



Published in final edited form as:

*Clin Trials*. 2011 August ; 8(4): 432–439. doi:10.1177/1740774511410994.

## Some essential considerations in the design and conduct of non-inferiority trials

Thomas R Fleming<sup>a,b</sup>, Katherine Odem-Davis<sup>a,b</sup>, Mark D Rothmann<sup>c</sup>, and Yuan Li Shen<sup>c</sup>

<sup>a</sup>Department of Biostatistics, University of Washington, Seattle, WA, USA

<sup>b</sup>Fred Hutchinson Cancer Research Center, Seattle, WA, USA

<sup>c</sup>Division of Biometrics V, U.S. Food and Drug Administration, Silver Spring, MD, USA

### Abstract

**Background**—Suppose a standard therapy (Standard) has been established to provide a clinically important reduction in risk of irreversible morbidity or mortality. In that setting, the safety and efficacy of an experimental intervention likely would be assessed in a clinical trial providing a comparison with Standard rather than a placebo arm. Such a trial often is designed to assess whether the efficacy of the experimental intervention is not unacceptably worse than that of Standard, and is called a non-inferiority trial. Formally, the non-inferiority trial usually is designed to rule out a non-inferiority margin, defined as the minimum threshold for what would constitute an unacceptable loss of efficacy.

**Purpose**—Even though the literature has many important articles identifying various approaches to the design and conduct of non-inferiority trials, confusion remains especially regarding key considerations for selecting the non-inferiority margin. The purpose of this article is to provide improved clarity regarding these considerations.

**Methods**—We present scientific insights into many factors that should be addressed in the design and conduct of non-inferiority trials to enhance their integrity and reliability, and provide motivation for key considerations that guide the selection of non-inferiority margins. We also provide illustrations and insights from recent experiences.

**Results**—Two considerations are essential, and should be addressed in separate steps, in the formulation of the non-inferiority margin. First, the margin should be formulated using adjustments to account for bias or lack of reliability in the estimate of the effect of Standard in the non-inferiority trial setting. Second, the non-inferiority margin should be formulated to achieve preservation of an appropriate percentage of the effect of Standard.

**Limitations**—The considerations, in particular regarding the importance of preservation of effect, might not apply to settings where it would be ethical as well as clinically relevant to include both Standard and placebo arms in the trial for direct comparisons with the experimental intervention arm.

**Conclusions**—Non-inferiority trials with non-rigorous margins allow substantial risk for accepting inadequately effective experimental regimens, leading to the risk of erosion in quality of health care. The design and conduct of non-inferiority trials, including selection of non-inferiority margins, should account for many factors that can induce bias in the estimated effect of Standard

---

© The Author(s), 2011

**Author for correspondence:** Thomas R Fleming, Department of Biostatistics, University of Washington, Box 357232, Seattle, WA 98195-7232, USA. tfleming@u.washington.edu.

**Disclaimer** The opinions expressed in this paper are those of the authors and not necessarily those of the US FDA.

in the non-inferiority trial and thus lead to bias in the estimated effect of the experimental treatment, for the need to ensure the experimental treatment preserves a clinically acceptable fraction of Standard's effect, and for the particular vulnerability of the integrity of a non-inferiority trial to the irregularities in trial conduct. Due to the inherent uncertainties in non-inferiority trials, alternative designs should be pursued whenever possible.

---

## Introduction

Consider clinical settings for treatment or prevention of disease where an intervention has been reliably established to provide a clinically important reduction in the risk of irreversible morbidity or mortality. We will refer to that effective intervention as Standard. In such settings, there may be an interest in evaluating an experimental therapy thought to potentially provide efficacy similar to Standard, while being likely to provide substantial improvements in safety, tolerability, or feasibility of allowing sustained delivery. When it would not be ethically or clinically appropriate to deprive patients of the established therapy, a proper design to evaluate the experimental intervention would be a randomized trial with Standard as the control regimen. Establishing the experimental therapy to be safe and to have efficacy superior to Standard would be a preferred approach for obtaining reliable evidence that the experimental intervention has a favorable benefit-to-risk profile. However, it may be sufficient to obtain evidence of safety and efficacy of this experimental treatment through a trial designed to assess whether its efficacy is not unacceptably worse than that of Standard. A study with these objectives has been called a 'non-inferiority trial,' and usually is designed to formally rule out the non-inferiority 'margin,' denoted  $\delta$ , defined as the minimum threshold for what would constitute an unacceptable loss of efficacy [1–14].

A key challenge in designing a non-inferiority trial is the need to provide a rigorous scientific justification for the choice of  $\delta$ . In this article, we present insights into two essential considerations that should guide two separate steps in the formulation of the margin. The first consideration relates to the need to address the many factors that can induce bias in the evaluation of the effect of Standard in the non-inferiority trial setting, when this evaluation is based on data from previous trials evaluating that regimen. Such bias is important because it would lead to bias in the estimate of the true effect of the experimental intervention. The second relates to the need to ensure that this test treatment preserves a clinically acceptable fraction of Standard's effect. We also discuss the particular vulnerability of the integrity of a non-inferiority trial to irregularities in trial conduct. Even though the literature has many important articles identifying various approaches to the design and conduct of non-inferiority trials, confusion remains especially regarding key considerations for selecting the non-inferiority margin. While recognizing the fact that differences of opinions in the clinical trials community regarding these issues exist, our purpose in this article is to provide improved clarity regarding the reasoning for these considerations.

## Two fundamental considerations in the formulation of the non-inferiority margin, $\delta$

Similar efficacy between the experimental intervention and Standard in the non-inferiority trial does not allow one to distinguish between the two regimens being similarly effective or, conversely, similarly ineffective, unless there is reliable evidence about the effect of Standard relative to best supportive care or placebo in the setting of the non-inferiority trial. When the non-inferiority trial does not have a placebo control as a third arm in the trial, Standard's effect, (i.e., the true effect of Standard relative to a placebo) in the non-inferiority trial setting usually is evaluated in an indirect manner. Specifically, it often is assumed that Standard's estimated effect from earlier (typically randomized) trials provides an unbiased

estimate of its effect in the non-inferiority trial. However, there are many mechanisms or factors that cause substantial risk for bias with this approach, either because the estimate of Standard's effect in the earlier trials is biased, or due to failure of the 'constancy assumption' that indicates Standard has the same true effect in the non-inferiority trial as it had in the earlier trials.

One factor that can cause the constancy assumption to be violated is that Standard's effect can vary across observed or unobserved patient characteristics or covariates, and there may be differing distributions of these characteristics between the previous studies and the non-inferiority trial. Extensive research has been undertaken in many disease settings to identify genetic factors that influence the magnitude of treatment effect. Successes include evidence that the effect of the drug, trastuzumab, in breast cancer patients depends on tumor levels of her2-neu over-expression, and recently it has been suggested that the level of effect (if any) of epidermal growth factor receptor-inhibiting drugs in colorectal cancer patients depends strongly on whether tumors express the wild type or the mutated version of the KRAS gene [15,16]. In these examples, her2-neu tumor levels and KRAS gene expression type are defined as treatment effect modifiers since the size of the effect of the corresponding treatments differs substantially according to the levels of these variables. Non-genetic factors also can be treatment effect modifiers. For example, there is considerable evidence that the magnitude of the absolute reduction in mortality provided by penicillin-like antibiotics in community acquired bacterial pneumonia is strongly dependent on the age and bacteremia status of the patient and, in first-line non-small cell lung cancer, benefit for bevacizumab and pemetrexed is limited to patients with predominantly non-squamous disease [17–19]. If patient characteristics that are treatment effect modifiers are not distributed similarly in the non-inferiority trial and in the earlier trials used to estimate the effect of Standard, the resulting estimate of Standard's effect likely will not be applicable to the population in the non-inferiority trial, leading to violation of the constancy assumption. To illustrate the importance of this phenomenon in the setting of antibiotics for community acquired bacterial pneumonia; suppose the formulation of the non-inferiority margin is based on earlier trials that establish Standard has large effects on the measure of absolute reduction in mortality in a population at highest risk of death such as in the elderly or those with bacteremia. A new experimental antibiotic that truly is ineffective in all patients may mistakenly be judged to be effective if it is evaluated in a non-inferiority comparison with Standard that is conducted in only young patients at low risk for major morbidity or mortality, if the Standard is ineffective or has much less effect on the absolute risk of death in such low risk patients. Unfortunately, this scenario has similarities to a widely implemented approach in the evaluation of antibiotics in community acquired bacterial pneumonia. Even though evidence-based non-inferiority margins can be provided in this setting only for the endpoint of mortality and these margins are of substantial size only in high risk populations, recent non-inferiority trials in community acquired bacterial pneumonia typically have recruited lower risk populations as documented in presentations at the U.S. Food and Drug Administration Anti-infective Drugs Advisory Committee meetings held on April 1–2, 2008 and on December 9, 2009, (for example, see the slide presentations by Dr Sumati Nambiar of the Food and Drug Administration [20,21]). While a proper approach in this illustration would be to ensure the non-inferiority trial is conducted in elderly or bacteremic patients, in many non-inferiority settings such an approach could not be implemented because important effect modifiers are not recognized.

Other factors that differ between the non-inferiority trial and the earlier trials used to estimate the effect of Standard provide additional reasons why the true effect of Standard may differ between these two settings. Examples of such factors include the enhancement of concomitant medications or supportive care in the non-inferiority trial setting, changes in

disease etiology, changes in trial endpoints, and changes in the dose or schedule of the Standard regimen due to evolution of its use in clinical practice.

Whether or not the constancy assumption is valid, there are additional concerns about factors that would cause bias in the estimates of Standard's effect in the earlier trials, in turn inducing bias in the estimates of Standard's effect in the non-inferiority trial. These factors include processes for selecting the Standard regimen, for selecting the information sources for estimating its effect, or for identifying covariates that are apparent effect modifiers. Specific factors include (i) publication bias, (ii) selecting the evidence from historical trials that will yield more favorable estimates of Standard's effect, (e.g., as illustrated in section B of [12]), and (iii) 'random high' bias. 'Random high' bias arises when one selects the best from among many estimated outcomes, since that which appears to be best tends to be an overestimate of its true value [22]. The effects of 'random high' bias are of particular concern for non-inferiority trials because only therapies having particularly favorable estimated effects would be considered to be the control regimen (i.e., Standard). Further, the non-inferiority study population also may be restricted to the subgroup of patients estimated to most benefit from the control regimen, where this subgroup often is identified through rather extensive exploratory analyses of data from the earlier trials used to estimate the effect of Standard [23,24]. The setting of previously untreated metastatic pancreas cancer provides an illustration of this concern about random high bias in the estimate of Standard's effect. While the estimated effect of the addition of erlotinib to gemcitabine in this setting was only modest, it was best among a large number of agents that were evaluated in combination with gemcitabine in previously untreated metastatic pancreas cancer [25]. Therefore, if gemcitabine plus an experimental treatment is compared with gemcitabine plus Standard, where erlotinib is chosen to be Standard, there likely is 'random high' bias in the estimate of erlotinib's effect which in turn will lead to a biased overestimate of the effect of the experimental therapy.

Because there are many mechanisms or factors that can lead to bias in estimating the effect of Standard in the non-inferiority trial, the following consideration is integral in determining the non-inferiority margin:

## Consideration A

The non-inferiority margin should be formulated using adjustments to account for bias or lack of reliability in the estimate of the effect of Standard in the non-inferiority trial setting.

The non-inferiority trial is not a unique setting where there is a need for adjustments to account for bias or lack of reliability that is inherently present due to the type of study design being used. For example, suppose the efficacy of an experimental intervention were assessed instead in a non-randomized superiority trial comparing this test treatment with a standard-of-care control, where imbalances in prognostic baseline covariates inherently undermine the integrity of the evaluation. Even if adjustments were made for confounding caused by known and recorded covariates, it still would be necessary to account for the risk of bias in estimates of efficacy of the experimental treatment that arises from imbalances in unaddressed *prognostic* baseline covariates. These challenges in non-randomized superiority trials are conceptually similar to those arising when evaluating the efficacy of an experimental intervention with a non-inferiority trial design. In the non-inferiority trial setting, the unbiasedness and reliability of estimates of the efficacy of Standard, (and hence of the test treatment), are inherently undermined by unaddressed treatment effect modifiers and by processes providing additional risks of obtaining biased estimates of the true effect (relative to placebo) of Standard in the non-inferiority trial. This inherent risk of bias and lack of reliability when non-inferiority trial designs are used to evaluate the efficacy of an

experimental treatment arises even though the non-inferiority trial and the earlier trials providing estimates of Standard's effect may be randomized. Several of these bias-inducing factors in trials using a non-inferiority design provide a tendency for the true effect of Standard in the non-inferiority trial to be overestimated, in turn leading to overestimation of the true effect of the test therapy. In essence, just as there is no scientifically rigorous basis to ensure integrity of a non-randomized trial designed to distinguish between an experimental treatment being ineffective as opposed to having moderate yet clinically relevant effects, there are inherent inadequacies when using a non-inferiority trial design, even if randomized, to assess efficacy of an experimental intervention when the anchor for that assessment is not placebo but rather Standard whose effect (relative to placebo) in the non-inferiority trial is inherently unknown.

Some approaches to account for bias or lack of reliability in the estimate of the effect of Standard in the non-inferiority trial setting, often resulting from factors inducing a tendency to overestimate Standard's true effect, include: (i) obtaining a confidence interval for the effect of Standard by aggregating data from earlier controlled trials evaluating Standard, and then assuming the lower limit of that confidence interval to be the true effect of Standard in the non-inferiority trial setting, according to the '95-95' method [7,8,12,26]; (ii) reducing, by a multiplicative attenuation factor, the estimate of the effect of Standard obtained from those earlier trials [7]; or (iii) adjusting the variance as well as the estimate of the effect of Standard from those earlier trials, due to lack of information about how to specify this attenuation factor, [23]. Approaches should be pursued not only to account for but also to reduce the magnitude of bias. For example, whenever possible, non-inferiority trials should be conducted in populations and under conditions comparable to those for the historical trials used to estimate the effect of Standard.

There is a second consideration that also is integral in determining the non-inferiority margin. Since the established effect of Standard is sufficiently clinically important that it would have been unethical to have conducted a trial where the control patients would have received best supportive care or placebo rather than Standard, the margin,  $d$ , should be selected in a manner to ensure that a substantial fraction of the effect of Standard is preserved by any regimen that may be used. To be specific, the following also is integral.

## Consideration B

The non-inferiority margin should be formulated to achieve preservation of an appropriate percentage of the effect of Standard.

Some argue that the preservation of a set fraction or amount of the Standard effect inappropriately creates a higher standard than would be required in a placebo-controlled trial [27,28]. However, this need for preservation of effect, resulting in a higher bar for efficacy for new interventions, is clinically and ethically justified once clinically meaningful benefit has been achieved by Standard in settings of irreversible morbidity or mortality [29]. In such settings, meaningful loss of such efficacy by an experimental intervention constitutes harm relative to the level of risk of clinical outcomes experienced by patients receiving Standard. For example, in patients with community acquired bacterial pneumonia who are non-bacteremic and above age 50, it has been documented that 21-day mortality that is 50% in patients receiving no specific treatment is reduced to 16% by sulfon-amide derivatives and penicillin (i.e., Standard), [17]. If it were to be established that a test treatment also has superior efficacy to no specific treatment, but that its mortality benefit were estimated to be only one-third of Standard's magnitude of effect in community acquired bacterial pneumonia patients who would be eligible to receive Standard, the efficacy profile of this test treatment would represent harm relative to that of Standard. In 1995, President Clinton and Vice President Gore issued the following policy [30,31]:

It is essential for public health protection that a new therapy be as effective as alternatives that are already approved for marketing when:

The disease to be treated is life-threatening or capable of causing irreversible morbidity (e.g., stroke or heart attack); or The disease to be treated is a contagious illness that poses serious consequences to the health of others (e.g., sexually transmitted disease).'

To address Consideration B, the margin,  $\delta$ , can be formulated based on a statistical test for preservation of a fraction of the effect of Standard using estimates of Standard's effect and corresponding variances obtained from the earlier trials evaluating Standard [7,8,12,23]. Frequently, this preservation fraction has been chosen to be  $\frac{1}{2}$ . However, the threshold for the amount of loss of the effect of Standard that is acceptable will depend on the clinical relevance of the endpoint. Cases where Standard has a large effect on very clinically meaningful endpoints, where a large fraction of the active control effect may need to be preserved, include sulfonamide derivatives or penicillin in community acquired bacterial pneumonia [17], enoxaparin and warfarin in symptomatic venous thromboembolism [32–35], or warfarin in atrial fibrillation [36,37]. The amount of loss of effect of Standard that is acceptable also will depend on the level of improvements in safety, tolerability, and convenience of administration of the experimental regimen relative to Standard. For example, in settings of atrial fibrillation and symptomatic venous thromboembolism, consider an experimental regimen that has a substantially improved safety profile relative to Warfarin by reducing the bleeding risks, and that meaningfully improves convenience of administration by eliminating the need for regular monitoring of blood coagulation that is required with use of Warfarin. In such settings, it may be adequate for the experimental regimen to preserve a smaller fraction of Warfarin's effect. Furthermore, for the subset of patients in these clinical settings who are unwilling or unable to receive Warfarin, there would not be any preservation of effect issues, and so a placebo controlled superiority trial would be an appropriate approach for evaluating the experimental regimen in such patients.

### Quality of trial conduct issues in non-inferiority trials

Irregularities in quality of the conduct of the non-inferiority trial induce increased risk of both bias and variability. These irregularities include failure to achieve complete and timely enrollment of the targeted population, violations in eligibility criteria, lack of adherence to Standard at a level that matches best achievable in a real world setting, lack of adherence to the experimental intervention, cross-ins from one regimen to the other, and missing data on outcome measures. While such irregularities are of concern in superiority trials, they are even more problematic in a non-inferiority trial since they often dilute the sensitivity to true differences between the experimental intervention and Standard regimens, leading to an increased risk of falsely declaring non-inferiority in settings where the test treatment truly is clinically inferior to Standard. As stated in [1]: 'Many flaws on the design or conduct of the trial will tend to bias the results toward a conclusion of equivalence.'

Often, 'per protocol' analyses have been proposed as an approach to restore sensitivity to true treatment differences. However, such analyses have the same flaws in non-inferiority trials that exist in superiority settings. Only 'as randomized' analyses, where all randomized patients are followed to their outcome or to the end of study (i.e., to the planned duration of maximum follow-up or the analysis cutoff date), preserve the integrity of the randomization and, due to their unconditional nature, address the questions of most important scientific relevance. Therefore, the preferred approach to enhancing the integrity and interpretability of the non-inferiority trial should be to establish performance standards for measures of quality of trial conduct (e.g., targets for enrollment and eligibility rates, event rate, adherence and retention rates, cross-in rates, and currentness of data capture) when

designing the trial, and then to provide careful oversight during the trial to ensure these standards are met [38], with the 'as randomized' analysis being primary and with analyses such as 'per protocol' or analyses based on more sophisticated statistical models accounting for irregularities being supportive.

## Summary

Although the majority of non-inferiority trials are randomized to reduce the risk of bias arising from unaddressed prognostic covariates, there remains an inherent risk of bias in such trials arising from unaddressed treatment effect modifiers and from other sources. These sources of bias often induce a tendency to overestimate Standard's true effect in the non-inferiority trial, leading to an inflated estimate of the true effect of the experimental intervention [23,24]. Specific illustrations of some factors resulting in such overestimates include: (i) estimating Standard's effect in the non-inferiority trial using trials evaluating Standard that were conducted prior to the evolution of more effective supportive care, (ii) enrolling patients in the non-inferiority trial who are less responsive to Standard than those enrolled in the earlier trials, (iii) use of an antibiotic (as Standard) shown to be effective in earlier trials, but that is likely less effective in the non-inferiority trial setting due to emergence of resistance, (iv) selection of the Standard control regimen and the patient subgroup to be enrolled in the non-inferiority trial based on extensive exploratory analyses of earlier trials, and (v) lack of attention by those conducting the non-inferiority trial to ensure Standard is delivered in an optimal manner. Non-inferiority trials must be designed, conducted, and analyzed in a manner to address this inherent risk of bias.

Based on this recognition, when the efficacy of an experimental treatment will be assessed in a non-inferiority trial by a comparison with a Standard that has been established to provide a clinically important reduction in the risk of irreversible morbidity or mortality, it is important to the integrity of this non-inferiority trial that it be conducted with high quality, and that the formulation of the non-inferiority margin,  $\delta$ , involves two *separate* steps. First, the non-inferiority margin should be formulated using adjustments to account for bias or lack of reliability in the estimate of the effect of Standard in the non-inferiority trial setting, resulting from many factors including those listed above that induce a tendency to overestimate that effect. The second step should be to obtain the non-inferiority margin,  $\delta$ , based on the preservation of a specified fraction of the effect of Standard, since loss of such efficacy would constitute harm. Detailed illustrations of the application of these principles have been provided for many clinical settings, including for the evaluation of bivalirudin versus glycoprotein IIb/IIIa inhibitors in percutaneous coronary intervention [12, Section 2.3], and for several clinical settings in the Appendix of the draft Food and Drug Administration Guidance Document on Non-Inferiority Clinical Trials [13].

It should be noted that the consideration regarding preservation of effect might not apply when Standard has not been shown to reduce the risk of irreversible morbidity or mortality, or when it would be ethical as well as clinically relevant to include both Standard and placebo arms in the trial for direct comparisons with the experimental intervention arm. However, when a placebo arm ethically could not be included, an inherent weakness of the resulting two-arm non-inferiority trial comparing a test treatment with Standard is that it is not possible to design the trial (and specifically, to derive a non-zero non-inferiority margin) from first scientific principles (i.e., based on a direct randomized comparison to placebo or best supportive care) that will ensure the reliability of the benefit-to-risk assessment of the experimental therapy. Thus the design of such trials, including approaches to address Considerations A and B, should be carefully scrutinized and debated on a trial-by-trial basis.

Commonly, there is a tension between wanting the non-inferiority margin,  $\delta$ , to be large enough to allow for timely completion of the non-inferiority trial, and to be small enough to enhance trial integrity and the preservation of a substantial proportion of the demonstrated benefits provided by currently available regimens. To be candid, the choice of margins that are much too large often is not based on misunderstandings or differences of judgment between informed and unbiased clinicians and scientists, but rather on the clear recognition that wider margins allow sponsors to conduct smaller trials as well as trials that will have a substantially higher probability of providing 'positive' non-inferiority conclusions. Allowing margins to be chosen in such a manner is dangerous to public health interests since this allows substantial risk for erosion in the quality of health care through the replacement of effective standard regimens by experimental regimens having inferior benefit-to-risk profiles.

## Conclusion

Non-Inferiority trials should be conducted in a manner to meet high performance standards for quality of trial conduct, and the non-inferiority margins should be based on clinical evidence and unbiased judgment, with recognition of the importance of achieving reliable research results and the preservation of benefits provided by currently available regimens. Furthermore, due to the inherent uncertainties of non-inferiority trials that do not have placebo arms, alternative designs should be pursued whenever possible.

## Acknowledgments

**Funding** For TR Fleming and K Odem-Davis, the sources of financial support for research described in this article are NIH/NIAID grants entitled 'Statistical Issues in AIDS Research' (R37 AI 29168) and 'Clinical Research on AIDS Training Grant' (T32 AI07450).

## References

1. ICH E-9 – International Conference on Harmonisation. [(accessed 6 July 2011)] Statistical principles for clinical trials. Federal Register, 62 FR 25712 (09 May 1997). Available at: <http://www.gpo.gov/fdsys/pkg/FR-1997-05-09/pdf/97-12139.pdf>
2. ICH E-10 - International Conference on Harmonisation. [(accessed 06 July 2011)] Choice of control group in clinical trials. Available at: <http://www.fda.gov/RegulatoryInformation/Guidances/ucm125802.htm>
3. Temple, R. Difficulties in evaluating positive control trials Proceedings of the American Statistical Association, Biopharmaceutical Section, Section I. 1983. p. 1-7.
4. Fleming TR. Treatment evaluation in active control studies. *Cancer Treat Rep.* 1987; 71:1061–65. [PubMed: 3119201]
5. Fleming TR. Evaluation of active control trials in acquired immune deficiency syndrome. *J AIDS.* 1990; 3:82–87.
6. Temple R, Ellenberg S. Placebo-controlled trials and active-control trials in the evaluation of new treatments. Part 1: ethical and scientific issues. *Ann Intern Med.* 2000; 133:455–63. [PubMed: 10975964]
7. Rothmann M, Li N, Chen G, et al. Design and analysis of non-inferiority mortality trials in oncology. *Stat Med.* 2003; 22:239–64. [PubMed: 12520560]
8. Hung HMJ, Wang SJ, Tsong Y, et al. Some fundamental issues with non-inferiority testing in active controlled trials. *Stat Med.* 2003; 22:213–25. [PubMed: 12520558]
9. Piaggio G, Elbourne DR, Altman DG, et al. Reporting of noninferiority and equivalence randomized trials: an extension of the CONSORT statement. *JAMA.* 2006; 295:1152–60. [PubMed: 16522836]
10. Kaul S, Diamond GA. Good enough: a primer on the analysis and interpretation of noninferiority trials. *Ann Intern Med.* 2006; 145:62–69. [PubMed: 16818930]



11. Freidlin B, Korn EL, George SL, Gray R. Randomized clinical trial design for assessing non-inferiority when superiority is expected. *J Clin Oncol.* 2007; 25:5019–23. [PubMed: 17971602]
12. Fleming TR. Current issues in non-inferiority trials. *Stat Med.* 2008; 27:317–32. [PubMed: 17340597]
13. Draft FDA Guidance Document. [(accessed 06 July 2011)] Guidance for industry non-inferiority clinical trials. Available at:  
<http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM202140.pdf>
14. Powers JH. Non-inferiority and equivalence trials: deciphering the similarity of medical interventions. *Stat Med.* 2008; 27:343–52. [PubMed: 18186148]
15. U. S. FDA. [(accessed 15 May 2011)] Herceptin product labeling. Available at:  
[http://www.accessdata.fda.gov/drugsatfda\\_docs/label/2008/103792s51751bl.pdf](http://www.accessdata.fda.gov/drugsatfda_docs/label/2008/103792s51751bl.pdf)
16. Karapetis CS, Khambata-Ford S, Jonker DJ, et al. K-ras mutations and benefit from cetuximab in advanced colorectal cancer. *New Engl J Med.* 2008; 359:1757–65. [PubMed: 18946061]
17. Fleming TR, Powers JH. Issues in noninferiority trials: the evidence in community-acquired pneumonia. *Clin Infect Dis.* 2008; 47:S108–20. [PubMed: 18986275]
18. U.S. FDA. [(accessed 15 May 2011)] Avastin product labeling. Available at:  
[http://www.accessdata.fda.gov/drugsatfda\\_docs/label/2009/125085s01681bl.pdf](http://www.accessdata.fda.gov/drugsatfda_docs/label/2009/125085s01681bl.pdf)
19. U.S. FDA. [(accessed 15 May 2011)] Alimta product labeling. Available at:  
[http://www.accessdata.fda.gov/drugsatfda\\_docs/label/2009/021462s0211bl.pdf](http://www.accessdata.fda.gov/drugsatfda_docs/label/2009/021462s0211bl.pdf)
20. U.S. FDA. FDA presentation slides for Anti-Infective Drugs Advisory Committee meeting; 09 December 2009; Available at:  
<http://www.fda.gov/downloads/AdvisoryCommittees/CommitteesMeetingMaterials/Drugs/Anti-InfectiveDrugsAdvisoryCommittee/UCM195620.pdf>
21. U.S. FDA. FDA presentation slides for Anti-Infective Drugs Advisory Committee meeting; 01 April 2008–02 April 2008; Available at:  
[http://www.fda.gov/ohrms/dockets/ac/08/slides/2008-4343s1-01-FDA-corepresentation\\_files/frame.htm](http://www.fda.gov/ohrms/dockets/ac/08/slides/2008-4343s1-01-FDA-corepresentation_files/frame.htm)
22. Fleming TR. Clinical trials: discerning hype from substance. *Ann Intern Med.* 2010; 153:400–06. [PubMed: 20855804]
23. Davis, KS. Dissertation. University of Washington, Department of Biostatistics; 2010. Non-constancy, estimation bias, biocreep, and an alternative to current methods used in non-inferiority trials.
24. Rothmann, M.; Wiens, B.; Chan, I. Design and Analysis of Non-inferiority Trials. Chapman Hall/CRC Press; Boca Raton: 2011.
25. Moore MJ, Goldstein D, Hamm J, et al. Erlotinib plus gemcitabine compared with gemcitabine alone in patients with advanced pancreatic cancer: a phase III trial of the National Cancer Institute of Canada Clinical Trials Group. *J Clin Oncol.* 2007; 25:1960–66. [PubMed: 17452677]
26. CBER/FDA Memorandum. Summary of CBER considerations on selected aspects of active controlled trial design and analysis for the evaluation of thrombolytics in acute MI. June. 1999
27. Snapinn S, Jiang Q. Preservation of effect and the regulatory approval of new treatments on the basis of non-inferiority trials. *Stat Med.* 2008; 27:383–91.
28. Peterson P, Carroll K, Chuang-Stein C, et al. PISC expert team white paper: toward a consistent standard of evidence when evaluating the efficacy of an experimental treatment from a randomized active-controlled trial. *Stat Biopharm Res.* 2010; 2:522–32.
29. Brown D. Comment on 'PISC expert team white paper'. *Stat Biopharm Res.* 2010; 2:535–37.
30. Office of the Federal Register. [(accessed 06 July 2011)] Statement regarding the demonstrations of effectiveness of human drug products and devices (Notice). *Federal Register* 60:147 (01 August 1995) 39180-39181. Available at:  
<http://www.gpo.gov/fdsys/pkg/FR-1995-08-01/pdf/95-18877.pdf>
31. Clinton, B.; Gore, A. Reinventing Regulation of Drugs and Medical Devices. National Performance Review; Washington, DC: 1995.

32. Schulman S, Rhedin AS, Lindmarker P, et al. A comparison of six weeks with six months of oral anticoagulant therapy after a first episode of venous thromboembolism. Duration of Anticoagulation Trial Study Group. *New Engl J Med.* 1995; 332:1661–65. [PubMed: 7760866]
33. Kearon C, Gent M, Hirsh J, et al. A Comparison of three months of anticoagulation with extended anticoagulation for a first episode of idiopathic venous thromboembolism. *New Engl J Med.* 1999; 340:901–07. [PubMed: 10089183]
34. Schulman S, Granqvist S, Holmstrom M, et al. The duration of oral anticoagulant therapy after a second episode of venous thromboembolism. The Duration of Anticoagulation Trial Study Group. *New Engl J Med.* 1997; 336:393–98. [PubMed: 9010144]
35. van Dongen CJ, van den Belt AGM, Prins MH, Lensing AWA. Fixed dose subcutaneous low molecular weight heparins versus adjusted dose unfractionated heparin for venous thromboembolism. *Cochran Database of Systematic Reviews.* 2004; (4) Art. No.: DOI: 10.1002/14651858.CD001100.pub2.
36. Kaul S, Diamond GA, Weintraub WS. Trials and tribulations of non-inferiority: the ximelagatran experience. *J Am Coll Cardiol.* 2005; 46:1986–95. [PubMed: 16325029]
37. Lawrence, J.; Hung, J.; Mahjoob, K. [(accessed 15 May 2011)] Statistical review and evaluation for Exanta (ximelagatran) 36 mg bid oral formulation, clinical studies, NDA 21-686 2004. Available at: [http://www.fda.gov/ohrms/dockets/ac/04/briefing/2004-4069B1\\_07\\_FDA-Background-C-R-stat%20Review.pdf](http://www.fda.gov/ohrms/dockets/ac/04/briefing/2004-4069B1_07_FDA-Background-C-R-stat%20Review.pdf)
38. Fleming T. Addressing missing data in clinical trials. *Ann Intern Med.* 2011; 154:113–117. [PubMed: 21242367]