



Noninferiority Randomized Controlled Trials

Kevin S. Kim^{1,2}, An-Wen Chan^{3,4}, Emilie P. Belley-Côté^{2,5} and Aaron M. Drucker^{3,4}

From 2005 to 2015, the publication of noninferiority trials increased by six-fold. Noninferiority trials assess whether a new treatment's efficacy is comparable with that of the standard of care and have several appeals. Noninferiority trials can evaluate for both noninferiority and superiority of a new treatment. In addition, multiple treatment modalities exist, and new treatments may be advantageous for reasons beyond efficacy. Common elements of trial design such as the research question, outcomes, statistical analysis, and interpretation of results differ meaningfully between noninferiority trials and superiority trials. The noninferiority margin, constancy assumption, and assay sensitivity are unique aspects of noninferiority trials. As with all randomized controlled trials, patient engagement in and reporting of noninferiority trials are also important. In this review, we discuss the methodological considerations and limitations of noninferiority trials.

Journal of Investigative Dermatology (2022) **142**, 1773–1777; doi:10.1016/j.jid.2022.04.015

INTRODUCTION

Most randomized trials test whether a new intervention is more efficacious than a placebo or active comparator. However, regulatory agencies do not require a demonstration of superiority over existing therapies to receive market approval. New interventions might aim to improve other aspects of treatment, such as better safety, reduced cost, or increased convenience. In those circumstances, noninferiority trials help to establish whether a new treatment is as efficacious as the standard of care. Noninferiority trials are commonly seen in conditions such as psoriasis and atopic dermatitis, where multiple efficacious treatments exist. Although we use the term efficacy throughout this article, noninferiority trial design is not exclusive to explanatory trials. The noninferiority trials we cite as examples compared the effectiveness of treatments already used in clinical practice. Williams et al. (2015) discuss the explanatory-pragmatic continuum, pragmatic trials, and its considerations in a previous article in the “Research Techniques Made Simple” series.

Planning a noninferiority trial involves several unique considerations. Table 1 summarizes the differences between noninferiority and traditional superiority trials, with each row reflecting the progression from developing the research question to interpreting the results.

RESEARCH QUESTION

Constructing the research question is an important first step of any randomized trial, whether the intent is to assess superiority or noninferiority. The research question is framed differently in noninferiority trials than in traditional superiority trials (Table 1). For example, the research question in the BLISTER (Bullous Pemphigoid Steroids and Tetracyclines) study, a noninferiority trial comparing doxycycline with prednisolone for bullous pemphigoid, can be summarized as, “In adults diagnosed with bullous pemphigoid, is starting treatment with oral doxycycline noninferior to oral prednisolone for blister control at 6 weeks?” (Williams et al., 2017). A noninferiority design was sensible for this study because prednisolone was known to be efficacious but was associated with a substantial risk of serious adverse events; conversely, doxycycline had a more favorable safety profile, and although its efficacy was unclear, it was unlikely to be more efficacious than prednisolone.

OUTCOMES IN NONINFERIORITY TRIALS

The primary outcome in a noninferiority trial often measures the efficacy of the treatment relative to that of an active control that is considered standard care. The active control should have shown efficacy for this outcome in a previous trial. When two outcomes are both important (e.g., efficacy, safety) a

¹Department of Health Research Methodology, Evidence, and Impact, Faculty of Health Sciences, McMaster University, Hamilton, Ontario, Canada;

²Population Health Research Institute, Hamilton, Ontario, Canada; ³Division of Dermatology, Department of Medicine, University of Toronto, Toronto, Ontario, Canada; ⁴Women's College Research Institute, Women's College Hospital, Toronto, Ontario, Canada; and ⁵Division of Cardiology, Department of Medicine, Faculty of Health Sciences, McMaster University, Hamilton, Ontario, Canada

Correspondence: Kevin Kim, McMaster University Medical Centre, 2C Area, 1280 Mainstreet West, Hamilton, Ontario L8S 4K1, Canada. E-mail: kevin.kim@medportal.ca

Abbreviations: BCC, basal cell carcinoma; CI, confidence interval; ITT, intention-to-treat; MAL-PDT, methyl aminolevulinate photodynamic therapy; PP, per-protocol; SAP, statistical analysis plan

SUMMARY POINTS

- Noninferiority trials evaluate whether a new intervention with potential secondary benefits, such as reduced side effects, is similarly efficacious to the standard of care according to a prespecified noninferiority margin.
- The noninferiority margin is critical in the design, analysis, and interpretation of the primary outcome and should reflect both statistical and clinical considerations, including patient input.
- Internal validity is harder to ensure in noninferiority trials than in superiority trials, and a placebo group is helpful whenever possible.
- It is unclear how noninferiority margins should be established, leading to varied noninferiority margins between trials answering similar research questions.

co-primary outcome may be used. In the BLISTER trial, the investigators chose co-primary outcomes assessing the noninferiority in blister control and superiority in adverse events.

Secondary outcomes in noninferiority trials capture the potential advantages of the new intervention. For example, the investigators of the trials **D**albavancin for **I**nfections of the **S**kin **C**ompared to **V**ancomycin at an **E**arly **R**esponse 1 and 2 showed weekly dalbavancin to be noninferior to the standard of care (daily vancomycin-linezolid) for an early clinical response indicating treatment success when managing acute bacterial skin infections (Boucher et al., 2014). Patients receiving dalbavancin also had a shorter course of treatment (i.e., convenience) and experienced fewer treatment-related adverse events (i.e., safety).

NONINFERIORITY MARGIN

Defining the noninferiority margin

Noninferiority trials conclude whether an intervention retains acceptable efficacy compared with the active control using the noninferiority margin. The noninferiority margin defines how much worse a new treatment can be than the active control for it to be acceptable in clinical practice given its other potential benefits such as safety or cost.

The noninferiority margin is a critical element in the design, analysis, and interpretation of noninferiority trials. The reporting of noninferiority trials is standardized by the extension to the 2010 CONSolidated Standards of Reporting Trials guideline (Piaggio et al., 2012). This extension promotes the reporting of the margin and its rationale, but there are currently no standards for establishing the noninferiority margin (Piaggio et al., 2012). Margins are established by regulatory agencies in select areas of research, but most research questions require investigators to establish the noninferiority margin. Regulatory guidance defines the noninferiority margin as a value that should reflect both clinical and statistical considerations on the basis of past evidence (ICH, 2001), and patient values should be reflected in the

Table 1. Considerations for Designing a Noninferiority Trial Compared with a Superiority Trial Framework

Aspects of Randomized Controlled Trial	Noninferiority Trials	Superiority Trials
Purpose	To show that new intervention X is an acceptable alternative to active control Y and possibly preferable for benefits beyond efficacy.	To show that new intervention X is more efficacious than active control Y and should be the preferred treatment.
Research question	“In adults with chronic plaque psoriasis, is X noninferior to Y for PASI100 at 1 year?”	“In adults with chronic plaque psoriasis, is X superior to Y for PASI100 at 1 year?”
Null hypothesis	“X is inferior to Y by at least 10% (noninferiority margin) for PASI100 at 1 year”	“There is no difference between X and Y for PASI100 at 1 year.”
Main analysis	Both intention-to-treat and per-protocol	Intention-to-treat
Possible interpretations	Noninferiority testing: X is noninferior to Y X is inferior to Y Cannot conclude that X is noninferior to Y (inconclusive) Superiority testing: same as superiority trials	X is superior to Y X is inferior to Y There is no evidence of a significant difference between X and Y

margin whenever possible (Acuna et al., 2019; U.S. Food and Drug Administration, 2016).

The BLISTER trial investigators established the noninferiority margin after surveying dermatologists in the United Kingdom (Williams et al., 2017). Dermatologists were willing to accept a 25% relative reduction in blister control if there was at least a 20% relative reduction in serious side effects with doxycycline compared with the side effects with oral prednisolone.

Without a gold standard for determining noninferiority margins, a wide range of margins for trials addressing similar questions are possible. This can lead to trials with comparable treatment effect sizes being interpreted differently depending on the noninferiority margin used. For example, Szeimies et al. (2008) compared methyl aminolevulinic photodynamic therapy (MAL-PDT) with surgery for treating superficial basal cell carcinoma (BCC). The primary outcome was the clearance rate at 3 months, and the noninferiority margin was -15% absolute difference. MAL-PDT was noninferior to surgery at 3 months (92.2 vs. 99.2%, absolute difference = -7%, 95% confidence interval [CI] = -12.1 to -1.9). In another noninferiority trial, Sinx et al. (2020) compared curettage and imiquimod with surgery for nodular BCC (4–20 mm). The primary outcome was the clearance rate at 12 months but with a more conservative noninferiority margin of -8% absolute difference. At 12 months, the investigators could not conclude that the combination of curettage and imiquimod was noninferior to surgery (86.3 vs. 100%, absolute difference = -13.7%, 95% CI = -21.6% to -5.8%). If Szeimies et al. (2008) had used the same noninferiority margin as Sinx et al. (2020) (i.e., -8% absolute difference), they too would not have concluded noninferiority.

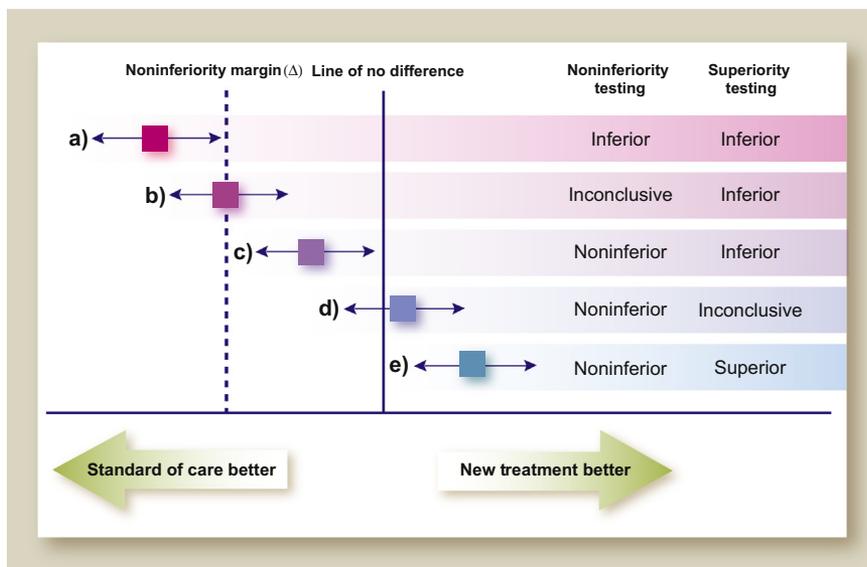


Figure 1. Interpretation of a noninferiority trial based on testing for noninferiority and superiority. Squares (■) represent point estimates, and arrows (↔) represent the confidence interval. (a) A new treatment is inferior to the standard of care for both noninferiority and superiority testing. (b) Results are inconclusive for noninferiority; new treatment is inferior to the standard of care on superiority testing. (c) A new treatment is both noninferior and inferior to the standard of care on the basis of noninferiority and superiority testing, respectively. (d) A new treatment is noninferior to the standard of care; results are inconclusive (no evidence of a difference) for superiority testing. (e) A new treatment is superior to the standard of care. Illustration assistance was provided by Jan Ruvido Stebbins, Ruvido Medical Illustration, Dexter, MI.

INTERPRETING NONINFERIORITY TRIALS

The primary analysis of a noninferiority trial is interpreted using the noninferiority margin and the lower bound of the primary outcome’s CI (Figure 1). A conclusion of noninferiority is made when the CI does not cross the noninferiority margin, regardless of the statistical line of no difference (Figure 1c–e). Noninferiority cannot be concluded if only the point estimate is less than the noninferiority margin.

An advantage of noninferiority trials is that if prespecified, they can test for both noninferiority and superiority for a given outcome. It is easy to interpret the results of a noninferiority trial when a new treatment shows superiority (Figure 1e). It is also easy to interpret the results when the new treatment is worse than the active control relative to both the noninferiority margin and the line of no difference (Figure 1a).

A new treatment may not be superior on the basis of a 5% level of significance ($P < 0.05$) but may be noninferior on the basis of the noninferiority margin (Figure 1c). This can be challenging because interpretations based on statistical significance differ from those based on clinical relevance (as reflected by noninferiority margins). Consider the trial by Szeimies et al. (2008) mentioned earlier comparing MAL-PDT with surgery for treating superficial BCC. With a noninferiority margin of -15% absolute difference, the authors concluded noninferiority of MAL-PDT to surgery for clearance of superficial BCC at 3 months (92.2 vs. 99.2%, absolute difference = -7%, 95% CI = -12.1 to -1.9). However, the CI also suggests that MAL-PDT is significantly less efficacious than surgery. Nevertheless, the conclusion of this trial is that MAL-PDT is noninferior to surgery because the noninferiority margin reflects both statistical and clinical considerations.

STATISTICAL CONSIDERATIONS

Hypothesis testing

Hypothesis testing in a traditional superiority trial is designed around rejecting the null hypothesis that the treatments do not differ (Table 1). When there is no significant difference, it

is appropriate to say, “We are unable to reject the null hypothesis that there is no difference between the new treatment and active control.” In contrast, hypothesis testing in a noninferiority trial is designed around rejecting the null hypothesis where two treatments differ more than a prespecified noninferiority margin (Table 1). When the difference exceeds the margin, it is appropriate to say, “We are unable to reject the null hypothesis that the difference between treatments is equal to or more than the prespecified noninferiority margin.”

The implication of the null hypothesis used in traditional superiority trials is that it can never be proven. In other words, a nonsignificant difference does not mean that there is no difference between treatments; it means that the analysis failed to reject the null hypothesis (Altman and Bland, 1995). For example, Bernard et al. (1992) compared roxithromycin with penicillin for the treatment of erysipelas, assessing overall efficacy at 30 days. Without a defined noninferiority margin, they concluded that roxithromycin was as efficacious as penicillin because there was no significant difference (84 vs. 76%, absolute difference = 8%, 95% CI = -11.1 to 26%) (Bernard et al., 1992). In a subsequent noninferiority trial,

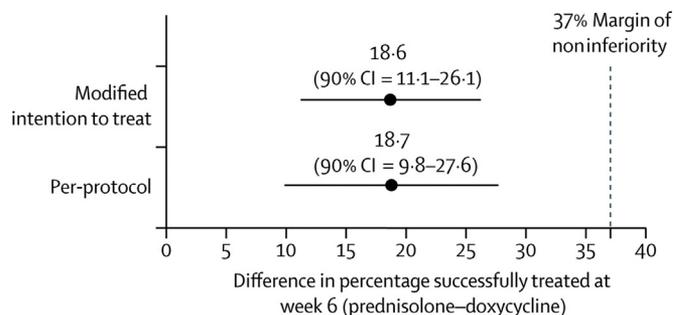


Figure 2. Figure 2 from the BLISTER study showing the CI from two approaches of analysis: noninferiority margin and line of no difference (Williams et al., 2017). BLISTER, Bullous Pemphigoid Steroids and Tetracyclines; CI, confidence interval.

Bernard et al. (2002) compared pristinamycin with penicillin with a predefined noninferiority margin of -10% absolute difference and showed that pristinamycin is both noninferior and superior to penicillin (81 vs. 67%, absolute difference = 14%, one-sided 97.5% CI = 3.3 to $\infty\%$) (Bernard et al., 2002). The earlier trial assessing roxithromycin would not have concluded noninferiority if a 10% absolute margin had been used to evaluate comparative efficacy against penicillin. As such, conclusions of noninferiority (e.g., as effective as or both treatments are effective or both are equally effective) should be avoided in superiority trials and reserved for interpretation of noninferiority or equivalence trials.

Intention-to-treat versus per-protocol analyses

Intention-to-treat (ITT) analysis includes all participants on the basis of the treatment group they were randomized to, whereas per-protocol (PP) analysis includes only participants who adhered to the treatment protocol. In superiority trials, ITT is the preferred method for analysis; it is a more conservative and robust approach because it mitigates the bias arising from nonadherence to treatment and loss to follow-up. However, ITT analyses can have the opposite effect in noninferiority trials by making the difference between groups smaller. This increases the likelihood of showing noninferiority, particularly if nonadherence to treatment protocol is high. There is currently no gold standard in the analysis of noninferiority trials, so investigators should include both ITT and PP analyses and interpret the results holistically (Mo et al., 2020). Any foreseeable analyses in the trial (e.g., primary analysis using ITT and PP) should be prespecified in the statistical analysis plan (SAP). A clear and comprehensive SAP supports the reproducibility of trial results and their conclusions. Gamble et al. (2017) published a guidance document in 2017 listing the recommended items that should be addressed in a SAP. In their SAP, BLISTER study investigators prespecified using both ITT and PP analyses to show the noninferiority of doxycycline to prednisolone (Williams et al., 2017). Figure 2 illustrates both the analyses and the noninferiority margin used to support their conclusion.

CONSTANCY ASSUMPTION AND ASSAY SENSITIVITY

The internal validity of noninferiority trials is contingent on the constancy assumption and assay sensitivity. Constancy assumption is the assumption that the effects of the standard of care from past trials remain constant in the current noninferiority trial. The effect estimate and event rate from past trials are often used to establish the noninferiority margin. If these values differ between past trials and the current noninferiority trial, it may result in an inaccurate noninferiority margin that could lead to false conclusions (i.e., type I or type II error). Assay sensitivity is the ability to discriminate an effective treatment from a less efficacious or ineffective treatment. For example, a trial has assay sensitivity if it could show that the active control is better than the placebo (if the placebo were to be included as a study group).

By definition, a trial showing superiority, regardless of its design, has assay sensitivity. However, a noninferiority trial showing noninferiority does not necessarily have assay sensitivity because various factors can reduce the expected efficacy of the active control (relative to that of the placebo) or reduce the effect size between the new and control treatments. Regulatory bodies thus recommend the inclusion of a placebo group (i.e., three-arm trial) where possible to show assay sensitivity in a noninferiority trial (U.S. Food and Drug Administration, 2016), but a placebo group is not always ethical. In the absence of a placebo group, assay sensitivity is more likely to be achieved if the noninferiority trial fulfills three criteria: (i) the active control has been shown to be efficacious in previous trials; (ii) the study design is not substantially different from those of previous trials in terms of the enrolled patient population, concomitant interventions, outcomes, and analysis; and (iii) trial quality is robust in terms of adherence to treatment and follow-up protocols (U.S. Food and Drug Administration, 2016).

PATIENT INVOLVEMENT

Patient engagement is particularly important in justifying and designing noninferiority trials. If a new intervention is thought to be more convenient, evidence of patient inconvenience with the standard of care should be established before embarking on a trial. Patients should contribute to choosing outcomes that measure the new intervention's acceptability. A new biological therapy may show noninferiority with a favorable safety profile, but patients may still choose the standard of care if the cost is exorbitant or if it requires injection. Most importantly, the noninferiority margin should reflect the level of efficacy patients are willing to forego for other benefits and not just the opinion of clinicians.

SUMMARY

Noninferiority trials have distinct advantages and limitations. Investigators and readers should be attentive to how noninferiority margins are established, how noninferiority trials are analyzed, and how to make appropriate conclusions about noninferiority. Noninferiority trials often evaluate the efficacy and other potential advantages of the new intervention. They should report both ITT and PP analyses to inform conclusions of noninferiority. Interpreting noninferiority trials can be challenging, and a figure including the CI, noninferiority margin, and line of no difference can facilitate understanding.

ORCIDs

Kevin S. Kim: <http://orcid.org/0000-0002-6117-8158>
An-Wen Chan: <http://orcid.org/0000-0002-4498-3382>
Emilie P. Belley-Côté: <http://orcid.org/0000-0002-5071-076X>
Aaron M. Drucker: <http://orcid.org/0000-0002-7388-9475>

CONFLICT OF INTEREST

AMD has received compensation from the *British Journal of Dermatology* (reviewer and section editor), the American Academy of Dermatology (guidelines writer), and the National Eczema Association (grant reviewer). EPBC has received grant support from Bayer, BMS-Pfizer, and Roche outside the submitted work. She holds a career award from the Department of Medicine at McMaster University (Hamilton, Canada) and a New Investigator Award from the Heart and Stroke Foundation. KSK is a recipient of the Canada

Graduate Scholarship—Doctoral Award from the Canadian Institutes of Health Research.

AUTHOR CONTRIBUTIONS

Conceptualization: KSK, AWN, AMD; Data Curation: KSK; Investigation: KSK, AWN, AMD; Project Administration: KSK; Visualization: KSK, AWN, EPBC, AMD; Supervision: AWN, EPBC, AMD; Writing - Original Draft Preparation: KSK; Writing - Review and Editing: KSK, AWN, EPBC, AMD

SUPPLEMENTARY MATERIAL

Supplementary material is linked to the online version of the paper at www.jidonline.org and at <https://doi.org/10.1016/j.jid.2022.04.015>.

REFERENCES

- Acuna SA, Chesney TR, Baxter NN. Incorporating patient preferences in noninferiority trials. *JAMA* 2019;322:305–6.
- Altman DG, Bland JM. Absence of evidence is not evidence of absence. *BMJ* 1995;311:485.
- Bernard P, Chosidow O, Vaillant L, French Erysipelas Study Group. Oral pristinamycin versus standard penicillin regimen to treat erysipelas in adults: randomised, non-inferiority, open trial. *BMJ* 2002;325:864.
- Bernard P, Plantin P, Roger H, Sassolas B, Villaret E, Legrain V, et al. Roxithromycin versus penicillin in the treatment of erysipelas in adults: a comparative study. *Br J Dermatol* 1992;127:155–9.
- Boucher HW, Wilcox M, Talbot GH, Puttagunta S, Das AF, Dunne MW. Once-weekly dalbavancin versus daily conventional therapy for skin infection. *N Engl J Med* 2014;370:2169–79.
- Gamble C, Krishan A, Stocken D, Lewis S, Juszcak E, Doré C, et al. Guidelines for the content of statistical analysis plans in clinical trials. *JAMA* 2017;318:2337–43.
- ICH. International Council for Harmonisation of Technical Requirements for Registration of Pharmaceutical for Human Use (ICH). <https://www.hsa.gov.sg/therapeutic-products/international-collaboration/ich#:~:text=Overview-,The%20International%20Council%20for%20Harmonisation%20of%20Technical%20Requirements%20for%20Pharmaceuticals,that%20safe%2C%20effective%2C%20and%20high; 2001> (accessed December 20, 2021).
- Mo Y, Lim C, Watson JA, White NJ, Cooper BS. Non-adherence in non-inferiority trials: pitfalls and recommendations [published correction appears in *BMJ* 2020;370:2692]. *BMJ* 2020;370:m2215.
- Piaggio G, Elbourne DR, Pocock SJ, Evans SJW, Altman DG, CONSORT Group. Reporting of noninferiority and equivalence randomized trials: extension of the CONSORT 2010 statement. *JAMA* 2012;308:2594–604.
- Sinx KAE, Nelemans PJ, Kelleners-Smeets NWJ, Winnepenninckx VJL, Arits AHMM, Mosterd K. Surgery versus combined treatment with curettage and imiquimod for nodular basal cell carcinoma: one-year results of a noninferiority, randomized, controlled trial. *J Am Acad Dermatol* 2020;83:469–76.
- Szeimies RM, Ibbotson S, Murrell DF, Rubel D, Frambach Y, De Berker D, et al. A clinical study comparing methyl aminolevulinic photodynamic therapy and surgery in small superficial basal cell carcinoma (8–20 mm), with a 12-month follow-up. *J Eur Acad Dermatol Venereol* 2008;22:1302–11.
- U.S. Food and Drug Administration. Non-Inferiority Clinical Trials to Establish Effectiveness. <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/non-inferiority-clinical-trials>, 2016. (Accessed December 20, 2021).
- Williams HC, Burden-Teh E, Nunn AJ. What is a pragmatic clinical trial? *J Invest Dermatol* 2015;135:1–3.
- Williams HC, Wojnarowska F, Kirtschig G, Mason J, Godec TR, Schmidt E, et al. Doxycycline versus prednisolone as an initial treatment strategy for bullous pemphigoid: a pragmatic, non-inferiority, randomised controlled trial [published correction appears in *Lancet* 2017;390:1948]. *Lancet* 2017;389:1630–8.