Noninferiority and Equivalence Trials in Medical Research

Patrick Schober, MD, PhD, MMedStat,* and Thomas R. Vetter, MD, MPH†

KEY POINT: Equivalence or noninferiority testing is required when the aim of a study is to show that a treatment works as good as, or not worse than, another treatment.

In this issue of *Anesthesia & Analgesia*, Tsan et al report results of a randomized, controlled trial in which authors compared the laryngeal view during conventional laryngoscopy in the bed-up-head-elevated (BUHE) position to the view obtained during videolaryngoscopy in supine position. These authors appropriately used a noninferiority design aiming to show that the laryngeal view is not significantly worse in the BUHE position.

As useful therapies or interventions are currently available for many conditions, research increasingly focuses on whether alternative treatments could be equally useful as (equivalent to), or not worse than (noninferior to), the standard treatment. Such studies are particularly useful when the alternative treatment has advantages, such as lower costs or fewer side effects, and could replace the current standard treatment, when the alternative treatment effect is equivalent or noninferior.

![Figure](https://example.com/figure.png)

**Figure.** The upper part of this figure shows hypothetical CIs, which are identical across all 3 panels. The interpretation depends on whether a superiority, equivalence, or noninferiority testing framework is applied (see text for details). The bold text under the x-axis corresponds to the rejection region, meaning that the null-hypothesis is rejected when the CI entirely lies in that region. Note that a “nonsignificant” result of a superiority test (CIs B, C, and D) does not necessarily correspond to an equivalent or noninferior treatment effect. The lower part of this figure is Figure 2 from Tsan et al, showing the mean difference and 2-sided 98% CI of the laryngeal view (expressed as POGO). The authors set the noninferiority margin at −15%. The entire CI lies in the noninferiority region, allowing the authors to claim noninferiority. Note this is analogous to the CI K in the upper part of this figure. CI indicates confidence interval; POGO, percentage of glottic opening.

From the *Department of Anesthesiology, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands; and †Department of Surgery and Perioperative Care, Dell Medical School at the University of Texas at Austin, Austin, Texas.

Address correspondence to Patrick Schober, MD, PhD, MMedStat, Department of Anesthesiology, Amsterdam UMC, Vrije Universiteit Amsterdam, De Boelelaan 1117, 1081 HV Amsterdam, the Netherlands. Address e-mail to p.schober@amsterdamumc.nl.

Copyright © 2020 International Anesthesia Research Society
Researchers commonly use superiority testing to assess whether one treatment is better than the other, and are then tempted to conclude that the treatments are “similar,” “comparable,” or “equivalent” when the between-group difference is not statistically significant. However, this is a classic misinterpretation. A nonsignificant result of a superiority test only demonstrates that there is insufficient evidence to claim a difference, but it does not exclude relevant differences. Other statistical methods are hence required when the study aim is not to show superiority, but equivalence or noninferiority.

Two different treatments would likely never have identical effects, and even if so, this would be fundamentally impossible to prove. Therefore, an equivalence trial aims to show that the effects differ by no more than a clinically acceptable amount. This amount is termed the equivalence margin, and it must be specified a priori to avoid post hoc decision making. Similarly, a noninferiority trial assesses whether the alternative treatment is no worse than the standard treatment by more than a prespecified noninferiority margin.

While equivalence or noninferiority can be tested with hypothesis tests, the intuitive and statistically equivalent confidence interval (CI)–based approach is often preferred. This concept is illustrated in the Figure here, which shows hypothetical CIs of the mean difference of a continuous outcome between 2 treatments, where a negative difference indicates a worse performance of the treatment of interest.

In superiority testing, a 2-sided $[100 \times (1 - \alpha)]\%$ CI that does not contain the null-difference (A) corresponds to a “statistically significant” difference in a 2-sided hypothesis test on significance level $\alpha$. In contrast, a CI that contains the null value (B, C, and D) is compatible with lack of a difference, and is thus “nonsignificant.”

In equivalence or noninferiority testing, the CI is compared to the equivalence or noninferiority margin ($\delta$), respectively. A 2-sided $[100 \times (1 - 2\alpha)]\%$ CI that lies entirely in the equivalence region (G, between $-\delta$ and $+\delta$) or noninferiority region (K and L, to the right of $-\delta$) indicates equivalence or noninferiority on significance level $\alpha$, respectively. Otherwise, equivalence or noninferiority cannot be claimed.

Note the difference in the above confidence level between 2-sided superiority testing versus equivalence and noninferiority testing, because equivalence and noninferiority testing actually test 1-sided hypotheses. However, in practice, 2-sided 95% CIs are commonly used for all these tests, as a more conservative 0.025 $\alpha$ level is often recommended for equivalence or noninferiority testing. Tsan et al1 use a 0.01 $\alpha$ level, corresponding to a $[100 \times (1 - 0.02)]\% = 98\%$ CI, as shown in Figure 2 of their manuscript, and included in our Figure here.

REFERENCES