

Non-Inferiority Trials in Stroke Research: What Are They, and How Should We Interpret Them?

Linxin Li,¹ Vasileios-Arsenios Lioutas,² Ralph K. Akyea,³ Stefan Gerner,⁴ Kui Kai Lau,⁵ Emily Ramage,⁶ Aristeidis H. Katsanos,⁷ George Howard,⁸ Philip M. Bath⁹

¹Wolfson Centre for Prevention of Stroke and Dementia, Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, UK

²Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA, USA

³Centre for Academic Primary Care, Lifespan & Population Health Unit, School of Medicine, University of Nottingham, Nottingham, UK

⁴Department of Neurology, University Hospital Erlangen, Germany

⁵University of Hong Kong, Hong Kong, China

⁶Florey Institute of Neuroscience and Mental Health, Victoria, Australia

⁷McMaster University/Population Health Research Institute, Hamilton, ON, Canada

⁸University of Alabama at Birmingham, Birmingham, AL, USA

⁹Stroke Trials Unit, Mental Health & Clinical Neuroscience, University of Nottingham, Nottingham, UK

Randomized clinical trials are important in both clinical and academic stroke communities with increasing numbers of new design concepts emerging. One of the "less traditional" designs that have gained increasing interest in the last decade is non-inferiority trials. Whilst the concept might appear straightforward, the design and interpretation of non-inferiority trials can be challenging. In this review, we will use exemplars from clinical trials in the stroke field to provide an overview of the advantages and limitations of non-inferiority trials and how they should be interpreted in stroke research.

Keywords Non-inferiority; Stroke; Trial design; Clinical trial; Methodology

Correspondence: Linxin Li
Wolfson Centre for Prevention of Stroke and Dementia, Nuffield Department of Clinical Neurosciences, University of Oxford, Wolfson Building, John Radcliffe Hospital, Oxford. OX3 9DU, UK
Tel: +44-(0)1865-61127
E-mail: Linxin.li@ndcn.ox.ac.uk
<https://orcid.org/0000-0002-3636-8355>

Received: September 24, 2024

Revised: November 8, 2024

Accepted: December 27, 2024

Introduction

Randomized clinical trials (RCT) play a prominent role in both clinical and academic stroke communities. They have gained increasing interest in the last few decades with new design concepts emerging. One of the "less traditional" designs is non-inferiority trials, which have increased from only one or two per year in the stroke field to more than 20 since 2019.

Non-inferiority trials are clinical studies that aim to show that a new treatment is not worse than an established treatment by a difference which is known as the non-inferiority margin. The concept of non-inferiority trial emerged in the late 20th century and started to gain increasing recognition in the 1990s.¹ They have been conducted in various fields, including infectious disease, oncology, and cardiovascular disease.²⁻⁴ Whilst the concept

might appear straightforward, the design and interpretation of non-inferiority trials can be challenging. In this review, we will use exemplars from specific clinical trials to provide an overview of the advantages and limitations of non-inferiority trials and how they should be interpreted in stroke research.

Superiority, equivalence, and non-inferiority RCTs

In general, RCTs are divided into three types and they each are testing distinct hypothesis which are important to recognize when designing or interpreting the trial results.

Superiority

Superiority trials are most commonly used in stroke research.

They aim to demonstrate that a new treatment or intervention is superior to an existing standard treatment or placebo.⁵ In statistical terms, "superior" means that the results will reject the null hypothesis that the new treatment is not superior to the control. Investigators typically choose the expected difference between the comparison groups (Δ), an accepted type 1 error rate (α), and the power to decide on the sample size. The observed difference might be bigger or smaller than Δ , but as long as the lower 95% confidence interval (CI) is above 0 (for absolute difference) or 1 (for relative risk or odds ratio), they might reject the null hypothesis.

Equivalence

Equivalence trials aim to demonstrate that two treatments or interventions have similar effectiveness within a pre-defined margin of difference.⁵ The goal is to show that the treatments are essentially equivalent in terms of clinical outcomes. Because it is impossible to show exact equivalence unless an infinite sample size is possible, investigators usually choose an equivalence margin (δ). Two-sided 90% CI are typically used in these trials and if the CI was strictly within $[-\delta, +\delta]$, the comparison groups are "equivalent." Unlike the Δ in superiority trials, δ is not only important for the sample size but also for the analysis (Figure 1). Given that the equivalence margin can be challenging to set, and the sample size is usually much larger than superiority or non-inferiority trials, especially if the equivalence margin is small, they are rarely performed.⁶

Non-inferiority

Non-inferiority trials aim to demonstrate that a new treatment

or intervention is not "inferior" to the comparison group. In other words, the goal is to show that the new treatment is not unacceptably worse than an established treatment by a predefined margin (δ).⁵ In this design, the null hypothesis is less intuitive as it states that the new treatment of intervention is worse than the comparator by more than the non-inferiority margin δ . The alternative hypothesis states that the difference between the two groups is less than δ . A 95% CI or 97.5% CI is usually used in these trials and if the lower boundary of the CI is above δ (or upper boundary if relative risk difference was measured), non-inferiority can be claimed. Similar to equivalence trials, δ is important for both sample size and analysis (Figure 1). The difference between the two is that in non-inferiority trials, only one side of the CIs matters.

For the different clinical trial concepts, an explanation of the associated hypothesis and types of errors in statistics are summarized in Table 1.

Why do we need non-inferiority trials?

Non-inferiority trials are designed for studies in which it is ethically inappropriate to include a control arm. In clinical practice, there are new treatments or interventions that might not be expected to be more effective than an existing approach but may have some other advantages, such as greater availability, reduced cost, better safety profile, or easier administration. In these scenarios, the new treatments or interventions might have less efficacy compared to standard approaches but that potential lost efficacy is acceptable given the advantages, hence they are called "non-inferior." It is important to note that "neutral" (null)

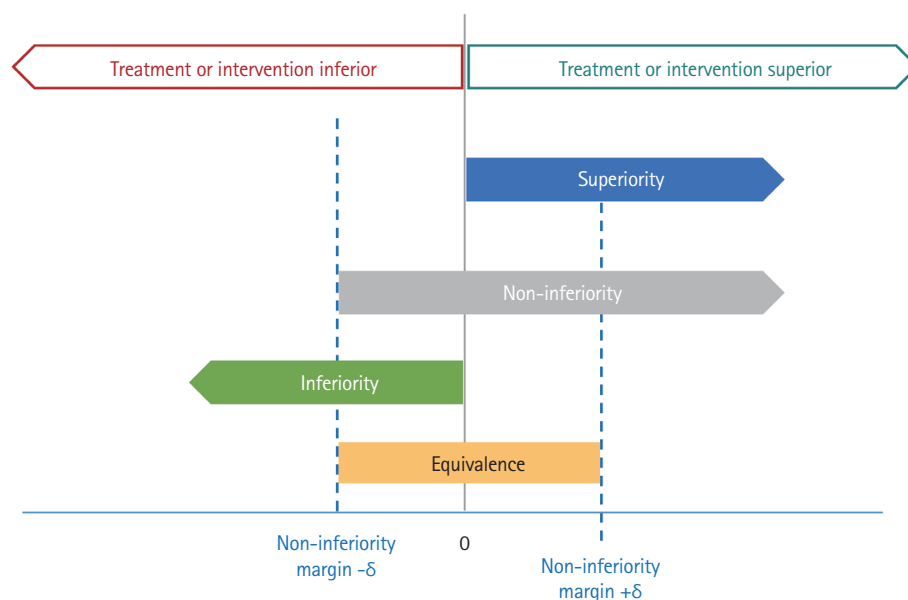


Figure 1. Treatment/intervention effects and 95% confidence interval in defining the different clinical trial concepts.

Table 1. The hypothesis and the types of errors in statistics for the different clinical trial concepts

Trial concept	Null hypothesis	Alternative hypothesis	Type-1 error	Type-II error
Superiority	The new treatment is not better than the standard treatment or placebo	The new treatment is better than the standard treatment or placebo	Concluding superiority of new intervention when there is no superiority (that is, a false positive)	Not concluding superiority of new intervention when there is a superiority (that is, a false negative)
Equivalence	The new intervention or treatment is either superior or inferior to the standard intervention or treatment	The new treatment or intervention is equivalent to the standard treatment or intervention	Concluding the new intervention or treatment is equivalent to the standard intervention when it is not equivalent (false positive)	Concluding the new treatment or intervention is not equivalent to the standard intervention when it actually is equivalent (false negative)
Non-inferiority	The new treatment or intervention is worse than the standard intervention by more than a pre-specified margin, known as the non-inferiority margin (δ)	The new treatment or intervention is not worse (i.e., it is either superior or equivalent) to the standard intervention within the non-inferiority margin (δ)	Concluding the new treatment or intervention is not inferior when the new intervention is in fact not non-inferior (false positive)	Concluding inferiority for an intervention or treatment which is actually non-inferior (false negative)

superiority trials are not necessarily showing "non-inferiority." In this context, the conclusion refers to a failure to reject the null hypothesis of "no difference" but they might lack statistical power to rule out important differences. Therefore, non-inferiority trials are crucial for determining whether a new treatment/intervention is not worse than a reference approach by more than an acceptable amount (i.e., margin δ). It is also worth noting that non-inferiority trials are also important in drug development because they allow the introduction of new treatments for conditions with effective but suboptimal existing therapies in terms of administration, cost, or side effects.

How are non-inferiority trials designed?

What is the non-inferiority margin δ ?

Non-inferiority trials necessitate evidence from previous randomized trials that establish the superiority of the active control, which becomes the standard of care, over no intervention. This basic requirement ensures that the active control is established, thus serving as a solid benchmark against which new treatments can be compared by non-inferiority analysis.

In the second step, an adequate difference needs to be defined under which a new treatment is stated as non-inferior; this difference, as already mentioned in the previous section, is called the non-inferiority margin δ and can be chosen as an absolute risk difference or relative risk difference (risk ratio). It represents the minimal clinically and/or statistically acceptable difference in efficacy between the new treatment and the active control where the new treatment would be considered as "not unacceptably worse" or "non-inferior." The establishment of this margin is crucial as it quantifies the extent to which the new treatment may deviate from the established treatment without being considered inferior. Table 2 summarizes the margins used in selected key stroke trials.

The selection of δ should be based upon a combination of statistical reasoning and clinical judgment as suggested by international guidelines.^{7,8} Yet these considerations are only poorly reflected by stated methods of most non-inferiority trials.^{4,9,10}

Steps in the selection of the non-inferiority margin δ

The selection of δ fundamentally relies on historical data from prior randomized trials that evaluate the superiority of the active control over a placebo. This historical evidence serves as a benchmark, informing the extent to which the new treatment can be deemed non-inferior while still retaining a significant portion of the active control's therapeutic benefit. For instance, in stroke trials where anticoagulants are compared, the selection of δ might be influenced by previous studies demonstrating the efficacy of a standard anticoagulant over placebo in preventing stroke recurrence. The chosen δ must ensure that the new anticoagulant preserves a substantial part of this demonstrated efficacy, a principle that underscores the clinical relevance of the margin.

Moreover, the clinical judgment of experts in stroke management plays a pivotal role in defining δ . These professionals assess the clinical implications of different levels of treatment efficacy, considering patient outcomes, side effect profiles, and the practical aspects of treatment administration. Their insights ensure that δ is set at a level where any potential reduction in efficacy from the new treatment is counterbalanced by other clinical benefits, such as reduced side effects or improved patient adherence.

Statistical considerations also guide the determination of δ , with the aim of ensuring that the non-inferiority trial is adequately powered to detect a meaningful difference between the new treatment and the active control. This involves sophisticated statistical models that account for the variability in outcome

measures and the expected performance of the new treatment. These models help in quantifying the uncertainty around δ and ensuring that the trial's conclusions are robust.

Regulatory guidelines from bodies such as the U.S. Food and Drug Administration and European Medicines Agency provide frameworks within which δ is selected, often advocating for a

Table 2. Margins used in selected stroke trials of non-inferiority design

Trial name	Primary outcome	Margin type	Margin
Acute and hyperacute setting			
Bridging vs. direct endovascular treatment			
DIRECT-MT	mRS 0–2	ARD	4%
DEVT	mRS 0–2	ARD	10%
DIRECT-SAFE	mRS 0–2 or return to baseline	ARD	10%
SWIFT DIRECT	mRS 0–2	ARD	12%
SKIP	mRS 0–2	OR	0.74
MR-CLEAN-NO IV	mRS shift	OR	0.80
DIRECT-MT	mRS shift	OR	0.80
Tenecteplase vs. alteplase			
NOR-TEST 2, part A	mRS 0–1	ARD	3%
TRACE-2	mRS 0–1	ARD	4%
AcT	mRS 0–2	ARD	5%
Other acute trials			
ENCHANTED	mRS 2–6	OR	1.14
EXTEND-IA	Substantial reperfusion	ARD	2%
ARAMIS	mRS 0–1	ARD	5%
rhPro-UK	mRS 0–1	ARD	10%
CAIST	mRS 0–2	ARD	10%
EDO	mRS 0–1	ARD	11%
FRIDA	mRS 0–1	ARD	16%
Secondary prevention setting			
Direct oral anticoagulant vs. warfarin			
RELY	Stroke or systemic embolism	RR	1.46
ROCKET-AF	Stroke or systemic embolism	RR	1.46
ARISTOTLE	Stroke or systemic embolism	RR	1.44
ENGAGE AF-TIMI 48	Stroke or systemic embolism	HR	1.38
Other secondary prevention trials			
PERFORM	Ischaemic stroke, MI, or vascular death	HR	1.05
PICASSO	Stroke, MI, or vascular death	HR	1.25
INSURE	Stroke	HR	1.25
CSPS2	Stroke	HR	1.33
S-ACCESS	Ischaemic stroke	HR	1.33
JASAP	Ischaemic stroke	HR	1.37
PRASTRO-I	Ischaemic stroke, MI, or vascular death	RR	1.35
PRoFESS	Stroke	OR	1.08
JASAP	Ischaemic stroke	ARD	2.00%
EVA-3S	Stroke or death	ARD	2.00%
SPACE	Ipsilateral stroke or death	ARD	2.50%
INSURE	Stroke	ARD	2.25%
TIMING	Ischaemic stroke, symptomatic ICH, or death	ARD	3.00%

mRS, modified Rankin Scale; MI, myocardial infarction; ICH, intracerebral haemorrhage; ARD, absolute risk difference; OR, odds ratio; RR, risk ratio; HR, hazard ratio.

conservative approach.^{7,8} These guidelines emphasize the importance of protecting patient safety by ensuring that new treatments do not significantly compromise efficacy. They may offer specific recommendations on the fraction of the historical effect size that should be preserved by the new treatment, serving as a crucial reference point in the margin-setting process.

An illustrative example of these principles in action can be seen in the non-inferiority trials comparing direct oral anticoagulants (DOACs) to warfarin for stroke prevention in patients with atrial fibrillation.^{11–14} Historical trials establishing warfarin's superiority over placebo provided a quantitative foundation for setting δ . Expert consensus, informed by clinical experience and patient outcome priorities, played a role in determining the acceptable trade-offs between efficacy, safety, and convenience. Statistical analyses ensured that the chosen δ was justifiable based on the historical effect sizes and the expected variability in treatment outcomes. Regulatory guidelines shaped the conservativeness of δ , emphasizing patient safety and the preservation of a substantial fraction of warfarin's efficacy. The therapeutic context, including the known limitations of warfarin such as its dietary restrictions and need for regular monitoring, justified the exploration of DOACs as alternatives, provided they could meet the efficacy threshold defined by δ .

Other factors associated with the choice of the non-inferiority margin δ

The therapeutic context, including the availability of alternative treatments, the severity of the condition treated, and the potential benefits of the new treatment, also influences the choice of δ . In conditions where treatment options are limited or the disease burden is high, a slightly larger δ might be justified if the new treatment offers significant non-efficacy-related advantages. Conversely, in scenarios where effective treatments are already available, a smaller δ would be necessary to justify the adoption of a new treatment.

The nature of the outcome events is sometimes important to consider when deciding on the δ , especially when a composite outcome is considered. Including a "softer" outcome which might be driving the composite outcome (for example imaging findings of new lesion) might make one feel more willing to accept a larger δ .

The duration of follow-up can be important too. In general, the shorter the follow-up, the more conservative the δ should be. Moreover, it is also important to recognize that whilst the relative risk difference can remain constant over time, the absolute risk difference can differ significantly when comparing a 30-day outcome versus a 5-year outcome. This will also have implications in deciding whether an absolute risk difference or a

relative difference should be chosen for the δ .

Examples of why non-inferiority trials are done and how the margins are selected

Here we illustrate a few examples of why a non-inferiority design was chosen in the field of clinical stroke research and also how they determined the chosen margin.

Convenience and better safety profile

In patients with atrial fibrillation, the standard treatment before the four DOAC trials (RE-LY [Randomized Evaluation of Long-Term Anticoagulation Therapy]; ROCKET AF [Rivaroxaban Once Daily Oral Direct Factor Xa Inhibition Compared with Vitamin K Antagonism for Prevention of Stroke and Embolism Trial in Atrial Fibrillation]; ARISTOTLE [Apixaban for Reduction in Stroke and Other Thromboembolic Events in Atrial Fibrillation]; ENGAGEAF-TIMI [Effective Anticoagulation with Factor Xa Next Generation in Atrial Fibrillation–Thrombolysis in Myocardial Infarction]^{11–14}) was warfarin, which is a vitamin K antagonist. Whilst warfarin is highly effective, it requires frequent laboratory monitoring, has multiple interactions with food and drugs and is associated with an increased risk of major bleeding. Therefore, the investigators were looking for new anticoagulant agents that might be comparable or even slightly worse compared to warfarin in terms of efficacy but would potentially be safer and more convenient to use and hence well-suited for a non-inferiority design.

All of the four trials set their non-inferiority margin δ based on the efficacy of vitamin K antagonists as compared with control therapy which was derived from a meta-analysis of relevant trials.¹⁵ In this meta-analysis, warfarin was associated with a 62% relative reduction in the risk of stroke or systemic embolism (95% CI 48–72). RE-LY and ROCKET AF used a margin of 1.46, which was half of the lower boundary of the 95% CI (i.e., reduction of 0.48) comparing control and warfarin (control risk 1, warfarin risk 1–0.48, relative risk = $1/(1-0.48)=1.92$, and half the risk increase would be $1+0.92/2=1.46$).^{11,13} ARISTOTLE and ENGAGE AF-TIMI 48 used a margin of 1.44 (or 1.38 on a log scale), which required that the new drugs preserve at least 50% of the relative reduction in the risk of stroke or systemic embolism associated with warfarin (i.e., relative risk reduction should be $62/2=31\%$, so the risk would be $[1-0.31]=0.69$ for the new drug, hence the relative risk would be $1/0.69=1.44$).^{12,14}

Convenience

Intravenous thrombolysis with alteplase is standard medical therapy for patients with acute ischemic stroke. Tenecteplase is

a genetically modified variant of alteplase and has a well-characterized mechanism of action. Moreover, it has a longer plasma half-life and is administered as a bolus rather than as an infusion. This ease of administration gives tenecteplase a unique practical advantage, making it an attractive replacement for alteplase. Consequently, several trials comparing alteplase with tenecteplase have applied the non-inferiority design.

For example, the AcT (Alteplase Compared to Tenecteplase in Patients With Acute Ischemic Stroke) trial to determine whether intravenous tenecteplase, at a dose of 0.25 mg/kg, is non-inferior to alteplase in all patients with acute ischemic stroke who meet criteria for intravenous thrombolysis.¹⁶ The investigators chose 5% as the non-inferiority margin. This margin was based on a meta-analysis comparing alteplase with placebo, which showed that treatment with alteplase resulted in better functional outcomes compared to placebo (absolute difference 9.8%, 95% CI 5.4–14.3).¹⁷ The chosen 5% assumed that at least half of the point estimate of effect for intravenous alteplase versus control will be preserved. This non-inferiority margin is also less than the lower 95% CI boundary on the point estimate of alteplase versus placebo.

Logistical benefits and resources use/cost

Current guidelines recommend intravenous thrombolysis before endovascular treatment for all eligible patients with anterior circulation large artery occlusion.^{18–20} However, the value of intravenous thrombolysis in patients presenting directly to endovascular treatment capable centers has been questioned, especially due to the low chance of recanalization and increased chance of bleeding. Omitting thrombolysis in this specific population could reduce healthcare costs and workflow delays.

The IRIS (Improving Reperfusion strategies in Ischemic Stroke) collaboration, which included six trials (DEVT [Direct Endovascular Thrombectomy vs Combined IVT and Endovascular Thrombectomy for Patients With Acute Large Vessel Occlusion in the Anterior Circulation]; DIRECT-MT [Direct Intraarterial Thrombectomy in Order to Revascularize Acute Ischemic Stroke Patients With Large Vessel Occlusion Efficiently in Chinese Tertiary Hospitals]; DIRECT-SAFE [DIRECT Endovascular Clot Retrieval versus Standard Bridging Therapy]; MR CLEAN-NO IV [Multi-center Randomized Clinical Trial of Endovascular Treatment for Acute Ischemic Stroke in the Netherlands, No IV]; SKIP [The Randomized Study of EVT With Versus Without Intravenous Recombinant Tissue-Type Plasminogen Activator in Acute Stroke With ICA and M1 Occlusion]; SWIFT-DIRECT [Solitaire™ With the Intention For Thrombectomy Plus Intravenous t-PA Versus DIRECT Solitaire™ Stent-retriever Thrombectomy in Acute Anterior Circulation Stroke]) comparing bridging versus direct endovascular

lar treatment²¹ chose 5% as their margin based on an international survey and the European Stroke Organization–European Society for Minimally Invasive Neurological Therapy guidelines.^{20,22} They also performed a meta-analysis of all existing trials and based on the observed pooled rate of functional independence (i.e., modified Rankin scale 0–2) in the control group (49%), a 5% higher rate in favor of the combined treatment (54%) would result in an odds ratio of 0.82 (i.e., $[49 \times 46]/[54 \times 51]$), which was chosen as their non-inferiority margin.²¹

Used in combination with superiority trials

Occasionally, a non-inferiority design was used in a hierarchical way together with a superiority design. For example, in the antiplatelet arm of the PROfESS (Prevention Regimen for Effectively Avoiding Second Strokes) trial, the investigators aimed to compare the fixed combination of low-dose aspirin (A) and extended-release dipyridamole (A+D) with clopidogrel (C).²³ The investigators planned to test, first, the non-inferiority of A+D compared to C and, if this was satisfied by demonstrating that the hazard ratio (HR) for A+D compared to C was less than a predefined non-inferiority margin δ , to then test whether A+D was superior to C. The δ was set at 1.075 and was based on estimates from previous trials and meta-analysis.^{24,25} Using data from the CAPRIE (Clopidogrel versus Aspirin in Patients at Risk of Ischaemic Events) trial²³ and the meta-analysis from the Antithrombotic Trialists' Collaboration,²⁵ the odds ratio for clopidogrel being better than placebo for the outcome of nonfatal stroke was translated into an excess risk for placebo versus clopidogrel: 1.377 with a 95% CI of 1.155 to 1.645. The selected non-inferiority δ of HR 1.075 assumes that A+D retains more than half of the clopidogrel effect over placebo if the lower boundary was used (i.e., 0.155/2). Given the importance of controlling for type I error in this kind of design, trials are also recommended to report how they controlled type I error (e.g., by the Benjamini and Hochberg procedure).²⁶

How to interpret non-inferiority trial results

Correct interpretation of a non-inferiority trial results requires some level of familiarity with the basics of hypothesis testing as outlined in detail in the segment "superiority, equivalence, non-inferiority and inferiority." A main challenge for many readers of scientific literature is that we are primed to think in "traditional" superiority analysis terms using conventional null hypothesis language ("there is no difference..."). It should be kept in mind that in strict statistical terms we either reject the null hypothesis which means that there is a difference or fail to reject the null

which is often perceived as the same as “accepting” the null, but it is not; we cannot really say that the two groups are equal. By contrast, non-inferiority design aims to show that the new intervention is not significantly worse than the established one and consequently frames the question (null hypothesis) differently: ‘the new treatment is worse plus an additional “cushion” (δ)’, which provides the non-inferiority margin. It also provides a specific direction of the assumed difference (worse) and is expressed as H_0 : intervention + δ < control. Thus, the alternative is H_A : intervention + $\delta \geq$ control. At a basic level, the interpretation of a non-inferiority is straightforward and necessitates a clear definition of the primary endpoint and the non-inferiority margin. One simply needs to estimate the difference between the two treatments and calculate the CI around it. The focus of a non-inferiority trial is not to characterize the point estimate of the treatment effect but rather the lower bound of its CI. If this lies above the predefined non-inferiority margin, the intervention is deemed “non-inferior.” It should be emphasized that failing to establish non-inferiority at this stage does not equate to inferiority: rather, all we can say is that the intervention is “not non-inferior.” Non-inferiority can be followed by “traditional” testing for superiority (or inferiority, depending on the direction of the effect) without incurring a statistical penalty for multiple testing. In this instance, the interest shifts from the non-inferiority margin to the null value: if the lower margin of the CI lies above

the null value, then the intervention is considered both non-inferior and superior. It bears noting that the null value when examining absolute differences is 0, whereas for relative differences (such as odds, risk ratios, or HR) it is 1. It should also be noted that given the premise of non-inferiority trials it is possible for a trial to technically show non-inferiority for the efficacy endpoint but deemed inferior on the basis of significantly higher rate of safety events. Overall, there are five basic possible outcomes of a non-inferiority trial, illustrated in Figure 2.

We present representative non-inferiority stroke trials highlighting examples of all the above scenarios (Figure 2).

Non-inferior and superior

The ARISTOTLE trial is a characteristic example of non-inferiority and superiority.¹² ARISTOTLE was a double-blind, randomized controlled non-inferiority trial comparing compared apixaban with warfarin for the prevention of stroke or systemic embolism in patients with atrial fibrillation. The non-inferiority margin was determined as the upper boundary of the 95% CI for the relative risk would be less than 1.38. For the primary endpoint of the trial, the HR was 0.79 with a 95% CI of 0.66 to 0.95. At first sight, the upper boundary of the CI is well below the predetermined 1.38 thus establishing non-inferiority. Moreover, the upper boundary of the CI is also less than 1 therefore apixaban also met the criteria of superiority. Inspection of the safety endpoints shows a

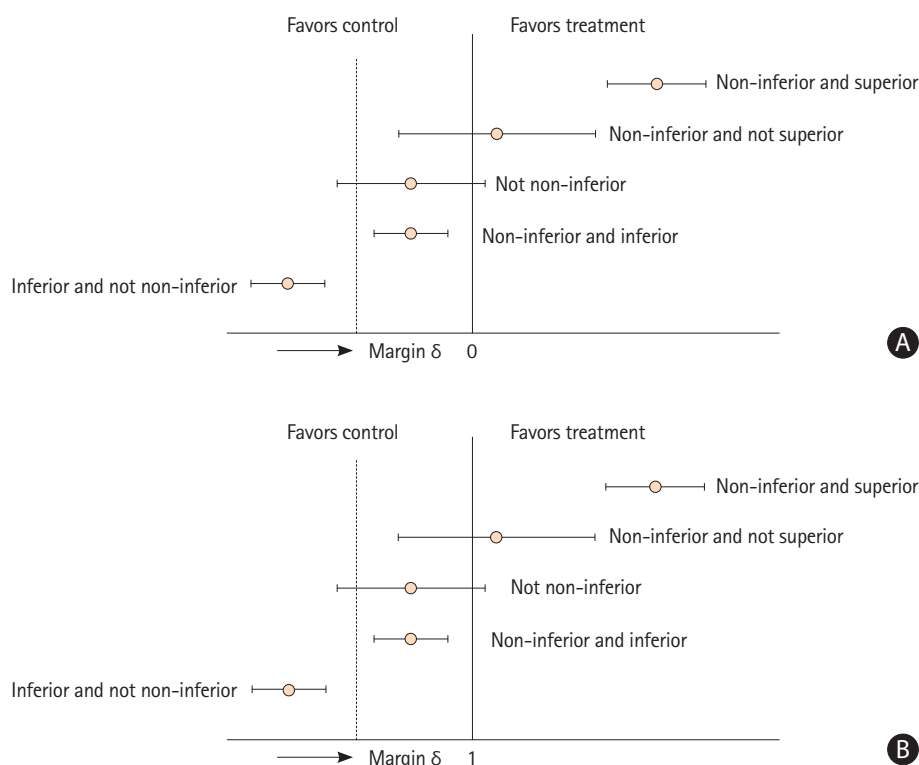


Figure 2. Possible scenarios to interpret the findings of a non-inferiority trial. (A) Absolute difference. (B) Relative risk.

favorable safety profile for apixaban, upholding the non-inferiority conclusion.

Non-inferior and not superior

This is a quintessential non-inferiority scenario, i.e., a study that confirms that the new intervention is not unacceptably worse, without meeting criteria for superiority and at the same time without any safety concerns. The TRACE-2 (Tenecteplase versus Alteplase in Acute Ischemic Cerebrovascular Events) trial was a phase 3, multicenter, prospective, open-label, blinded-endpoint, randomized controlled, non-inferiority trial that compared intravenous tenecteplase (new intervention) with tissue plasminogen activator.²⁷ Tenecteplase would be declared non-inferior if the lower 97.5% one-sided CI of the risk ratio for the primary outcome did not cross 0.937. The final results yielded a risk ratio of 1.07 with a 95% CI of 0.98 to 1.16. Given that the lower boundary (i.e., 0.98) did not cross 0.937, tenecteplase was found to be non-inferior. Superiority could not be established given that the 95% CI crossed the null value of a risk ratio of 1.00.

Non-inferiority not established, also known as inconclusive

In the SoSTART (Start or Stop Anticoagulants Randomized Trial), the primary aim was to establish the non-inferiority of starting oral anticoagulation versus avoiding in intracerebral hemorrhage patients who have atrial fibrillation.²⁸ Non-inferiority would be confirmed if the upper limit of the 95% CI of the adjusted HR for the effect of starting oral anticoagulation on recurrent symptomatic spontaneous intracranial hemorrhage is less than 3.2. The adjusted HR for recurrent hemorrhage in those starting oral anticoagulation was estimated at 2.42 with a 95% CI of 0.72 to 8.09. Given that 8.09 exceeds the predefined margin of 3.2, non-inferiority could not be established. Common reasons to explain this scenario included underpowered study, higher than expected variability, and a too narrowed margin. In SoSTART, the event rate observed were lower than assumed (4% in the avoid arm vs. 4.2%–8.6% based on previous data) so the estimate of effect became less precise than expected. It has been suggested that using an adaptive design in future trials might allow for adjustment based on interim results hence ensuring more adequate power in these scenarios.

Non-inferior and inferior

This situation occurs primarily when relatively generous margins were set and therefore in general should be interpreted as “somewhat inferior.” Sometimes, the inferiority is interpreted due to significant harm alongside with the primary efficacy outcome. For example, the trial conducted by Amadeus Investigators is a

multicenter, randomized, open-label non-inferiority trial with blinded assessment of outcome.²⁹ The investigators compared fixed-dose idraparinux (factor X inhibitor) with dose-adjusted oral vitamin K antagonist therapy for prevention of thromboembolism in patients with atrial fibrillation. The primary efficacy outcome was the composite of all stroke or systemic embolism. The non-inferiority boundary was set at 1.5 for the HR comparing the two groups. The trial was stopped early due to an excess of clinically relevant bleeding with idraparinux. For the primary outcome, the HR was 0.71 with a 95% CI of 0.39 to 1.30. Given that the upper boundary (i.e., 1.30) did not exceed the predefined margin (i.e., 1.5), the trial achieved non-inferiority, although due to significant bleeding, the overall interpretation of the trial was considered as non-inferior and inferior.

Not non-inferior and inferior

The INSURE (Indobufen versus Aspirin in Acute Ischemic Stroke) trial was a randomized, double-blind non-inferiority trial which examined the hypothesis that indobufen is non-inferior to aspirin in reducing the risk of new stroke at 90 days in patients with moderate-to-severe ischemic stroke.³⁰ If the upper limit of the 95% CI for the event risk in the treatment group was less than 1.25, indobufen would be considered noninferior to aspirin. The HR for the primary endpoint for the indobufen group was, 1.23 with a 95% CI of 1.01 to 1.50. Firstly, the upper limit of the 95% CI crossed the non-inferiority margin of 1.25, establishing that indobufen is not non-inferior to aspirin. Further, the lower margin of the 95% CI was greater than 1.00, suggesting that indobufen was also inferior to aspirin. Therefore indobufen was both not non-inferior and inferior to aspirin.

Other challenges to consider when interpreting non-inferiority trial results

Absolute or relative risk difference

An important assumption required for an accurate interpretation of non-inferiority trials is the assumption of inter-trial constancy, which allows the superiority of the active control compared to placebo in the current trial to be inferred based on historical trials. Consequently, heterogeneity between trials can lead to incorrectly claiming non-inferiority when the treatment was ineffective or even harmful. The choice of the δ can sometimes also modify this potential challenge.

As discussed earlier, the non-inferiority margin δ can be chosen as an absolute risk difference or relative risk difference (usually presented as a risk ratio). When the constancy assumption is difficult to assess or when the baseline risk or event rate is expected to vary between studies or patient population, it is rec-

Table 3. Hypothetical example

	Expected risks	Observed risks
Control group (%)	20	10
New drug/intervention (%)	25	15
Absolute risk difference (%)	5	5
Relative risk difference	1.25	1.50
Sample size	546	970

In a trial where the margin δ was set using absolute difference based on expected risks, the relative risk difference can be calculated accordingly (i.e., 5%). However, if in reality, the observed risks were much lower, the same absolute risk difference would appear to be more generous. Moreover, there would be an inflation of the relative risk difference (from 1.25 to 1.50). If the relative risk from the expected risks were used instead (i.e., 1.25), the relevant absolute risk difference in the trial would need to be 2.5% (10%*1.25=10%; instead of 5%), which would make the results more conservative but at the same time the sample size also increased.

ommended to use a relative risk difference approach to account for changes in event rates over time. Using fixed risk ratios is usually a more conservative approach when the event rate is unpredictable or when the observed rate is lower than expected,³¹ but it also inevitably results in larger sample sizes. On the other hand, δ based on absolute risk difference can potentially introduce a bias towards non-inferiority by resulting in an underpowered trial due to lower than expected event rates.^{2,30} It might be helpful to check the concordance of non-inferiority conclusion using both approaches when an absolute risk difference was used in the original trial. Table 3 illustrated a hypothetical example.

Quality of trial conduct

Quality of trial conduct is always important for RCTs. However, this is also a particular weakness for non-inferiority trials when there is poor adherence, loss to follow-up or treatment crossover.³² In a superiority trial, any such inadequacy in the execution of the trial would tend to make the estimate more conservative, which means that the trial would be more likely not to be able to show the differences between the groups. In a non-inferiority trial, anything that tends to make the two treatment groups more similar is by definition more likely to result in falsely concluding non-inferiority. Whilst this is also an obvious problem for superiority trials, they can pose higher risks to patients in a non-superiority trial setting.

Choices of analytic approaches

Analytic approaches can also be challenging in a non-inferiority trial. In a superiority trial, an intention-to-treat analysis is recommended by guidelines.²⁶ For a superiority trial, this approach is considered conservative as including these patients who did not receive the treatment is more likely to narrow the difference

between the two treatment groups. However, in a non-inferiority trial, this approach may, for the same reason, result in a bias towards a false positive conclusion of non-inferiority.³² As a result, a per-protocol analysis approach might be helpful in a non-inferiority trial setting, although it might violate the randomization principle causing unmeasured confounding and might also result in smaller sample sizes. It is commonly recommended that both analyses are reported for non-inferiority trials with the intention-to-treat analysis as the primary approach and per-protocol analysis as a sensitivity analysis.³³

Power deflation

Power deflation is a challenge where the statistical power of a non-inferiority trial decreases if the new treatment or intervention unexpectedly performs better than the control arm. For example, if we set a non-inferiority margin of 10% and the rate for the primary outcome expected in the control arm is 70%, we are hoping to show that the rate of the primary outcome of the new treatment is no less than 60%. Assuming a power of 80% and significance level at 0.05, 785 participants per group is needed. If in reality the new treatment performs better than the control (e.g., 80% instead of 70%), the estimated effect size is now larger resulting in a wider CI. Therefore a bigger sample size (now $n=2,285$) will be required to reduce this CI to achieve the "non-inferiority" concept. This highlights how non-inferiority trials can be sensitive to variations in the performance of the new treatment. On the other hand, if a superiority trial was designed in the first instance and if the expected effect size is also 10% (i.e., 80%–70%), the sample size required would be 394.

Evolving standard of care

Another concern specific to non-inferiority trials is related to the evolving standard of care.³⁴ For example, drug A is compared to a placebo and demonstrated superiority in a superiority trial and becomes standard of care. Then a new drug B is compared to drug A in a non-inferiority design and demonstrated non-inferiority as compared to A with some cost advantages. Now drug B is standard of care. Subsequently, a new drug C is compared to drug B also with a non-inferiority design (with a new δ' which is a further "worsening" from drug A compared to drug B). In this case, drug C might be less effective than drug A or even placebo hence a further comparison of a new drug D with drug C might not be clinically ethical. In the area of infectious disease, a true biological biocrep is also well-recognized. In this case, bacteria can mutate and become more resistant to standard care. Using a non-inferiority design to test for new drugs could potentially introduce treatment that is less and less effective against strengthening bacteria.^{5,35}

Conclusion

Non-inferiority trials can provide important information about the effectiveness of a new treatment compared to an established treatment. However, the determination of non-inferiority margin requires careful consideration and should be selected based on statistical reasoning and clinical judgment. The results of non-inferiority trials should be interpreted with caution and must account for the non-inferiority hypotheses and any superiority hypotheses.

Funding statement

None

Conflicts of interest

The authors have no financial conflicts of interest.

Author contribution

Conceptualization: LL, GH, PMB. Study design: LL. Methodology: LL. Data collection: all authors. Writing—original draft: all authors. Writing—review & editing: all authors. Approval of final manuscript: all authors.

Acknowledgments

This project is a joint effort with special acknowledgment to the World Stroke Organization Future Leaders Program. No designated funding was required for the manuscript.

References

- Lang J, Cetre JC, Picot N, Lanta M, Briantais P, Vital S, et al. Immunogenicity and safety in adults of a new chromatographically purified Vero-cell rabies vaccine (CPRV): a randomized, double-blind trial with purified Vero-cell rabies vaccine (PVRV). *Biologicals* 1998;26:299–308.
- Head SJ, Kaul S, Bogers AJ, Kappetein AP. Non-inferiority study design: lessons to be learned from cardiovascular trials. *Eur Heart J* 2012;33:1318–1324.
- Burotto M, Prasad V, Fojo T. Non-inferiority trials: why oncologists must remain wary. *Lancet Oncol* 2015;16:364–366.
- Bai AD, Komorowski AS, Lo CKL, Tandon P, Li XX, Mokashi V, et al. Methodological and reporting quality of noninferiority randomized controlled trials comparing antibiotic therapies: a systematic review. *Clin Infect Dis* 2021;73:e1696–e1705.
- Schumi J, Wittes JT. Through the looking glass: understanding non-inferiority. *Trials* 2011;12:106.
- Chow SC, Shao J, Wang H, Lohknygina Y. Sample Size Calculations in Clinical Research. 3rd ed. Boca Raton, FL: CRC Press, 2017.
- European Medicines Agency. Choice of a non-inferiority margin – scientific guideline [Internet]. Amsterdam: European Medicines Agency; 2005 [accessed September 1, 2024]. Available from: <https://www.ema.europa.eu/en/choice-non-inferiority-margin-scientific-guideline>.
- U.S. FDA. Non-inferiority clinical trials [Internet]. Silver Spring: U.S. FDA; 2016 [accessed September 1, 2024]. Available from: <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/non-inferiority-clinical-trials>.
- Ito C, Hashimoto A, Uemura K, Oba K. Misleading reporting (spin) in noninferiority randomized clinical trials in oncology with statistically not significant results: a systematic review. *JAMA Netw Open* 2021;4:e2135765.
- Rehal S, Morris TP, Fielding K, Carpenter JR, Phillips PP. Non-inferiority trials: are they inferior? A systematic review of reporting in major medical journals. *BMJ Open* 2016;6:e012594.
- Patel MR, Mahaffey KW, Garg J, Pan G, Singer DE, Hacke W, et al. Rivaroxaban versus warfarin in nonvalvular atrial fibrillation. *N Engl J Med* 2011;365:883–891.
- Granger CB, Alexander JH, McMurray JJ, Lopes RD, Hylek EM, Hanna M, et al. Apixaban versus warfarin in patients with atrial fibrillation. *N Engl J Med* 2011;365:981–992.
- Connolly SJ, Ezekowitz MD, Yusuf S, Eikelboom J, Oldgren J, Parekh A, et al. Dabigatran versus warfarin in patients with atrial fibrillation. *N Engl J Med* 2009;361:1139–1151.
- Giugliano RP, Ruff CT, Braunwald E, Murphy SA, Wiviott SD, Halperin JL, et al. Edoxaban versus warfarin in patients with atrial fibrillation. *N Engl J Med* 2013;369:2093–2104.
- Jackson K, Gersh BJ, Stockbridge N, Fleming TR, Temple R, Califf RM, et al. Antithrombotic drug development for atrial fibrillation: proceedings, Washington, DC, July 25–27, 2005. *Am Heart J* 2008;155:829–840.
- Menon BK, Buck BH, Singh N, Deschaintre Y, Almekhlafi MA, Coutts SB, et al. Intravenous tenecteplase compared with alteplase for acute ischaemic stroke in Canada (AcT): a pragmatic, multicentre, open-label, registry-linked, randomised, controlled, non-inferiority trial. *Lancet* 2022;400:161–169.
- Emberson J, Lees KR, Lyden P, Blackwell L, Albers G, Bluhmki E, et al. Effect of treatment delay, age, and stroke severity on the effects of intravenous thrombolysis with alteplase for acute ischaemic stroke: a meta-analysis of individual patient data from randomised trials. *Lancet* 2014;384:1929–1935.
- Berge E, Whiteley W, Audebert H, De Marchis GM, Fonseca

- AC, Padiglioni C, et al. European Stroke Organisation (ESO) guidelines on intravenous thrombolysis for acute ischaemic stroke. *Eur Stroke J* 2021;6:I-LXII.
19. Turc G, Bhogal P, Fischer U, Khatri P, Lobotesis K, Mazighi M, et al. European Stroke Organisation (ESO) – European Society for Minimally Invasive Neurological Therapy (ESMINT) guidelines on mechanical thrombectomy in acute ischemic stroke. *J Neurointerv Surg* 2023;15:e8.
 20. Turc G, Tsvigoulis G, Audebert HJ, Boogaarts H, Bhogal P, De Marchis GM, et al. European Stroke Organisation–European Society for Minimally Invasive Neurological Therapy expedited recommendation on indication for intravenous thrombolysis before mechanical thrombectomy in patients with acute ischaemic stroke and anterior circulation large vessel occlusion. *Eur Stroke J* 2022;7:I-XXVI.
 21. Majoie CB, Cavalcante F, Gralla J, Yang P, Kaesmacher J, Treurniet KM, et al. 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.
 22. Kaesmacher J, Mujanovic A, Treurniet K, Kappelhof M, Meinel TR, Yang P, et al. Perceived acceptable uncertainty regarding comparability of endovascular treatment alone versus intravenous thrombolysis plus endovascular treatment. *J Neurointerv Surg* 2023;15:227–232.
 23. Sacco RL, Diener HC, Yusuf S, Cotton D, Ounpuu S, Lawton WA, et al. Aspirin and extended-release dipyridamole versus clopidogrel for recurrent stroke. *N Engl J Med* 2008;359:1238–1251.
 24. CAPRIE Steering Committee. A randomised, blinded, trial of clopidogrel versus aspirin in patients at risk of ischaemic events (CAPRIE). *Lancet* 1996;348:1329–1339.
 25. Antithrombotic Trialists' Collaboration. Collaborative meta-analysis of randomised trials of antiplatelet therapy for prevention of death, myocardial infarction, and stroke in high risk patients. *BMJ* 2002;324:71–86.
 26. Schulz KF, Altman DG, Moher D; CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *BMJ* 2010;340:c332.
 27. Wang Y, Li S, Pan Y, Li H, Parsons MW, Campbell BCV, et al. 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.
 28. SoSTART Collaboration. Effects of oral anticoagulation for atrial fibrillation after spontaneous intracranial haemorrhage in the UK: a randomised, open-label, assessor-masked, pilot-phase, non-inferiority trial. *Lancet Neurol* 2021;20:842–853.
 29. Amadeus Investigators; Bousser MG, Bouthier J, Büller HR, Cohen AT, Crijns H, Davidson BL, et al. Comparison of idraparinux with vitamin K antagonists for prevention of thromboembolism in patients with atrial fibrillation: a randomised, open-label, non-inferiority trial. *Lancet* 2008;371:315–321.
 30. Pan Y, Meng X, Yuan B, Johnston SC, Li H, Bath PM, et al. Indobufen versus aspirin in patients with acute ischaemic stroke in China (INSURE): a randomised, double-blind, double-dummy, active control, non-inferiority trial. *Lancet Neurol* 2023;22:485–493.
 31. Kaul S, Diamond GA, Weintraub WS. Trials and tribulations of non-inferiority: the ximelagatran experience. *J Am Coll Cardiol* 2005;46:1986–1995.
 32. Mo Y, Lim C, Watson JA, White NJ, Cooper BS. Non-adherence in non-inferiority trials: pitfalls and recommendations. *BMJ* 2020;370:m2215.
 33. Piaggio G, Elbourne DR, Pocock SJ, Evans SJ, Altman DG; CONSORT Group. Reporting of noninferiority and equivalence randomized trials: extension of the CONSORT 2010 statement. *JAMA* 2012;308:2594–2604.
 34. D'Agostino RB Sr, Massaro JM, Sullivan LM. Non-inferiority trials: design concepts and issues – the encounters of academic consultants in statistics. *Stat Med* 2003;22:169–186.
 35. Fleming TR, Powers JH. Issues in noninferiority trials: the evidence in community-acquired pneumonia. *Clin Infect Dis* 2008;47(Suppl 3):S108–S120.