Reading and Designing Non-Inferiority Trials in Anaesthesia and Critical Care: A Practical Guide with Worked Examples

Ayesha Sara

PharmD (Doctor of Pharmacy), Clinical Pharmacist, KIMS Hospital, Secunderabad, India.

Corresponding author: ayeshasara28@gmail.com

Received: 27 April 2026; Revised: 06 July 2026; Accepted: 13 July 2026; Published: 29 August 2026

Abstract

Background. Much of the recent trial literature in anaesthesia and critical care asks whether a new technique, device or drug is not unacceptably worse than an established one, rather than whether it is better. Non-inferiority trials answer that question, and they carry a set of traps that do not exist in superiority trials: the conclusion depends on a margin chosen by the investigators, the usual protections of intention-to-treat analysis run in the wrong direction, and a sequence of individually valid trials can erode a standard of care.

Objective. To provide a practical guide to reading and designing non-inferiority trials, illustrated with worked examples and quantified by simulation.

Key points. Sample size is inversely proportional to the square of the margin, so halving the margin quadruples the trial: at a 10 % control event rate and 90 % power, a 5 percentage-point margin requires 757 patients per arm and a 2-point margin requires 4,729. If the new treatment is genuinely 1 percentage point worse rather than exactly equivalent, those figures become 1,234 and 19,744 respectively. Simulation shows that non-adherence inflates the rate at which a truly inferior treatment is declared non-inferior: with a treatment truly inferior by exactly the margin, an intention-to-treat analysis declared non-inferiority in 2.4 % of trials under full adherence, rising to 8.6 % with 10 % crossover and 42.4 % with 30 % crossover. Repeated non-inferiority comparisons permit biocreep, with a worst-case event rate rising from 10 % to 20 % after five generations at a 2-point margin.

Conclusions. A non-inferiority result is interpretable only in the light of its margin, its adherence and the evidence that the active control works. Nine reading questions and a design checklist are provided.

Keywords: Non-Inferiority, Clinical Trial Design, Anaesthesia, Critical Care, Research Methodology, Statistical Interpretation.

1. WHY THIS MATTERS

Most of the interventions an anaesthetist chooses between already work. The question is rarely whether a regional block, an airway device, a sedative or a monitoring strategy is better than nothing, but whether a newer option that is cheaper, quicker, safer in some other respect or simply easier to teach is close enough to the established one to adopt. That is a non-inferiority question, and it now accounts for a substantial share of trials in the specialty.

Non-inferiority trials are harder to read than superiority trials and much harder to design, for one structural reason: their conclusion depends on a number the investigators chose. In a superiority trial the null hypothesis is fixed by nature — no difference — and the reader need not accept anything from the investigators except the data. In a non-inferiority trial the null hypothesis is that the new treatment is worse by at least some margin, and that margin is a judgement. Choose it generously and almost anything passes; choose it strictly and the trial becomes unaffordable.

This guide sets out what a reader should check, what a designer must decide, and what the arithmetic costs. Every number reported is computed rather than asserted, and the simulations behind Sections 4 and 5 use a control event rate of 10 %, which is representative of outcomes such as block failure, unplanned conversion or device insertion failure in anaesthetic practice.

2. THE LOGIC AND THE VOCABULARY

Table 1. Terms used in this guide

Term	Meaning	Where it causes trouble
Non-inferiority margin (Δ)	The largest loss of effect the investigators are prepared to accept in exchange for the new treatment's other advantages	Chosen by investigators; a generous margin makes the trial easy to pass
Active control	The established treatment used as comparator	If its own benefit is uncertain, non-inferiority to it means little
Assay sensitivity	The trial's ability to have detected a difference had one existed	Cannot be verified internally; a sloppy trial tends toward non-inferiority
Constancy assumption	The active control works as well in this trial as in the historical trials from which the margin was derived	Fails when populations, co-interventions or era differ
Intention to treat	Analysis by assigned treatment regardless of what was received	Conservative in superiority trials; anti-conservative here
Per protocol	Analysis restricted to adherent patients	Restores the contrast but breaks randomisation
Biocreep	Successive treatments each non-inferior to the last, drifting away from the original standard	Invisible within any single trial

2.1 The hypothesis is reversed

In a superiority trial the null hypothesis is that the difference is zero and the trial tries to reject it. In a non-inferiority trial the null is that the new treatment is worse by at least Δ, and the trial tries to reject *that*. Writing θ for the difference in event rate (new minus control, with a higher rate being worse):

H_0: $\theta \geq \Delta$ versus H_1: $\theta < \Delta$ (1)

Non-inferiority is concluded when the upper limit of the two-sided 95 % confidence interval for θ lies entirely below Δ. Everything in this guide follows from that sentence, including the traps.

2.2 Reading the Interval

Figure 1 shows the seven results a non-inferiority trial can produce. The habit of looking for whether the interval crosses zero — the correct instinct in a superiority trial — is the wrong instinct here.

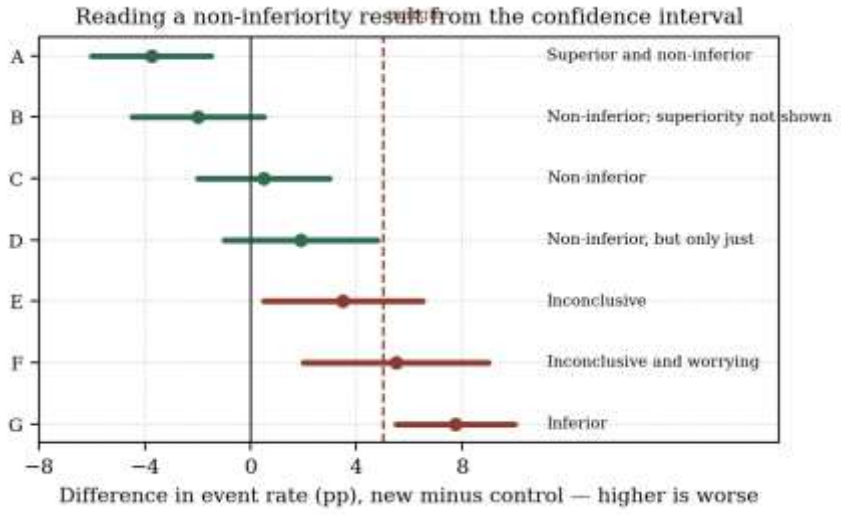

Figure 1. The seven interpretations of a non-inferiority result, according to where the confidence interval sits relative to zero and the margin.

Scenario A shows a new treatment that is both non-inferior and superior, which is a legitimate conclusion from a non-inferiority trial provided superiority was pre-specified as a secondary question. Scenarios B to D are non-inferior, with D only just so — the interval reaching almost to the margin, which should prompt the reader to ask how the margin was chosen. Scenarios E and F are inconclusive: the interval crosses the margin, so neither non-inferiority nor inferiority is established, and a trial reporting these as a negative result is being read carelessly. Only in G, where the entire interval lies above zero and beyond the margin, is the new treatment shown to be inferior.

Two errors are common. The first is reading E or F as evidence of equivalence, on the grounds that the interval includes zero; it does, but it also includes unacceptable harm, which is what matters. The second is reading a narrow interval sitting just below the margin as reassuring without checking what the margin was.

3. CHOOSING THE MARGIN

The margin is the single most consequential decision in the design, and the one most often made by convention. Two principles govern it.

The clinical principle. The margin must be a loss of effect that clinicians and patients would genuinely accept in exchange for whatever the new treatment offers. A margin of 5 percentage points in block failure means accepting that one additional patient in twenty will have a failed block; whether that is a fair price for a technique that takes ten minutes less is a clinical judgement and should be justified as one, preferably with patient involvement.

The statistical principle. The margin must preserve a defined fraction of the active control's own benefit over doing nothing. If the established treatment reduces the event rate by 12 percentage points compared with no treatment, a margin of 6 points would permit a new treatment that retains only half of that benefit. Figure 2 shows the relationship.

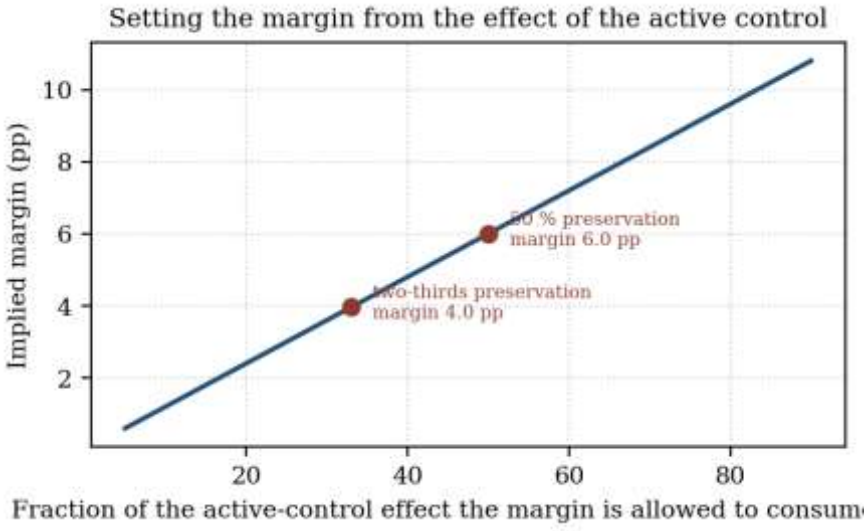

Figure 2. The margin implied by preserving a given fraction of the active control's effect, for an illustrative control effect of 12 percentage points.

The statistical principle requires knowing how well the active control works, which requires placebo-controlled trials of the active control that are often old, conducted in different populations, and sometimes absent altogether. Where they are absent, the margin cannot be anchored and non-inferiority to the active control establishes very little — possibly only that both treatments are equally ineffective. This is the deepest problem in non-inferiority design and it has no technical solution.

4. WHAT THE MARGIN COSTS: A WORKED EXAMPLE

Consider a trial comparing a new peripheral nerve block technique against an established one. The established technique fails in 10 % of patients. The new technique is faster to perform and the investigators are willing to accept some increase in failure in exchange. Required sample size per arm for one-sided 2.5 % significance and 90 % power is

$$n = (z_{1-\alpha} + z_{1-\beta})^2 \, [\, p_c(1 - p_c) + p_t(1 - p_t) \,] \, / \, (\Delta - \delta)^2 \qquad (2)$$

Where δ is the difference genuinely expected between the treatments. Table 2 evaluates it for two assumptions: that the new technique is exactly as good as the old (δ = 0), and that it is in truth 1 percentage point worse (δ = 0.01).

Table 2. Patients required per arm, control failure rate 10 %, one-sided 2.5 %, 90 % power

Margin Δ	Assuming exact equivalence	Assuming the new technique is 1 pp worse	Ratio
2 pp	4,729	19,744	4.2×
3 pp	2,102	4,936	2.3×
5 pp	757	1,234	1.6×
7.5 pp	337	468	1.4×
10 pp	190	244	1.3×

Note. pp = percentage points. The right-hand column shows how sharply the requirement escalates when the new treatment is not assumed to be exactly as good as the old — which is the realistic assumption, since a treatment expected to be exactly equivalent would not need testing.

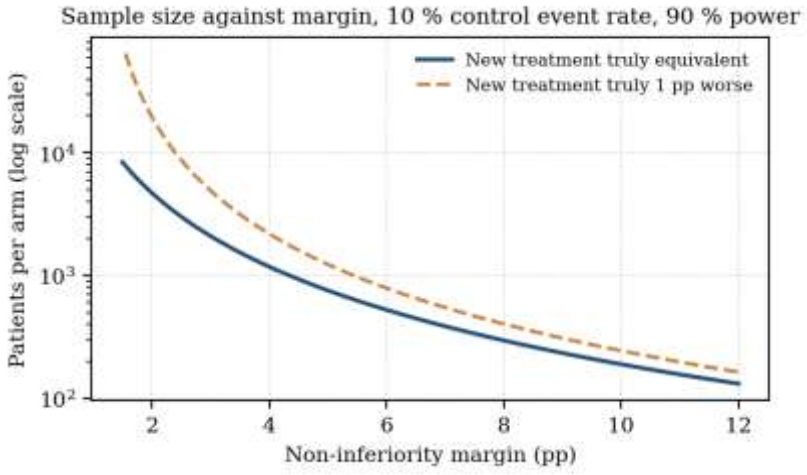

Figure 3. Patients required per arm against the non-inferiority margin, on a logarithmic scale.

Two lessons follow. Sample size scales with the inverse square of the margin, so halving the margin quadruples the trial; the jump from a 5-point to a 2-point margin multiplies the requirement more than sixfold. And the common design shortcut of assuming exact equivalence is optimistic: allowing for a true 1-point inferiority multiplies the requirement by 4.2 at a 2-point margin. A trial powered on the assumption of exact equivalence, in which the new treatment turns out to be slightly worse, is underpowered for its own question and will return an inconclusive interval of the kind shown as scenario E in Figure 1.

This arithmetic is also why generous margins are chosen. A 10-point margin makes the trial feasible at 190 patients per arm; it also means the trial can declare non-inferiority for a technique that doubles the failure rate. Readers should treat a large margin as a finding about the trial, not a detail of it.

5. THE ADHERENCE TRAP

In a superiority trial, patients who do not receive their assigned treatment dilute the difference between arms, pushing the estimate toward the null and making it harder to declare a difference. Intention-to-treat

analysis is therefore conservative, which is one reason it is the default. In a non-inferiority trial the same dilution pushes the estimate toward the null too — and the null is now what the investigators want to show. The protective property is inverted.

Figure 4 quantifies this by simulation for the worked example, with a 5-point margin and 757 patients per arm. A proportion of patients in each arm receives the other treatment, and the trial is analysed by intention to treat.

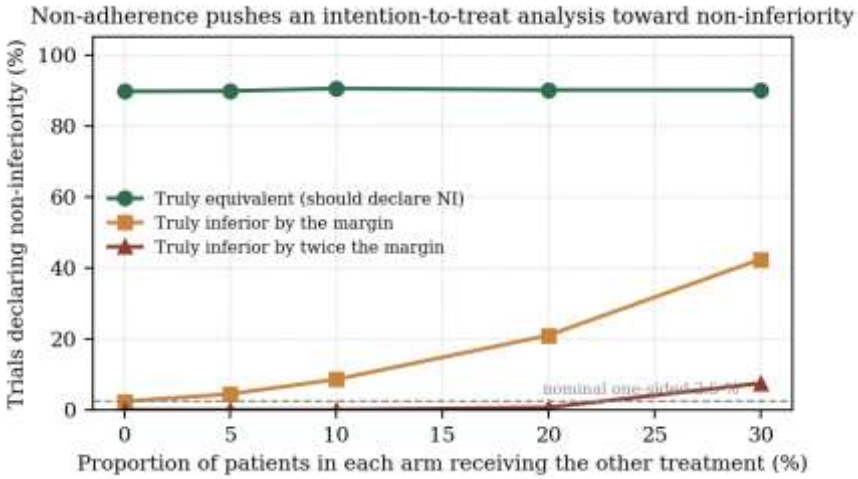

Figure 4. Proportion of simulated trials declaring non-inferiority, by degree of crossover and by the true effect of the new treatment.

Table 3. Trials declaring non-inferiority (%), by crossover and true effect

Crossover in each arm	New treatment truly equivalent	Truly inferior by exactly the margin	Truly inferior by twice the margin
0 %	89.7	2.4	0.0
5 %	89.8	4.5	0.0
10 %	90.5	8.6	0.0
20 %	90.1	21.0	0.7
30 %	90.1	42.4	7.5

Note. 4,000 simulated trials per cell, 757 patients per arm, 5 percentage-point margin, intention-to-treat analysis. The second column confirms that power against a genuinely equivalent treatment is unaffected by crossover; the third and fourth show the inflation of the error that matters.

The second column shows that crossover does not reduce the trial's power when the new treatment really is equivalent: it stays at about 90 % throughout. The third column shows what it does to the error that matters. A treatment that is truly inferior by exactly the margin — the boundary case the design is supposed to reject — is wrongly declared non-inferior in 2.4 % of trials under perfect adherence, which is the nominal one-sided error rate, but in 8.6 % with 10 % crossover and 42.4 % with 30 % crossover. At that level of non-adherence, a trial is close to a coin toss on the question it was built to answer.

Crossover rates of 10 to 20 % are entirely ordinary in pragmatic anaesthetic trials, where a block may be abandoned, a device may fail to insert and the alternative is used, or a clinician may deviate for a reason recorded as clinical judgement. The implication is not that intention to treat should be abandoned — per-protocol analysis breaks randomisation and introduces its own bias, which can run in either direction — but that both analyses must be reported and must agree. A non-inferiority trial that reports only intention to treat, or in which the two analyses disagree, has not established non-inferiority.

6. BIOCREEP

Each non-inferiority trial is judged against its own immediate comparator. If a new treatment is accepted as non-inferior to the standard, and then becomes the comparator for the next trial, the chain of individually valid conclusions can drift away from the original standard. Figure 5 traces the worst case.

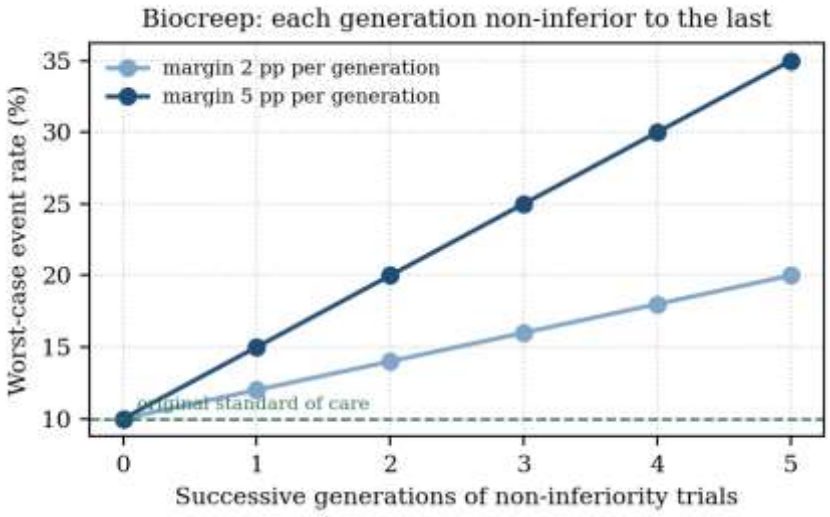

Figure 5. Worst-case drift in event rate across successive generations of non-inferiority trials.

At a 2-point margin, five generations of trials each correctly concluding non-inferiority permit a worst-case event rate of 20 % against an original 10 %: the failure rate has doubled and no individual trial did anything wrong. At a 5-point margin the same five generations reach 35 %.

The drift is a worst case rather than a prediction, since a new treatment declared non-inferior is usually not as bad as the margin allows, and later trials sometimes show superiority. But nothing in the design of any individual trial detects the drift, because each compares only against its immediate predecessor. The defences are external: anchoring the margin to the original placebo-controlled evidence rather than to the current comparator, retaining the original standard as a third arm where feasible, and periodically re-establishing the absolute effect of current practice.

7. READING AND DESIGNING: TWO CHECKLISTS

Table 4. Nine questions for the reader

#	Question	Why it matters
1	Was the trial pre-specified as non-inferiority?	A superiority trial reinterpreted after a null result is not a non-inferiority trial
2	What is the margin, and how was it justified?	An unjustified margin makes the conclusion uninterpretable
3	Does the margin preserve a stated fraction of the control's own effect?	Anchors the margin in evidence rather than convenience
4	Is there good evidence that the active control works?	Non-inferiority to an ineffective control establishes nothing
5	Where does the confidence interval sit relative to zero and the margin?	The P value alone cannot distinguish Figure 1's scenarios
6	What was the crossover rate?	Table 3 shows the error inflation it produces
7	Are intention-to-treat and per-protocol analyses both reported, and do they agree?	Disagreement undermines the conclusion

8	Was the trial conducted well enough to have detected a difference?	Poor conduct biases toward non-inferiority
9	Was the comparator itself established by a previous non-inferiority trial?	If so, consider cumulative drift

Table 5. Design checklist

Decision	Recommended practice
Margin	Justify clinically and statistically; state the fraction of the control effect preserved; involve patients in the judgement
Sample size	Do not assume exact equivalence; power for a plausible small true inferiority (Table 2)
Analysis population	Pre-specify both intention-to-treat and per-protocol as co-primary; require agreement
Adherence	Set a monitored ceiling on crossover; report it by arm; pre-specify a sensitivity analysis if exceeded
Superiority	Pre-specify testing for superiority if non-inferiority is met; this requires no alpha adjustment in that order
Assay sensitivity	Report process measures demonstrating the trial could have detected a difference
Reporting	Follow the CONSORT extension for non-inferiority trials; present the interval against the margin graphically

8. LIMITATIONS OF THIS GUIDE

The worked examples use a binary outcome with a 10 % control event rate and a normal approximation to the risk difference, which is appropriate for the sample sizes discussed but not for very rare outcomes or very small trials. Continuous and time-to-event outcomes follow the same logic with different formulae, and the qualitative conclusions carry over, but the specific numbers in Tables 2 and 3 do not.

The adherence simulation models crossover as occurring at random and independently of prognosis. In practice patients who cross over differ systematically from those who do not — a block is abandoned because it is proving difficult, which correlates with the chance of failure — and such informative crossover would generally worsen the picture in Table 3 rather than improve it. The magnitudes reported are therefore optimistic.

The biocreep analysis is a worst-case arithmetic exercise rather than an empirical observation, and the extent to which drift has actually occurred in any anaesthetic field is unknown. Establishing it empirically would require tracing comparator chains across successive trials in a defined area, which would be a valuable study and has not been done here.

Finally, this guide addresses the statistical structure of non-inferiority trials and not the prior question of whether a non-inferiority design is the right choice. A new treatment with no advantage in cost, safety, speed or acceptability should not be studied for non-inferiority at all, since establishing that it is not much worse than the alternative provides no reason to adopt it.

9. CONCLUSION

A non-inferiority trial answers a genuinely useful question and answers it in a way that rewards carelessness. Its conclusion turns on a margin the investigators selected, its sample size escalates with the inverse square of that margin, and the analytic convention that protects superiority trials from sloppy conduct actively assists a non-inferiority trial toward its preferred conclusion — a treatment truly inferior by exactly the margin was declared non-inferior in 42 % of simulated trials at 30 % crossover.

The reader's defence is to look at three things before anything else: the margin and its justification, the position of the confidence interval relative to that margin, and the crossover rate with both analysis populations reported. A trial that does not supply all three has not demonstrated what its abstract claims, however large it is and however small its P value.

Declarations

Ethical approval: Not applicable.

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Conflicts of interes: The authors declare that there are no conflicts of interest regarding the publication of this paper.

Data availability: The data that support the findings of this study are available from the corresponding author upon reasonable request.

Author contributions:

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Ayesha Sara	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

10. REFERENCES

1. G. Piaggio, D. R. Elbourne, S. J. Pocock, S. Evans, and D. C. Altman, 'Reporting of noninferiority and equivalence randomized trials: extension of the CONSORT 2010 statement', JAMA, vol. 308, no. 24, pp. 2594-2604, 2012. doi.org/10.1001/jama.2012.87802
2. L. Mauri and D. Agostino, 'Challenges in the design and interpretation of noninferiority trials', N Engl J Med, vol. 377, no. 14, pp. 1357-1367, 2017. doi.org/10.1056/NEJMra1510063
3. S. Kaul and G. A. Diamond, 'Good enough: a primer on the analysis and interpretation of noninferiority trials', Ann Intern Med, vol. 145, no. 1, pp. 62-69, 2006. doi.org/10.7326/0003-4819-145-1-200607040-00011
4. T. R. Fleming, 'Current issues in non-inferiority trials', Stat Med, vol. 27, no. 3, pp. 317-332, 2008. doi.org/10.1002/sim.2855
5. S. M. Snapinn, 'Noninferiority trials', Noninferiority trials. Curr Control Trials Cardiovasc Med, vol. 1, no. 1, pp. 19-21, 2000. doi.org/10.1186/cvm-1-1-019
6. D. Agostino, R. B. Massaro, and J. M. Sullivan, 'Non-inferiority trials: design concepts and issues - the encounters of academic consultants in statistics', Stat Med, vol. 22, no. 2, pp. 169-186, 2003. doi.org/10.1002/sim.1425
7. S. Hahn, 'Understanding noninferiority trials', Korean J Pediatr, vol. 55, no. 11, pp. 403-407, 2012. doi.org/10.3345/kjp.2012.55.11.403
8. S. J. Head, S. Kaul, A. Bogers, and A. P. Kappetein, 'Non-inferiority study design: lessons to be learned from cardiovascular trials', Eur Heart J, vol. 33, no. 11, pp. 1318-1324, 2012. doi.org/10.1093/eurheartj/ehs099
9. J. Schumi and J. T. Wittes, 'Through the looking glass: understanding non-inferiority', Trials, vol. 12, 2011. doi.org/10.1186/1745-6215-12-106
10. M. M. Sanchez and X. Chen, 'Choosing the analysis population in non-inferiority studies: per protocol or intent-to-treat', Stat Med, vol. 25, no. 7, pp. 1169-1181, 2006. doi.org/10.1002/sim.2244
11. E. Brittain and D. Lin, 'A comparison of intent-to-treat and per-protocol results in antibiotic non-inferiority trials', Stat Med, vol. 24, no. 1, pp. 1-10, 2005. doi.org/10.1002/sim.1934
12. S. Everson-Stewart and S. S. Emerson, 'Bio-creep in non-inferiority clinical trials', Stat Med, vol. 29, no. 27, pp. 2769-2780, 2010. doi.org/10.1002/sim.4053
13. T. A. Althunian, A. De Boer, R. Groenwold, and O. H. Klungel, 'Defining the noninferiority margin and analysing noninferiority: an overview', Br J Clin Pharmacol, vol. 83, no. 8, pp. 1636-1642, 2017. doi.org/10.1111/bcp.13280

14. S. Rehal, T. P. Morris, K. Fielding, J. R. Carpenter, and P. Phillips, 'Non-inferiority trials: are they inferior? A systematic review of reporting in major medical journals', BMJ Open, vol. 6, no. 10, 2016. doi.org/10.1136/bmjopen-2016-012594