Reviewer’s report

Title: Adaptation of Chinese and English versions of the Ankylosing Spondylitis Quality of Life (ASQoL) scale for use in Singapore

Version: 0 Date: 05 May 2017

Reviewer: Peter Klooster

Reviewer’s report:

This is a potentially relevant and interesting study on the translation and validation of the ASQoL for use in Singapore. In its current form, however, I feel that more information needs to be provided on several steps of the study and analyses to comprehensively judge the findings. Also, the overall claim and conclusion drawn currently seem somewhat overly optimistic to me given the several limitations of the current study and the actual results of the (very basic) analyses done.

Abstract

1) The defined objective in the abstract is quite different from the aim described at the end of the introduction, and in my opinion formulated somewhat too ambitious and vague. To establish true equivalence of a translated scale one should use rigorous advanced statistical modeling techniques in appropriate and large samples (e.g., by using multi-group factor analyses and/or IRT) which were not done in this study. Moreover, what does it mean to have "the same psychometric properties", do the authors mean that it should meet (similar) basic or minimum psychometric standards?

Methods

2) Please describe the test conditions of the measurements in the cross-sectional and test-retest study (e.g., paper and pencil questionnaires? Completed independently at home? Instructions? etc.).

Measures

3) Please provide all relevant psychometric properties of the ASQoL found in the original validations. This is essential to interpret and compare the current findings on convergent validity, internal consistency and test-retest reliably.

4) For all other measures, measures the authors should briefly describe their psychometric properties, preferably and if relevant, supplemented by internal consistency in the current study.
5) Did the patients complete the other measures in the same primary language as the ASQoL and were these validated for use in Singapore English and Chinese populations? If so, please provide references or state this more clearly if this was not the case (and add a reflection on this in the discussion).

Translation

6) Why exactly is this "dual-panel" methodology better than other traditional forward-backward methodologies (which are still generally considered the gold standard) and how was this established? A reference to one expert opinion paper in the discussion does not seem to suffice here.

Field testing

7) This section provides very little information on the actual procedure used here. What was the exact (standardized?) protocol of the CDIs? Was this based on established procedures? If so, please provide reference. Did the patients complete the questionnaires while thinking aloud? How were "difficulties and hesitation" observed and what (types of) questions were asked to examine comprehension and acceptability.

Statistical analysis

8) Cronbach's alpha >0.70 is considered adequate for group-level use of the scores only.

9) Were there no a-priori hypotheses with respect to the absolute and differential strength of the associations with other measures? This is highly desirable for adequate assessments of construct validity (see for instance Terwee et al) and just concluding that the scores are "moderately" correlated with a whole range of other patient-reported measures that are intended to measure different concepts is not very informative.

10) The fact that all other measures were patient-reported only (with the exception of PhGA) should be mentioned as an important limitation of this study in the discussion.

11) Using the same measures for both convergent validity assessment (correlations) and known-groups validity (dichotomized with t-test) essentially provides very little additional information. Again, it would have been very informative if know-groups could have been distinguished based on more clinical or objective standards.

12) Spearman's correlation coefficient is not a suitable measure for test-retest reliability as this only assesses the association between rank orderings of patient scores and not agreement between scores (for instance if all patients have changed substantially, but in the same direction, this will still result in a high coefficient). Especially given the relatively small sample size of the test-retest sample (which is actually smaller than those recommended for instance by Terwee et
al), the authors should report the intraclass correlation coefficient for agreement with 95% confidence intervals. Preferably supplemented with a Bland-Altman plot with limits of agreement. Together, these analyses would provide a much better picture of actual agreement between test-retest scores and systematic and unsystematic changes in scores.

13) The authors should define cut-off points for both internal consistency and test-retest reliability for different purposes and interpret the findings accordingly in their conclusions. Generally, coefficients >0.70 are considered adequate for group-level use of an instrument, while coefficients >0.90 are considered necessary for individual level use (see also Terwee et al). Both the internal consistency (0.88) and test-retest reliability (0.81) do meet the latter cut-off.

14) Please describe how and where the sample for the test-retest study was recruited and was this sample similar to the cross-sectional sample with respect to demographic and clinical characteristics?

Results

15) Field testing

Please provide more detailed information on actual comments and difficulties patients did have with the ASQoL instead of just stating that "overall patients found the instrument easy to understand and relevant". and "…no changes were necessary".

Discussion

16) The authors conclude that findings are comparable to those found for the original UK measure, but this cannot be concluded from the current paper (e.g., not without describing these psychometric properties in the UK).

17) Please provide more thorough attention to the limitations and possible impactions of the current study and assumptions made. Some of these that should at the very least be described are the reliance on patient-reported measures only for validity, and the combining of English and Chinese version in the analyses.

18) The authors conclude that the scale can be used in routine practice (abstract), clinical trials and cohort studies (discussion), but it is not made clear which types of psychometric properties need to be established and which minimal standards a scale must meet, to be confidently used for these different types of applications. E.g., is a reliability of 0.81 sufficiently high for (individual?) use in clinical practice and can you really recommend the measure for cohort studies and RCTs without having examined responsiveness. Overall, I feel that recommending the scale on the current findings is overly optimistic and a more conservative claims and conclusions should be drawn given the considerable limitations, examined properties and outcomes of the current study.

Are the methods appropriate and well described?
If not, please specify what is required in your comments to the authors.

No

**Does the work include the necessary controls?**
If not, please specify which controls are required in your comments to the authors.

Yes

**Are the conclusions drawn adequately supported by the data shown?**
If not, please explain in your comments to the authors.

No

**Are you able to assess any statistics in the manuscript or would you recommend an additional statistical review?**
If an additional statistical review is recommended, please specify what aspects require further assessment in your comments to the editors.

I am able to assess the statistics

**Quality of written English**
Please indicate the quality of language in the manuscript:

Acceptable

**Declaration of competing interests**
Please complete a declaration of competing interests, considering the following questions:

1. Have you in the past five years received reimbursements, fees, funding, or salary from an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

2. Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

3. Do you hold or are you currently applying for any patents relating to the content of the manuscript?

4. Have you received reimbursements, fees, funding, or salary from an organization that holds or has applied for patents relating to the content of the manuscript?

5. Do you have any other financial competing interests?

6. Do you have any non-financial competing interests in relation to this paper?
If you can answer no to all of the above, write 'I declare that I have no competing interests' below. If your reply is yes to any, please give details below.

I declare that I have no competing interests

I agree to the open peer review policy of the journal. I understand that my name will be included on my report to the authors and, if the manuscript is accepted for publication, my named report including any attachments I upload will be posted on the website along with the authors' responses. I agree for my report to be made available under an Open Access Creative Commons CC-BY license (http://creativecommons.org/licenses/by/4.0/). I understand that any comments which I do not wish to be included in my named report can be included as confidential comments to the editors, which will not be published.

I agree to the open peer review policy of the journal

Do you want to get recognition for reviewing this manuscript? Add a record of this review to Publons to track and showcase your reviewing expertise across the world's journals. Signing up is quick, easy and free!

No