

- 2 Marsh HW. The influence of student, course, and instructor characteristics in evaluation of university teaching. *Am Educ Res J* 1980;17 :219–37.
- 3 Marsh HW, Roche LA. Effects of grading leniency and low workload on students' evaluations of teaching: popular myth, bias, validity, or innocent bystanders? *J Educ Psychol* 2000;92:202–28.
- 4 Marsh HW. Students' evaluations of university teaching: dimensionality, reliability, validity, potential biases and usefulness. In: Perry RP, Smart JC, ed. *The Scholarship of Teaching and Learning in Higher Education: an Evidence-based Perspective*. Dordrecht: Springer 2007;319–83.
- 5 Schwarz N. How the questions shape the answers. *Am Psychol* 1999; 54:93–105.

## Using equivalence designs to improve methodological rigor in medical education trials

Martin G Tolsgaard<sup>1</sup> & Charlotte Ringsted<sup>2</sup>

Editor – Experimental studies comparing the effects of one type of training against those of no training are often deemed to be of little value to the development of theory.<sup>1</sup> Few such papers are now published in leading journals dedicated to medical education. However, comparing different types of educational intervention with one another stands a greater chance of both providing empirical evidence of their relative effectiveness and strengthening the underlying theoretical frameworks that might yield even better ways of training. Furthermore, in such comparisons no trainees are restricted from training, a situation that might be considered unethical in contexts in which the effectiveness of training is well established.<sup>2</sup> Examples of research comparing different types

of intervention include studies on peer teaching versus teaching provided by faculty staff,<sup>3</sup> dyad versus single-learner training<sup>4</sup> and high-versus low-fidelity simulation training.<sup>5</sup> These studies often use superiority designs to test whether or not one type of training is better than another. Unfortunately, when these studies fail to demonstrate statistically significant differences, equivalence between the interventions is often concluded. Equivalence – or non-inferiority – cannot be established based on negative results. Rather, it requires certain methodological considerations that are not accounted for in a superiority trial.

An equivalence trial requires a pre-defined educationally or clinically relevant maximum difference, also called delta, below which the new and existing type of training can be regarded as equivalent. Sample size calculations should be based on delta values, which often results in samples larger than those in corresponding superiority trials. Finally, to assess equivalence, the researcher compares the confidence interval for any observed difference in outcome with the delta value, rather than simply comparing the mean values of the interventions being examined.

Hence, a failed superiority study cannot be converted to an equivalence trial as delta values must be determined *a priori*, and equivalence should not be inferred from negative results, which may just as well be explained by small sample sizes or inadequate measurement instruments. An extension of the CONSORT statement has now been developed for reporting equivalence trials<sup>6</sup> and current literature provides several useful methodological guidelines for conducting this type of research.<sup>7</sup>

The adoption of equivalence trials in medical education research would undoubtedly improve the scientific rigor of many comparative studies in which the objective is to assess the effectiveness or efficiency of new types of training that may yield learning outcomes equivalent to those of existing methods of training but at lower costs, with less time, better adherence to training, higher feasibility or greater user satisfaction. To date, only a very limited number of equivalence trials have been published over the last decade in the field of medical education<sup>8,9</sup> despite both theoretical and methodological support for this approach.

<sup>1</sup>Centre for Clinical Education and Juliane Marie Centre, Copenhagen University Hospital Rigshospitalet, Copenhagen, Denmark

<sup>2</sup>Department of Anesthesia and The Wilson Centre. University of Toronto and The University Health Network, Toronto, Canada

*Correspondence:* Martin G Tolsgaard, Dept. 4001, Juliane Marie Centre Copenhagen University Hospital Rigshospitalet, Blegdamsvej 9, 2100 Copenhagen O, Denmark. Tel: 00 45 61 30 30 72; E-mail: martintolsgaard@gmail.com

doi: 10.1111/medu.12313

## REFERENCES

- 1 Egan M, Mainous AG III. The tension between educational equivalency and equipoise in medical education research. *Fam Med* 2012;**44** (1):5–6.
- 2 Ziv A, Wolpe PR, Small SD, Glick S. Simulation-based medical education: an ethical imperative. *Acad Med* 2003;**78** (8):783–8.
- 3 Ross MT, Cameron HS. Peer-assisted learning: a planning and implementation framework: AMEE Guide no. 30. *Med Teach* 2007;**29** (6):527–45.
- 4 Tolsgaard MG, Bjørck S, Rasmussen MB, Gustafsson A, Ringsted C. Improving efficiency of clinical skills training: a randomised trial. *J Gen Intern Med* 2013;**18**:1072–7.
- 5 Norman G, Dore K, Grierson L. The minimal relationship between simulation fidelity and transfer of learning. *Med Educ* 2012;**46**:636–47.
- 6 Piaggio G, Elbourne DR, Pocock SJ, Evans SJ, Altman DG; CONSORT Group. Reporting of non-inferiority and equivalence randomised trials: extension of the CONSORT 2010 statement. *JAMA* 2012;**308** (24):2594–604.
- 7 Pocock SJ. The pros and cons of non-inferiority trials. *Fundam Clin Pharmacol* 2003;**17** (4):483–90.
- 8 Mpotos N, Lemoyne S, Calle PA, Deschepper E, Valcke M, Monsieurs KG. Combining video instruction followed by voice feedback in a self-learning station for acquisition of basic life support skills: a randomised non-inferiority trial. *Resuscitation* 2011;**82** (7):896–901.
- 9 Watson K, Wright A, Morris N *et al.* Can simulation replace part of clinical time? Two parallel randomised controlled trials. *Med Educ* 2012;**46**:657–67.