Superiority vs. Non-Inferiority

Porntep Amornritvanich, MD, FAPSC

Journal of Clinical Epidemiology

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

Published: 16th November 2024

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

Superiority vs. Non-Inferiority (NI) Trials

Aspect	Superiority Trial	Non-Inferiority (NI) Trial
Purpose	To show the new intervention is better than control	To show the new intervention is not worse (within a margin)
Hypothesis	H_o: No difference or new is worse	H_o: New is worse beyond NI margin
	H_1: New is better	H_1: New is not worse
Typical Use	New treatment expected to improve outcomes	New treatment offers other benefits (safety, cost, convenience)
Result Interpretation	Reject H_o if statistically significant difference	Conclude NI if difference is within pre-specified NI margin
Sample Size	Usually smaller (needs detectable difference)	Usually larger (prove not worse)
Risk of Bias	Less sensitive to bias	Highly sensitive — bias may falsely show non-inferiority

Key Message 

- Superiority = Proving “better”

- Non-Inferiority = Proving “good enough” with extra perks

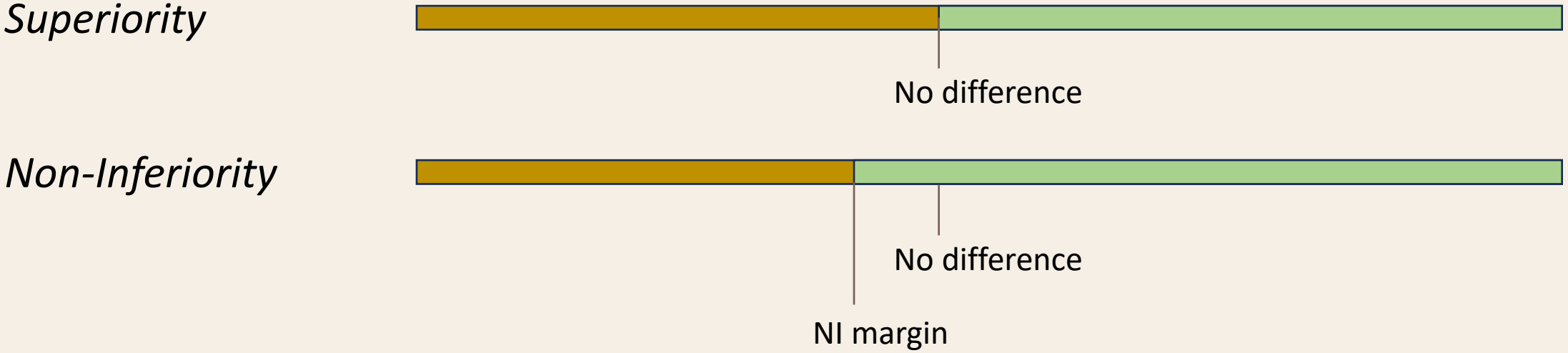
  (safety, cost, convenience)

NI margin 

- Δ (Delta) = The largest acceptable difference where the new treatment is “not unacceptably worse”
- It reflects the clinical judgment of “how much worse is still acceptable” if benefits like safety, cost, or convenience exist

Superiority vs. Non-Inferiority Hypothesis

Null hypothesis/**Alternative Hypothesis**



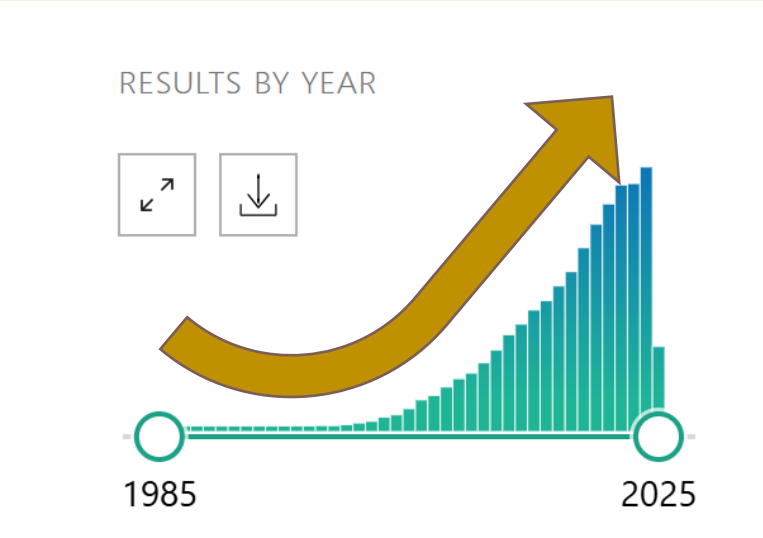
25th March 2025

PubMed®

noninferiority

Search

Advanced Create alert Create RSS User Guide



22,002 results

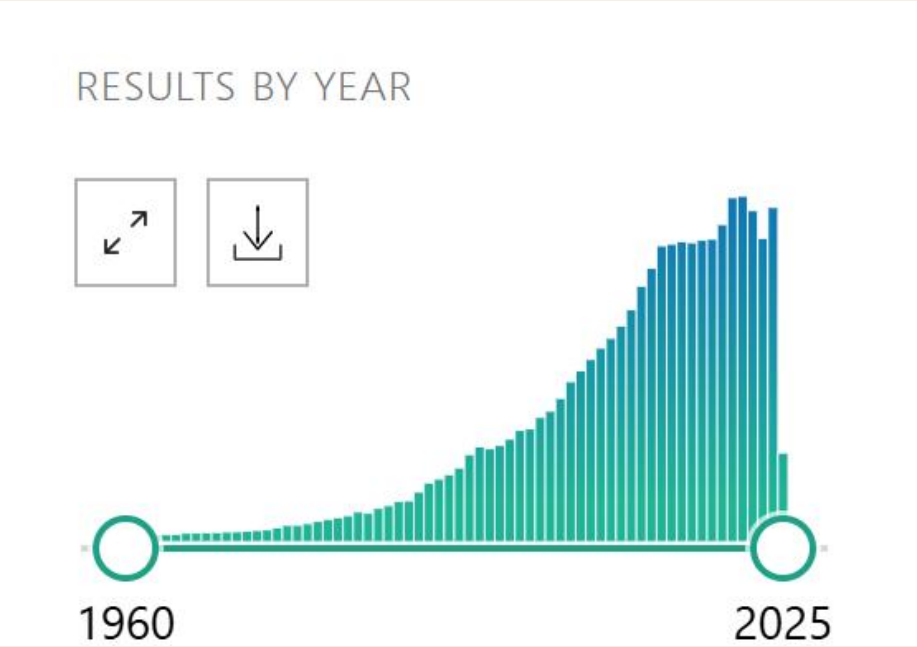
25th March 2025

PubMed®

randomised controlled trial

Search

Advanced Create alert Create RSS User Guide



861,674 results

The NEW ENGLAND JOURNAL of MEDICINE

14th July 2016

ORIGINAL ARTICLE

Initiation Strategies for Renal-Replacement Therapy in the Intensive Care Unit

Stéphane Gaudry, M.D., David Hajage, M.D., Fréderique Schortgen, M.D., Laurent Martin-Lefevre, M.D., Bertrand Pons, M.D., Eric Boulet, M.D., Alexandre Boyer, M.D., Guillaume Chevrel, M.D., Nicolas Lerolle, M.D., Ph.D., Dorothée Carpentier, M.D., Nicolas de Prost, M.D., Ph.D., Alexandre Lautrette, M.D., Anne Bretagnol, M.D., Julien Mayaux, M.D., Saad Nseir, M.D., Ph.D., Bruno Megarbane, M.D., Ph.D., Marina Thirion, M.D., Jean-Marie Forel, M.D., Julien Maizel, M.D., Ph.D., Hodane Yonis, M.D., Philippe Markowicz, M.D., Guillaume Thiery, M.D., Florence Tubach, M.D., Ph.D., Jean-Damien Ricard, M.D., Ph.D., and Didier Dreyfuss, M.D., for the AKIKI Study Group*

620 patients

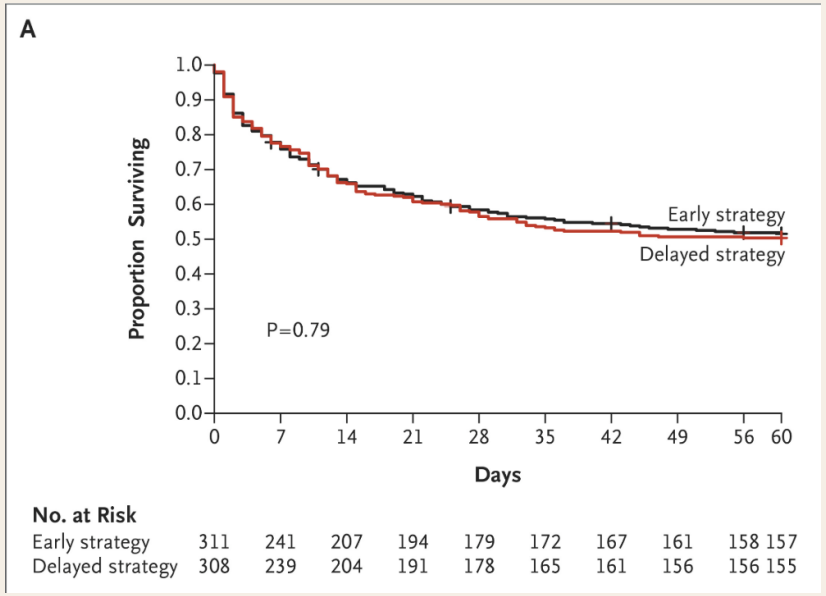
AKI stage III

- Cr rise > 3-fold or
- Urine output ≤ 0.3 mg/kg/hr more than 24 hr.

Early RRT

Delayed RRT

Outcome:
overall survival at 60 days



CONCLUSIONS

In a trial involving critically ill patients with severe acute kidney injury, we found no significant difference with regard to mortality between an early and a delayed strategy for the initiation of renal-replacement therapy. A delayed strategy averted the need for renal-replacement therapy in an appreciable number of patients. (Funded by the French Ministry of Health; ClinicalTrials.gov number, NCT01932190.)

Research question

Among superiority RCTs with non-significant results comparing two active interventions, how many were actually more suitable for a NI design?

Search strategy

- **Database:** PubMed
- **Population:** Human studies only
- **Study type:** Randomized Controlled Trials (RCTs)
- **Time frame:** Published in 2021

- **Scope:** RCTs comparing two active interventions published in high-impact journals from multiple specialties

Search strategy

- Exclude
  - Placebo-controlled trial
  - Study mentions “Non-inferior” in the title

How Did They Judge Potential NI Situations?

- 3 independent reviewers
- Used predefined criteria:
  - Did the experimental intervention offer real-world advantages (safety, cost, convenience)?
  - Did the trial test a less intensive or de-escalation strategy?
  - Did the study mention comparable efficacy or alternative benefits in the conclusion?
- Reviewers discussed disagreements until reaching consensus.

Assessment for Trials Identified as Potential NI Situations

- Registration and Protocol Check
- Author Survey
- Search for Similar NI Trials

Assessment for Trials Identified as Potential NI Situations

- Registration and Protocol Check
  - They checked the trial registry and protocol (if available)
  - Looked for any mention of non-inferiority (NI) design in the original trial plan
- Author Survey
- Search for Similar NI Trials

Assessment for Trials Identified as Potential NI Situations

- Registration and Protocol Check
- Author Survey
- Search for Similar NI Trials

- They emailed the trial authors to ask:
  - Did you consider planning your study as an NI trial? If yes, why not use it?
  - If you had done an NI trial, what NI margin would you choose?
- Sent 2-3 reminders if no response

Assessment for Trials Identified as Potential NI Situations

- Registration and Protocol Check
- Author Survey
- Search for Similar NI Trials
  - Checked if any existing NI trials comparing the same interventions were published
  - Reviewed intro, discussion, and references of the trial
  - If not found, searched PubMed using keywords for population, interventions, and NI design

"SPIN" in potential NI trials

- Spin = *Misleading readers by presenting non-significant results as favorable*

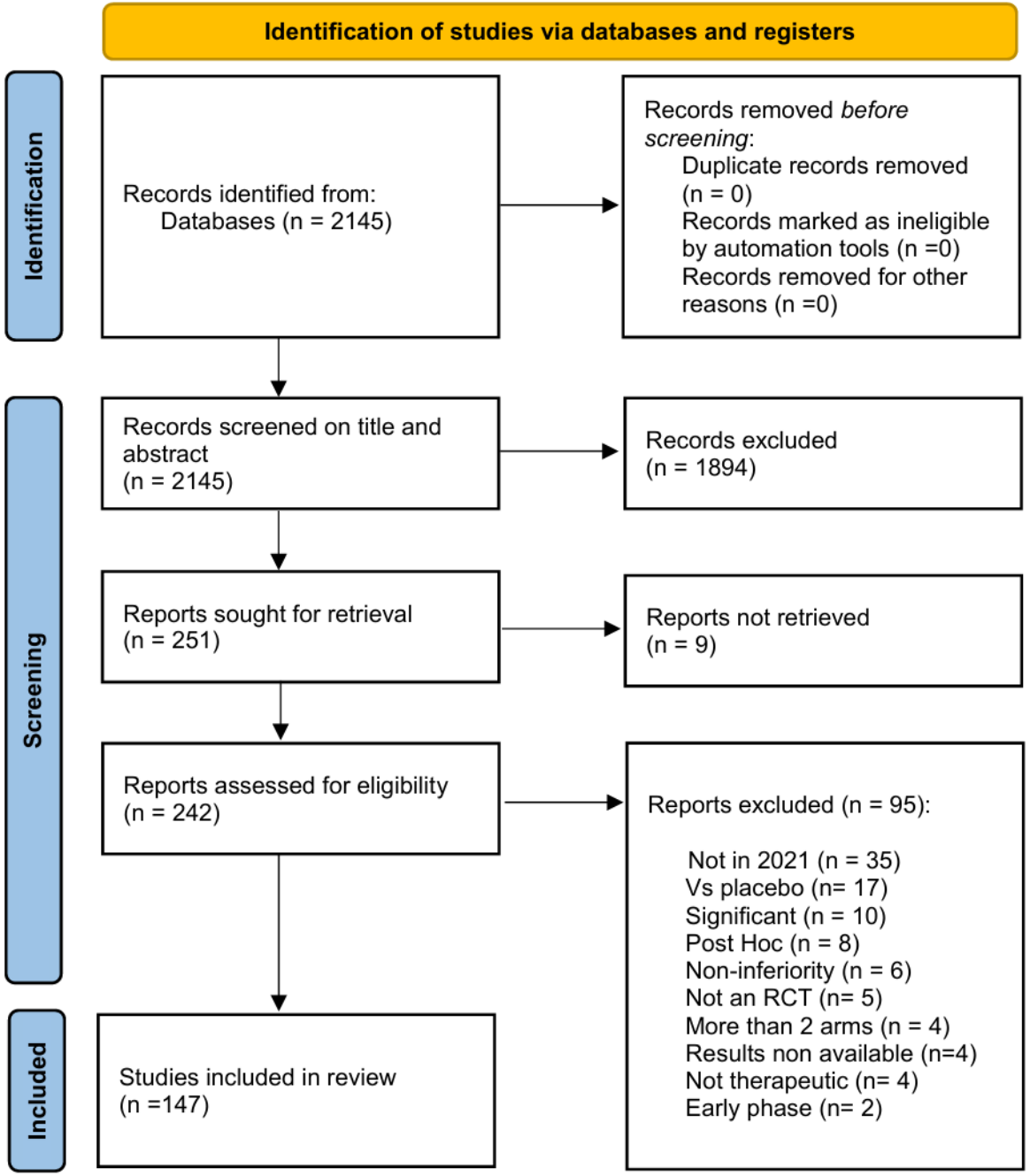


Figure 1. Flow diagram of the selection of articles.

Characteristics of Trials That Could Be Non-Inferiority (NI Situations)

Characteristic	NI-like Trials (n = 19)	Other Trials (n = 128)
Median Sample Size (IQR)	120 (68–218)	260 (135–546)
High Risk of Bias	68%	36%
Spin in Text	58%	19%
Spin in Abstract	47%	25%
Median Journal Impact Factor	8.7	15.6

Types of Spin in Abstract and Main-Text

Spin Category	NI Trials (n=19)	Non-NI Trials (n=128)
Any Spin (Abstract)	52.6% (10/19)	11.7% (15/128)
Any Spin (Main Text)	57.9% (11/19)	18.8% (24/128)
Strict Recommendation	5.30%	3.10%
Recommendation based on advantages	26.30%	3.10%
Recommendation based on comparable efficacy	26.30%	2.30%
Mention of comparable efficacy	10.50%	5.50%
Recommendation based on secondary outcomes	0%	4.70%
Recommendation based on subgroup analysis	0%	1.60%

Author survey results

- Key points: Only 3 of 19 authors considered NI design

- Reasons they didn't use NI:
  - Feasibility concerns (sample size too large)
  - Rare disease population

- Most authors **didn't even think about NI design**

Discussion

- 12.9% of negative superiority trials could have been designed as non-inferiority (NI) trials

- These trials had:
  - Higher risk of bias
  - Smaller sample size
  - More frequent spin in conclusions (57.9% vs. 18.8%)

- NI design was rarely considered upfront — authors cited feasibility issues (rare disease, sample size needs)

Discussion

- Spin often misleads by suggesting non-inferiority based on non-significant superiority results
- Poor methodology and less statistician involvement may explain bias and spin
- Published in lower impact factor journals — reflecting methodological concerns
- Sample size often too small for proper NI design (median 120 vs. 260)

Strength and Limitations

- **Strengths:**
  - **First study** to systematically check if negative superiority trials **could have been NI trials**
  - Clear **predefined criteria** for judging potential NI situations
  - **Independent review** by three experienced methodologists with consensus decision
  - Broad coverage across **top journals and specialties**
- **Limitations**:
  - **Subjective judgment** — deciding if a trial fits NI was reviewer-dependent (but resolved by consensus)
  - Only checked trials **published in 2021** → limits generalizability
  - Unable to access **protocols** for every trial → possible misclassification if authors had an NI mindset but didn't label it
  - **Spin assessment** focused only on the **abstract and main text** conclusions, not the entire discussion

Strength and Limitations

- **Strengths:**
  - **First study** to systematically check if negative superiority trials **could have been NI trials**
  - Clear **predefined criteria** for judging potential NI situations
  - **Independent review** by three experienced methodologists with consensus decision
  - Broad coverage across **top journals and specialties**

Strength and Limitations

- **Limitations**:
  - **Subjective judgment** — deciding if a trial fits NI was reviewer-dependent (but resolved by consensus)
  - Only checked trials **published in 2021** → limits generalizability
  - Unable to access **protocols** for every trial → possible misclassification if authors had an NI mindset but didn't label it
  - **Spin assessment** focused only on the **abstract and main text** conclusions, not the entire discussion

Take-home messages

- Not all RCTs should be superiority trials — choose design based on clinical question
- Non-inferiority design is underused even when it fits the real-world aim
- Spin is common and can mislead — we must critically read conclusions
- Plan NI trials carefully with a clear margin (Δ) when “good enough” is the goal
- Better design = better science = better patient care

Thank you