Journal of Clinical Epidemiology 177 (2025) 111613

Journal of Clinical Epidemiology

ORIGINAL RESEARCH

Some superiority trials with nonsignificant results published in high impact factor journals correspond to noninferiority situations: a research-on-research study

Deivanes Rajendrabose, Lucie Collet, Camille Reinaud, Maxime Beydon, Xiaojun Jiang, Sahra Hmissi, Antonin Vermillac, Thomas Degonzague, David Hajage, Agnès Dechartres*

Sorbonne Université, INSERM, Institut Pierre Louis d'Epidémiologie et de Santé Publique, AP-HP, Hôpital Pitié-Salpêtrière, Département de Santé Publique, Centre de Pharmacoépidémiologie de l'AP-HP (Cephepi), Paris, France

Accepted 12 November 2024; Published online 16 November 2024

Abstract

Objectives: Many negative randomized controlled trials (RCTs) report spin in their conclusions to highlight the benefits of the experimental arm, which could correspond to a noninferiority (NI) objective. We aimed to evaluate whether some negative superiority RCTs comparing 2 active interventions could correspond to an NI situation and to explore associated trial characteristics.

Study Design and Setting: We searched PubMed for superiority RCTs comparing 2 active interventions with non−statistically significant results for the primary outcome that were published in 2021 in the 5 journals with the highest impact factor in each medical specialty. Three reviewers independently evaluated whether trials could correspond to an NI situation (ie, an evaluation of efficacy as the primary outcome, with the experimental intervention presenting advantages including better safety profile, ease of administration, or decreased cost as compared with the control intervention).

Results: Of the 147 trials included, 19 (12.9%, 95% CI [7.9%, 19.4%]) corresponded to a potential NI situation. As compared with trials not in a potential NI situation, they were published in a journal with a lower impact factor (median impact factor 8.7 vs 15.6), were more frequently rated at high or some concerns regarding risk of bias (n = 14, 73.7% vs n = 69, 53.9%) and reported spin in the article conclusions (n = 11, 57.9% vs n = 24, 18.8%).

Conclusion: A non-negligible proportion of superiority negative trials comparing 2 active interventions could correspond to an NI situation. These trials seemed at increased risk of bias and frequently reported spin in the conclusions, which may distort the interpretation of results.

Keywords: Noninferiority; Randomized controlled trial; Trial design; Spin; Risk of bias; Nonsignificance

Plain Language Summary

Noninferiority trials are designed to show that a new intervention is not worse in terms of efficacy than the reference intervention. It is adapted when the new intervention has an advantage in terms of safety, ease of use or cost over the reference one. However, the literature displayed some superiority negative trials comparing 2 active interventions that could correspond to a potential noninferiority situation. Our study aimed to assess whether some superiority trials with nonsignificant results for the primary outcome could correspond to an NI situation and to explore associated trial characteristics. Our findings indicate that a non-negligible proportion of superiority negative trials could correspond to a noninferiority situation. Moreover, those trials seemed at increased risk of bias and frequently reported spin in the conclusions.

* Corresponding author. Sorbonne Université, INSERM U1136, équipe PEPITES, Institut Pierre Louis d'épidémiologie et de santé publique, Hôpital Pitié-Salpêtrière, 47-83 Boulevard de l'Hôpital, 75013, Paris, France.

E-mail address: agnes.dechartres@aphp.fr (A. Dechartres).

https://doi.org/10.1016/j.jclinepi.2024.111613

What is new?

- Twelve-point-nine percent of negative trials comparing 2 active interventions could have had a noninferiority design.
- Negative trials that corresponded to a noninferiority situation were frequently rated at high or some concerns regarding risk of bias and reported spin in the article conclusions.
- The choice of the trial design should be better justified.

1. Introduction

Noninferiority (NI) trials are designed to show that a new intervention is not worse in terms of efficacy than the reference intervention [1]. This type of design is adapted when the experimental intervention has advantages in terms of safety, ease of administration or reduced cost as compared with the reference intervention [2]. The NI margin is a key element of the NI design and corresponds to the loss of efficacy that can be accepted given the advantages of the experimental intervention [3]. Ideally, the margin should be small to guarantee the clinical relevance of the new intervention effect, but the smaller the margin the larger the sample size needed to show NI, which may limit the feasibility of the trial.

NI trials have increased in frequency: the number has been multiplied by a factor of 6 between 2005 and 2015, reaching 600 trials listed under the rubric ''noninferiority'' in MEDLINE in 2015 [4]. However, this number is small as compared with the overall number of randomized controlled trials (RCTs) published in 2015, representing about 11,000 publications [5], even though many interventions have been developed and evaluated to limit side effects or simplify modes of administration. In the literature, there are examples of negative superiority trials comparing 2 active interventions (ie, finding no statistically significant difference for the primary outcome), the experimental intervention corresponding to a therapeutic de-escalation as compared with the control one [6,7]. These trials frequently highlight the advantages of the experimental intervention in the article conclusions despite the lack of statistical difference, which may be considered ''spin'' (ie, a way of misleading readers by distorting the interpretation of results) [8,9]. The choice of a superiority design does not allow for this type of conclusion, which may indicate situations in which the authors failed to use an NI design.

In this study, we aimed to assess whether some superiority trials with nonsignificant results for the primary outcome could correspond to an NI situation and to explore associated trial characteristics.

2. Methods

2.1. Search strategy

We searched PubMed for all RCTs published in 2021 in the 5 journals in each medical specialty with the highest impact factor according to the Journal Citation Reports (Appendix 1). The search algorithm made with an expert of systematic review methodology (AD) is reported in Appendix 2.

2.2. Selection of relevant studies

We manually selected superiority trials comparing 2 active interventions with negative results (ie, non−statistically significant) for the primary outcome. We excluded trials that did not evaluate therapeutic interventions (eg, diagnostic evaluation); early-phase clinical trials (ie, phase I, II or II/III); studies that were not designed as superiority trials for the primary analysis (ie, equivalence and NI trials); trials comparing the experimental intervention to a placebo without active treatment, no intervention or usual care; trials with more than 2 arms; cluster, crossover or pilot trials; and post-hoc or follow-up analyses. Two reviewers (DR, XJ) independently selected the articles using Rayyan [10]. They manually screened titles and abstracts first and then full texts whenever necessary. Any disagreements were resolved by consultation with a third reviewer (AD).

2.3. Data extraction

Two reviewers (DR, and SH, AV, TD, or CR) independently extracted the following data for each trial using a standardized data extraction form. There were few disagreements and these were resolved by consultation with a third reviewer (AD).

- General characteristics: first and corresponding authors, journal, whether it was a specialty journal and its impact factor, whether statisticians or epidemiologists were involved (if a department of epidemiology, biostatistics, public health, or a clinical research unit was mentioned in the affiliations) [11], funding source, conflict of interest, registration, and whether the protocol was available.
- Setting: main country; whether it was a single-center or a multicenter trial.
- Medical condition: including whether it was a rare or an orphan disease.
- Methodological characteristics: masking, type of primary analysis (eg, intent-to-treat, modified intent-to-treat, per protocol analysis), mention of a switch from an NI to a superiority design or from superiority to

an NI design, primary outcome, and number of patients randomized and analyzed.
- Characteristics of experimental and control interventions: type of intervention (that is, pharmacological, nonpharmacological or both and which one: surgical procedure, medical device, psychotherapy, diet, education/promotion, or other.
- Discussion and conclusions: discussion about the possibility of an NI trial at the time of study conception, mention of existing NI trials on the same topic in the discussion or in the reference list.

2.4. Assessment of risk of bias and spin in the conclusions

Two reviewers (DR and LC) independently assessed the risk of bias for each trial with the revised version of the Cochrane Risk of Bias tool (RoB 2.0) for RCTs [12]. They also independently assessed the presence of spin in the abstract or main-text conclusions. All disagreements were resolved by consultation with a third reviewer (AD).

We considered spin in the article conclusions when the authors recommended the experimental intervention despite a negative result for the primary outcome analysis, by highlighting the potential advantages of the intervention, by mentioning a superior efficacy, by mentioning a comparable efficacy of both arms, or according to statistically significant secondary outcomes or subgroup analysis results [13].

2.5. Assessment of a potential NI situation

Three authors (DR, MB, LC) with a master's degree in clinical epidemiology but different background (2 residents in medicine, 1 in pharmacy) independently assessed each trial to evaluate whether the trial could correspond to an NI situation, that is, an evaluation of the efficacy as a primary outcome, with the experimental intervention presenting advantages as compared with the reference one based on different criteria, as follows.

- better safety profile and/or fewer drug interactions
- ease of administration (eg, oral pills instead of intravenous injections)
- ''therapeutic de-escalation'': fewer doses (eg, 1 daily dose instead of 3) or modified scheme of administration (individually tailored instead of fixed-schedule administration)
- intervention less invasive (eg, celioscopy rather than laparotomy)
- intervention less costly
- saving time and resources (eg, outpatient vs inpatient management)
- other advantages (availability, compliance, comfort, quality of life, esthetics etc.)

Any disagreements were resolved by consultation with a senior researcher (AD) who also checked all potential NI situations for validation.

2.6. For trials identified as corresponding to a potential NI situation

2.6.1. Registration and protocol verification

We checked the date of registration and the initial design as registered in the registry archives and checked the protocol when available. We assessed whether NI was mentioned in the original design of the trial.

2.6.2. Author survey

We contacted the corresponding authors to ask the following.

- whether they considered planning an NI trial, and if so, why they did not choose this design
- what values would they choose for the NI margin if they had to plan this type of trial

In case of nonresponse, we mailed 2 or 3 reminders.

2.6.3. Search of existing NI trials on the same topic

Finally, we evaluated whether NI trials comparing the same 2 interventions had been published. First, we screened the introduction and discussion parts of the selected publication and the reference list to determine whether NI trials were mentioned. If not, we searched PubMed using keywords for the population, interventions evaluated, and NI design.

2.7. Statistical analysis

Descriptive analysis included frequencies (percentages) for categorical variables and medians (IQR) for continuous variables. We compared the characteristics of trials corresponding to a potential NI situation vs the other trials. We did not perform statistical tests because our work was exploratory and because of the low number of potential NI situations expected and the number of characteristics evaluated. We used R version 4.1.2 for statistical analysis.

3. Results

3.1. Selection process

From the 2145 references identified, we excluded 1894 on the basis of the title and abstract. Among the remaining 251 references, 9 reports were not retrieved. We finally screened the full text of 242 publications and included 147 trials corresponding to our eligibility criteria (Fig 1).

3.2. Main characteristics of included trials

Most trials (79.6%, $n = 117$) were published in specialty journals (Table 1). The median impact factor was 14.8 (IQR

8.7-44.6). The trials were conducted mostly in Europe (n = 63, 42.9%) and North America (n = 32, 21.8%) and were predominantly multicenter trials (n = 101, 68.7%). The main medical specialties were anesthesiology (23 trials, 15.6%), oncology (20 trials, 13.6%), and intensive care (18 trials, 12.2%). Four trials concerned an orphan disease. The experimental interventions were pharmacological in 50 trials (34%), nonpharmacological in 90 (61.2%), and both in 7 (4.8%). The most frequent types of experimental interventions evaluated were drugs (57 trials, 38.8%) and surgical procedures (27 trials, 18.4%).

Regarding risk of bias (Appendix 3), 43.5% (n = 64) of trials were at low overall risk of bias, 43.5% (n = 64) had some concerns and the remaining 12.9% (n = 19) were at high risk of bias. Domains most frequently rated at high or some concerns were "deviations from the intended interventions" (n = 44, 29.9%), "missing outcome data" (n = 34, 23.1%), and "selection of the reported result" (n = 31, 21.1%).

3.3. Assessment of potential NI situations

We identified 19 trials that could correspond to a potential NI situation, which represents a rate of 12.9% (95% CI [7.9%, 19.4%]).

The advantages of the experimental over the control intervention were principally saving time and resources (n = 10) (eg, spinal anesthesia vs general anesthesia for hip surgery), better safety profile (n = 6) (eg, apremilast vs methotrexate for palmoplantar psoriasis), decreasing cost (n = 5) (eg, single dental implant with cantilever vs 2 adjacent implants), therapeutic de-escalation (n = 4) (eg, more-delayed vs standard-delayed renal replacement therapy in acute kidney injury), and lack of invasiveness

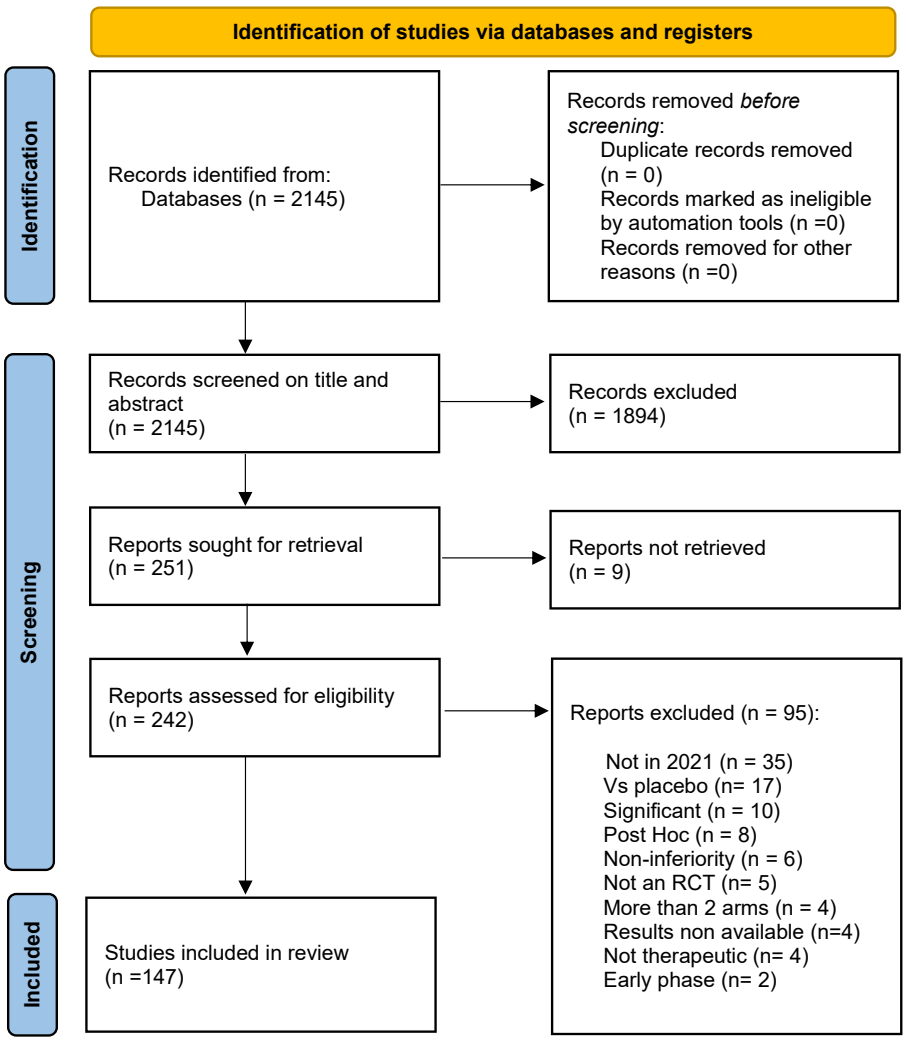

Figure 1. Flow diagram of the selection of articles.

Table 1. Characteristics of included randomized controlled trials overall and by noninferiority situation status

General characteristics	Overall (*N* = 147)	NI situation (*N* = 19)	No NI situation (*N* = 128)
Type of journal			
General	30 (20.4%)	6 (31.6%)	24 (18.8%)
Specialty	117 (79.6%)	13 (68.4%)	104 (81.3%)
Impact factor			
Median (IQR)	14.8 (8.7-44.6)	8.7 (7.3-67.8)	15.6 (9.1-44.5)
Involvement of a statistician or an epidemiologist			
Yes	73 (49.7%)	8 (42.1%)	65 (50.8%)
Funding source			
Public only	63 (42.9%)	6 (31.6%)	57 (44.5%)
Private only	30 (20.4%)	3 (15.8%)	27 (21.1%)
Public and private	32 (21.8%)	6 (31.6%)	26 (20.3%)
No specific funding	8 (5.4%)	2 (10.5%)	6 (4.7%)
Not reported	14 (9.5%)	2 (10.5%)	12 (9.4%)
Conflicts of interest reported			
Yes	99 (67.3%)	15 (78.9%)	84 (65.6%)
Registration reported			
Yes	142 (96.6%)	18 (94.7%)	124 (96.9%)
Setting			
Geographic area			
Europe	63 (42.9%)	9 (47.4%)	54 (42.2%)
North America	32 (21.8%)	6 (31.6%)	26 (20.3%)
International	23 (15.6%)	3 (15.8%)	20 (15.6%)
China	8 (5.4%)	-	8 (6.3%)
Other	20 (13.7%)	1 (5.3%)	17 (13.3%)
Not reported	1 (0.7%)	-	1 (0.8%)
Centers			
Multicenter	101 (68.7%)	14 (73.7%)	87 (68.0%)
Single center	40 (27.2%)	4 (21.1%)	36 (28.1%)
Not reported	6 (4.1%)	1 (5.3%)	5 (3.9%)
Number of patients randomized			
Median (IQR)	241 (120-547)	120 (74-264)	260 (140-557)
Medical Conditions			
Main medical specialties			
Anesthesiology	23 (15.6%)	1 (5.3%)	21 (16.4%)
Oncology	20 (13.6%)	1 (5.3%)	19 (14.8%)
Intensive Care	18 (12.2%)	1 (5.3%)	17 (13.3%)
Surgery	15 (10.2%)	4 (21.1%)	11 (8.6%)
Cardiology	8 (5.4%)	-	9 (7.0%)
Particular condition (rare/orphan disease)			
Yes	4 (2.7%)	1 (5.3%)	3 (2.3%)
Interventions			
Type of experimental interventions			
Pharmacological	50 (34%)	7 (36.8%)	43 (33.6%)
Nonpharmacological	90 (61.2%)	12 (63.2%)	78 (60.9%)
Both	7 (4.8%)	-	7 (5.5%)
Experimental interventions[a]			
Drugs	57 (38.8%)	7 (36.8%)	50 (39.1%)
Surgical procedures	27 (18.4%)	7 (36.8%)	20 (15.6%)
Monitoring techniques	18 (12.2%)	1 (5.3%)	17 (13.3%)
Medical procedures	14 (9.5%)	-	14 (10.9%)

(*Continued*)

Table 1. Continued

General characteristics	Overall (*N* = 147)	NI situation (*N* = 19)	No NI situation (*N* = 128)
Medical devices	10 (6.8%)	1 (5.3%)	9 (7.0%)
Education/promotion	9 (6.1%)	-	9 (7.0%)
Diet	5 (3.4%)	1 (5.3%)	4 (3.1%)
Physiotherapy	4 (2.7%)	1 (5.3%)	3 (2.3%)
Psychotherapy	2 (1.4%)	-	2 (1.6%)
Radiotherapy	4 (2.7%)	-	4 (3.1%)
Others	8 (5.4%)	1 (5.3%)	7 (5.5%)
Type of control interventions			
Pharmacological	48 (32.7%)	6 (31.6%)	42 (32.8%)
Nonpharmacological	97 (66%)	13 (68.4%)	84 (65.6%)
Both	2 (1.3%)	-	2 (1.6%)
Control interventions[a]			
Drugs	50 (34.0%)	6 (31.6%)	44 (34.4%)
Surgical procedures	25 (17.0%)	7 (36.8%)	18 (14.1%)
Monitoring techniques	18 (12.2%)	1 (5.3%)	17 (13.3%)
Medical procedures	17 (11.6%)	1 (5.3%)	16 (12.5%)
Medical devices	10 (6.8%)	1 (5.3%)	9 (7.0%)
Education/promotion	7 (4.8%)	-	7 (5.5%)
Diet	5 (3.4%)	1 (5.3%)	4 (3.1%)
Physiotherapy	4 (2.7%)	1 (5.3%)	3 (2.3%)
Psychotherapy	3 (2.0%)	-	3 (2.3%)
Radiotherapy	3 (2.0%)	-	3 (2.3%)
Others	8 (5.4%)	1 (5.3%)	7 (5.5%)

[a] The total number exceeds 100% because of different modalities for a same trial.

(n = 4) (eg, carotid artery stenting vs open carotid artery surgery in severe carotid artery stenosis) (Appendix 4).

3.4. Characteristics of trials corresponding and not corresponding to a potential NI situation

As compared with trials not corresponding to a potential NI situation, those that did correspond seemed to be published in a journal with a lower impact factor (median 8.7 [7.2-67.8] vs 15.6 [9.1-44.5]) and to less frequently involve epidemiologists or statisticians among the authors (n = 8, 42.1% vs n = 65, 50.8%) (Table 1). They also less frequently seemed to report masking of patients (none vs n = 38, 29.7%) and intent-to-treat analysis (n = 10, 52.6% vs n = 83, 64.8%) (Appendix 5). As well, their risk of bias seemed higher (n = 4, 21% vs n = 15, 12% for high risk of bias and n = 10, 53% vs n = 54, 42% for some concerns) (Fig 2). Finally, their sample size seemed smaller (median 120 [74-264) vs 260 [140-557] patients) (Table 1).

For the interpretation of results, as compared with trials not corresponding to a potential NI situation, those that did more frequently presented a form of spin in the conclusions, whether in the main text (n = 11, 57.9% vs n = 24, 18.8%) or in the abstract (n = 10, 52.6% vs n = 15, 11.7%) (Table 2). The main types of spin encountered in the article conclusions for trials corresponding to an NI situation were recommendations based on the comparable efficacy of both arms (n = 5, 26.3%) and recommendations based on the advantages of the experimental intervention arm (n = 5, 26.3%).

3.5. Registration and protocol checking for trials corresponding to a potential NI situation

All trials corresponding to a potential NI situation except 1 reported registration. Among those 18 trials, 16 (88.9%) were prospectively registered. All were planned with a superiority design without any mention of an NI design. Seven of the 19 trials (36.8%) had a protocol available, and the initial planned design was a superiority design.

3.6. Author survey

We received a response for 14 of the 19 trials (73.4%). The NI design had been considered for 3 trials (21.4%), including one for which the author mentioned that for the rare syndrome they were studying, the trial would have required a larger sample size. For 1 additional trial, the author asked us whether he should have considered an NI design. Only 2 trials (different from the ones that considered an NI design) proposed an NI margin.

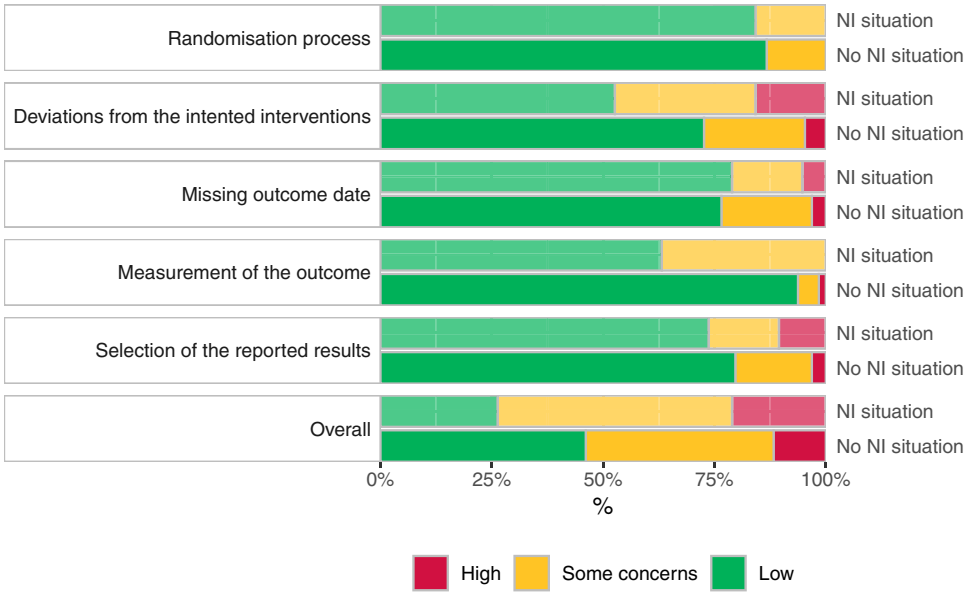

Figure 2. Risk of bias for trials corresponding and not corresponding to a noninferiority situation.

3.7. Search of existing NI trials on the same research question

Among the trials corresponding to a potential NI situation, none mentioned other NI trials studying the same research question in the text or reference list. By searching PubMed, we identified published NI trials corresponding to the same research question for 4 potential NI situations (21.1%): for 1 study [14], we identified 3 NI trials [15–17] and for the other 3 studies [18–20], 1 NI trial was identified ([21–23]). Only 1 of these trials was published after 2021 (23). The primary outcome was the same for only 1 (for the others, there were different time points or it was a composite outcome) so it is difficult to compare the NI margin in these trials with the results presented by the authors.

4. Discussion

In this research-on-research study of 147 trials with a nonsignificant result for the primary outcome, we identified

Table 2. Analysis of the types of spin in the abstract and main-text conclusions of trials corresponding and not corresponding to a noninferiority situation

	Abstract		Main-text conclusions	
General characteristics	**NI situation (N = 19)**	**No NI situation (N = 128)**	**NI situation (N = 19)**	**No NI situation (N = 128)**
Studies with at least 1 incidence of spin[a]	10 (52.6%)	15 (11.7%)	11 (57.9%)	24 (18.8%)
Strict recommendation	1 (5.3%)	1 (0.8%)	1 (5.3%)	4 (3.1%)
Recommendation based on the advantages of the experimental intervention arm	2 (10.5%)	3 (2.3%)	5 (26.3%)	4 (3.1%)
Recommendation based on the comparable efficacy of both arms	3 (15.8%)	3 (2.3%)	5 (26.3%)	3 (2.3%)
Mention of the comparable efficacy of both arms	3 (15.8%)	4 (3.1%)	2 (10.5%)	7 (5.5%)
Recommendation based on superior efficacy of the experimental arm	1 (5.3%)	-	-	-
Recommendation based on secondary outcomes significant results	-	4 (3.1%)	-	6 (4.7%)
Recommendation based on subgroup analysis significant results	-	2 (1.6%)	-	2 (1.6%)

[a] One trial can present several forms of spin.

19 corresponding to a potential NI situation. These trials seem to have greater risk of bias and lower sample size than the other trials. Also, they seemed to more frequently report spin in study conclusions whether in abstract or main text than the other trials. When we contacted the corresponding authors of these potential NI trials, the NI design had been considered for 3 trials (21.4%). Our PubMed search also identified 4 NI situations with existing NI trials for the same research question.

Our study has a particularly original concept because it is the first, to our knowledge, to evaluate whether negative superiority trials comparing 2 active interventions could correspond to NI situations. To identify negative trials, we manually screened all abstracts, and full texts whenever necessary, to ensure that the result was statistically nonsignificant for the primary outcome because of no specific keywords targeting negative trials. We used rigorous methods with the selection of relevant studies and data extraction conducted in duplicate. Potential NI situations were assessed independently by 3 authors with diverse medical backgrounds to limit interpretation biases.

Our study has some limitations. We did not register the protocol. We restricted our search to the general and specialty journals with the top 5 impact factors to focus on well-reported trials [5,24] and because searching for all the published trials was not feasible. These trials may be more likely to have adequate methods [24] including a design well-adapted to the research question, which could underestimate the proportion of potential NI situations. We searched for only trials published over a 1-year period for practical reasons, which may have limited our ability to identify potential NI situations because we expected these trials to be relatively seldom. Our study includes trials identified as a potential NI situation comparing the experimental intervention to an active control, without the latter necessarily being the standard of care or the reference intervention for a particular pathology although active comparators must be the standard of care [25]. Nevertheless, there was no known standard of care in the pathology. In addition, we excluded 3-arm trials comparing 2 active interventions with a placebo arm, which could theoretically correspond to an NI situation between the 2 active interventions compared. However, this type of design has some ethical issues [26−28] because a placebo is assigned to patients when two alternate active strategies already exist. This situation represents a small amount of trials in the literature, so it might not really affect the proportion of potential NI situations identified.

In our study, the 12.9% proportion of NI situations identified seems non-negligible, even though we do not have any comparable figure in the literature. We decided to identify the trials that ‘‘could’’ have been designed as NI trials and not ‘‘should’’ have been designed: in fact, if the objective of the trial was to demonstrate the superiority of the experimental over the control intervention, it had to be designed that way [29]. During our author survey, in some cases, the choice of the design could be influenced by some elements such as the pathology, with rare diseases, which could result in patient recruitment issues. NI trials generally require a larger sample size than superiority trials, so an NI trial may have been considered unfeasible.

Misinterpretation of statistically nonsignificant results in superiority trials is common [30]. No evidence of difference does not mean evidence of no difference [30], so concluding that 2 interventions are equivalent or noninferior with a statistically nonsignificant result in a superiority trial may be considered spin [13,31]. In our study, the 19 potential NI situations seemed more likely to report spin in the conclusions, with 57.9% of these trials using a form of spin in the main-text conclusions. They either recommended the experimental intervention by highlighting its advantages over the control or simply mentioned comparable efficacy for both arms. In contrast, trials not identified as a potential NI situation had a lower risk of spin, with a rate of 18.8%. We considered only trials with negative results for the primary outcome because a common spin is to conclude equivalence or NI with nonsignificant results, but we cannot exclude that some trials with statistically significant results for the primary outcome could also correspond to potential NI situations. It would have been interesting to assess whether some of these trials could correspond to a potential NI situation. This could serve as negative controls in our analysis as we expected the rate of these trials close to 0.

Risk of bias seemed higher for trials corresponding than not corresponding to an NI situation. Those trials seemed less frequently masking patients or outcome assessors and reporting an intent-to-treat analysis. This observation is consistent with our assessment suggesting that an NI trial could have been more adapted to answer the research question than a superiority trial. The choice of a less appropriate design, the higher risk of bias, and the more frequent reporting of spin in the conclusions are also consistent with a slightly lower involvement of methodologists or biostatisticians in these trials. All these findings may also explain the lower impact factor of the journals in which trials corresponding to an NI situation were published because high−impact factor journals generally have higher methodological expectations [32]. Trials corresponding to an NI situation may also remain unpublished and because we focused on published trials only, we may have underestimated the rate of trials corresponding to an NI situation in our study. Another approach could have been to work on a cohort of registered trials. However, the amount of details at registration might be insufficient to judge whether an NI design would have been possible.

The sample sizes observed in NI-situation trials were relatively low, with a median of 120 patients, and lower than the sample sizes observed in non−NI-situation trials, which was 260, although the sample size needed for an NI trial may be higher than for a superiority trial. This larger sample size is due to the choice of the NI margin,

which is often smaller than the expected difference that would have been chosen in a superiority trial on the same topic [33,34]. Some authors have suggested conducting phase II NI trials [35]. The principle is to relax some of the sample size calculation parameters (ie, higher type I error rate, lower power, or wider NI margin) to require less patients [33]. These trials will provide a lower level of evidence but might help identify a signal of potential NI(33) before conducting a confirmatory larger NI trial requiring more patients.

5. Conclusion

This research-on-research study showed that a non-negligible proportion of negative superiority trials could have been designed as NI trials. As compared with trials not corresponding to an NI situation, those corresponding to an NI situation seemed to have a higher risk of bias and to present some forms of spin in articles, which may distort the interpretation of results by readers.

CRediT authorship contribution statement

Deivanes Rajendrabose: Writing − original draft, Investigation, Formal analysis, Data curation, Conceptualization. **Lucie Collet:** Writing − review & editing, Investigation. **Camille Reinaud:** Writing − review & editing, Investigation. **Maxime Beydon:** Writing − review & editing, Investigation. **Xiaojun Jiang:** Writing − review & editing, Investigation. **Sahra Hmissi:** Writing − review & editing, Investigation. **Antonin Vermillac:** Writing − review & editing, Investigation. **Thomas Degonzague:** Writing − review & editing, Investigation. **David Hajage:** Writing − review & editing, Methodology, Formal analysis. **Agnès Dechartres:** Writing − review & editing, Validation, Supervision, Methodology, Conceptualization.

Declaration of competing interest

There are no competing interests for any author.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jclinepi.2024.111613.

Data availability

Raw data are available on Open Science Framework (https://osf.io/xt8nk/?view_only=fdef4fda88ac4ce4ad086978b8521ddf)

References

[1] EMA. EMA 2005 GUIDELINE ON. The choice of the NON-inferiority margin. 2024. Available at: https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-choice-non-inferiority-margin_en.pdf. Accessed December 2, 2024.

[2] Piaggio G, Elbourne DR, Altman DG, Pocock SJ, Evans SJW. CONSORT group for the. Reporting of noninferiority and equivalence randomized trials: an extension of the CONSORT statement. JAMA 2006;295(10):1152−60.

[3] Leening MJG, Mahmoud KD. Non-efficacy benefits and non-inferiority margins: a scoping review of contemporary high-impact non-inferiority trials in clinical cardiology. Eur J Epidemiol 2021;36(11):1103−9.

[4] Mauri L, D'Agostino RB. Challenges in the design and interpretation of noninferiority trials. In: Drazen JM, Harrington DP, McMurray JJV, Ware JH, Woodcock J, editors. N Engl J Med 2017;377(14):1357−67.

[5] Vinkers CH, Lamberink HJ, Tijdink JK, Heus P, Bouter L, Glasziou P, et al. The methodological quality of 176,620 randomized controlled trials published between 1966 and 2018 reveals a positive trend but also an urgent need for improvement. In: Siegerink B, editor. PLoS Biol 2021;19(4):e3001162.

[6] Charles P, Terrier B, Perrodeau É, Cohen P, Faguer S, Huart A, et al. Comparison of individually tailored versus fixed-schedule rituximab regimen to maintain ANCA-associated vasculitis remission: results of a multicentre, randomised controlled, phase III trial (MAINRITSAN2). Ann Rheum Dis 2018;77(8):1143−9.

[7] Gaudry S, Hajage D, Schortgen F, Martin-Lefevre L, Pons B, Boulet E, et al. Initiation strategies for renal-replacement therapy in the intensive care unit. N Engl J Med 2016;375(2):122−33.

[8] Fletcher RH, Black B. "Spin" in scientific writing: scientific mischief and legal jeopardy. Med Law 2007;26(3):511−25.

[9] Junger D. The rhetoric of research. Embrace scientific rhetoric for its power. BMJ 1995;311(6996):61.

[10] Olofsson H, Brolund A, Hellberg C, Silverstein R, Stenström K, Österberg M, et al. Can abstract screening workload be reduced using text mining? User experiences of the tool Rayyan. Res Syn Meth 2017;8(3):275−80.

[11] Delgado-Rodriguez M, Ruiz-Canela M, De Irala-Estevez J, Llorca J, Martinez-Gonzalez A. Participation of epidemiologists and/or biostatisticians and methodological quality of published controlled clinical trials. J Epidemiol Community Health 2001;55(8):569−72.

[12] Current version of Risk of bias tool RoB 2. Available at: https://www.riskofbias.info/welcome/rob-2-0-tool/current-version-of-rob-2. Accessed December 2, 2024.

[13] Boutron I. Reporting and interpretation of randomized controlled trials with statistically nonsignificant results for primary outcomes. JAMA 2010;303(20):2058.

[14] Cosyn J, Eeckhout C, Christiaens V, Eghbali A, Vervaeke S, Younes F, et al. A multi-centre randomized controlled trial comparing connective tissue graft with collagen matrix to increase soft tissue thickness at the buccal aspect of single implants: 3-month results. J Clin Periodontol 2021;48(12):1502−15.

[15] Thoma DS, Zeltner M, Hilbe M, Hämmerle CHF, Hüsler J, Jung RE. Randomized controlled clinical study evaluating effectiveness and safety of a volume-stable collagen matrix compared to autogenous connective tissue grafts for soft tissue augmentation at implant sites. J Clin Periodontol 2016;43(10):874−85.

[16] Zeltner M, Jung RE, Hämmerle CHF, Hüsler J, Thoma DS. Randomized controlled clinical study comparing a volume-stable collagen matrix to autogenous connective tissue grafts for soft tissue augmentation at implant sites: linear volumetric soft tissue changes up to 3 months. J Clin Periodontol 2017;44(4):446−53.

[17] McGuire MK, Scheyer ET, Lipton DI, Gunsolley JC. Randomized, controlled, clinical trial to evaluate a xenogeneic collagen matrix as an alternative to free gingival grafting for oral soft tissue augmentation: a 6- to 8-year follow-up. J Periodontol 2021;92(8):1088−95.

[18] Combined Randomised and Observational Study of Surgery for Fractures in the Distal Radius in the Elderly (CROSSFIRE) Study Group, Lawson A, Naylor JM, Buchbinder R, Ivers R, Balogh ZJ. Surgical plating vs closed reduction for fractures in the distal radius in older patients: a randomized clinical trial. JAMA Surg 2021;156(3):229−37.

[19] Halliday A, Bulbulia R, Bonati LH, Chester J, Cradduck-Bamford A, Peto R, et al. Second asymptomatic carotid surgery trial (ACST-2): a randomised comparison of carotid artery stenting versus carotid endarterectomy. Lancet 2021;398(10305):1065−73.

[20] Slagter KW, Meijer HJA, Hentenaar DFM, Vissink A, Raghoebar GM. Immediate single-tooth implant placement with simultaneous bone augmentation versus delayed implant placement after alveolar ridge preservation in bony defect sites in the esthetic region: a 5-year randomized controlled trial. J Periodontol 2021;92(12):1738−48.

[21] Eckstein HH, Ringleb P, Allenberg JR, Berger J, Fraedrich G, Hacke W, et al. Results of the Stent-Protected Angioplasty versus Carotid Endarterectomy (SPACE) study to treat symptomatic stenoses at 2 years: a multinational, prospective, randomised trial. Lancet Neurol 2008;7(10):893−902.

[22] Slagter KW, Meijer HJA, Bakker NA, Vissink A, Raghoebar GM. Immediate single-tooth implant placement in bony defects in the esthetic zone: a 1-year randomized controlled trial. J Periodontol 2016;87(6):619−29.

[23] ter Meulen DP, Mulders MAM, Kruiswijk AA, Kret EJ, Slichter ME, van Dongen JM, et al. Effectiveness and cost-effectiveness of surgery versus casting for elderly patients with D isplaced intra- A rticular type C distal R adius fractures: protocol of a randomised controlled T rial with economic evaluation (the DART study). BMJ Open 2022;12(4):e051658.

[24] Dechartres A, Trinquart L, Atal I, Moher D, Dickersin K, Boutron I, et al. Evolution of poor reporting and inadequate methods over time in 20 920 randomised controlled trials included in Cochrane reviews: research on research study. BMJ 2017;357:j2490.

[25] Le Henanff A, Giraudeau B, Baron G, Ravaud P. Quality of reporting of noninferiority and equivalence randomized trials. JAMA 2006;295(10):1147.

[26] Rothman KJ, Michels KB. The continuing unethical use of placebo controls. N Engl J Med 1994;331(6):394−8.

[27] Taubes G. Use of placebo controls in clinical trials disputed. Science 1995;267(5194):25−6.

[28] Freedman B, Weijer C, Glass KC. Placebo orthodoxy in clinical research I: empirical and methodological myths. J Law Med Ethics 1996;24(3):243−51.

[29] Leon AC. Comparative effectiveness clinical trials in psychiatry: superiority, noninferiority, and the role of active comparators. J Clin Psychiatry 2011;72(10):1344−9.

[30] Hemming K, Javid I, Taljaard M. A review of high impact journals found that misinterpretation of non-statistically significant results from randomized trials was common. J Clin Epidemiol 2022;145:112−20.

[31] Boutron I, Altman DG, Hopewell S, Vera-Badillo F, Tannock I, Ravaud P. Impact of spin in the abstracts of articles reporting results of randomized controlled trials in the field of cancer: the SPIIN randomized controlled trial. J Clin Orthod 2014;32(36):4120−6.

[32] Ahmed Ali U, Reiber BMM, Ten Hove JR, Van Der Sluis PC, Gooszen HG, Boermeester MA, et al. Journal impact factor and methodological quality of surgical randomized controlled trials: an empirical study. Langenbeck's Arch Surg 2017;402(7):1015−22.

[33] Jung SH. Design of phase II non-inferiority trials. Contemp Clin Trials Commun 2017;7:23−7.

[34] Dunn DT, Copas AJ, Brocklehurst P. Superiority and non-inferiority: two sides of the same coin? Trials 2018;19(1):499.

[35] Neuenschwander B, Rouyrre N, Hollaender N, Zuber E, Branson M. A proof of concept phase II non-inferiority criterion. Statist Med 2011;30(13):1618−27.