

EDITORIAL AND COMMENT

Non-Inferiority Trials in Medicine: Practice Changing or a Self-Fulfilling Prophecy?

Vinay Prasad, MD, MPH^{1,2,3}

¹Division of Hematology and Medical Oncology, Knight Cancer Institute, Oregon Health and Science University, Portland, OR, USA; ²Department of Preventive Medicine and Public Health, Oregon Health and Science University, Portland, OR, USA; ³Center for Health Care Ethics, Oregon Health and Science University, Portland, OR, USA.

J Gen Intern Med 33(1):3–5
DOI: 10.1007/s11606-017-4191-y
© Society of General Internal Medicine 2017

Each week, a practicing physician may read the latest medical journal and find a new intervention claiming to be non-inferior to a therapy she uses, while offering the advantage of freedom from monitoring, lower cost, or less toxicity. In an ideal world, such trials could be accepted at face value, but a growing body of evidence suggests that when it comes to non-inferiority trials: reader beware.

The ubiquity of non-inferiority trials raises important questions: how often do they find non-inferiority? Are industry-sponsored trials more likely to reach favorable conclusions? Is the definition of non-inferiority chosen for good clinical reasons? Are there interventions that are actually inferior, but are somehow deemed “non-inferior”? Many of these questions are answered by the tour-de-force article by Aberegg, Hersh, and Samore, which examines 183 non-inferiority comparisons from 163 clinical trials appearing in the five highest impact factor medical journals.¹

WHAT ARE NON-INFERIORITY TRIALS AND HOW ARE THEY INTERPRETED?

In any randomized comparison, there is some difference (d) between the experimental and control arms, and a 95% confidence interval around this difference, which provides a sense of the precision of the estimate. In a non-inferiority trial, the researcher specifies, before the trial begins, the difference (or delta) between the old and new treatments that is small enough for the trialist to claim no difference. After the study is complete, the researcher reports the observed delta, along with the 95% confidence interval for this difference. To be non-inferior, the lower bound of the confidence interval must be greater than the non-inferiority margin or delta. In other words, the experimental therapy must be no worse than the margin to be deemed non-inferior. It is important to note the importance of specifying this delta before the trial is conducted.

Aberegg and colleagues present an elegant figure for possible outcomes and interpretations from non-inferiority trials and note a curious inconsistency in the CONSORT recommendation for their interpretation. One possibility (scenario 1) is that the experimental arm is superior to the conventional treatment, with a sufficiently large delta and a 95% CI for the difference entirely above 0. Another possibility is that the experimental treatment is significantly worse than the conventional treatment (scenario 4), with a sufficiently large delta and a 95% CI of the difference entirely below 0. Here, out of fairness, we should say the experimental treatment is inferior to the active control, but the CONSORT guidance is to call this non-inferior if it falls above the margin. In another inconsistency, if the 95% confidence interval is below 0, but crosses the margin, CONSORT calls this inconclusive, though the correct interpretation would again be inferior (scenario 7). CONSORT's interpretation is asymmetric.

HOW OFTEN DO NON-INFERIORITY TRIALS FIND NON-INFERIORITY? WHAT ABOUT THOSE THAT ARE INDUSTRY SPONSORED?

Aberegg and colleagues find that an astonishing 77% (141/183) of non-inferiority comparisons find the experimental arm is either superior or non-inferior. Even more sobering is the finding that only 2% conclude the novel intervention is inferior. This percentage is greater than superiority trials in medicine, which find new interventions superior only half the time,² and these figures are similar to non-inferiority estimates from all indexed journals.³

If non-inferiority trials almost never conclude an intervention is inferior, one wonders if they represent unbiased research. In a non-inferiority trial one can assure oneself of a favorable outcome (i.e., non-inferiority), by selecting a permissive margin. When such a high percentage of trials find non-inferiority or remain inconclusive, one wonders if these trials were fairly designed or if instead their sponsors were so wedded to the outcome the trial was designed from the outset to get the desired answer. Have non-inferiority trials become a self-fulfilling prophecy?

When it comes to industry-sponsored trials, the answer may be yes. A recent paper by Flacco and colleagues, shows that

97% of industry-sponsored non-inferiority trials reach favorable conclusions.⁴ This exceeds Aberegg and colleagues' estimates of non-inferiority trials in general and raises the provocative questions of whether industry-sponsored non-inferiority trials offer any value—aside from capturing market share.

DO NON-INFERIORITY TRIALS HIDE INFERIOR INTERVENTIONS?

Aberegg and colleagues do something that has not been done before. They calculate two-sided 95% confidence intervals surrounding the difference between experimental and active control arms and are able to show that some interventions called non-inferior are actually inferior. In fact, 2% of trials called non-inferior have a 95% CI that excludes no difference. These are interventions that any unbiased observer would say are inferior. Another 8% of trials are called inconclusive because the 95% confidence interval spans the margin, though the confidence interval excludes no difference, and these too should be called inferior. Aberegg exposes a major limitation in the CONSORT guidance. Ten percent of studies called non-inferior or inconclusive should be called inferior.

In fairness to the CONSORT authors, this predicament was partially anticipated. In the caption to Fig. 1 of the CONSORT guidance, the authors acknowledge that in both these scenarios the “new treatment is significantly worse than the standard.”⁵ However, the CONSORT authors' prediction that the scenario occurring in 2% of comparisons (where inferior interventions are called non-inferior) would be “unlikely because it would require a very large sample size” seems slightly off the mark.

IS THE MARGIN SELECTED FOR GOOD REASON?

Simply calling inferior interventions non-inferior does not fully explain the high rate of non-inferior trials finding non-inferiority. Instead, we also have to look at the delta or margin. The delta or margin is how much worse the new intervention can be to the older intervention before it is no longer considered non-inferior. The easiest way to ensure a successful non-inferiority trial is to pick a very large margin.

The US Food and Drug Administration offers clear guidance on how big the margin can be. The margin cannot be bigger than the benefit that the active control arm has over placebo. In other words, a margin cannot allow a drug that is totally inert to be called non-inferior.⁶ Moreover, the FDA points out that it is desirable for the margin to be much smaller than the benefit of the active control over placebo. They write a good

margin is the “the largest loss of effect that would be clinically acceptable.”⁶ Aberegg and colleagues show that 58% of margins provide no reason as to why they are selected. Another 17% are vague and non-reproducible, and only 25% provide a concrete explanation. These rates are not acceptable. If the conclusion that a new treatment is non-inferior is based on selection of the margin, clear explication of how the margin was selected is essential.

Consider a recent non-inferiority trial sponsored by GlaxoSmithKline, the makers of pazopanib. In the study, pazopanib was found to be non-inferior on the endpoint of progression-free survival to sunitinib—the standard of care—for adults with metastatic kidney cancer. The margin selected was 0.25, meaning that pazopanib could be up to 25% worse. Given that sunitinib only improves PFS 6 months⁷ over the largely ineffective therapy of interferon, this means that pazopanib may only provide 4.5 months of PFS benefit and still be called non-inferior. In fact, the upper bound of the 95% CI of the delta was an HR of 1.22 for pazopanib. Aberegg et al. found some trials allowed margins as large as 0.25.

IS THERE A JUSTIFICATION FOR RUNNING A NON-INFERIORITY TRIAL IN THE FIRST PLACE?

Non-inferiority trials only make sense if the experimental therapy is cheaper, more convenient, less invasive, or less toxic than the active control. If a new therapy fails to offer benefit on any of these metrics, testing it in a non-inferiority trial is unethical. Why would anyone want an equally (or more) toxic, equally (or more) costly, equally (or more) invasive, and equally (or more) inconvenient intervention that may be worse than an established alternative? Yet, Aberegg and colleagues find that only 70% of non-inferiority studies explicitly said why the new therapy has an advantage, and in 11% of cases, no advantage could be deduced. This suggests that some fraction of these studies should not have been performed. It is worth nothing that several of the findings of the Aberegg et al. study were confirmed in another recent paper.⁸

DO PATIENTS KNOW ALL THIS?

Complementing the current paper, recently Doshi et al. studied informed consent forms from non-inferiority studies of antibiotics leading to EMA approval.⁹ They found that frequently neither experts in methodology nor patient advocates could identify the purpose of the study from the forms. Just 1/50 studies conveyed the purpose in the eyes of methodologists, and just 7 of 50 in the eyes of patients. These results raise the question of whether consent is truly informed and, if not, whether the trial is even ethical.

WHAT SHOULD A CLINICIAN TAKE AWAY?

For the practicing doctor the paper by Aberegg and colleagues provides an important lesson. Non-inferiority trials are often positive because the margins are chosen for no clearly given reason and because there is an asymmetry in how the trials are interpreted. It is likely many of these trials seek to capture market share rather than answer a meaningful clinical question.

For this reason, we cannot accept the conclusions of non-inferiority trials to guide our practice without close examination. When you read a non-inferiority study, ask yourself: is the new therapy cheaper, more convenient, less invasive, or less toxic than the older one? If the answer is no, read no further. Next, ask yourself how much treatment effect am I willing to lose to have this option? 5%, 10%? Then look at the margin tested. Finally, ask yourself if the intervention was actually inferior. Aberegg and colleagues should be congratulated for performing a rigorous empirical evaluation that forces us all to think more about non-inferiority trials.

Corresponding Author: Vinay Prasad, MD, MPH; Division of Hematology and Medical Oncology, Knight Cancer Institute Oregon Health and Science University, Portland, OR, USA (e-mail: prasad@ohsu.edu).

Funding Dr. Prasad is funded by the Laura and John Arnold Foundation to Study Low Value Practices.

Compliance with Ethical Standards:

Conflict of interest: Dr. Prasad receives royalties from his book *Ending Medical Reversal*.

REFERENCES

1. **Aberegg SK, Hersh AM, Samore MH.** Empirical consequences of current recommendations for the design and interpretation of noninferiority trials. *J Gen Intern Med.* 2017. doi: <https://doi.org/10.1007/s11606-017-4161-4>.
2. **Djulgovic B, Kumar A, Glasziou P, et al.** Medical research: trial unpredictability yields predictable therapy gains. *Nature.* 2013;500(7463):395-96. <https://doi.org/10.1038/500395a>.
3. **Suda KJ, Hurley AM, McKibbin T, et al.** Publication of noninferiority clinical trials: changes over a 20-year interval. *Pharmacotherapy.* 2011;31(9):833-9. doi: <https://doi.org/10.1592/phco.31.9.833> [published Online First: 2011/09/20].
4. **Flacco ME, Manzoli L, Boccia S, et al.** Head-to-head randomized trials are mostly industry sponsored and almost always favor the industry sponsor. *J Clin Epidemiol.* 2015;68(7):811-20. <https://doi.org/10.1016/j.jclinepi.2014.12.016> [published Online First: 2015/03/10].
5. **Piaggio G, Elbourne DR, Pocock SJ, et al.** Reporting of noninferiority and equivalence randomized trials: extension of the consort 2010 statement. *JAMA.* 2012;308(24):2594-604. <https://doi.org/10.1001/jama.2012.87802>.
6. Food and Drug Administration. Guidance for industry: non-inferiority clinical trials to establish effectiveness. In: Center for Drug Evaluation and Research (CDER) CfBEaRC, ed. Food and Drug Administration webpage; 2016:56.
7. **Motzer RJ, Hutson TE, Cella D, et al.** Pazopanib versus sunitinib in metastatic renal-cell carcinoma. *N Engl J Med.* 2013;369(8):722-31. <https://doi.org/10.1056/NEJMoa1303989>.
8. **Rehal S, Morris TP, Fielding K, et al.** Non-inferiority trials: are they inferior? A systematic review of reporting in major medical journals. *BMJ Open.* 2016;6(10): e012594. <https://doi.org/10.1136/bmjopen-2016-012594>.
9. **Doshi P, Hur P, Jones M, et al.** Informed consent to study purpose in randomized clinical trials of antibiotics, 1991 through 2011. *JAMA Intern Med.* 2017. <https://doi.org/10.1001/jamainternmed.2017.3820>.