Delta: The Value That Matters in Fatigue Risk Management

Amanda C. M. Lamp; Michael J. Rempe; Gregory L. Belenky

INTRODUCTION: Noninferiority or equivalence testing are often used when comparing a novel pharmaceutical, operation, or procedure to the current standard designated as safe. Noninferiority and equivalence testing require estimates of a metric called delta: the margin of meaningful difference. Inappropriate delta margins can lead to invalid conclusions, thereby creating uncertainty about a study's scientific credibility. We recommend that a working group be convened with the following goals: 1) to evaluate delta values currently in use in aviation; 2) to determine if it is possible to develop a systematic, evidence-based, and replicable process to derive delta values based on statistical properties from population data, rather than a mixture of evidence- and opinion-based processes; and 3) based on the findings of the second goal, update the current delta values in use in aviation. This working group should include, at a minimum, government agencies and other key stakeholders using these values within operational settings.

KEYWORDS: aviation, safety performance indicators, delta values, noninferiority testing, equivalence testing, sleepiness, fatigue, safety, flight operations.

Lamp ACM, Rempe MJ, Belenky GL. *Delta: the value that matters in fatigue risk management*. Aerosp Med Hum Perform. 2021; 92(2):127–128.

Noninferiority or equivalence testing are often used when comparing a new or novel pharmaceutical, operation, protocol, piece of equipment, or procedure to the current standard designated as safe.[1] The aim of equivalence testing is to test whether or not a new protocol, operation, etc. has analogous results to a control or what is otherwise considered a standard protocol.[9] The aim of noninferiority testing is to test whether a new or novel group is found to be not inferior to (i.e., equal to or better than; as safe as or safer than) the standard of care or standard of safety.[2] The literature on traditional difference testing acknowledges that just because two groups are not statistically different does not mean they are equivalent.[5,6] This is critical to underline because the consequences of such a problematic assumption can be dire, e.g., in healthcare, assuming that two drugs are not statistically different and are therefore equivalent can lead to significant, previously undetected adverse effects. Therefore, equivalence testing or noninferiority testing must be performed to verify the equivalence or noninferiority (which encompasses equivalence and superiority) of the novel in comparison to the standard.

Both equivalence and noninferiority can be detected through the combination of graphical inspection of data and a modified *t*-test. To detect equivalence, the procedure is to use a two one-sided *t*-test and, to detect noninferiority, a one-sided *t*-test is used. In both cases, the parameter that is tested is the difference between the experimental and control values of a specific outcome, e.g., a safety performance indicator (SPI).[4] The modified *t*-tests that support equivalence and noninferiority testing require estimates of a metric typically referred to as delta: the margin (+ delta / − delta) above or below which the parameters being compared are no longer considered equivalent and/or noninferior.[3] Delta is expressed in SPI units and designates the margin of meaningful difference (also expressed as the smallest clinically or operationally relevant change[2]). This means that the range within the + and − delta is expected to be a clinically or operationally acceptable difference that does not have a significant effect on the safety or performance of the drug, device, procedure, protocol, or operation under investigation.[8] Therefore,

From the Occupational Sleep Medicine Group, Sleep and Performance Research Center, Department of Medical Education and Clinical Sciences, Elson S. Floyd College of Medicine, Washington State University, Spokane, WA, USA.

This manuscript was received for review in September 2020. It was accepted for publication in November 2020.

Address correspondence to: Amanda Lamp, Elson S. Floyd College of Medicine, Washington State University, 412 E. Spokane Falls Blvd., Spokane, WA 99202, USA; alamp@wsu.edu.

DOI: https://doi.org/10.3357/AMHP.5768.2021

when the mean and 95% (standard) confidence interval of the data fall within this range (i.e., between the + delta / − delta) it is considered tolerable and inconsequential.[10]

The statistical analyses available for equivalence and noninferiority testing with the required delta values have been used extensively within the pharmaceutical industry and by the Food and Drug Administration (FDA), specifically to approve a new pharmaceutical or medical device. The FDA also has published industry-wide guidance for any investigator conducting equivalence or noninferiority trials.[7] The FDA has a systematic, documented process for setting delta values. Typically, delta values are set from prior data and/or by a committee of experts prior to beginning an evaluation of a drug or device. To the authors' knowledge there is no systematic, widely available, transparent, documented process for setting delta values for Fatigue Risk Management in aviation even though there are a standard set of delta values used for assessing cognitive performance, sleep, fatigue, and sleepiness. The current delta margins, such as those used for assessment of sleep and alertness in medical trainees, are informed by consensus statements, experts in the respective fields, and studies on chronic sleep restriction, but not estimated specifically from the populations being studied[1] or, as in aviation, have been estimated from the population being studied but could be strengthened by industry-wide consensus from experts in the field with additional sources of airline data not used in the original study.[10] Inappropriate delta margins can yield inappropriate assumptions about statistical power and lead to invalid conclusions, thereby creating uncertainty about the scientific credibility of the study.[8] Thus, caution must be taken in, and sound logic must be applied to, the critical step of choosing the delta margins a priori for all SPIs due to potential safety consequences. The delta values should also be estimated from the population being studied.

We recommend that a working group be convened with the following goals: 1) to evaluate delta values currently in use in aviation that use the SPIs of sleep, fatigue, and performance; 2) to determine if it is possible to develop a systematic, evidence-based, and replicable process to derive delta values based on statistical properties from population data, rather than a mixture of evidence- and opinion-based processes; and 3) based on the findings of the second goal, update the current delta values in use in aviation that use the SPIs of sleep, fatigue, and performance. This working group should periodically re-evaluate the delta values as more data and technology become available. This working group should include, at a minimum, government agencies overseeing the use of these values (e.g., the Federal Aviation Administration) and other key stakeholders using these values (e.g., commercial airline representatives, union representatives, and respective sleep and performance scientists) within operational settings to put delta values on firm empirical footing.

ACKNOWLEDGMENTS

Financial Disclosure Statement: The authors have no competing interests to declare.

Authors and affiliation: Amanda C. M. Lamp, M.S., Ph.D., Michael J. Rempe, M.S., Ph.D., and Gregory L. Belenky, M.D., Occupational Sleep Medicine Group, Sleep and Performance Research Center, Department of Medical Education and Clinical Sciences, Elson S. Floyd College of Medicine, Washington State University, Spokane, WA, USA.

REFERENCES

1. Basner M, Asch DA, Shea JA, Bellini LM, Carlin M, et al. Sleep and alertness in a duty-hour flexibility trial in internal medicine. N Engl J Med. 2019; 380(10):915–923.
2. Greene CJ, Morland LA, Durkalski VL, Frueh BC. Noninferiority and equivalence designs: issues and implications for mental health research. J Trauma Stress. 2008; 21(5):433–439.
3. Gupta SK. Non-inferiority clinical trials: practical issues and current regulatory perspective. Indian J Pharmacol. 2011; 43(4):371–374.
4. Lamp A, Chen JMC, McCullough D, Belenky G. Equal to or better than: the application of statistical non-inferiority to fatigue risk management. Accid Anal Prev. 2019; 126:184–190.
5. Mascha EJ, Sessler DI. Equivalence and noninferiority testing in regression models and repeated-measures designs. Anesth Analg. 2011; 112(3):678–687.
6. Piaggio G, Elbourne DR, Pocock SJ, Evans SJW, Altman DG; CONSORT Group. Reporting of noninferiority and equivalence randomized trials: extension of the CONSORT 2010 statement. JAMA. 2012; 308(24):2594–2604.
7. U.S. Food and Drug Administration. Non-inferiority clinical trials to establish effectiveness. Guidance for Industry. [Accessed August 5, 2020]. Available from https://www.fda.gov/media/78504/download.
8. Walker E, Nowacki AS. Understanding equivalence and noninferiority testing. J Gen Intern Med. 2011; 26(2):192–196.
9. Wiens BL. Choosing an equivalence limit for noninferiority or equivalence studies. Control Clin Trials. 2002; 23(1):2–14.
10. Wu LJ, Gander PH, van den Berg M, Signal TL. Equivalence testing as a tool for fatigue risk management in aviation. Aerosp Med Hum Perform. 2018; 89(4):383–388.