BMJ Open

Defining non-inferiority margins in randomised controlled surgical trials: a protocol for a systematic review

Christoph Kuemmerli ,[1,2] Iain J Gallagher,[3] Richard Skipworth,[4] Barry Laird[5]

To cite: Kuemmerli C, Gallagher IJ, Skipworth R, *et al*. Defining non-inferiority margins in randomised controlled surgical trials: a protocol for a systematic review. *BMJ Open* 2024;**14**:e089587. doi:10.1136/bmjopen-2024-089587

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (https://doi.org/10.1136/bmjopen-2024-089587).

Received 03 June 2024
Accepted 09 August 2024

[1]Molecular, Genetic and Population Health Sciences, The University of Edinburgh Medical School, Edinburgh, UK
[2]Department of Surgery, University Digestive Health Care Center, Clarunis, University Hospital Basel, Basel, Switzerland
[3]Centre for Biomedicine & Global Health, Edinburgh Napier University, Edinburgh, UK
[4]Department of General Surgery, Royal Infirmary of Edinburgh, Edinburgh, UK
[5]European Palliative Care Research Centre, Edinburgh Cancer Research Centre, Edinburgh, UK

Correspondence to
Dr Christoph Kuemmerli; Christoph.kuemmerli@clarunis.ch

ABSTRACT

Background The reporting of randomised controlled non-inferiority (NI) drug trials is poor with less than 50% of published trials reporting a justification of the NI margin. This is despite the introduction of the Consolidated Standards of Reporting Trials (CONSORT) extension on reporting of NI and equivalence in randomised trials. It is critical to set the appropriate NI margin as this choice dictates the conclusions of the trial. Methods to estimate the margin are heterogeneous but generally based on clinical judgement and statistical reasoning, and hence tailored to each clinical situation. Yet an appraisal of NI in clinical trials has not been undertaken. Therefore the aim of this systematic review is to assess the reporting and methodological quality of defining the NI margin. Surgical NI trials have been chosen as our prototype to assess this.

Methods We will conduct a systematic review of published randomised controlled trials in abdominal surgery that use an NI design. Key eligibility criteria will be: surgical intervention in at least one trial arm; adult patients and a sample size of 100 or more. Ovid MEDLINE, EMBASE and the Cochrane Central Register of Controlled Trials will be searched from inception until the date of the search. Identified studies will be assessed for reporting according to the CONSORT recommendations. The outcomes are the description of the methods for defining the NI margin, and the robustness of the NI margin estimation. The latter will be based on simulations using alternative assumptions for model parameters. The results of the simulation will be compared with the trial authors' conclusions.

Anticipated results The review will describe and appraise the design and reporting of surgical NI trials including shortcomings thereof and allow a comparison with pharmaceutical trials. These findings will inform researchers on the appropriate design and pitfalls when conducting surgical randomised controlled trials with an NI design and promote thorough and standardised reporting of study findings.

Ethics and dissemination Ethical approval is not required and any changes to the protocol will be communicated via the registration platform. The final manuscript will be submitted to a journal for publication and the findings will be disseminated through conference presentations to inform researchers and the public.

STRENGTHS AND LIMITATIONS OF THIS STUDY

⇒ This systematic review will identify the potential for improving reporting quality and the design of surgical non-inferiority trials.
⇒ The search strategy will be comprehensive without language restrictions and input from experienced trialists and a librarian.
⇒ Simulation analysis will enable the assessment of the robustness beyond descriptive statistics.
⇒ Only trials published after 2006, the publication of the Consolidated Standards of Reporting Trials extension for non-inferiority trials, will be included.
⇒ Only abdominal surgical non-inferiority trials will be analysed limiting the generalisability of the findings.

INTRODUCTION

Randomised controlled non-inferiority (NI) trials are increasingly used because it is unethical to deny effective treatments that were assessed in former superiority trials. NI trials investigate whether the efficacy of an intervention is not any worse than that of an active comparator. The NI margin is used to assess whether an intervention will preserve a clinically significant fraction of the effect of the active comparator and assess the largest clinically acceptable difference between the intervention and the active comparator. It is critical to set the appropriate NI margin for an NI trial as the choice of the NI margin dictates the conclusions of the trial. Methods to estimate the margin are heterogeneous but generally based on clinical judgement and statistical reasoning and hence tailored to each clinical situation.

The reporting of methodological aspects of randomised controlled NI drug trials is poor with less than 50% of published trials reporting a justification of the NI margin.[1–3] If the rationale was reported it was often not based on historical evidence,[4] that would establish the foundation of the statistical reasoning. The introduction of the Consolidated Standards of Reporting Trials (CONSORT) extension on reporting of NI and equivalence randomised trials did not substantially improve reporting quality.[1 2 5 6] While the publication of guidance per se had

limited impact, Rehal and colleagues argued that the heterogeneity of guidelines is a contributing factor to the reporting deficit.[3] Aupais and colleagues described the limitation of NI trials in small cohorts due to large CIs that also depend on the method of defining the NI margin.[7] Other work showed that there is variation in the conclusion of NI trials when different methods of estimating the CI were used.[8]

For drug trials, recommendations on how to determine NI margins from agencies like the Food and Drug Administration or the European Medicines Agency exist. However, there is no explicit guidance for non-pharmacological trials and the reporting quality of non-pharmalogical trials conducted is unknown. Through a systematic review, we aim to examine this issue further and:

1. Describe the frequency and diversity of said NI margins defining methods.
2. Assess reporting and methodological quality that have been used to define NI margins in surgical randomised controlled trials.
3. Asses how robust the different techniques used to define NI margins are.

By robust we refer to the assumptions used in generating the margins and the associated variability. Ultimately, this variability might change the outcome direction and lead to different conclusions.

METHODS AND ANALYSIS

This article was prepared according to the Preferred Reporting Items for Systematic Review and Meta-Analysis (PRISMA) Protocols recommendations. The checklist can be found in the online supplemental file 1. For the systematic review results we will follow the PRISMA guideline. There is currently no guideline for this type of methodological research. The eligibility criteria are:

- Randomised controlled trials with an NI design.
- At least two arms of which at least one must be a surgical treatment and the trial should assess an abdominal surgical intervention. A surgical intervention is a surgical procedure (eg, appendectomy or cholecystectomy), or a surgical technique (eg, open or laparoscopic access).
- Any control or intervention will be included as long as at least one trial arm is a surgical intervention as defined above.
- Date of publication from 2006 (date of the publication of the CONSORT extension) until the date of the search.
- Sample size ≥100 participants.
- Adults (participants ≥18 years old).

The rationale for excluding small trials (n<100) is manifold. First, smaller trials rather identify if there is an effect than whether the effect is superior or non-inferior. Second, these trials often are of lower quality. Third, they generally have a low statistical power and therefore a reduced chance of detecting a true effect. At the same time, a low power minimises the chance that a statistically significant result reflects the true effect.

Altogether, the impact of such small trials on clinical practice may be insignificant and might also be disregarded due to the risk of a misleading interpretation of the trial findings. Also, we rely on a minimal reporting quality to enable simulation studies and the lower quality is also reflected by the poorer reporting quality.

Trials comparing medicines will not be included. Psychological or behavioural interventions and secondary reports of previously published trials will be excluded. There are no language restrictions and a primary results report must be published in or after 1 January 2006 until the date of the search. No secondary reports from randomised controlled trials will be included.

The following databases will be searched:

- MEDLINE.
- EMBASE.
- Cochrane Central Register of Controlled Trials (CENTRAL).

Data will be collected from result reports, protocols and statistical analysis plans. The search strategy can be found in the online supplemental file 2.

Study selection

Covidence will be used to import findings from all databases. Duplicates will be identified automatically but a sample of 25% of articles will be checked manually. Screening will be hierarchical in the following order: title and abstracts and full-text article. One reviewer (CK) will prescreen all articles and assess the automatically removed duplicates, trials with less than 100 participants will be removed. After this prescreening, all remaining articles will be screened independently by CK and IJG or RS. Any differences will be discussed and resolved by consultation with BL. Ultimately, full-text articles will be screened after exporting to EndNote, V.20 independently by two reviewers.

Data management

Summary tables will be used to display results and data will be displayed descriptively. The reporting quality will be compared with the CONSORT statement extension for NI trials, referring to the latter as the minimum requirement of reporting. The author conclusion will be balanced against the obtained results, especially, whether there is a discrepancy between the study results and conclusion (eg, some studies claim NI although their data does not support NI based on their design and analysis). An overview of all data points that will be extracted from publications can be found in the online supplemental file 3. A dedicated case report form will be used to extract all data by two independent reviewers in duplicate.

Outcomes

The main outcome will be a description of the methods for defining the NI margin. This outcome will include the reporting quality thereof based on estimated or previously

known effect sizes of the studies intervention or control, respectively. The quality is measured by assessing the effect size, CIs of the effect size and adherence to the CONSORT statement extension for NI trials.

A secondary outcome is the assessment of the robustness of the NI margin estimation. A simulation using alternative assumptions for the effect size and CIs of the effect size will be performed. Information about the rationale for using an NI design, number of patients per arm, number of events if applicable, mean or median including SD, IQR or range per arm, will be sought. The information used to estimate the NI margin will be collected from each report. If possible, further data from prior trials will be included in the simulation studies.

For the simulation analysis the method to define the NI margin will be appraised by changing the model assumptions and if possible additional previously available information from superiority trials will be included. In detail, the historical effect size of the active comparator will be reappraised based on the available literature. The literature will be identified through a search of the original manuscript because the purpose of this simulation is to show how the results may differ if the assumptions were changed. Furthermore, Ovid MEDLINE will be searched for meta-analyses and systematic reviews to facilitate the identification of previous randomised controlled trials and pooled effect estimates of the historical evidence. No attempt will be made to systematically identify all the historical evidence for this part of the analysis.

Alternative methods to calculate the NI margin will be carried out as described by Althunian and colleagues.[4] The preserved fraction will not exceed 50% and as the point estimate of the primary endpoint, either the relative or absolute difference will be chosen. The results of the simulation will be compared with the reported NI margin and both will be compared with the authors' conclusions.

The alternative methods used to calculate the CI for proportions will be Wald, Wald continuity corrected, Wald interval with an adjustment according to Agresti and Caffo,[9] Miettinen and Nurminen[10] and Newcombe or Wilson score.[11] The Wald interval is the most basic one and performs well if the proportions, for example, patients with an event, are close to 0.5 because it assumes a normal sampling distribution. If not, the coverage of the CI is poor. The Wilson score interval improves the poor coverage of the Wald interval.[12] The Agresti Caffo adds two to the number of events and the sample size and also reaches a better coverage than the Wald method. The Miettinen and Nurminen interval is based on a restricted maximum likelihood estimation and the Newcombe interval is a more accurate approximation to the binomial distribution especially near the boundaries of 0 and 1.

All analyses will be performed using R Statistical Software (V.4.2.3; R Core Team 2023). Risk of bias in the individual trials will be assessed using the Cochrane Risk of Bias (RoB) 2 tool.[13] As this search does not cover a specific disease area, publication bias will not be assessed.

Altogether, our findings will inform the design of surgical NI trials and help reduce misinterpretation of study findings. They will also support the initiative to improve NI trial design by institutional guidance similar to that for drug trials. The assessment of established methods of defining NI and the robustness of these methods should aid researchers and clinicians in defining the NI margins for randomised controlled trials in healthcare to identify the most appropriate research design.

ETHICS AND DISSEMINATION

The results of this systematic review will be submitted for peer-reviewed journal publication. Findings will be presented at conferences with a mainly surgical target audience. Ethical approval is not required and any changes to the protocol will be communicated via the Open Science Framework (OSF) platform (https://osf.io/3gx4b/). Post hoc analyses will be referred to as such in the result report. Any additional publications will be tagged as secondary analyses.

Patient and public involvement

There was no patient or public involvement in the development of the research question, planning, design or conduct of this research.

X Christoph Kuemmerli @ChristophKuemme

Acknowledgements We thank Dr Marshall Dozier, Lead for Library Academic Support, who helped develop the search strategy and identify the information sources.

Contributors CK developed the idea and the concept, contributed to the methodology, developed and tested the search strategy, will conduct the search, screen articles and extract data and drafted the manuscript. IJG refined concept and methods for the review, will screen articles and extract data and reviewed the manuscript. RS developed the methodology, will screen articles and extract data and reviewed the manuscript. BL shaped the idea and the concept, developed methods, refined the search strategy, reviewed the manuscript, supervised the work and is the guarantor.

Funding The review was sponsored by the University of Edinburgh. The sponsor had no role in developing the protocol.

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

ORCID iD
Christoph Kuemmerli http://orcid.org/0000-0002-7109-3545

REFERENCES

1 Wangge G, Klungel OH, Roes KCB, *et al*. Room for improvement in conducting and reporting non-inferiority randomized controlled trials on drugs: a systematic review. *PLoS One* 2010;5:e13550.
2 Schiller P, Burchardi N, Niestroj M, *et al*. Quality of reporting of clinical non-inferiority and equivalence randomised trials--update and extension. *Trials* 2012;13:214.
3 Rehal S, Morris TP, Fielding K, *et al*. Non-inferiority trials: are they inferior? A systematic review of reporting in major medical journals. *BMJ Open* 2016;6:e012594.
4 Althunian TA, de Boer A, Klungel OH, *et al*. Methods of defining the non-inferiority margin in randomized, double-blind controlled trials: a systematic review. *Trials* 2017;18:107.
5 Piaggio G, Elbourne DR, Altman DG, *et al*. Reporting of noninferiority and equivalence randomized trials: an extension of the CONSORT statement. *JAMA* 2006;295:1152–60.
6 Lo CKL, Komorowski AS, Hall CW, *et al*. Methodological and reporting quality of noninferiority randomized controlled trials comparing antiretroviral therapies: a systematic review. *Clin Infect Dis* 2023;77:1023–31.
7 Aupiais C, Zohar S, Taverny G, *et al*. Exploring how non-inferiority and equivalence are assessed in paediatrics: a systematic review. *Arch Dis Child* 2018;103:1067–75.
8 Bai AD, Komorowski AS, Lo CKL, *et al*. Confidence interval of risk difference by different statistical methods and its impact on the study conclusion in antibiotic non-inferiority trials. *Trials* 2021;22:708.
9 Agresti A, Caffo B. Simple and effective confidence intervals for proportions and differences of proportions result from adding two successes and two failures. *Am Stat* 2000;54:280.
10 Miettinen O, Nurminen M. Comparative analysis of two rates. *Stat Med* 1985;4:213–26.
11 Newcombe RG. Interval estimation for the difference between independent proportions: comparison of eleven methods. *Stat Med* 1998;873–90.
12 Wilson EB. Probable Inference, the Law of Succession, and Statistical Inference. *J Am Stat Assoc* 1927;22:209–12.
13 Sterne JAC, Savović J, Page MJ, *et al*. RoB 2: a revised tool for assessing risk of bias in randomised trials. *BMJ* 2019;366:l4898.