebo arm was probably not ethically feasible in most studies. Nonetheless, the non-inferiority result of the drugs in NI trials might bear two meanings: both drugs are equally effective, or both drugs are equally ineffective against placebo. In this sense, a placebo arm in an NI trial will enable evaluation whether both drugs in the trial are effective, if the trial shows non-inferiority. Alternatively, if the use of a placebo arm is not possible, the trial should choose a margin that assures that the estimated effect of the new drug is likely to be superior to placebo, under the constancy assumption for the active comparator. The readers, not only the investigators, also need to be aware of this issue of assay sensitivity in interpreting the

result of NI trials. They need to consider the type of endpoints; the number of patients in the final analysis; reasons of patient's dropouts; the similarity of the trial with the previous trial(s) that established the efficacy profile of the comparator; and the constancy assumption of the data used as reference for the NI margin. Based on our review, two of the latter were only being reported in a small numbers of the articles.

Less than five percent of the trials in our review mentioned whether the trials were designed similar to relevant past trial(s). Thus, it was difficult to assess whether the historical data that were used for determining the NI margin were reliable. Since the validity of the NI margin is related to the interpretation of the NI trials, clear reporting of the method of NI margin determination and the constancy assumption is essential for every NI trial publication. It is impossible to check the validity of the constancy assumption without a parallel placebo arm. However, at minimum, it is possible to check whether the current NI trial was similar to previous trial(s) that estimated the efficacy of the active comparator.[25]

We found no difference between reporting before and after the release of the extension of the CONSORT statement on NI trials. Furthermore, in general, there is no difference in adherence to the CONSORT statement between the high-impact and the low-impact journals. The overall low adherence to the statement might be due to unfamiliarity of the authors, referees, and editors of all of the journals with the statement extension. Researchers and editors of journals should be more aware of this extension and should comply with its recommendations. We realized that it might be too early to see full adherence of the CONSORT statement extension after 3 years, but due to the reputation of the CONSORT statement itself, we considered it reasonable to expect a certain degree of improvement.

Our review has some limitations. First, we excluded several trials since we only used a random sample of all NI trials that we identified. However, as this was a random sample, this will not have influences our results. Second, we only used PubMed to identify NI trials; therefore, we might have missed some trials. However, we assume that NI trials retrieved from PubMed do not have different methodological characteristics than NI trials in other databases, so we do not think that this influenced our results. Third, since the

terms that we used to search for NI trials were not standard MESH terms and our search for those terms was limited to the abstract of the articles, our search might not have captured all NI drug trials available in PubMed. Also for this selection, we expect that the NI trials that we found are not different from the NI trials that we did not capture with our search. A strength of our study is that we did not only focus on the NI margin, as previous reviews [17,18,19] did, but also evaluated other methodological aspects of NI trials. In addition, we evaluated the quality of reporting using the current guidelines from the CONSORT statement.

In conclusion, the conduct and reporting of NI trials can be further improved. Particularly, in terms of maximizing the use of blinding, the use of both ITT and PP analysis, reporting the similarity with the previous comparator's trials to guarantee a valid constancy assumption and reporting the method to determine NI margin.

References:

(1) 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 Feb;47(1):28-36; 99-107.

(2) 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 Jan 30;22 (2):169-86.

(3) Regulatory watch: Non-inferiority-trial discussions impact new drug applications. Nat Rev Drug Discov 2009 Jan;8(1):10.

(4) Garattini S, Bertele V. Non-inferiority trials are unethical because they disregard patients' interests. The Lancet 2007;370(9602):1875-77.

(5) Jones B, Jarvis P, Lewis JA, Ebbutt AF. Trials to assess equivalence: the importance of rigorous methods. BMJ 1996 Jul 6;313(7048):36-9.

(6) Senn S. Active control equivalence studies. Statistical Issues in Drug Development, 2nd Edition Glasgow, UK: Wiley; 2007. p. 235-47.

(7) Kay R. Equivalence and non-inferiority. In: Kay R, editor. Statistical thinking for non-statisticians in drug regulation: Wiley, John & Sons, Incorporated; 2007. p. 173-91.

(8) ICH Expert Working Group. (2001) Choice of control group and related issues in clinical trials. E-10. Available at : http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E10/Step4/E10_Guideline.pdf

(9) Koch A, Rohmel J. Hypothesis testing in the "gold standard" design for proving the efficacy of an experimental treatment relative to placebo and a reference. J Biopharm Stat 2004 May;14(2):315-25.

(10)Piaggio G. Reporting of Noninferiority and Equivalence Randomized Trials: An Extension of the CONSORT Statement JAMA: The Journal of the American Medical Association 2006;295(10):1152-60.

(11) ICH Expert Working Group. (1998) Stastical principles for clinical trials. E9 . Available at : http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E9/Step4/E9_Guideline.pdf

(12) Committee for Medicinal Products for Human Use (CHMP). (2005) Guideline on the choice of the non-inferiority margin. EMEA/CPMP/EWP/2158/99 . Available at : http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003636.pdf

(13) Center for Drug Evaluation and Research (CDER) Center for Biologics Evaluation and Research (CBER). (2010) DRAFT GUIDANCE : Guidance for industry non-inferiority clinical trials . Available at : http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM202140.pdf

(14) Rados C. Inside Clinical Trials: Testing Medical Products in People. 2003;37.

International Conference on Harmonisation. General considerations for clinical trials. E8 1997. Available at : http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E8/Step4/E8_Guideline.pdf

(16) Snapinn SM. Noninferiority trials. Curr Control Trials Cardiovasc Med 2000;1(1):19-21.

(17) Lange S, Freitag G. Choice of delta: requirements and reality--results of a systematic review. Biom J 2005 Feb;47(1):12-27; discussion 99-107.

(18) Le Henanff A, Giraudeau B, Baron G, Ravaud P. Quality of reporting of noninferiority and equivalence randomized trials. JAMA 2006 Mar 8;295(10):1147-51.

(19) McCaffrey N, Merlin T HJ. Lack of clinical rationale provided for non-inferiority margins. -Poster presentation at 3rd Annual Health Technology Assessment international (HTAi) conference. July 3-5, 2006. Adelaide, South Australia. 2006.

(20) Committee for Proprietary Medicinal Products (CPMP). Note for guidance on evaluation of medicinal products indicated for treatment of bacterial infections. CPMP/ EWP/558/95 Rev 1 2004.

(21) Committee for Human Medicinal Products (CHMP). Note for guidance on the clinical evaluation of vaccines. 2005. Available at : http://www.ema.europa.eu/docs/en_GB/ document_library/Scientific_guideline/2009/09/WC500003875.pdf

(22) Brittain E, Lin D. A comparison of intent-to-treat and per-protocol results in antibiotic non-inferiority trials. Stat Med 2005 Jan 15;24(1):1-10.

(23) Matilde Sanchez M, Chen X. Choosing the analysis population in non-inferiority studies: per protocol or intent-to-treat. Stat Med 2006 Apr 15;25(7):1169-81.

(24) Sheng D, Kim MY. The effects of non-compliance on intent-to-treat analysis of equivalence trials. Stat Med 2006 Apr 15;25(7):1183-99.

(25) Fleming TR. Current issues in non-inferiority trials. Stat Med 2008;27(3):317-32.

The challenges of determining non-inferiority margins:

a case study of non-inferiority randomized controlled trials of novel oral

Grace Wangge
Kit CB Roes
Anthonius de Boer
Arno W. Hoes
Mirjam J. Knol

Modified from
CMAJ. 2012 Aug 20.

Introduction

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

We have seen a large increase in publications on NI 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 an NI trial is pre-specifying an NI margin: a threshold below which it can be established that the new drug is not worse than its comparator. The 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 (46%) trials reported the method they used to determine the NI margin and these methods varied considerably.[1] In 22 % of the trials the margin was determined merely based on investigator's own assumption, while in 8.6% of the 232 trials the margin was stated as an acceptable clinical difference according to the literature.[2] These observations are worrisome, as the choice of the NI margin determines the conclusion of the trial and, thus, clinical decision making.

Here, we explain the method to determine an NI margin, as outline in the draft US Food and Drug Administration (FDA) guideline on NI trials.[3] In addition, we present a case study on the NI margins used in trials on novel anticoagulants, drugs for which many NI trials are performed. The case study demonstrates substantial variability in the NI margins applied in those trials.

Key points

- The aim of a non-inferiority trial is to show that a new drug is not worse than its comparator.
- How a non-inferiority margin is chosen is often not explained; methods can be highly variable, resulting in inconsistent conclusions of non-inferiority.
- A non-inferiority 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.

Determining an NI margin

Most of the guidelines on NI trials [4-6] 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 issued drafted guideline on NI trials issued by the FDA.[3] The guideline was composed based on previous guidelines [4-6] and methodological publications on NI trials [7-10] published since the 1980s. The guideline is only one example of determining a NI 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 use of the fixed margin method or 95%-95% method, which is seen as the most straightforward and most readily understood approach. The method starts by identifying M1 and M2. M1 is the effect of the active control compared with placebo, assumed present in the NI trial. M1 is chosen as a conservative estimate (least effect size possible) of the effect of the active comparator, which is the upper-bound of the 95% 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

to the active control. For example, if it were concluded that it would be necessary for a test drug to preserve 75% of a mortality effect, M2 would be 25% of M1, the loss of effect that must be ruled out. Determination of M2 assures that the test drug will be superior to placebo.

Determining M1, as the first step in defining an NI margin, can be based on one or more placebo-controlled trials of the active comparator that have a design similar to the current NI 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 NI margin, meaning it is more difficult to conclude non-inferiority. The formula to calculate M2 for a risk difference (RD) is:

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

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:

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

The results of the NI trial are compared with the prespecified NI margin (M2) as follows: if the upper bound of the 95% CI for the effect estimate is smaller than the NI margin, non-inferiority is concluded. For example, if an NI trial shows that the RR of the new drug compared to the active comparator is 0.90 (95% CI: 0.68-1.20) and the NI margin is 1.25, it is concluded that the new drug is non-inferior to the active comparator.

Determining M2 is also related to how much of the treatment effect is judged necessary to preserve, 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

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

NI trials

Recently, new classes of anticoagulants, direct thrombin inhibitors (DTI) and direct inhibitors of factor Xa (DXAI) have been developed. These new drugs were claimed to be as effective as conventional therapies, such as heparins or low-molecular weight heparins (LMWH), but with a more convenient route of administration and no requirement for monitoring after discharge from hospital. DTI and DXAI were first registered for prevention

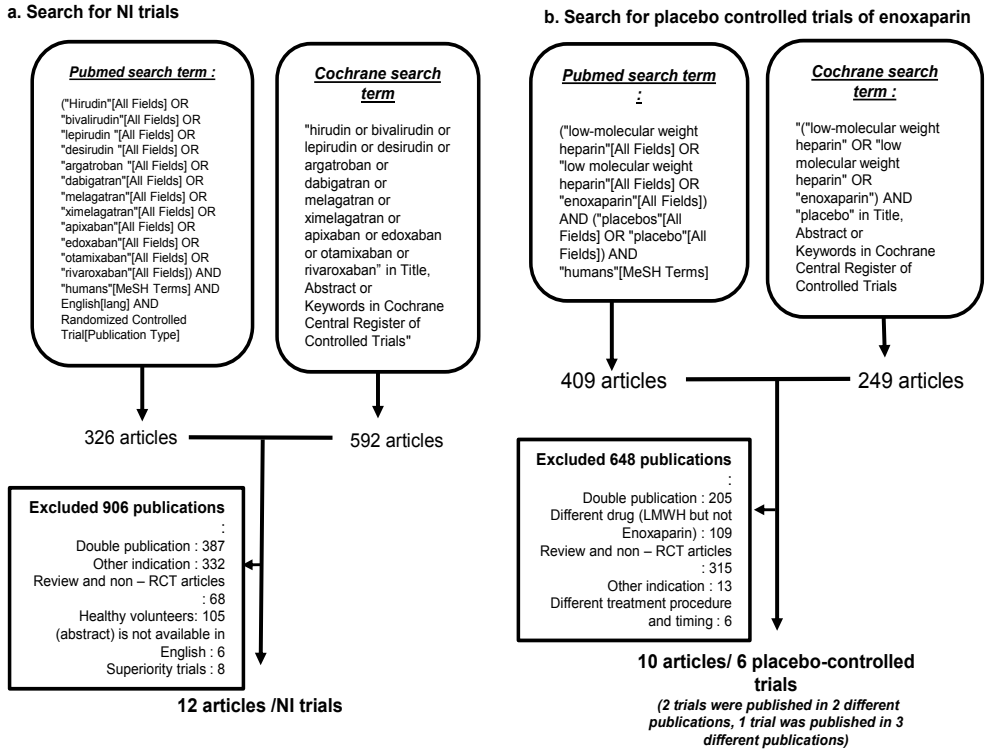


Figure. 1. Search strategy and publication selection for the case study. NI = non-inferiority.

Table 1a. Characteristics of the NI trials of oral anticoagulants

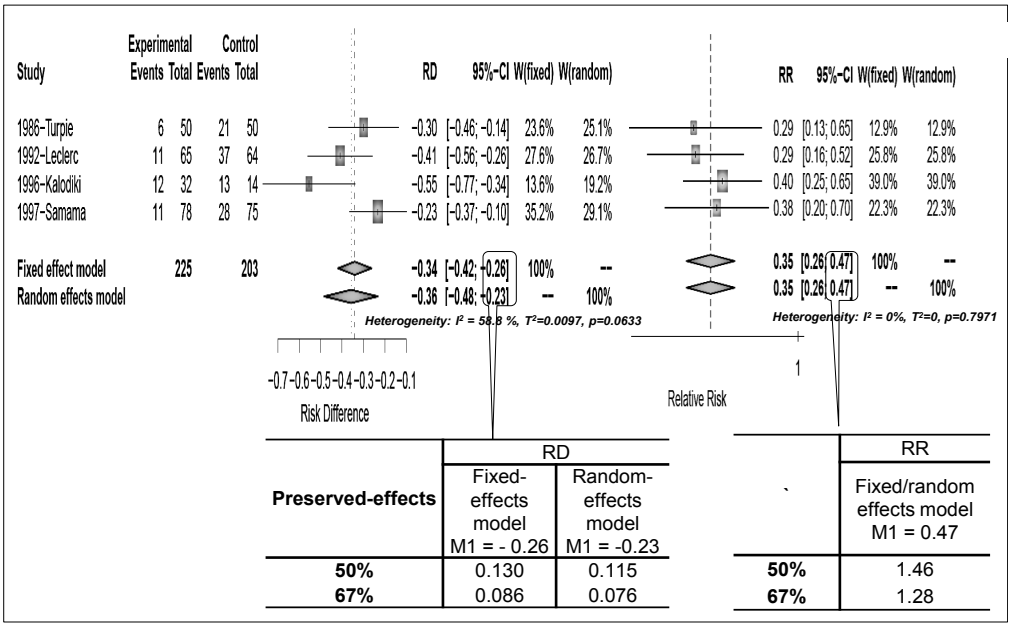
Trial/ Author	Name of drug	Duration of therapy and fol.up (days)	Dosage of enoxaparin	Primary endpoint	NI margin	Method to determine NI margin	Start of recruit-ment	Mean age	Female (%)	N
Colwell, et.al.[11]	Ximela-gatran	7-12	30 mg bid	Total VTE (symptomatic or venographic DVT or PE)	RD = 5%	- Independent clinical expert committee - Same NI margin as published NI trial on Tinzarapin vs Enoxaparin	Mar 2000	65	52	1838
EXPRESS[19]	Dabigatra n	8-12	40 mg qd	Major VTE (proximal DVT measured with venogra-phy, pulmonary embolism and/or death where PE could not be ruled out)	RD = 2%	Not described	Apr 2001	67	50	2835
REMODEL[12]	Dabigatra n	6-10	40 mg qd	Composite of total VTE (symptomatic or veno-graphic DVT or symptomatic PE) and all-cause mortality	RD = 9.2%	- 67% preserved-effect of difference between enoxaparin and placebo - Based on one published placebo controlled trial	Nov 2004	68	65	2183
RE MOBILIZE [13]	Dabigatra n	12-15	30 mg bid	Composite of total VTE (symptomatic or veno-graphic DVT or symptomatic PE) and all-cause mortality	RD = 9.2%	- 67% preserved-effect of difference between enoxaparin and placebo - Based on one published placebo controlled trial	Nov 2004	66	58	2615
Re-Novate[14]	Dabigatra n	28-35	40 mg qd	Composite of total VTE (symptomatic or veno-graphic DVT or symptomatic PE) and all-cause mortality	RD = 7.7%	- 67% preserved-effect of difference between enoxaparin and placebo - Based on pooled analysis of 3 placebo controlled trials	Dec 2004	64	56	3613
EXTEND[25]	Ximela-gatran	32-38	40 mg qd	Major VTE (proximal DVT as diagnosed at end-of-treatment, any clinically suspected and objectively confirmed DVT, measured by bilateral compres-sion ultrasound clinically suspected and objectively confirmed PE and VTE-related death or death where VTE-related causes could not be excluded)	RD = 2%	Not described	Sep 2005	65	54	1158
RECORD 1[26]	Rivaroxa-ban	35	40 mg qd	Composite of DVT measured with venography, nonfatal PE and all-cause mortality	RD = 3.5%	Not described	Feb 2006	63	56	4541
RECORD 3[27]	Rivaroxa-ban	10 - 14	40 mg qd	Composite of DVT measured with venography, nonfatal PE and all-cause mortality	RD = 4%	Not described	Feb 2006	68	68	2531
RECORD 4[28]	Rivaroxa-ban	11-15	30 mg bid	Composite of DVT measured with venography, nonfatal PE and all-cause mortality	RD = 4%	Not described	Jun 2006	65	65	3148
ADVANCE 1[20]	Apixaban	10 - 14	30 mg bid	Composite of DVT measured with venography, nonfatal PE and all-cause mortality	RD = 5.6% RR = 1.25	Not described	Nov-20061	66	61	3195
ADVANCE 3[29]	Apixaban	32-38	40 mg qd	Composite of DVT measured with venography, nonfatal PE and all-cause mortality	RR = 1.25	Not described	Mar 2007	61	53	5407
ADVANCE 2[30]	Apixaban	10 - 14	40 mg qd	Composite of DVT measured with venography, nonfatal PE and all-cause mortality	RD = 5.6% RR = 1.25	Not described	Jun 2007	67	72	3057

Table 1b. Placebo controlled trials of enoxaparin

Name of author	Date of publica-tion	Duration of therapy (days)	Dosage of enoxaparin	Primary endpoint	Mean age	Fe-male (%)	N
Turpie, et.al [15]	Oct-86	14	30 mg bid	DVT measured with venography	67	52	100
Leclerc, et.al [16]	Jan-92	14	30 mg bid	DVT measured with venography	69	60	131
Kalodiki, et.al [17]	Jun-96	8 - 12	40 mg qd	Composite of DVT measured with venography and PE	67	65	170
Samama, et.al [18]	Jan-97	10 ± 2	40 mg qd	Composite of DVT measured with venography and PE	69	42	93
Fuji, et.al (1) [24]	Jun-08	14	20 mg qd, 20 mg bid, 40 mg qd	Composite of DVT measured with venography and PE	62	88	419
Fuji, et.al (2) [24]	Jun-08	14	20 mg qd, 20 mg bid, 40 mg qd	Composite of DVT measured with venography and PE	70	84	364

of venous thromboembolism (VTE) in patients undergoing elective hip or knee replacement surgery. Many of these trials were NI trial. We found 12 such trials in PubMed and Cochrane -central register for controlled trials in May 2012 (fig 1). The characteristics of these trials are presented in table 1a.

All trials used enoxaparin, either 40 mg once daily or 30 mg twice daily, as the active comparator. Most trials used the risk difference to define the NI margin, and these margins ranged from 2.0 to 9.2%. Three trials used the relative risk to define the NI margin and in these trials the margin was 1.25. Only four of the 12 trials stated how they determined their NI margin. One trial stated that the NI margin was determined by an independent expert committee and it was the same NI margin that was used in a previous active-controlled trial of enoxaparin vs. tinzarapin.[11] Three trials used 67% preserved-effect of the (pooled) effect of one or three placebo controlled trials. [12-14]



Preserved-effects	RD Fixed-effects model M1 = - 0.26	RD Random-effects model M1 = -0.23	RR Fixed/random effects model M1 = 0.47
50%	0.130	0.115	1.46
67%	0.086	0.076	1.28

Fig. 2. Meta-analysis of placebo controlled trials of enoxaparin for risk difference (RD) and relative risk (RR); M1s and M2s.

Reference NI margin

We determined a reference NI 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 (Table 1b). The placebo-controlled trials were quite similar to the NI trials with respect to enoxaparin's dosage and duration patients' ages and gender distribution. However, death was not included as an outcome in the placebo-controlled trials, whereas most NI trials included all-cause mortality in their composite outcome. Because the NI trials in our case study started recruiting patients after 2000, we only included the 4 placebo -controlled trials [15-18] published before 2000 in the meta-analysis. We calculated the pooled RD and RR with 95% CIs using a fixed and random-effects model (fig 1). 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.

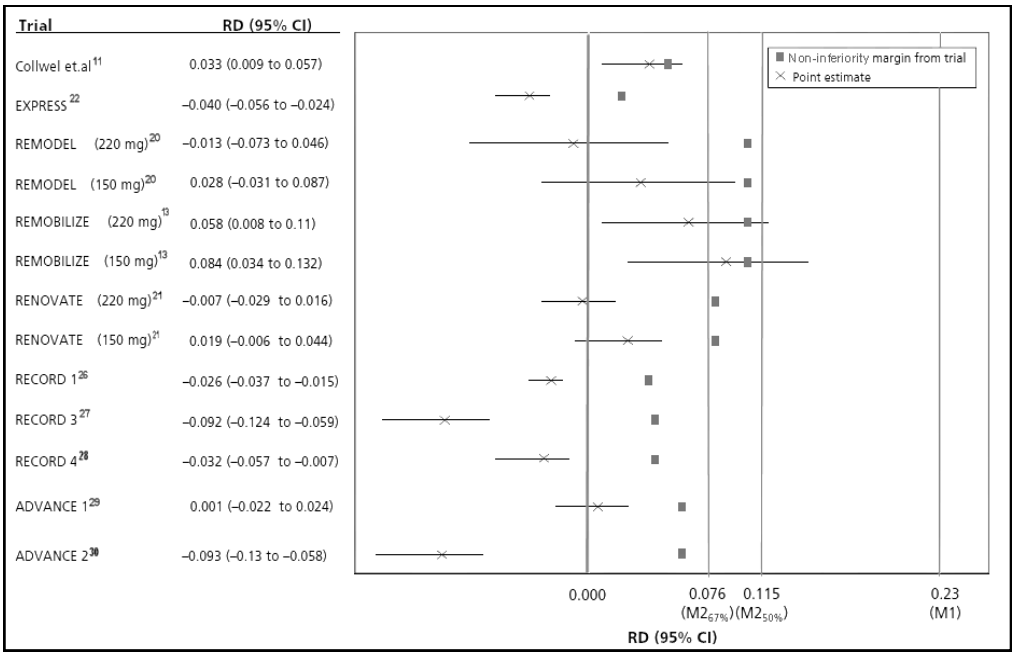


Fig 3a. Results of non-inferiority trials, non-inferiority margins of non-inferiority trials and 50% and 67% preserved-effects reference non-inferiority margin for risk difference

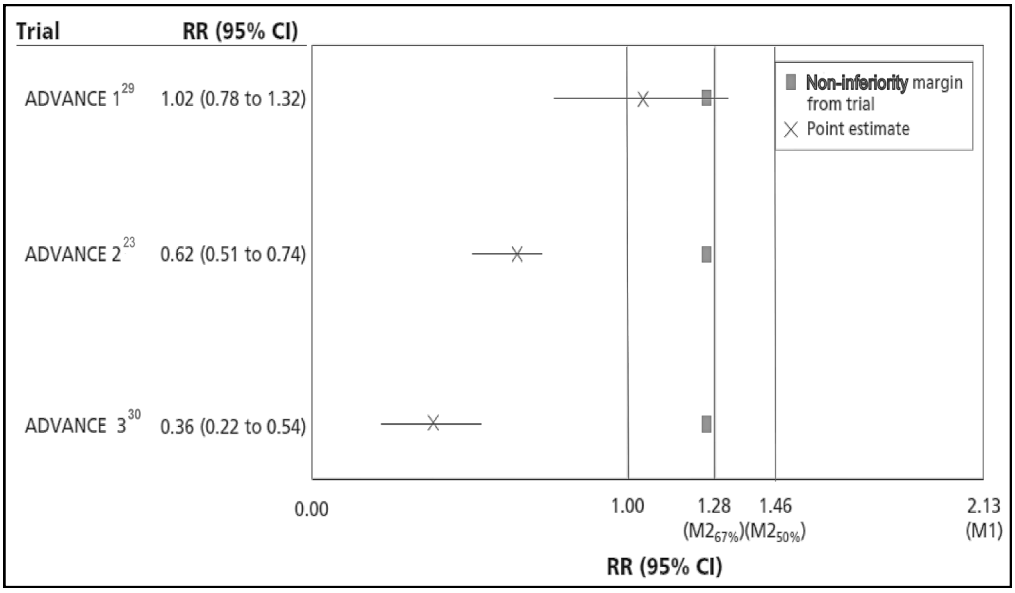


Fig 3b Results of non-inferiority trials, non-inferiority margins of non-inferiority trials and 50% and 67% preserved-effects reference non-inferiority margin for relative risk

Second, we calculated values for M2 using a 50% and 67% preserved-effect of M1 (fig 2). For example, calculating M2 with 50% preserved-effect for RD based on the fixed effects model resulted in the following calculation:

$$(1-0.5) * - (-0.26) = 0.130.$$

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

Comparison between reference and published NI margins

We plotted the point estimate and 95% CI of the NI trials, their NI margin and the reference NI margin to assess whether the conclusion of the trials would have been different had the reference NI margin been used (fig 3a for risk differences; fig 3b for relative risks). We did not include one of the trials in the figures [19] , because it was stopped early due to safety concerns and therefore lacked data on efficacy.

Fig 3a shows that the NI margins for the risk differences from the trials were stricter than the 50% preserved-effects reference NI margin (0.02 to 0.092 versus 0.115) and therefore the conclusion of non-inferiority in these trials does not change when using the reference NI margin, except for trial by Colwell and colleagues.[11] The NI margins in REMODEL[12], REMOBILIZE[13] and RENOVATE [14] were larger (i.e. less conservative) than the 67% preserved-effect reference NI margin (0.092 and 0.077 versus 0.076). In the REMODEL trial[12], dabigatran would not have been concluded as non-inferior to enoxaparin if the 67% preserved-effect reference NI margin had been used. Moreover, if the most conservative NI margin from EXPRESS[19] trial were used (0.02), the REMODEL [12] and RENOVATE[14] trials would not have concluded non-inferiority to enoxaparin.

Fig 3b shows that the NI margin in all trials was smaller (i.e. more conservative) than the 50% and 67% preserved-effect reference NI margin (1.25 versus 1.46 and 1.28). In the ADVANCE-1 trial [20], non-inferiority of apixaban was not concluded by the authors due to inconsistency between results for the RD and RR. If the 50% preserved-effect reference NI margin was used for both the RD and RR, apixaban would have been found non-inferior to enoxaparin.

Lessons learned

We found substantial variation in NI margins used in NI trials of oral anticoagulant medications compared with enoxaparin for prophylaxis of venous thromboembolism after orthopaedic surgery. Such variation could lead to inconsistent conclusions on non-inferiority and the efficacy of the studied drugs compared with placebo. Furthermore, when determining a NI 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 NI 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 NI 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 endpoints (e.g., stricter margins for irreversible outcome, such as death), the treatment effect of active comparator versus placebo (e.g, using a larger margin for larger effects), adverse effects of the test drug (e.g., using a larger margin if test drug has fewer serious adverse effects than available therapies), the availability other drugs (e.g., using a stricter margin if other efficacious and safe drugs are available) and overall cost and benefit-risk assessment.[3,21] Although all NI trials in our case study were similar in terms of these considerations substantial variation in the NI margin existed between the trials, suggesting that the different clinical judgements and perceptions of the investigators played a role.

Furthermore, for valid inference of a NI 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.[22] In our case study, although the placebo-controlled trials were quite similar to the NI trials, they did differ in their definition of outcomes. The question, therefore, remains as to whether the NI trials and placebo-controlled trials were similar enough. This is another subjective judgement inherent to NI 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 an NI trial can raise concerns over the conclusion of non-inferiority.

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 authorisation, either in a peer-reviewed publication or on an independent website (e.g., www .clinicaltrials .gov).[23] Such disclosure of data will certainly help improve the quality of future trials.

The draft FDA guideline does not explicitly state whether the NI 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 benefits and harms of an intervention, which is crucial in NI 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 NI trials, the RDs and RRs can yield opposite conclusions regarding non-inferiority if the rate of events seen in in the active comparator group differs from the assumed event rate that was used to define the NI margin. In a superiority trial, this cannot occur.

Substantial variation in NI margins exists among NI trials of anticoagulant medications for prophylaxis of venous thromboembolism after orthopaedic surgery, which could lead to inconsistent conclusions of a drug's NI 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 NI margin.

References :

(1) Wangge G, Klungel OH, Roes KCB, de Boer A, Hoes AW, Knol MJ. Interpretation and Inference in Noninferiority Randomized Controlled Trials in Drug Research. Clin Pharmacol Ther 2010 print;88(3):420-23.

(2) Wangge G, Klungel OH, Roes KCB, de Boer A, Hoes AW, Knol MJ. Room for Improvement in Conducting and Reporting Non-Inferiority Randomized Controlled Trials on Drugs: A Systematic Review. PLoS ONE 2010 10/27;5(10):e13550.

(3) Center for Drug Evaluation and Research (CDER) Center for Biologics Evaluation and Research (CBER). (2010) DRAFT GUIDANCE : Guidance for industry non-inferiority clinical trials. Available at : http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM202140.pdf

(4) ICH Expert Working Group. (1998) Stastical principles for clinical trials. E9 . Available at : http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E9/Step4/E9_Guideline.pdf (5) ICH Expert Working Group. Choice of Control Group and Related Issues in Clinical Trials. E-10 2001.

(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/Guide lines/Efficacy/E10/Step4/E10_Guideline.pdf

(6) Committee for Medicinal Products for Human Use (CHMP). (2005) Guideline on the choice of the non-inferiority margin. EMEA/CPMP/EWP/2158/99 . Available at : http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003636.pdf

(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 Jan 30;22 (2):169-86.

(8) Fleming TR. Design and interpretation of equivalence trials. Am Heart J 2000 Apr;139 (4):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 Feb;47(1):28-36; 99-107.

(10) Lange S, Freitag G. Choice of delta: requirements and reality--results of a systematic review. Biom J 2005 Feb;47(1):12-27; 99-107.

(11) Colwell CW, Berkowitz SD, Davidson BL, Lotke PA, Ginsberg JS, Lieberman JR, 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 Throm Haemost 2003;1(10):2119-30.

(12) Eriksson BI, Dahl OE, Rosencher N, Kurth AA, van Dijk CN, Frostick SP, et al. Oral dabigatran etexilate vs. subcutaneous enoxaparin for the prevention of venous thromboembolism after total knee replacement: the RE-MODEL randomized trial. Journal of thrombosis and haemostasis : J Throm Haemost 2007;5(11):2178-85.

(13) Ginsberg JS, Davidson BL, Comp PC, Francis CW, Friedman RJ, Huo MH, et al. Oral Thrombin Inhibitor Dabigatran Etexilate vs North American Enoxaparin Regimen for Prevention of Venous Thromboembolism After Knee Arthroplasty Surgery. J Arthroplasty 2009 1;24(1):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, noninferiority trial. Lancet 2007;370:949-56.

(15) Turpie AGG, Levine MN, Hirsh J, Carter CJ, Jay RM, Powers PJ, et al. A randomized controlled trial of a low-molecular-weight heparin (enoxaparin) to prevent deep-vein thrombosis in patients undergoing elective hip surgery. N Engl J Med. 1986;315 (15):925-9.

(16) Leclerc JR, Geerts WH, Desjardins L, Jobin F, Laroche F, Delorme F, et al. Prevention of deep vein thrombosis after major knee surgery--a randomized, double-blind trial

comparing a low molecular weight heparin fragment (enoxaparin) to placebo. Thromb Haemost. 1992 Apr 2;67(4):417-23.

(17) Kalodiki EP, Hoppensteadt DA, Nicolaides AN, Fareed J, Gill K, Regan F, et al. Deep venous thrombosis prophylaxis with low molecular weight heparin and elastic compression in patients having total hip replacement. A randomised controlled trial. Int Angiol. 1996 Jun;15(2):162-8.

(18) Samama CM, Clergue F, Barre J, Montefiore A, Ill P, Samii K. Low molecular weight heparin associated with spinal anaesthesia and gradual compression stockings in total hip replacement surgery. arar study group. British Journal of Anaesthesia. 1997 78 (6):660-5.

(19) Eriksson BI, Agnelli G, Cohen AT, Dahl OE, Lassen MR, Mouret P, 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(12):2490-6.

(20) Lassen MR, Gallus A, Raskob GE, Pineo G, Chen D, Ramirez LM. Apixaban versus enoxaparin for thromboprophylaxis after hip replacement. N Engl J Med. 2010 ;363 (26):2487-98.

(21) Kaul S, Diamond GA. Making sense of noninferiority: A clinical and statistical perspective on its application to cardiovascular clinical trials. Prog Cardiovasc Dis. 2007 2;49(4):284-99.

(22) Fleming TR. Design and interpretation of equivalence trials. Am Heart J. 2000 Apr;139 (4):S171-6.

(23) News Release : New Joint Industry Clinical Trials Transparency Position Requires Companies to Disclose all Clinical Trials in Patients, (10 November 2009, 2009).

(24) Fuji T, Ochi T, Niwa S, Fujita S. Prevention of postoperative venous thromboembolism in japanese patients undergoing total hip or knee arthroplasty: Two randomized, double-blind, placebo-controlled studies with three dosage regimens of enoxaparin. J Orthop Sci 2008 Sep 01(5):442-51.

(25) Agnelli G, Eriksson BI, Cohen AT, Bergqvist D, Dahl OE, Lassen MR, et al. Safety assessment of new antithrombotic agents: Lessons from the EXTEND study on ximelagatran. Thromb Res. 2009;123(3):488-97.

(26) Eriksson BI, Borris LC, Friedman RJ, Huisman MV, Kakkar AK, Bandel TJ, et al. Rivaroxaban versus Enoxaparin for Thromboprophylaxis after Hip Arthroplasty. N Engl J Med. 2008 Jun 26;358(26):2765-75.

(27) Lassen MR, Ageno W, Borris LC, Lieberman JR, Rosencher N, Bandel TJ, et al. Rivaroxaban versus enoxaparin for thromboprophylaxis after total knee arthroplasty. N Engl J Med. 2008 Jun 26;358(26):2776-86.

(28) Turpie AGG, Lassen MR, Davidson BL, Bauer KA, Gent M, Kwong LM, et al. Rivaroxaban versus enoxaparin for thromboprophylaxis after total knee arthroplasty (RECORD4): A randomised trial. Lancet. 2009 May 16;373(9676):1673-80.

(29) Lassen MR, Gallus A, Raskob GE, Pineo G, Chen D, Ramirez LM. Apixaban versus enoxaparin for thromboprophylaxis after hip replacement. N Engl J Med. 2010 Dec 23;363(26):2487-98.

(30) Lassen MR, Raskob GE, Gallus A, Pineo G, Chen D, Hornick P. Apixaban versus enoxaparin for thromboprophylaxis after knee replacement (ADVANCE-2): A randomised double-blind trial. Lancet. 2010 Mar 6;375(9717):807-15.

Expert-opinion on non-inferiority margin:

a case study of oral anticoagulant agents for prophylaxis of venous thromboembolic events after orthopedic surgery

Grace Wangge
Anthonius de Boer
Arno W. Hoes
Olaf H. Klungel
Mirjam J. Knol

Submitted for publication

ABSTRACT

Background: Determination of the non-inferiority (NI) margin in NI trials should be based on both statistical and clinical considerations. The choice of an NI margin is usually presented without providing the underlying rationale. Therefore, it often remains unclear how the investigator incorporates clinical judgment while deciding on the NI margin. This clinical judgment is subjective and different experts may come up with contradicting judgments. In this study, we evaluated how experts' (clinical) considerations guide their choice of NI margin, using the example of oral anticoagulant therapy.

Methodology and Principal Findings: A case study of a hypothetical new anticoagulant that should be tested against enoxaparin for prophylaxis of VTE after orthopedic surgery in a phase III randomized clinical trial was developed. Experts were asked to decide on the NI margin for the new NI trial via an online questionnaire. They were asked to give their choices of the NI margin in two study-sections: before and after additional information on the statistical NI margin was presented. Furthermore, we asked the experts why they chose a specific NI margin. A large variation is existed in NI margins provided by the 25 experts included in our study. Nine experts chose risk difference (RD) as the effect measure for the NI margin, eight experts chose relative risk (RR), and six experts chose both RR and RD. The median NI margin of RD was 1.8 % (interquartile range (IQR) 1 to 2 % and the median NI margin of RR was 1.3 (IQR 1.05 to 1.5). After information on statistical consideration for the NI margin was provided to the experts, the median NI margin of RD increased to 9% (IQR 7.7 to 10 %), while for RR the median of NI margin was 1.25 (IQR 1.2 to 1.5). Clear reasons underlying the choice of NI margin was given by 60% of the experts, even though additional information on the statistical NI margin was presented.

Conclusion: We conclude that presently subjectivity plays an important role in the determination of NI margins. In order to increase objectivity, more guidance is needed to improve adequate and consistent determination of clinically acceptable NI margins.

Introduction

Non-inferiority (NI) trials aim to show that a new drug is not worse than its comparator, which typically is an active drug. The crucial but difficult step in designing an NI trial is pre-specifying an NI margin, a limit by which it can be established that the new drug is similar or not worse than its comparator. The NI margin should be determined in such a way that the new drug can be shown, albeit indirectly, to be effective relative to placebo, also taking into account the uncertainty in the effect size of the active control versus placebo.

The guidelines on NI trials[1-4] state that NI margins should be based on both clinical and statistical judgments. However, details on how such a margin should be determined are not clearly specified. The recently issued draft FDA guideline on NI trials provides a clearer insight on how a margin can be defined based on statistical judgment, but not on clinical judgment, while acceptance of a specific NI margin by regulators and, most notably, by prescribers typically also, and perhaps predominantly, involves clinical consideration.[4]

Interestingly, in 22% of publications of the NI trials we previously reviewed, the margin was based on clinical considerations of the investigator only.[5] How investigators incorporate this clinical judgment remains unknown. These implicit clinical judgments might have been derived from clinical experience. However, these judgments remain subjective and different clinicians may propose contradicting judgments. Thus, it is important to study how this clinical judgment can be incorporated in the NI margin determination.

In this study, we evaluated the reasoning of experts in determining an NI margin. We created a hypothetical new anticoagulant, 'Escheraban', which was tested against enoxaparin for prophylaxis of venous thromboembolic events (VTE) after orthopedic surgery in a phase III NI trial. We asked clinicians, regulators, and researchers from pharmaceutical industry what the NI margin should be via an online questionnaire. Furthermore, we asked the expert about the reasons for their choice of NI margin.

Methods

Procedure and participants

We identified (inter)national clinical experts and researchers by searching for corresponding authors of publications on randomized trials of direct thrombin inhibitor (DTI), direct XA inhibitor (DXAI) and enoxaparin in Pubmed and Cochrane-center register of clinical trials from 2006 to 2011. Additionally, we approached our own networks and asked responding experts to provide names of important experts in the field. For regulatory experts, we contacted the members of the scientific advice working party from the European Medicine Agency (SAWP-EMA).

We sent an invitation email containing an individual token and a link to the questionnaire to 178 experts. A maximum of three reminders, with an interval of three weeks each, were sent.

Questionnaire

We developed an online questionnaire in the Lime-survey system [6], an open source survey application which is hosted on the Utrecht pharmacy panel for education and research (UPPER)-server (see appendix). The questionnaire consisted of three parts: the characteristics and experience of the experts, study-section I, and study-section II. The online-questionnaire was anonymous and personal data of the respondents were used to track responder status only.

In the first part of the questionnaire, experts were asked about their age, gender and their profession. For clinicians, we also asked whether they were involved as a consultant for pharmaceutical industry and/or regulatory bodies. Furthermore, we asked the experts whether they had any experience in anticoagulants trials, trials in VTE or orthopedic surgery, NI trials in general, NI trials in anticoagulants, NI trials in VTE, and NI trials in orthopedic surgery.

In the second part (study-section I), we presented a brief explanation on NI margins and a hypothetical plan for a future NI trial of 'Escheraban', a new oral anticoagulant (direct XA inhibitor) indicated for the prophylaxis of VTE in orthopedic

surgery, with enoxaparin as the active comparator. The efficacy endpoint was the composite of established deep vein thrombosis, or pulmonary embolism (confirmed by state-of-the-art imaging), or death from any cause during the intended treatment period (14 days). The primary safety outcome was bleeding during the treatment period or the two days thereafter.

Based only on the above information, experts were asked to define the effect measure (risk-difference (RD) and/or relative-risk (RR)) and the excess risk they would accept as the NI margin for this NI trial; for example a relative risk of 1.2 or risk difference of 10%. Furthermore, in an open question, we asked the rationale for their choice of the NI margin.

In the third part (study-section II), we presented a brief explanation on how an NI margin can be determined based on statistical considerations, in this case by using the fixed-margin approach described in the draft FDA guidelines for NI trials 2010.[4] We presented two possible choices of NI margins for the 'Escheraban' trial based on a meta-analysis of placebo controlled trials of enoxaparin, using of 50 % preserved effect. These possible NI margins were an RD of 9 % and an RR of 1.25. Subsequently, respondents were asked again to define the effect measure, the NI margin they would use and their rationale for the NI margin.

Data analysis

Two authors (GW and MK) assessed and classified the experts' choices of NI margin and their rationale. If necessary, the value of the NI margin was converted by GW and MK to a valid value. For example, if an expert mentioned that he chose a RR of 10 % as the NI margin, we converted the value of 10 % to 1.1. This was done for four NI margins in study-section I and one value in study-section II. Questionnaires were excluded from the analysis if neither study-section I nor study-section II was filled in.

Results

Questionnaire and general characteristics

From 33 experts who agreed to participate in this study, we excluded responses

Table 1. General characteristics of respondents

Characteristic	N (%) unless stated otherwise
Age, median (range)	51 (33-62)
Male gender	22 (88)
Main profession	
Clinicians	8 (32)
Regulators	7 (28)
Researchers working in academia	6 (24)
Researchers in industry	2 (8)
Other	2 (8)
Experience in	
Trials on the effect on anticoagulants	18 (72)
Trials on the effect of some intervention on the risk of venous thromboembolic events	12 (48)
Trials in patients undergoing any orthopedic surgery	7 (28)
Design or conduct of a NI trial	18 (72)
Designing NI trials on the effect of anticoagulants	13 (52)
Designing NI trials on the effect of some intervention on the risk of VTE	7 (28)
Evaluation of NI trials on the effect of anticoagulants	17 (68)
Evaluation of NI trials on the effect of some intervention on the risk of venous thromboembolic events	9 (36)

from eight of the experts due to incompleteness of the questionnaire.

The mean age of the experts was 51 years, 88% was male, and 32% were clinicians (Table 1). Most of the experts had experience in trials of anticoagulants (72%), design and conduct of NI trials (72%), and design and conduct of NI trials in anticoagulants (52%). Only 28% of the experts had experience in trials of patients undergoing orthopedic surgery.

Choices of NI margin

In table 2 and figure 1, we present the choices of the experts on the NI margin in both study-sections. In study-section I, Nine experts chose risk difference (RD) as the effect measure for the NI margin, eight experts chose relative risk (RR), and six experts chose both RR and RD. The median NI margin of RD was 1.8 % (interquartile range (IQR) 1 to 2 % and the median NI margin of RR was 1.3 (IQR 1.05 to 1.5). In study-section II, in total 14 experts chose RD as the effect measure with a median of 9.0% (IQR 5.5 to 10 %) and 13 experts chose RR as the effect measure with a median of 1.25 (IQR 1.2 to 1.3).

Figure 1 shows the variation in NI margins between the experts. The variation in

Table 2. Choices of the experts

Study-section I	Frequency (%)	Study-section II	Frequency (%)
Choice of effect measure		**Choice of effect measure**	
RD	9 (36)	RD	9 (36)
RR	8 (32)	RR	8 (32)
Both RD and RR	6 (24)	Both RD and RR	5 (20)
Not given	2 (8)	Not given	3 (12)
Reasons of experts on the choice of NI margin in study - section I		**Reasons of experts on the choice of NI margin in study-section II**	
To be as narrow as possible	3 (12)	To be as narrow as possible	3 (12)
The value was considered as clinically relevant	3 (12)	Based on preserved effect of placebo controlled trial(s)	9 (36)
Copied value from previous NI trials	3 (12)	The value was considered as clinically relevant	3 (12)
Cannot decide due to lack of data	3 (12)	Cannot decide due to lack of data	2 (8)
Based on preserved effect of placebo controlled trial(s)	2 (8)	Not clear	4 (16)
Based on similar trials in other therapeutic areas	1 (4)	Not given	4 (16)
Not clear	10 (40)		

NI margin with RD as the effect measure was lower in study-section I than in study -section II, and the chosen NI margins in study-section II were larger (further from 0) than in study-section I. In contrast, the variation in NI margin with RR as the effect measure was larger in study-section I than in study-section II, while the median. RR was similar in both study sections.

The dynamics of the choice of experts on NI margins from study-section I to study-section II are presented in more detail in Table 3. Most of the experts changed their choice of NI margin (15 experts, 60%). Four experts (16%) chose the same effect measure, but their NI margin values became more lenient/ larger; while one expert (4%) chose a stricter NI margin in study-section II. Only four experts (16%) did not change their choice of NI margin.

Rationale for the choice of NI margins

In table 2, the underlying reasons for choosing an NI margin are shown. In study-section I: less than half of the experts had a clear rationale for how they chose the NI margin (12 experts, 48%): three (12%) experts stated they

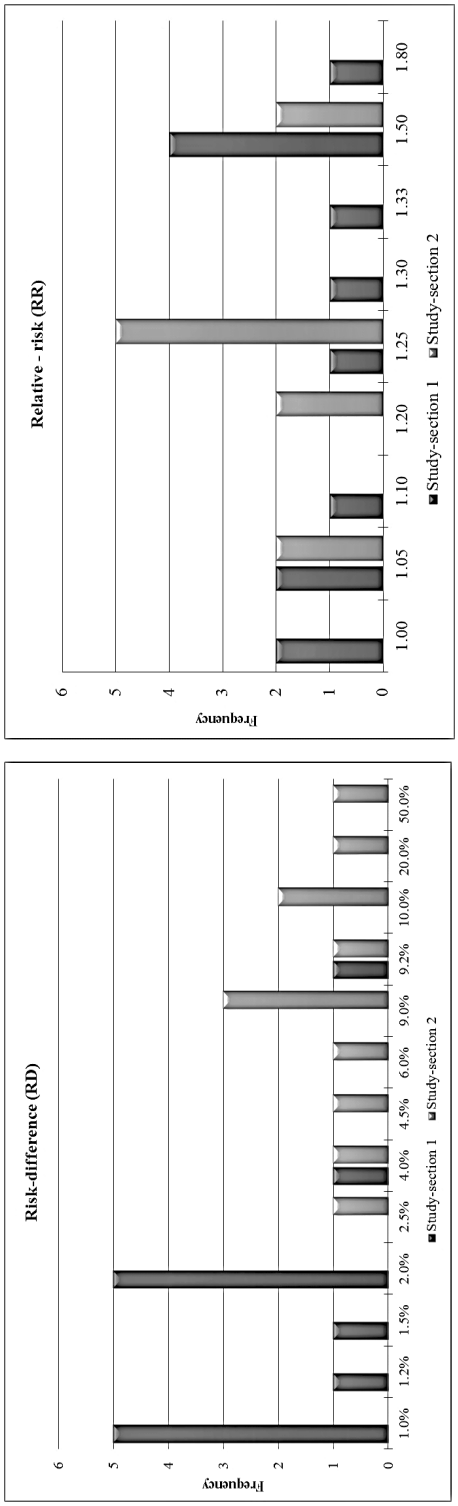


Figure 1. The choices of the experts on NI margins

In study-section 1, one expert gave the effect measure of both RD and RR, but did not give any value because he needed more information to determine the NI margin values. In study-section 2, another expert did the same

wanted the NI margin to be as narrow as possible; three (12%) experts copied the value from previous NI trials; three (12%) experts considered their choice of NI margin as "clinically relevant"; two (8%) experts used a specific preserved fraction of the effect seen in placebo controlled trials; and one (4%) expert chose the NI margin based on similar trials in other therapeutic areas. Of the two experts that used a certain preserved effect: one expert chose 67% preserved effect, while the other expert was not clear on how much preserved effect he would use.

In study-section II, fifteen experts (60%) reported a clear rationale for their NI margin: nine (36 %) experts based their choice on a specific preserved effect; three (12%) stated they wanted the NI margin to be as narrow as possible and three(12%) experts considered their choice of NI margin as clinically relevant.

Of the nine experts that used a certain preserved effect, seven experts chose 50% preserved

Table 3. Dynamics on NI margin choices from study-section I to study-section II		
		Frequency (%)
Same NI margin		4 (16)
Change of NI margin	Same effect measure, but value more lean	4 (16)
	Same effect measure, but value more strict	1 (4)
	Same effect measures, but values more strict or more lean #	2 (8)
	Different effect measure	8 (32)
Only gave NI margin in study-section I		4 (16)
Dynamics cannot be defined*		2 (8)

Note : # Experts gave NI margins for both RR and RD. One expert gave a more lenient RD in study-section II, while the value of RR was the same with his choice in study-section I. Another expert gave the same RD in study-section II as in study-section I, while the value of RR was stricter in study-section II.
* The experts did not present a NI margin in either one of the study-sections, but gave their rationale not to do so ("cannot decide due to lack of data")

effect, one expert chose 67% preserved effect, and one expert chose 75 % preserved effect.

Discussion and conclusion

This study has three major findings. First, NI margins provided by experts showed a large variation even after additional information on the statistical NI margin was given. Second, most of the experts did not have a clear reasoning for their choice of NI margin. Third, there was no clear preference on the choice of effect measure (risk difference or relative risk) to base an NI margin on.

The large variation in NI margins in this study corresponds with the large variation found in published NI trials even when similar drugs are compared (e.g. new oral anticoagulants versus enoxaparin).[7] This large variation still existed when additional information on the statistical margin was presented. Moreover, we also observed that stricter NI margins using absolute risk (RD) were chosen by the majority of the experts in study-section I compared with study-section II. Experts might have intuitively chosen to minimize the risk as much as possible, when the exact size of the estimated treatment effect between the active comparator and the placebo was unknown. This seems sensible from a clinical point-of-view, but illustrates that NI margin determination is highly

subjective. In certain cases, choosing the 'safest' NI margin might not be logistically feasible because it generate an unachievable large sample size. Although such a pragmatic consideration was not observed in our study, it can be expected that it might influence the choice of the NI margin. Taken this all together, we believe it is important to clearly describe the process of determination of NI margin in any research proposal or publication of an NI trial, including any reasoning that was done by the investigator for choosing a specific NI margin.

In addition to the large variation in NI margin, only few experts provided a clear reasoning for the choice of NI margin in study-section I. This number increased in study-section II, where more experts based their choice of NI margin on a certain preservation of the effect of the active comparator. When the concept of preserved effect was introduced in study-section II, some experts were willing to increase the NI margin value. This shows that the use of preserved effect might be helpful to make a decision on the NI margin, but such merely statistical guidance may not convince prescribers in accepting the margin as begin acceptable from a clinical perspective.

Furthermore, we observed no clear preference on the choice of effect measure to base an NI margin on. We expected that an absolute measure like RD would be chosen by the majority of the experts because it is particularly useful when considering trade-offs between benefits and harms of an intervention, which is crucial in NI trials, but only 36% chose the risk difference. Importantly, the absolute and relative risk can yield opposite conclusions regarding non-inferiority if the observed event rate in the active comparator group differs from the assumed event rate that was used to define the NI margin. In a superiority trial, this cannot occur.

The small number of respondents is the major limitation of our study. A larger number of respondents are, however, unlikely to change our major findings that a large variation in the choice of NI margins exists and that clinical considerations can be very subjective. This study is the first to report how experts determine an NI margin in NI trials and may contribute to further research in the field.

We conclude that presently subjectivity plays an important role in the determination of NI margins. In order to increase objectivity, more guidance is needed to

improve adequate and consistent determination of clinically acceptable NI margins.

References

(1) ICH Expert Working Group. (1998) Stastical principles for clinical trials. E9 . Available at : http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E9/Step4/E9_Guideline.pdf

(2) ICH Expert Working Group. (2001) Choice of control group and related issues in clinical trials. E-10 .Available at : http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E10/Step4/E10_Guideline.pdf

(3) Committee for Medicinal Products for Human Use (CHMP). (2005) Guideline on the choice of the non-inferiority margin. EMEA/CPMP/EWP/2158/99 . Available at : http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003636.pdf

(4) Center for Drug Evaluation and Research (CDER) Center for Biologics Evaluation and Research (CBER). (2010) DRAFT GUIDANCE : Guidance for industry non-inferiority clinical trials . Available at : http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM202140.pdf

(5) Wangge G, Klungel OH, Roes KCB, de Boer A, Hoes AW, Knol MJ. Room for Improvement in Conducting and Reporting Non-Inferiority Randomized Controlled Trials on Drugs: A Systematic Review. PLoS ONE 2010 10/27;5(10):e13550.

(6) LimeSurvey Project Team - Carsten Schmitz. LimeSurvey: An Open Source survey tool - LimeSurvey Project Hamburg, Germany. 2012; Available at: http://www.limesurvey.org, 2011.

(7) Wangge G, Roes KC, de Boer A, Hoes AW, Knol MJ. The challenges of determining noninferiority margins: a case study of noninferiority randomized controlled trials of novel oral anticoagulants. CMAJ 2012 Aug 20.

Appendix– Questionnaire

Clinical considerations in determining a non-inferiority margin

Dear colleague,

You are invited to participate in a survey on "Clinical considerations in determining a non-inferiority margin".

This survey is part of a study to identify challenges in determining a non-inferiority margin using the case study of oral anticoagulants for prophylaxis of venous thromboembolic events (VTE) after orthopedic surgery. This survey forms part of a PhD-project. In a first part, we looked at statistical considerations when choosing a non-inferiority margin. In this part of the project, we aim to identify clinical considerations in choosing a non-inferiority (NI) margin and we would appreciate your participation in this survey.

The aim of the current study is to ask clinical and regulatory experts what they consider the appropriate NI margin for a future NI trial on oral anticoagulants, direct thrombin inhibitors (DTI) and direct XA inhibitors (DXAI), for prevention of VTE in patients undergoing elective hip or knee replacement surgery and what their motivations are to arrive at this NI margin.

Each participant will be asked to fill in the NI margin they find appropriate in a hypothetical scenario before and after additional information is given. In a second round, the results of the first round will be presented and again each participant will be asked to fill in their preferred NI margin.

We invite you to participate in this survey, because:

- You are a clinical expert in the field of anticoagulants, or
- You are an expert in the field of non-inferiority trials from a regulatory perspective, or
- You are a scientist from pharmaceutical industry on the topic of anticoagulants

This online questionnaire will take approximately 15 minutes of your time. This survey is for scientific purposes only. The information you provide will handled anonymously and kep confidential and will not be distributed to other people than those in the research group listed below.

This study is a PhD project and part of *The Escher project: science-driven drug regulation and innovative research throughout phased drug development,* from Top Institute Pharma (http://www.tipharma.com)

About us

We are a research group from the division of Pharmacoepidemiology and Clinical Pharmacology, Utrecht University and from the Julius Center, University Medical Center Utrecht, The Netherlands.

Best regards,
Grace Wangge,MD
PhD student

Other members of the research group are: Prof. A. de Boer, MD, PhD, Prof. A.W. Hoes, MD, PhD, O.H. Klungel, PharmD, PhD, M.J. Knol, PhD.

There are 18 questions in this survey

Your consent

Do you agree to participate in this study?
Please choose only one of the following:

- Yes
- No

Questions below only had been asked to people who agree to participate in the study

About you

First, we would like to know some information about you as an expert.

2. What is your age in years?
Please write your answer here: _____

3. Your gender:
Please choose **only one** of the following:

- Female
- Male

4. Your main profession (Please choose one):

- Regulator
- Clinician
- Researcher working in academia
- Researcher working in pharmaceutical industry
- Other(please specify): _____

5. Are you also a consultant for the pharmaceutical industry?

This question will only appeared if the answer was 'Regulator' or 'Clinician' or 'Researcher working in academia' at question '4 ' (Your main profession (Please choose one):)

Please choose **only one** of the following:

- Yes
- No

6. Are you also a member of a regulatory agency (eg EMA, FDA)

This question will only appeared if the answer was 'Clinician' or 'Researcher working in academia' at question '4' (Your main profession (Please choose one):

Please choose **only one** of the following:

- Yes
- No

7. Were you ever involved in (the development of) any trial in either one of the topics below:

	Yes	Uncertain	No
Trials on the effect of anticoagulants	○	○	○
Trials on the effect of some intervention on the risk of venous thromboembolic events	○	○	○
Trials on patients undergoing any orthopaedic surgery	○	○	○

8. Were you ever involved in (the design or conduct of) a non-inferiority trial?
Please choose **only one** of the following:

- Yes
- No

9. Were you ever involved in (the development of) a non-inferiority trial on one of the topics below:

	Yes	Uncertain	No
Trials on the effect of anticoagulants	○	○	○
Trials on the effect of some intervention on the risk of venous thromboembolic events	○	○	○
Trials on patients undergoing any orthopaedic surgery	○	○	○

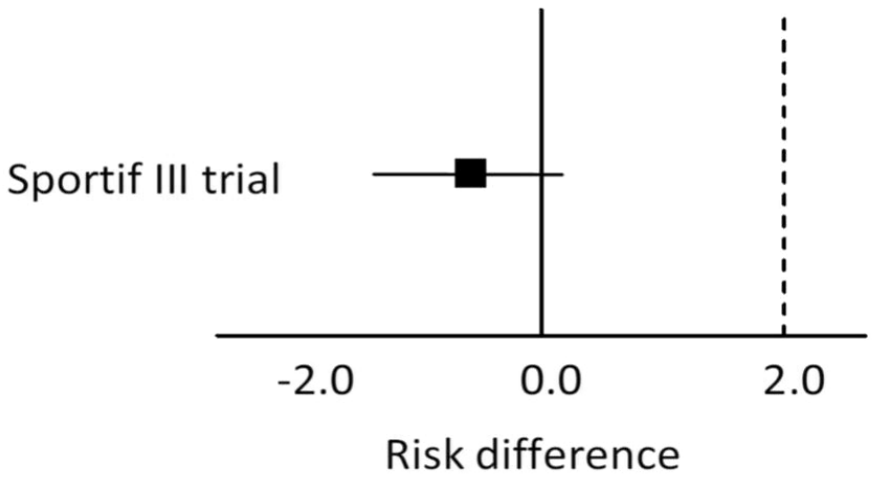
10. Were you ever involved in the evaluation of a protocol (e.g. as a member of an ethics committee or grant committee) or a manuscript (as a journal referee or editor) of a non-inferiority trial on one of the topics below:

	Yes	Uncertain	No
Trials on the effect of anticoagulants	○	○	○
Trials on the effect of some intervention on the risk of venous thromboembolic events	○	○	○
Trials on patients undergoing any orthopaedic surgery	○	○	○

Non-inferiority margin

The aim of a non-inferiority (NI) trial is to show that a new treatment is not worse than its comparator, which typically is an active drug. NI trials can be used in a situation when a new drug considered has a similar efficacy profile as its comparator but may offer other advantages over the existing drug such as a novel method of administration or a better safety profile. In a regulatory setting, NI trials can be used to provide primary, but indirect, evidence of a new drug's efficacy in cases where a placebo is not ethically justified.

The main step in designing an NI trial is pre-specifying an NI margin, i.e. a limit by which it can be established that the new drug is not worse than its comparator. An example is the SPORTIF III trial *(Lancet 2003;362:1691-98)* that compared ximelagatran with warfarin for prevention of thromboembolism in patients with nonvalvular atrial fibrillation. The NI margin in this trial was determined to be a risk difference of 2%. This means that an excess risk of 2% in the ximelagatran group as compared with the warfarin group was considered acceptable to declare that ximelagatran was not worse than warfarin. The trial showed that the risk of thromboembolism was xx% in the ximelagratran group and xx% in the warfarin group. Thus the risk difference is -0.7% and the 95% confidence interval -1.4 to 0.1%. The upper boundary of the confidence interval (0.1%) is below 2%, thus non-inferiority was concluded (see Figure below).



In this example, the NI margin was defined based on a risk difference. An NI margin can also be defined on a relative risk scale, for example, the NI margin is a relative risk of 1.2.

Case study

As many of the trials in direct thrombin inhibitors (DTI) and direct XA inhibitors (DXAI) for prevention of VTE in patients undergoing elective hip or knee replacement surgery were NI trials, we use this as a case study to identify challenges in determining an NI margin.

Suppose that we would like to conduct an NI trial on a new and promising DXAI for prevention of VTE after orthopedic surgery. We will first provide some details on the proposed study. Then we will ask you to define an NI margin for this future trial. Next, we will provide more background information on previous studies in this field and ask you again to define an NI margin for the future trial.

Future trial

Suppose we want to conduct an NI trial on a new potent, selective, oral, DXAI, called escheraban, for prevention of VTE after orthopedic surgery. In previous phase II trials, escheraban has showed to be similarly safe and effective with enoxaparin.

The trial is a randomized, non-inferiority, double-blind, controlled trial in adult patients undergoing orthopedic surgery. Enoxaparin is used as the active comparator. The efficacy endpoint is the composite of established deep vein thrombosis, or pulmonary embolism (conformed by state-of-the-art imaging), or death from any cause during the intended treatment period (14 days). The primary safety outcome is bleeding during the treatment period or until 2 days after the last dose of study medication is administered. Treatment with *escheraban* will be started 12-24h after surgery.

Your opinion on NI margin in *escheraban* trial:

11. Which effect measurement would you choose for the *escheraban* trial?

- Risk difference (RD)
- Relative risk (RR)
- Both RD and RR
- Other (please mention): _____

12. What excess risk of the composite efficacy endpoint in the *escheraban* group do you find acceptable to declare *escheraban* non-inferior to *enoxaparin*? (Please fill in the number below based on your choice of effect measurement above)

13. Please explain why you choose this specific NI margin. Provide your separate arguments briefly below:

Please write your answer here:

_____
_____
_____
_____
_____

Background information

Most of the guidelines on NI trials only state that a margin should be based on both clinical and statistical considerations. However, a recently issued draft FDA guideline on NI trials (http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM202140.pdf) includes a more specific guidance on how to determine an NI margin. The NI margin should be determined such that the new drug can be shown to be effective relative to placebo and needs to account for the uncertainty in the effect size of

the active control versus placebo. To this aim the guideline introduced the concept of M1 and M2.

M1 resembles the most conservative effect of the active control compared with placebo, assumed present in the NI trial. The most conservative effect is defined as the upper bound of the confidence interval of the (pooled) effect estimate of the difference between the active control and placebo. M1 is typically determined from earlier trials comparing the active control with placebo or a state-of-the-art meta-analysis of such trials. M2 reflects the clinical judgment about how much of M1 should be preserved to be the largest clinically acceptable difference (degree of inferiority) of the test drug compared to the active control. For example, if it were concluded that it would be necessary for a test drug to preserve at least 75% of an effect on a specific outcome, M2 would be 25% of M1, i.e. the loss of effect that must be ruled out. Determination of M2 provides (some) reassurance that the test drug will be superior to placebo and that the effect is clinically relevant.

We determined M1 and M2 for the future trial on *escheraban* following the steps of the draft FDA guideline.

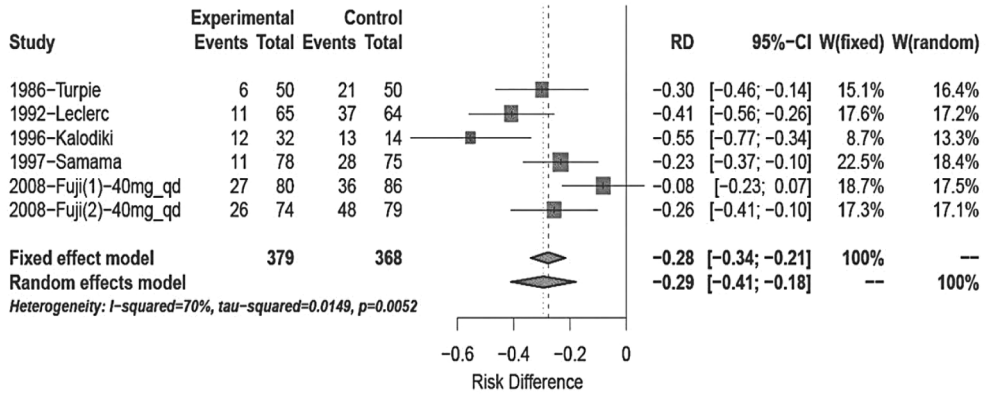
Determining M1

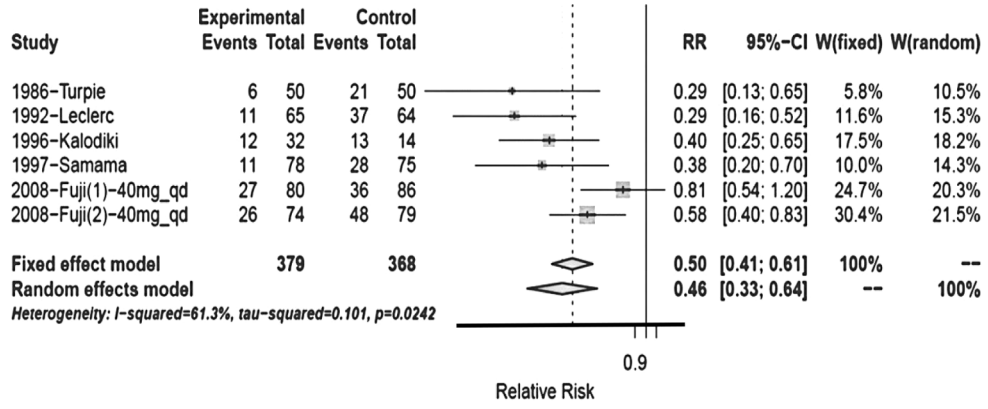
First, we searched for placebo-controlled trials on *enoxaparin*, the active comparator in the future trial.

From Pubmed and Cochrane Central Register of Controlled Trials, we found six placebo controlled trials:

Name of author	Date of publication	Duration of therapy	Dosage of enoxaparin	Primary endpoint	Mean age	Female subjects (%)	N
Turpie, et.al	Oct-86	14	30 mg bid	DVT measured with venography	67	52	100
Leclerc, et.al	Jan-92	14	30 mg bid	DVT measured with venography	69	60	131
Kalodiki, et.al	Jun-96	8 - 12	40 mg qd	Composite of DVT measured with venography and PE	67	65	170
Samama, et.al	Jan-97	10 ± 2	40 mg qd	Composite of DVT measured with venography and PE	69	42	93
Fuji, et.al (1)	Jun-08	14	20 mg qd, 20 mg bid, 40 mg qd	Composite of DVT measured with venography and PE	62	88	419
Fuji, et.al (2)	Jun-08	14	20 mg qd, 20 mg bid, 40 mg qd	Composite of DVT measured with venography and PE	70	84	364

Using the data of the placebo-controlled trials, we calculated a pooled risk difference and relative risk:





We decided to take the upper-bound of the 95% confidence interval of the pooled effect estimate based on the random effects model as M1. Thus, M1 based on a RD is -18% and M1 based on a RR is 0.64.

Determining M2

We calculated M2 by preserving 50% of the M1's effect.

Based on this 50% preserved effect, M2 based on a RD is 9% (18% divided by 2).

The M2 based on a RR is 1.25. This is calculated based on the formula recommended by the draft FDA guideline:

$$(1/M1)^{(1 - \text{preserved-effects})} = (1/0.64)^{(1 - 0.5)}$$

14. Based on the new information, which effect measurement would you choose for the NI margin in *escheraban* trial?

- RD
- RR
- Both RD and RR
- Other (please mention): _____

15. What excess risk of the composite efficacy endpoint in the *escheraban* group do you find acceptable to declare *escheraban* non-inferior to *enoxaparin*? (Please fill in the number below based on your choice of effect measurement above)

16. Please explain why you choose this specific NI margin. Provide your separate arguments briefly below:
Please write your answer here:

_____
_____
_____
_____
_____

Thank you for your participation

17. If you have any colleague you would like to recommend to participate in this study, please list their name and email address below:
Please write your answer here:

18. If you have any comments or questions on this study, please put them below:
Please write your answer here:

Questions below only appeared for people who refused to participate in this study
You refused to participate in this study
2. Please, give the reason why you do not want to participate in this study
Please write your answer here:

3. Can you suggest a colleague that would be interested to participate in this survey? (please provide name and email address)
Please write your answer here:

This is the end of the survey

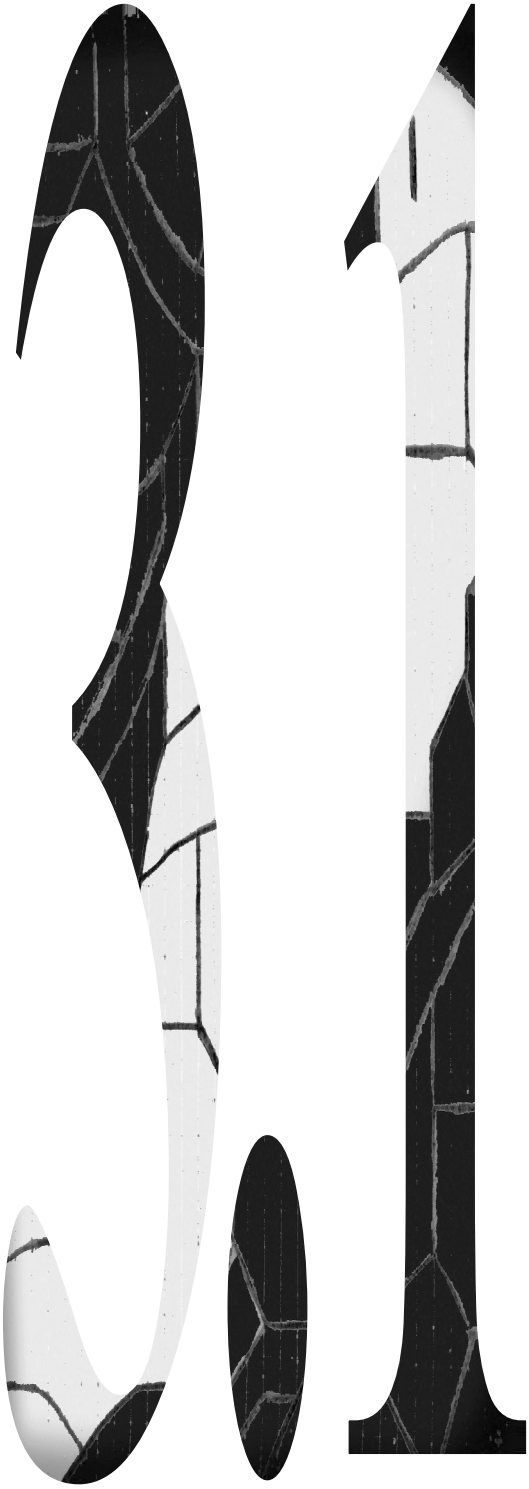
Thank you for your contribution

Submit your survey.

Regulatory challenges in non-inferiority trials

We adore chaos because we love to produce order.

MC Escher, Dutch graphic artist (1898– 1972)



Regulatory scientific advice on non-inferiority drug trials

Grace Wangge*

Michelle Putzeist*

Mirjam J. Knol

C.C. Gispen—De Wied

Antonius de Boer

Arno W. Hoes

Hubert G. Leufkens

Aukje K. Mantel-Teeuwisse

Submitted for publication

** equal contribution*

ABSTRACT

Background : The active-controlled trial with a non-inferiority (NI) design has gained popularity in recent years. NI trials have methodological challenges, especially in determining the NI margin. Regulatory guidelines provide some general statements on how an NI trial should be conducted. Apart from the guidelines, regulators provide opportunities for companies to discuss critical trial issues prior to the trial's conduct; so-called Scientific Advice (SA). In our current study, we identified questions on NI trials that were posed by applicants of European SA in 2008 and 2009, and the responses given by the European medicines agency (EMA) to identify potential issues that may benefit from a more explicit guidance.

Methodology and Principal Findings: We included in our analysis 156 final-advice letters given to 94 different applicants. Our analysis of final advice letters in 2008 and 2009 yielded two major findings: (1) questions 'whether' and 'how' to conduct an NI trial were frequently asked by applicants, but 'how' questions were more frequent than 'whether' questions. (74 % vs. 26 %); (2) the choice of the NI margin seems to be EMA's main concern in NI trials (36 % of total regulatory answers). In 40 % of the EMA answers, they recommended the use of a stricter margin; and in 10 % of the EMA answers on NI margin, they questioned the justification of proposed NI margin.

Conclusion : We conclude that difficulties still exist in selecting the appropriate methodology of NI trials. Straightforward and harmonized guidance on NI trials is needed, such as when to conduct NI trials and how to determine the NI margin. Regulatory guidelines (either as one general guideline or special sections on NI trials in disease-specific guidelines) may not be feasible to cover all therapeutic areas; in that case regulatory scientific advice may be used as an opportunity for tailored advice.

Introduction

Randomized *placebo*-controlled trials (RCT) are considered the gold standard to confirm a drug's efficacy. Nowadays, *active*-controlled trials are often performed instead of or in addition to placebo-controlled trials as the basis for marketing authorization and reimbursement decisions. A previous study showed that for 48 % of new medicines approved between 1999 and 2005 at least one active-controlled trial was conducted during the development phase.[1]

An active-controlled trial may have a non-inferiority (NI) design. An NI trial intends to demonstrate that the new drug is not worse than its comparator (an active drug previously shown to be more effective than placebo) to a certain limit (NI margin), while, thus, indirectly showing that the new treatment is effective (i.e. more effective than placebo). However, NI trials pose several methodological challenges, especially in determining the NI margin. Previously we found that in 22 % of the NI-trials the choice of NI margin was merely based on assumptions made by the investigators.[2]

The ICH E9[3], the ICH E10[4], the European Medicines Agency (EMA) guidelines [5,6] and US Food Drug Administration (FDA) draft guideline on NI trials[7] are the currently available guidelines that advice on the appropriate conduct of NI trials in general. Most of the guidelines only have general statements on how a NI trial should be conducted. Nevertheless, the FDA and EMA provide more explicit guidance in guidelines for trials in certain therapeutic areas (such as diabetes mellitus and infectious diseases) on how to use NI trial methodology.[8,9] Interestingly, in those guidelines where a specific NI margin is given, discrepancies exist between FDA and EMA. For example, in the 2008 draft FDA guidance for diabetes mellitus, an NI margin of 0.3 % *or* 0.4 % HbA1C reduction is suggested, while the 2011 EMA guideline suggests an NI margin of 0.3 %.[8]

Apart from guidelines, regulators nowadays provide opportunities for companies to discuss critical trial issues prior to the trial's conduct, to improve the quality of pre-

registration trials. An important part of such dialogue is formed by so-called Scientific Advice (SA). In Europe, SA can be sought either from the EMA or from one or more of the national regulatory agencies.[10] Regulatory SA can be asked as often as deemed necessary by an applicant, who is not obliged to adhere to, the advices received or committed to accept any result of an SA procedure. In a previous study, we found that one of the top five questions posed by the applicants was on study design.[11] However, we did not assess the questions and the responses of the regulators specifically related to the NI design in more detail. Therefore, it is largely unknown whether companies often ask questions specifically related to NI trial design, and what the nature of and answers to these questions are.

In this study, we identified questions on NI trials that were posed by applicants of European SA in 2008 and 2009, and the responses given by the EMA, to identify potential issues that may benefit from a more explicit regulatory guidance.

Methods

With the keyword “inferior”, we searched among final-advice letters from the EMA, represented by Committee for medicinal products for human use (CHMP), in the years 2008–2009 in the Dutch Medicines Evaluation Board (MEB) database, to identify which documents discussed NI trials. At the time of our study, information on SA in more recent years was not fully available.

Each final-advice letter consisted of questions asked by the applicants followed by a company position (CP) and a CHMP response. The CP is an elaboration of the question from applicants. We excluded documents and each unit of question-CP-CHMP response that did not discuss NI trials on efficacy or discussed bioequivalence trials.

The following information was collected for each SA-application: whether it was a follow-up to a previous SA application, whether the drug was classified as orphan drug, and

indication of the drug. The drugs were categorized by their therapeutic target group according to the Anatomical Therapeutic Chemical (ATC) classification system.[12] In case an ATC classification was missing, the anatomical main group was determined based on the intended indication of the product.

Each question-CP and CHMP response was scored according to the topic of interest. The topics of interest were divided into two types, 'general' and 'specific'. General topics covered discussion about the strategic/overall development process of a drug. The specific topics consisted of NI trial unique topics and topics not related to NI trials. NI trial unique topic included questions on whether an NI trial should be conducted or not ('whether' question: NI study design) and topics that discussed technical issues about how an NI-trial should be conducted ('how' questions; e.g. type of comparator, NI margin, NI sample size calculations, intention to treat (ITT) or per-protocol (PP) analysis; and switching (from non-inferiority to a superiority design or vice versa). NI trials not unique topics discussed aspects of a clinical trial that were not specific to NI trials, for example trial inclusion-exclusion criteria and type of endpoints (See Table 1 for further details and examples).

In each question-CP or CHMP response, multiple topics can be discussed. All topics were included separately in the analyses. Additional topics that were found in the CHMP response, but not in the accompanying question or company position were classified as 'extra information'.

Author GW searched and extracted all questions, company positions, and answers documents, while classification was done by both GW and MP. In case of discrepancies (n=5), AM-T and MK were consulted to reach consensus. Subsequently, data were analyzed by GW and MP in a descriptive way. In addition, the proportions of the topics according to their therapeutic target group were assessed.

Table 1. Examples of topics related to NI trials in question-CP-CHMP response units identified in final-advice letters

Topics			Definition	Example		
				Question	Company position	CHMP answer
Unique to NI trials	WHETHER-topic	NI study design	discusses NI trial design	Does the agency support the selected non-inferiority trial design for the phase 3 study in patients with YYY?	An open-labelled, randomised, multi-centre, parallel group study will be conducted in YYY adult patients for a period of 6-8 months.The primary objective of this	CHMP agrees with the non-inferiority design. However, a two-arm non-inferiority study is considered inadvisable due to concerns relating to assay-sensitivity. For these reasons, inclusion of a placebo control is strongly recommended in addition to the
	HOW- topic	Type of comparator	a) Discusses the type of comparator or b) the use of additional placebo arm in the trial	Does the agency agree on the use of DEF as the active comparator in this non-inferiority trial?	DEF is recommended by the XXX as the current therapy of choice for YYY.	The choice of active comparator is justified.
		NI margin	discusses the NI margin	Does the SAWP consider an X % non-inferiority margin acceptable?	Minimum clinical relevant difference to detect is suggested to X % (non-inferiority margin) preserving 2/3 of the effect of ABC vs. placebo on primary endpoint.	CHMP agrees
		NI data analysis[$]	a) discusses data analysis of an NI trial, including sample size calculation for NI trial or b) Controlling type I error in NI trial	In this trial, the primary analysis is performed on the per-protocol population (with a secondary analysis performed on the intent-to-treat population), with sample size of ZZ patients.Does CHMP agree?	Sample size was calculated based on 5% better response rate in the ABC arm, a non-inferiority margin of X %, one-sided significance level 2.5%, and power of 80% .	CHMP disagrees. A sample size of ZYX patients is necessary to exclude a significant activity difference. We recommend to do the primary analysis in both PP and ITT analysis.
		ITT or PP [$]	discusses the choice of per-protocol (PP) or Intention to treat (ITT)			
		Switching	discusses a plan to switch from NI to superiority or Superiority to NI	Is the test procedure - testing first for non inferiority followed by superiority – acceptable?	The sponsor believes that the described procedure is in line with EMEA/CHMP guidelines for switching between superiority and non-inferiority.	Switching from non-inferiority to superiority as proposed is acceptable.
Non-unique to NI trials			discusses topics related to conduct of the trial, that are not specifically related to NI trials	We propose modifying the primary composite endpoint to include hospitalization in this non-inferiority trial.	The proposal to exchange the two composite endpoints would address the lower than anticipated event rate and would still allow the assessment of the impact of ABC on a composite endpoint.	Reluctance exists from clinical grounds to agree to the proposal. Adding hospitalization would add a component that has a different relevance than the components that were originally agreed as making a valid composite primary endpoint.
General			discusses " Strategy question"	Does the Agency agree that a single Phase 3 study with the proposed primary endpoint and statistical evaluation will provide sufficient data for approval?	Since, at this stage, no approved drug is available for YYY patients, it is considered acceptable to base potential approval of ABC for this disease on the proposed Phase 3 program.	Approval of ABC based on the current program will be possible. A new agent or indication should have a safety/efficacy profile non-inferior to marketed comparators.

Note : $ An example that in one question/CP multiple topics can be discussed.
To ensure confidentiality, the data presented in the table are not the original sentences found in the SA applications.

Table 2.General characteristics of SA applications		
	Based on number of questions	
	Based on number of documents N = 156 (%)	Based on number of questions N = 278 (%)
Follow up application	36 (23)	51 (18)
Orphan drugs	14 (9)	23 (8)
Therapeutic target group		
Antineoplastic and immunomodulating products	34 (22)	63 (23)
Alimentary tract and metabolism	27 (17)	48 (17)
Anti – infective drugs	25 (16)	47 (17)
Blood and blood-forming organs	15 (10)	31 (11)
Respiratory system	10 (6)	19 (7)
Musculoskeletal system	13 (8)	16 (6)
Systemic hormonal preparations, excluding sex hormones and insulins	7 (5)	14 (5)
Nervous system	7 (5)	10 (3)
Others	18 (11)	30 (11)

Results

Search result and general characteristics

In total, there were 350 final-advice documents in the year 2008 and 345 documents in 2009 and 166 of these contained the keyword 'inferior' in the database (75 documents in 2008 and 91 documents in 2009). We excluded nine documents in 2008 and one document in 2009, because they were not related to NI trials. We finally included 156 documents in our analysis, consisting of 66 final-advice letters from 2008 and 90 final-advice letters from 2009. These final-advice letters were given to 94 different applicants. In total, the documents contained 278 question-CP-CHMP response units related to NI trials.

Characteristics of the documents and questions are described in Table 2. Of the therapeutic groups, antineoplastic and immunomodulating products were discussed most often (22% of included final-advice letters), followed by alimentary tract and metabolism products (17 %) and anti-infectives (16 %).

Table 3. Frequency of topics unique to NI trials appearing in Question-CP- CHMP answers based on therapeutic target group

Topic		All therapeutic groups N (% of total)		Antineoplastic and immunomodulating drugs N (% of total)		Alimentary tract and metabolism drugs N (% of total)		Anti – infective drugs N (% of total)		Other drugs N (% of total)	
		Q-CP	CHMP answers	Q-CP	CHMP answers	Q-CP	CHMP answers	Q-CP	CHMP answers	Q-CP	CHMP answers
Whether topic	NI study design	92 26)	67 (28)	21(25)	11 (20)	11(22)	10 (26)	4 (20)	3 (21)	56 (28)	43 (32)
HOW - topics	Type of comparator	34(10)	22 (9)	8 (9)	5 (9)	4 (8)	6 (15)	4 (20)	1 (7)	18 (9)	10 (7)
	NI margin	98(28)	86 (36)	21(25)	21(38)	15(31)	11 (28)	8 (40)	8 (57)	54 (27)	46 (35)
	NI data analysis	87(25)	50 (21)	25(29)	15 (27)	15(31)	10 (26)	3 (15)	2 (14)	44 (22)	23 (17)
	ITT or PP	22 (6)	8 (3)	5 (6)	0 (0)	1 (2)	0 (0)	0 (0)	0 (0)	16 (8)	8 (6)
	Switching	21 (6)	9 (4)	5 (6)	3 (6)	3 (6)	2 (5)	1 (5)	0 (0)	12 (6)	4 (3)
TOTAL		354 (100)	242 (100)	85 (100)	55 (100)	49 (100)	39 (100)	20 (100)	14 (100)	200 (100)	134 (100)

Topics of discussion

Within the 278 questions-CP units related to NI trials, a total of 587 different topics were discussed. Of thes, 101 were classified as general topics, asking advice regarding the overall development strategy which may include a non-inferiority RCT. Issues that were specific, but not unique to NI trial design were identified 132 times. The remaining 354 topics were unique to NI trials. In CHMP answers, a total of 400 different topics were discussed. Of those 242 topics were unique to NI trials.

Among the NI trial unique topics, both topics of 'whether' and 'how' to conduct an NI trial frequently appeared in the questions-CPs and CHMP answers , but 'how' questions (74% of total NI topics asked and 72% of total CHMP answers) were more frequently asked than 'whether' questions (26% of total topics asked and 28 % of total CHMP NI unique answers). Among the 'how' topics the NI margin was most frequently discussed in questions-CP and CHMP answers (98 (28 %) questions-CPs and 86(36 %) CHMP-answers of all NI trial unique topics). In 42 out of 86 (49%) CHMP answers

that discussed NI margin, the CHMP supported the NI margin proposal from the applicants; while in another 35 out of 86 (41%) answers, CHMP recommended a stricter margin. In the remaining 9 out of 86 (10%) CHMP answers that discussed NI margin) the justification of the NI margin was questioned by CHMP, but no specific advice on its magnitude was given. The topic of switching appeared least often in question-CP units (6 % of total topics asked); and the topic of 'ITT or PP' appeared least often in CHMP-answers (3 % of total CHMP answers) (See table 3).

In addition, table 3 shows the differences and similarities in questions-CP and CHMP answers between the three most often discussed therapeutic target groups. Among anti-neoplastic and immunomodulating products and alimentary tract and metabolism products, the topics of NI data analysis (29% and 31% respectively within the therapeutic area) mostly appeared in question-CP, while for anti-infective drugs the NI margin was discussed most often (40 % of total topics asked within the therapeutic area). Among other drugs, most questions-CPs were about NI study design (28 % of total topics asked).

Among CHMP answers, the NI margin was the topic that mostly appeared in all three therapeutic target groups (38 % of total CHMP answers in anti-neoplastic and immunomodulating drugs, 28 % in alimentary and metabolism drugs and 57 % in anti-infective drugs).The NI margin was also most often discussed in the CHMP answers for other drugs (35 % of total CHMP answers).

"Extra information"(i.e. unsolicited answers) given by the CHMP, more often pertained to 'how' to do an NI study rather than 'whether' to perform an NI trial (data not shown). Only in alimentary tract and metabolism products, extra information is mostly given about 'whether' to do an NI-trial.

Discussion

Our content analysis of 2008 and 2009 scientific advices on NI trials provided by the EMA, showed that questions on 'whether' and 'how' to conduct an NI trial were frequently asked by applicants. In addition, NI margin seems to be the main concern of EMA in NI trials.

Interestingly, more than 25 % of the questions were 'whether' questions, and thus it seemed that the doubts of the company about the need of an NI trial frequently exist. These results illustrate that more explicit guidance on fundamental issues in NI trials, such as in which situation an NI trial can or should be applied are necessary. However, we realize one general guideline may not be feasible for all therapeutic areas, for example when efficacy of the current standard therapy against placebo is not fully established, e.g. anti-depressants.[13]

Our second finding shows that NI margins and data analysis were the most frequently discussed specific topics. This finding applied to all therapeutic areas. Furthermore, in 40 % of the CHMP answers on NI margins, a stricter margin was recommended. This concern was previously acknowledged by the European regulators. [14,15] The large proportion of 'how'-questions confirms that the methodology of NI-trials, in particular NI margin determination, is not straightforward.[2] These facts strengthen the need of the applicants' additional guidance on technical issues such as previously given by the EMA guidance[6] and draft FDA guidelines on NI trials.[7]

Our subgroup analysis showed that that NI trial design for alimentary tract and metabolism products is of specific concern to CHMP since CHMP often recommends a NI design for these products, without the applicant asking for guidance on this point. Apparently, in this therapeutic area, the use of NI trials to confirm drug efficacy is still complex. Recently, CHMP released revised guidance in 2011 on anti-diabetic drugs[8] which recommends beside the use of superiority trials, the use of NI trials in diabetes patients. This may help to clarify in which cases NI trials should be performed.

In the 2011 guidance on anti-diabetic drugs described above, a recommendation on a NI margin of 0.3% HbA1C was included. A similar specific requirement was previously proposed by EMA for anti-infective drugs, where a specific value of NI margin (10 %) was recommended.[16] Although the numbers are small, we found that the specific requirements still resulted in questions on the NI margin in anti-infective drugs. Recently, in a 2011 updated version, the value of 10 % was replaced by a general statement in the guideline on how an NI margin should be determined.[17] This approach is in line with the draft FDA guideline 2010[7] that recommends determining an NI margin based on historical data instead of using a single fixed value as an NI margin. Whether this new approach will lead to a reduction or an increase in scientific advice questions related to NI trials remains to be established. In the meantime, awareness of regulators about the difficulties faced by applicants is essential and dialogue between both parties, for example by means of the scientific advice process, can support the regulators in improving guidance on NI trials.

We conclude that difficulties still exist in selecting the appropriate methodology of NI trials. Straightforward and harmonized guidance on NI trials is needed, such as when to conduct NI trials and how to determine the NI margin. Regulatory guidelines (either as one general guideline or special sections on NI trials in disease-specific guidelines) may not be feasible to cover all therapeutic areas; in that case regulatory scientific advice may be used as an opportunity for tailored advice.

References

1. van Luijn JC, van Loenen AC, Gribnau FW, Leufkens HG. (2008) Choice of comparator in active control trials of new drugs Ann Pharmacother 42(11): 1605 -12.
2. Wangge G, Klungel OH, Roes KCB, de Boer A, Hoes AW, et al. (2010) Interpretation and inference in noninferiority randomized controlled trials in drug research. Clin Pharmacol

Ther 88(3): 420-423.

3. ICH Expert Working Group. (1998) Stastical principles for clinical trials. E9 . Available at : http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E9/Step4/E9_Guideline.pdf
4. ICH Expert Working Group. (2001) Choice of control group and related issues in clinical trials. E-10 .Available at : http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E10/Step4/E10_Guideline.pdf
5. Committee for Medicinal Products for Human Use (CHMP). (2005) Guideline on the choice of the non-inferiority margin. EMEA/CPMP/EWP/2158/99 . Available at : http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003636.pdf
6. Committee for Proprietary Medicinal Products (CPMP). (2000) Points to consider on switching between superiority and non-inferiority. . Available: http://www.emea.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003658.pdf
7. Center for Drug Evaluation and Research (CDER) Center for Biologics Evaluation and Research (CBER). (2010) DRAFT GUIDANCE : Guidance for industry non-inferiority clinical trials . Available at : http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM202140.pdf
8. Committee for Medicinal Products for Human Use (CHMP). (2011) Guideline on clinical investigation of medicinal products in the treatment of diabetes mellitus. CPMP/EWP/1080/00 Rev 1* . Available at : http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2011/10/WC500115945.pdf
9. Center for Drug Evaluation and Research (CDER). (2010) Guidance for industry. antibacterial drug products: Use of noninferiority trials to support approval. Available at: http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm070951.pdf .

10. European Medicines Agency. (2010) European medicines agency guidance for companies requesting scientific advice and protocol assistance. EMEA-H-4260-01-Rev 6 .Available at : http://www.ema.europa.eu/docs/en_GB/document_library/Regulatory_and_procedural_guideline/2009/10/WC500004089.pdf
11. Putzeist M, Mantel-Teeuwisse AK, Wied CCG, Hoes AW, Leufkens HG. (2010) Regulatory scientific advice in drug development: Does company size make a difference? Eur J Clin Pharmacol. 2011 Feb;67(2):157-64. Epub 2010 Nov 4.
12. WHO Collaborating Centre for Drug Statistics Methodology. (2012) ATC/DDD index. . Available: http://www.whocc.no/atc_ddd_index/ via the Internet. 31 January 2012.
13. Temple R, Ellenberg SS. (2000) Placebo-controlled trials and active-control trials in the evaluation of new treatments. part 1: Ethical and scientific issues. Annals of Internal Medicine 133(6): 455-63.
14. Brown D, Day S, Hemmings R, Wright D. (2008) Assessing the impact of ICH E9. Pharmaceutical Statistics 7(2): 77-87.
15. Lewis JA. (2002) The european regulatory experience. Stat Med 21(19): 2931-8.
16. Committee for Proprietary Medicinal Products (CPMP). (1995) Guideline on the evaluation of medicinal products indicated for treatment of bacterial infections. .
17. Committee for Medicinal Products for Human Use (CHMP). (2011) Guideline on the evaluation of medicinal products indicated for treatment of bacterial infections. CPMP/EWP/558/95 Rev 2 .Available at : http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003417.pdf

Phase IV non-inferiority trials and additional claims of benefit

Grace Wangge*

Rose Marie DLC Bernabe*

Mirjam J. Knol

Olaf H. Klungel

Johannes JM van Delden

Antonius de Boer

Arno W. Hoes

Jan Raaijmakers

Ghislaine JMW van Thiel

Submitted for publication

** equal contribution*

ABSTRACT

Background : Non-inferiority (NI) trials in drug research are used to demonstrate that a new treatment is not less effective than an active comparator. Since phase IV trials typically aim at informing a clinical decision, the value of a phase IV non-inferiority trial hinges also on its clinical relevance. In such trials, clinical relevance would refer to the added benefit claims of a specific drug, apart from efficacy, relative to its comparator drug in the trial.

Methodology and principal findings : In this study, we reviewed 41 phase IV trials and extracted information on whether the authors mentioned any additional benefit beyond the NI (efficacy) claim of the drug and whether the additional benefit was proven in the trial. We checked whether the additional claim was based on descriptions only or on formal statistical analyses. Our results showed that 22 out of the 41 NI trials mentioned additional benefit of the test drug and most of these claims were related to the safety profile. Of all the post-authorization NI trials that claimed additional benefit, 10 out of 22 NI trials used formal statistical analyses to show additional benefit, and only one included a sample size calculation for the additional benefit prior to the trial.

Conclusion : We conclude that there is room for improvement in terms of designing phase IV NI trials with added benefit claims and in proving these additional claims.

Introduction

Non-inferiority (NI) trials in drug research are used to demonstrate that a new treatment is not less effective than an active (i.e. effective) comparator.[1,2] Thus, an NI trial, which is mostly defined according to efficacy parameters, indirectly shows that the new treatment is also effective. However, the clinical significance of phase IV (i.e., "studies, other than routine surveillance, performed after drug approval and related to the approved indication"[3]) NI trials do not solely pertain to efficacy endpoints that were already established in pre-authorization trials. Rather, phase IV trials aim at "informing a decision"[4], or in ethics, such a trial should disturb equipoise, i.e., the "state of indifference or disagreement in the expert medical community about the net preferred medically established procedure".[5] As such, in principle, all NI trials should have additional benefit claims. Consequently, NI trials performed after authorizations have a reinforced obligation to make additional claims, apart from the primary (effectiveness) endpoint, for the results of such trials to be clinically relevant. Such additional claims may relate to improved safety, but also optimization of the method of administration, improved compliance, and cost-effectiveness. Since the value of late stage NI trials depends on these additional claims, appropriate study design and/or tests to demonstrate scientific validity of such claims is truly important. Whether and how these claims are scientifically justified in the NI trials currently performed is, however, unknown.

In this study, we reviewed 41 published post-authorization NI trials and determined whether these trials reported benefit claims beyond clinical efficacy and how these additional claims were supported or proven in the trials.

Methods

We included all post-authorization NI trial publications among the 232 publications used for our earlier review on NI trials.[6] In that review, we performed a search in PUBMED using the search terms, "non-inferior*", "noninferior*" or "active control and "equivalence", in combination with the MeSH term "humans" and

"Randomized Controlled Trial" as publication type. This search resulted in 669 articles and, based on pragmatic consideration rather than formal sample size calculations, we randomly selected 300 for our review. Subsequently, we excluded studies on bioequivalence, phase I studies, non-drugs trials, and articles that did not have full-text in English which resulted in 227 articles that reported 232 NI trials.

We extracted the phase of the trial according to statements in the publications or the referred clinical-trial database (e.g. clinicaltrials.gov). We could only identify the phase of 91 NI trials. Of the 91 trials, 15 were phase IV trials. For the remaining 141 NI trials, we compared the start date of the trial with the marketing approval date of the studied drug. The marketing approval dates were obtained from public domains. The first date of the marketing approval anywhere in the world was considered as the date of the drug's approval. If the trial started later than the drug's worldwide marketing approval date, we considered it a phase IV trial. Of these 141 NI trials, we identified 35 post-authorization trials. Hence, in total we found 50 post-authorization trials. We excluded trials that were aiming for the registration of a new indication (i.e., phase IIIB trials) by checking the aim of the trials stated in the article and by double-checking in the public domain via FDA and EMA websites. In total, we excluded nine phase IIIB trials. In the end, we included 41 phase IV NI trials in our analysis.

From each article, we extracted information on the type of drug, type of trial initiator, number of trial subjects, and the conclusion of the trial. We categorized the trials either as pharmaceutical-industry-initiated or non-pharmaceutical-industry-initiated. A trial is initiated by a pharmaceutical industry if besides the sponsoring there was active involvement of the pharmaceutical industry in the trial process. This involvement included any inputs of the pharmaceutical industry in writing the trial protocol, trial monitoring, data analysis, and reporting. If it is stated in the article that the pharmaceutical industry only gave unrestricted funding or grant, without any other involvement, we classified the trial as non-pharmaceutical industry-initiated.

Furthermore, we extracted information on whether the authors mentioned any additional benefit beyond the NI claim of the drug and whether the additional benefit was substantiated in the trial via descriptions (e.g., via simple distribution tables) or formal

Table 1. Characteristics of the NI trials

	N (%) (unless stated otherwise)
I. Type of Drugs	
Anti-infective	9 (22)
Cardiovascular system	9 (22)
Systemic hormonal preparations	5 (12)
Vaccines	5 (12)
Musculo-skeletal system	2 (5)
Nervous system	3 (7)
Antineoplastic	2 (5)
Others	6 (15)
II. Type of trial initiators	
Non-pharmaceutical industry	12 (29)
Pharmaceutical industry	25 (61)
Not clear	4 (10)
III. Number of trial subjects (median (interquartile range))	316 (196 -629)
IV. Conclusion of the trial	
Non-inferiority	30 (73)
Superiority	2 (5)
Inferiority	6 (15)
Others	3 (7)
V. Mentioned additional benefit	22 (54)

statistical analyses. For example, if the author mentioned that the additional benefit of the new drug was its better safety profile, we evaluated whether the safety data were presented descriptively, or if any formal testing to establish statistical significance was used to test the difference in safety profile between the two drugs. In addition, we determined whether sample size calculations for additional benefit (if any) were and extracted the authors' conclusion on the additional benefit.

Table 2 Characteristics of additional benefit claims

Additional benefit (N =22)	N	Presentation of additional benefit		Conclusion on additional benefit		
		Statistical test	Descrip-tively	Proven	Not proven	Not explicitly discussed
Convenient method of administration	1	0	0	0	0	1
Better safety profile	12	5	7	7	3	2
Better compliance	3	1	2	3	0	0
Less costly	1	0	1	0	0	1
Convenient method of administration and better safety profile	5	4	1	2	2	1

Table 3 Additional benefit claims based on types of sponsor

Type of initiators	Additional benefit (% type of sponsor)						
	Not mentioned	Convenient method of administration	Better safety profile	Better compli-ance	Less costly	Convenient method of administra-tion & better safety profile	Convenient method of administra-tion, better safety profile, better re-sistance profile
Non-pharmaceuti-cal industry (n=12)	7 (59)	0	4 (33)	0	0	0	1 (8)
Pharmaceuti-cal industry (n = 25)	11 (44)	1 (4)	8 (32)	1 (4)	1 (4)	3(12)	0
Not clear (n = 4)	1(25)	0	0	2 (50)	0	1 (25)	0

GW and RB extracted all data and, in case of discrepancies, reached consensus by discussion. All statistical analyses were performed using SPSS 19 (SPSS Inc, USA; www.spss.com).

Results

Description of the trials

Cardiovascular drugs and anti-infective drugs were the most frequently studied drugs (22 % for each; Table 1). The majority of all the trials were initiated by the pharmaceutical industry (61 %). In 73 % of the NI trials, the tested drugs were concluded to be non-inferior to their comparators.

Additional benefit

Of the 41 NI trials, 22 (54 %) mentioned additional benefit of the test drug (Table 2). Among those 22 trials, the additional benefit of “better safety profile” was most often claimed (12 trials; 55 %). Twelve trials (55 %) stated that the claimed additional benefits of the test drug were proven in the current trial. In 10 trials (45%), formal tests were used to explore statistical significance of the claimed additional benefit, but only one performed a

sample size calculation for the claimed additional benefit prior to the start of the trial (7).

Of the 25 NI trials with pharmaceutical industry involvement, 14 (56 %) mentioned additional benefit of the test drug, while among the 12 non-pharmaceutical industry initiated NI trials, five (42 %) mentioned additional benefit of the test drug (Table 3). Fourteen of the 25 NI trials with industry involvement claimed several types of additional benefit; in five of these, statistical testing was performed, while eight simply discussed the additional benefit claims, and one did not discuss the additional benefit claim at all. For the five non-pharmaceutical industry initiated NI trials that claimed additional benefit, "better safety profile" was most often claimed (four trials). Four of the five latter trials used statistical tests to explore the additional benefit claim.

Discussion

In our study of 41 phase IV NI trials, 54% reported beneficial claims in addition to the NI claim and 55% of these claims were related to safety profile. Of all post-authorization NI trials that claimed additional benefit, 45% performed tests to show statistical significance, and only one included a pre-study sample size calculation for the additional claim.

In the introduction, we stated that a phase IV trial should aim at "informing a clinical decision." We defined "informing a decision" to refer to clinically relevant differences that would allow physicians to reasonably choose one drug over another. As such, we have hinged our definition on the obligation of the physician to choose the best-suited therapy given the patient's condition. However, these clinically relevant differences also matter in the decision-making processes of the other stakeholders such as the regulators, patient groups, pharmaceutical industry, and third party payers. The importance of these clinically relevant differences is illustrated by the emergence of relative effectiveness as an important issue in the post-authorization stage, especially for third party payers such as the health insurance agencies.[4] The European Commission's High Level Pharmaceutical Forum defines relative effectiveness as "the extent to which an intervention does more good than harm compared to one or more intervention

alternatives for achieving the desired results when provided under the usual circumstances of health care practice". [4,8] Ultimately, the aim of relative effectiveness assessment is "to compare healthcare interventions in practice in order to classify them according to their practical therapeutic value".[8] We can expect this issue to sharpen as drug registration moves towards a "live license approach," i.e., an approach where launch is limited, and the widening of the scope of the license depends on post-authorization trial results. [9] In the latter case, relative effectiveness matters not only for the payers but also for the regulators. Clearly, pharmaceutical companies would need to demonstrate more than ever the added value of a new drug, or in our terms, they need to demonstrate clinically relevant differences.

Our results demonstrate that this need to establish clinically relevant differences in post-authorization NI trials through added benefit claims remains to be met. The issue is emphasized by the fact that among those that made additional benefit claims, only half used formal testing to establish statistical significance, and the other half merely presented their claims descriptively. It is questionable if it is acceptable to base decisions/ judgments of clinical relevance if claims are not sufficiently supported by evidence, such as those trials that only provide descriptions of the additional benefit claims. Some may argue that some additional benefits, such as the convenience of an oral route of administration compared to that of the intravenous route, may be obvious; hence, there is no need for evidentiary support. However, even for such claims, evidence is needed, as patients' preferences may be different. Oral route might be more convenient in the physician's perspective, but for the patient, the shape or the taste of the pill may be real issues, and therefore, the intravenous route could be better.

Apart from these scientific and regulatory issues with post-authorization NI trials without added benefit claims, or those with added benefit claims but without (or with questionable) scientific evidence, there is also an issue with the ethical justification of these trials. It is ethical for a trial to begin with the assumption of equipoise with the aim of disturbing it. Equipoise justifies the inclusion of patient-participants since the state of equipoise retains the possibility of a medically endorsable therapeutic benefit. Disturbing equipoise unambiguously establishes the value of an intervention, and hence, a trial that

aims to disturb equipoise also aims to "improve preventive, diagnostic and therapeutic interventions (methods, procedures and treatments)".[10] A trial that does not show that intervention A is in some way better than intervention B does not contribute to the improvement of therapeutic interventions. Hence, a phase IV NI trial that does not aim to assess benefit claims of the new drug does not disturb nor is it expected to disturb equipoise precisely because its goal is simply to show that A is not worse than B, and not that A is in some way better than B, a goal that does not even partly resolve the state of indifference and/or disagreement in the expert medical community. As such, a phase IV NI trial without added benefit claims may have ethical justification issues. In our study, only half of the NI trials claimed such additional benefits.

Of the 25 pharmaceutical industry-initiated trials, about half (56%) claimed multiple additional benefits. The variety of additional benefit claims made by the industry seems encouraging, as this may be a sign of how the industry tries to resolve the relative effectiveness obstacle. However, the absence of statistical testing and the reliance on mere descriptions of the alleged benefit in majority of the pharmaceutical industry initiated post-authorization NI trials bring us back to the evidence-problem we discussed earlier.

Lastly, the limited (in terms of number and variety) additional benefit claims in NI trials from independent investigators and in government initiated trials may be an indication that non-industry bodies are still generally more concerned about the narrower concepts of safety and effectiveness (as opposed to the wider benefit-risk assessment, which includes factors beyond safety and efficacy [11]). This is understandable and useful for regulatory purposes; but this situation does not help ease the impending relative efficacy and live license hurdles.

Based on the foregoing discussions, it is clear at this point that post-authorization NI trials need to be designed such that potentially, the resulting data are capable of disturbing equipoise and hence address issues such as relative effectiveness. This may be enhanced by closer and earlier collaboration between stakeholders. [12;13]

Our small sample size is a limitation of this study. In addition, clinical relevance cannot be directly investigated using our data, and as such, further research is needed.

Our study clearly shows that post-marketing NI trials vary considerably in their aims and claims. Importantly, only about half of the trials claimed additional benefit. Consequently, post-authorization NI trials need to be more robust, i.e., these trials must produce information that is directly useful to the clinical setting. Moreover, these trials must show scientific validity if they are to claim any additional value that physicians can bank on. Hence, there is room for improvement in terms of designing phase IV NI trials with additional benefit claims and in proving these additional claims.

References

(1) Committee for Medicinal Products for Human Use (CHMP). (2005) Guideline on the choice of the non-inferiority margin. EMEA/CPMP/EWP/2158/99 . Available at : http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003636.pdf

(2) Center for Drug Evaluation and Research (CDER) Center for Biologics Evaluation and Research (CBER). (2010) DRAFT GUIDANCE : Guidance for industry non-inferiority clinical trials . Available at : http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM202140.pdf

(3) International Conference on Harmonisation. General considerations for clinical trials. E8 1997. Available at : http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E8/Step4/E8_Guideline.pdf

(4) Eichler H, Bloechl-Daum B, Abadie E, Barnett D, König F, Pearson S. Relative efficacy of drugs: an emerging issue between regulatory agencies and third-party payers Nature Reviews Drug Discovery 2010 02-26.

(5) van der Graaf R, van Delden JJ. Equipoise should be amended, not abandoned. Clinical Trials 2011 August 01;8(4):408-16.

(6) Wangge G, Klungel OH, Roes KCB, de Boer A, Hoes AW, Knol MJ. Interpretation and Inference in Noninferiority Randomized Controlled Trials in Drug Research. Clin Pharmacol Ther 2010 print;88(3):420-23.

(7) Marano N, Plikaytis BD, Martin SW, Rose C, Semenova VA, Martin SK, et al. Effects of a reduced dose schedule and intramuscular administration of anthrax vaccine adsorbed on immunogenicity and safety at 7 months: a randomized trial JAMA 2008 Oct 1;300 (13):1532-43.

(8) European Commission. Core principles on relative effectiveness. 2008. Available at: http://ec.europa.eu/enterprise/sectors/healthcare/files/docs/rea_principles_en.pdf. Accessed 6/15/2012.

(9) Price Waterhouse Cooper. Pharma 2020: The vision: Which path will you take? Available at: http://www.pwc.com/gx/en/pharma-life-sciences/pharma-2020/pharma-2020-vision-path.jhtml. Accessed 6/15/2012.

(10) World Medical Association. WMA Declaration of Helsinki - Ethical Principles for Medical Research Involving Human Subjects. Available at: http://www.wma.net/en/30publications/10policies/b3/. Accessed 6/15/2012, 2012.

(11) Breckenridge A. The big society and the regulation of medicines. Available at: http://www.marionegri.it/mn/it/sezioni/intranet/docs/diapositive/03Breckenridge.pdf. Accessed 6/15/2012.

(12) Eichler HG, Pignatti F, Flamion B, Leufkens H, Breckenridge A. Balancing early market access to new drugs with the need for benefit/risk data: a mounting dilemma Nat Rev Drug Discov 2008 Oct;7(10):818-26.

(13) New development paradigms: building regulatory confidence for the early release of medicines. ; 2011. Available at : http://cirsci.org/system/files/private/CMR%20International%20Institute%20for%20Regulatory%20Science%20October%202010%20Workshop%20Synopsis_0.pdf. Accessed 6/15/2012.

General discussion

Should we ban non-inferiority trials altogether ?

Introduction

In their article published in 2007, Garattini and Bertele condemned non-inferiority (NI) trials as unethical because NI trials do not have the intention to show that a new drug is better than the standard drug and even the new drug might be worse.[1] Garattini and Bertele went as far as to suggest that the scientific community should ban NI (and equivalence) trials, even when measures are taken to improve the methodological problems inherent to NI trials.

Since that publication, however, the yearly number of published NI trials (listed in Pubmed-Medline) did not decrease, but increased from 173 to over 200. In addition, new guidelines on NI trials have been released, i.e. the CONSORT statement extension on NI trials 2008[2] and the draft FDA guideline on NI trials 2010.[3] These indicate not only a growing interest in NI trials, but also continued efforts of overcoming the methodological challenges of NI trials.

In this article, we will discuss whether the arguments given by Garattani and Bertele to ban NI trials are valid. We will focus our discussion on the ethical, methodological and regulatory aspects of NI trials.

Why should we ban NI trials?

The first reason to ban NI trials is that these trials do not have any intention to show that a new drug is better than an active standard treatment. This is considered as unethical. Why we risk trials' subjects to inconveniences and side effects when this will not lead to a better drugs on the market. It is even possible that NI trials accept that the new drug is somewhat less effective than its comparator, quantified by the NI margin, i.e. the clinically acceptable lower limit of the 95% confidence interval of the effect measure. For example, when the new drug is expected to reduce the incidence of an outcome compared to placebo, one *expects* in an NI trial comparing this drug to an *active* comparator a risk difference of 0 (or higher), while typically an lower limit of the confidence interval below 0, e.g. - 0.1 (i.e. a 10% NI margin) is accepted; thus accepting the probability that the new

drug is 10% worse than its comparator.

Many NI trials are not only performed because the new drug is thought to have a similar efficacy profile as its active comparator but also because it may offer advantages over the active comparator drug, such as a more attractive method of administration (e.g. oral instead of intravenously) or a superior safety profile.

Ideally, the additional benefit claims of a new drug should be proven superior compared to its active comparator. This can be done in an independent superiority trial or in combination with an NI trial that aims to prove NI of the drug's intended effect. In the latter situation, the trial should have sufficient power to prove both the non-inferiority of the new drugs intended effect and superiority for the additional benefit. In that sense, we do not need NI trials, because there is always a superiority counterpart in a 'so-called' NI trial, namely for the additional benefit.

Some argue that proving additional benefit might not be an issue in the situation where the aim of the NI trial is to find alternatives for patients who respond sub optimally to a standard treatment. The sub-population might respond better to a drug for the same indication but with a different mechanism of action, e.g. hypertensive Caucasians or women might respond sub optimally to diuretic therapy, but favourable to ACE inhibitors. A superiority trial comparing diuretics to ACE-inhibitors in the specific sub-population, however, seems a better alternative.

The second reason to ban NI trials is the methodological argument that an NI margin cannot be validly and objectively determined. So far, most of the efforts to overcome the methodological challenges in NI trials have concentrated on this issue. An NI margin is a clinically acceptable limit within which it can still be concluded that the new drug is similar or not worse than its comparator. Theoretically, an NI margin should be chosen in such a way that the new drug can be considered effective relative to placebo (although a placebo-controlled patient group is not included in an NI trial). This NI margin needs to account for the uncertainty in the effect size of the active control versus placebo.

Methods for determining the NI margin can vary considerably. In 22% of publications of the 232 NI trials we have reviewed, NI margins were determined merely based on subjective (clinical) considerations of the investigator. In 20 (8.7%) trials, the NI

margins were obtained from other publications or reviews. In 18 (7.7%) trials, the NI margins were obtained from available guidelines and in 17 (7.3%) trials the NI margins were calculated by the investigators based on data from previous trials.[4]

In addition, we observed that different clinical judgement and perception of the investigators play an important role in the process of determining an NI margin. Importantly, however, such, rather subjective, clinical judgment has been acknowledged by regulators as the key step in determining NI margins[3,5], since it helps in preventing *biocreep* i.e., moving gradually to less effective treatments.

In one study we used an online survey to ask 25 experts (including clinicians from academic and non-academic hospitals, regulators, and researchers in pharmaceutical industry) to choose an appropriate NI margin for a hypothetical NI trial on a new oral anticoagulant indicated for prophylaxis of venous thromboembolic events in post-orthopedic surgery. We found that a large variation in NI margins existed, even after we gave a suggestion of an NI margin following the draft FDA guideline on NI trials. Additionally, most experts provided no clinical reasoning for their choice of the NI margin. [Chapter 2.4]

The complexity of NI margin determination was also shown in questions posed by applicants who requested scientific advice from the European Medicines Agency (EMA). In our content analysis of the scientific advice pertaining to NI trials given by the EMA in 2008 and 2009, we found that questions on the NI margin were the most frequently asked questions by the applicants.[Chapter 3.1] In addition, most of the proposed NI margins from applicants were questioned or not approved by EMA. This is remarkable, since the guidelines on how the NI margin should be determined were already available prior to 2008, such as the guideline on the choice of the NI margin released by Committee for Medicinal Products for Human use (CHMP).[5]

Principally, determination of the NI margin is closely linked to the issue of assay sensitivity and constancy assumption. Assay sensitivity is the ability of a clinical trial to distinguish an effective treatment from an ineffective treatment. A drug is considered effective if it shows a significant treatment effect as compared with placebo. In a superiority trial, a significant difference between two treatments directly confirms assay sensitivity. In

an NI trial the efficacy of both drugs over placebo is not directly shown. A result of non-inferiority can be interpreted as both drugs were effective, but it could also mean that both drugs were ineffective, i.e. similar to placebo. To prove assay sensitivity, investigators should either include a placebo arm in the NI trial or discuss how they have arrived at the conclusion that the trial had assay sensitivity, for example by discussing the results of all placebo-controlled trials of the active comparator. Without such discussion, readers cannot reliably judge whether the conclusions from the trial are valid and relevant for treatment decisions. We observed that only 6.0% of NI trial included a placebo arm to evaluate assay sensitivity and none discussed assay sensitivity.[6]

An artefact of assay sensitivity that was feared by experts is *biocreep*.[7,8] Referring to the previous example, by accepting that the new drug is 10% worse than its comparator, if the drug is later being used as an active comparator in a next NI trial, the new drug may be 20% worse than the standard drug. This declining of the efficacy might continue in the next NI trials and we might end up with a drug that is actually worse than placebo.

Besides assay sensitivity, the estimated treatment effect between the active comparator and the placebo should be (still) accurate for the NI trial at hand. This is called the constancy assumption. We found only 3.9% of published NI trials discussed the constancy assumption.[6] It is important to notice that the constancy assumption cannot be assessed with total objectivity. However, it can be supported by a proper meta-analysis and by demonstration of similarity between the current trial (for example similarity in the main inclusion criteria) and the placebo-controlled trials used for setting the NI margin. Unfortunately, a meta-analysis is not a perfect solution either, since it is not always easy to decide which trials are similar “enough” to be used for NI margin determination.[9]

Why should we NOT ban NI trials?

First, the ethical argument that one should not expose patients to a drug in that does not have the intention to show that it has any additional benefits could be refuted by the fact that even when a trial was set out to prove that a new drug has additional benefit

(such as mode of administration), we still need an NI part of the study that assesses non-inferiority of the primary efficacy outcome. So, although one could claim that in any trial a superiority aim should be included, non-inferiority for other outcomes remains an important (additional) goal. As such, a requirement for superiority claims is not a good reason to ban NI aspects of trials.

The second argument not to ban NI trials is that we need a few drugs with similar efficacy in the market, so that patients, doctors and third-party payers have alternatives to choose from. NI trials provide an opportunity to test these alternative drugs, although one could argue how many drugs are clinically necessary as alternatives.

Ideally, doctors wish to choose from various drugs with similar efficacy, but with various clinically important additional benefit claims. In a review of phase IV NI trials, we showed that among industry-initiated trials, about 56% claimed multiple additional benefits. [Chapter 3.2] The variety of these additional benefit claims seems encouraging. It may be a sign of how the industry tries to answer the need for alternative drug options, albeit that there are multiple examples from the past of claimed benefits that turned out to be irrelevant for patients. Nevertheless, we have to carefully consider how many alternative drugs we actually need on the market. Two drugs with similar efficacy profile from the same class as alternatives might be enough. Regulators should have a key role in the decision whether additional alternative drugs should be released to the market.

The third argument is the possibility to overcome the main methodological limitation of NI trials; i.e. the difficulty in determining NI margins. It is clear from the previous section that the main culprits in determining a NI margin are assay sensitivity and constancy assumption. Assessment of assay sensitivity relies heavily on clinical evidence. The use of data from similar but 'outdated' placebo-controlled trials might not be avoidable, but with sufficient knowledge on the current evidence base of the drugs and the disease itself, the size of the estimated treatment effect between the active comparator and the placebo can be more accurately defined. In addition, this may lead to consensus that a specific NI margin is clinically acceptable. Thus, we also need to incorporate clinical judgment to determine an NI margin. The fear of *biocreep* seems somewhat overstated, [10,11] indicating that clinical judgment might have prevented the drugs tested with NI

trials to gradually move to less effective treatments.

The subjectivity in clinical judgment should not be a major concern as it is not solely a problem of NI trials. Superiority trials are also not free from subjectivity. Defining the smallest difference to be detected in superiority trials depends on the experience and the perspective (individual, professional, or societal) of the investigators. It may also depend on feasibility grounds.[12,13]

Efforts to reduce subjectivity have been studied more extensively in the field of social science and psychology. In clinical trials, similar methodologies could be applied. These efforts include patient's perception and use of a systematic scoring system in defining a minimal clinically important difference.[13,15] Thus, there is still room for improvement in the methodology of NI trials.

The fourth argument not to ban NI trials is the existence of many regulatory guidelines that can act as a safety net for NI trials. The first regulatory guidance on NI trials in drugs was the 'guideline on the evaluation of medicinal products indicated for treatment of bacterial infections', which was released by the European regulators in 1995. [16] It stated that each trial that is indicated for the treatment or prophylaxis of infection should be adequately powered to show at least non-inferiority to an acceptable active comparative regimen or superiority to placebo (whenever considered to be possible) or, possibly, both. In addition, it also mentioned to use an NI margin of 10% for anti-infective agents. This guideline was later followed by similar guidelines in other therapeutic areas, such as in anti-diabetic drugs.[17,18] These guidelines have been revised recently. In 2011, the 10% NI margin was no longer mentioned in the anti-infective agents' guideline. [19] In addition, general guidelines, such as the CHMP guideline on NI trials and draft FDA guideline on NI trials are available.[3,5] Beyond these guidelines, specific issues in NI trials can also be solved with dialogue between regulators and investigators/sponsors, such as via a scientific advice procedure[20] or pre-Investigational new drug (IND) consultations. [21]

Conclusion

Although NI trials may be criticized based on ethical, methodological and regulatory arguments, NI trials should NOT be banned. We can see the main reason to ban NI trials is the unethical concerns about exposing patients to drugs without the intention to show that these drugs have additional benefit. However, even when one believes that showing superiority on an outcome is an integrate part of all trials, a non-inferiority aim for another outcome could be important.

In addition, there is still ample room to improve the determination of the NI margin. To support it, dialogue with regulators to solve specific issues in NI trials could be improved, for example through scientific advice. Regulators need to have attention for unproved claims of additional benefit in NI trials to avoid misuse of the results of NI trials as a cover for unethical marketing.

To ban NI trials altogether may hinder the development of alternative drug therapies. The new oral anticoagulant drugs[22,23],may serve as an example, irrespective of whether these drugs eventually prove to be an acceptable or even preferable alternative to current standard treatment.

References

(1) Garattini S, Bertele V. Non-inferiority trials are unethical because they disregard patients' interests. The Lancet 2007;370(9602):1875-7.

(2) Piaggio G. Reporting of Noninferiority and Equivalence Randomized Trials: An Extension of the CONSORT Statement JAMA: The Journal of the American Medical Association 2006;295(10):1152-60.

(3) Center for Drug Evaluation and Research (CDER) Center for Biologics Evaluation and Research (CBER). DRAFT GUIDANCE : Guidance for Industry Non-Inferiority Clinical Trials 2010.

(4) Wangge G, Klungel OH, Roes KCB, de Boer A, Hoes AW, Knol MJ. Room for Improvement in Conducting and Reporting Non-Inferiority Randomized Controlled Trials on Drugs: A Systematic Review. PLoS ONE 2010 10/27;5(10):e13550.

(5) Committee for Medicinal Products for Human Use (CHMP). Guideline on the choice of

the non-inferiority margin. EMEA/CPMP/EWP/2158/99 2005.

(6) Wangge G, Klungel OH, Roes KCB, de Boer A, Hoes AW, Knol MJ. Interpretation and Inference in Noninferiority Randomized Controlled Trials in Drug Research. Clin Pharmacol Ther 2010 print;88(3):420-423.

(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 Jan 30;22 (2):169-186.

(8) Fleming TR. Current issues in non-inferiority trials. Stat Med 2008;27(3):317-32.

(9) Wangge G, Roes KC, de Boer A, Hoes AW, Knol MJ. The challenges of determining noninferiority margins: a case study of noninferiority randomized controlled trials of novel oral anticoagulants. CMAJ 2012 Aug 20.

(10) Beryl P, Vach W. Is there a danger of "biocreep" with non-inferiority trials? Trials 2011 Dec 13;12 Suppl 1:A29.

(11) Everson-Stewart S, Emerson SS. Bio-creep in non-inferiority clinical trials Stat Med 2010; 29(27):2769-80.

(12) Gayet-Ageron A, Agoritsas T, Combescure C, Bagamery K, Courvoisier D, Perneger T. What differences are detected by superiority trials or ruled out by noninferiority trials? A cross-sectional study on a random sample of two-hundred two-arms parallel group randomized clinical trials. BMC Med Res Methodol. 2010 Oct 15;10:93.

(13) Man-Son-Hing M, Laupacis A, O'Rourke K, Molnar FJ, Mahon J, Chan KB, et al. Determination of the clinical importance of study results. J Gen Intern Med 2002 Jun;17(6):469-476.

(14) Peterson, LS. "Clinical" Significance: "Clinical" Significance and "Practical" Significance are NOT the Same Things. ; 2008. Paper presentation in Annual meeting of the Southwest Educational Research Association,New Orleans, February 7, 2008.

(15) Campbell TC. An Introduction to Clinical Significance: An Alternative Index of Intervention Effect for Group Experimental Designs. Journal of Early Intervention 2005;27(3):210-27.

(16) Committee for Proprietary Medicinal Products (CPMP). Guideline on the evaluation of medicinal products indicated for treatment of bacterial infections. 1995.

(17) Committee for Medicinal Products for Human Use (CHMP). Guideline on clinical investigation of medicinal products in the treatment of diabetes mellitus. CPMP/ EWP/1080/00 Rev 1* 2011.

(18) Center for Drug Evaluation and Research (CDER). Diabetes Mellitus: Developing Drugs and Therapeutic Biologics for Treatment and PreventionDRAFT GUIDANCE. 2008.

(19) Committee for Medicinal Products for Human Use (CHMP). Guideline on the

evaluation of medicinal products indicated for treatment of bacterial infections. CPMP/EWP/558/95 rev 2 2011.

(20) European Medicines Agency. European Medicines Agency Guidance for Companies requesting Scientific Advice and Protocol Assistance. EMEA-H-4260-01-Rev 6 2010.

(21) Pre-IND Consultation Program Available at: http://www.fda.gov/Drugs/DevelopmentApprovalProcess/HowDrugsareDevelopedandApproved/ApprovalApplications/InvestigationalNewDrugINDApplication/Overview/default.htm. Accessed 5/9/2012,.

(22) Miller CS, Grandi SM, Shimony A, Filion KB, Eisenberg MJ. Meta-Analysis of Efficacy and Safety of New Oral Anticoagulants (Dabigatran, Rivaroxaban, Apixaban) Versus Warfarin in Patients With Atrial Fibrillation. Am J Cardiol. 2012 Aug 1;110(3):453-60.

(23) Sobieraj DM, Coleman CI, Tongbram V, Chen W, Colby J, Lee S, et al. Comparative Effectiveness of Low-Molecular-Weight Heparins versus Other Anticoagulants in Major Orthopedic Surgery: A Systematic Review and Meta-analysis. Pharmacotherapy. 2012 Jun 28. doi: 10.1002/j.1875-9114.2012.01106.x. [Epub ahead of print]

Summary

Ringkasan

amenvatting

Summary

A randomized clinical trial (RCT) is the gold standard to evaluate the intended effects of drugs. In such trials a drug can be compared with a placebo or with another active compound for the same indication. RCTs can be used to demonstrate that a drug is superior to placebo or an active comparator (superiority trial) or that a drug is not worse than an active comparator (non-inferiority (NI) trial).

NI trials can be used in a situation when a new drug is expected to have a similar efficacy as its comparator but may offer other advantages over the existing drug such as a more convinient method of administration (for example oral drugs versus parenteral route) or has less side effects.

The concept of NI trials has been developed in the 1970s and was inspired by the methodology of (bio) equivalence trials. In bioequivalence trials, two medicinal products containing the same active substance are compared for their bioavailability, typically an original drug and a generic drug that contains the same active substance are compared. During that time, the terms non-inferiority and bioequivalence were used interchangeably. NI trials became popular in the 1990s, especially after the introduction of several regulatory guidelines on active-controlled trials. This is shown by a major increase of publications on NI trials since the release of the first guideline, the International conference on harmonization (ICH) E9, in 1998.

Currently, other regulatory guidelines on NI trials are available, such as the ICH E10, Committee for Medicinal Products for Human Use (CHMP) guidelines on the choice of NI margin, and the US Food and drugs administration (FDA) draft guideline on NI trials. Furthermore, the Consolidated Standards of Reporting Trials (CONSORT) organization has provided a specific guideline on how to report NI trials.

From a methodological perspective, NI trials have special challenges in design and analysis that can influence proper interpretation of its result. First, there are different methods to determine the limit where we can say that the drug is not worse than its active comparator (NI margin). Second, there is a difficulty in interpreting NI trials because of their lack of ability to distinguish an effective drug from an ineffective drug i.e. assay sensitivity, without relying on evidence outside the trial. Third, the validity of the historical data that were used to base the NI margin on, i.e. constancy assumption sometimes is questionable.

The objective of this thesis was to look deeper into the challenges of the methodology of NI trials, and the role of regulatory guidelines in it.

This objective is in line with the context of Escher (T6-202), a project from the Top Institute Pharma, the umbrella of this project. 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 manner. The Escher project focuses on delivering evidence and credibility for regulatory reform and policy recommendations in drug research.

The first part of this thesis (**Chapter 2**) focuses on the methodological challenges of NI trials. In **chapters 2.1** and **2.2**, we give an overview how an NI trial can be designed and interpreted. We systematically reviewed the publications of 232 NI trials to identify how NI trials are currently designed, analyzed, and reported.

We found that a large proportion of NI trials did not use blinding, which is essential for unbiased interpretation of any RCT. The NI margin was reported in almost all of the trials, but less than half of the NI trials reported the method to determine their NI margins. In 22 % of the publications of the NI trials, the margin was merely based on subjective considerations of the investigator without any explanation whether statistical or clinical reasoning was used.

Furthermore, we found that approximately 8% of the trials were interpreted incorrectly, and less than 10% of the trials discussed assay sensitivity or the validity of the constancy assumption. No difference was seen in the quality of reporting before and after the release of the CONSORT statement extension 2006 or between the high-impact and low-impact journals.

These findings provide evidence that the interpretation of and inference from NI trials are complicated, and publications do not routinely contain the information needed, especially information on how the NI margin was determined.

The guidelines on NI trials state that NI margins should be based on both statistical and clinical judgments. The recently issued US FDA draft guideline on NI trials provides a clearer insight on how a margin can be defined based on statistical judgment, but not on clinical judgment, while acceptance of a specific NI margin by regulators and by clinicians, involves clinical consideration. In **chapter 2.3** we explain in more detail the method to determine an NI margin based on statistical judgment. We used a case study of NI margins used in trials on novel oral anticoagulants that are indicated as prophylaxis of venous thromboembolism after orthopedic surgery patients.

From the case study we learned that there is a large variation in NI margins used in the NI trials of oral anticoagulants, which could lead to inconsistent conclusions on non-inferiority, and thus efficacy versus placebo, of the studied drugs. This is undesirable both

from a clinical and regulatory perspective.

In addition, we noted some issues that are not explicitly described in the draft FDA guidelines yet, including the amount of effect that should be preserved, how similar the characteristics of the placebo-controlled trials and NI trials need to be, and whether a risk difference or relative risk should or could be used to calculate the margin.

Next to statistical judgment, clinical judgment from the investigator plays an important role in determining an NI margin. How investigators incorporate this clinical judgment remains unknown. These implicit clinical judgments might have been derived from clinical experience.

However, these judgments remain subjective and different clinicians may propose contradicting judgments. Thus, it is important to study how this clinical judgment can be incorporated in the NI margin determination. In **chapter 2.4**, using an online questionnaire and a case scenario of a hypothetical new anticoagulant that should be tested against enoxaparin for prophylaxis of venous thromboembolism after orthopedic surgery, we gathered experts' opinions on a clinically relevant NI margin. We evaluated how experts' (clinical) considerations guide their choice of NI margin. Experts were asked to give their choices of the NI margin in two study-sections: before and after additional information on the statistical NI margin was presented. Furthermore, we asked the experts to motivate their choice for why they chose a specific NI margin.

Again, we found a large variation in NI margins provided by the 25 experts included in our study. More experts gave clear reasons underlying their choice of NI margin when additional information on the statistical NI margin was presented. Among the reasons were the NI margin was chosen because the experts want it to be as narrow as possible; the NI margin was copied from previous NI trials and the NI margin was considered "clinically relevant". We concluded that presently subjectivity plays an important role in the determination of NI margins. In order to increase objectivity, more guidance is needed to improve adequate and consistent determination of clinically acceptable NI margins.

The second part of this thesis (**chapter 3**) focuses on the regulatory perspective of NI trials. In **chapter 3.1** we reported the content analysis of final-advice letters of scientific advices (SA) from the European Medicines Agency (EMA). The scientific advice is a medium where regulators provide opportunities for companies to discuss critical trial issues prior to the trial's conduct. We identified questions on NI trials that were posed by companies in 2008 and 2009, and the responses given by the EMA to identify potential issues that may benefit from a more explicit guidance.

Our evaluation yielded two major findings. First, questions 'whether' and 'how' to

conduct an NI trial were frequently asked by applicants (60 % out of 587 different topics that were asked for), but 'how' questions (74% of total questions asked and 72% of total EMA responses) were more frequent than 'whether' questions (26% of total questions asked and 28 % of total EMA responses). Second, the choice of the NI margin seems to be EMA's main concern in NI trials. In 40 % of the EMA answers, they recommended the use of a stricter margin; and in 10 % of the EMA answers on NI margins, they questioned the justification of the proposed NI margin. We conclude that difficulties still exist in selecting the appropriate methodology of NI trials.

Straightforward and harmonized guidance on NI trials is needed, such as when to conduct NI trials and how to determine the NI margin. Regulatory guidelines (either as one general guideline or special sections on NI trials in disease-specific guidelines) may not be feasible to cover all therapeutic areas; in that case regulatory scientific advice may be used as an opportunity for tailored advice.

The value of NI trial hinges on its clinical relevance. Clinical relevance refers to the added benefit claims of a specific drug, apart from efficacy, relative to its comparator drug in the trial. Since phase IV (post-authorization) trials typically aim at informing a clinical decision, these trials ideally should aim at "informing a clinical decision." We defined "informing a decision" to refer to clinically relevant differences that would allow physicians (and also regulators, patient groups, pharmaceutical industry or third-party payers) to reasonably choose one drug over another. Thus, in **chapter 3.2**, we looked at 41 phase IV NI trials' publications found in our earlier review (chapter 2.1 and 2.2) and extracted information on whether the authors mentioned any additional benefit beyond the NI claim of the drug and whether the additional benefit was proven in the trial.

Our results demonstrate that the need to establish clinically relevant differences in phase IV NI trials has yet to be met. The issue is emphasized by the fact that among 54 % of phase IV NI trials that made additional benefit claims, only half used formal testing to establish statistical significance and the other half merely described these claims. Obviously it is not acceptable to base decisions on clinical relevance if claims are not sufficiently supported by valid evidence.

Furthermore, we found that the pharmaceutical industry-initiated trials claimed a variety of additional benefits while non-pharmaceutical initiated trials mostly claimed additional safety benefit. The variety of additional benefit claims made by the industry seems encouraging, as this may be a sign of how the industry tries to resolve the clinically relevant needs. However, the absence of statistical tests and the reliance on mere descriptions of the claimed additional benefit in the NI trials initiated by pharmaceutical industry has made this 'good' intention questionable.

Finally, this thesis ends with a discussion (**chapter 4**) where we discussed whether we should or should not ban NI trials. We based our discussion on arguments by Garattini and Bertele in 2007 that condemned NI trials as unethical because they do not have the intention to show that a new drug is better than the standard drug. Garattini and Bertele went as far as to suggest that the scientific community should ban NI trials, even when measures are taken to improve the methodological problems inherent to NI trials.

We discussed the advantages and disadvantages of NI trials based on ethical, methodological and regulatory arguments. We suggest that the non-inferiority aim of showing efficacy of a drug in NI trials should be studied together with aims directed at showing superiority of aspects of the drug over its active comparator e.g. on side effects, or ease of use of drugs. Furthermore, these other study aims should be studied in a scientific sound way (e.g. with adequate statistical power). When these conditions have been met we think it is not unethical to perform NI trials.

In summary, we think there is still ample room to improve the determination of the NI margin. Additionally, regulators need to have attention for unproven claims of additional benefit in NI trials to avoid misuse of the results of NI trials as a cover for unethical marketing.

Ringkasan

Untuk membuktikan apakah suatu obat baru benar-benar efektif, uji klinis terkontrol acak buta berganda merupakan metode yang terbaik. Dalam uji ini, pengacakan membuat karakteristik kelompok pasien yang mendapatkan obat baru yang diuji setara dengan karakteristik kelompok pasien yang mendapatkan obat pembanding. Kesetaraan ini akan membuat perbedaan efektifitas yang ditemui (berdasarkan hasil perhitungan statistik) diantara kedua kelompok pasien dalam uji, bermakna. Buta berganda berarti peneliti maupun pasien tidak mengetahui obat mana yang dikonsumsi, sehingga mengurangi subyektifitas dalam penafsiran hasil akhir terapi.

Tujuan utama uji klinis adalah menunjukkan bahwa obat dengan indikasi tertentu lebih efektif daripada plasebo (obat tanpa zat aktif) atau obat lain. Hal ini dikenal sebagai uji superioritas. Akan tetapi obat juga dapat diuji untuk menunjukkan bahwa obat yang baru tidak lebih buruk efektifitasnya dibandingkan obat standar. Uji seperti ini dikenal sebagai uji non-inferioritas (NI). Penting diingat bahwa dalam setiap uji klinis terkontrol, efek samping yang tidak diinginkan juga selalu diamati dan dipelajari.

Uji NI dilakukan, jika obat baru diharapkan memiliki khasiat yang sama dengan obat lainnya, tetapi menawarkan keuntungan yang lain, misalnya mempunyai cara pemakaian yang lebih nyaman untuk pasien (obat minum dibandingkan obat suntik) atau memiliki efek samping yang lebih sedikit.

Uji NI berkembang pada tahun 1970-an dan terinspirasi dari metodologi yang dipakai dalam uji bioekivalensi. Dalam uji tersebut, dua obat yang memiliki molekul aktif yang sama, tetapi berbeda bentuk (misalnya, kapsul dan tablet) saling dibandingkan bioavailabilitas-nya (kadar zat aktif yang terkandung dalam darah) pada orang sehat. Uji ini biasanya dilakukan pada obat generik.

Jika kita melihat berdasarkan jumlah publikasi uji NI per tahunnya, terlihat bahwa uji ini menjadi sangat populer sejak tahun 1990-an. Hal ini sebagian terkait dengan dikeluarkannya pedoman-pedoman regulasi yang mengatur uji NI. Pedoman-pedoman ini berasal dari badan regulasi Eropa dan Amerika Serikat (*European medicines agency (EMA)* dan *US Food Drug and Administration (US-FDA)*).

Dari sudut pandang metodologis, percobaan NI memiliki tantangan dalam desain dan analisis yang dapat mempengaruhi penafsiran hasil akhirnya. Pertama, banyak metode yang berbeda untuk menentukan batas di mana kita dapat mengatakan bahwa obat tidak lebih buruk dari komparator aktif (marjin NI). Kedua, ada kesulitan dalam menafsirkan uji NI karena kurangnya kemampuan uji ini untuk membedakan antara obat yang efektif dari obat yang tidak efektif (dikenal sebagai *assay sensitivity*), tanpa bergantung pada bukti lain

di luar situasi uji klinis itu sendiri. Ketiga, data historis yang digunakan sebagai dasar untuk menentukan marjin NI, yaitu asumsi kekonstanan (*constancy assumption*), kadang-kadang diragukan kebenarannya.

Tesis ini bertujuan untuk melihat secara lebih mendalam tantangan dalam metodologi uji NI, dan peran pedoman regulasi di dalamnya, sehingga kualitas uji NI dapat ditingkatkan di masa mendatang. Tujuan ini sejalan dengan konteks proyek Escher (T6-202), sebuah proyek dari *Top Institut Pharma,* Belanda yang merupakan payung penelitian proyek ini. Proyek Escher menyatukan universitas dan industri farmasi dengan tujuan untuk memberikan motivasi bagi berkembangnya riset dan pengembangan farmasi.

Dalam **bab 2.1** dan **2.2**, kami memberikan gambaran umum bagaimana uji NI dirancang dan diinterpretasikan. Menggunakan review sistematik, kami menganalisis publikasi dari 232 uji NI untuk mengidentifikasi bagaimana uji NI saat ini dirancang, dianalisis, dan dilaporkan.

Kami menemukan bahwa dalam 34 % uji NI, peneliti dan pasien tidak buta terhadap jenis obat yang diterima. Padahal hal ini penting untuk mencegah bias dalam interpretasi hasil akhir setiap penelitian. Sebagian besar publikasi melaporkan marjin NI yang mereka gunakan, tetapi kurang dari setengah jumlah publikasi tersebut melaporkan bagaimana cara mereka menentukan marjin NI. Dalam 22% publikasi uji NI, marjin ditentukan hanya berdasarkan pertimbangan klinis peneliti. Selanjutnya, kami menemukan bahwa sekitari 8% hasil penelitian diinterpretasikan secara tidak tepat, dan kurang dari 10% dari uji NI membahas *assay sensitivity* atau asumsi kekonstanan. Tidak ada perbedaan yang terlihat pada kualitas publikasi sebelum dan setelah rilisnya pedoman *CONSORT* tahun 2006 atau di antara publikasi yang dimuat dalam jurnal ilmiah berdampak-tinggi dan berdampak rendah (*High impact* dan *low impact factor journals*).

Temuan ini memberikan bukti rumitnya penafsiran dan kesimpulan dari uji NI. Hal ini diperberat dengan kenyataan bahwa publikasi uji NI sering tidak memuat informasi yang dibutuhkan, terutama informasi tentang bagaimana marjin NI ditentukan.

Pedoman-pedoman regulasi mengenai uji NI memang memberikan rekomendasi umum bagaimana marjin NI dapat dihitung. Akan tetapi menarik untuk diamati bahwa, pedoman-pedoman ini, seperti pedoman yang cukup detil dari *US FDA*, ber fokus pada bagaimana marjin NI dapat ditentukan berdasarkan pertimbangan statistik namun kurang memperhatikan pertimbangan klinis. Pertimbangan klinis justru penting karena merupakan sarana yang baik untuk memperkirakan efek obat yang sebenarnya di klinik.

Dalam **bab 2.3** kami menjelaskan bagaimana menentukan marjin NI berdasarkan pertimbangan statistik. Sebagai contoh kasus kami menggunakan uji NI pada obat antikoagulan oral baru dengan indikasi pencegahan penyakit tromboemboli setelah

operasi bedah tulang yang dibandingkan dengan pengobatan standar. Ternyata kami menemukan bahwa marjin NI yang digunakan dalam uji-uji ini sangat bervariasi. Hal ini dapat mengakibatkan kesimpulan yang salah tentang apakah obat baru yang diuji benar-benar tidak lebih buruk dari obat kontrol dan sejauh mana obat baru ini sebenarnya lebih efektif daripada plasebo. Selanjutnya, bab ini berfokus beberapa isu yang tidak secara eksplisit dijelaskan dalam rancangan pedoman *US FDA* tentang uji NI, termasuk berapa besar efek antara obat standar dan placebo yang harus dipertahankan,jenis *effect measure* yang harus digunakan (risiko relatif atau perbedaan risiko absolut) dan asumsi kekonstanan.

Seperti dikemukakan sebelumnya, di samping pertimbangan statistik, pertimbangan klinis mempunyai peran penting dalam menentukan marjin NI. Bagaimana dalam prakteknya pertimbangan klinis diperhitungkan ketika menentukan margin NI tidaklah jelas. Untuk menyelidiki hal ini, lebih lanjut dalam **bab 2.4**, kami menyajikan hasil survei menggunakan kuesioner melalui internet. Dalam studi ini, para ahli menjawab pertanyaan berdasarkan sebuah kasus hipotetikal tentang antikoagulan baru yang diuji terhadap enoxaparin untuk profilaksis penyakit tromboemboli setelah operasi bedah tulang. Untuk menentukan ini, para ahli diminta untuk memberikan pilihan marjin NI mereka dalam dua bagian: sebelum dan sesudah informasi tambahan tentang marjin NI yang telah ditentukan secara statistik diberikan. Selanjutnya, kami meminta para ahli menyebutkan alasan mereka memilih marjin NI tersebut.

Sekali lagi, kami menemukan variasi yang besar dalam pilihan marjin NI dari 25 ahli yang ikut dalam penelitian kami. Lebih banyak ahli memberikan alasan jelas mengenai pilihan marjin NI mereka setelah informasi tentang marjin NI yang ditentukan berdasarkan pertimbangan statistik diberikan. Kami menyimpulkan subjektivitas memainkan peran penting dalam penentuan marjin NI. Untuk meningkatkan objektivitas, diperlukan regulasi bagaimana menentukan marjin NI yang dapat diterima secara klinis. Hal ini dapat membantu terciptanya marjin NI yang lebih konsisten dan bermakna secara klinis.

Bab 3.1 dan **3.2** membahas aspek regulasi uji NI. Dalam **bab 3.1** disajikan sebuah studi dimana kami membahas saran yang diberikan oleh *EMA* untuk industri farmasi dalam pelaksanaan uji NI. Saran-saran ini berasal dari tahun 2008 dan 2009. Kami menemukan bahwa industry farmasi sering bertanya tentang bagaimana uji NIseharusnya dilakukan dan bagaimana uji NI harus dirancang. Dalam 40% dari jawaban EMA, mereka merekomendasikan penggunaan marjin yang lebih ketat daripada yang diajukan pemohon. Dalam 10% dari jawaban EMA pada marjin NI, mereka mempertanyakan alasan pemilihan marjin NI yang diusulkan. Penelitian ini menunjukkan bahwa industri farmasi juga memiliki banyak pertanyaan metodologis tentang pelaksanaan uji NI.

Perbaikan pedoman regulasi mungkin diperlukan untuk mengurangi ambiguitas dari industri farmasi dalam pelaksanaan uji NI, namun saran dari otoritas regulasi juga tetap diperlukan terutama ketika uji NI akan dilakukan pada area terapi baru.

Seperti telah diungkapkan sebelumnya relevansi klinis dari uji NI merupakan masalah penting. Relevansi klinis mengacu pada klaim akan manfaat tambahan obat tertentu, selain efikasi, relatif terhadap obat pembanding misalnya efek samping yang lebih ringan atau manfaat lain dari obat baru yang memberikan nilai terapeutik tambahan. Hal ini penting terutama pada uji NI yang dilakukan pada fase 4, atau uji obat yang dilakukan setelah dikeluarkannya ijin peredaran suatu obat. Obat baru yang akan diedarkan di pasaran seyogyanya tidak hanya memiliki efikasi yang sama dengan obat standar tetapi juga harus dapat memberikan nilai terapeutik tambahan tertentu. Dengan demikian, uji fase 4 harus memberikan informasi yang relevan untuk memungkinkan dokter (dan juga regulator, kelompok pasien, industri farmasi atau pihak asuransi) untuk memilih obat yang paling tepat bagi pasien dari semua alternatif obat yang ada.

Dalam **bab 3.2**, kami menyajikan analisa 41 uji NI fase 4 untuk melihat sejauh mana uji-uji NI menggambarkan manfaat tambahan selain klaim bahwa obat bersifat non-inferior terhadap obat pembanding dan apakah klaim tersebut terbukti dalam uji yang dilakukan. Ternyata diantara uji yang mengklaim manfaat tambahan, hanya setengah melakukan uji statistik (54% dari 41 uji NI). Selanjutnya, kami menemukan setengah dari percobaan yang dilakukan oleh industri farmasi mengklaim berbagai manfaat teraupetik tambahan sementara lembaga non-industri farmasi pada umumnya lebih peduli mengenai masalah keamanan obat . Berbagai klaim tambahan yang dibuat oleh industri tampaknya menggembirakan, karena hal ini menjadi tanda bagaimana industri farmasi mencoba untuk kebutuhan akan informasi yang dapat membantu keputusan klinis. Namun, tanpa disertai adanya uji statistik dan penggambaran manfaat tambahan yang hanya dilakukan secara deskriptif membuat timbulnya kekhawatiran penyalahgunaan uji-uji NI ini untuk tujuan pemasaran semata.

Tesis ini diakhiri dengan sebuah diskusi (**bab 4**) di mana kami membahas sejauh mana kebutuhan kita akan uji NI. Pada tahun 2007, Garattini dan Bertele mengutuk uji NI tidak etis karena tidak memiliki niat untuk menunjukkan bahwa obat baru lebih baik dari obat standar. Garattini dan Bertele menunjukkan bahwa komunitas ilmiah harus melenyapkan uji NI, bahkan ketika banyak upaya dilakukan untuk memperbaiki masalah metodologis yang terkait dengan uji NI.

Dalam diskusi kami membahas pro dan kontra dari uji NI dan mengakhiri perdebatan ini dengan kesimpulan bahwa setiap uji NI yang dilakukan harus disertai dengan pengujian akan kebenaran nilai terapeutik tambahan yang diklaim. Klaim-klaim

tersebut harus dibuktikan secara ilmiah, sama dengan klaim NI. Jika kondisi ini terpenuhi, kita dapat mengatakan uji NI sebagai etis dan relevan dilakukan dalam proses pengembangan farmasi.

Secara garis besar dapat disimpulkan, terdapat tantangan dalam metodologi dan regulasi yang harus dihadapi jika kita ingin memperbaiki kualitas uji NI. Selain membuktikan bahwa suatu obat bersifat non-inferior, pembuktian nilai terapeutik dari obat tersebut harus ditentukan berdasarkan uji statistik yang memadai.

Samenvatting

Voor het bestuderen van de effectiviteit van een geneesmiddel is het gerandomiseerde dubbel blinde onderzoek de beste methode. In een dergelijk onderzoek is de randomisatie, het geen betekent dat het lot bepaalt wie welke behandeling krijgt, bedoeld om de groepen prognostisch vergelijkbaar te maken voor de uitkomst die wordt bestudeerd. Als de behandelgroepen prognostisch vergelijkbaar zijn, zijn eventuele verschillen in effecten die met voldoende statistische zekerheid zijn aangetoond in principe toe te schrijven aan de toegepaste geneesmiddelen. De blindering zorgt ervoor dat de onderzoeker noch de patiënt weet wie welk middel gebruikt zodat patiënten zich in de behandelgroepen niet verschillend gaan gedragen (gedrag dat de uitkomst van de studie kan beïnvloeden) en de patiënten en onderzoekers eventuele effecten van de middelen niet verschillend interpreteren (subjectiviteit trachten uit te sluiten). Deze maatregelen hebben dus tot doel de effecten van geneesmiddelen kwalitatief en kwantitatief zo goed mogelijk te schatten.

Het primaire doel van een gerandomiseerd onderzoek is meestal te laten zien dat het geneesmiddel voor een bepaalde indicatie effectiever is dan placebo of een ander actief middel. Dit noemen we ook wel " superiority trials". Ook kan het doel zijn te laten zien dat een middel niet minder effectief is dan een ander geneesmiddel. Dit zijn de zogenaamde "*non-inferiority (NI) trials*". Van belang is aan te geven dat in gerandomiseerde onderzoeken ook altijd ongewenste effecten ook wel bijwerkingen worden bestudeerd.

*NI trial*s worden uitgevoerd indien van een geneesmiddel wordt verwacht dat het een vergelijkbare effectiviteit heeft als een ander middel maar op andere vlakken voordelen biedt zoals bijvoorbeeld een makkelijkere toedieningsweg (via de mond in plaats van injecties) of minder bijwerkingen.

De *NI trial* als methode is in de zeventiger jaren ontwikkeld en was geïnspireerd door de methodologie van de zogenaamde bio-equivalentie studies. In dergelijke studies worden twee geneesmiddelen die hetzelfde actieve molecuul bezitten maar een andere toedieningsvorm (bijvoorbeeld een capsule versus een tablet) met elkaar vergeleken voor wat betreft hun biologische beschikbaarheid (de hoeveelheid van de actieve stof die in het bloed komt) bij gezonde mensen. Deze studies worden meestal uitgevoerd indien een goedkopere variant (generiek geneesmiddel) van een merkgeneesmiddel wordt ontwikkeld.

Gebaseerd op het aantal publicaties waarin de resultaten van *NI trial*s worden gepresenteerd blijkt deze studieopzet met name vanaf de negentiger jaren populair. Dit is

mede gerelateerd aan de publicatie van een aantal richtlijnen met adviezen over *NI trial*s. Deze richtlijnen zijn afkomstig van de Europese en Amerikaanse registratie-autoriteiten, respectievelijk de European Medicines Agency (EMA) en de Food and Drug Administration (FDA).

Vanuit een methodologisch perspectief zijn er een aantal uitdagingen bij de opzet, analyse en interpretatie van *NI trial*s. Allereerst moet een grens worden vastgesteld om te kunnen beslissen of een geneesmiddel niet minder effectief is dan het middel waarmee het wordt vergeleken. In het Engels wordt deze grens ook wel de "*NI margin*" genoemd. Voor het bepalen van deze grens zijn verschillende methoden beschikbaar. Ten tweede is het lastig om in een *NI trial* goed vast te stellen wat het effect van een geneesmiddel is. Indien een geneesmiddel met een placebo wordt vergeleken is goed vast te stellen wat de effectiviteit van een geneesmiddel is, wat het middel meer doet dan placebo. In een *NI trial* wordt een geneesmiddel met een ander geneesmiddel vergeleken. Omdat de placebo ontbreekt wordt de effectiviteit indirect vastgesteld. Indien we van het controle middel weten dat het beter was dan placebo kun je indirect afleiden dat als het middel niet slechter is dat het controle middel dat het middel ook beter is dan placebo. Of je met een studie op indirecte wijze betrouwbaar kunt bestuderen of het middel beter is dan placebo wordt "*assay sensitivity*" genoemd. Ten slotte is een belangrijk probleem dat goed nagegaan moet worden of het veronderstelde effect van het controle middel ten opzichte van placebo ook geldt in de huidige *NI trial*. Mogelijk is dit effect anders indien de patiëntengroep anders is (bijvoorbeeld meer ouderen of patiënten met andere ziektes) of de betreffende aandoening ook al op andere wijzen beter wordt behandeld. De aanname dat een controle middel hetzelfde effect heeft ten opzichte van placebo in de huidige *NI trial* in vergelijking met de studie waarin dit eerder daadwerkelijk is bestudeerd is de zogenaamde "*constancy assumption*".

Het doel van dit proefschrift was de methodologische en regulatoire uitdagingen van *NI trial*s verder te bestuderen ten einde de toepassing van deze trials in de toekomst zonodig te verbeteren.

Dit project is onderdeel van een groter project, het zogenaamde Escher project van TI Pharma. Het Escher project heeft tot doel methodologische problemen en barrières op het gebied van regelgeving bij de ontwikkeling van geneesmiddelen te identificeren en aanbevelingen te doen ter verbetering.

In de **hoofdstukken 2.1 en 2.2** staan de methodologische uitdagingen van de *NI trial* centraal. In deze hoofdstukken wordt uiteengezet hoe *NI trials* worden opgezet en geïnterpreteerd. Tevens werden 232 gepubliceerde *NI trials* systematisch geëvalueerd om na te gaan hoe zij waren opgezet, geanalyseerd en gerapporteerd.

Het viel op dat 34% van de trials niet waren geblindeerd hetgeen tot vertekening van de resultaten kan leiden. In bijna alle trials werd de *NI margin* gepresenteerd maar slechts bij 46% werd uitgelegd hoe de grens was vastgesteld. Het bleek dat bij 22% van de trials de onderzoekers hun eigen criteria daarvoor gebruikten zonder nadere uitleg over statistische en klinische overwegingen. Verder bleek dat in 8% van de studies op basis van de *NI margin* en de studieresultaten een verkeerde conclusie over het al of niet inferieur zijn van het middel werd getrokken. Ook werd in minder dan 10% van de studies aandacht besteed aan de *assay sensitivity* en de *constancy assumption*. Als deze niet worden bediscussieerd is het in feite niet goed beoordeelbaar of het onderzochte middel effectiever is dan placebo en dus al helemaal niet of het effect van het middel klinisch relevant is. Een analyse om na te gaan of er verschillen waren in de kwaliteit van het rapporteren van *NI trials* voorafgaand en na publicatie van een belangrijke richtlijn over *NI trials* (CONSORT statement) in 2006 en in hoge kwaliteit versus lage kwaliteit tijdschriften gaf aan dat deze verschillen er niet waren.

Deze bevindingen geven dus aan dat de interpretatie van *NI trial*s niet eenvoudig is en dat in publicaties relevante informatie om de klinische relevantie van de bevindingen te kunnen inschatten niet worden gepresenteerd. De richtlijnen over *NI trial*s doen aanbevelingen hoe een *NI margin* kan worden vastgesteld. Interessant genoeg geeft de belangrijke FDA richtlijn vooral hoe de *NI margin* op grond van statistische overwegingen kan worden bepaald en minder aandacht aan klinische overwegingen. Deze klinische overwegingen zijn nu juist belangrijk voor zorgverleners om de waarde van het middel voor de klinische praktijk goed in te kunnen schatten.

In **hoofdstuk 2.3** wordt in eerste instantie op dit onderwerp verder doorgegaan door nader uit te leggen hoe op basis van statistische overwegingen een *NI margin* kan worden bepaald. Verder werd ter illustratie een voorbeeld gepresenteerd van gepubliceerde *NI trials* waarin nieuwe orale antistollingsmiddelen voor de indicatie preventie trombo-embolische aandoeningen na orthopedische chirurgie werden vergeleken met de standaard behandeling voor deze indicatie. Het bleek dat er een grote variatie was in de gekozen *NI margin*s in een situatie waarin deze eigenlijk gelijk zouden moeten zijn. Dit kan er toe leiden dat er verkeerde conclusies worden getrokken over het al of niet slechter zijn van de nieuwe middelen ten op zichte van het controle geneesmiddel en in welke mate deze middelen effectiever zijn dan placebo. Verder wordt in dit hoofdstuk aandacht besteed aan tekortkomingen in het huidige FDA concept richtlijnvoorstel op het gebied van het vaststellen van de *NI margin*, met inbegrip van hoeveel effect tussen standaard en placebo medicatie die moet worden gehandhaafd, de hieraan gerelateerde effectmaten (relative risico's of absolute risicoverschillen) en de

constancy assumption.

Zoals eerder aangegeven zijn naast statistische overwegingen ook klinische overwegingen bij het vaststellen van de *NI margin* belangrijk. Hoe in de praktijk de klinische overwegingen meewegen bij het vaststellen van de *NI margin* is niet duidelijk. Om dit verder uit te zoeken wordt in **hoofdstuk 2.4** een studie gepresenteerd waarin op basis van een via internet toegestuurde vragenlijst experts op dit gebied werden ondervraagd. In dit onderzoek werd aan de experts een hypothetisch nieuw antistollingsmiddel gepresenteerd en deze zou vergeleken moeten worden met een bestaand antistollingmiddel (enoxaparine) voor de indicatie preventie thrombo-embolische aandoeningen na orthopedische chirurgie. Getracht werd na te gaan welke klinische overwegingen de experts gebruiken voor het vaststellen van de *NI margin*. Om dit na te gaan werden twee situaties gepresenteerd. Eerst kregen de experts geen enkele informatie over statistische overwegingen om de *NI margin* vast te stellen, in de tweede situatie werd hierover wel informatie verstrekt. Bij de beantwoording werd aan de experts steeds gevraagd hun antwoorden te motiveren.

De *NI margin*s die de 25 deskundigen die aan het onderzoek deelnamen voorstelden vertoonden een grote variatie. De motivatie welke klinische overwegingen werden gehanteerd bij het vaststellen van de *NI margin* waren duidelijker in de tweede situatie waarin de statistische overwegingen tevens werden gepresenteerd. Als motivaties werden onder andere genoemd de *NI margin* zo klein mogelijk maken, de *NI margin* overnemen van een vorige NI studie en een *NI margin* kiezen die als klinisch relevant kan worden beschouwd. Het bleek dus uit het onderzoek dat subjectiviteit een grote rol speelt bij het vaststellen van de *NI margin*. Het verdient aanbeveling in richtlijnen over *NI trial*s meer aandacht te besteden aan de overwegingen die een rol zouden moeten spelen bij het vaststellen van de *NI margin*. Mogelijk zou dit tot meer consistente en klinisch relevante keuzes van de *NI margin* leiden.

In **hoofstukken 3.1 en 3.2** staan de regulatoire aspecten van *NI trial*s centraal. In **hoodstuk 3.1** wordt een studie gepresenteerd waarin de adviezen die door de EMA worden gegeven aan de farmaceutische industrieover *NI trial*s worden besproken. De adviezen zijn afkomstig uit de jaren 2008 en 2009. De belangrijkste bevindingen waren dat firma's vaker vragen stellen over hoe een *NI trial* moet worden uitgevoerd dan of een *NI trial* überhaupt wel de aangewezen studieopzet is. En het was opvallend dat bij vragen over de *NI margin* de EMA in 40% van de gevallen een striktere *NI margin* adviseerde dan door de farmaceutische industrie voorgesteld en bij 10% vraagt om de keuze van de *NI margin* nader te motiveren. Uit dit onderzoek bleek dus dat er bij de farmaceutische industrie ook recent nog veel methodologische vragen zijn over de uitvoering van *NI trial*s.

Mogelijk kunnen verbeterde richtlijnen voor de uitvoering van *NI trial*s de ervaren onduidelijkheden bij de farmaceutische industrie terugdringen echter het wetenschappelijk advies van registratie-autoriteiten zal waarschijnlijk ook in de toekomst belangrijk blijven, bijvoorbeeld wanneer op nieuwe therapeutische gebieden *NI trial*s worden uitgevoerd.

Zoals eerder gesteld is klinische relevantie bij *NI trial*s een belangrijk onderwerp. Alhoewel een *NI trial* als primaire doel heeft aan te tonen dat een nieuw middel niet minder effectief is dan een controle middel zou het gewenst zijn om in dezelfde studie ook te bestuderen in hoeverre het nieuwe middel op andere gebieden zoals bijwerkingen, toepasbaarheid of gebruikersgemak een therapeutische meerwaarde heeft. Vooral zogenaamde fase 4 trials die worden uitgevoerd nadat een middel al op de markt is toegelaten zouden niet alleen NI als doel moeten hebben maar ook het aantonen van therapeutische meerwaarde op specifieke gebieden. Deze studies moeten die informatie opleveren die relevant zijn voor onder andere zorgverleners, patiënten, richtlijncommissies om geneesmiddelkeuzes op te baseren.

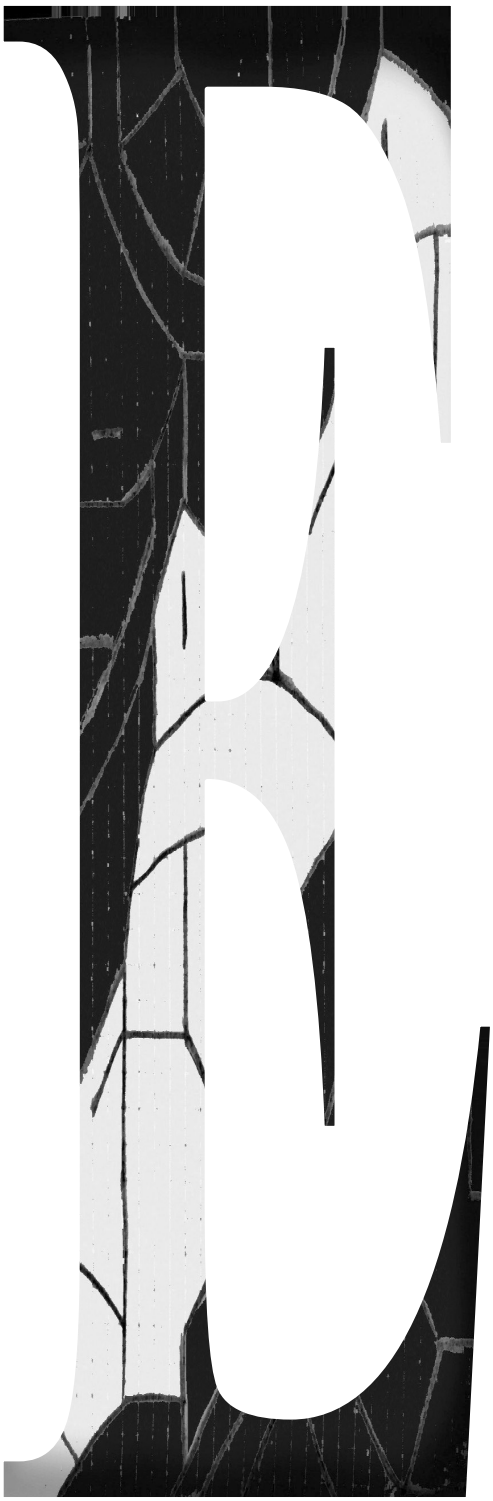
In **hoofdstuk 3.2** wordt een studie gepresenteerd waarin 41 gepubliceerde fase 4 *NI trial*s werden geanalyseerd in hoeverre zij naast de NI resultaten ook extra andere gunstige resultaten op basis van bijwerkingen, toepasbaarheid en gebruikersgemak presenteren en in hoeverre deze resultaten wetenschappelijk zijn onderbouwd. Het bleek dat bij ongeveer de helft (54 %) van de 41 studies waarin claims over gunstige effecten werden gepresenteerd statistisch waren onderbouwd. Verder viel op dat niet-industrie gesponsorde studies ten opzichte van wel industrie gesponsorde studies vaker op het gebied van geneesmiddelveiligheid extra gunstige effecten presenteerden. Alhoewel het dus goed is dat er in fase 4 trials naast de NI resultaten regelmatig ook andere therapeutische claims worden gepresenteerd is er een zorg van misbruik voor marketing doeleinden.

In het afsluitende **hoofdstuk 4** (algemene discussie) wordt de vraag centraal gesteld in hoeverre er wel behoefte is aan *NI trials*. In 2007 stelden Garattini en Bertele dat *NI trial*s onethisch zijn omdat zij niet de intentie hebben te laten zien dat een nieuw geneesmiddel een therapeutische meerwaarde heeft ten opzichte van een controle middel. Zij vonden zelfs dat de *NI trials* verboden zouden moeten worden zelf als de methodologische problemen van de *NI trials* zoals eerder besproken zouden worden opgelost.

In de discussie bespreken we de voor- en nadelen van *NI trial*s en eindigen deze afweging met de conclusie dat er in *NI trial*s naast NI claims ook altijd aspecten van therapeutische meerwaarde moeten worden bestudeerd. Deze extra claims moeten wel dezelfde wetenschappelijk onderbouwing kennen als de NI claim. Indien aan deze

voorwaarde is voldaan achten wij *NI trial*s niet onethisch en relevant voor het geneesmiddelonderzoek.

Samenvattend zijn er dus voldoende methodologische en regulatoire uitdagingen om de *NI trial* te verbeteren. Naast NI zouden er ook altijd aspecten van therapeutische meerwaarde bepaald moeten worden en deze moeten met voldoende statistische zekerheid worden onderbouwd.



pilog

Seoul, South Korea, April 2008

All the members of Nanta family are sleeping when the two thieves arrived. The thieves think they would have an easy night. They never guessed that their they had chosen a wrong victim.: Nanta family was not just an ordinary family but a family of Kungfu fighters. The thieves are slowly tiptoeing, passing the old grandpa who is sleeping on the bench……..

_____

In the middle rows of the theater, a mobile phone vibrates. The owner, a South East Asian women checks the caller ID. Her eyebrows frowned. This is it, she thinks. Without answering the call, she slips through the exit curtains and goes outside the theater hall. She is looking for an empty spot where she can answer the call freely. She finally finds a spot near the Emergency stairs. She picks up the phone.

Hallo, Grace speaking, says she.

Hallo Grace? We are calling from the Netherlands…

_____

That was one of the story that started all of this. The whole story behind this thesis was actually started when I found in a vacancy for a PhD position with the Escher project in Utrecht University. As a person who was working with clinical trials in Asia, the aim of the project has fascinated me the most : 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 manner. I found that this aim rows to the same direction as my long-term life goal to develop and simplify the process of clinical trials.

I sent out my application and after several emails, I got an interview. When the interview was scheduled, I was watching a performance by a famous South Korean theater with my Asia-Pacific colleagues after 2 days of annual company meetings. I rarely told the story to anyone, but it was a weird situation when I have to walk up and down the emergency stairs , in a theater in South Korea, while answering what is “confounding by indication’ on the phone to persons sitting somewhere in Holland.

Apparently, I gave satisfying answers that night, because in a few months, after several follow up interviews, I was accepted in the Escher program 2.1. The story later has been my inspiration when I felt I am exhausted and low in motivation to reach the finish line of my PhD.

My emergency stairs drama is not the only story that kept me going. There are other stories that kept me go on and finish this thesis. For that, I will be forever grateful to the story tellers, who with their own consciences nor without , have told, showed and put me inside their stories (of life).

So, first, I would like to thank my supervisors. Prof. de Boer, thank you so much for being such an attentive and patient tutor for me. Thank you for listening and responding to my emails rapidly. I learned from you, not only how to conduct research, but also how a good supervisor should be. I hope someday, when I got the opportunity to become a supervisor, I will always remember to commit my time and guidance for my students. Prof. Hoes, thank you for all your inspirations and brilliant ideas. I first knew you as the 'swimming pool' professor (thanks to the case-control parable), and it is an honour that I have had the opportunity to get to know you beyond that. I learned from you, how to think critically and express ideas creatively, but yet understandable to others. Dear Mirjam, how can I thank you enough? Since the selection process you were helping me. Thank you for sharing your knowledge and your constant guidance. Thank you for your patience for my broken English and sometimes, my silence and my flight of ideas that had resulted in my inconsistency. Thank you for showing me the right directions and be firm on me. I learned a lot from you how to be a daily supervisor. Olaf, thank you for your help and time in the project. Thank you for sharing your knowledge on observational studies and your criticism on the articles.

I would like to thank the members of my reading committee : Prof. H.G.M Leufkens, Prof. Yolanda van der Graaf, Prof. H.J. Out, Prof. P.A. de Graeff and Dr. O. M Dekkers for the time you have spent in reading my thesis. I would also like to thank Prof. Kit Roes, Dr. Christine Gispen-de Wied, Aukje Mantel-Teeuwisse, PhD, Ghislaine van Tiel, PhD, Prof. J. van Delden, Prof. Jan Raaijmakers and Pierre Verweij, PhD for their valuable contributions in the articles and this project.

Lieve Addie, Ineke, Suzanne, Anja en Monique, hartelijk dank voor jullie hulp tijdens mijn studie, met name wanneer ik een afspraak moest regelen . Dank U Liz Pardoen, voor al Uw hulp tijdens mijn onderzoek in het CBG.

To my paranimphs: *Muchas Gracias* Francesco Hernandez for being my brother in science. Bro, this last 4 years would have been different without all the laughs and 'loco momento'. The finish line is not so bad, so run and join me soon. *Salamat* Rose Bernabe ! It's been a honor to share thoughts and first-coauthorship with you. Hopefully, someday, we will continue this collaboration in South-East Asian level. Thank you for the time we have shared inside and outside the Escher Project.

In the Escher projects, I was honored to share my stories with the following PhD

students : Arna Arnardottir, Hans Ebbers, Jacoline Bouvy, Ruud Boessen, Yan Miao, Frederika van de Baan, Gudrun Stefansdottir, Margriet Franken, Paul Smink and Sita de Vries. Thank you all for all the 'get together' inside and outside the project. Thank you Michelle Putzeist for your collaboration in our regulatory science article. Thank you also to Ellen, Sonja, Ingrid and Peter for backing us up.

For all my colleagues in the division of Pharmacoepidemiology and Clinical Pharmacology, Utrecht University : thank you for the time, friendship, all the coffee breaks and lunch we have shared together. Thank you Jamal, Sanni, Floriaan and Rolf for letting me join the "causality" journal club. Thank you for Karen Velthove , Bas Peters, Talitha Verhoef, Arjo Roersch and Joelle Höebert for our time together in our beloved N803 and the 'penthouse'.

In Holland, I was blessed with my extended families. Thank you family Priadi for making my Christmas merrier and my days in Holland full of family love. Thank you for all my families in *Keluarga Katolik Indonesia*, especially the families Pechler, Ririhena, Scheep, Janssen and Parus. Thank you *mbak* Christine, *zus* Nona and Pastor Klemens Hayons. Thank you also for the English mass community in Utrecht. The choir practices and the masses are the light and strength of my soul.

The story of this thesis and my life in Holland for the last 4 years would not be completed without stories and gossip-sessions from my friends . Thank you so much Rahmilyna Putri, Hendy Krisyanto, Tifanny Istamto, Edwin Sutanudjaja, Riza Nugraha and Aileen Yang for being there to cheer me up anytime I was down. This is all because of you, so let's raise our glass ! Thank you for my Sanurian sisters : Nadya Ayu Mawar Wulan, Paula Nagel (plus Eric and Helena Nagel), Anti Kaandorp, Mully Setyorini, Herny Rusli, Messalinda Putri, Alin Narun, Ryna Pridadi, Dede Permadi van Raaij , Shinta Tan and Irene Hadiprayitno for sharing the *Serviam* spirit and our memories in high school. Thank you for the support from my UU-ING friends, especially Sascha Muenzing, , Sean Leather, Balaji Prabhu, Batari Bulan, Fanja Bergman, Weng Chuan Peng and Naushad Hossain! Thank you for the members of UU-ING cookery groups , esp. Mary Kelly, for letting me channel my passion in cooking. Thank you for the Karaoke group, the movie-goers group, the Utrechtsers (Rizki Pandu Permana, Melinda Aprelia, Vitto Christaldi, Detty , Jeroen Zandbergen, Edy-Daisy Husaeni, mbak Kiki, Bene–Thomas Dirkmaat, Vita Sketelenburg and all members of PPI Utrecht), the Randstad-ers and the Groningen-ers (Tita Listowardojo and Laksmi Darmoyono). Thank you Esther de Vries, Sandra Leon-Lopez and all my NIHES friends. I am glad that we still have contact after so many years. Thank you Henny Mardiani for all the idealism-chats, may we both prevails in our (im)possible dream. Thank you dr. Lina Ratulangie, (the best ex-Boss ever!) , for letting me apply for this PhD position. Thank you

my CRA-Indonesian colleagues (QQ,Anna, Eping, Carin, Ekun,*mbak* Noni) for inspiring me to make clinical trials in Indonesia better.

Thank you for the 'Rabbits 95 van Holland' (and it extended members) who were in the Netherlands during the period of my stay. It is amazing that 10 years after graduating from the medical school, I can still share the moments with you guys as if we are still the innocent medical students. Your stories and passions about clinical medicine has been a constant reminder for me that in the end we do this for the benefit of our patients. Patients were always been our greatest teachers and to them we owe our medical knowledge. *Terima kasih* Anastasia Wulandari, Agus Rizal A. Hamid, Aprilianto Eddy Wiria, FX Widiarso, Esthika Dewiasty, Sunanto Ng, Sugiharto, Ichnandy A Rachman, Sita Arumi and Ita Fatati for all the precious moments. Thank you for Henny Wati, Eva Suarthana, Made Klaci Ramadhani, Gunawan Purdianto and Isabelle Soerjamataram who were never ceased their support, even though they were living far-far away from Holland.

I thank God for my family, especially my parents for their life stories that has helped me to compose this chapter of my life. *Terima kasih* Papa, for all your stories and examples on how to be a good doctor and a good scholar. Thank you for your love of knowledge that you have shared with me. I hope you watch me from 'above' and be proud of me. Miss you a lot down here, Pa. *Mama, merci beaucoup pour vos prières.* Your stories about living in France has inspired me a lot and gave me motivations to push myself to be better. Thank you for sharing your love for languages and travelling. Thank you for being the best mom in the world. For my sisters, Margaret Wangge, Anastasia Intan Wangge and Maria Berlian Wangge, thank you for your love and support. Sorry that I was not around during the last few years and be there for you. Thank you for keeping me in the loop of our sisterhood stories. Thank you all for taking care of Jollie and Odie. Thank you Charles, for guarding my sister Maggie and help her creating a new chapter of her life. I cannot wait to see Nate and read him a special auntie Grace's bedtime story.

Last but certainly not least, I like to thank the readers of my articles, my thesis and of this long epilog. Stories are meant to be told and shared. Thus, without you, these stories about non-inferiority trials will be meaningless. I hope some of you will be inspired by stories in this book and create many memorable stories of your own.

…….grandpa Nanta is still sleeping, but somehow he manages to hit and kick the thieves. With his Kungfu skills he surely gives the thieves a lesson of their life time. The thieves jump out of the window and run for their life. They scream and warn everybody: there is a big, scary, sleeping monster in the Nanta's house.

List of publications

Wangge G, Klungel OH, Roes KC, de Boer A, Hoes AW, Knol MJ. Room for improvement in conducting and reporting non-inferiority randomized controlledtrials on drugs: a systematic review. PLoS One. 2010 Oct 27;5(10):e13550. Review.

Wangge G, Klungel OH, Roes KC, de Boer A, Hoes AW, Knol MJ. Interpretation and inference in noninferiority randomized controlled trials in drug research. ClinPharmacolTher. 2010 Sep;88(3):420-3. Epub 2010 Jul 28.

Wangge G, Roes KC, de Boer A, Hoes AW, Knol MJ. The challenges of determining noninferiority margins: a case study of noninferiority randomized controlled trials of novel oral anticoagulants. CMAJ 2012 Aug 20.

Publication not related to this project :

Dijk JM, Wangge G, Graaf Y, Bots ML, Grobbee DE, Algra A; SMART-Study Group. Hemoglobin and atherosclerosis in patients with manifest arterial disease. The SMART-study. Atherosclerosis. 2006 Oct;188(2):444-9. Epub 2006 Jan 4.(2nd author)

I do indeed believe that there is a certain contrast between, say, people in scientific professions and people working in the arts.

Often there is even mutual suspicion and irritation, and in some cases one group greatly undervalues the other.

Fortunately there is no one who actually has only feeling or only thinking properties. They intermingle like the colors of the rainbow and cannot be sharply divided.

So let us then try to climb the mountain, not by stepping on what is below us, but to pull us up at what is above us.

MC Escher, Dutch graphic artist (1898– 1972)

About the author

Grace Wangge was born on the 16th of October 1977 in Jakarta, Indonesia. She completed her high school in 1995 from Sancta Ursula High School in Jakarta, Indonesia. On the same year, she continued her study in the Faculty of Medicine, University of Indonesia, Jakarta, Indonesia. She graduated as a general practitioner in 2001, and then worked as a physician in several clinics and hospitals in Jakarta and west Java.

In 2002, she got a scholarship to continue her master education in clinical epidemiology in Netherlands Institute of Health Sciences (NIHES), Rotterdam, The Netherlands.

After graduating in 2003, she went back to Indonesia and worked with several multinational pharmaceutical companies in Jakarta as a clinical research associate and a medical advisor.

From 2008 to 2012, she continued her PhD in the division of Pharmaco-epidemiology and Clinical Pharmacology, Utrecht Institute of Pharmaceutical Science, Utrecht University, The Netherlands. The results obtained during this last period are described in this thesis.