Author's response to reviews

Title: Effect of Head and Limb Orientation on Trunk Muscle Activation During Abdominal Hollowing in Chronic Low Back Pain

Authors:

Kevin Parfrey (kparfrey@hotmail.com)
Sean Gibbons (stabilityphysio@gmail.com)
Eric J Drinkwater (edrinkwater@csu.edu.au)
David G Behm (dbehm@mun.ca)

Version: 5 Date: 30 January 2014

Author's response to reviews: see over
We continue to disagree with the reviewer that the only assessment of the “degree of evidence for the presence of any effect... is measured by the p-value”. The p-value is only one of several measures of distribution. While the p-value is the most common due to its link with the null hypothesis test, our argument around presenting confidence limits in lieu of p-values is twofold:

1. We are purposefully avoiding testing a null hypothesis

To conclude that a “non-significant finding” indicates “no effect” violates the central premise of the null hypotheses and represents a fundamental misunderstanding of the null hypothesis test. A null hypothesis can be “accepted” or “fail to be accepted” but a null hypothesis cannot be “rejected”; a non-significant result “fails to find a difference” but cannot conclude “there was no difference”. While this may seem to argue semantics, it is the difference between failing to find something versus finding nothing. Therefore no research using a null hypothesis test can indicate that “there was no difference between groups” and represents a flaw in the conclusion of a large volume of published research. The real-world damage of this error in clinical medicine is highlighted by Prinz, Schlange, & Asadullah (2011).

Even the US Supreme Court has ruled that “(the defendant’s) argument rests on the premise that statistical significance is the only reliable indication of causation. This premise is flawed.” (Matrixx v Siracusano, p. 11). The papers “The Cult of Statistical Significance” (Ziliak & McCloskey, 2009) and “On p-values” (Iacobucci, 2005) highlight the flaw in the null hypothesis test very well.

2. Presenting both p-values and confidence limits is redundant

The reviewer suggests we present both p-values and confidence limits. We have previously indicated that the p-value is only one of several methods of illustrating imprecision of the mean estimated difference. Confidence limits and effect sizes are a more intuitive way for clinicians to interpret the effects of an intervention by presenting the range of possible effects rather than a single mean. Furthermore, if statistical significance is still desired by the reader, it can be inferred from the confidence limit: a mean with a confidence limit that crosses the zero is not statistically significant. Therefore, presenting confidence limits is more information rich and further presentation of p-values is redundant.

In summary, we are not presenting p-values because we feel that there is no value in null hypothesis testing since it can only fail to be rejected. Even if the reviewer disagrees with this position and insists on observing statistical significance, statistical significance can be inferred from the confidence limits. Confidence limits have the added advantage over p-values of demonstrating the expected range of responses (with 95% confidence) to clinicians. Therefore, if inference about statistical significance can be made from a confidence limit but confidence limits also provide more intuitive information about the range of possible clinical effects of the intervention, we see no role in presenting p-values.

We are by no means alone in expressing research findings with confidence limits rather than p-values with many clinical journals intonating that confidence limits “…are, in general, the best reporting strategy.” (APA, 2010, p. 34). Journals and associations re-thinking their reliance on p-values and null hypothesis testing include the American Psychological Association (APA, 2010) the American College of Sports Medicine (Hopkins, et al., 2009) and Nature (Prinz, Schlange & Asadullah, 2011). We particularly would like to highlight Iacobucci (2005) as a concise and interesting summary.

To provide the readers with some additional context we added the following information to the results section.

Statistical significance can be inferred from the 95% confidence limits. If 95% confidence limits cross the zero, the mean must have a p>0.05, because the lower limit is less than zero while the upper limit is greater than zero. If, however, the 95% confidence limits in the figures are both on the same side of the zero, the mean has a p<0.05 (Figures 2-3).

Furthermore we added the following explanation in the discussion:

The expression of our results may not be familiar to all readers. We purposefully omitted p-values and discussion of statistical significance. While both p-values and 95% confidence limits can be used to infer statistical significance (see last paragraph of results section), 95% confidence limits are much more information-rich to the clinician. The p-value only represents the probability of the mean response to a treatment not being zero. Concluding that a mean was “unlikely to be zero” is not a clinically useful conclusion because it does not express the variability in responses in clinically meaningful units. However, expressing results using a mean in conjunction with upper and lower confidence limits allows clinicians to easily interpret the likely effects they can expect from an intervention.