Author's response to reviews

Title: Reliability and validity study of Persian modified version of MUSIC (musculoskeletal intervention center) - Norrtalje questionnaire

Authors:

Akbar Alipour (akbar.alipour@ki.se)
Mostafa Ghaffari (mostafa.ghaffari@ki.se)
Irene Jensen (irene.jensen@ki.se)
Batoul Shariati (batoul_shariati@yahoo.com)
Eva Vingard (eva.vingard@medsci.uu.se)

Version: 2 Date: 29 April 2007

Author's response to reviews: see over
Thank you very much for your comments. They were very useful and we have revised our manuscript accordingly.

1. About the concept of musculoskeletal disorders:
Based on your suggestion we added two paragraphs (in introduction and discussion section) to describe more about this concept.

2. Golden standard for comparing in validity study:
Totally there are not any golden standards for musculoskeletal disorders.
Based on your suggestion we described this problem in discussion section.

3. About the number of questions:
We are completely agreed with you about the huge number of items (297). This is an expanded questionnaire with different sections. Deleting one domain (scale) or sub domain does not affect the validity of questionnaire and it depends on research group and the aims of using questionnaire. On the other hand, as you see this questionnaire is about different regions of body (70 questions). Usually in practice we use restricted part of body region and it decrease the number of questionnaire.
Finally as you referred there is new short version of some part of questionnaire (General Health questionnaire GHQ-12) that can be substitute in this case and we referred in text.

4. The interval of test-retest:
If re-test is administered immediately or after only a short period, the subjects, remembering their earlier responses tend to answer the questions in the same way and here we have assessed memory rather than reliability (reliability is underestimated).
If we increase the interval, as you said maybe we have had different changes in subject that affect answers which are unrelated to the test reliability (reliability is overestimated). With the exception of three persons (we had repeat test 23 days after test) the interval was 21 days, so we changed 3-4 weeks to three weeks.

On the other hand, in this study we followed the subjects between test retest from the point of factors that could affect their answer (for example: any changes in their job and tasks, workplace physical and psychosocial exposures, new musculoskeletal complaint, new disease based on sickness absence registration system of company).

Based on your suggestion we added some sentences in text in this regard.

5. About ICC in reliability:
Beside the advantages of ICC, we are agreeing with you about ICC restriction in reliability studies. There is no doubt that ICC is the ratio of the between subjects variance divided by the total variance and in some instances can produce misleadingly high level of reliability.

We referred to this restriction and extended our description.

In this study with considering the kind of questions and their answer we used Interclass Correlation Coefficients (ICCs) for the rating scale, and kappa coefficients for dichotomous answers and categorical data. From this point for most of items we used Kappa. With considering that probably in former version we did not refer to this point maybe it confuse reader, so we described more details and referred to this important point.

We are trying to have a separate booklet for validity and reliability study of this questionnaire and there we will have all tables that indicate about agreement plot or compute repeatability. In this manuscript with considering word and table restriction hardly we could refer to this points in more details.

6. Based on your suggestion I shortened Table 2 and added referred details in footnote.

7. We again checked the references and wrote based on BMC format.
Reviewer 2:

Thank you very much for your comments. They were very useful and we have revised our manuscript accordingly.

1. As we referred in text and figure 1, our test-retest study for reliability was completed after changes in translation and focus group discussion.

2. Our decision for items with poor ICC:
   Totally there were a few items with poor ICC
   It could be different reasons for poor ICC:
   As you referred some part came back to job tasks of participants (for example appreciating work by customers is confusing for workers), or regular modest exercise (some workers considered their job as exercise but we have not had this problem for office or technician workers).
   Based on final decision we suggest deleting items with low ICC (only 5 items).
   This deleting does not affect total validity of questionnaire.
   Based on your suggestion we discussed more about in text.