JAMA Guide to Statistics and Methods

Noninferiority Trials
Is a New Treatment Almost as Effective as Another?

Amy H. Kaji, MD, PhD; Roger J. Lewis, MD, PhD

Sometimes the goal of comparing a new treatment with a standard treatment is not to find an approach that is more effective but to find a therapy that has other advantages, such as lower cost, fewer adverse effects, or greater convenience with at least similar efficacy to the standard treatment. With other advantages, a treatment that is almost as effective as a standard treatment might be preferred in practice or for some patients. The purpose of a noninferiority trial is to rigorously evaluate a new treatment against an accepted and effective treatment with the goal of demonstrating that it is at least almost as good (ie, not inferior).

Related article page 2340

In this issue of *JAMA*, Salminen et al describe the results of a multicenter noninferiority trial of 530 adults with computed tomography–confirmed acute appendicitis who were randomized either to early appendectomy (the standard treatment) or to antibiotic therapy alone (a potentially less burdensome experimental treatment).[1]

Use of the Method

Why Are Noninferiority Trials Conducted?

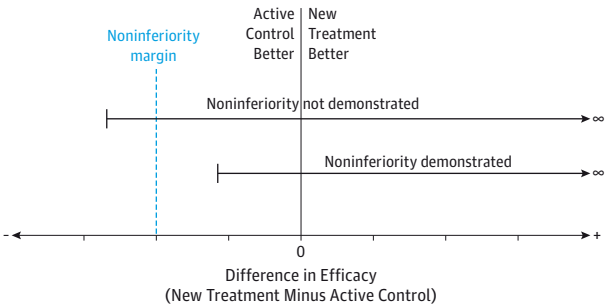
In a traditional clinical trial, a new treatment is compared with a standard treatment or placebo with the goal of demonstrating that the new treatment has greater efficacy. The null hypothesis for such a trial is that the 2 treatments have the same effect. Rejection of this hypothesis, implying that the effects are different, is signaled by a statistically significant *P* value or, alternatively, by a 2-tailed confidence interval that excludes no effect. While the new treatment could be either superior or inferior, the typical trial aims to demonstrate superiority of the new treatment and is known as a superiority trial. Since a superiority trial is capable of identifying both harmful and beneficial effects of a new therapy vs a control (ie, a current therapy), a 2-tailed 95% CI can be used to indicate the upper and lower limits of the difference in treatment effect that are consistent with the observed data. The null hypothesis is rejected, indicating that the new therapy differs from the control, if the confidence interval does not include the result that indicates absence of effect (eg, a risk ratio of 1 or a risk difference of 0).[2] This is equivalent to a statistically significant *P* value.

Although superiority or inferiority of a new treatment can be demonstrated by a superiority trial, it would generally be incorrect to conclude that the absence of a significant difference in a superiority trial demonstrates that the therapies have similar effects; absence of evidence of a difference is not reliable evidence that there is no difference. An active-controlled noninferiority trial is needed to determine whether a new intervention, which offers other advantages such as decreased toxicity or cost, does not have lesser efficacy than an established treatment.[3-6] Noninferiority trials use known effective treatments as controls because there is little to be gained by demonstrating that a new therapy is not inferior to a sham or placebo treatment.

The objective of a noninferiority trial is to demonstrate that the intervention being evaluated achieves the efficacy of the established therapy within a predetermined acceptable noninferiority margin (**Figure**). The magnitude of this margin depends on what would be a clinically important difference, the expected event rates, and, possibly, regulatory requirements. Other determinants of the noninferiority margin include the known effect of the standard treatment vs placebo; the severity of the disease; toxicity, inconvenience, or cost of the standard treatment; and the primary end point. A smaller noninferiority margin is likely appropriate if the disease under investigation is severe or if the primary end point is death.[3-6]

The sample size required to reliably demonstrate noninferiority depends on both the choice of the noninferiority margin and the assumed true relative effects of the compared treatments.[3-6] An active-controlled noninferiority trial often requires a larger sample size than a superiority trial because the noninferiority margins used in noninferiority studies are generally smaller than the differences sought in superiority trials. Just as important is the assumed effect of the experimental treatment relative to the active-control treatment. The assumed effect may be that the experimental treatment is worse than the control but by a smaller amount than the noninferiority margin, that the 2 treatments are equivalent, or even that

Figure. Two Different Possible Results of a Noninferiority Trial, Summarized by 1-Tailed Confidence Intervals for the Relative Efficacy of the New and Active-Control Treatments



In the top example, the lower limit of the confidence interval lies to the left of the noninferiority margin, demonstrating that the results are consistent with greater inferiority (worse efficacy) than allowed by the noninferiority margin. Thus, the new treatment may be inferior and noninferiority is not demonstrated. In the lower example, the lower limit of the confidence interval lies to the right of the noninferiority margin, demonstrating noninferiority of the new treatment relative to the active-control treatment. The overall result of the trial is defined by the lower limit of the 1-sided confidence interval rather than by the point estimate for the treatment effect, so point estimates are not shown.

the experimental treatment is more effective. These 3 options will result in larger, intermediate, and smaller required sample sizes, respectively, to achieve the same trial power—the chance of demonstrating noninferiority—because they assume progressively better efficacy of the experimental treatment.

Because a noninferiority trial only aims to demonstrate noninferiority and does not aim to distinguish noninferiority from superiority, it is analyzed using a 1-sided confidence interval (Figure) or hypothesis test. Typically, a 1-sided 95% or 97.5% CI (−L to ∞; negative values represent inferiority of the experimental treatment) is constructed for the difference between the 2 treatments, and the lower limit, −L, is compared with the noninferiority margin. Noninferiority is demonstrated if the lower confidence limit lies above or to the right of the noninferiority margin.[3-6]

What Are the Limitations of Noninferiority Trials?

A negative noninferiority trial does not in general demonstrate inferiority of the experimental treatment, just as a negative superiority trial does not demonstrate equivalence of 2 treatments.

A noninferiority trial is similar to an equivalence trial in that the objective of both is to demonstrate that the intervention matches the action of the established therapy within a prespecified margin. However, the objective of a noninferiority trial is only to demonstrate that the experimental treatment is not substantially worse than the standard treatment, whereas that of an equivalence trial is to demonstrate that the experimental treatment is neither worse than nor better than the standard treatment.[3]

Why Was a Noninferiority Trial Conducted in This Case?

Ever since McBurney demonstrated reduced morbidity from pelvic infections with appendectomy, the standard treatment for acute appendicitis has been surgery, which requires general anesthesia, incurs increased cost, and is associated with postoperative complications, such as wound infections and adhesions. Thus, a less invasive approach with similar efficacy might be preferred by many patients and physicians. Three randomized trials summarized in a recent Cochrane analysis demonstrated equipoise as to whether appendicitis can successfully be treated with antibiotics alone rather than surgery.[7] Because appendectomy is viewed as the standard treatment, it was considered the active control with which the less invasive experimental antibiotic treatment was to be compared.

To design the clinical trial, Salminen et al assumed a surgical treatment success rate of 99% and prespecified a noninferiority margin of −24% based on clinical considerations. This is equivalent to saying that if the rate of treatment success with antibiotics alone could be shown to be no worse than 24% worse than the rate with surgery, then the antibiotic-only strategy would be clinically noninferior. As this study demonstrates, the selection of the noninferiority margin is often subjective rather than based on specific criteria.

How Should the Results Be Interpreted?

The results demonstrated that all but 1 of 273 patients randomized to the surgery group underwent successful appendectomy, resulting in a treatment efficacy of 99.6%. In the antibiotic treatment group, 186 of 256 patients available for follow-up had treatment successes, for a success rate of 72.7%; 70 of the 256 patients underwent surgical intervention within 1 year of initial presentation. Thus, the point estimate for the difference in success rate with the antibiotic-only strategy was −27.0% and the associated 1-tailed 95% CI would range from −31.6% to infinity. Because that interval includes efficacy values worse than the noninferiority margin of −24%, noninferiority cannot be demonstrated.

Caveats to Consider When Looking at a Noninferiority Trial

Noninferiority active-controlled trials often require a larger sample size than placebo-controlled trials, in part because the chosen noninferiority margins are often small. The required sample size for a noninferiority trial is highly dependent on the noninferiority margin and the assumed effect of the new treatment; this assumed effect must be clearly stated and realistic.

The primary analysis for a superiority trial should be based on the intention-to-treat (ITT) principle because it is generally conservative in the setting of imperfect adherence to treatment. However, analyzing a noninferiority trial by ITT could make an inferior treatment appear to be noninferior if poor patient adherence resulted in both treatments being similarly ineffective. Thus, when analyzing a noninferiority trial, both ITT and per-protocol analyses should be conducted. The results are most meaningful when both approaches demonstrate noninferiority.

A noninferiority trial does not distinguish between a new treatment that is noninferior and one that is truly superior and cannot demonstrate equivalence.

ARTICLE INFORMATION

Author Affiliations: Department of Emergency Medicine, Harbor-UCLA Medical Center, Torrance, California (Kaji, Lewis); David Geffen School of Medicine at UCLA, Torrance, California (Kaji, Lewis); Los Angeles Biomedical Research Institute, Los Angeles, California (Kaji, Lewis); Berry Consultants, LLC (Lewis).

Corresponding Author: Roger J. Lewis, MD, PhD, Department of Emergency Medicine, Harbor-UCLA Medical Center, 1000 W Carson St, Box 21, Torrance, CA 90509 (roger@emedharbor.edu).

Section Editors: Roger J. Lewis, MD, PhD, Department of Emergency Medicine, Harbor-UCLA Medical Center and David Geffen School of Medicine at UCLA; and Edward H. Livingston, MD, Deputy Editor, *JAMA*.

Conflict of Interest Disclosures: The authors have completed and submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest and none were reported.

REFERENCES

1. Salminen P, Paajanen H, Rautio T, et al. Antibiotic therapy vs appendectomy for treatment of uncomplicated acute appendicitis: the APPAC randomized clinical trial. *JAMA*. doi: 10.1001/jama.2015.6154.

2. Young KD, Lewis RJ. What is confidence? I: the use and interpretation of confidence intervals. *Ann Emerg Med*. 1997;30(3):307-310.

3. Kaji AH, Lewis RJ. Are we looking for superiority, equivalence, or noninferiority? *Ann Emerg Med*. 2010;55(5):408-411.

4. Mulla SM, Scott IA, Jackevicius CA, et al. How to use a non-inferiority trial: Users' Guides to the Medical Literature. *JAMA*. 2012;308:2605-2611.

5. Tamayo-Sarver JH, Albert JM, Tamayo-Sarver M, Cydulka RK. Advanced statistics: how to determine whether your intervention is different, at least as effective as, or equivalent. *Acad Emerg Med*. 2005;12(6):536-542.

6. Piaggio G, Elbourne DR, Altman DG, Pocock SJ, Evans SJW; CONSORT Group. Reporting of noninferiority and equivalence randomized trials: an extension of the CONSORT statement. *JAMA*. 2006;295(10):1152-1160.

7. Wilms IM, de Hoog DE, de Visser DC, Janzing HM. Appendectomy vs antibiotic treatment for acute appendicitis. *Cochrane Database Syst Rev*. 2011;(11):CD008359.