



Superiority vs. equivalence/non-inferiority: study design and meeting abstracts

Peter J. Strouse¹

Received: 28 August 2018 / Accepted: 2 September 2018 / Published online: 12 September 2018
© Springer-Verlag GmbH Germany, part of Springer Nature 2018

In a superiority study, the authors hypothesize that one study is superior to another — i.e. A is superior to B. When there is a statistically significant difference between the studies (usually defined as $P < 0.05$), superiority of one study has been demonstrated. When there is not a statistically significant difference between the studies ($P \geq 0.05$), this does not necessarily indicate that the two studies are equivalent (A is equivalent to B or A is non-inferior to B) or that the opposite result is true (B is superior to A). Rather, it simply indicates that superiority was not demonstrated. Failure to demonstrate superiority could indicate one of two things: (1) the result is correct (A is truly not superior to B) and/or (2) the study lacked sufficient power/sample size (we don't know whether A is superior to B).

To demonstrate equivalency or non-inferiority, it is required to define what the acceptable difference between the studies is for the two studies to be considered equivalent and to do a power analysis/sample size calculation to determine the number of subjects necessary to demonstrate that there is no statistical significance between the study results [1]. The study must then include a sufficient number of patients as determined by the power analysis/sample size calculation. If the sample size is limited, the power of the study can be increased by requiring replicated reads of the diagnostic test, such as by having multiple radiology readers independently interpreting the diagnostic tests.

In this issue of *Pediatric Radiology*, May et al. [2] investigated scientific abstracts from the 2016 International Pediatric

Radiology Conjoint Meeting and Exhibition (IPR) relative to study design and appropriateness of conclusions. Alarming, they found a prevalence of false inference of study non-inferiority from what they determined to be superiority studies. According to May et al. [2], of the 194 abstracts presented at IPR 2016, 112 studies were of diagnostic accuracy comparing 2 or more diagnostic modalities, and 36 of these abstracts made “unfounded inferences of equivalence or similarity in diagnostic imaging performance.”

May et al. [2] used the 2016 IPR abstracts as a convenience sample, albeit highly relevant to our profession of pediatric radiology. Assuredly, their results are not specific to the 2016 IPR meeting nor are they specific to the field of pediatric radiology. However, this is a common challenge when studying pediatric pathology because patient numbers in diseases of childhood are typically smaller as compared to adult diseases such as heart disease, diabetes and adult malignancies.

Study design is very important. Investigators need to consider the construct of their study (i.e. superiority vs. non-inferiority), what their required sample size is to reach statistical significance and what valid conclusions can be reached from the study, as constructed and with sample size studied. In assessing results and reaching conclusions, word choice is highly important. Failure to demonstrate superiority does not necessarily mean demonstration of equivalency or non-inferiority unless the sample size is sufficient to have done so. Certainly, there is no intent of investigators to deceive but naively chosen words might infer conclusions that are not substantiated.

Finally, a comment on meeting abstracts is necessary. As alluded to here, any scientific study should include a power analysis/sample size calculation to determine the adequacy of the patient population to reach statistically significant conclusions. Sample size calculation should be included in a meeting presentation. Sample size calculation should be included in a published paper. However, should sample size calculations be required in scientific

✉ Peter J. Strouse
Rad-Ped-Radiol-Journal@med.umich.edu

¹ Section of Pediatric Radiology,
C. S. Mott Children's Hospital,
Department of Radiology,
University of Michigan Health System,
Room 3-231, 1540 E. Hospital Drive,
Ann Arbor, MI 48109-4252, USA

abstracts for professional meetings such as the Society for Pediatric Radiology (SPR), as May et al. [2] suggest? If there is space, including sample size calculation in a meeting abstract is definitely useful. But space is limited. Abstracts are a concise description of the study and by necessity lack in detail. Abstracts are bounded by word or character limits. CONSORT (consolidated standards of reporting trials) [3] and STARD (standards for reporting of diagnostic accuracy) [4] checklists for scientific meeting abstracts do NOT include sample size calculation. Neither the SPR [5], nor the Radiological Society of North America (RSNA) [6], American Roentgen Ray Society (ARRS) [7], American Institute of Ultrasound in Medicine (AIUM) [8], American Association of Physicists in Medicine (AAPM) [9] or International Society for Magnetic Resonance in Medicine (ISMRM) [10] make note of a requirement of including sample size in their instructions for preparation of submitted abstracts.

Without advanced training in study design, these concepts (superiority vs. non-inferiority and implications of statistical significance and proper conclusions) are difficult to grasp. This would be an excellent topic for an educational review in print or at the SPR annual meeting.

Acknowledgments The author thanks Geetika Khanna and Ronald L. Thomas for their helpful review and comments.

Compliance with ethical standards

Conflicts of interest None

References

1. Ahn S, Park SH, Lee KH (2013) How to demonstrate similarity by using noninferiority and equivalence statistical testing in radiology research. *Radiology* 267:328–338
2. May LA, Guilleman RP, Zhang W, Orth RC (2018) Absence of evidence is not evidence of absence: a pervasive fallacy of inference in pediatric radiology scientific abstracts. *Pediatr Radiol*. <https://doi.org/10.1007/s00247-018-4222-9>
3. Hopewell S, Clarke M, Moher D et al (2008) CONSORT for reporting randomised trials in journal and conference abstracts. *Lancet* 371:281–283
4. Cohen JF, Korevaar DA, Gatsonis CA et al (2017) STARD for abstracts: essential items for reporting diagnostic accuracy studies in journal or conference abstracts. *BMJ* 358:j3751
5. Society for Pediatric Radiology (2018) Instructions for authors. <https://www.pedrad.org/Portals/5/Events/2019/Annual%20Meeting/Instructions%20for%20Authors.pdf?ver=2018-08-13-133624-440>. Accessed 27 Aug 2018
6. Radiological Society of North America (2018) RSNA abstract guidelines and information. <https://abstract.rsna.org/pdf/2018%20RSNA%20Abstract%20Submissions%20Guidelines.pdf>. Accessed 27 Aug 2018
7. American Roentgen Ray Society (2018) ARRS 2019 annual meeting abstract submission program. https://abstracts.arrs.org/Content/ARRS_Abstract_Instructions.pdf. Accessed 27 Aug 2018
8. American Institute of Ultrasound in Medicine (2018) Abstract scorecard. <https://www.conferenceabstracts.com/cfp2/login.asp?EventKey=OVLBWATW>. Accessed 27 Aug 2018
9. American Association of Physicists in Medicine (2016) Annual meeting online submission (AMOS) system instructions. <https://www.aapm.org/meetings/2017AM/AbstractSubmission.asp>. Accessed 27 Aug 2018
10. International Society for Magnetic Resonance in Medicine (2018) 2018 call for abstracts. <https://www.ismrm.org/18m/2018-call-for-abstracts/>. Accessed 27 Aug 2018