

JOURNAL SERIES:

Statistics Review Part 8: Non-Inferiority Trial Design

by Andrew M. North, PharmD, BCPS, MBA, Amanda Margolis, PharmD, MS, BCACP, and Kevin Look, PharmD, PhD

Objectives

1. Distinguish fundamental aspects of non-inferiority trial designs from that of superiority trials
2. Describe the role of a non-inferiority margin and the methods for how one is determined prior to trial initiation
3. Recognize the appropriate methods of data handling for non-inferiority trial designs
4. List main limitations and controversies related to non-inferiority trial designs

Clinicians must have foundational skills for evaluation of non-inferiority (NI) trial designs given the recent surge of FDA drug approvals based upon this type of study and the inherent widespread issues observed with reporting quality. A recent review of clinical trials showed only one NI trial published before 1998; however, there was a clear annual increase in the number of NI trials published since that time with over 100 published per year between 2007 and 2009.¹ Twenty-five percent of new drug applications submitted to the FDA between 2002 and 2009 included NI clinical evidence, and two-thirds of drug approvals during this period included NI trials.² The importance for clinicians to evaluate key aspects of NI trials is further illustrated by past evaluations suggesting many NI trials claimed inappropriate conclusions or had other significant methodological flaws.³⁻⁶ Additionally, NI trials have more complex designs and statistical considerations than traditional superiority studies. The purpose of this article is to increase awareness of foundational elements clinicians should consider when evaluating NI trials.

Rationale for Non-Inferiority Design

Unlike superiority trial designs where the purpose is to demonstrate that an intervention or treatment is better than either placebo or another treatment, the goal of NI trial designs is to establish that a new treatment is not ‘unacceptably worse’ than another treatment (often a standard-of-care).⁷ This results in a functional change concerning hypothesis testing when compared to traditional hypothesis tests (i.e., superiority testing).⁸ (See paper 4 in this series for a review of traditional hypothesis testing & Figure 1 in this manuscript.) In NI studies the null hypothesis states the difference between treatments will fall outside a pre-specified margin (see Figure 2). This margin serves as the maximum treatment effect that can be lost before the intervention being tested is considered ‘inferior’ to the control treatment and is utilized for testing of statistical significance to demonstrate that the NI objective is met.

The original impetus behind NI trials was to provide a means for conducting clinical trials against an established standard-of-care in cases where it would be unethical to compare an agent to placebo (e.g., cardiovascular and oncology trials).⁷ If drug sponsors were only allowed to perform superiority trials for the purpose of FDA-approval, it would become increasingly difficult to show incremental superiority against previous therapies over time, which could result in clinically viable options being blocked

from entry into the market. Given that NI trials allow an intervention to be “acceptably worse” than a standard-of-care, the tested intervention often yields some ancillary benefits which are considered favorable for patients and/or clinicians. Such possible advantages for new treatments could include lower costs, greater adherence potential (e.g., once daily dosing vs twice-daily), increased administration convenience (e.g., oral agent vs subcutaneous injection), increased safety (e.g., decreased intracranial hemorrhage risk), or decreased monitoring (e.g., reduced or a lack of laboratory monitoring parameters). The maximum amount of efficacy patients and clinicians are willing to sacrifice varies between disease states and is highly reliant upon these gains.⁷

Considerations of Non-Inferiority Trial Methodology Non-Inferiority Margins

One aspect of NI trials which is not present in superiority trials is the use of a “non-inferiority margin” (also referred to as Δ). Evaluation of the NI margin is a critical aspect when reviewing NI studies (see Box 1). The NI margin effectively serves as the maximum potential decrease in efficacy for which the studied intervention would be considered “non-inferior” to the reference intervention.⁸ Put in layman’s terms, if the NI margin is set as 20% *a priori* for the primary efficacy endpoint, this means that the treatment under evaluation will be considered “non-inferior” as long as

FIGURE 1. A Glance at Hypothesis Testing for Superiority and Non-inferiority Trial Designs

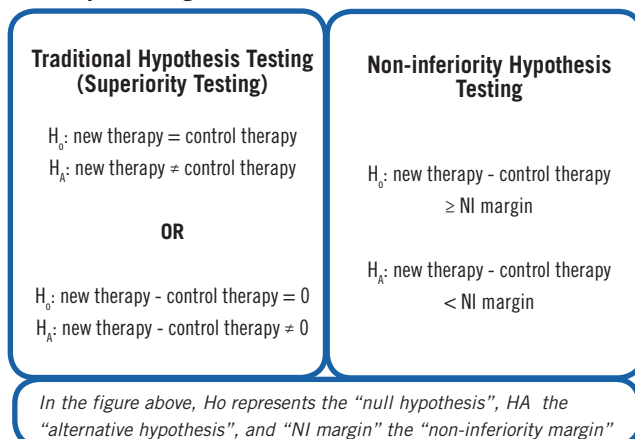
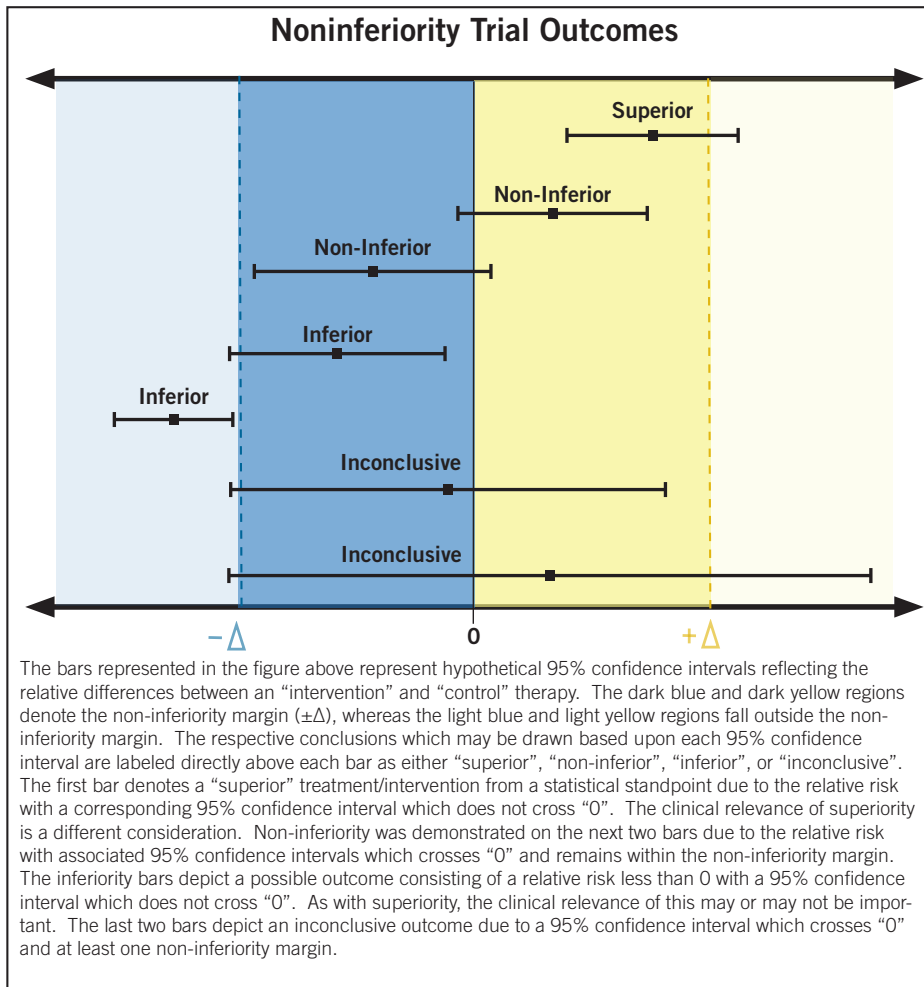


FIGURE 2. Possible Non-inferiority Clinical Trial Result Outcomes



the treatment is not more than 20% less effective than the reference treatment. For example, suppose investigators were using a 20% NI margin to evaluate a new treatment for secondary prevention of strokes and the results of the study were a 5% annual stroke rate for the standard-of-care treatment and a 5.75% rate for the new treatment. In this scenario, the new treatment would be considered inferior but within the prespecified NI margin or “non-inferior” to the standard-of-care because it was 15% less effective than the standard-of-care, which falls within the 20% NI margin.

Study investigators have several different modalities for selecting an NI margin including clinical opinion and statistical approaches, but the best approach is typically one that blends the two. The method in which the NI margin was derived prior to initiation of the study should be described in every NI study, and concerns should be raised if this

information is not provided.

Ideally, development of the NI margin should include a statistical derivation component, and several steps are essential to calculating the margin via this means. This approach requires the presence of previously conducted studies comparing the reference treatment to placebo either in the form of individual clinical trials or a meta-analysis.⁷ These placebo-controlled trials, or comparisons of two active drugs, should preferably have similar inclusion and exclusion criteria, concomitant treatments, and study duration to the planned NI trial. Additionally, the studies used should be relevant from a standard-of-care aspect for the disease state being assessed. For example, suppose investigators were planning to study a new alternative agent to clopidogrel in conjunction with aspirin use after percutaneous coronary intervention (PCI) for ST-segment elevation myocardial infarctions (STEMIs). It would not be advantageous to include any studies

conducted prior to the advent of PCI for computation of the NI margin given PCI is associated with major mortality reduction and is standard-of-care for patients who can have it performed within the recommended timeframe. Based upon the past superiority placebo-controlled trials identified, the lower bound of the 95% confidence interval is used to determine the largest possible NI margin and provides reassurance that the studied intervention is likely at least better than placebo.⁷ Investigators must now determine how much of the effects of the standard-of-care treatment would be reasonable to forgo and still consider the studied treatment non-inferior.

Clinical experience and opinion may be suitable to serve as the basis for an NI margin in the case of a lack of past superiority trials or as an adjunct to the statistical approach mentioned directly above. Use of literature on the control treatment/standard of care’s endpoint event rate or efficacy is preferred in order to adequately power the study. When there is a lack of past trials to estimate the event rate of the control group, the professional experiences of the investigators or a panel of clinical experts may be utilized to assist with development of an NI margin. Patient groups may also provide insight as to the acceptable decrease in efficacy with consideration to the possible ancillary benefits (e.g., less side effects, etc.).

Overly generous NI margins have the potential to allow for a greater loss of efficacy than patients and clinicians are willing to accept. Even more concerning, an excessive NI margin could demonstrate that a treatment is non-inferior to another when the treatment being studied is actually no more effective than placebo. Using the example above involving a new treatment for the secondary prevention of strokes, had a 75% NI margin been utilized, the treatment would have been found to be “non-inferior” if the event rates were as high as 8.75%. Because this example involves comparison to a current “standard of care” for the prevention of secondary events, clinicians may not consider an agent with an event rate as high as 8.75% non-inferior to an established agent with a 5% event rate. In addition, large NI margins could also inappropriately

BOX 1. Literature Evaluation Criteria Specific to Non-Inferiority Studies to Consider¹³

1. Is the NI margin established *a priori*?
2. Is the size of the NI margin clinically relevant and appropriate? Do the authors provide reasonable justification?
3. Has the selected control arm been proven to be effective versus placebo?
4. Is the sample size calculated taking into account the size of the NI margin?
5. Is the two-sided confidence interval margin around the difference between the treatment and control arms presented taking into account the size of the NI margin?
*Does the author test for location of the upper/lower bound relative to the NI margin (depending on whether higher or lower outcomes are preferred)?
6. Are the results based on a per-protocol analysis?
7. Are the intention-to-treat results also presented, and are the per-protocol and intention-to-treat results consistent?
8. Are the conclusions of the trial conditional on the NI margin as a reference?
*It is preferable to state “the new therapy is not worse than the standard of care by more than the pre-specified amount determined *a priori* by the NI margin” versus “the new therapy is non-inferior to the standard of care”

reduce the number of patients needed for enrollment in a study. On the other hand, overly stringent NI margins can prevent an effective therapy from being deemed non-inferior.⁸ The FDA suggests NI margins should be no larger than 50% or one-half of the difference in efficacy between the control agent and placebo.⁹ Therefore, in the example above with an established 5%

standard-of-care event rate, the NI margin should be no larger than 2.5% (Upper boundary of the 95% CI as 7.5%).

Data Handling & Statistical Analysis

Unlike superiority trial designs where intention-to-treat (ITT) analysis serves as the gold standard for outcome measures, NI trials should ideally be performed using per protocol (PP) analysis, with the results being compared to ITT analysis to check for consistency. Unlike ITT analysis which includes all patients who were randomized regardless of drop-outs or crossing of treatment arms, PP analysis compares treatment arms based upon the patients who complete the study with their original treatment randomization/allocation. The most conservative form of data analysis should be reported to prevent inappropriate labeling of a drug being “as effective” as another treatment.

Because the goal of NI trials is to show the new therapy is not unacceptably worse, the use of ITT analysis is inappropriate as it may bias the results by making the two treatments seem more similar. For superiority trial designs, ITT serves as the more conservative estimate of outcomes, as a lack of adherence to the assigned treatment arm will make the outcome seem more similar between the intervention and control groups. As the use of ITT analysis is more conservative, it allows clinicians to be more confident in the results.⁷ Use of the wrong data handling procedure can lead to inappropriately depicting a treatment as being non-inferior because of protocol deviations creating a smaller difference than there truly is. Therefore, NI analyses should ideally be performed using the PP population.^{10,11} PP analysis will thereby prevent an issue called “assay sensitivity” which refers to concerns such as flawed allocation and the inability to be certain that the statistical conclusions are valid.¹⁰ ITT analysis should still be provided as a supporting measure for non-inferiority; in fact, both the FDA and European drug regulating bodies recommend that both PP and ITT analysis be provided.¹⁰ For NI trial designs, in the case where the PP and ITT analysis yield the same result of non-inferiority, greater merit can be attributed to the results.¹² On the other hand, if PP and ITT yield dissonant results, the conclusion

that non-inferiority was achieved is weakened and scrutiny should exist.^{7,10,12}

Supplemental Testing for Superiority

With NI trial designs, investigators have the option to pre-specify the intent to perform a post-hoc superiority test if non-inferiority is demonstrated. As with superiority trial designs, investigators should always provide the metrics used to calculate power for the NI analysis in order to provide reassurance that Type II statistical error is unlikely to be the reason for failure to reject the null hypothesis. In the case where the NI null hypothesis is rejected (i.e., non-inferiority is found), the superiority analysis may proceed. In contrast, post-hoc testing for non-inferiority in a superiority trial is considered unethical. For example, if a trial designed to show superiority of one agent over another failed to demonstrate superiority, it is inappropriate to perform a “non-inferiority” analysis.

Limitations of Non-Inferiority Trials

Several concerns have been expressed about NI trials. One large issue with NI trial designs is that the statistical techniques underlying the results are much more difficult to confirm than in superiority trials. Additionally, there are many more interpretations of NI trial results in the case where non-inferiority is not achieved, which are beyond the scope of this article.

An issue known as “constancy” may pose a threat to whether a comparison was made to a reasonable reference treatment. As mentioned above, this is hopefully accounted for by knowledge of historical event rates with the reference treatment; however, there is no definitive guarantee that the intervention arm is better than placebo given the lack of a placebo arm. Inappropriate conclusion of non-inferiority can be prevented by using a standard-of-care treatment with established efficacy as the comparison. One concern is whether the reference treatment was suboptimally administered, thereby predisposing the study to a non-inferiority conclusion. This could either be in the form of lower-than-normal

dosing of the reference arm treatments or could be evident though other means such as therapeutic drug monitoring (e.g., low warfarin time in therapeutic range). Clinicians can often compare the NI study being evaluated to past studies to see if any such discrepancies exist.

Last, the concern of “biocreep” has been raised given the massive influx of NI trials during the past couple of decades. Biocreep is ultimately the issue that inferior treatments could be deemed non-inferior and become the standard-of-care for subsequent NI trials. Given the utilization of an NI margin, it has been theorized that after several NI trials it might be unknown if a studied treatment is even better than placebo. To provide an example, suppose that a theoretical Drug A is deemed to be superior to placebo. Several years later, Drug B is found to be non-inferior to drug A with a certain NI margin. Then, Drug C is compared to Drug B via a NI study at some point later. During each of these studies, the new agent is found to be “not acceptably” worse than the previous agent. It may then be unknown if Drug C is more effective than placebo.

Summary

This article reviewed several key concepts related to NI trial design, overarching reasons behind the use of NI trial designs, considerations clinicians should have when evaluating NI trials (such as a close look at the NI margin and the methods for data handling), and some of the main limitations of NI trial designs. ●

Practice Questions

- The best way of describing the fundamental goal of non-inferiority trial designs when comparing two treatments is:
 - No difference exists between treatment arms
 - One treatment is no less effective than another
 - One treatment is not “acceptably worse” than another
 - One treatment is more safe/effective than the other
- Which of the following types of data handling provides the most conservative estimate of endpoints for non-inferiority trial designs?

- Censoring of patients as they drop out
 - Intention-to-treat (ITT)
 - Per protocol (PP)
- Which of the following is a descriptor for the theorized concern that several NI trials on the same disease state could result in treatments no more efficacious than placebo?
 - Assay sensitivity
 - Biocreep
 - Constancy
 - Non-inferiority margin

Answers:

- c** With NI trial designs, it is impossible to prove that no difference exists between two treatments. For this reason, one of the hallmark features of NI trial designs is the use of an NI margin which allows for a studied treatment/intervention to be “acceptably worse” than another treatment. The studied agent often has other advantages such as less side effects or easier regimens for adherence which allow for clinicians and patients to accept a certain degree of reduced efficacy.
- c** Per protocol (PP) analysis is the most conservative because it looks at patients who completed the study based upon the original treatment arms they were assigned to. For NI trial designs, the most conservative estimate of outcomes is one which lends towards treatment arms being less similar (similar to ITT analysis use for superiority trials).
- b** Biocreep refers to the potential phenomenon where after several NI trials there may be a question as to whether treatments are actually more effective than placebo.

Andrew North is a Specialty Practice Pharmacist in emergency medicine at The Ohio State University Wexner Medical Center in Columbus, Ohio and a previous PGY-2 Critical Care Pharmacy Resident at UW Health in Madison, Wisconsin. Amanda Margolis is a Lecturer at the University of Wisconsin School of Pharmacy and a Clinical Pharmacist at the William S. Middleton Memorial Veterans Hospital, Madison, WI. Kevin Look is an Assistant Professor in the Social and Administrative Sciences Division at the University of Wisconsin School of Pharmacy, Madison, WI.

Acknowledgments: The authors would like to credit and thank Lee Vermeulen, RPh, MS, FCCP who is the Director for the UW Health Center for Clinical Knowledge Management with development of the aspects located in Box 1 of this manuscript. The authors would like to credit and thank Sally Griffith-Oh who works with the Instructional and Informational Technology Department at the UW-Madison School of Pharmacy for development of Figure 2 in this manuscript.

References and suggestions for further review:

- Suda KJ, Hurley AM, McKibbin T, et al. Publication of noninferiority clinical trials: changes over a 20-year interval. *Pharmacotherapy*. 2011;31(9):833-839.
- Office. USGA. FDA's Consideration of Evidence from Certain Clinical Trials. <http://www.gao.gov/assets/310/308304.html>. Accessed July 12, 2014.
- Greene WL, Concato J, Feinstein AR. Claims of equivalence in medical research: are they supported by the evidence? *Ann Intern Med*. 2000;132(9):715-722.
- Wangge G, Klungel OH, B. RKC, et al. Interpretation and inference in noninferiority randomized controlled research trials in drug research. *Clin Pharmacol Ther*. 2010;88(3):420-423.
- Wangge G, Klungel OH, Roes KC, et al. Room for improvement in conducting and reporting non-inferiority randomized controlled trials on drugs: a systematic review. *Clin Pharmacol Ther*. 2010;5(10):e13550.
- Le Henanff A, Giraudeau B, Baron G, et al. Quality of reporting of noninferiority and equivalence randomized trials. *JAMA*. 2006;295(10):1147-1151.
- Head SJ, Kaul S, Bogers AJ, et al. Non-inferiority study design: lessons to be learned from cardiovascular trials. *Eur Heart J*. 2012;33(11):1318-1324.
- Prins MH, Lensing AW. Derivation of the non-inferiority margin for the evaluation of direct oral anticoagulants in the treatment of venous thromboembolism. *Thromb J*. 2013;11(1):13.
- Administration USDoHaHSUSFaD. Guidance for Industry: Non-Inferiority Clinical Trials (Draft Guidance). 2010. <http://www.fda.gov/downloads/Drugs/Guidances/UCM202140.pdf>.
- Schumi J, Wittes JT. Through the looking glass: understanding non-inferiority. *Trials*. 2011;12:106.
- Christensen E. Methodology of superiority vs. equivalence trials and non-inferiority trials. *J Hepatol*. 2007;46(5):947-954.
- Mulla SM, Scott IA, Jackevicius CA, et al. How to Use a Noninferiority Trial: Users' Guides to the Medical Literature. *JAMA*. 2012;308(24):2605-2611.
- Piaggio G, Elbourne DR, Pocock SJ, et al. Reporting of noninferiority and equivalence randomized trials: extension of the CONSORT 2010 statement. *JAMA*. 2012;308(24):2594-2604.