Journal of Clinical Epidemiology 110 (2019) 82−89

Journal of Clinical Epidemiology

ORIGINAL ARTICLE

Most noninferiority trials were not designed to preserve active comparator treatment effects

Michael Tsui[a], Sunita Rehal[b], Vipul Jairath[c,d,1], Brennan C. Kahan[e,*,1]

[a]*Schulich School of Medicine and Dentistry, 1151 Richmond St, London, Ontario N6A 5C1, Canada*
[b]*MRC Clinical Trials Unit at UCL, Institute of Clinical Trials and Methodology, 90 High Holborn, London WC1V 6LJ, UK*
[c]*Department of Medicine, University of Western Ontario, 1151 Richmond St, London, Ontario N6A 5C1, Canada*
[d]*Department of Epidemiology and Biostatistics, University of Western Ontario, 1151 Richmond St, London, Ontario N6A 5C1, Canada*
[e]*Pragmatic Clinical Trials Unit, Queen Mary University of London, London, E1 2AB UK*

Accepted 1 March 2019; Published online 8 March 2019

Abstract

Objectives: To evaluate whether noninferiority trials are designed to adequately preserve the historical treatment effect of their active comparators.

Study Design and Setting: We reviewed 162 noninferiority trials published in high-impact medical journals. We assessed whether trials were designed to ensure that interventions could only be declared noninferior if they preserved at least 50% of the active comparator's historical treatment effect.

Results: Only 25 of 162 trials (15%) were designed so that interventions could only be declared noninferior if they preserved at least 50% of the active comparator's historical treatment effect. Most trials did not provide evidence that the active comparator was effective ($n = 101$), provided inadequate evidence ($n = 18$), or used a noninferiority margin that was too wide ($n = 18$). In a subset of 61 noninferiority trials which referenced a prior randomized trial or meta-analysis evaluating the active comparator, only 25 (41%) used a noninferiority margin small enough to preserve at least 50% of the active comparator's treatment effect. Overall, 14 of 162 noninferiority trials (9%) would have allowed the intervention to be declared noninferior even if it was worse than either placebo or another historical control.

Conclusion: Most noninferiority trials published in major medical journals could allow erroneous declarations of noninferiority.

Keywords: Noninferiority trials; Equivalence trial; Randomized controlled trial; Clinical trial

1. Introduction

Most randomized controlled trials (RCTs) are designed to demonstrate superiority of an intervention, whereas noninferiority trials are designed to demonstrate that an intervention is not worse than a standard treatment (referred to as the ''active comparator'') by more than a specific margin [1−11]. These latter designs are often used to assess new interventions that may not be more effective than existing treatments but may offer other advantages such as an improved safety profile, ease of administration, or reduced costs [2,4,6,8,12,13].

Interventions found to be noninferior are assumed to be almost as effective as their active comparators, and therefore acceptable for clinical use. However, the validity of this assumption relies on two key criteria in the design of these trials [1−3,6,7,14−20]. First, that the superiority of the active comparator to placebo or other historical control has been previously established in a randomized trial or meta-analysis, and second, that the noninferiority margin (the maximum acceptable threshold by which the new intervention can be worse than the active comparator and nevertheless considered noninferior) is narrow enough to preserve a sufficient amount of the active comparator's treatment effect.

Violation of either of these two critical aspects of the noninferiority trial design can compromise the validity of

Funding: Michael Tsui is supported by the Schulich Medicine & Dentistry's, Canada Summer Research Training Program. No other authors received funding for this research.

Conflict of interest statement: Sunita Rehal took a job at Roche after this work was completed, and now is a shareholder of F. Hoffmann-La Roche Ltd. All other authors declare they have no conflicts of interest.

[1] Contributed equally.

* Corresponding author: Pragmatic Clinical Trials Unit, Queen Mary University of London, 58 Turner St, London, UK.

E-mail address: b.kahan@qmul.ac.uk (B.C. Kahan).

https://doi.org/10.1016/j.jclinepi.2019.03.003

What is new?

Key findings

- In this review of 162 noninferiority trials published in high-impact medical journals, only 25 (15%) provided definitive evidence that interventions could only be declared noninferior if they were preserved at least 50% of the active comparator's historical treatment effect. The remaining trials may have allowed erroneous declarations of noninferiority.

What this adds to what was known?

- Most trials (101/162, 62%) did not provide evidence that the active comparator was effective. Of the 61 trials which referenced a previous study in support of the active comparator, only 70% (43/61) showed a statistically significant benefit of the active comparator.
- Fourteen trials (9%) were designed so the intervention could be declared noninferior despite being less effective than either placebo or another historical control.

What is the implication and what should change now?

- Noninferiority trials require improved design to ensure new interventions are sufficiently effective.

the trial and its conclusions [2,5,6,21,22]. First, it is meaningless to declare noninferiority of a new intervention when no prior evidence of effectiveness for its active comparator exists. In this case there can be no certainty whether the new intervention is at all effective. Second, an inappropriately wide noninferiority margin may allow a conclusion of noninferiority despite a worse performance of the new intervention compared to historical control. Both scenarios have the potential for the erroneous adoption of ineffective interventions into routine clinical practice. The aim of this study was to assess whether noninferiority trials published in major medical journals were designed to adequately preserve historical treatment effects through the use of appropriate active comparators and noninferiority margins.

2. Methods

2.1. Review of noninferiority trials

We utilized noninferiority trials previously described in a systematic review that addressed the adequacy of reporting of these trials [23]. Briefly, articles were initially identified in a MEDLINE search of titles and abstracts published in high-impact factor (>10) medical journals (general or internal medicine) between January 2010 and May 2015, using the terms "non-inferior," "noninferior," "non-inferiority," and "noninferiority". Articles were eligible for inclusion if they reported results of a randomized noninferiority trial and excluded if they were systematic reviews, meta-analyses, or commentaries, if the primary analysis was not for noninferiority, or if the trial used Bayesian methods. During data extraction and analysis, we made a post hoc decision to exclude six trials which included a placebo arm, as the preservation of historical treatment effects was less relevant in these studies because the new intervention could be directly compared to placebo.

For the present study, we extracted data related to the primary noninferiority comparison for each trial into standardized, prepiloted forms. We also examined published protocols and supplementary appendices when these were referred to in the main text of the article. We used the following strategy to identify a single primary noninferiority comparison for data extraction when the noninferiority trial under review had specified multiple primary comparisons: (1) if the noninferiority trial had performed sample size calculation for only one comparison, we used this comparison; (2) if the noninferiority trial performed sample size calculations for multiple comparisons, or if no sample size calculations were performed, we used the first comparison listed in the abstract.

One author extracted data from all trials. To assess agreement among authors, a second author independently extracted data from a random sample of 17 reports (10% of all trials). Agreement on data extraction was 98% for the sample. Any discrepancies in the random sample were resolved through discussion.

2.2. Review of articles demonstrating effectiveness of the active comparator

We assessed each noninferiority trial for prior evidence for the effectiveness of the active comparator (RCTs or meta-analyses of RCTs, hereafter referred to as "active comparator studies") by reviewing the reference list and trial protocol if available. Data from active comparator studies were extracted onto a second standardized form. We only considered active comparator studies which included an outcome which was similar to the primary noninferiority endpoint. The outcomes we used for both the noninferiority trials and their corresponding active comparator study are available in Online Appendix 2. We extracted the size of the treatment effect of the active comparator vs. placebo or other historical control, alongside its confidence interval and *P*-value, if available. These were extracted for the outcome that was most similar to the primary noninferiority endpoint.

When noninferiority trials had referenced multiple studies as evidence for the effectiveness of the active

comparator, we used a prespecified algorithm to select a single active comparator study for data extraction. We opted to select a single RCT or meta-analysis because combining results from multiple active comparator studies into a single estimate of effectiveness could have been problematic because of differences between studies in control groups or outcome definitions. The algorithm specified a set of hierarchical criteria on which to select studies. Studies were compared sequentially for each criterion until one study was selected. The criteria, in order of rank, were (1) similarity of the active comparator to that used in the noninferiority trial (e.g., based on dose); (2) placebo-controlled studies were preferred over studies which used a historical treatment for comparison; (3) similarity of the outcome to that used in the noninferiority trial (e.g., based on timepoint of the outcome); (4) meta-analyses of RCTs were preferred over single RCTs; and (5) more recent studies were preferred to older studies.

2.3. Estimating the percentage of preserved effect of the active comparator

In the subset of noninferiority trials which referenced an active comparator study, we assessed the minimum effect of the active comparator that the new intervention could preserve while still being declared noninferior (Fig. 1); we refer to this as the ''percentage of preserved effect'' (%PE). The %PE is based on the size of the noninferiority margin (*NI margin*) and the estimated treatment effect of the active comparator vs. placebo or another historical control (*active comparator effect*). For absolute differences (i.e., means, percentages), the %PE is calculated as *%PE* = (*active comparator effect* + *NI margin*)/*active comparator effect* (where the active comparator effect and the NI margin are in different directions), and for relative differences (i.e., hazard, risk, or odds ratios), it is calculated as ln(*active comparator effect* + *NI margin*)/ln(*active comparator effect*), where ''ln'' denotes the natural logarithm [3]. For example, imagine a noninferiority trial designed to assess a new intervention for prevention of breast cancer recurrence, where a previous RCT showed the active comparator increased 5-year survival by 10 percentage points vs. placebo. If the noninferiority margin for the new intervention is an absolute difference of −2.5 percentage points (i.e., the lower bound of the confidence interval must be greater than −2.5 percentage points for the new intervention to be declared noninferior), then %PE = (10 + [−2.5])/10 = 75%, meaning the new intervention must retain at least 75% of the active comparator's effectiveness compared to placebo to be considered noninferior. However, if the noninferiority margin for the trial is −12.5 percentage points, the %PE = −25%, meaning the new intervention could be worse than placebo and still be declared noninferior.

A %PE value of 100% indicates a new intervention would need to completely retain the effect of the active comparator to be declared noninferior, whereas a value of 0% indicates a new intervention would not need to retain

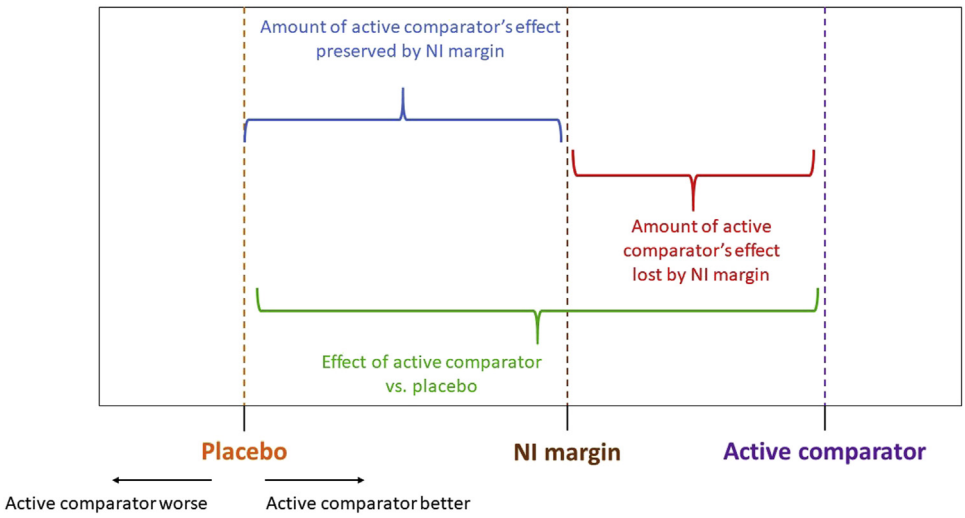

Fig. 1. Preservation of the active comparator's treatment effect through the size of the noninferiority margin. The amount by which the active comparator's treatment effect is preserved through the size of the noninferiority margin (%PE) is calculated as: *%PE* = (*active comparator effect* − *NI margin*)/*active comparator effect*. The distance between the placebo (orange line) and the active comparator (purple line) represents the *active comparator effect* (effect of the active comparator vs. placebo), and the distance between the active comparator (purple line) and the noninferiority (NI) margin (brown line) is the maximum amount of the active comparator's effect that can be lost when the intervention is declared noninferior. The distance between the placebo (orange line) and the NI margin (brown line) is the minimum amount of the active comparator's effect that is preserved when the intervention is declared noninferior. For example, if the active comparator was 10 percentage points better than placebo, and the noninferiority margin was −2.5 percentage points, the %PE would be %PE = (10−2.5)/10 = 75%, meaning the new intervention must retain at least 75% of the active comparator's effectiveness compared to placebo to be considered noninferior. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

any effect of the active comparator to conclude that it is noninferior. A negative %PE indicates that the noninferiority margin is so wide that the new intervention could be declared noninferior despite being worse than the historical control.

When active comparator studies estimate the active comparator to be worse than the historical treatment, the %PE cannot be calculated. In these cases, no matter the size of the NI margin, the new intervention could be declared noninferior, while being worse than the historical treatment.

2.4. Outcome measures

Our main outcome measure was the number of trials that were designed to ensure that interventions could only be declared noninferior if they were at least 50% as effective as the active comparator (i.e., preserve at least 50% of the active comparator's treatment effect). Although the amount of treatment effect that should be preserved depends on the clinical context of each new intervention, our choice of 50% reflects FDA's recommendation for the minimum amount of active comparator's effect to be preserved [24]. We classified noninferiority trials as meeting this outcome if (1) they referenced a previous randomized trial or meta-analysis of randomized trials which evaluated the effectiveness of the active comparator; (2) this previous study found the active comparator to be significantly better than placebo or another historical treatment; and (3) the specified noninferiority margin was sufficiently narrow to preserve at least 50% of the active comparator's treatment effect from the previous study (i.e., %PE $\geq$ 50%).

A secondary outcome measure was the number of trials which were designed to allow new interventions to be declared noninferior even if they were less effective than placebo or another historical control. Noninferiority trials were classified in this way if (1) the active comparator was estimated to be worse than placebo or historical control in a previous active comparator study, regardless of whether this difference was statistically significant; or (2) the noninferiority margin was so wide that none of the active comparator's treatment effect would be preserved (i.e., %PE $<$ 0).

2.5. Sensitivity analyses for outcome measures

We conducted three sensitivity analyses for the two outcomes listed previously. In the first sensitivity analysis, we calculated the %PE using the lower confidence interval limit of the active comparator effect, as is the recommended approach for determining the noninferiority margin by regulatory authorities such as the FDA [24]. The use of the lower confidence interval limit provides a more conservative estimate of the benefit the active comparator and minimizes the likelihood that an ineffective intervention is declared noninferior.

In the second sensitivity analysis, we restricted the analysis to noninferiority trials which referenced an active comparator study. This analysis was performed to investigate whether our results were driven by poor reporting of previous active comparator studies by noninferiority trials.

Finally, in the third sensitivity analysis, we restricted the analysis to noninferiority trials, which referenced a single active comparator study to evaluate whether results may have been influenced by our algorithm, which selected a single active comparator study when multiple studies were referenced.

2.6. Exploratory analysis of association between justification of noninferiority margin and %PE

As an exploratory analysis, we assessed the association between justifications for the chosen noninferiority margin and the %PE. This was restricted to the subset of trials for which the %PE could be calculated.

3. Results

3.1. Characteristics of included studies

The original MEDLINE search identified 168 eligible noninferiority trials. We excluded a further six trials which included a placebo group, leaving 162 trials included in this review. A list of trials included is available in Online Appendix 1, and characteristics of included trials are available in Online Appendix 2.

Of 162 included noninferiority trials, 61 referred to at least one active comparator study. Forty-three trials referenced a single active comparator study, and 18 trials reference multiple studies. For these 18 trials, we used our prespecified algorithm to select a single study for analysis. The reasons for selection were similarity of the active comparator (n = 3), placebo-controlled study (n = 2), similarity of the outcome (n = 5), meta-analysis (n = 6), and most recent study (n = 2).

Of the 61 selected active comparator studies, 40 were RCTs (17 placebo control, 23 historical control) and 21 were meta-analyses of RCTs (12 placebo control, nine historical control).

3.2. Outcome measures

Of the 162 noninferiority trials included, only 25 (15%) were designed so that interventions could only be declared noninferior if they preserved at least 50% of the active comparator's historical treatment effect (main outcome) (Fig. 2). Overall, 14 trials (9%) were designed so the intervention could be declared noninferior even if it was less effective than either placebo or another historical control (secondary outcome). This was because the active comparator was estimated to be worse than placebo or historical control (n = 5) or the %PE was $<$0% (n = 9).

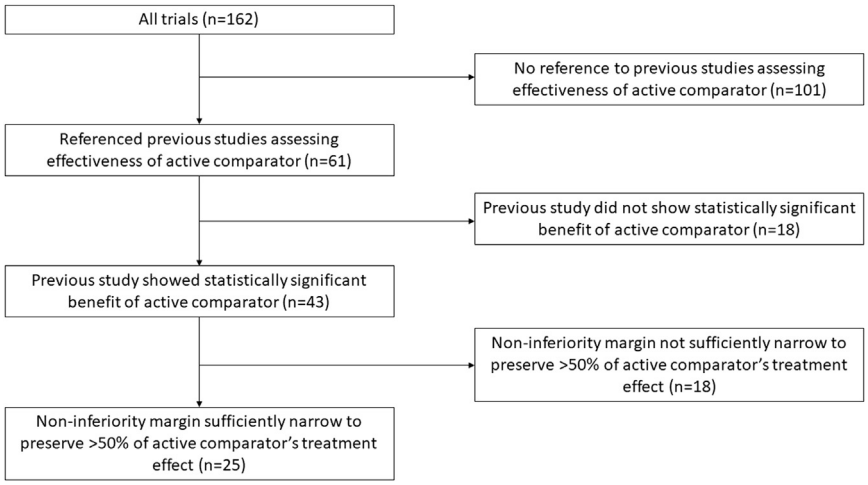

Fig. 2. Trials designed to preserve at least 50% of the active comparator's treatment effect (main outcome measure).

Of note, out of 61 selected active comparator studies, only 43 (70%) showed a statistically significant benefit of the active comparator over placebo or historical control, one study (2%) showed the active comparator was significantly worse, 9 (15%) found no statistical difference, and 8 (13%) did not include sufficient information to assess this.

3.3. Sensitivity analyses for outcome measures

In our first sensitivity analysis that calculated %PE using the confidence interval limit toward the null effect for the active comparator, we excluded 14 trials that did not include sufficient information to calculate a confidence interval, leaving 148 trials for analysis. Of these, 8 (5%) were designed so that interventions could only be declared noninferior if they were at least 50% as effective as the active comparator. Twenty-nine trials (20%) were designed so the intervention could be declared noninferior even if it was less effective than either placebo or another historical control.

Second, restricting the analysis to the set of 61 noninferiority trials, which referenced an active comparator study, 25 (41%) were designed so that interventions could only be declared noninferior if they were at least 50% as effective as the active comparator, and 14 (23%) would have allowed interventions to be declared noninferior even if they were less effective than placebo or another historical control.

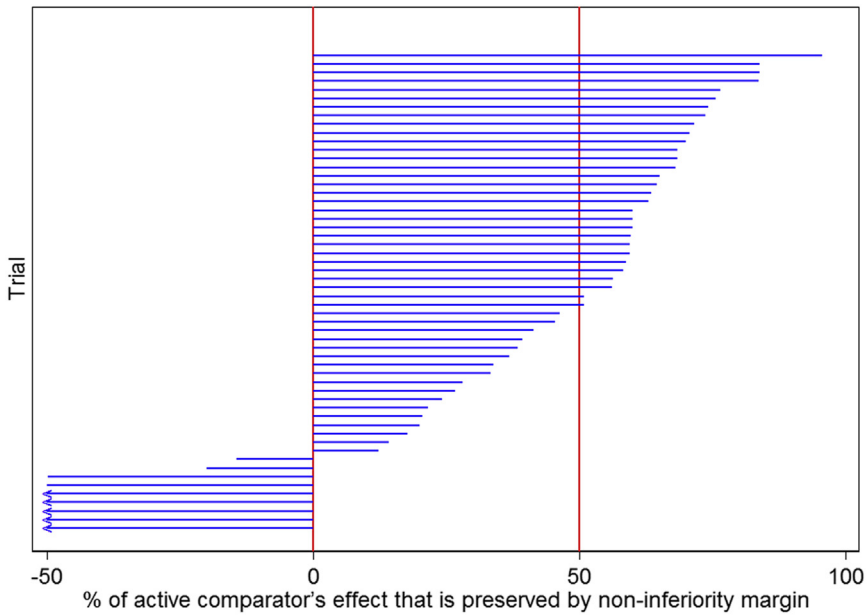

Fig. 3. Percentage of effect of active comparator vs. historical control that is preserved by noninferiority margin. Values less than −50% were censored to −50%, marked with "<".

Table 1. Association between justification for noninferiority margin and percentage of the active comparator's treatment effect, which is preserved by the noninferiority margin

Justification	Total (n = 162)	Evaluable %PE	Median %PE (IQR)
No justification provided	55 (34)	18	31% (−20% to 58%)
Author's judgment only	28 (17)	10	38% (17% to 68%)
Judgment of independent experts	13 (8)	1	21% (NA)
Based on effect size from active comparator study	22 (14)	19	60% (46% to 71%)
Statistical or clinical justification based on other sources	20 (12)	0	NA (NA)
Margin recommended by disease-specific guidelines	13 (8)	4	54% (−280% to 70%)
Other[a]	11 (7)	4	43% (−12% to 62%)

Abbreviations: %PE, % preserved effect; IQR, interquartile range.
Interquartile ranges were calculated for categories with four or more trials.
[a] Other justifications include statistical reasoning of authors not supported by references, margins used in previous similar noninferiority trials, feasible sample size calculations, and special protocol authorization from regulatory agency.

Finally, further restricting the analysis to the 43 noninferiority trials, which referenced a single active comparator effect study, these numbers were 17 (40%) and 10 (23%).

3.4. Exploratory analysis of association between justification of noninferiority margin and %PE

Fifty-six noninferiority trials had a calculable %PE and were included in this analysis. The median %PE among in these trials was 54% (IQR 21% to 67%) (Fig. 3). In nine trials (16%) the %PE was <0; the %PE was only >75% in six trials (11%).

The median %PE was 31% (IQR −20% to 58%) for trials that offered no justification for their margin and 38% (17% to 68%) for trials that relied on authors' clinical judgment only (Table 1). Conversely, trials that based their margin on an active comparator study had a median %PE of 60% (46% to 71%).

4. Conclusions

We found that few noninferiority trials (15%) were designed so that interventions could only be declared noninferior if they preserved at least 50% of the active comparator's historical treatment effect. Furthermore, 9% of trials were designed so interventions could be declared noninferior even if they were less effective than either placebo or another historical control. A common assumption of noninferiority trials is that interventions found to be noninferior are effective and acceptable use clinical use. However, our results cast doubt on this assumption; a finding of noninferiority would not indicate the intervention is effective in most cases. These results were driven by the use of active comparators for which effectiveness had not been demonstrated and the use of inappropriately wide noninferiority margins.

The use of active comparators, which have not been shown to be effective in previous RCTs complicates a noninferiority finding in most instances, as there is no certainty whether the new intervention is at all effective; both treatments may be equally ineffective. Some trials may have not referenced active comparator studies even if they existed, and so we may have overestimated the extent of this issue. In these instances, it is difficult to disentangle poor methodological design from incomplete reporting. This demonstrates the importance of good reporting [25,26]. However, we found that only 41% of trials, which reported an active comparator study, were designed to preserve at least 50% of its treatment effect, indicating that incomplete reporting may only be a small part of the problem.

In some settings, the standard of care may not have been shown to be effective in previous RCTs. Investigators may feel that using a different control arm to the standard of care is unethical, but that a noninferiority design is still appropriate because of other benefits of the intervention, such as reduced costs or ease of administration. There is little consensus as to whether noninferiority is the most appropriate design in this setting; however, if it is used then investigators should clearly report that the active control has not been previously shown to be effective, and that a noninferiority finding does not necessarily indicate effectiveness.

Many trials used a noninferiority margin that was too wide to ensure preservation of a sufficient proportion of the active comparator's effectiveness. For some trials, the noninferiority margin was wide enough to allow the new intervention to be declared noninferior even if it was no better than placebo or another historical control. The use of inappropriately wide margins seems to be driven in part by the way that noninferiority margins are selected. We found that over half of all trials offered no justification for their margins or relied solely on the authors' own judgment, the latter of which were often vague and did not

clarify why the chosen margin was clinically meaningful. Furthermore, few trials that referenced an active comparator study chose their margin based on this effect.

There were several limitations to our study. First, it is difficult to distinguish to what extent our results were driven by incomplete reporting of active comparator studies rather than by poor methodological design. Had previous active comparator studies been more completely reported, our results would likely have shown a higher proportion of trials designed to preserve the active comparator's treatment effect. We note, however, that even among trials reporting previous active comparator studies, most were not designed to preserve the active comparator's treatment effect. Second, the prespecified algorithm we used to select a single active comparator study for analysis was designed based on the factors we felt were most important (e.g., similarity of active comparators), which is subjective. Had the algorithm been designed differently, our results may have differed. However, the algorithm was only necessary in a small number of trials ($n = 18$), and our third sensitivity analysis that limited our sample to trials with only a single active comparator study showed almost identical results. Finally, the cutoff we used to define the adequate preservation of the active comparator's treatment effect ($>50\%$) was subjective. In some settings, preserving less than 50% of the treatment effect may be clinically acceptable, whereas in others it may not be sufficient.

In conclusion, we found the design of most noninferiority trials in this review could allow erroneous declarations of noninferiority. This was driven by use of active comparators for which effectiveness had not been demonstrated and selection of inappropriately wide noninferiority margins.

CRediT authorship contribution statement

Michael Tsui: Methodology, Formal analysis, Investigation, Writing - review & editing, Writing - original draft. **Sunita Rehal:** Investigation, Writing - review & editing. **Vipul Jairath:** Conceptualization, Methodology, Writing - review & editing. **Brennan C. Kahan:** Conceptualization, Methodology, Investigation, Writing - review & editing, Formal analysis.

Acknowledgements

Writing assistance was provided by Lisa Shackelton PhD.

Authors' contributions: M.T. contributed to the study design, data collection, data analysis, article drafting, and editing. S.R. contributed to data collection and article editing. V.J. contributed to study conception, design, and article editing. B.K. contributed to study conception, design, data collection, data analysis, and article editing. B.K. is the guarantor of the article.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jclinepi.2019.03.003.

References

[1] James Hung HM, Wang S-J, Tsong Y, Lawrence J, O'Neil RT. Some fundamental issues with non-inferiority testing in active controlled trials. Stat Med 2003;22:213−25.
[2] Mauri L, D'Agostino RB. Challenges in the design and interpretation of noninferiority trials. N Engl J Med 2017;377:1357−67.
[3] Kaul S, Diamond GA. Good enough: a primer on the analysis and interpretation of noninferiority trials. Ann Intern Med 2006;145:62−9.
[4] Schumi J, Wittes JT. Through the looking glass: understanding non-inferiority. Trials 2011;12:106.
[5] Head SJ, Kaul S, Bogers AJJC, Kappetein AP. Non-inferiority study design: lessons to be learned from cardiovascular trials. Eur Heart J 2012;33:1318−24.
[6] Macaya F, Ryan N, Salinas P, Pocock SJ. Challenges in the design and interpretation of noninferiority trials: insights from recent stent trials. J Am Coll Cardiol 2017;70:894−903.
[7] Weintraub WS. Cutting through the statistical fog: understanding and evaluating non-inferiority trials. Int J Clin Pract 2010;64:1359−66.
[8] Kaji AH, Lewis RJ. Noninferiority trials: is a new treatment almost as effective as another? JAMA 2015;313:2371−2.
[9] Mulla SM, Scott IA, Jackevicius CA, You JJ, Guyatt GH. How to use a noninferiority trial: users' guides to the medical literature. JAMA 2012;308:2605−11.
[10] Althunian TA, Boer A de, Groenwold RHH, Klungel OH. Using a single noninferiority margin or preserved fraction for an entire pharmacological class was found to be inappropriate. J Clin Epidemiol 2018;104:15−23.
[11] Abulizi X, Flandre P. Choice of treatment-effect measures when non-inferiority margins originally defined in absolute difference translated into relative difference influenced the results of clinical trials. J Clin Epidemiol 2018;96:63−72.
[12] Hills RK. Non-inferiority trials: No better? No worse? No change? No pain? Br J Haematol 2017;176(6):883−7.
[13] Riechelmann RP, Alex A, Cruz L, Bariani GM, Hoff PM. Non-inferiority cancer clinical trials: scope and purposes underlying their design. Ann Oncol 2013;24(7):1942−7.
[14] D'Agostino RB, Massaro JM, Sullivan LM. Non-inferiority trials: design concepts and issues - the encounters of academic consultants in statistics. Stat Med 2003;22:169−86.
[15] Henanff AL, Giraudeau B, Baron G, Ravaud P. Quality of reporting of noninferiority and equivalence randomized trials. JAMA 2006;295:1147−51.
[16] European Medicines Agency, Committee for Medicinal Products for Human Use. Guideline on the choice of the non-inferiority margin. 2005. Available at http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003636.pdf; Accessed April 16, 2018.
[17] Hernandez AV, Pasupuleti V, Deshpande A, Thota P, Collins JA, Vidal JE. Deficient reporting and interpretation of non-inferiority randomized clinical trials in HIV patients: a systematic review. PLoS One 2013;8:e63272.
[18] Althunian TA, de Boer A, Klungel OH, Insani WN, Groenwold RHH. Methods of defining the non-inferiority margin in randomized, double-blind controlled trials: a systematic review. Trials 2017;18:107.
[19] Wangge G, Klungel OH, Roes KCB, Boer A de, Hoes AW, Knol MJ. Room for improvement in conducting and reporting non-inferiority randomized controlled trials on drugs: a systematic review. PLoS One 2010;5:e13550.

[20] Donken R, de Melker HE, Rots NY, Berbers G, Knol MJ. Comparing vaccines: a systematic review of the use of the non-inferiority margin in vaccine trials. Vaccine 2015;33(12):1426−32.

[21] Hou Y, Wu X, Li K. Issues on the selection of non-inferiority margin in clinical trials. Chin Med J (Engl) 2009;122:466−70.

[22] Lange S, Freitag G. Special invited papers section: therapeutic equivalence − clinical issues and statistical methodology in noninferiority trials. Biom J 2005;47(1):12−27.

[23] 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(10): e012594.

[24] U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research. Non-inferiority clinical trials to establish effectiveness: guidance for industry. 2016. Available at https://www.fda.gov/downloads/Drugs/Guidances/UCM202140.pdf; Accessed April 16, 2018.

[25] Piaggio G, Elbourne DR, Pocock SJ, Evans SJW, Altman DG. Group for the C. Reporting of noninferiority and equivalence randomized trials: extension of the CONSORT 2010 statement. JAMA 2012;308:2594−604.

[26] 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.