

CRITICAL REVIEW

Ensuring Superior Reporting of Radiation Therapy Noninferiority Trials: A Systematic Review



Andrew J. Arifin, MD,^a Vivian S. Tan, MD,^a Michael Yan, MD,^b
Andrew Warner, MSc,^a R. G. Boldt, MLIS,^a Hanbo Chen, MD,^c
George B. Rodrigues, MD, PhD,^a David A. Palma, MD, PhD,^a and
Alexander V. Louie, MD, PhD^{c,*}

^aDivision of Radiation Oncology, London Regional Cancer Program, London, Ontario, Canada; ^bRadiation Medicine Program, Princess Margaret Cancer Centre, Toronto, Ontario, Canada; and ^cDepartment of Radiation Oncology, Odette Cancer Centre, Sunnybrook Health Sciences Centre, Toronto, Ontario, Canada

Received 14 November 2022; accepted 9 January 2023

Abstract

Purpose: Although the frequency of noninferiority trials is increasing, the consistency of the reporting of these trials can vary. The aim of this systematic review was to assess the reporting quality of radiation therapy noninferiority trials.

Methods and Materials: The PubMed, Embase, and Cochrane databases were queried for randomized controlled radiation therapy trials with noninferiority hypotheses published in English between January 2000 and July 2022, and this was performed by an information scientist. Descriptive statistics were used to summarize data.

Results: Of 423 records screened, 59 (14%) were included after full-text review. All were published after 2003 and open label. The most common primary cancer type was breast (n = 15, 25%). Altered radiation fractionation (n = 26, 45%) and radiation de-escalation (n = 11, 19%) were the most common types of interventions. The most common primary endpoints were locoregional control (n = 17, 29%) and progression-free survival (n = 14, 24%). Fifty-three (90%) reported the noninferiority margin, and only 9 (17%) provided statistical justification for the margin. The median absolute noninferiority margin was 9% (interquartile range, 5%-10%), and the median relative margin was 1.51 (interquartile range, 1.33-2.04). Sample size calculations and confidence intervals were reported in 54 studies (92%). Both intention-to-treat and per-protocol analyses were reported in 27 studies (46%). In 31 trials (53%), noninferiority of the primary endpoint was reached.

Conclusions: There was variability in the reporting of key components of noninferiority trials. We encourage consideration of additional statistical reasoning such as guidelines or previous trials in the selection of the noninferiority margin, reporting both absolute and relative margins, and the avoidance of statistically vague or misleading language in the reporting of future noninferiority trials.

Crown Copyright © 2023 Published by Elsevier Inc. on behalf of American Society for Radiation Oncology. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Sources of support: This work had no specific funding.

Disclosures: Dr Arifin is a board member of the Canadian Association of Radiation Oncology. Dr Palma reports research funding from the Ontario Institute for Cancer Research and a consultant relationship with Need Inc., unrelated to the present work. Dr Louie has received honoraria from AstraZeneca for advisory board participation and speaker's fees. No other disclosures were reported.

Research data are available upon reasonable request.

*Corresponding author: Alexander V. Louie, MD, PhD; E-mail: Alexander.Louie@sunnybrook.ca

Introduction

Noninferiority trials aim to demonstrate that an experimental treatment is not worse than the standard treatment by a prespecified threshold called the noninferiority margin. These studies are often conducted when the experimental treatment is more convenient

<https://doi.org/10.1016/j.adro.2023.101178>

2452-1094/Crown Copyright © 2023 Published by Elsevier Inc. on behalf of American Society for Radiation Oncology. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

for patients, less toxic, more readily available, less costly, and/or when it is unethical to perform a placebo-controlled trial.¹

In a superiority trial, the null hypothesis asserts that 2 arms are the same. If the lower bound of the 95% confidence interval (CI) of the treatment difference is above zero, one can reject the null hypothesis (Fig 1A). In

contrast, the null hypothesis in a noninferiority trial states that the experimental arm is worse than the control arm by a specified margin (δ). There are 6 possible outcomes from a noninferiority trial as shown in Fig 1B. If the lower bound of the 95% CI of the treatment difference is above the noninferiority margin, one can conclude noninferiority. Depending on if the 95% CI lies wholly above or

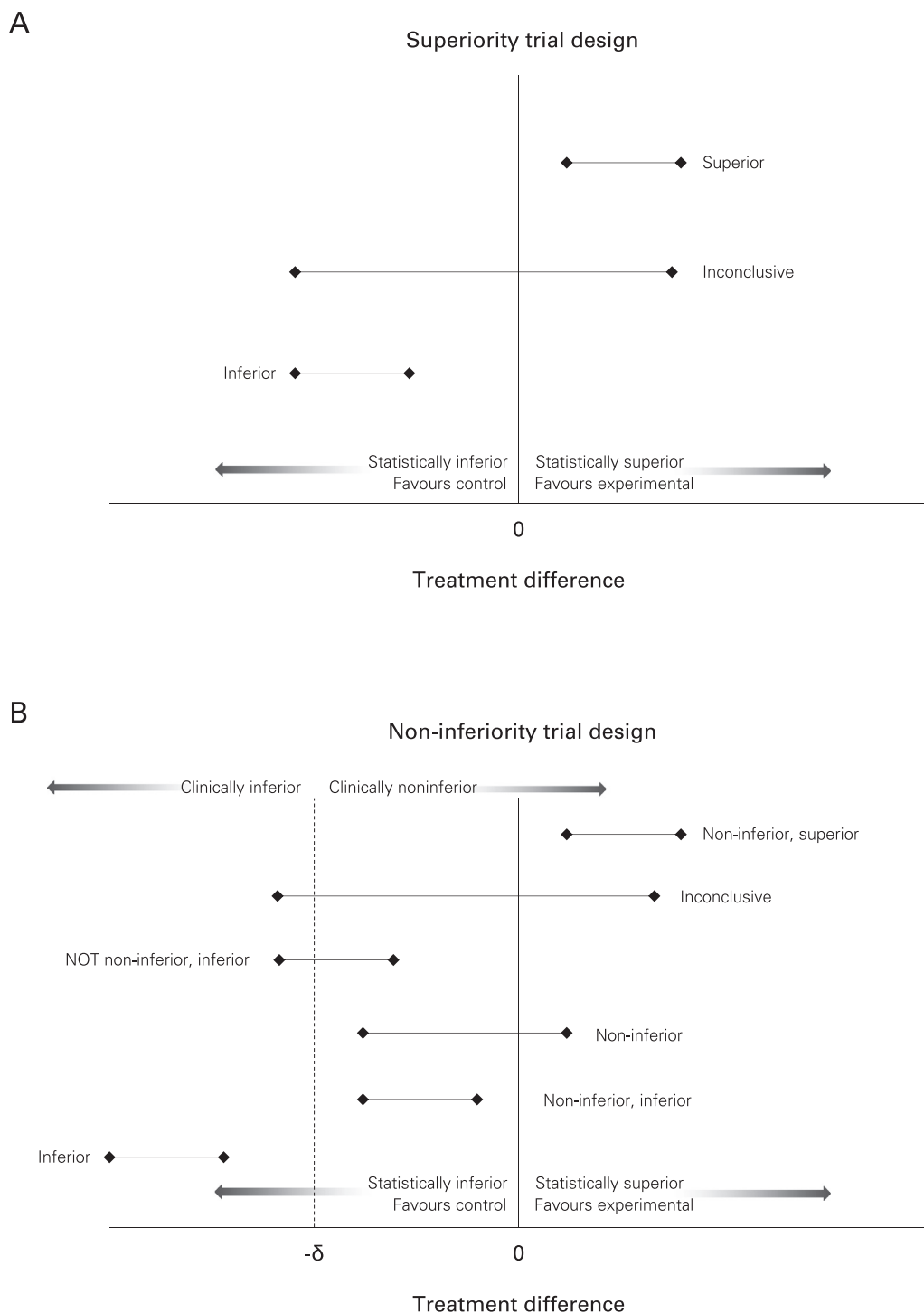


Figure 1 Conclusions from the 95% confidence intervals of treatment differences in superiority trials (A) and noninferiority trials (B).

below 0, one can also conclude statistical superiority or inferiority, respectively.

As with other types of trials, the methodological quality of noninferiority trials should be appraised before drawing conclusions. A 2006 review of noninferiority trials published between 2003 and 2004 showed that only 20.3% of studies fulfilled reporting requirements to adequately allow readers to make conclusions.² To improve the quality of reporting, the Consolidated Standards of Reporting Trials (CONSORT) group published a statement regarding reporting standards for noninferiority and equivalence clinical trials.¹ A summary of the recommendations from this report are listed in [Table 1](#).

Previous reviews of noninferiority trials in cancer have mainly focused on pharmacologic trials.³ To our knowledge, none have examined those involving radiation therapy. Noninferiority trials are important in radiation oncology as many trials test different schedules to make treatments more convenient or less toxic. This review aims to evaluate the reporting quality of noninferiority clinical trials involving radiation therapy by analyzing the reported data and to describe the characteristics of these studies.

Methods and Materials

This systematic review was performed and reported according to the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) statement.⁴ The prespecified protocol was registered with the International Prospective Register of Systematic Reviews (PROSPERO), CRD42021270644.

Search

A literature search of the PubMed, Embase, and Cochrane databases of randomized controlled radiation therapy trials with noninferiority hypotheses published in English between January 1, 2000, and July 18, 2022, was performed by an information scientist (RGB) on July 18, 2022. The exact search strategy is detailed in [Appendix E1](#).

Study selection

Population

We included publications of randomized controlled trials of pediatric and adult patients. We did not include abstracts, study protocols, follow-up, or interim analyses.

Intervention

Trials must have described a noninferiority hypothesis. Although the initial protocol stated that we intended to review both noninferiority and equivalence trials, this was amended to only include noninferiority trials. The hypothesis must be relevant to radiation therapy; studies in which the same dose/volume of radiation therapy was provided to all patients were excluded (eg, studies examining difference in concurrent systemic treatments). All forms of radiation therapy were included (eg, external beam radiation therapy, stereotactic radiation therapy, brachytherapy), except for radionuclide therapy.

Table 1 Summary of methodological and statistical reporting recommendations from the CONSORT 2010 extension for noninferiority and equivalence trials

Section/topic	Checklist item
Title	<ul style="list-style-type: none"> • Identification of the study as a noninferiority or equivalence trial
Introduction	<ul style="list-style-type: none"> • Rationale for a noninferiority study • Specification of a noninferiority margin with the rationale for its choice
Methods	<ul style="list-style-type: none"> • Description of trial design • Eligibility criteria • Description of interventions and whether the reference treatment is identical to that in any trial that established efficacy • Specify primary and secondary outcomes and whether hypotheses for each are noninferiority or superiority • Sample size calculation using a noninferiority criterion • Method used for randomization • Blinding details • Statistical methods, including whether a 1- or 2-sided confidence interval approach was used
Results	<ul style="list-style-type: none"> • For the primary noninferiority outcome, report results in relation to the noninferiority margin with measures of precision (eg, confidence intervals) • For outcomes for which noninferiority was hypothesized, a figure showing confidence intervals and the margin may be useful
Abbreviation: CONSORT = Consolidated Standards of Reporting Trials.	

Outcomes

Trials must have reported a clinical outcome (eg, survival, toxicity, or response to treatment). We excluded planning studies in which the primary outcome was a dosimetric quantity.

Two reviewers (AJA, VST) independently reviewed title and abstracts for eligibility of independent full-text review. A third reviewer (AVL) was available in case of discrepancies.

Data collection and analysis

One researcher (AJA) performed data collection and analysis. Data pertaining to study size, primary cancer type, type of comparison, endpoints, and statistical measures were collected. Descriptive statistics were performed to summarize data. Risk of bias assessment was not performed as study biases would not affect the outcomes of our review. A meta-analysis was not performed in line with the objective of this review. Covidence software was used for data management (Veritas Health Innovation, Melbourne, Australia).

Results

Of 423 records screened, 59 trials (14%) were included after full-text review. A diagram summarizing the screening and selection process is shown in Fig 2. Study characteristics are summarized in Table 2. The median number of participants was 486 (range, 40–4823). All studies were open label and were published after 2003. One trial (2%) was funded by industry exclusively, while 3 (5%) had joint funding from public and private sources. Four studies (7%) did not provide a rationale for a noninferiority design. The most common primary cancer type was breast ($n = 15$, 25%). The majority of studies ($n = 53$, 90%) had 2 treatment arms. Altered radiation fractionation ($n = 26$, 45%) and radiation de-escalation ($n = 11$, 19%) were the most common types of interventions. Nine studies (16%) compared radiation to another treatment modality (eg, surgery, radiofrequency ablation), and 8 studies (14%) examined the omission of radiation.

Endpoints and statistical data are summarized in Table 3. The most common primary endpoints were locoregional control ($n = 17$, 29%) and progression-free survival ($n = 14$, 24%). Fifty-three (90%) reported the noninferiority margin, and only 9 (17%) provided statistical justification for the margin based on previous clinical trials or published data. The median absolute noninferiority margin was 9% (interquartile range [IQR], 5%–10%), and the median relative margin was 1.51 (IQR, 1.33–2.04). Sample size calculations and CIs were reported in 54 studies (92%). Both intention-to-treat and per-protocol analyses were reported in 27 studies (46%).

In 31 trials (53%), noninferiority of the primary endpoint was reached. Authors concluded noninferiority in 34 trials (58%), and there was a discrepancy between the conclusion of noninferiority and statistical results in 3 studies (5%).

Discussion

In this systematic review of radiation noninferiority clinical trials, we found that the reporting of key methodological components was inconsistent. Noninferiority margins, CIs, and *P* values were not always reported, making it impossible to interpret results of these trials. Despite lacking the statistical rationale, a conclusion of noninferiority was claimed on the basis of inappropriate metrics in 3 studies. In light of these findings, we stress the importance of trialists reviewing CONSORT guidelines before the design of a noninferiority trial and reporting their data.

Selection of the noninferiority margin is the most important aspect in the design of a noninferiority trial as it is used to confirm or reject the hypothesis. A previous systematic review of noninferiority clinical trials of oncologic drugs showed that the median noninferiority margin was large at 12.5%.³ This is similar to the median noninferiority margin in our study of 9%. A larger noninferiority margin makes it easier to conclude noninferiority and can therefore be problematic if not appropriate. In contrast, a smaller margin would require a larger sample size to conclude noninferiority. Although reporting guidelines recommend that authors report the method to set the margin,¹ only a minority of studies (10%) in our review reported statistical justification for the noninferiority margin. The European Medicines Agency and Food and Drug Administration provide guidance on deciding the margin for trials involving drugs.^{5,6} The margin is statistically defined as the lower bound 95% CI of the standard treatment effect compared with placebo based on historic clinical trials. A more conservative margin can also be considered to account for differences between historic trial conditions and the current trial; the Food and Drug Administration suggests the noninferiority margin to be 50% of the lower bound 95% CI of the historic standard treatment effect. These guidelines are difficult to apply to trials involving treatments that are historically not compared with placebo, such as in radiation oncology. Without statistical justification for the noninferiority margin, many authors relied on expert opinion and stakeholder analyses alone to derive their margins. This was in keeping with trials of medical devices which rely on expert opinion to select a noninferiority margin.⁷

Furthermore, margins can be expressed as absolute (eg, 2% decrease) or relative values (eg, hazard ratio of 1.3). Many studies ($n = 27$, 51%) in our review reported only absolute margins. Absolute margins can bias toward

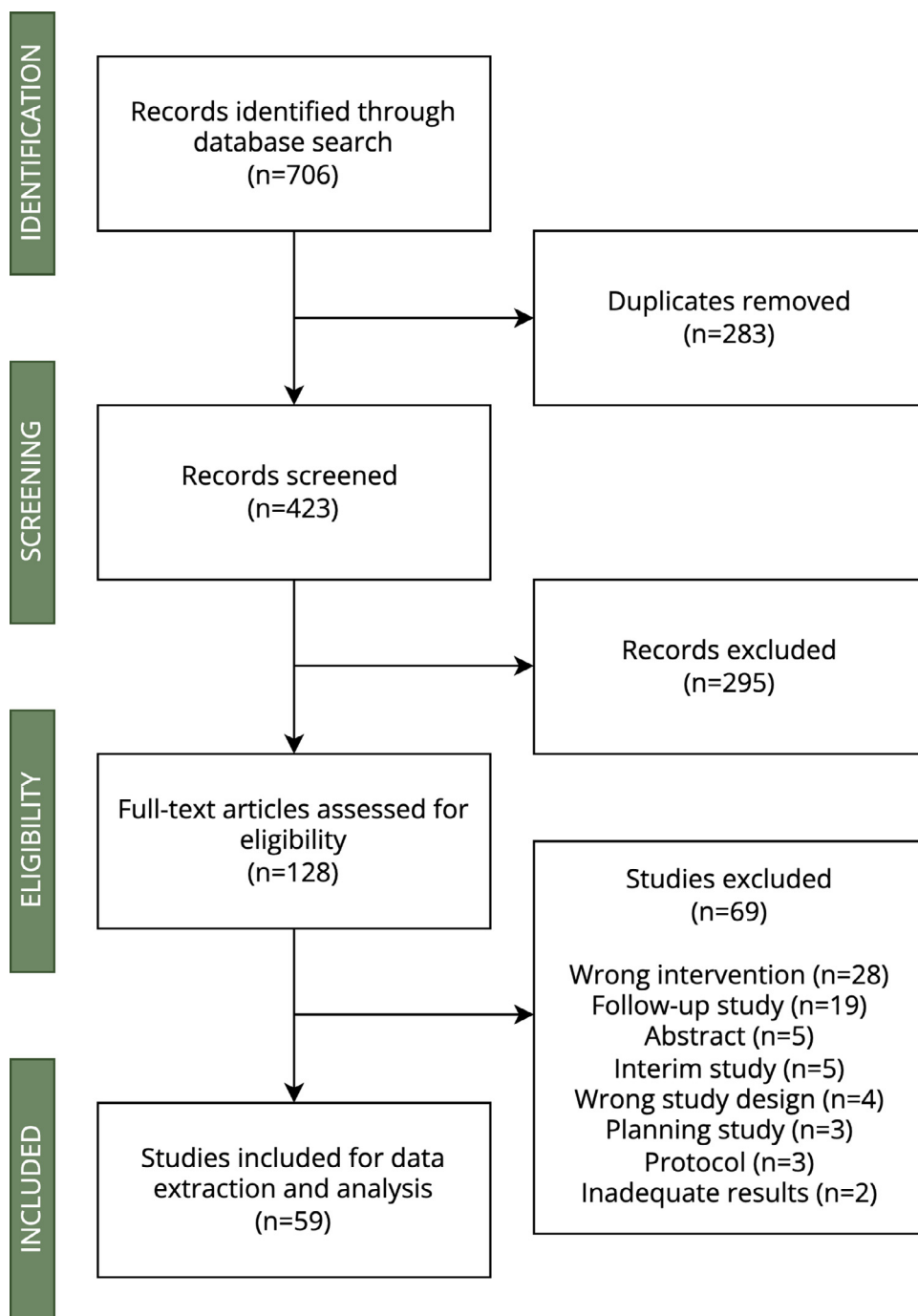


Figure 2 Summary of the screening and selection process.

noninferiority when event rates are lower than expected, whereas relative margins correspond to the same relative risk independent of event rates.⁸ A recent systematic review and meta-analysis of coronary stent noninferiority trials showed that the majority of trials only reported absolute margins (55 of 58, 94.8%), and the majority of those (n = 43) overestimated the control event rate, making the noninferiority margin more permissive.⁹ When the authors performed a reanalysis of the trials with adjusted margins, they found that 17 of the 50 trials

(34%) that met noninferiority using the absolute margin did not meet criteria using the relative margin. Absolute margins can be more practical as it increases power, but this is contingent on accurate control event rate estimation.

Previous reviews of noninferiority clinical trials in other settings have also found variability in reporting. A review of all noninferiority and equivalence trials published between 2003 and 2004 found that only 20.4% of studies provided justification for the noninferiority

Table 2 Summary of study characteristics

Characteristic	Value	No. (total = 59)	%
Date of publication	2000-2004	1	2
	2005-2009	3	5
	2010-2014	9	15
	2015-2019	26	44
	2020-2022	20	34
Country/region in which the study was performed	Australia	2	3
	Canada	2	3
	China	7	12
	Egypt	2	3
	Europe	10	17
	Japan	2	3
	United Kingdom	6	10
	United States	3	5
	Other*	3	5
	Multiple	22	37
Study funding	Government or academic	51	86
	Industry	1	2
	Mixed	3	5
	None/not specified	4	7
Blinding	None	59	100
Population	Adults	54	92
	Pediatrics	4	7
	Mixed	1	2
Primary cancer type	Breast	15	25
	CNS	3	5
	Gastrointestinal	4	7
	Gynecologic	2	3
	Head and neck	4	7
	Hematologic	10	17
	Prostate	8	14
	Multiple	9	15
	Other [†]	2	3
Purpose of conducting a noninferiority trial	Fewer adverse events	28	47
	More convenient	19	32
	Other	4	7
	Multiple	4	7
	Not specified	4	7
Number of treatment arms	2	53	90
	3	6	10

(continued on next page)

Table 2 (Continued)

Characteristic	Value	No. (total = 59)	%
Type of interventions	Radiation de-escalation	11	19
	Altered fractionation	26	45
	Alternate modality	9	16
	Omission of radiation	8	14
	Other [†]	4	7
Systemic therapy	Concurrent	4	7
	Sequential	11	19
	Optional	6	10
	Not allowed	33	56
	Study examined concurrent versus sequential systemic therapy	1	2
	Systemic therapy alone was a comparator	4	7
Radiation modality	Photon	52	88
	Proton	1	2
	Multiple/study compared modalities	6	10
Radiation technique	Field-based	14	24
	3D-CRT	4	7
	IMRT/VMAT	10	17
	SRS/SABR	1	2
	Not specified	8	14
	Multiple/study compared techniques	22	37
Fractionation	Conventional	19	32
	Hypofractionated	2	3
	Hyperfractionated	3	5
	Stereotactic	3	5
	Study compared fractionation schemes	32	54

Abbreviations: 3D-CRT = 3-dimensional conformal radiation therapy; CNS = central nervous system; IMRT = intensity modulated radiation therapy; SRS = stereotactic radiosurgery; VMAT = volumetric modulated arc therapy.

* Other countries/regions: Korea, India, Iran.

† Other primary cancer types: bladder, seminoma.

‡ Other types of interventions: delay of surgery after radiation, timing of radiation, difference in systemic therapy, difference in radiation volumes.

margin, and only 42.6% of studies reported both intention-to-treat and per-protocol analyses.² Most studies (n = 156, 96%) reported a prespecified noninferiority or equivalence margin. However, the authors were only able to adequately assess noninferiority and equivalence in 33 (20%) studies. Even among this small subgroup of studies, 4 reports (12%) misleadingly concluded noninferiority or equivalence. In a 2013 review of noninferiority trials involving oncologic drugs, the authors found that 62 of 75 studies (83%) reported a prespecified noninferiority margin.³ The authors found that the number of studies that did not report a noninferiority margin did not change after the publication of the CONSORT guidelines.

We found that 3 studies concluded noninferiority despite not reporting CIs of the primary endpoint. In

addition, some authors used statistically vague terminology such as “comparable” and “as effective” in concluding statements of trials in which noninferiority was not reached. This misleading reporting in clinical trials has been termed “spin.”¹⁰ A recent systematic review of oncologic noninferiority clinical trials that did not meet statistical significance for noninferiority showed that 75% had spin.¹¹ Compared with a previous review of spin, the authors reported the prevalence of spin in noninferiority clinical trials was higher than superiority clinical trials. Spin strategies included emphasizing trends for primary endpoints, conclusions based on secondary endpoints, or conclusions based on subgroup analyses. Spin was more likely associated with trials without for-profit funding, without data managers, and with novel treatments. The

Table 3 Summary of endpoints and statistical reporting

Characteristic	Value	No.	%
Primary endpoint	Progression-free survival	14	24
	Locoregional survival	17	29
	Disease-free survival	4	7
	Overall survival	8	14
	Toxicity	6	10
	Response (eg, pain response)	6	10
	Other	4	7
Were adverse events reported?	Yes	59	100
Was a noninferiority margin specified?	Yes	53	90
	No	6	10
Was statistical justification of the noninferiority margin specified?	Yes	9	17
	No	44	83
Was a sample size calculation performed and rationalized?	Yes	57	97
	No	2	3
Were confidence intervals reported?	Yes	54	92
	No	5	8
Confidence interval type	2-sided	18	33
	1-sided	12	22
	Not specified	24	44
Confidence interval size	97.5%	1	2
	95%	41	76
	90%	11	20
	Other: 91%	1	2
Was a <i>P</i> value reported?	Yes	56	95
	No	3	5
Type of analysis reported	ITT	22	37
	Modified ITT	2	3
	PP	8	14
	Both ITT and PP	27	46

Abbreviations: ITT = intention-to-treat; PP = per-protocol.

authors posited that trials with external funding were held to stricter standards, hence less likely to have spin. They also suggested that trials with novel treatments had higher spin because a negative trial could result in the treatment not becoming standard of care, or the report not being published. Authors should be cautious when making conclusions based on analyses outside of the primary endpoint as this could be easily misconstrued.

With the increasing frequency of noninferiority trials, clinicians should also be wary of bio-creep, a phenomenon that describes a situation in which an ineffective or even harmful treatment may be deemed effective.¹² This can happen when there is a series of noninferiority trials in which a new drug is slightly worse than another, and

this cycle may eventually lead to a drug that will eventually be ineffective or harmful compared with the original standard. For example, a new treatment B is found to be noninferior to treatment A and becomes the new standard of care. A subsequent trial uses treatment B as the active control against a new treatment C, which is found to be noninferior to treatment B. It would be wrong to conclude that treatment C is also noninferior to the original treatment A. Although this phenomenon has mostly been discussed theoretically, simulations have suggested that this is possible, but can be avoided by choosing an active control that has been compared with placebo, choosing an appropriate noninferiority margin, and accurately estimating the control event rate.¹³

To our knowledge, this is the first systematic review to examine the reporting quality of noninferiority clinical trials involving radiation therapy. Given the focused nature of this review, we were also able to describe radiation-specific details of the studies. Limitations include that our review focused on only English language articles, and that we did not assess the statistical rigor of the reported data as this was outside of the scope of this review.

Conclusion

There was variability in the reporting of key components of noninferiority trials including the noninferiority margin. Adherence to standards of data reporting and statistical methodology are important to ensure proper interpretation of trial results.

Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:[10.1016/j.adro.2023.101178](https://doi.org/10.1016/j.adro.2023.101178).

References

1. CONSORT Group. Reporting of noninferiority and equivalence randomized trials: Extension of the CONSORT 2010 statement. *JAMA*. 2012;308:2594-2604.
2. Le Henaff A, Giraudeau B, Baron G, Ravaud P. Quality of reporting of noninferiority and equivalence randomized trials. *JAMA*. 2006;295:1147-1151.
3. Riechelmann RP, Alex A, Cruz L, Bariani GM, Hoff PM. Non-inferiority cancer clinical trials: Scope and purposes underlying their design. *Ann Oncol*. 2013;24:1942-1947.
4. Page MJ, Moher D, Bossuyt PM, et al. PRISMA 2020 explanation and elaboration: Updated guidance and exemplars for reporting systematic reviews. *BMJ*. 2021;372:n160.
5. European Medicines Agency. *Guideline on the choice of the non-inferiority margin*. 2005. Available at; https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-choice-non-inferiority-margin_en.pdf. Accessed August 1, 2022.
6. US Food and Drug Administration. Non-inferiority clinical trials to establish effectiveness 2016. Available at: <https://www.fda.gov/media/78504/download>. Accessed August 1, 2022.
7. Lin CJ, Saver JL. Noninferiority margins in trials of thrombectomy devices for acute ischemic stroke: Is the bar being set too low? *Stroke*. 2019;50:3519-3526.
8. Kaul S, Diamond GA. Good enough: A primer on the analysis and interpretation of noninferiority trials. *Ann Intern Med*. 2006;145:62-69.
9. Simonato M, Ben-Yehuda O, Vincent F, Zhang Z, Redfors B. Consequences of inaccurate assumptions in coronary stent noninferiority trials: A systematic review and meta-analysis. *JAMA Cardiol*. 2022;7:320-327.
10. Boutron I, Dutton S, Ravaud P, Altman DG. Reporting and interpretation of randomized controlled trials with statistically nonsignificant results for primary outcomes. *JAMA*. 2010;303:2058-2064.
11. Ito C, Hashimoto A, Uemura K, Oba K. Misleading reporting (spin) in noninferiority randomized clinical trials in oncology with statistically not significant results: A systematic review. *JAMA Netw Open*. 2021;4: e2135765.
12. Everson-Stewart S, Emerson SS. Bio-creep in non-inferiority clinical trials. *Stat Med*. 2010;29:2769-2780.
13. Odem-Davis K, Fleming TR. A simulation study evaluating bio-creep risk in serial non-inferiority clinical trials for preservation of effect. *Stat Biopharm Res*. 2015;7:12-24.