Stroke

TOPICAL REVIEW

Section Editors: Seana Gall, PhD, Linxin Li, DPhil, and Mathew J. Reeves, PhD

Understanding Noninferiority Trials: What Stroke Specialists Should Know

Fabiano Cavalcante, MD, MSc; Kilian M. Treurniet, MD, PhD; Manon Kappelhof, MD, PhD; Johannes Kaesmacher, MD, PhD; Hester F. Lingsma, MSc, PhD; Jeffrey L. Saver, MD; Jan Gralla, MD, MSc; Urs Fischer, MD; Charles B. Majoie, MD, PhD; Yvo B.W.E.M. Roos, MD, PhD

ABSTRACT: Noninferiority trials aim to prove that the efficacy, defined in terms of a key clinical outcome, of a new treatment is not meaningfully worse than that of an established active control. Noninferiority trials are important when other aspects of care can be improved, such as convenience, toxicity, costs, and safety (nonefficacy benefits). While the motivation for a noninferiority trial is straightforward, the design, execution, and interpretation of these trials is not a trivial task. Several safeguards that protect superiority trials from incorrect conclusions do not apply or even work in reverse for noninferiority trials. This review aims to provide stroke clinicians and researchers with a general overview of noninferiority trials and a deeper understanding of 10 pitfalls they should consider when designing and interpreting such trials.

GRAPHIC ABSTRACT: A graphic abstract is available for this article.

Key Words: clinical trials ■ research design ■ statistics ■ stroke

"Let us cast lots, half the patients I will treat without bloodletting and the other half you will treat according to your knowledge." The 1648 challenge by van Helmont is often cited as the first description of a randomized trial. As the aim was to withdraw a potentially ineffective treatment, the trial by van Helmont could also be considered the first noninferiority trial.

Today, most clinical trials aim to investigate whether a new intervention is superior to standard practice. However, with the increasing success rates of modern treatments, identifying new treatments that improve efficacy outcomes is becoming more difficult or even unnecessary for many conditions.[1] Meanwhile, expanded treatment options and increased life expectancy have led to rising concerns regarding health care resources worldwide. Therefore, researchers are increasingly shifting their attention to new treatments and diagnostic studies that could bring other benefits such as convenience in administration, lower costs, dose reduction, or reducing toxicity, without an unacceptable loss of efficacy. Here, the question is no longer whether the new intervention is superior (more efficacious) but rather whether it can be considered noninferior (ie, not meaningfully worse in terms of efficacy) to the standard of care.[2–6] Noninferiority trials are also becoming increasingly important in the field of stroke, where they are essential tools in providing regulatory approval for new endovascular devices and medical treatments. In recent years, noninferiority trials have investigated topics such as thrombectomy devices,[7–11] direct oral anticoagulants for atrial fibrillation,[12,13] withholding intravenous thrombolysis in patients who can undergo endovascular treatment,[14–20] and tenecteplase as an alternative to alteplase.[21–24]

Although the rationale for a noninferiority trial is intuitive, statistically demonstrating noninferiority involves additional particularities.[2,4,6,25,26] Clinicians who are mostly familiar with superiority trials should be aware of several pitfalls when interpreting the results of noninferiority trials

Correspondence to: Fabiano Cavalcante, MD, MSc, Department of Radiology and Nuclear Medicine, Amsterdam UMC, University of Amsterdam, Room G1-230, Meibergdreef 9, Amsterdam 1105AZ, the Netherlands. Email f.cavalcante@amsterdamumc.nl

This article is Part 5 in a 5-part series. Parts 1-3 appear in the November 2024 issue of *Stroke*. Part 4 appears in the January 2025 issue.
For Sources of Funding and Disclosures, see pages 550–551.

Stroke is available at www.ahajournals.org/journal/str

as they become more frequent in the field. Researchers considering executing noninferiority trials should also be aware of these intricacies to prevent a mismatch between the chosen trial design and the research question they try to answer. Here, we aim to provide an overview of how noninferiority trials work and highlight 10 key pitfalls in designing and interpreting these trials.

SUPERIORITY, EQUIVALENCE, AND NONINFERIORITY

Superiority Trials

Superiority trials test whether a new treatment improves a key outcome when compared with standard care. Superiority is statistically assessed by establishing a null hypothesis, stating that there are no outcome differences between the groups. The null hypothesis is then tested and possibly rejected if the data provide enough confidence for a difference between the studied treatments. In the usual frequentist analyses, this is based on assessing whether the CI around the estimated treatment effect excludes the point of no difference (Figure 1). Different statistical frameworks can be used, but ultimately, all methods will test whether one can confidently reject the null hypothesis.

Importantly, demonstrating that treatment A is better than treatment B requires showing both that treatment A is statistically superior to treatment B and that the magnitude of this difference is clinically worthwhile. If the benefit is less than the minimal clinically important difference (MCID), the new treatment will be statistically superior but not clinically superior. A small, clinically irrelevant difference can be statistically significant as long as one can be confident enough that the observed difference is not caused by chance. Consequently, interpretation of superiority trials requires consideration of the clinical and statistical significance of any observed treatment effect.

Noninferiority and Equivalence Trials

In contrast to superiority trials, noninferiority and equivalence trials aim to show that an alternative treatment strategy is not meaningfully different from the standard of care in terms of efficacy. Importantly, efficacy is defined by one or a few key clinical outcomes (eg, survival rate or 90-day functional outcome). However, patients and health care systems are interested in many other aspects of care (convenience, toxicity, side effects, costs, treatment duration, and hospitalization requirements). In this context, a new treatment providing nonefficacy benefits can be preferable to an existing treatment, as long as its efficacy is not meaningfully worse.

Statistically proving that 2 treatments are exactly equal would require data from the entire population universe.[27] Because clinical trials are limited to just a sample of the population, researchers must accept some degree of uncertainty regarding the effect estimates of the tested treatments. Noninferiority trials aim to determine whether a potential difference between the groups is at least not larger than an acceptable noninferiority margin (−Δ). Here, the null hypothesis is no longer null but instead states that the difference between the treatments is larger than the predefined noninferiority margin (−Δ). If the data provide enough confidence that the difference in treatment effect is not larger than −Δ, the null hypothesis can be rejected and noninferiority can be claimed (Figure 1).

The aim of noninferiority trials is to exclude the possibility that the new treatment is meaningfully worse (<−Δ) than the existing treatment. If the new treatment is shown to be noninferior, a nested analysis can then test whether the new treatment is also superior to the control. The new treatment can be both noninferior and superior or noninferior but not superior. Equivalence trials, however, aim to prove that a new treatment is neither worse (<−Δ) nor better (>+Δ) than a standard treatment. Such trials are common in pharmacological studies of bioequivalence between 2 formulations of the same drug, for example, by showing that they yield comparable blood

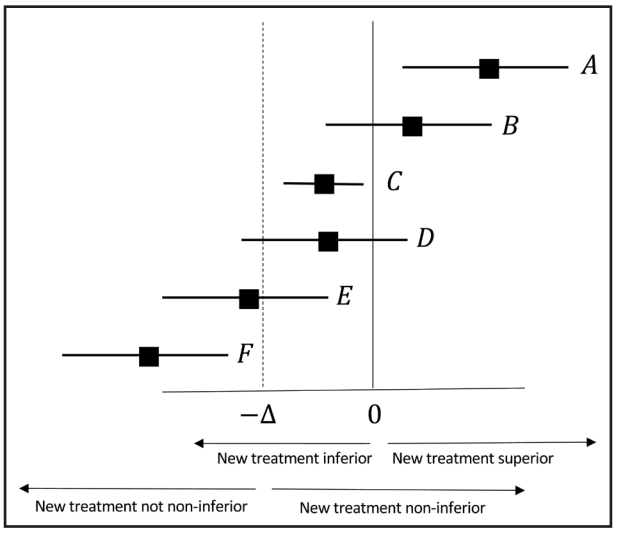

Figure 1. Treatment effect estimates and CIs for a hypothetical new treatment.
The point of no difference is indicated as 0 (equivalent to 1 on the odds ratio scale) and the noninferiority margin as −Δ. To conclude superiority (or inferiority), the CI should not include the point of no difference. To conclude noninferiority, the CI should not include the noninferiority margin. As the point of no difference does not play any role in determining noninferiority, it becomes clear that inferiority and noninferiority are not opposing concepts. A: new treatment is superior and noninferior. B: new treatment is noninferior, but superiority cannot be claimed. C: new treatment is simultaneously inferior and noninferior, illustrating that noninferiority and inferiority are unrelated. D: new treatment for which no conclusions can be drawn regarding superiority or noninferiority. E: noninferiority cannot be claimed, and inferiority has been established. F: noninferiority has been rejected (ie, the new treatment is not noninferior).

 Stroke. 2025;56:543–552. DOI: 10.1161/STROKEAHA.124.048024

concentrations.[28] With the exception of their 2-sided nature, equivalence trials raise considerations similar to noninferiority trials. As they are less frequent in the context of stroke, we will focus on noninferiority trials from now on.

In the following section, we highlight 10 key pitfalls that should be considered when designing and interpreting noninferiority trials (Table).

Table. Summary of Recommendations for Interpreting and Designing Noninferiority Trials

Nonefficacy benefits
A noninferiority trial should be motivated by the possibility of obtaining nonefficacy benefits. Carefully assess how realistic the claims of nonefficacy benefits are and test to confirm these benefits whenever possible.
The noninferiority margin
Consider how the noninferiority margin was defined and how realistically the margin would correspond to a clinically acceptable difference.
The constancy assumption
Noninferiority trials should only test new treatments against controls that are proven effective. If historical data are used, assess how realistic it is that the control treatment is still effective vs placebo in the current population.
Assay sensitivity
Be aware of study issues that potentially limit the ability of the trial to detect real differences between treatment arms as they can lead to incorrect claims of noninferiority.
Spinning of conclusions: not noninferior does not mean inferior
Avoid incorrect interpretation and spinning of results. Failure to demonstrate noninferiority of a new treatment does not confirm superiority of the control treatment.
Impact of expected vs observed treatment effect, and the role of interim analyses
When designing noninferiority trials, optimistic estimations of benefit in favor of the experimental treatment can affect power calculations and result in small sample size requirements and potentially inconclusive results. Carefully consider interim analysis plans to match a noninferiority paradigm.
Study population: ITT and per protocol
Noninferiority trials should report both ITT and per-protocol analyses. Crossovers, protocol violations, and loss to follow-up may artificially reduce the differences between groups, increasing the risk of false noninferiority claims.
Absolute and relative risk measurements
Interpret the results using both relative and absolute risk metrics. Small differences in relative risks can represent large differences in absolute risks and vice versa. This relationship is influenced by the observed rate of events.
Impact of treatment effect heterogeneity
Assessment of treatment effect heterogeneity requires judicious consideration in noninferiority trials. Ensure that treatment effects are consistent throughout the study population and that a noninferiority result in the overall population is not the result of relevant treatment effect heterogeneity.
Follow recommended planning, executing, and reporting guidelines
Reporting and assessment of noninferiority trials should adhere to guidelines (CONSORT) and recommendations from regulatory medical agencies (FDA and EMA).

CONSORT indicates Consolidated Standards of Reporting Trials; EMA, European Medical Agency; FDA, Food and Drug Administration; and ITT, intention to treat.

TEN PITFALLS IN INTERPRETING AND DESIGNING NONINFERIORITY TRIALS

Consider the Rationale for a Noninferiority Design: Nonefficacy Benefits

The first question that should be asked when interpreting a noninferiority trial is why it was conducted to begin with. In the past, noninferiority trials have been criticized with some claiming that they are unethical.[29] After all, no patient is interested in receiving a possibly worse treatment, even if the efficacy difference is not meaningful.

These criticisms overlook the underlying motivation for a noninferiority trial: a new treatment could provide other relevant benefits (nonefficacy benefits) that are not captured in the efficacy outcomes. Without such trials, it would be hard to provide evidence for safe de-escalation of medical care.[30] For example, noninferiority trials provided the evidence to safely reduce the duration of antituberculosis treatment, replace toxic drugs from chemotherapy schemes, and switch from an intravenous antibiotic treatment requiring hospitalization to oral antibiotics.[31,32] Ultimately, patients are interested in less invasive treatments, faster return to home, or fewer side effects, as long as the efficacy is not meaningfully worse.

In the context of stroke, recent noninferiority trials have investigated thrombolysis with tenecteplase as an alternative to alteplase. For these trials, easier administration (single bolus versus 60-minute infusion) is a clear nonefficacy benefit.[21,22,33] Other trials investigated skipping thrombolysis for patients who can directly receive endovascular treatment, with expected nonefficacy benefits including less hemorrhagic complications, faster procedures, and reduced costs.[14–20]

The expected nonefficacy benefits are often mentioned in noninferiority trials; however, testing for these benefits is not always performed. Some benefits are not testable or debatable, but critical readers should carefully assess how realistic the claimed nonefficacy benefits are and whether they were tested when possible (eg, if researchers claim that a new treatment has fewer side effects, the trial should report results confirming this benefit).

Consider the Noninferiority Margin: The Price of Certainty

The noninferiority margin ($-\Delta$) is a key concept in noninferiority trials. Importantly, it does not represent how much worse the new treatment is but rather how much uncertainty one is willing to accept around the difference between the effects of the treatments.

Correct interpretation of noninferiority trials includes assessing how acceptable the noninferiority margin is and how it was defined by the investigators. Defining a noninferiority margin involves a trade-off between acceptable

uncertainty and the required sample sizes. If the margin is too generous, a new treatment could be considered noninferior while causing a substantial reduction in efficacy. Conversely, if the margin is too strict, proving noninferiority will require infeasibly large sample sizes. In fact, if a noninferiority margin of 1.3% absolute risk reduction is used, determining noninferiority between 2 identical treatments can require tens of thousands of patients (Figure 2).[34,35] Despite the importance of the noninferiority margin, reviews found that 54% to 80% of published noninferiority studies had issues regarding justification of the noninferiority margin.[36–39]

Due to variability in approaches to define a clinically meaningful difference, interpretation of 2 studies with identical results can differ if different margins ($-\Delta$) are selected. In a recent example, 2 oncology trials published in the same issue of *Lancet* in 2019 found almost identical treatment effects comparing 6- versus 12-month trastuzumab treatment in patients with HER2 (human epidermal growth factor receptor 2)-positive breast cancer (hazard ratio, 1.07 [95% CI, 0.93–1.24] in the PERSEPHONE trial [Duration of Trastuzumab With Chemotherapy in Women With Early Stage Breast Cancer: Six Months Versus Twelve] and 1.08 [95% CI, 0.93–1.25] in the PHARE trial [Protocol of Herceptin Adjuvant With Reduced Exposure]).[40,41] However, the trials had different prespecified noninferiority margins (1.15 and 1.32), leading to opposite conclusions.

In the field of stroke, a recent study surveyed clinicians regarding acceptable noninferiority boundaries for skipping intravenous alteplase in patients presenting directly to endovascular treatment–capable centers. The study illustrated the subjective nature of meaningfully worse. Overall, the median acceptable uncertainty was an absolute difference of 3% in the rate of 90-day functional independence (mode, 5% [interquartile range, 1%–5%]). Respondents from an interventionalist background accepted higher uncertainty margins, while those with senior positions and working in centers with >200 thrombectomies annually required lower margins.[42] The level of certainty required to consider withholding thrombolytics also varied between countries where thrombolytic costs were covered by insurance/governments or not.[43]

Recently, Saver and Mistry[44] proposed 3 different frameworks for defining noninferiority margins based on the underlying motivation for the noninferiority trial. When a trial aims to demonstrate that a new treatment is at least as good as the standard and the outcome of interest is continuous, the MCID can be defined as the reduction of effect that would not have a detectable real-world value to patients (indistinguishability framework). For binary outcomes, however, a true MCID cannot be determined as a change in outcome would always have value to the affected patient.[45] In these cases, the noninferiority margin can be selected to preserve a reasonable fraction of the active control treatment effect (reasonable comparability framework). Alternatively, researchers may be interested in the distinct advantages provided by a new treatment over a standard of care. In these cases, the noninferiority margin should be seen as the degree to which one is willing to accept a loss of efficacy in exchange for offsetting benefits (balanced advantages

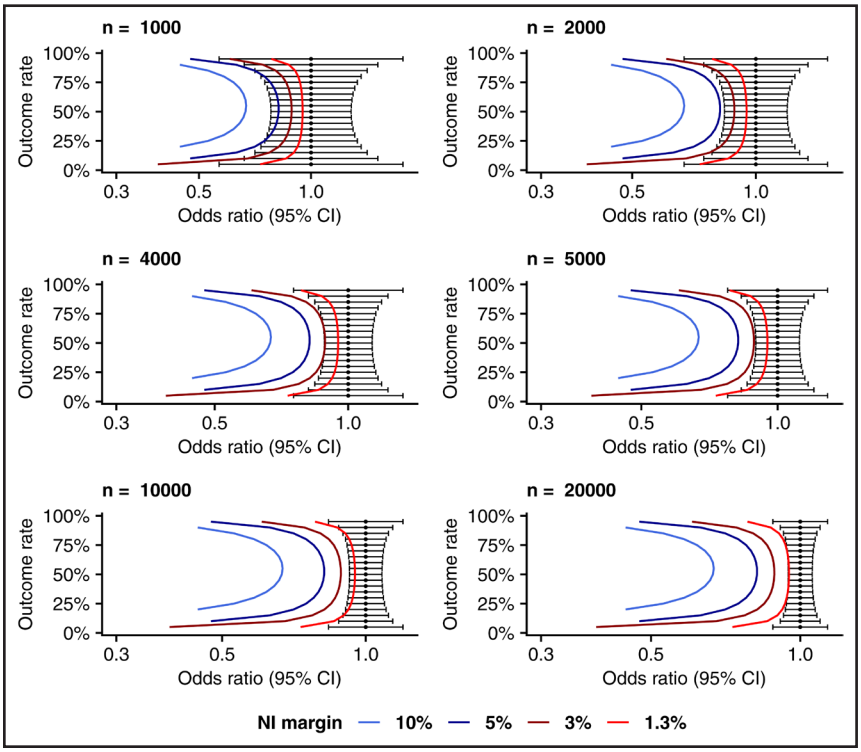

Figure 2. Noninferiority (NI) margins, expressed as odds ratios, corresponding to a 1.3% (light red), 3% (dark red), 5% (dark blue), and 10% (light blue) absolute risk difference between control and experimental treatment at different expected rates of outcome in the control arm.
The point estimates and 95% CIs represent the treatment effect of 2 identical treatments. Demonstrating NI with strict margins requires large sample sizes, even for identical treatments. The shape of the NI margin illustrates how the expected rate of events can influence the relative scale NI margin based on the predefined absolute risk difference.

framework). It is sensible to consider different strategies to define the noninferiority margin and different levels of strictness according to the aims of the study, specific diseases, and patient population.

Three approaches are commonly used to select a noninferiority margin.[3,7,24] The first is to define a margin based on patient perception of the smallest change in outcome that patients would consider valuable. This yields the MCID–outcome specific. For example, an analysis based, in part, on patients' behavior found the MCID–outcome specific for a novel neuroprotective treatment of ≥1.3% increase in modified Rankin Scale score of 0 to 2 (functional independence) at 90 days.[46] The second approach is based on clinicians' or stakeholders' perspectives on the difference needed to alter clinical practice, the MCID–practice changing. For example, for trials evaluating the labor- and capital-intensive intervention of endovascular thrombectomy, surveys suggested that clinician-experts considered the MCID–practice changing for increase in 90-day functional independence to be between 2.9% and 5.9%.[47]

These approaches are often based on qualitative studies and surveys, a naturally subjective approach. In fact, MCIDs derived from stroke expert surveys were shown to be higher than MCIDs derived from econometric modeling or from observations of physician behavior and medical guidelines.[46] Furthermore, the noninferiority margins selected in thrombectomy-related trials frequently exceed the MCIDs determined acceptable by surveys of stroke clinicians.[48] The tendency to inflate MCIDs introduces the risk of a margin that is larger than the difference between the control treatment and placebo, possibly leading to incorrect claims of noninferiority. For example, if a standard treatment improves the efficacy outcome by 7% versus a placebo, selecting a 10% noninferiority margin for a new treatment could lead to a noninferiority claim even if the new treatment is not better than placebo.[49]

A third approach is based on the previously established effect size of the standard treatment arm compared with placebo. The reasonably comparable effect size is defined as the noninferiority margin that ensures a new treatment both when (1) compared with no treatment, provides a benefit; and (2) compared with an existing active treatment, provides a benefit that may exceed or equal the active treatment and, at worst, is greater than the largest loss of effect considered reasonably comparable. For example, in EXTEND-IA TNK (Tenecteplase Versus Alteplase Before Endovascular Therapy for Ischemic Stroke),[21] a noninferiority trial comparing tenecteplase to alteplase, the investigators identified the rate of the primary outcome (reperfusion on the first angiographic assessment) based on historical data to be 7.5% (95% CI, 4.6%–11.5%). A margin of 2.3% was selected, thus preserving 50% of the most conservative estimate (4.6%) of the control treatment's efficacy. Alternative methods may be necessary to identify the ideal noninferiority margin, especially in situations where multiple estimates of the active control effect are available.[25,27,50] Furthermore, reliance on historical data can be problematic, as the responsiveness of the population may change over time, and there is no guarantee that the treatment effect of the active control in the trial population will be equal to historical results.

Noninferiority trials should carefully explain the rationale for their selected noninferiority margin, and clinicians interpreting noninferiority trials should be aware of the subjective nature of these margins.

Be Aware of Reliance on Historical Data: The Constancy Assumption

Unlike superiority trials, the efficacy of the control is crucial in noninferiority trials. However, efficacy of the active control is often based on previous trials. It is important to ensure that a control treatment proven effective in the past remains effective in current settings, considering differences in patient and health care service characteristics. Hypothetically, an ideal noninferiority trial would have 3 arms: placebo, active control, and new treatment. This would allow researchers to estimate the effect of the active control versus placebo, avoiding the risk of comparing 2 treatments that are not different from placebo, and thus noninferior to each other.[51] However, allocating patients to a placebo arm when an established effective treatment exists is not ethical. Without a placebo arm, noninferiority must rely on the constancy assumption,[3] the idea that a treatment proven to be effective in the past will preserve its effectiveness in a new setting. This assumption may not hold in face of changes in patient characteristics, medical practice, and health care systems.[52,53]

In stroke, endovascular treatment for large vessel occlusions has profoundly changed medical practice and health care organization. Despite historical data showing the efficacy of intravenous thrombolysis in the general stroke population, the introduction of thrombectomy to the treatment paradigm made historical data less applicable to patients who would receive thrombectomy immediately after intravenous thrombolysis. Clinical equipoise was evidenced by the execution of 6 randomized clinical trials investigating intravenous alteplase plus thrombectomy versus thrombectomy alone, the majority of which had a noninferiority design.[14,16–20] Interestingly, the hypothesis of these trials (that omitting alteplase would not be harmful) should have alerted the investigators that the constancy assumption might no longer hold. An individual participant data–level meta-analysis of these 6 trials found that time from onset to randomization and lytic start modified treatment effect, with alteplase plus thrombectomy being superior to thrombectomy alone only in patients presenting with short onset-thrombolyic time (<2.5 hours), the effect of alteplase before intravenous thrombolysis was shown to

quickly decrease and to be possibly harmful for patients presenting beyond 3 hours from onset.[15,54] The evidence in this population raises an important question for trials investigating noninferiority of tenecteplase versus alteplase under the assumption that bridging alteplase is superior to thrombectomy alone. Since alteplase cannot be considered a proven active control for patients presenting directly to thrombectomy-capable centers in the 2.5- to 4.5-hour window, comparing a new drug to alteplase in these patients could result in overoptimistic noninferiority claims with both drugs being not meaningfully different from each other (and, therefore, noninferior to each other) but also not meaningfully superior to thrombectomy alone. Furthermore, deriving recommendations from such trials to the general population could lead to bio-creep (approval of new treatments as a consequence of confirming noninferiority versus a standard treatment that is itself not better than placebo).[55] Commendably, stroke trials investigating tenecteplase noted this possibility and took measures to reduce this risk. The TWIST (Tenecteplase in Wake-Up Ischemic Stroke Trial)[23] and TIMELESS (Thrombolysis in Imaging Eligible, Late Window Patients to Assess the Efficacy and Safety of Tenecteplase)[24] trials opted for a superiority design, avoiding the need to rely on the constancy assumption. Among the noninferiority trials, TRACE-2 (Tenecteplase Reperfusion Therapy in Acute Ischemic Cerebrovascular Events-II)[22] and TASTE[33] excluded thrombectomy patients and noted that including such cases could confound the results while EXTEND-IA TNK[21] selected a primary outcome (reperfusion at the first angiogram) measured before endovascular treatment and, therefore, was not influenced by it. Importantly, a noninferiority trial in this specific population would still be a valid tool to advocate a switch from alteplase to tenecteplase as it ensures that tenecteplase is not more harmful than alteplase. What remains unsolved is whether any of the drugs provide a clear benefit over placebo in the specific population of patients who can directly undergo thrombectomy. This example illustrates how violations of the constancy assumption can occur in real research and highlights the potential risks.

Noninferiority trials should only compare new treatments to active controls proven superior to placebo in a similar population or clinical setting.[25–27] If the control treatment is not better than placebo, any new (ineffective) treatment can potentially be considered noninferior.

Consider Study Design and Execution: Assay Sensitivity

Assay sensitivity is the ability to detect a difference between more and less effective treatments and is an essential feature for any trial internal validity.[2] For example, in a trial investigating the effect of a treatment reducing the ischemic lesion volume after stroke, adequate measurement of the volumes should be ensured. If a software is used that produces flawed measurements, the study may fail to detect an existing treatment difference.

In a positive superiority trial, assay sensitivity is automatically ensured as the difference between treatments is the primary analysis result. In the absence of assay sensitivity, a superiority trial will be inconclusive, even if investigating truly effective treatments (type II error).[2,56] However, lack of assay sensitivity is particularly problematic for noninferiority trials. Since these trials aim to exclude differences between treatments, a lack of assay sensitivity (inability to detect differences between the treatments) makes it easier to (incorrectly) claim noninferiority.

Most of the pitfalls described in this review involve some type of assay sensitivity loss. Therefore, researchers and clinicians should carefully consider the design and execution of noninferiority trials and their ability to detect differences between the tested treatments.

Be Aware of Spinning in Result Interpretation: Noninferiority and Inferiority Are Not Opposites

An important distinction is that noninferiority and superiority (or inferiority) are assessed using 2 unrelated parameters (Figure 1). To define superiority (or inferiority), one observes if the treatment effect excludes the point of no difference. To define noninferiority, the point of no difference is irrelevant, the assessment is based exclusively on the noninferiority margin ($-\Delta$). Understanding this distinction is important to correctly interpret noninferiority trial results and understand the unusual (but valid) possibility of a new treatment being simultaneously noninferior (in relation to $-\Delta$) and inferior (in relation to the point of no difference; Figure 1). In other words, "the opposite of 'non-inferior' is not 'inferior'; it is…'not non-inferior'."[6] Clinicians should be careful when interpreting inconclusive noninferiority trials, with a recent review showing that conclusions of noninferiority trials are often inappropriately spinned. [57] Failure to prove noninferiority of an alternative treatment does not confirm superiority of the active control.

Be Aware of Differences Between Expected and Observed Benefit

In principle, a noninferiority trial design is appropriate when one suspects that a new treatment is at least not meaningfully worse but not necessarily better. In practice, however, noninferiority trials have been designed to investigate treatments that clinicians were confident to also be superior. Since noninferiority trials evaluate the new treatment against a lower threshold ($-\Delta$) than the point of no difference, noninferiority can be established with smaller sample sizes (Figure 1). It may be tempting to design a noninferiority trial when investigating a

 Stroke. 2025;56:543–552. DOI: 10.1161/STROKEAHA.124.048024

treatment believed to be truly superior. However, investigators should be aware of the limitations of doing so. By missing the opportunity to enroll an adequate number of patients, researchers may not demonstrate the superiority of a truly superior treatment. Furthermore, noninferiority alone may not warrant changes in clinical practice.

Interim analysis plans also deserve special attention in noninferiority trials. Researchers may be compelled to stop a noninferiority trial early if noninferiority has been demonstrated in interim analysis. This is a common practice in superiority trials, as it becomes unethical to continue randomizing patients to a worse control treatment. However, these ethical concerns are not present if noninferiority has been demonstrated: the control arm is receiving a treatment that is equally effective.[26,58] In fact, early termination of noninferiority trials based on the primary efficacy outcome is not recommended.[26,56,58] Reasons for continuing the trial include the subjective nature of the noninferiority margin, which may raise questions about the trial's conclusion. For example, a recent trial investigating skipping intravenous thrombolysis before endovascular treatment was stopped early after an interim analysis proved noninferiority at the prespecified margin of 10%.[20] While the results of the trial were positive, part of the stroke community questioned its noninferiority margin and thereby its conclusion. Continuing the trial further may provide more information and, in some cases, allow for superiority tests with possible practice-changing consequences.[58,59] Obviously, early stopping is warranted if interim analysis of a noninferiority trial shows meaningful superiority or inferiority of the experimental treatment. In such cases, interim analysis is crucial to prevent unethical exposure of patients to inferior therapies and reduce the running costs and duration of noninferiority trials.[58] In this sense, safety outcomes and a superiority hypothesis are recommended as stopping rules for noninferiority trials.[26] Importantly, the use of stopping rules designed for superiority studies may be inadequate in noninferiority trials.[58]

Researchers designing noninferiority trials should be careful when considering large benefits for the experimental treatment and carefully design interim analysis plans and stopping rules to match the research context.

Consider the Study Population: Impact of Crossovers and Protocol Violations

Superiority trials have strong reasons to conduct intention-to-treat (ITT) analyses. This approach more closely resembles real-life situations, such as crossovers, treatment discontinuation, loss to follow-up, and other practical sources of bias and noise. In practice, ITT analysis makes it harder to claim superiority and, therefore, acts as a safeguard against biased or poorly conducted trials falsely claiming superiority (type I error). For noninferiority trials, however, biases caused by compliance, attrition, and measurement problems will favor a noninferiority conclusion by artificially reducing the difference between the treatment arms.[51,60,61] In an extreme example, a trial that is so poorly conducted as to completely mix up both treatment arms due to protocol violations and crossovers would eventually end up with 2 arms that are uniform.[6] Simulation studies have shown that in an ITT analysis, if 10% of participants cross over to the opposite arm, the probability of claiming noninferiority increases by 8% to 10% from a nominal value of 2.5%.[49,60] However, it is not entirely clear what population analysis should be considered, and some evidence suggests that despite the limitations, ITT might be more conservative due to higher variance and consequent wider CIs.[60,62,63]

Both the US Food and Drug Administration and European Medical Agency warn against conclusions based only on ITT analysis in noninferiority trials.[5,25,59] Noninferiority trials should also report per-protocol or as-treated results, and noninferiority claims should be interpreted in both analyses. Recently proposed alternatives to correct for nonadherence may become more frequent if proven useful.[25,49,59]

Consider How the Difference in Treatment Effects Is Being Measured: Absolute Versus Relative Risk Measurements

To assess noninferiority, researchers predefine a noninferiority margin and then quantify whether the treatment effect difference between the new treatment and active control is equal to or larger than the noninferiority margin. Treatment effect can be expressed either as an absolute risk difference or as a relative treatment effect (relative risks and odds ratios).

The decision regarding the comparative framework to be used is not trivial and has important implications for interpreting the results.[25] Relative measures like odds ratios are often used as they derive naturally from regression modeling estimates. However, relative rates will depend on baseline risks in the control arm. For example, in trials investigating endovascular treatment alone versus intravenous thrombolytics followed by endovascular treatment, the rate of favorable outcome in the control arm was ≈50%. An acceptable absolute risk difference of −5% would translate into a noninferiority margin of 0.9 ([0.50−0.05]/0.50) on the relative risk scale; however, if the favorable outcome rate in the control arm was 90%, the same 0.9 noninferiority margin on the relative risk scale would represent an (likely unacceptable) absolute risk difference of −9%.

This relative risk inflation can substantially impact interpretation and is particularly problematic because noninferiority trials often rely on historical data to estimate an acceptable noninferiority margin. In ISAR-SAFE (Intracoronary Stenting and Antithrombotic Regimen: Safety And Efficacy of 6 Months Dual Antiplatelet

Therapy After Drug-Eluting Stenting),[64] a trial evaluating a shorter duration of dual antiplatelet therapy after coronary stenting, the expected rate of adverse outcomes was ≈10%, with a noninferiority margin defined as an absolute risk difference of 2%. However, the study had an observed rate of adverse effects of only 1.6%. The statistical analysis results supported the noninferiority claim with an upper 95% CI limit for the risk difference between the groups of 0.5% (within the 2% margin); however, this would represent accepting a relative 30% ([1.6+0.5]/1.6) more events in the intervention arm than in the control arm, which led the authors to bypass a claim of noninferiority, even though the noninferiority margin was statistically met.

It is worthwhile to examine the treatment effect both in absolute and relative terms when interpreting noninferiority trials. Small absolute risk differences can represent larger relative risks depending on the rate of events in the control arm. For example, a 2% absolute reduction from a 95% baseline rate of success represents a small relative change, whereas a 2% absolute reduction from a 4% baseline rate of success represents a large relative change. Interestingly, the risk inflation phenomenon behaves differently when odds ratios are used instead of relative risks. The noninferiority margin corresponding to a constant absolute risk difference in the odds ratio scale is the strictest when the rate of events in the control arm is closer to 50%. For example, when the rate of events in the control arm is 50%, the noninferiority margin corresponding to a 5% absolute risk difference is 0.82 on the odds ratio scale. However, the same 5% absolute risk difference would translate to a more liberal noninferiority margin of 0.75, if the rate of events was closer to 20% or 80% (Figure 2).

Consider Heterogeneity of Treatment Effect: The Impact of Clinically Relevant Interactions

Heterogeneity of treatment effects is an increasingly important aspect of clinical trials, given the emerging interest in personalized health care. Subgroup analyses require careful consideration, as the conventional approach to test treatment effect heterogeneity is often underpowered and can inflate type 1 error rates due to multiple testing and selection bias. These concerns can be accentuated in noninferiority trials. While in superiority trials the control arm is generally an inert placebo, noninferiority trials compare 2 active treatments, and, therefore, it is not unreasonable to expect situations in which the experimental intervention is actively harmful to one group but beneficial to another.[65]

In trials that successfully demonstrate noninferiority of the experimental intervention in the overall population, it is important to ensure that both treatments are not unacceptably different across patient subgroups and that the overall results are not driven by clinically relevant interactions with significant benefit in some patients and harm in others. For trials that failed to demonstrate noninferiority, subgroup analyses are still relevant and can provide important insights to explain why noninferiority was not established.[65]

Adhere to Existing Guidelines When Reporting and Appraising Noninferiority Studies

Finally, noninferiority trials should adhere to the existing guidelines. Despite the potential pitfalls, noninferiority trials are accepted as evidence for policy changes by the US Food and Drug Administration[25] and the European Medical Agency[27,59] when rigorously conducted, with both agencies providing recommendations on design, execution, and reporting. Other relevant guidelines include the Consolidated Standards of Reporting Trials statement and expansion to cover noninferiority trials.[26,66] Adherence to guidelines remains unsatisfactory, with issues mainly arising in justification for the noninferiority margin, reporting of both ITT and per-protocol populations, and agreement between the reported results and study conclusions.[15,16,18,31]

CONCLUSIONS

Noninferiority trials are becoming more frequent and relevant in stroke research. Interpreting and conducting these trials requires special attention and a clear understanding of the distinctions between noninferiority and superiority trials. Stroke researchers and clinicians should be aware of these distinctions and of several potential pitfalls when designing and interpreting noninferiority trials.

ARTICLE INFORMATION

Affiliations

Department of Radiology and Nuclear Medicine (F.C., K.M.T., M.K., C.B.M.) and Department of Neurology (Y.B.W.E.M.R.), Amsterdam University Medical Centers, University of Amsterdam, Amsterdam Neuroscience, the Netherlands. Department of Radiology, Haaglanden Medical Center, The Hague, the Netherlands (K.M.T.). Institute of Diagnostic and Interventional Neuroradiology (J.K., J.G.) and Department of Neurology, Stroke Research Center Bern (U.F.), University Hospital of Bern, University of Bern, Switzerland. Diagnostic and Interventional Neuroradiology, Tours, France (J.K.). Le Studium Loire Valley Institute for Advanced Studies, Tours, France (J.K.). Department of Public Health, Erasmus Medical Center, Rotterdam, the Netherlands (H.F.L.). Department of Neurology, David Geffen School of Medicine at UCLA (J.L.S.).

Acknowledgments

The authors acknowledge the CONTRAST (Collaboration for New Treatments of Acute Stroke) and IRIS (Improving Reperfusion Strategies in Ischemic Stroke) collaborations for their support.

Sources of Funding

None.

Disclosures

Dr Kaesmacher reports grants from Le Studium and grants from the Swiss National Science Foundation (SNSF) to other. Dr Saver reports stock options in Neuronics Medical; compensation from Biogen for consultant services; compensation from BrainQ for consultant services; compensation from Johnson & Johnson

 Stroke. 2025;56:543–552. DOI: 10.1161/STROKEAHA.124.048024

Health Care Systems, Inc, for consultant services; compensation from Roche for consultant services; stock options in MindRhythm; compensation from Stream Medical for consultant services; compensation from Bayer for consultant services; compensation from Medtronic USA, Inc, for consultant services; compensation from MIVI Neuroscience for data and safety monitoring services; compensation from CSL Behring for consultant services; compensation from Aeromics for consultant services; compensation from Boehringer Ingelheim (prevention only) for consultant services; compensation from Abbott Laboratories for consultant services; stock options in Rapid Medical; compensation from Occlutech for data and safety monitoring services; and compensation from BrainsGate for consultant services. Dr Gralla reports compensation from Johnson & Johnson Health Care Systems, Inc, for consultant services and compensation from Medtronic USA, Inc, for other services. Dr Fischer reports grants from Stryker to other; grants from Schweizerische Herzstiftung; grants from Penumbra, Inc, to other; grants from SNSF; compensation from phenox, Inc, for end point review committee services; grants from Medtronic to other; grants from phenox, Inc, to other; grants from Rapid Medical, Ltd, to other; grants from Boehringer Ingelheim; grants from Stryker to other; and grants from Medtronic to other. Dr Majoie reports grants from European Commission to other; grants from Health Evaluation Netherlands to other; grants from Stryker Corporation to other; stock holdings in Nicolab; grants from Dutch Heart Foundation to other; and grants from Boehringer Ingelheim to other. Dr Roos reports stock holdings in Nicolab. The other authors report no conflicts.

REFERENCES

1. Brownlee S, Chalkidou K, Doust J, Elshaug AG, Glasziou P, Heath I, Nagpal S, Saini V, Srivastava D, Chalmers K, et al. Evidence for overuse of medical services around the world. *Lancet*. 2017;390:156–168. doi: 10.1016/S0140-6736(16)32585-5
2. Dunn DT, Copas AJ, Brocklehurst P. Superiority and non-inferiority: two sides of the same coin? *Trials*. 2018;19:499. doi: 10.1186/s13063-018-2885-z
3. Temple R, Ellenberg SS. Placebo-controlled trials and active-control trials in the evaluation of new treatments. Part 1: ethical and scientific issues. *Ann Intern Med*. 2000;133:455–463. doi: 10.7326/0003-4819-133-6-200009190-00014
4. Mauri L, D'Agostino RBS. Challenges in the design and interpretation of noninferiority trials. *N Engl J Med*. 2017;377:1357–1367. doi: 10.1056/NEJMra1510063
5. Huitfeldt B, Hummel J; European Federation of Statisticians in the Pharmaceutical Industry (EFSPI). The draft FDA guideline on non-inferiority clinical trials: a critical review from European pharmaceutical industry statisticians. *Pharm Stat*. 2011;10:414–419. doi: 10.1002/pst.508
6. Schumi J, Wittes JT. Through the looking glass: understanding non-inferiority. *Trials*. 2011;12:106. doi: 10.1186/1745-6215-12-106
7. Gupta R, Saver JL, Levy E, Zaidat OO, Yavagal D, Liebeskind DS, Khaldi A, Gross B, Lang M, Narayanan S, et al; TIGER Trial Investigators. New class of radially adjustable stentrievers for acute ischemic stroke: primary results of the multicenter TIGER trial. *Stroke*. 2021;52:1534–1544. doi: 10.1161/STROKEAHA.121.034436
8. Zaidat OO, Bozorgchami H, Ribó M, Saver JL, Mattle HP, Chapot R, Narata AP, Francois O, Jadhav AP, Grossberg JA, et al. Primary results of the multicenter ARISE II study (Analysis of Revascularization in Ischemic Stroke With EMBOTRAP). *Stroke*. 2018;49:1107–1115. doi: 10.1161/STROKEAHA.117.020125
9. Lapergue B, Blanc R, Costalat V, Desal H, Saleme S, Spelle L, Marnat G, Shotar E, Eugene F, Mazighi M, et al; ASTER2 Trial Investigators. Effect of thrombectomy with combined contact aspiration and stent retriever vs stent retriever alone on revascularization in patients with acute ischemic stroke and large vessel occlusion: the ASTER2 randomized clinical trial. *JAMA*. 2021;326:1158–1169. doi: 10.1001/jama.2021.13827
10. Turk AS, Siddiqui A, Fifi JT, De Leacy RA, Fiorella DJ, Gu E, Levy EI, Snyder KV, Hanel RA, Aghaebrahim A, et al. Aspiration thrombectomy versus stent retriever thrombectomy as first-line approach for large vessel occlusion (COMPASS): a multicentre, randomised, open label, blinded outcome, non-inferiority trial. *Lancet*. 2019;393:998–1008. doi: 10.1016/S0140-6736(19)30297-1
11. Nogueira RG, Frei D, Kirmani JF, Zaidat O, Lopes D, Turk AS, Heck D, Mason B, Haussen DC, Levy EI, et al; Penumbra Separator 3D Investigators. Safety and efficacy of a 3-dimensional stent retriever with aspiration-based thrombectomy vs aspiration-based thrombectomy alone in acute ischemic stroke intervention: a randomized clinical trial. *JAMA Neurol*. 2018;75:304–311. doi: 10.1001/jamaneurol.2017.3967
12. Granger CB, Alexander JH, McMurray JJV, Lopes RD, Hylek EM, Hanna M, Al-Khalidi HR, Ansell J, Atar D, Avezum A, et al; ARISTOTLE Committees and Investigators. Apixaban versus warfarin in patients with atrial fibrillation. *N Engl J Med*. 2011;365:981–992. doi: 10.1056/NEJMoa1107039
13. Patel MR, Mahaffey KW, Garg J, Pan G, Singer DE, Hacke W, Breithardt G, Halperin JL, Hankey GJ, Piccini JP, et al; ROCKET AF Investigators. Rivaroxaban versus warfarin in nonvalvular atrial fibrillation. *N Engl J Med*. 2011;365:883–891. doi: 10.1056/NEJMoa1009638
14. LeCouffe NE, Kappelhof M, Treurniet KM, Rinkel LA, Bruggeman AE, Berkhemer OA, Wolff L, Voorst H van, Tolhuisen ML, Dippel DWJ, et al. A randomized trial of intravenous alteplase before endovascular treatment for stroke. *N Engl J Med*. 2021;385:1833–1844. doi: 10.1056/NEJMoa2107727
15. Majoie CB, Cavalcante F, Gralla J, Yang P, Kaesmacher J, Treurniet KM, Kappelhof M, Yan B, Suzuki K, Zhang Y, et al; IRIS collaborators. Value of intravenous thrombolysis in endovascular treatment for large-vessel anterior circulation stroke: individual participant data meta-analysis of six randomised trials. *Lancet*. 2023;402:965–974. doi: 10.1016/S0140-6736(23)01142-X
16. Fischer U, Kaesmacher J, Strbian D, Eker O, Cognard C, Plattner PS, Bütikofer L, Mordasini P, Deppeler S, Pereira VM, et al; SWIFT DIRECT Collaborators. Thrombectomy alone versus intravenous alteplase plus thrombectomy in patients with stroke: an open-label, blinded-outcome, randomised non-inferiority trial. *Lancet*. 2022;400:104–115. doi: 10.1016/S0140-6736(22)00537-2
17. Mitchell PJ, Yan B, Churilov L, Dowling RJ, Bush SJ, Bivard A, Huo XC, Wang G, Zhang SY, Ton MD, et al; DIRECT-SAFE Investigators. Endovascular thrombectomy versus standard bridging thrombolytic with endovascular thrombectomy within 4·5 h of stroke onset: an open-label, blinded-endpoint, randomised non-inferiority trial. *Lancet*. 2022;400:116–125. doi: 10.1016/S0140-6736(22)00564-5
18. Suzuki K, Matsumaru Y, Takeuchi M, Morimoto M, Kanazawa R, Takayama Y, Kamiya Y, Shigeta K, Okubo S, Hayakawa M, et al. Effect of mechanical thrombectomy without vs with intravenous thrombolysis on functional outcome among patients with acute ischemic stroke: the SKIP randomized clinical trial. *JAMA*. 2021;325:244–253. doi: 10.1001/jama.2020.23522
19. Yang P, Zhang Y, Zhang L, Zhang Y, Treurniet KM, Chen W, Peng Y, Han H, Wang J, Wang S, et al; DIRECT-MT Investigators. Endovascular thrombectomy with or without intravenous alteplase in acute stroke. *N Engl J Med*. 2020;382:1981–1993. doi: 10.1056/NEJMoa2001123
20. Zi W, Qiu Z, Li F, Sang H, Wu D, Luo W, Liu S, Yuan J, Song J, Shi Z, et al; DEVT Trial Investigators. Effect of endovascular treatment alone vs intravenous alteplase plus endovascular treatment on functional independence in patients with acute ischemic stroke: the DEVT randomized clinical trial. *JAMA*. 2021;325:234–243. doi: 10.1001/jama.2020.23523
21. Campbell BCV, Mitchell PJ, Churilov L, Yassi N, Kleinig TJ, Dowling RJ, Yan B, Bush SJ, Dewey HM, Thijs V, et al; EXTEND-IA TNK Investigators. Tenecteplase versus alteplase before thrombectomy for ischemic stroke. *N Engl J Med*. 2018;378:1573–1582. doi: 10.1056/NEJMoa1716405
22. Wang Y, Li S, Pan Y, Li H, Parsons MW, Campbell BCV, Schwamm LH, Fisher M, Che F, Dai H, et al; TRACE-2 Investigators. Tenecteplase versus alteplase in acute ischaemic cerebrovascular events (TRACE-2): a phase 3, multicentre, open-label, randomised controlled, non-inferiority trial. *Lancet*. 2023;401:645–654. doi: 10.1016/S0140-6736(22)02600-9
23. Roaldsen MB, Eltoft A, Wilsgaard T, Christensen H, Engelter ST, Indredavik B, Jatužis D, Karelis G, Kõrv J, Lundström E, et al; TWIST Investigators. Safety and efficacy of tenecteplase in patients with wake-up stroke assessed by non-contrast CT (TWIST): a multicentre, open-label, randomised controlled trial. *Lancet Neurol*. 2023;22:117–126. doi: 10.1016/S1474-4422(22)00484-7
24. Albers GW, Jumaa M, Purdon B, Zaidi SF, Streib C, Shuaib A, Sangha N, Kim M, Froehler MT, Schwartz NE, et al; TIMELESS Investigators. Tenecteplase for stroke at 4.5 to 24 hours with perfusion-imaging selection. *N Engl J Med*. 2024;390:701–711. doi: 10.1056/NEJMoa2310392
25. U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER). Non-inferiority clinical trials to establish effectiveness guidance for industry. 2016.
26. 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;308:2594–2604. doi: 10.1001/jama.2012.87802
27. European Medicines Agency. Committee for medicinal products for human use (CHMP) guideline on the choice of the non-inferiority margin. *Stat Med*. 2006;25:1628–1638.
28. Ring A, Lang B, Kazaroho C, Labes D, Schall R, Schütz H. Sample size determination in bioequivalence studies using statistical assurance. *Br J Clin Pharmacol*. 2019;85:2369–2377. doi: 10.1111/bcp.14055

29. Garattini S, Bertele' V. Non-inferiority trials are unethical because they disregard patients' interests. *Lancet*. 2007;370:1875–1877. doi: 10.1016/S0140-6736(07)61604-3
30. Augustsson H, Ingvarsson S, Nilsen P, von Thiele Schwarz U, Muli I, Dervish J, Hasson H. Determinants for the use and de-implementation of low-value care in health care: a scoping review. *Implement Sci Commun*. 2021;2:13. doi: 10.1186/s43058-021-00110-3
31. Nunn AJ, Meredith SK, Spigelman MK, Ginsberg AM, Gillespie SH. The ethics of non-inferiority trials. *Lancet*. 2008;371:895; author reply 896–895; author reply 897. doi: 10.1016/S0140-6736(08)60404-3
32. Leening MJG, Mahmoud KD. Non-efficacy benefits and non-inferiority margins: a scoping review of contemporary high-impact non-inferiority trials in clinical cardiology. *Eur J Epidemiol*. 2021;36:1103–1109. doi: 10.1007/s10654-021-00820-x
33. Parsons MW, Yogendrakumar V, Churilov L, Garcia-Esperon C, Campbell BCV, Russell ML, Sharma G, Chen C, Lin L, Chew BL, et al; TASTE investigators. Tenecteplase versus alteplase for thrombolysis in patients selected by use of perfusion imaging within 4·5 h of onset of ischaemic stroke (TASTE): a multicentre, randomised, controlled, phase 3 non-inferiority trial. *Lancet Neurol*. 2024;23:775–786. doi: 10.1016/S1474-4422(24)00206-0
34. Fiehler J, Zeleňák K, Zapf A. The price of certainty: when is a new therapy good enough? *J NeuroInterv Surg*. 2021;13:1065–1066. doi: 10.1136/neurintsurg-2021-018392
35. Julious SA, Owen RJ. A comparison of methods for sample size estimation for non-inferiority studies with binary outcomes. *Stat Methods Med Res*. 2011;20:595–612. doi: 10.1177/0962280210378945
36. Komorowski AS, Bai AD, Cvetkovic A, Mourad O, Lo CKL, Li XX, Mokashi V, Findlater A, Duncan DB, Fuller C, et al; McMaster Infectious Diseases Fellow Research Group. Methodological and reporting quality of non-inferiority randomized controlled trials comparing antifungal therapies: a systematic review. *Clin Microbiol Infect*. 2022;28:640–648. doi: 10.1016/j.cmi.2021.11.003
37. Schiller P, Burchardi N, Niestroj M, Kieser M. Quality of reporting of clinical non-inferiority and equivalence randomised trials--update and extension. *Trials*. 2012;13:214. doi: 10.1186/1745-6215-13-214
38. Le Henanff A, Giraudeau B, Baron G, Ravaud P. Quality of reporting of noninferiority and equivalence randomized trials. *JAMA*. 2006;295:1147–1151. doi: 10.1001/jama.295.10.1147
39. Rehal S, Morris TP, Fielding K, Carpenter JR, Phillips PPJ. Non-inferiority trials: are they inferior? A systematic review of reporting in major medical journals. *BMJ Open*. 2016;6:e012594. doi: 10.1136/bmjopen-2016-012594
40. Earl HM, Hiller L, Vallier AL, Loi S, McAdam K, Hughes-Davies L, Harnett AN, Ah-See ML, Simcock R, Rea D, et al; PERSEPHONE Steering Committee and Trial Investigators. 6 versus 12 months of adjuvant trastuzumab for HER2-positive early breast cancer (PERSEPHONE): 4-year disease-free survival results of a randomised phase 3 non-inferiority trial. *Lancet*. 2019;393:2599–2612. doi: 10.1016/S0140-6736(19)30650-6
41. Pivot X, Romieu G, Debled M, Pierga JY, Kerbrat P, Bachelot T, Lortholary A, Espié M, Fumoleau P, Serin D, et al; PHARE trial investigators. 6 months versus 12 months of adjuvant trastuzumab in early breast cancer (PHARE): final analysis of a multicentre, open-label, phase 3 randomised trial. *Lancet*. 2019;393:2591–2598. doi: 10.1016/S0140-6736(19)30653-1
42. Kaesmacher J, Mujanovic A, Treurniet K, Kappelhof M, Meinel TR, Yang P, Liu J, Zhang Y, Zi W, Yang Q, et al. Perceived acceptable uncertainty regarding comparability of endovascular treatment alone versus intravenous thrombolysis plus endovascular treatment. *J Neurointerv Surg*. 2023;15:227–232. doi: 10.1136/neurintsurg-2022-018665
43. Ye Z, Busse JW, Hill MD, Lindsay MP, Guyatt GH, Prasad K, Agarwal A, Beattie C, Beattie J, Dodd C, et al. Endovascular thrombectomy and intravenous alteplase in patients with acute ischemic stroke due to large vessel occlusion: a clinical practice guideline. *J Evid Based Med*. 2022;15:263–271. doi: 10.1111/jebm.12493
44. Saver JL, Mistry E. Time to recognize three classes of non-inferiority trial margins. *J Neurointerv Surg*. 2023;15:2–4. doi: 10.1136/neurintsurg-2022-018879
45. Pong S, Urner M, Fowler RA, Mitsakakis N, Seto W, Hutchison JS, Science M, Daneman N. Testing for non-inferior mortality: a systematic review of non-inferiority margin sizes and trial characteristics. *BMJ Open*. 2021;11:e044480. doi: 10.1136/bmjopen-2020-044480
46. Cranston JS, Kaplan BD, Saver JL. Minimal clinically important difference for safe and simple novel acute ischemic stroke therapies. *Stroke*. 2017;48:2946–2951. doi: 10.1161/STROKEAHA.117.017496
47. Lin CJ, Saver JL. The minimal clinically important difference for achievement of substantial reperfusion with endovascular thrombectomy devices in acute ischemic stroke treatment. *Front Neurol*. 2020;11:524220. doi: 10.3389/fneur.2020.524220
48. Lin CJ, Saver JL. Noninferiority margins in trials of thrombectomy devices for acute ischemic stroke: is the bar being set too low? *Stroke*. 2019;50:3519–3526. doi: 10.1161/STROKEAHA.119.026717
49. Mo Y, Lim C, Mukaka M, Cooper BS. Statistical considerations in the design and analysis of non-inferiority trials with binary endpoints in the presence of non-adherence: a simulation study. *Wellcome Open Res*. 2020;4:207. doi: 10.12688/wellcomeopenres.15636.2
50. Rinkel LA, Ospel JM, Brown SB, Campbell BCV, Dippel DWJ, Demchuk AM, Majoie CBLM, Mitchell PJ, Bracard S, Guillemin F, et al; HERMES Collaborators. What is a meaningful difference when using infarct volume as the primary outcome?: results from the HERMES database. *Stroke*. 2024;55:866–873. doi: 10.1161/STROKEAHA.123.044353
51. Liu Q, Li Y, Odem-Davis K. On robustness of noninferiority clinical trial designs against bias, variability, and nonconstancy. *J Biopharm Stat*. 2015;25:206–225. doi: 10.1080/10543406.2014.923738
52. Diamond GA, Kaul S. An Orwellian discourse on the meaning and measurement of noninferiority. *Am J Cardiol*. 2007;99:284–287. doi: 10.1016/j.amjcard.2006.07.090
53. Albers GW, Diener HC, Frison L, Grind M, Horrow J, Nevinson M, Olsson SB, Partridge S, Petersen P, Vahanian A, et al. Trials and tribulations of noninferiority: the ximelagatran experience. *J Am Coll Cardiol*. 2006;48:1058. doi: 10.1016/j.jacc.2006.06.004
54. Kaesmacher J, Cavalcante F, Kappelhof M, Treurniet KM, Rinkel L, Liu J, Yan B, Zi W, Kimura K, Eker OF, et al; IRIS Collaborators. Time to treatment with intravenous thrombolysis before thrombectomy and functional outcomes in acute ischemic stroke: a meta-analysis. *JAMA*. 2024;331:764–777. doi: 10.1001/jama.2024.0589
55. Everson-Stewart S, Emerson SS. Bio-creep in non-inferiority clinical trials. *Stat Med*. 2010;29:2769–2780. doi: 10.1002/sim.4053
56. D'Agostino RB, Massaro JM, Sullivan LM. Non-inferiority trials: design concepts and issues - the encounters of academic consultants in statistics. *Statist Med*. 2002;22:169–186. doi: 10.1002/sim.1425
57. Ito C, Hashimoto A, Uemura K, Oba K. Misleading reporting (spin) in non-inferiority randomized clinical trials in oncology with statistically not significant results: a systematic review. *JAMA Netw Open*. 2021;4:e2135765. doi: 10.1001/jamanetworkopen.2021.35765
58. Korn EL, Freidlin B. Interim monitoring for non-inferiority trials: minimizing patient exposure to inferior therapies. *Ann Oncol*. 2018;29:573–577. doi: 10.1093/annonc/mdx788
59. Committee for Proprietary Medicinal Products (CPMP). Points to consider on switching between superiority and non-inferiority: points to consider on switching between superiority and non-inferiority. *Br J Clin Pharmacol*. 2001;52:223–228. doi: 10.1046/j.0306-5251.2001.01397-3.x
60. Mo Y, Lim C, Watson JA, White NJ, Cooper BS. Non-adherence in non-inferiority trials: pitfalls and recommendations. *BMJ*. 2020;370:m2215. doi: 10.1136/bmj.m2215
61. Matilde Sanchez M, Chen X. Choosing the analysis population in non-inferiority studies: per protocol or intent-to-treat. *Stat Med*. 2006;25:1169–1181. doi: 10.1002/sim.2244
62. Bai AD, Komorowski AS, Lo CKL, Tandon P, Li XX, Mokashi V, Cvetkovic A, Findlater A, Liang L, Tomlinson G, et al; McMaster Infectious Diseases Fellow Research Group. Intention-to-treat analysis may be more conservative than per protocol analysis in antibiotic non-inferiority trials: a systematic review. *BMC Med Res Methodol*. 2021;21:75. doi: 10.1186/s12874-021-01260-7
63. Dodd M, Fielding K, Carpenter JR, Thompson JA, Elbourne D. Statistical methods for non-adherence in non-inferiority trials: useful and used? A systematic review. *BMJ Open*. 2022;12:e052656. doi: 10.1136/bmjopen-2021-052656
64. Schulz-Schupke S, Byrne RA, Ten Berg JM, Neumann F-J, Han Y, Adriaenssens T, Tolg R, Seyfarth M, Maeng M, Zrenner B, et al; Intracoronary Stenting and Antithrombotic Regimen: Safety And Efficacy of 6 Months Dual Antiplatelet Therapy After Drug-Eluting Stenting (ISAR-SAFE) Trial Investigators. ISAR-SAFE: a randomized, double-blind, placebo-controlled trial of 6 vs. 12 months of clopidogrel therapy after drug-eluting stenting. *Eur Heart J*. 2015;36:1252–1263. doi: 10.1093/eurheartj/ehu523
65. Fazzari MJ, Kim MY. Subgroup discovery in non-inferiority trials. *Stat Med*. 2021;40:5174–5187. doi: 10.1002/sim.9118
66. Schulz KF, Altman DG, Moher D; CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *BMC Med*. 2010;8:18. doi: 10.1186/1741-7015-8-18