Reviewer's report

Title: Inter- and intra-observer reliability of clinical movement-control tests for marines

Version: 4 Date: 18 October 2012

Reviewer: Harald Ekedahl

Reviewer's report:

This study investigated inter-observer and intra-observer reliability and validity of movement-control tests in a cohort of (male) marines. The tests have previously proven reliable, however not for this particular study group. Moreover, the validity of the tests was still to be investigated.

For the reliability analysis, Cohen’s Kappa coefficient was used. Logistic regression analysis was used to prove validity (sensitivity/specificity) of the test. The reference variables were three (worst pain the last 6 months, average pain the last 6 months and present pain) pain rating scales for different body regions. The subjects were assessed at baseline and followed up one week later.

The results proved moderate to almost perfect inter-reliability at baseline and at the follow-up. The intra-observer reliability was only fair to moderate. Three tests showed a coefficient >0.4. The authors suggest that the results from the regression analyses indicate that combinations of low- and high-threshold movement-control tests had some discriminative validity for previous back pain, but not for present pain. One variable, BKFO, seem to have both adequate reliability and validity.

Major compulsory revisions:

1) The authors investigated reliability of six tests. One major concern in reliability studies is the variability between observers when assessing the subjects. The two observers showed no such variability. Inter-observer reliability is proven. However, the tests show different results one week later. The authors need to interpret these results in the discussion in order to create hypothesis for future research.

2) The results from the tests do not mirror the present pain. Thus, the tests show low validity in assessing present back pain. The authors suggest adequate validity of the tests. This is based on average back pain the last 6 months. This variable is gold standard. Further description is needed; the validity of this or a similar variable would be of interest.

3) The authors need to explain why the tests correlate with average pain the last 6 months when no correlation with present pain is shown. How can a test correlate with average pain the last 6 months when the test is not consistent after one week.

Minor essential revisons:
1) The title only mentions reliability. Validity is not mentioned.

2) The pros and cons of using a cohort of (male) marines are not fully discussed. Can the results be used for other categories? Previous studies were on civilians. Further discussion would enlighten the reader.

3) Further studies are needed to in order to consider the tests valid. Both cross-sectional and longitudinal studies. Please discuss.

4) Gold standard is not discussed. Since the tests are not pain provoking test, self-reported disability instead of pain measure might be suggested as dependent variable in future studies.

This article is of great interest. Self-reported pain and disability are considered as the most valuable outcomes. Yet, rehabilitation often focuses upon motor control and thus, validated clinical tests are essential.

**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 have no competing interests in relation to this paper.