Reviewer's report

Title: Measurement properties of the Western Ontario Shoulder Instability Index (WOSI) in Dutch patients with shoulder instability.

Version: 4  Date: 22 December 2013

Reviewer: Björn Salomonsson

Reviewer's report:

This is an interesting study on shoulder instability scores. However to my surprise I received 2 almost identical papers for review the same week from different journals, where only the main focus on WOSI or OSIS scores was exchanged.

There are several good reasons for presenting two separate works on two separate scores, but in this case I will start with the common questions for both of the works and then go for the specific questions of each study separately.

To conclude my impression that this is 2 parts of one study:
It is the same authors.
It is the same ethical decision number.
It is identical patients description.
It is identical methods.
It mentions the other part of the study in the discussion text.

Regarding both studies I have the same major points:

Score validation and other statistical analyses may not be considered as common knowledge in the large population of orthopaedics that would be interested as readers of these manuscripts. You introduce rather complex analyses with very little explanation on why and how it should be used, the references are not that much easier to the average reader like me.

If the less common analyses could be explained in a more practical way to the readers I believe the manuscripts would gain a lot in usefulness, as well as spread the knowledge regarding the score validation process.

A) Major Compulsory Revisions
Mixing the paper version score with a digital version must be validated!

This mix of formats is done both for WOSI and OSIS.

To my knowledge these different formats first must be validated to each other to be used as interchangeable units? Just like any other translation in language must be validated, the translation in formats also need validation just as much.
How was the different cohorts distributed? First paper format then change to digital in the second half, or paper format at first visit and digital in re-test situation?

Are any of the other scores also mixing paper and digital versions?

In description of the WOSI digital version you write that it was a 0-10 scale instead of a 0 mm-100 mm VAS scale, and that is also a rather special change of the format of the WOSI score. Were there no intermediate values or just the 0-10 as whole steps?

Would it be possible to validate the paper forms to the digital format of the scores as well within these studies, then that would be very valuable to the validity of the study.

B) Minor Essential Revisions

Floor and Ceiling effects, do you measure in the score or in the patients!

In both texts you discuss the floor and ceiling of the patients (in discussion it is obvious) and say that your SDC limits for F/C: “measurement error cannot be detected in these patients” and “detection of improvement and deteriorations that are distinct from measurement errors at follow-up”.

That indicates to me that you measure F/C in the patients, and not in the score validation?

To me measurement errors of the score would be e.g. by the ruler measuring mm on the VAS scale in WOSI. I think that in validation of the score I would choose to have F/C of the absolute result of the score.

But the clinical F/C by the way you have measured is very valuable in the clinical study, and should ideally be done in every study to check the actual cohort as well as responsiveness?

However, you exemplify the SDC as measuring for change at follow-up, but you did not present any change/follow-up, and no Effect size or SRM for responsiveness. Is your calculation of the SDC comparable to an actual change at follow up to discuss detection of improvement or deterioration. To me this is confusing because it is two different thing in one analysis, the score itself, and the SDC in the patients, and difficult to understand how it should be interpreted.

C) Minor Essential Revisions

SDC

Please explain in more detail the rational for using SDC as Floor and Ceiling effects.

Your SDC results in about 20 % of the upper and lower scores as F/C?

Do you really intend to propose that a score of 80 % is clinically the same as 95% ?

I agree that it might be possible for two individuals to have every other measure
equal and still score 80 % or 95 % in WOSI (or similar in OSIS). But it is difficult to see that if it is the means in two groups, then the difference would be substantial. I think this could be explained how you use the SDC to the readers to better understand the analysis as individual measures or in groups.

Is the SDC homogenous through the whole scale of the scores?
We know that e.g. Effect sizes vary in different parts of a score/scale. And it may also be more difficult to earn the very last point in a score at the extremities of the scale, than to have change in the central part. How does the SDC behave regarding F/C effects?

You also discuss change measurements with regard to SDC?
Perhaps you could be more specific on that you have not had any change to measure in the scores (no follow-up after treatment) and (if I understand correctly) that this change is only a calculation/estimation from one measure at one time, without any change seen in the patients. And how this is useful instead of a measured change.

D) Discretionary Revisions
CFA
In the manuscripts you conclude that the CFA was unreliable och not useful if I understand correctly?

Pleas inform on how it could have been useful if it would have a better fit?
I suggest that the analysis description could be shortened and just confirm that it was not very useful, but that you inform on the conclusions that this led to (that score changes are individual and that the total WOSI score is recommended instead of domains).

The full CFA analysis description and tables could be presented in a digital supplement instead of in the manuscript, with the rather large tables as well.

E) Major Compulsory Revisions
Not using the full information of the data you really possess?
In both the manuscripts you say that the WOSI/OSIS score would need to be compared further but that the other score is not yet fully validated and published. This may be true but it is a bit unfair to the readers when you publish this in parallel manuscripts, and you do propose this as limitations to your studies and that would like to propose which of the scores that have the superior measurement properties?

It would not be impossible to do that in at least one of the manuscripts?
And if you have a possibility to measure The effect size or SRM it would be advantageous as well.
It does seem likely that it would come available (from the same study cohort?):
Determination and comparison of the smallest detectable change (SDC) and the minimal important change (MIC) of four-shoulder patient-reported outcome measures (PROMs).

van Kampen DA, Willems WJ, van Beers LW, Castelein RM, Scholtes VA, Terwee CB.

Other points: Discretionary Revisions

WOSI is Canadian, not from USA to my knowledge (PROM/WOSI descriptions)

Chronbachs alpha is related to the number of questions as well, the more questions the higher correlation, that might be mentioned?

WOSI manuscript: Minor Essential Revisions

Patients part: Copy-Paste error from other manuscript, OSIS should be WOSI.

Results part:

Reliability: Values does not match, WOSI retest score should be 959? As in table 4

SEM and SDC (abstract and results part) of SEM in WOSI is % not points I think, that would be easier to understand in the text.

Discretionary Revisions

Table 4:

The difference in test/re-test values for the 3 last domains seems very strange compared to the total score and the change? And the ICC does not seem to be affected as well.
Were there really those differences in the domains without worsening the ICC?

Table 5:

See C/F and SDC above.
Your C/F limits is half of the score scale, it must be explained better how this is interpreted.

References OK

Level of interest: An article whose findings are important to those with closely related research interests

Quality of written English: Not suitable for publication unless extensively edited
**Statistical review:** Yes, and I have assessed the statistics in my report.

**Declaration of competing interests:**

I declare that I have no competing interests