Circulation

ORIGINAL RESEARCH ARTICLE

Noninferiority Designed Cardiovascular Trials in Highest-Impact Journals
Main Findings, Methodological Quality, and Time Trends

BACKGROUND: Noninferiority trials are increasingly being performed. However, little is known about their methodological quality. We sought to characterize noninferiority cardiovascular trials published in the highest-impact journals, features that may bias results toward noninferiority, features related to reporting of noninferiority trials, and the time trends.

METHODS: We identified cardiovascular noninferiority trials published in *JAMA*, *Lancet*, or *New England Journal of Medicine* from 1990 to 2016. Two independent reviewers extracted the data. Data elements included the noninferiority margin and the success of studies in achieving noninferiority. The proportion of trials showing major or minor features that may have affected the noninferiority inference was determined. Major factors included the lack of presenting the results in both intention-to-treat and per-protocol/as-treated cohorts, $\alpha > 0.05$, the new intervention not being compared with the best alternative, not justifying the noninferiority margin, and exclusion or loss of ≥10% of the cohort. Minor factors included suboptimal blinding, allocation concealment, and others.

RESULTS: From 2544 screened studies, we identified 111 noninferiority cardiovascular trials. Noninferiority margins varied widely: risk differences of 0.4% to 25%, hazard ratios of 1.05 to 2.85, odds ratios of 1.1 to 2.0, and relative risks of 1.1 to 1.8. Eighty-six trials claimed noninferiority, of which 20 showed superiority, whereas 23 (21.1%) did not show noninferiority, of which 8 also demonstrated inferiority. Only 7 (6.3%) trials were considered low risk for all the major and minor biasing factors. Among common major factors for bias, 41 (37%) did not confirm the findings in both intention-to-treat and per-protocol/as-treated cohorts and 4 (3.6%) reported discrepant results between intention-to-treat and per-protocol analyses. Forty-three (38.7%) did not justify the noninferiority margin. Overall, 27 (24.3%) underenrolled or had >10% exclusions. Sixty trials (54.0%) were open label. Allocation concealment was not maintained or unclear in 11 (9.9%). Publication of noninferiority trials increased over time ($P<0.001$). Fifty-two (46.8%) were published after 2010 and had a lower risk of methodological or reporting limitations for major ($P=0.03$) and minor factors ($P=0.002$).

CONCLUSIONS: Noninferiority trials in highest-impact journals commonly conclude noninferiority of the tested intervention, but vary markedly in the selected noninferiority margin, and frequently have limitations that may impact the inference related to noninferiority.

Behnood Bikdeli, MD, MS*

John W. Welsh, BA*

Yasir Akram, MD, MPH

Natdanai Punnanithinont, MD, MPH

Ike Lee, BA

Nihar R. Desai, MD, MPH

Sanjay Kaul, MD

Gregg W. Stone, MD

Joseph S. Ross, MD, MHS

Harlan M. Krumholz, MD, SM

Clinical Perspective

What Is New?

- Noninferiority cardiovascular trials are increasingly being published by the highest-impact journals.
- Most of these trials concluded that the tested interventions were noninferior to the compared therapy.
- Many trials had methodological or reporting limitations that may impact confidence in these conclusions.

What Are the Clinical Implications?

- Given the increasing utilization of noninferiority design, the advantages and limitations of this type of trial design deserve greater attention in the cardiovascular community.
- Better adherence to methodological and reporting standards for noninferiority designed trials can improve the confidence of the community in their findings.

Noninferiority trials have a distinctly different objective than superiority trials.[1] These trials are designed to determine whether a new intervention is not less effective than the standard of care as measured using a prespecified difference in efficacy determined by the investigators.[1,2] Historically, physicians only adopted a new therapy if it had efficacy advantages. However, with the introduction of the noninferiority concept, if there are other benefits in convenience, cost, or safety while the efficacy is not materially worse (noninferior) than the standard of care, a new intervention may be considered as a reasonable alternative treatment.[1,2]

The use of noninferiority design has increased in recent years.[3] Noninferiority trial design has specifically appealed to funders and investigators in cardiovascular medicine because of the limitations with several standard-of-care cardiovascular therapeutics. For example, although effective interventions existed for stroke prevention in atrial fibrillation and for valvular disease, the use of these therapies was compromised by burdensome logistics, or safety and tolerability concerns.[4,5] Consequently, pivotal noninferiority trials led to the approval of non–vitamin K antagonist oral anticoagulants for patients with nonvalvular atrial fibrillation,[6–8] and percutaneous techniques for some valvular diseases, as well.[9]

Noninferiority trials, however, possess distinct vulnerabilities that may undermine their validity, what would be considered sources of bias.[10] In superiority trials, methodological limitations such as high rates of postenrollment exclusions, loss to follow-up, or low treatment adherence lead to inadequate power, thereby biasing the trial results toward the null (negative findings). Investigators seeking to test a superiority hypothesis are strongly incentivized to guard against these problems. However, for a noninferior-

ity trial, these same factors might increase the likelihood that the trial will find no difference between treatment arms, potentially resulting in a false claim of noninferiority. Another important distinction, unlike superiority trials that are tested with fairly uniform statistical criteria, positive or negative results from noninferiority trials are highly dependent on the investigator-defined criteria for noninferiority, which may vary substantially.[1,11–14] Several other factors related to reporting of noninferiority trials may also impact the interpretations by the scientific community and regulatory agencies.[13]

We sought to determine the number, main findings, and features causing methodological or reporting limitations related to noninferiority design among cardiovascular noninferiority trials published between 1990 and 2016 in 3 highest-impact medical journals: *Journal of the American Medical Association* (*JAMA*), the *Lancet*, and the *New England Journal of Medicine* (*NEJM*). Publications in these journals, on average, have more rigorous assessment standards and frequently address the clinical questions with patient-important outcomes.[15] We made use of recently published criteria for assessment of the risk of methodological or reporting limitations for noninferiority trials.[2,13,14]

METHODS

The data used for this study will be available for interested research investigators, after submitting a full study protocol (to B.B. or H.M.K.) and mutual agreement. No patient interview or access to medical records was needed in this study. The data were extracted from publicly available published literature. As such, obtaining an institutional review board approval was not required.

Search Strategy and Study Selection Criteria

We used a sensitive search strategy in PubMed to identify cardiovascular randomized trials, from January 1, 1990 to December 31, 2016, published in *JAMA*, the *Lancet*, and *NEJM* (see the online-only Data Supplement Appendix). The initial time point was selected by consensus and thought to capture all existing noninferiority cardiovascular trials.

We manually screened all the retrieved studies to identify cardiovascular trials that tested for noninferiority for a primary clinical end point. This choice allowed us to keep optimal sensitivity to include all noninferiority trials, including those with a less clear statement of noninferiority design in the title or abstract, including the BARI trial (Bypass Angioplasty Revascularization Investigation) and the EVEREST II trial (Endovascular Valve Edge-to-Edge Repair Study II). For trials with factorial design or papers that reported >1 trial, we did 1 abstraction per each primary hypothesis of noninferiority. We excluded substudies, post hoc analyses, and follow-up studies that were not the main report of a randomized trial. We also excluded trials related to diseases of the lymphatics, systemic vasculitides, and obesity trials that did not use a primary clinical cardiovascular end point.

Assessment of the Noninferiority Criteria

We investigated whether a trial achieved noninferiority, both noninferiority and superiority, had inconclusive results, or showed statistical inferiority, and categorized the studies based on their point estimates and confidence intervals.[2,13] If the studies had coprimary end points, or multiple active arms (eg, for various doses[8]), we reported the most favorable results. We recorded the noninferiority margin in each trial, whether it was based on absolute risk difference between the groups or relative terms (hazard ratios, relative risk, or odds ratios), and whether justification was provided for the noninferiority margin (preservation of a fraction of efficacy of standard of care versus placebo, the so-called putative placebo analysis, or at least as acceptable difference stated by clinical consensus among investigators).[13] We determined the reported power for detection of noninferiority, and the calculated and actual study sample size, as well.

Assessment of the Methodological Quality and Methodological or Reporting Limitations

Based on coauthors consensus after review of the existing recommendations, including the CONSORT statement (Consolidated Standards of Reporting Trials) for noninferiority trials,[2,13] we identified a set of characteristics to assess the quality of noninferiority trials and factors that posed major or minor limitations for methodology or reporting of noninferiority trials.[2,13,14] The group of authors included experts in outcomes research, clinical trial design, and regulatory supervision of clinical trials. Major methodological or reporting limiting factors included the following: Reporting the primary end point results based on only the intention-to-treat or modified intention-to-treat, or per-protocol/as-treated cohorts, choosing an α-level >0.05, new intervention not being compared with the best alternative, lack of a valid justification (putative placebo analysis, or consensus on therapeutic interchangeability) for the noninferiority margin, and underenrollment, early loss-to follow-up, early termination of the intervention, or other reasons for exclusion of >10% of the population from the study cohort. For trials that reported the primary end point in 2 separate analyses of intention-to-treat and as-treated (or per-protocol) cohorts, we determined whether results were concordant.[16] Minor methodological or reporting limiting factors included no blinding of patients and site investigators,[13] lack of blinded adjudication of the outcomes, lack of allocation sequence concealment, not mentioning of the noninferiority design in the study title or abstract (many times the only part of an article read by physicians),[13] and the absence of a methods paper for the study (or lack of simultaneous publication of the study protocol as a supplement). We considered studies that did not meet any of the major or minor limitations to be at low risk.

Other Variables

We determined the type of the tested intervention (drug, device, or other) and the types of ancillary advantages of the new intervention in comparison with standard of care (cost advantages, lower treatment burden, and presumed patient-important benefits such as quality-of-life improvements). We recorded whether a study was prematurely discontinued. We

assessed, in the abstract of the included trials, whether there were discrepancies between claims of noninferiority reported by the authors, and the data to support such claims.

Analysis of Trends

We determined the temporal trends in publication of noninferiority trials. The Food and Drug Administration (FDA) released a guidance statement about noninferiority trials in 2010 (followed by an update in 2016).[14] We explored if the risk of bias among included studies changed before versus after the end of 2010.

Data Extraction and Statistical Analysis

Four authors worked independently to extract the data. In case of discrepancy, the full text was revisited by the lead authors (B.B. and J.W.), and, if the discrepancy was unresolved, the case was discussed with 2 coauthors (J.S.R. and H.M.K.). We described normally distributed variables as means (SD) and nonnormally distributed variables as medians (interquartile ranges [IQR]). We compared normally distributed data using t test and used related counterparts for not normally distributed data, where needed. We described categorical variables as frequencies and compared them by using χ^2 tests, also repeating the results with an exact test. Analyses were performed using Stata Version 12.0 (StataCorp LP).

RESULTS

Our PubMed search resulted in 2544 eligible publications, of which 110 (including 111 trials) met the entry criteria (Table 1 and Table I in the online-only Data Supplement). Studies were published between 1992 and 2016. The majority (n=78, 70.3%) justified the noninferiority design based on presumed ancillary benefits for patient outcomes other than the primary end point, a few had reduced treatment burden (n=6 [5.5%], eg, minimally invasive surgery rather than open surgery) or direct cost advantages (n=4, 3.6%). The remainder had more than one presumed ancillary benefit. Nine (8.1%) trials tested noninferiority of a new intervention against placebo or no additional treatment.

The median number of randomly assigned patients was 3006 (IQR, 1021–6068). The median calculated power to detect noninferiority was 86% (IQR, 80%–90%). The median number of patients entered in the investigator-defined primary end point analysis was 2707 (IQR, 1021–5966). Nine studies (8.1%) were prematurely discontinued (4 for safety, 3 for slow enrollment, 1 for futility, and 1 because of the compromised integrity).

Noninferiority Margin

The noninferiority margins were based on absolute risk difference in 60 (54.0%) trials, and on relative measures in 50 (45.0%) trials (29 based on hazard ratios, 14 based on relative risk, and odds ratios in the remaining

Table 1. Basic Study Characteristics of Included Noninferiority Trials

Studied Characteristics	Trials (N=111)
Journal, n (%)	
NEJM	66 (59.5)
JAMA	30 (27)
The Lancet	15 (13.5)
Study field, n (%)	
General cardiology/ interventional cardiology/ cardiac surgery	66 (59.5)
Stroke prevention or treatment	15 (13.5)
VTE (prevention or treatment)	30 (27.0)
Number of patient-enrolling sites, median (IQR)	110 (19–381)
Locations of enrolling sites, n (%)	
United States and Canada only	18 (16.2)
International only	35 (31.5)
Combination	58 (52.3)
Trial intervention, n (%)	
Drug	66 (59.5)
Device	31 (28)
Other	14 (12.5)
Source of funding, n (%)	
Government	8 (7)
Nonprofit foundations	8 (7)
Private industry	79 (71.2)
Any combination	14 (12.6)
No disclosure	2 (1.8)
Advantage of the tested intervention, n (%)	
Lower direct costs	4 (3.6)
Reduced treatment burden	6 (5.4)
Potential benefit on patient outcomes	81 (73.0)
Any combination	19 (17.1)
A placebo arm or no-intervention arm included, n (%)	10 (9)

IQR indicates interquartile rang; *JAMA*, *Journal of the American Medical Association*; *NEJM*, *New England Journal of Medicine*; and VTE, venous thromboembolism.

7). One trial's margin was based on relative difference but not further classified,[16] and one trial did not report the selected margin.[17] The noninferiority margins in the trials varied widely (absolute risk differences between 0.4% and 25%, hazard ratios between 1.05 and 2.85, relative risks between 1.1 and 1.8, and odds ratios between 1.1 and 2.0).

Among trials with a published design paper/study protocol, we identified discrepancies or missing information between the design paper/protocol and the final published paper in 7 cases. These included examples of changing the noninferiority margin in the final publication, and the absence of some or all the details related to the chosen noninferiority margin in the final publication (Table II in the online-only Data Supplement).

Results for the Primary End Point

From the 111 trials, in 2 trials, although the overall results appeared favorable, the 2 bounds of the CIs could not be assessed. Of 109 eligible trials, 86 (78.9%) claimed noninferiority (20 of which also showed superiority), whereas 23 (21.1%) did not (16 that had inconclusive results and 7 that showed worse outcomes with the tested intervention; Figure 1). In 95 (85.6%) trials, the primary end point analyses were based on intention-to-treat or modified intention-to-treat, 11 (9.9%) trials used a per-protocol analysis, and 5 (4.5%) used a different or undefined analysis type. The vast majority (90%) of the trials appropriately reported the findings in the abstract. They claimed noninferiority, superiority, or inferiority, where they were achieved. Similarly, they mentioned lack of noninferiority when it was not achieved. However, on a few occasions, there was either discrepancy about the abstract claims and the study findings, or insufficient information was provided (Table III in the online-only Data Supplement).

METHODOLOGICAL AND REPORTING LIMITATIONS

Table 2 and Figure 2 demonstrate the frequency with which the trials met the major and minor limitations. Seventy trials (63.1%) presented their results for the primary end point using a second analysis of a different cohort (eg, per-protocol, if the primary analysis was intention-to-treat), the results were consistent in 66 trials, whereas in 4 examples the primary end point results were not consistent in the second analysis.

Sixty-eight trials (61.3%) provided justification for the noninferiority margin, whereas 43 (38.7%) did not. In 110 (99.0%) trials, the one-sided α was ≤0.05. Of these 110, in 44 trials the one-sided α was also ≤0.025. A post hoc analysis suggested smaller α-levels for drug trials than for device trials (P=0.02 for the Wilcoxon rank-sum test). The intervention was tested against the best available alternative in all trials. In 27 studies (24.3%), >10% of the participants were lost early, including because of loss to follow-up, early discontinuation of the intervention, or other reasons (Table 2). Overall, 27 trials (24.3%) were considered at low risk of bias according to all the major biasing factors.

The noninferiority design was reported in the title or abstract in 97 trials (87.3%). For 67 (60.4%) trials, either a design paper was published or the protocol was published along with the manuscript. Among these 67 trials, in 8 (11.9%), the noninferiority margin was discrepant or not disclosed in either the manuscript or the design paper (protocol). The majority of the trials (n=60, 54%) were open label; 12 (10.8%) were single blind, and 39 (35.1%) were double blind. Allocation concealment was maintained in 100 trials (90%). Blinded adjudication of the primary end point was performed in 91 trials (82%). Twenty-five trials (22.5%) were considered at low risk of bias according to all the minor factors (Figure 2). Collectively, 7 trials (6.3%) were considered at low risk of bias according to all major and minor criteria.

A post hoc analysis did not show a difference for major (P=0.28 for the Kruskal-Wallis test) or minor factors (P=0.11
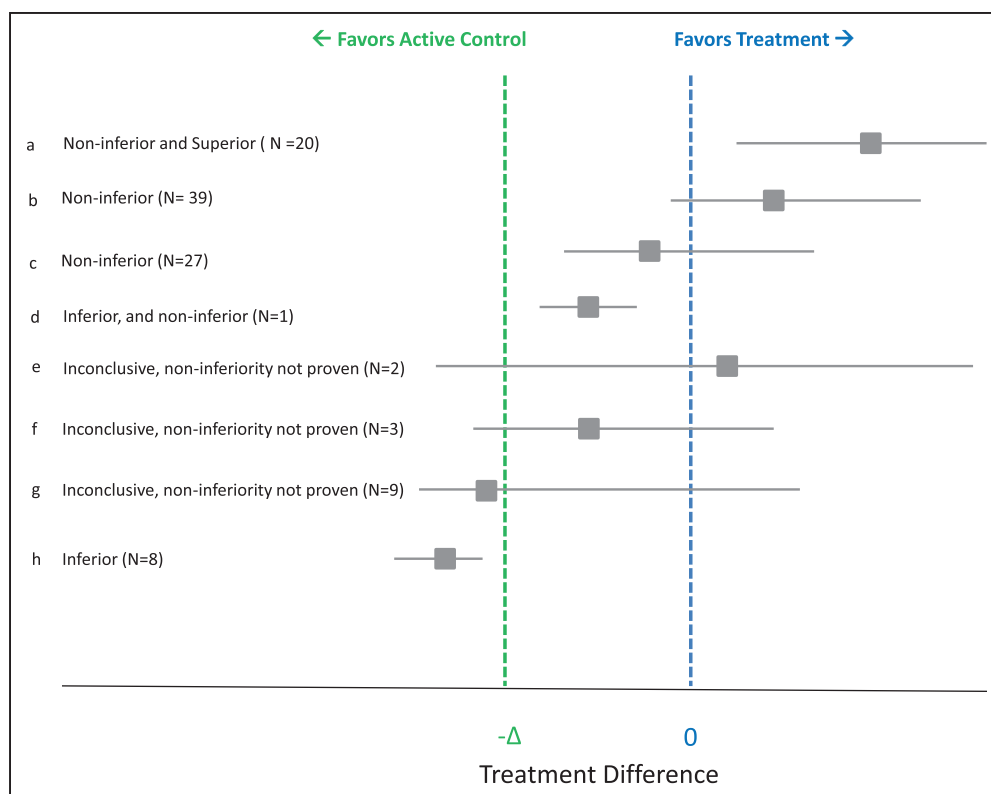
Figure 1. Results of primary end points in included studies.
Effect size for included studies is shown as absolute risk difference (ARD), odds ratio (OR), relative risk (RR), or hazard ratio (HR) according to the original studies. The line of no difference denotes 0 for ARD, and 1 for OR, RR, or HR. Delta reflects the margin the trialists chose for noninferiority in each study. If the 95% CIs fall above the line of no difference, the tested intervention is noninferior and superior, if the CIs are not crossing −Δ, the intervention is noninferior, and if both CIs are below −Δ (the noninferiority margin), the tested intervention is inferior. a, The intervention was not only noninferior, but also superior; b, noninferior with favorable point estimate; c, noninferior with worse point estimate; d, statistically inferior but noninferior per investigators' criteria; e through g, inconclusive results; h, inferior but not noninferior.

for the Kruskal-Wallis test) between the 3 journals. However, post hoc assessment of availability of a design paper or study protocol for noninferiority trials showed that 74.2% of studies in *NEJM* had a published methods paper or online available study protocol, in comparison with 53.3% of trials in *JAMA*, and 33.3% of the trials in the *Lancet* (*P*=0.001 for the Fisher exact test).

Temporal Trends

Over time, there was an increase in cardiovascular noninferiority trials published in these journals (*P*<0.001 for trend, Figure 3). A post hoc analysis showed that, among the 111 included trials, 52 (46.8) were published after 2010. Of these, 21 (40.3%) did not disclose sufficient justification for the chosen noninferiority margin and 10 (19.2%) did not replicate the findings in both intention-to-treat and as-treated cohorts, 9 (17.3%) did not have a published design paper or the study protocol as an online supplement to the final manuscript, 27 (52.9%) were open label, 5 (9.6%) did not adhere to allocation sequence concealment recommendations, 10 (19.2%) did not adhere to blinded outcome adjudication, and 2 (3.85%) did not report the noninferiority design in the title or abstract. As such, trials published after 2010 represented a lower risk of methodological

or reporting limitations according to major, and minor factors, as well (*P*=0.03 and *P*=0.002, respectively, for the Wilcoxon rank-sum test of limiting factors across trials before versus after 2010).

DISCUSSION

Noninferiority trials in cardiovascular medicine are increasingly being published in the highest-impact journals and most commonly conclude noninferiority of the new intervention tested. Most of these trials were large, multicentered, and adhered to several of the quality metrics we evaluated. This finding is particularly important given that many of the noninferiority trials have been the basis of approval for new treatments.[6–9,18–21] However, we noted important methodological limitations, including the use of wide noninferiority margins, sizeable underenrollment or postrandomization exclusions, and absence of confirmatory analyses or discordant results between the intention-to-treat versus astreated analyses. Over one-third of trials (39.6%) did not have a published methods paper or study protocol. Of those with a methods paper or protocol, 1 in 8 had discrepancies related to the noninferiority margin, in comparison with the published manuscript. Several tri-

Table 2. Factors Related to Risk of Bias Across the Included Studies

Factors	Relevance	Definition	Trials at Risk of Bias, n (%)
Major factors			
Chosen population for primary analysis	Noninferiority trials are prone to specific biases. Although intention-to-treat most accurately reflects the outcomes in all patients with the treatment as a strategy, there would be limitations in noninferiority design when many patients do not receive the treatment or if there is significant crossover (causing false claims of noninferiority). A per-protocol or as-treated population may provide more reliable data regarding safety of a new therapy, but also efficacy in non-inferiority trials. Ideally, the primary results should be reported for these 2 different populations.	Analysis would be adequate if the primary end point is analyzed in (1) the intention-to-treat or modified intention-to-treat population, and 2) either the per-protocol or as-treated population. The analysis plan and the definitions of these populations should be prespecified in the protocol or statistical analysis plan.	41 (36.9%)
Justification for noninferiority margin	Contrary to superiority trials, findings of noninferiority trials are fundamentally dependent on investigator-defined noninferiority margins. Studies may report sufficient statistical justification, clinical judgment justification only, or provide no information. Unreasonably wide margins that are not justified could potentially lead to claims of noninferiority for interventions that might be worse than the standard of care.	The noninferiority margin is considered sufficiently justified if it is selected based on preservation of at least a major portion of the efficacy of standard of care versus placebo or no treatment (typically 50%). The noninferiority margin is also considered justified if there is explicit discussion of the clinical judgment deciding on the therapeutic interchangeability.	43 (38.7%)
One-sided type I error	In noninferiority trials, the investigators already accept a reasonable margin of reduced efficacy (noninferiority margin). It remains crucial that stringent criteria are followed to avoid additional sources of bias that could lead into potentially false claims of noninferiority.	One-sided α of ≤0.05 would be considered acceptable. One-sided α of 0.025 (consistent with a 2-sided 95% CI) is preferable.	1 (1.0%)
New intervention being compared with the best available alternative	Declaring noninferiority to an intervention that is not the standard of care or the best available option is not appropriate.	Design is acceptable if the new intervention is compared against standard of care (typically the best available option for the studied condition).	0
Underenrollment or early postrandomization exclusions >10% of the cohort	Underenrollment, or exclusion of a large proportion of patients, could undermine the credibility of the results. Specifically, if absolute risk is used as the effect measure, unanticipated and unaccounted for underenrollment or postrandomization exclusions could lead to low event rates and, therefore, false claims of noninferiority.	Underenrollment or unanticipated and unaccounted for postrandomization exclusions rate of <10% of population from the final analysis (including exclusions immediately after enrollment, or loss to follow-up) when using absolute risk for noninferiority margin is desirable.	27 (24.3%)
Minor factors			
Declaration of noninferiority design in the title or the abstract	The abstract is the most visible part of a manuscript and many times the only part of an article read by physicians.	Reporting of noninferiority design either in the title or the abstract would be considered adequate.	14 (12.6%)
A design paper published or the study protocol available with the full text	Publication of a design paper provides the opportunity to describe the study plan typically in greater detail than the pivotal manuscript, which allows critical review by other investigators.	Publication of a design paper before publication of the main study would be considered adequate. Alternatively, the detailed study protocol can appear in a Supplemental Appendix of the main manuscript.	44 (39.6%)
Allocation concealment	Allocation concealment helps prevent selection bias related to intervention assignment. It protects the allocation sequence before and until assignment, and can always be successfully done regardless of the type of intervention.	A clear statement about concealing the allocation of the patients would be considered adequate.	11 (9.9%)
Blinding	Blinding attempts to prevent performance bias (eg, provision of dissimilar ancillary therapies) and detection/ascertainment bias (provision of dissimilar tests) after assignment.	Single-blind or double-blind studies would be considered adequate.	60 (54.0%) open label, 12 (10.8%) single blind
Blinded adjudication of outcomes	Even for device trials where blinding is difficult to achieve, blinded adjudication of events is usually feasible.	Clear description of blind adjudication of outcomes would be considered adequate.	20 (18.0%)

als suffered from a constellation of factors that could predispose the results toward noninferiority, potentially undermining the confidence in the trial results.

Noninferiority trials are inherently more susceptible to limitations and bias and are more difficult to con-duct and interpret,[22] in comparison with superiority trials. First, most biases applying to superiority trials, the open-label design, lack of allocation sequence concealment, and blinded adjudication can similarly undermine the noninferiority trials. Second, many
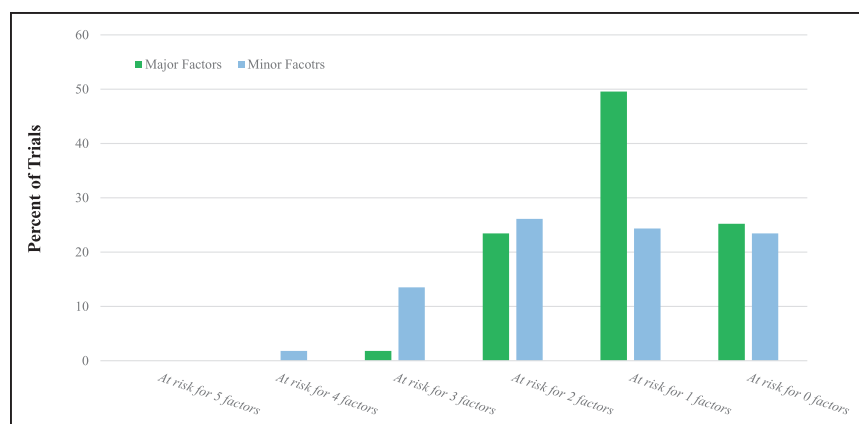
Figure 2. Distribution of trials according to limitations for major factors and minor factors.

specific biases can impact noninferiority trials and compromise their findings. Use of large type I error margins leads to claims of noninferiority for interventions that might be meaningfully worse than the standard of care. The intention-to-treat principle keeps the integrity of randomization. However, in noninferiority trials, early discontinuations, poor treatment adherence, crossovers, and other exclusions (regardless of why they happen)[22] may lead to false claims of noninferiority,[23] especially if absolute risk difference is used as the effect measure.[24] The per-protocol (or as-treated) analyses address these issues but are limited by breaking the randomization. Concordant results from the intention-to-treat and per-protocol analyses are reassuring, but the discrepancy of the 2 analyses[25] raises uncertainty around a claim of noninferiority. Use of absolute risk difference as the effect measure is also associated with increased risk of falsely claiming for noninferiority, especially when there are a lower number of events than expected.[26] Use of relative effect measures (odds ratios, relative risk, and hazard ratios), in general, are more helpful, but could be problematic if the actual event rate far exceeds the expected event rate, which may rarely happen. Ideally, the use of both absolute and relative measures is preferable.[27]

Because noninferiority trials heavily rely on the investigator-defined margin of noninferiority, the choice of the margin and adequate justification of such a margin would be critical.[2,13,14,28] The noninferiority margin should be specified based on both statistical reasoning and clinical judgment. Choosing an extremely wide margin may lead to a claim of noninferiority for an intervention that is markedly worse than the standard of care. In contrast, choosing an extremely stringent margin may lead to failure to detect noninferiority for an intervention with acceptable efficacy and the potential to reduce costs or adverse events. Appropriate noninferiority margins are typically chosen based on putative placebo analysis, or as a clear explanation of expert consensus where robust historical data are not available. Clear reporting of such information would help the readers to better understand the clinical relevance and trade-offs. In our study, the majority of the included trials provided adequate justification of the noninferiority margin,[7] whereas the information was missing or insufficient in some cases. We identified a few cases of discrepancy for the inferiority margin between the final publication and the previously prepared design paper/protocol. Although such discrepancies did not lead to critical changes in the trial conclusion in most cases, our sample of trials in the highest-impact journals likely represents a best case scenario of the cardiovascular literature. It is possible that other noninferiority trials published in other journals are prone to more prominent
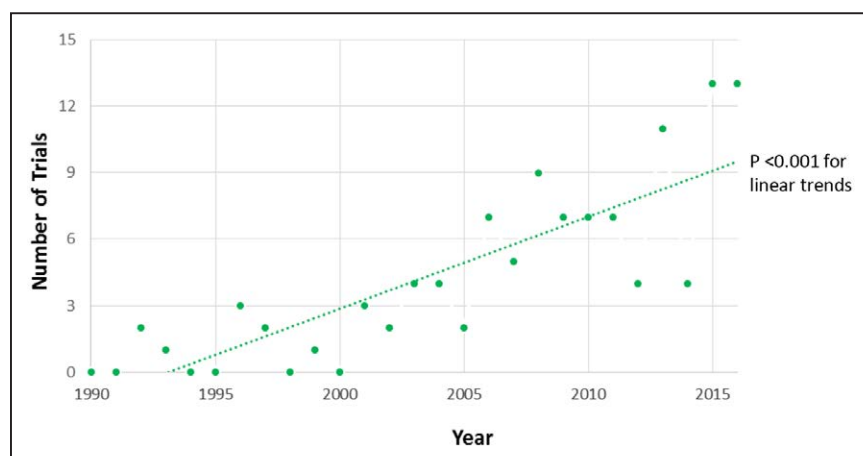

Figure 3. Trend in publication of cardiovascular noninferiority trials over time.

discrepancies between a design paper/ protocol and the final publication. If various practical reasons warrant a change in the predefined noninferiority margin, this should be clearly and thoroughly discussed by the trialists in the final publication.

The increased publication of noninferiority trials over time may be indicative of the shift in research priorities. We hypothesize that for some cardiovascular conditions, a ceiling effect for efficacy with current treatment paradigms might have been reached. However, because of costs, convenience, or safety concerns, the investigators and funders shift the attention toward alternatives with similar efficacy but with ancillary benefits. Of note, we identified a lower risk of bias for trials in recent years, concurrent with release of the updated version of the CONSORT statement and the FDA guidance statement for noninferiority trials.[13,14] The observed trend of the published trials may have better fitted a nonlinear or segmented regression model. However, because of the post hoc nature of this observation, we refrained from changing the regression model to avoid overfitting.

The ancillary benefits for some of the interventions were obvious (eg, outpatient treatment rather than inpatient treatment, or minimally invasive surgery rather than open surgery). For others, the benefits were presumed or less clear. The approval of additional alternatives could potentially put more pressure on the entire class of a given intervention, thereby making the intervention more affordable. Ideally, a formal assessment of the superiority for the ancillary benefits should be considered.[29]

In the design of our study, we consulted with the guidelines from the FDA for assessment of noninferiority trials. In fact, our studied variables were similar to what was explored in FDA panels related to noninferiority trials in other fields.[30] Based on our investigation, we propose careful attention to several methodological and reporting factors for trialists and readers of noninferiority trials (Table 3). Assessment of the information submitted to clinicaltrials.gov is another interesting topic that has been addressed in other studies.[12]

Our decision to investigate the studies published in highest-impact journals needs further explanation. This decision was, in part, attributable to practical considerations given the time-consuming nature of manual screening of the articles and extraction of several data elements that required assessment and interpretation. In addition, it is expected that studies published in *NEJM/Lancet/JAMA* are generally larger, and address clinically important (rather than surrogate) outcomes. Many of the high-quality practice-changing cardiovascular trials get published in these journals.[31] To further explore this issue, we conducted a search of the 3 high-impact cardiology journals (*Circulation*, *European Heart Journal*, and *Journal of the American College of Cardiology*). During the same study period (1990–2016), we identified 10 694 citations, and 329 were manually screened further for noninferiority or equivalence trials. Ultimately, 92 relevant trials were identified in those cardiology journals. The median (IQR) number of patients in studies published in *NEJM/Lancet/JAMA* was significantly higher than studies published in the 3 cardiology journals (2707 [IQR,1021–5966] versus 628 [IQR, 300–1400], *P*<0.0001). Furthermore, contrary to noninferiority trials in *NEJM/Lancet/JAMA*, several of the noninferiority trials in the 3 cardiology journals used primary surrogate end points (rather than clinical end points) for testing the noninferiority of the primary end point. In fact, other studies have suggested that methodological and reporting characteristics of studies published in those specialty journals are similar, if not worse, in comparison with *NEJM/Lancet/JAMA*.[31]

Our study has some limitations. First, in assessing the risk of methodological or reporting limitations across the studies, we considered the best results from the trials with multiple arms or coprimary end points. Therefore, our findings likely represent the best-case scenario for the noninferiority cardiovascular trials. Second, there were additional factors that we would have preferred to look into, such as the use of run-in phase, details about crossovers, and treatment adherence (especially in the control group). However, these measures were inconsistently reported across the trials. Third, we acknowledge that some of the minor factors we studied (such as reporting the noninferiority design in the title or abstract) do not necessarily represent bias. However, it is important that a noninferiority trial is clearly conveyed as such. For this reason, such factors are included in the CONSORT criteria for assessment of risk of bias, and we included them in our study. Third, there might be variations among investigators for the definitions and operationalization of some of the quality metrics that we assessed. We tried to address this challenge by incorporating a diverse team inclusive of investigators with expertise in outcomes research, clinical trial design and methodology, and regulatory aspects for noninferiority trials. Fourth, a plethora of other factors (including changing the definition of the primary end point from protocol to publication, changing the noninferiority margin over time, or changing the prespecified analytical plan of the trial, and the difference between the estimated and actual power of the study for detection of noninferiority) would be additional informative aspects that could shed light on methodological features and risk of bias among noninferiority trials. Although we did not formally assess such factors in the current study, they could be subject to future investigations. Fifth, our study does not include noninferiority trials published in 2017 or 2018. Although the addition of more

Table 3. Key Methodological and Reporting Factors to Consider for Noninferiority Trials

Methodological and Reporting Factors	Comment
Design and conduct	
One-sided type I error (α) should be ideally set at 0.025 or lower	This will correspond to a 2-sided type I error of 0.05.
Thinking of strategies to improve treatment adherence and minimize crossovers	If there is a concern for notable crossover or variable treatment adherence, the primary end point should be tested with both absolute and relative effect measures.
The new intervention should be tested against the best available alternative or standard of care. Assume that intervention A is the standard of care. Studies may show that a new intervention B is noninferior to A. Subsequently, a second study may show noninferiority of an intervention C to B. However, this is insufficient, because we do not know if C is noninferior to A. This phenomenon is referred to as noninferiority creep, and should be avoided.	The exception is safety noninferiority trials in which an intervention with promise for certain efficacy end points is tested for a primary safety outcome against no intervention or placebo.
Attempting to ensure blinding, and at least allocation sequence concealment, as well-blinded outcome adjudication.	Blinding is ideal but not always feasible. However, allocation sequence concealment is always feasible. Blinded adjudication should also be feasible in most scenarios.
Selecting the noninferiority margin based on clinical justification, statistical justification, or both, at the design phase. Also, ideally planning to show the superiority of the tested intervention for the ancillary benefit(s).	Because, in noninferiority trials, a margin of reduced efficacy is allowed, it would be valuable to show that the ancillary benefits of the tested intervention are evident, rather than presumed. If the ancillary benefits are evident (eg, minimally invasive procedure vs open surgery, or home treatment vs in-hospital treatment), this part could be skipped.
Analysis	
Reporting the results in the intention-to-treat cohort and showing consistency of the findings in per-protocol (or as-treated) analysis	The results are reassuring when the intention-to-treat analysis (which keeps the randomization intact) and per-protocol (or as-treated) analysis yield similar findings.
Testing (confirming) noninferiority based on both absolute and relative effect measures	Absolute effect measures (such as risk difference) are helpful because they relay clinically meaningful information (eg, 5% absolute difference in all-cause mortality). However, absolute effect measures may become challenging, especially if there is a high crossover rate, lower than expected event rates, or low treatment adherence. Providing confirmatory analyses with relative effect measures in those scenarios will be helpful.
Reporting	
Either a design paper should be published before publication of the main study results, or the study protocol should be publicly released for review and assessment of consistency with the final publication.	The details of study methodology (including the sample size or the noninferiority margin) may need to be changed based on the event rate, feedback from the regulatory agencies, or other reasons. Although this may be inevitable in some scenarios, clear explanations for these changes are recommended in the main publication.
Reporting the noninferiority design and margin in the abstract and full-text	
Reporting the statistical and clinical justification for the choice of the noninferiority margin	
Aligning the conclusions with the statistical study findings	

studies could have increased the study sample, we are unaware of a reason to expect distinctly different results for those 2 years in comparison with other recent trials. Finally, the results might be different for noninferiority trials related to fields other than cardiovascular medicine. However, the factors that predispose trials to bias are similar across the medical and surgical specialties.[32,33]

In conclusion, the number of noninferiority cardiovascular trials published in high-impact journals has increased over time. Although most claimed noninferiority of the tested intervention compared with control, many contained methodological or reporting limitations that pose risks for bias, potentially undermining the validity of their conclusions. Raising the awareness about such limitations, and better adherence to recommendations by the FDA and the CONSORT guidelines for the design, conduct, and reporting of noninferior-

ity trials will enhance confidence in the effectiveness of emerging alternative therapies.

Correspondence

Harlan M. Krumholz, MD, SM, 1 Church St, Ste 200, New Haven, CT 06510. Email harlan.krumholz@yale.edu

Affiliations

Columbia University Medical Center/ New York-Presbyterian Hospital (B.B., G.W.S.). Center for Outcomes Research and Evaluation (CORE) (B.B., J.W.W., N.R.D., J.S.R., H.M.K.), Section of Cardiovascular Medicine (N.R.D., H.M.K.), Section of General Medicine, Department of Internal Medicine (J.S.R.), Robert Wood Johnson Foundation Clinical Scholars Program, Department of Internal

Medicine (J.S.R., H.M.K.), Yale University School of Medicine (I.L.), New Haven, CT. Cardiovascular Research Foundation, New York, NY (B.B., G.W.S.). Saint Vincent Hospital, Worcester, MA (Y.A.). Division of Cardiology, Medical College of Georgia, Augusta (N.P.). Division of Cardiology, Cedars-Sinai Medical Center, Los Angeles, CA (S.K.). Department of Health Policy and Management, Yale School of Public Health, New Haven, CT (H.M.K.).

REFERENCES

1. Gøtzsche PC. Lessons from and cautions about noninferiority and equivalence randomized trials. *JAMA.* 2006;295:1172–1174. doi: 10.1001/jama.295.10.1172
2. Piaggio G, Elbourne DR, Altman DG, Pocock SJ, Evans SJ; CONSORT Group. Reporting of noninferiority and equivalence randomized trials: an extension of the CONSORT statement. *JAMA.* 2006;295:1152–1160. doi: 10.1001/jama.295.10.1152
3. Suda KJ, Hurley AM, McKibbin T, Motl Moroney SE. Publication of noninferiority clinical trials: changes over a 20-year interval. *Pharmacotherapy.* 2011;31:833–839. doi: 10.1592/phco.31.9.833
4. Head SJ, Kaul S, Bogers AJ, Kappetein AP. Non-inferiority study design: lessons to be learned from cardiovascular trials. *Eur Heart J.* 2012;33:1318–1324. doi: 10.1093/eurheartj/ehs099
5. Kaul S, Diamond GA. Good enough: a primer on the analysis and interpretation of noninferiority trials. *Ann Intern Med.* 2006;145:62–69.
6. Patel MR, Mahaffey KW, Garg J, Pan G, Singer DE, Hacke W, Breithardt G, Halperin JL, Hankey GJ, Piccini JP, Becker RC, Nessel CC, Paolini JF, Berkowitz SD, Fox KA, Califf RM; ROCKET AF Investigators. Rivaroxaban versus warfarin in nonvalvular atrial fibrillation. *N Engl J Med.* 2011;365:883–891. doi: 10.1056/NEJMoa1009638
7. Granger CB, Alexander JH, McMurray JJ, Lopes RD, Hylek EM, Hanna M, Al-Khalidi HR, Ansell J, Atar D, Avezum A, Bahit MC, Diaz R, Easton JD, Ezekowitz JA, Flaker G, Garcia D, Geraldes M, Gersh BJ, Golitsyn S, Goto S, Hermosillo AG, Hohnloser SH, Horowitz J, Mohan P, Jansky P, Lewis BS, Lopez-Sendon JL, Pais P, Parkhomenko A, Verheugt FW, Zhu J, Wallentin L; ARISTOTLE Committees and Investigators. Apixaban versus warfarin in patients with atrial fibrillation. *N Engl J Med.* 2011;365:981–992. doi: 10.1056/NEJMoa1107039
8. Connolly SJ, Ezekowitz MD, Yusuf S, Eikelboom J, Oldgren J, Parekh A, Pogue J, Reilly PA, Themeles E, Varrone J, Wang S, Alings M, Xavier D, Zhu J, Diaz R, Lewis BS, Darius H, Diener HC, Joyner CD, Wallentin L; RE-LY Steering Committee and Investigators. Dabigatran versus warfarin in patients with atrial fibrillation. *N Engl J Med.* 2009;361:1139–1151. doi: 10.1056/NEJMoa0905561
9. Smith CR, Leon MB, Mack MJ, Miller DC, Moses JW, Svensson LG, Tuzcu EM, Webb JG, Fontana GP, Makkar RR, Williams M, Dewey T, Kapadia S, Babaliaros V, Thourani VH, Corso P, Pichard AD, Bavaria JE, Herrmann HC, Akin JJ, Anderson WN, Wang D, Pocock SJ; PARTNER Trial Investigators. Transcatheter versus surgical aortic-valve replacement in high-risk patients. *N Engl J Med.* 2011;364:2187–2198. doi: 10.1056/NEJMoa1103510
10. Kolte D, Aronow HD, Abbott JD. Noninferiority trials in interventional cardiology. *Circ Cardiovasc Interv.* 2017;10:e005760.
11. Mulla SM, Scott IA, Jackevicius CA, You JJ, Guyatt GH. How to use a noninferiority trial: users' guides to the medical literature. *JAMA.* 2012;308:2605–2611. doi: 10.1001/2012.jama.11235
12. Gopal AD, Desai NR, Tse T, Ross JS. Reporting of noninferiority trials in clinicaltrials.gov and corresponding publications. *JAMA.* 2015;313:1163–1165. doi: 10.1001/jama.2015.1697
13. Piaggio G, Elbourne DR, Pocock SJ, Evans SJ, Altman DG; CONSORT Group. Reporting of noninferiority and equivalence randomized trials: extension of the CONSORT 2010 statement. *JAMA.* 2012;308:2594–2604. doi: 10.1001/jama.2012.87802
14. U.S. Department of Health and Human Services, Food and Drug Administration. Non-Inferiority Clinical Trials to Establish Effectiveness. November 2016. https://www.fda.gov/downloads/Drugs/Guidances/UCM202140.pdf. Accessed August 27, 2017.
15. Bala MM, Akl EA, Sun X, Bassler D, Mertz D, Mejza F, Vandvik PO, Malaga G, Johnston BC, Dahm P, Alonso-Coello P, Diaz-Granados N, Srinathan SK, Hassouneh B, Briel M, Busse JW, You JJ, Walter SD, Altman DG, Guyatt GH. Randomized trials published in higher vs. lower impact journals differ in design, conduct, and analysis. *J Clin Epidemiol.* 2013;66:286–295. doi: 10.1016/j.jclinepi.2012.10.005
16. Feldman T, Foster E, Glower DD, Glower DG, Kar S, Rinaldi MJ, Fail PS, Smalling RW, Siegel R, Rose GA, Engeron E, Loghin C, Trento A, Skipper ER, Fudge T, Letsou GV, Massaro JM, Mauri L; EVEREST II Investigators. Percutaneous repair or surgery for mitral regurgitation. *N Engl J Med.* 2011;364:1395–1406. doi: 10.1056/NEJMoa1009355
17. Heijboer H, Büller HR, Lensing AW, Turpie AG, Colly LP, ten Cate JW. A comparison of real-time compression ultrasonography with impedance plethysmography for the diagnosis of deep-vein thrombosis in symptomatic outpatients. *N Engl J Med.* 1993;329:1365–1369. doi: 10.1056/NEJM199311043291901
18. Giugliano RP, Ruff CT, Braunwald E, Murphy SA, Wiviott SD, Halperin JL, Waldo AL, Ezekowitz MD, Weitz JI, Špinar J, Ruzyllo W, Ruda M, Koretsune Y, Betcher J, Shi M, Grip LT, Patel SP, Patel I, Hanyok JJ, Mercuri M, Antman EM; ENGAGE AF-TIMI 48 Investigators. Edoxaban versus warfarin in patients with atrial fibrillation. *N Engl J Med.* 2013;369:2093–2104. doi: 10.1056/NEJMoa1310907
19. Adams DH, Popma JJ, Reardon MJ, Yakubov SJ, Coselli JS, Deeb GM, Gleason TG, Buchbinder M, Hermiller J Jr, Kleiman NS, Chetcuti S, Heiser J, Merhi W, Zorn G, Tadros P, Robinson N, Petrossian G, Hughes GC, Harrison JK, Conte J, Maini B, Mumtaz M, Chenoweth S, Oh JK; U.S. CoreValve Clinical Investigators. Transcatheter aortic-valve replacement with a self-expanding prosthesis. *N Engl J Med.* 2014;370:1790–1798. doi: 10.1056/NEJMoa1400590
20. Leon MB, Smith CR, Mack M, Miller DC, Moses JW, Svensson LG, Tuzcu EM, Webb JG, Fontana GP, Makkar RR, Brown DL, Block PC, Guyton RA, Pichard AD, Bavaria JE, Herrmann HC, Douglas PS, Petersen JL, Akin JJ, Anderson WN, Wang D, Pocock S; PARTNER Trial Investigators. Transcatheter aortic-valve implantation for aortic stenosis in patients who cannot undergo surgery. *N Engl J Med.* 2010;363:1597–1607. doi: 10.1056/NEJMoa1008232
21. Leon MB, Smith CR, Mack MJ, Makkar RR, Svensson LG, Kodali SK, Thourani VH, Tuzcu EM, Miller DC, Herrmann HC, Doshi D, Cohen DJ, Pichard AD, Kapadia S, Dewey T, Babaliaros V, Szeto WY, Williams MR, Kereiakes D, Zajarias A, Greason KL, Whisenant BK, Hodson RW, Moses JW, Trento A, Brown DL, Fearon WF, Pibarot P, Hahn RT, Jaber WA, Anderson WN, Alu MC, Webb JG; PARTNER 2 Investigators. Transcatheter or surgical aortic-valve replacement in intermediate-risk patients. *N Engl J Med.* 2016;374:1609–1620. doi: 10.1056/NEJMoa1514616
22. Nissen SE, Wolski KE, Prcela L, Wadden T, Buse JB, Bakris G, Perez A, Smith SR. Effect of naltrexone-bupropion on major adverse cardiovascular events in overweight and obese patients with cardiovascular risk factors: a randomized clinical trial. *JAMA.* 2016;315:990–1004. doi: 10.1001/jama.2016.1558

23. Hernán MA, Robins JM. Per-protocol analyses of pragmatic trials. *N Engl J Med.* 2017;377:1391–1398. doi: 10.1056/NEJMsm1605385

24. Lassen MR, Raskob GE, Gallus A, Pineo G, Chen D, Portman RJ. Apixaban or enoxaparin for thromboprophylaxis after knee replacement. *N Engl J Med.* 2009;361:594–604. doi: 10.1056/NEJMoa0810773

25. Christiansen EH, Jensen LO, Thayssen P, Tilsted HH, Krusell LR, Hansen KN, Kaltoft A, Maeng M, Kristensen SD, Bøtker HE, Terkelsen CJ, Villadsen AB, Ravkilde J, Aarøe J, Madsen M, Thuesen L, Lassen JF; Scandinavian Organization for Randomized Trials with Clinical Outcome (SORT OUT) V investigators. Biolimus-eluting biodegradable polymer-coated stent versus durable polymer-coated sirolimus-eluting stent in unselected patients receiving percutaneous coronary intervention (SORT OUT V): a randomised non-inferiority trial. *Lancet.* 2013;381:661–669. doi: 10.1016/S0140-6736(12)61962-X

26. 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. doi: 10.1016/j.jacc.2017.06.039

27. Van De Werf F, Adgey J, Ardissino D, Armstrong PW, Aylward P, Barbash G, Betriu A, Binbrek AS, Califf R, Diaz R, Fanebust R, Fox K, Granger C, Heikkilä J, Husted S, Jansky P, Langer A, Lupi E, Maseri A, Meyer J, Mlczoch J, Mocceti D, Myburgh D, Oto A, Paolasso E, Pehrsson K, Seabra-Gomes R, Soares-Piegas L, Sùgrue D, Tendera M, Topol E, Toutouzas P, Vahanian A, Verheugt F, Wallentin L, White H. Sin-gle-bolus tenecteplase compared with front-loaded alteplase in acute myocardial infarction: the ASSENT-2 double-blind randomised trial. *Lancet.* 1999; 354:716–722.

28. ICH Harmonised Tripartite Guideline. Choice of control group and related issues in clinical trials. 2000:E10. https://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E10/Step4/E10_Guideline.pdf Accessed November 27, 2018.

29. Kaul S, Diamond GA. Making sense of noninferiority: a clinical and statistical perspective on its application to cardiovascular clinical trials. *Prog Cardiovasc Dis.* 2007;49:284–299. doi: 10.1016/j.pcad.2006.10.001

30. Golish SR. Pivotal trials of orthopedic surgical devices in the United States: predominance of two-arm non-inferiority designs. *Trials.* 2017;18:348. doi: 10.1186/s13063-017-2032-2

31. Bikdeli B, Punnanithinont N, Akram Y, Lee I. Desai NR, Ross JS, Krumholz HM. Two decades of cardiovascular trials with primary surrogate endpoints: 1990–2011. *J Am Heart Assoc.* 2017; 6:e005285. DOI: 10.1161/JAHA.116.005285.

32. Le Henanff A, Giraudeau B, Baron G, Ravaud P. Quality of reporting of noninferiority and equivalence randomized trials. *JAMA.* 2006;295:1147–1151. doi: 10.1001/jama.295.10.1147

33. Aberegg SK, Hersh AM, Samore MH. Empirical consequences of current recommendations for the design and interpretation of noninferiority trials. *J Gen Intern Med.* 2018;33:88–96. doi: 10.1007/s11606-017-4161-4