outcome and goes beyond the pure evidence of having only a more comfortable regimen for a patient. So noninferiority is being tested when superiority could be proved.

Noninferiority trials can represent a rigorous method to widen treatment options in proper settings, but physicians should carefully read and understand the deepest statistical and technical pieces of them to verify that a proper noninferiority margin was chosen and assess the extent of clinically irrelevant events.

See Article page 2119.

Commentary: How to overcome an inferiority complex

Paul Kurlansky, MD

Noninferiority trials require careful execution and interpretation. For a variety of reasons, this format has increased markedly in recent years, with a PubMed search revealing 1 study each in 1997 and 1998, with an acceleration to greater than 500 in 2015¹ (Figure 1). Using “noninferiority” as a search term in [ClinicalTrials.gov](#) currently yields 2441 studies.² Therefore, the expert review by Falk and Friede³ in this issue of the *Journal* is both timely and welcome. At the risk of redundancy, it is worthwhile to review several of the pertinent points that they have raised (although their guide to many of the technical aspects of designing and interpreting a noninferiority trial will prove invaluable researchers and clinicians), as well as to explore how the current literature meets many of the standards that they propose.

Interestingly enough, if I dust off my trusty (if ancient) *Merriam-Webster's Collegiate Dictionary* (or even my son's more recent edition), I discover that “noninferior” and “noninferiority” are not recorded as words in the

References

1. Suda KJ, Hurley AM, McKibbin T, Motl Moroney SE. Publication of noninferiority clinical trials: changes over a 20-year interval. *Pharmacotherapy*. 2011;31:833-9.
2. 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;308:2594-604.
3. Falk V, Friede T. Noninferiority trials: what's clinically (ir) relevant? *J Thorac Cardiovasc Surg*. 2021;161:2119-23.
4. European Medicines Agency. Guideline on the choice of the non-inferiority margin. Available at: www.ema.europa.eu/pdfs/human/ewp/215899en.pdf. Accessed April 6, 2020.

CENTRAL MESSAGE

Noninferiority trials have increased in prevalence and importance, yet require specialized understanding and care to properly conduct and interpret.

English language. Perhaps it is this very “nonintuitive” nature that makes these studies so challenging to perform and interpret appropriately. Unlike the classic superiority trial, where treatment A is compared with placebo or treatment B to determine if one is better (however defined) than the other, a noninferiority trial seeks to determine if A is no worse than B. Therefore, the null hypothesis, rather than being that there is no difference between A and B, is actually that A is worse than B. This shift turns many things somewhat upside down—demonstration of a difference may actually lead to acceptance rather than rejection of the null hypothesis. Why do this? Aside from the more cynical perspective that this approach paves the way for “me too” drugs that pharmaceutical companies develop for market share but that add little or no value to those currently

From the Division of Cardiac Surgery, Department of Surgery, Columbia University, New York, NY.

Disclosures: The author reported no conflicts of interest.

The *Journal* policy requires editors and reviewers to disclose conflicts of interest and to decline handling or reviewing manuscripts for which they may have a conflict of interest. The editors and reviewers of this article have no conflicts of interest.

Received for publication April 28, 2020; revisions received April 28, 2020; accepted for publication April 29, 2020; available ahead of print May 16, 2020.

Address for reprints: Paul Kurlansky, MD, Black Building 210, 650 West 168th St, New York, NY 10032 (E-mail: Pk2245@cumc.columbia.edu).

J Thorac Cardiovasc Surg 2021;161:2125-9

0022-5223/\$36.00

Copyright © 2020 by The American Association for Thoracic Surgery

<https://doi.org/10.1016/j.jtcvs.2020.04.153>

FIGURE 1. The rise in non-inferiority trials.

available, or for surgeons to self-promote techniques that provide marginal incremental benefit (remember, the therapy being tested does not need to be BETTER, it just needs not to be WORSE), there are actually several valid reasons for this study design. The therapeutic advantage may be considerable but unrelated to the key therapeutic outcome. If a less-invasive transcatheter procedure can replace a critically stenotic aortic valve with similar death and stroke as the classic surgical approach, but with less patient morbidity, as was demonstrated in the PARTNER⁴ and CoreValve⁵ trials, the therapeutic landscape can be markedly advanced. The benefit being proposed is not related to a better treatment outcome, but rather to an improvement in ease of delivery, patient comfort or compliance, cost, safety, associated complications, and so forth. Indeed, conduct of a placebo-controlled trial may well not be ethically acceptable if reasonably acceptable current therapy already exists. Point number 1: The rationale for this study design needs to be clearly enunciated.

The next most interesting aspect of this approach is that, unlike the classic superiority trial, one must a priori decide what effect size is considered significant—not merely in the power calculation for study size, but in the actual determination of results. As is well known, sample size can drive *P* values, so that, in the more traditional superiority approach we can wind up with the anomalous findings, such that a hazard ratio (HR) of 0.90 for the combined end point of death, stroke, or myocardial infarction can show no benefit for bilateral versus single internal thoracic artery grafting in one trial (confidence interval [CI], 0.79-1.03),⁶ whereas the nearly identical HR of 0.88 can document a benefit for arterial grafting in another study with a larger sample size (CI, 0.83-0.93).⁷ In the reverse logic of noninferiority, the larger the noninferiority margin, the more likely one is to be able to “prove” noninferiority and the fewer subjects will be needed to do so. The obvious problem is that one might wind up finding noninferiority for a treatment that by clinical standards, or even by

FIGURE 2. Hypothesis testing in noninferiority trials.

statistical metrics, is in fact inferior. Note the situation circled in red in Figure 2,⁸ which is frequently enigmatically labeled as “inferior and noninferior”—in other words, inferior, but not sufficiently so to meet the preestablished margin of inferiority. Additionally, although the treatment under study may meet the noninferiority criteria of the “active control,” depending on the margin of inferiority and the size of treatment effect in the original trials of that active control versus placebo, the study treatment may actually not prove significantly different from placebo. Aside from specific regulatory guidelines, the ideal situation leverages previous placebo-controlled randomized trials that provide CIs around the effect size of treatment benefit. The margin cannot be greater than the smallest effect size (lowest CI), and Falk and Friede’s³ suggestion of a 50% effect size is well accepted when such information is available. As noted by the authors, such calculations presume a certain degree of comparability between the circumstances of prior studies and the current era: the constancy assumption. Perhaps this issue can be illustrated by trying to estimate the survival benefit for coronary artery bypass grafting (CABG) versus medical therapy for stable coronary artery disease. Most of the landmark trials on this topic were performed in the 1980s, before the importance of the use of the left internal thoracic artery was well recognized, before the availability of statins, and before the invention of percutaneous coronary intervention (PCI). Patient demographics and treatments have changed so substantially that any noninferiority trial would need to make several inferential assumptions (and leaps of faith). Perhaps this is one of the reasons that the recent ISCHEMIA trial⁹ was designed as a comparative effectiveness rather than a noninferiority trial. Additional complexity relates to the selection of relative versus absolute differences, with the latter being extremely sensitive to variations from anticipated or expected event rates. If the event rates are considerably

less than anticipated from landmark placebo controlled trials, the inferiority margin may be inappropriately high, suggesting noninferiority when a relative event rate might suggest the opposite. In the absence of such prior trial-driven information, much may depend on the outcome of interest. For example, our tolerance of inferiority regarding mortality may be considerably less than it might be for the ability of a medication to effectively lower blood pressure. Indeed, if a given medication required only 1 dosage per day as opposed to 3 or 4, a lower effectiveness might be tolerated in anticipation of greater compliance and improved overall treatment effect. It is in this area that there is the opportunity to include the patient perspective in both the outcome of interest and the determination of margin. Recent evidence suggests that many patients fear stroke more than mortality,¹⁰ and consideration of these sorts of perspectives may help to drive the determination of margin and the choice of outcome. The patient perspective may also pose certain ethical issues when the subject may be risking inferior treatment, whereas the benefit, such as reduced cost, accrues to society but not to the patient. Point number 2: Selection and justification of margin are critical.

Regarding outcome, the choice of end points may determine the fate of the trial. The SYNTAX trial was a noninferiority trial designed to compare PCI with CABG for patients with severe coronary artery disease.¹¹ The trial was powered to reject the null hypothesis with a noninferiority margin of 5% and a (generous) 1-sided 5% significance level for the combined outcome of major adverse cardiac and cerebrovascular events, that is, death, stroke, myocardial infarction, or repeat revascularization. Because the event rate was higher in the PCI group (17.8% vs 12.4% for CABG, $P = .002$), the study did not meet the margin for noninferiority. Results were largely driven by an increased rate of repeat revascularization (13.5% vs 5.9%, $P < .001$), although death and myocardial infarction

were similar. There is a hierarchical logic to these trials such that if the primary end point for noninferiority is not met, post hoc analysis of findings, even in prespecified subgroups, cannot be viewed as conclusive. Therefore, in planning the EXCEL trial of PCI versus CABG for patients with left main coronary artery disease, investigators were careful to exclude repeat revascularization and limit their end points to death, stroke, or myocardial infarction.¹² As fate would have it, even this more limited approach generated a maelstrom of controversy when the 5-year results reported a noninferior event rate of 22.0% for the patients who underwent PCI and 19.2% for the patients who underwent CABG ($P = .13$), yet a higher death rate for PCI (13.0% vs 9.9%; CI, 0.2-6.1) without a difference in stroke or myocardial infarction. To make matters more complex, periprocedural myocardial infarction was more common after CABG, whereas late infarction was more common after PCI. A similar problem in definition of myocardial infarction as an outcome was noted in the recent ISCHEMIA trial. Leaving aside the controversies, the journey illustrates the challenges posed by composite end points. They increase the event rates and help to drive the power of the trial with fewer subjects, but they can create problems when not all end points are valued equally and not all move in the same direction. One final note regarding end points: The study end points are generally not those of the purported value of the new therapy—they may or may not be incorporated into the study protocol (such as the quality of life assessments incorporated into the PARTNER trial). Point number 3: Investigators must choose study end points wisely.

Study execution, which is essential for all clinical trials, takes on added importance in the noninferiority design. Incomplete data, dropouts, and failure of projected recruitment may all tend to bias the study toward the mean, because real differences may not emerge. Subjective observations are particularly vulnerable in nonblinded trials, which, by their nature, describe virtually all but those rare and brave sham operative trials. Discretionary judgments such as wound appearance or clinical choices such as transfusion or length of stay are vulnerable to lack of equipoise. But perhaps the most counterintuitive aspect of these studies derives from analysis of the results. Although the primary analysis of any prospective randomized trial resides in the intention-to-treat analysis, even though the “as-treated” results may differ. However, in the noninferiority environment, cross-over patients may seriously bias the study toward the mean and drive a false conclusion of noninferiority. Perhaps this can best be illustrated by the extreme hypothetical example of all the patients in group A actually receiving the same treatment as the patients in group B. Intention-to-treat analysis would convincingly demonstrate absolutely no difference! Therefore, the more circumspect approach for these trials is to report both the

intention-to-treat and the as-treated, or per protocol, results. Point number 4: Extra care must be exercised in the conduct and interpretation of these trials.

A theoretical concern that may or may not have begun to manifest is that of “biocreep”—the phenomenon of a given treatment B being approved or accepted on the basis of noninferiority to a certain therapy A that was tested in a placebo-controlled trial, and a later treatment C being accepted on the basis of noninferiority to treatment B, and so on until treatments that are no better or even worse than placebo are being approved. According to a report from the General Accounting Office from a decade ago, this had not yet surfaced as a problem; however, with the proliferation of these studies in recent years, there may be a genuine cause for concern.¹³ Point number 5: Noninferiority should never be confused with superiority.

Unfortunately, despite their somewhat unique design, these trials are not free of the many classic challenges involved in conducting surgical trials. Unlike medication studies, surgical trials do not involve a fixed and uniform dosage. Although pharmacodynamics and pharmacokinetics may vary among patients (which have their correlates for surgical trials as well), the medication itself does not vary. On the other hand, the skill and style in the performance of a surgical procedure may vary (as was dramatically demonstrated by the broad range of bilateral internal thoracic artery grafting performed in the ART trial). As noted earlier, blinding can be a difficult limitation, as can the integrity of randomization concealment. However, perhaps the most subtle but compelling challenge in surgical trials is the lack of true equipoise. Just as for a superiority trial, a skilled operator may be reluctant to randomize a patient who he thinks has a high probability of benefit and therefore only enrolls those of questionable benefit, biasing the study toward the null; in a noninferiority trial, he may avoid those patients who he thinks would derive the least benefit and select only those most likely to arrive at positive results. Either way, preselection bias may condemn the study to a patient cohort that is not reflective of those who generally undergo the procedure or of those generally assessed in clinical practice for this particular problem.

With an increasing portion of late-breaking conference trials and Food and Administration drug approvals being based on these trials, a sober assessment of study quality seems in order. In a recent systemic review of cancer drugs,¹⁴ researchers found 23 trials (enrolling 21,437 patients) that used overall survival as the primary end point. The noninferiority margins for the HR of overall survival ranged from 1.08 to 1.33, the latter potentially constituting a 33% higher mortality! The noninferiority design was missing from 39%, although, fortunately, the overall pooled HR for survival was 0.97 (CI, 0.92-1.02). Given these methodological shortcomings, perhaps we can rely on our major high-impact journals to protect us from pitfalls. Not really:

A recent review of 111 noninferiority cardiovascular trials published in *The Journal of the American Medical Association*, *Lancet*, and *The New England Journal of Medicine* between 1990 and 2016 revealed some disturbing findings.¹⁵ There were wide variations in noninferiority margins: risk differences of 0.4% to 25%, HRs of 1.05 to 2.85, odds ratios of 1.1 to 2.0, and relative risks of 1.1 to 1.8. Only 7 of 111 (6.3%) could be considered low risk of bias for all major and minor biasing factors, 37% did not confirm findings in both intention-to-treat and per-protocol/as-treated groups, 3.6% had discrepant results, and 38.7% had no justification for the inferiority margin.

Regarding noninferiority trials, given their nonintuitive nature, increasing prevalence and influence, and vulnerability to spurious conclusion, the careful analysis by Falk and Friede³ in the *Journal* not only calls these issues to our attention but also provides a vital resource in helping us to evaluate the data and plan future studies according to the most appropriate standards.

References

1. A Double Whammy of Non-Good News About Non-Inferiority Trials. Available at: <https://blogs.plos.org/absolutely-maybe/2019/08/31/a-double-whammy-of-non-good-news-about-non-inferiority-trials/>. Accessed April 27, 2020.
2. Available at: <https://clinicaltrials.gov/ct2/results?cond=&term=noninferiority&cntry=&state=&city=&dist=>. Accessed April 27, 2020.
3. Falk V, Friede T. Noninferiority trials: what's clinically (ir) relevant? *J Thorac Cardiovasc Surg*. 2021;161:2119-23.
4. Smith CR, Leon MB, Mack MJ, Miller DC, Moses JW, Svensson LG, et al; for the PARTNER Trial Investigators. Transcatheter versus surgical aortic-valve replacement in high-risk patients. *N Engl J Med*. 2011;364:2187-98.
5. Adams DH, Popma JJ, Reardon MJ, Yakubov SJ, Coselli JS, Deeb M, et al; for the U.S. CoreValve Clinical Investigators. Transcatheter aortic-valve replacement with a self-expanding prosthesis. *N Engl J Med*. 2014;370:1790-8.
6. Taggart DP, Benedetto U, Gerry S, Altman DG, Gray AM, Lees B, et al; Arterial Revascularization Trial Investigators. Bilateral versus single internal-thoracic-artery grafts at 10 years. *N Engl J Med*. 2019;380:437-46.
7. Samadashvili Z, Sundt TM III, Wechsler A, Chikwe J, Adams DH, Smith CR, et al. Multiple versus single arterial coronary bypass graft surgery for multivessel disease. *J Am Coll Cardiol*. 2019;74:1275-85.
8. Kurlansky P. Commentary: the superiority of noninferiority. *J Thorac Cardiovasc Surg*. 2020;159:1866-7.
9. Maron DJ, Hochman JS, Reynolds HR, Bangalore S, O'Brien SM, Boden WE, et al; for the ISCHEMIA Research Group. Initial invasive or conservative strategy for stable coronary disease. *N Engl J Med*. 2020;382:1395-407.
10. Hanger HC, Fogarty B, Wilkinson TJ, Sainsbury R. Stroke patients' view on stroke outcomes: death versus disability. *Clin Rehabil*. 2000;14:417-24.
11. Serruys PW, Morice MC, Kappetein AP, Colombo A, Holmes DR, Mack MJ, et al; SYNTAX Investigators. Percutaneous coronary intervention versus coronary-artery bypass grafting for severe coronary artery disease. *N Engl J Med*. 2009;360:961-72.
12. Stone GW, Sabik JF, Serruys PW, Simonton CA, Généreux P, Puskas J, et al; EXCEL Trial Investigators. Everolimus-eluting stents or bypass surgery for left main coronary artery disease. *N Engl J Med*. 2016;375:2223-35.
13. US Government Accountability Office. New drug approval: FDA's consideration of evidence from certain clinical trials (GAO-10-798); 2010. Available at: <http://www.gao.gov/products/GAO-10-798>. Accessed April 27, 2020.
14. Gyawali B, Tessema FA, Jung EJ, Kesselheim AS. Assessing the justification, funding, success and survival outcomes of randomized noninferiority trials of cancer drugs: a systematic review and pooled analysis. *JAMA Netw Open*. 2019;2:e199570.
15. Bikdeli B, Welsh JW, Akram Y, Punnanithont N, Lee I, Desai NR, et al. Non-inferiority designed cardiovascular trials in highest-impact journals. *Circulation*. 2019;140:379-89.