JACC: ADVANCES VOL. 3, NO. 7, 2024

EDITORIAL COMMENT

Reducing Inferiority in the Design, Conduct, Analysis, Reporting, and Interpretation of Noninferiority Trials

C. Michael Gibson, MS, MD

As drug, device, digital and diagnostic technologies mature, there is often a plateauing in efficacy, yet further improvements may be made in safety, convenience, or expense. Noninferiority trials are undertaken to assure, with a pre-specified degree of confidence, that a new technology is not associated with a clinically or statistically significant loss in efficacy as compared with established therapies. In a PubMed search using the term noninferiority trials, the number of noninferiority trials has increased dramatically over the last 20 years with nearly 483 being reported in 2023 alone (**Figure 1**). Clinicians often communicate the absolute risk (not the relative risk) of an event to patients; likewise, regulators often focus on absolute event rates. Although dimensionless and less susceptible to variability in absolute event rates, the clinical relevance of relative risk reductions is less clear. Trials that use an absolute difference rather than a relative difference to determine a margin are often more highly powered statistically. For these reasons, about three-quarters of the time, the acceptable margin for a new technology to be noninferior to an established one is based on an absolute difference in event rates as Greco et al[1] point out in this issue of *JACC: Advances*. A problem arises when the rates of events are much lower than expected, reducing the statistical power to truly exclude a meaningful absolute difference in event rates.

Indeed, the study by Greco et al[1] is important as it points out inferiorities in the analysis, reporting and interpretation of noninferiority trials: in about 25% of trials that used an absolute risk reduction or margin, despite initial claims of noninferiority, noninferiority was in fact not demonstrated when actual observed event rates were used (because event rates were lower than projected) and no mention was made of this failure to show statistically significant results.[1] In failing to cite lack of power and recalculated results, the original authors may have flouted the spirit of the adage that "absence of evidence does not constitute evidence of absence." While the authors included studies from 2017 and beyond, the results of the earlier ISAR Safe study demonstrate the enormous impact of the findings of Greco et al.[1] The event rates in the control arm were lower than expected, and when the authors used the methodology adopted by Greco et al,[1] the sample size for a noninferiority study ballooned to 40,000 patients (10 times larger than the original study).[2]

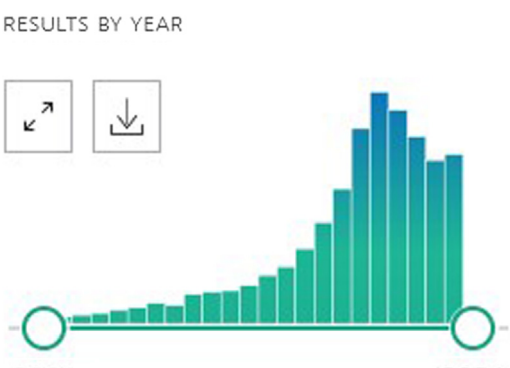
Greco et al[1] evaluated deficiencies in the analysis, reporting, and interpretation of noninferiority trials, but did not address deficiencies in their conduct. Nearly 70% of the trials were open label which may drive results to the null in a noninferiority trial. What the authors do not cite is the extent of missing data in the trials.[1] Missing data and the potential for informative censoring (people who tolerate the drug or device may differ from those who don't and drop out) may also drive the results toward noninferiority. Similarly, cross-over between treatment arms and non-adherence may drive the differences between strategies toward the null and increase the chance of declaring noninferiority. A per protocol analysis may address the cross-over issue but results in loss of randomization.

Another current and particularly vexing issue complicating the design of noninferiority trials is the

From the Baim Institute for Clinical Research, Harvard Medical School, Boston, Massachusetts, USA.
The author attests they are in compliance with human studies committees and animal welfare regulations of the author's institution and Food and Drug Administration guidelines, including patient consent where appropriate. For more information, visit the Author Center.

ISSN 2772-963X https://doi.org/10.1016/j.jacadv.2024.101026

FIGURE 1 The Number of Noninferiority Trials has Increased Dramatically Over the Last 20 Years With Nearly 483 Being Reported in 2023 Alone



use of composite endpoints in which efficacy and safety endpoints may go in opposite directions fully or partially cancelling each other out and potentially biasing the results toward noninferiority and masking important differences in the separate endpoints of efficacy and safety. Examples include the composite endpoints of net clinical adverse events (NACE) and ischemic/hemorrhagic stroke plus systemic embolization (SE) which merge ischemic and bleeding endpoints in noninferiority trials. Based upon the "constancy assumption", the FDA has recommended that subsequent noninferiority trials preserve the design and methods of the original trial upon which the magnitude of benefit of the original therapy over placebo was measured.[3] Because original trials of pharmacotherapy in atrial fibrillation used stroke/SE as the endpoint, subsequent trials have used the same endpoint to preserve the constancy of the endpoints and the margin of benefit of an original active treatment over "placebo."

While the stroke/SE endpoint fulfills the constancy assumption, this original endpoint has limitations. In trials evaluating antithrombins and devices to reduce bleeding, a reasonable expectation is that ischemic and hemorrhagic endpoints will go in opposite directions, driving the results to noninferiority and set the field up for "biocreep" or noninferiority creep. Indeed, recent guidance in the New England Journal of Medicine states that a noninferiority trial should not be based on divergent endpoints.[4] Additional limitations of a stroke/SE endpoint include the fact that efficacy (ischemic stroke) endpoints use intent to treat analyses, while safety endpoints (hemorrhagic strokes) typically use per protocol or as treated analyses, but despite differing analytic approaches these 2 endpoints are grouped together. The endpoint does not include morbid/mortal outcomes such as subdural and epidural intracranial hemorrhage. This endpoint gives ischemic and hemorrhagic stroke the same weight but hemorrhagic stroke is more often fatal and disabling (30%-55% mortality).[5] This endpoint does not account for gastrointestinal bleeding, critical organ, or fatal bleeding. Note also that the acceptable margin for efficacy and safety may be different; when safety and efficacy endpoints are combined, this nuance is masked. Despite the use of the stroke/SE endpoint in the ENGAGE trial, the Food and Drug Administration considered safety and efficacy separately to select the more effective but less safe dosing strategy of 60 mg vs 30 mg of Edoxaban, raising questions as to the utility of this composite endpoint.[6,7] Note also that the risk of patients may change, and the standard of care may change in subsequent trials which may violate the "constancy assumption", again raising questions about its utility. Finally, more input from patients (such as utility scores) as to how they value different endpoints, would be informative in constructing noninferiority margins. How many ischemic strokes would patients tolerate to yield fewer bleeding outcomes?

The current approach to noninferiority may itself be inferior at present. As has been recommended in the past, composite endpoints with diverging responses in individual components to pharmacologic or device therapy are not appropriate (eg, combing bleeding and efficacy endpoints) as they drive the result to noninferiority while masking important differences in the safety and efficacy of the strategies. Just as clinicians and patients do in their decision-making, safety and efficacy data should be considered separately, and decisions should be based upon the totality of what are often divergent data. Rather than using noninferiority, superiority in safety (bleeding) of a new strategy could be established,[8] and broader, more well powered patient centric endpoints such as death and hospitalization could be used to assess efficacy.[9] Careful monitoring of event rates is warranted, and should they be lower than expected, sample sizes could be increased, or patient risk could be increased. Given the frequency with which noninferiority trials may be underpowered and no longer statistically significant when analysis is

based upon actual event rates, regulators, journal editors, trialists and clinicians should require that both an absolute and a relative margin be assessed before a therapy is truly deemed to be noninferior to another. If there are concerns regarding the magnitude of missing data, tipping analyses should be performed to assess the robustness of the noninferiority findings. These modifications may represent a superior approach to the current deficiencies in noninferiority trials.

FUNDING SUPPORT AND AUTHOR DISCLOSURES

Dr Gibson has received research support from Johnson & Johnson and CSL Behring; and consulting support from AstraZeneca, CSL Behring, Johnson & Johnson, Janssen, and Bayer.

ADDRESS FOR CORRESPONDENCE: Dr C. Michael Gibson, Baim Institute, 930 Commonwealth Avenue, Boston Massachusetts 02115, USA. E-mail: charlesmichaelgibson@gmail.com.

REFERENCES

1. Greco A, Spagnolo M, Laudani C, et al. Assessment of noninferiority margins in cardiovascular medicine trials. *JACC Adv*. 2024;3:101021.

2. Schulz-Schüpke S, Byrne RA, ten Berg JM, et al. ISAR-SAFE: a randomized, double-blind, placebo-controlled trial of 6 vs. 12 months of clopidogrel therapy after drug-eluting stenting. *Eur Heart J*. 2015;36(20):1252–1263.

3. FDA. Non-inferiority clinical trials to establish Effectiveness guidance for Industry. 2016. Accessed May 26, 2024. https://www.fda.gov/media/78504/download

4. Mauri L, D'Agostino RBSr. Noninferiority trials. *N Engl J Med*. 2018;378(3):304–305.

5. Freeman WD, Aguilar MI. Management of warfarin-related intracerebral hemorrhage. *Expert Rev Neurother*. 2008;8(2):271–290.

6. Giugliano RP, Ruff CT, Braunwald E, et al. Edoxaban versus warfarin in patients with atrial fibrillation. *N Engl J Med*. 2013;369(22):2093–2104.

7. FDA Review of Edoxaban. Accessed May 26, 2024. https://www.accessdata.fda.gov/drugsatfda_docs/nda/2015/206316orig1orig2s000clinpharmr.pdf

8. Gibson CM, Mehran R, Bode C, et al. Prevention of bleeding in patients with atrial fibrillation Undergoing PCI. *N Engl J Med*. 2016;375(25):2423–2434.

9. Gibson CM, Pinto DS, Chi G, et al. Recurrent hospitalization Among patients with atrial fibrillation Undergoing Intracoronary stenting treated with 2 treatment strategies of Rivaroxaban or a Dose-Adjusted oral Vitamin K Antagonist treatment strategy. *Circulation*. 2017;135(4):323–333.

KEY WORDS acute coronary syndrome, anticoagulants, antiplatelets, atrial fibrillation, noninferiority