

# Evaluating the evaluation

Vance W Berger

National Cancer Institute, University  
of Maryland Baltimore County,  
Biometry Research Group,  
Rockville, MD, USA

## Dear editor

Zhang et al<sup>1</sup> sought to determine which adjustment method is the best. That is a laudable objective, but their approach leaves quite a bit to be desired. When we cut to the chase, we find that they pre-supposed that the analysis of covariance (ANCOVA) was ideal, and, presumably, confirmed this empirically by noting that the ANCOVA results were most aligned with the ANCOVA gold standard. This is fairly perplexing logic. Had any of the other methods been chosen instead as the gold standard, then that method would have been found to be the best by virtue of agreeing with its own results. This is hardly a compelling endorsement. Beyond that, even if the authors did use a more reasoned approach, how can one trial be used to validate an analysis?

An analysis is good or bad based on how well its results align with the underlying reality of the situation. In a simulation study, we would know this reality. In actual trials, we do not. There is no gold standard. Moreover, there is only one trial being considered. This is most assuredly not the way to compare analysis techniques. It is worth noting, however, that ANCOVA relies on normality, among other assumptions, for its validity. Since the data are never actually normally distributed, the method is never technically valid.<sup>2</sup> This indisputable fact should give us pause before we blindly accept so fanciful a method. There is a valid and exact method that is based on a ranking of the pairs, pre and post, without having to make any assumptions at all.<sup>3</sup> Surely, this method, which was developed for categorical data but applies equally well to continuous data, might have been considered as well. Finally, it is stated that “no methods are available for analysis of data that are ‘missing not at random’.” This is patently untrue<sup>4</sup> and almost reaches the level of lunacy attained two sentences later when it is stated that because there were not much missing data, the data were assumed to be missing at random. Missing data can never be demonstrated to be missing at random. This is an entirely academic construct with no application in the real world.<sup>5</sup>

## Disclosure

The author reports no conflict of interest in this work.

## References

1. Zhang S, Paul J, Nantha-Aree M, et al. Empirical comparison of four baseline covariate adjustment methods in analysis of continuous Outcomes in randomized controlled trials. *Clin Epidemiol*. 2014;6:227–235.

Correspondence: Vance W Berger  
National Cancer Institute, University  
of Maryland Baltimore County, Biometry  
Research Group, 9609 Medical Center  
Drive, Rockville, MD 20850, USA  
Tel +1 240 276 7142  
Email vb78c@nih.gov

2. Berger VW. Pros and cons of permutation tests in Clinical trials. *Stat Med.* 2000;19:1319–1328.
3. Berger VW, Zhou YY, Ivanova A, Tremmel L. Adjusting for ordinal covariates by inducing a partial ordering. *Biomet J.* 2004;46(1):48–55.
4. Lachin JM. Worst-rank score analysis with informatively missing observations in clinical trials. *Control Clin Trials.* 1999;20(5):408–422.
5. Berger VW. Conservative handling of missing data. *Contemp Clin Trials.* 2012;33:460.

## Authors' reply

Shiyuan Zhang  
Lehana Thabane

Department of Clinical Epidemiology and Biostatistics,  
McMaster University, Hamilton, ON, Canada

Correspondence: Lehana Thabane  
Department of Clinical Epidemiology and Biostatistics,  
McMaster University, Hamilton, ON, Canada  
Email thabanl@mcmaster.ca

## Dear editor

We thank Dr Vance Berger for his interest in and insightful discussion of our paper.<sup>1</sup> Essentially, we agree with his sentiments, however we need to clarify the central points of his discussion, namely: 1) the objective of the study; and 2) the handling of missing data in our statistical analyses.

The objective of the study: we designed and conducted the study to assess the sensitivity or robustness<sup>2</sup> of the findings from the original trial (MOBILE trial)<sup>3</sup> by varying a specific factor in the analysis, in this case the method of analysis. We judged robustness based on the magnitude, direction, and statistical significance of the effect estimate. The stated objectives of the study are clear on this goal. We did not intend to compare the methods on the basis of their statistical properties—something we agree with the author on, that can be done only through simulation. We also discuss this issue in the Discussion section, supplemented with findings from published simulation studies.

The “missing at random (MAR)” assumption in multiple imputation: again we agree with the author that the assumption of MAR cannot be verified. Given this potential limitation, a commonly used approach to handle missing data is to assess the impact of “missingness” on the findings through some sensitivity analyses. The goal is to check whether, under certain assumptions of missingness, the findings would remain robust if the missing data were imputed through some imputation strategy. We did this in our study and we found that the results remained robust irrespective of the method of handling missing data. We also discuss the limitations of the imputation methods used to assess robustness in the Discussion section. We make no claims about the validity of the MAR assumption.

We hope that this response provides some clarity on the objectives and context to our paper.

## Disclosure

The authors report no conflicts of interest in this communication.

## References

1. Zhang S, Paul J, Nantha-Aree M, et al. Empirical comparison of four baseline covariate adjustment methods in analysis of continuous outcomes in randomized controlled trials. *Clin Epidemiol.* 2014;6:227–235.
2. Thabane L, Mbuagbaw L, Zhang S, et al. CH (2013). A tutorial on sensitivity analyses in clinical trials: the what, why, when and how. *BMC Med Res Methodol.* 2013;13:92.
3. Paul JE, Nantha-Aree M, Buckley N, et al. Gabapentin does not improve multimodal analgesia outcomes for total knee arthroplasty: a randomized controlled trial. *Can J Anaesth.* 2013;60(5):423–431.

### Clinical Epidemiology

### Publish your work in this journal

Clinical Epidemiology is an international, peer-reviewed, open access, online journal focusing on disease and drug epidemiology, identification of risk factors and screening procedures to develop optimal preventative initiatives and programs. Specific topics include: diagnosis, prognosis, treatment, screening, prevention, risk factor modification,

Submit your manuscript here: <http://www.dovepress.com/clinical-epidemiology-journal>

systematic reviews, risk & safety of medical interventions, epidemiology & biostatistical methods, and evaluation of guidelines, translational medicine, health policies & economic evaluations. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use.

Dovepress