



Published in final edited form as:

*Disabil Rehabil.* 2015 ; 37(18): 1692–1693. doi:10.3109/09638288.2014.1002580.

## Conclusion of “Nordic Walking for Geriatric Rehabilitation: A Randomized Pilot Trial” is Based on Faulty Statistical Analysis and is Inaccurate

David B. Allison<sup>1,2,3</sup>, Michelle S. Williams<sup>4</sup>, Gregory A. Hand<sup>5</sup>, John M. Jakicic<sup>6</sup>, and Kevin R. Fontaine<sup>2,7</sup>

<sup>1</sup>Department of Biostatistics, University of Alabama at Birmingham

<sup>2</sup>Nutrition Obesity Research Center, University of Alabama at Birmingham

<sup>3</sup>Office of Energetics, University of Alabama at Birmingham

<sup>4</sup>Division of Preventive Medicine, University of Alabama at Birmingham

<sup>5</sup>School of Public Health, West Virginia University

<sup>6</sup>Department of Health and Physical Activity, University of Pittsburgh

<sup>7</sup>Department of Health Behavior, University of Alabama at Birmingham

---

Randomized controlled trials (RCTs) are generally recognized as the strongest method for inferring causal inferences about the effects of treatments. As the National Research Council of the National Academies of Science state, “a randomized trial, clinical trial, or true experiment, is considered the gold standard for determining the relationship of an agent to a health outcome or adverse side effect.” [1] Yet, any RCT can only be as good as its execution and reporting. If the data are not analyzed appropriately and the results not described accurately, then even well-designed RCTs can be misleading.

The recent paper by Figueiredo et al. [2] reports the results of an RCT comparing the effects of Nordic walking (NW) to those of usual overground walking (OW) on a number of outcome variables in older adults, including gait speed. The conclusion section of the paper’s abstract consists of the following sentence: “NW is 106% more effective in improving gait speed among elderly than OW,” and similar statements are made in the body of the text. For older adults, this purported conclusion seems of great value. Yet, is it supported by the data?

The instructions for authors for *Disability and Rehabilitation* [3] recommend adherence to the CONSORT Guidelines [4] when reporting RCTs. The CONSORT guidelines state, “Confidence intervals should be presented for the contrast between groups. A common error

---

Address correspondence to: David B. Allison, PhD, University of Alabama at Birmingham, Ryals Building, Room 140J, 1665 University Boulevard, Birmingham, Alabama 35294, Phone: (205) 975-9167, Dallison@uab.edu.

Declaration of interest

Supported in part by NIH grants P30DK056336, T32DK062710, R25DK099080, and R25HL124208. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH or any other organization.

is the presentation of separate confidence intervals for the outcome in each group rather than for the treatment effect.” [4] Unfortunately, Figueiredo et al. [2] do exactly what the CONSORT guidelines advise against. They show that gait speed changed significantly in the NW group, but did not change significantly in the OW group.

### What Would Be a Proper Statistical Test?

It is well established that in a parallel groups RCT (as the Figueiredo et al. [2] study was), the correct test for the effect of treatment is a test of the *between*-groups difference in the outcome. In contrast, showing that one group changed significantly from baseline and another did not is neither an equivalent, nor valid test of differences between two interventions. A thorough and excellent exposition of this point was provided by Bland and Altman [5] in a paper aptly titled ‘*Comparisons against baseline within randomised groups are often used and can be highly misleading.*’ As Bland and Altman show, such a testing procedure will often yield conclusions that are inconsistent with an appropriate between-groups test and would yield markedly inflated type-1 error rates under many circumstances.

Using the data (means and standard deviations) reported by Figueiredo et al. [2] in their Table II, we used a simple, freely available online module to calculate an ordinary t-test on the between-groups difference in change in gait speed. [6] The result is a *t* of 0.8879 with 24 degrees of freedom, which is nowhere near statistically significant (two-tailed P value of 0.38). Thus, an appropriate statistical test of the hypothesis that NW improved gait speed relative to OW should have led to the conclusion that the null hypothesis of no effect could not be rejected and, therefore, the study offers no compelling evidence that NW affects gait speed differently than does OW. Given the above, the stated conclusion of the paper is inaccurate.

### Concerns About the Proper Effect Size Metric

We also note that the effect size metric used by Figueiredo et al. [2] is a very unusual one. We have not seen the effect size metric the authors used before and they provide no reference for its justification. The effect size metric in their terms was:

$$\frac{\text{NW means difference} \div \text{SD at baseline}}{\text{OW means difference} \div \text{SD at baseline}}$$

It is notable that the authors state that “Despite the small sample size, Shapiro-Wilk, Skewness, and Kurtosis tests showed that all variables followed a normal distribution.” Hence, the effect size metric they are calculating is then (plausibly) a ratio of two normally distributed variables. Moreover, because each of the normally distributed variables in the ratio involves a difference score, their means are plausibly zero (and in the actual sample are close to zero). This is noteworthy because the ratio of two normally distributed variables with mean zero follows a Cauchy distribution, and a statistic with a Cauchy distribution is a poor choice for an effect size metric because a Cauchy distribution’s mean and variance do not exist (i.e., are undefined) [7]. Thus, no confident conclusions can be drawn about the

effect size Figueiredo et al. [2] drawn. Given that this calculation serves as the basis of their statement that, ‘NW is 106% more effective’, it renders this conclusion misleading.

There are many RCTs reported in the literature which offer good examples of using between-group tests and effect size metrics which have been well-studied and whose properties have been described by statistical scientists (e.g., [8]). Two good examples of papers in which appropriate between groups tests are conducted and established effect size metrics are used are references [9] and [10]. For example, in [9] the authors used a between groups test and also tested for group by time interactions (equivalent to a between groups test on change scores [11]) and used Cohen’s *d*, a standard and established effect size metric. Similarly, in [10], the authors also studied gait speed as an outcome, found “No significant effect of group, time, or group\*time adjusted for sex and baseline gait speed category”, and on that basis appropriately concluded “Both programmes were equally effective in maintaining walking capacity after discharge from stroke rehabilitation; or were equally ineffective in improving walking capacity.”

On the basis of the points above, the conclusion of the Figueiredo et al. [2] paper is incorrect. We hope that authors, readers, and editors of this and other journals are more cognizant of the issues raised herein and explicated so well by Bland and Altman [5].

## Acknowledgments

The authors would like to acknowledge Dr. Robert Matthews for his contribution to the preparation of this letter.

## References

1. National Research Council. Reference Manual on Scientific Evidence: Third Edition. Washington, DC: The National Academies Press; 2011. Available from: [/catalog/13163/reference-manual-on-scientific-evidence-third-edition](#)
2. Figueiredo S, Finch L, Mai J, Ahmed S, Huang A, Mayo NE. Nordic walking for geriatric rehabilitation: a randomized pilot trial. *Disability and Rehabilitation*. 2012 Oct 15; 35(12):968–975. [cited 2014 November 21]. [PubMed: 23066879]
3. Rehabilitation and Disability instructions for authors [Internet]. London: 2014. [cited 2014 November 21]; Available from: <http://informahealthcare.com/userimages/ContentEditor/1406303829764/Disability%20and%20Rehabilitation.pdf>
4. Moher D, Hopewell S, Schulz KF, Montori V, Gotzsche PC, Devereaux PJ, Elbourne D, Egger M, Altman DG. CONSORT 2010 Explanation and Elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ*. 2010 Mar 23; 340(mar23 1):c869–c869. [cited 2014 November 21]. [PubMed: 20332511]
5. Bland JM, Altman DG. Comparisons against baseline within randomised groups are often used and can be highly misleading. *Trials*. 2011 Dec 22.12(1):264. [cited 2014 November 21]. [PubMed: 22192231]
6. GraphPad Software, Inc. T test Calculator [Internet]. La Jolla, CA: 2014. [cited 2014 November 21]; Available from: <http://www.graphpad.com/quickcalcs/ttest1.cfm>
7. Wolfram [Internet]. Champaign, IL: 2007. [cited 2014 November 21]; Available from: <http://reference.wolfram.com/language/ref/CauchyDistribution.html>
8. Hedges, LV.; Olkin, I. *Statistical Methods for Meta-analysis*. Orlando, FL: Academic Press; 1985.
9. Henry M, Cohen SR, Lee V, Sauthier P, Provencher D, Drouin P, Gauthier P, Gotlieb W, Lau S, Drummond N, et al. The Meaning-Making intervention (MMi) appears to increase meaning in life in advanced ovarian cancer: a randomized controlled pilot study. *Psycho-Oncology*. 2010 Dec; 19(12):1340–1347. [PubMed: 20878857]

10. Mayo NE, MacKay-Lyons MJ, Scott SC, Moriello C, Brophy J. A randomized trial of two home-based exercise programmes to improve functional walking post-stroke. *Clinical Rehabilitation*. 2013 Jul; 27(7):659–671. [PubMed: 23503738]
11. Huck SW, McLean RA. Using a repeated measures ANOVA to analyze the data from a pretest-posttest design: A potentially confusing task. *Psychological Bulletin*. 1975; 82(4):511–518.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript