Reviewer's report

Title: Reliability, construct and discriminative validity of clinical testing in subjects with and without chronic neck pain

Version: 2  Date: 7 July 2014

Reviewer: David Walton

Reviewer's report:

Intro
1. Line 72 re: gaze stability widely described – requires a citation?

Design
2. Line 104: Why was examiner C not blinded to patient status?

Procedure
3. Line 150: Subjects performed SWAY and PPT tests? This sounds like they did it on themselves.

Sample Size
4. Line 160: In a two one-sided test analysis… Is this the correct term?

Statistical analysis
5. Line 173-174: I’m not sure that the formula for SEM is correct. Should it not be SDdiff x sqrt(1-ICC)? As written, the sqrt(2) factors cancel each other out in your MDC formula. Perhaps your formula is simply one I’m not familiar with.

Results
6. Line 195: A purely semantic comment, but PCS makes me intuitively think of the Pain Catastrophizing Scale. Perhaps others don’t have my same hang up, but the authors may consider calling it something like SF36PCS, or follow the same convention as used for the Mental Component Score immediately after. Not as elegant, but might prevent confusion?
7. Line 194: This is the first time the BDI acronym is used. In the pain field this would be assumed to be the Beck Depression Inventory. If that’s the case, you may need to justify the use of the scale for this study.
8. Line 197: Just a comment, but strange that 5 of 21 controls stated their neck pain condition changed over the course of a week. What happened there?
9. A general comment: when reporting comments, unless there’s a compelling reason for reporting p values to less than 1/100th of a point, round them to 2 decimal places (e.g. <0.01)
10. I tend to follow Paul Stratford’s lead on the GPE and assume that a range of -1 to +1 (admittedly on a 15 rather than 11 point scale) is adequately similar to consider the condition stable. I’m not overly prescriptive here, no question that restricting to a zero is a rigorous approach to test-retest reliability. The only
reason I make this comment is that by doing so you've reduced your sample to arguably underpowered and no doubt your CI’s will be large. Just a comment, do with it what you like.

11. When reporting your ICCs, consider using the word ‘to’ rather than a dash ‘-‘ which can look like a ‘minus’ sign to those not paying attention.

Discussion

12. As I'm reading the Discussion on the different tests I'm realizing that I have no idea at this point how the different tests were conducted (e.g. ROM, JPE). I know the description of the tests are in the Appendix, and I suspect the conservation of words in the test may have been due to word limitations. This may also be a comment for the editor to consider if word count is an issue. However, I would encourage the authors to consider adding in even some cursory descriptions of, for example, ROM and the tool used (‘Cervical Range of Motion was captured using a ....’). Without this information, and unless I looked at the Appendix first, I am unable to adequately interpret the results and discussion.

13. The JPE discussion is a little disjointed, in that one sentence states that previous ICCs were reported higher, and then two sentences later they are reported to be lower. In other words, it’s hard to know if your results are in keeping with previous work or not. Can this be somehow reoriented to be a bit more clear?

14. In your gaze stability discussion, I would urge caution in switching between ICC and % agreement, as not all readers will understand the difference.

15. As I continue to read through your discussion (SWAY), I'm even more convinced of my previous comment, that the readers need at least some cursory information on how these tests were performed, captured and analyzed. At least one of the authors has surely forgotten more than I currently know about postural sway analysis, and while I don’t think a full dissertation on the subject is required here, providing some basic information on the nature of data collection and analysis would certainly be helpful in the text.

Level of interest: An article of importance in its field

Quality of written English: Acceptable

Statistical review: Yes, and I have assessed the statistics in my report.

Declaration of competing interests:

I declare that I have no competing interest tests