Author's response to reviews

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

Authors:

Ying Ying Leung (katyccc@hotmail.com; katy.leung.y.y@singhealth.com.sg)

Weixian Lee (lee.weixian@yahoo.com.sg)

Nai Lee Lui (Lui.nai.lee@singhealth.com.sg)

Matthew Rouse (mrouse@galen-research.com)

Stephen McKenna (SMcKenna@galen-research.com)

Julian Thumboo (julian.thumboo@singhealth.com.sg)

Version: 2 Date: 27 Jun 2017

Author’s response to reviews:

25 Jun 2017

Editor in-Chief
BMC Musculoskeletal Disorders

Sir,

Re: BMSD-D-17-00196 R1

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

Thank you for your letter dated 20 Jun 2017. Following the Reviewers’ comments and advice, we have revised our original manuscript. We are pleased to re-submit our revised manuscript and our point-by-point responses to the Reviewers’ advice and comments. Changes are highlighted in the revised manuscript.

The revised manuscript has also been approved by all the authors and that they have given necessary attention to ensure the integrity of the work.

Thank you for your kind attention and we look forward to hearing your favorable reply.
Yours sincerely,

Dr. Ying-Ying Leung
Corresponding author
Department of Rheumatology & Immunology
Singapore General Hospital
Tel No.: (65) 63265276
Fax No.: (65) 62203321
E-Mail: katyccc@hotmail.com

Encl.

Point-to-point response to Reviewer(s)' advice and comments

Reviewer reports:

Reviewer 1: The authors revised the article and thereby answered my original comments.
Reply: Thank you.

Reviewer 2: The authors have reasonably dealt with most of the previous issues raised in the reviews. Moreover, the addition of the Rasch analysis makes the manuscript more interesting and informative. However, I am still of the opinion that the aim of the study (as mentioned in the abstract) is too ambitious and the overall conclusion much too far-reaching given the current study design, analyses performed and findings. I believe these need to be formulated more conservative before the manuscript can be considered for publication.

As I also mentioned in my previous review, to develop versions of the ASQoL that are really "equivalent" to the original requires much more rigorous analyses (e.g., CFA or IRT analysis of measurement invariance or absence of DIF between both the original UK and the Singapore Chinese and Singapore English versions of the scale) which are probably not feasible given the current sample size.

I think that based on the current study design and analyses you could claim at the max something like that the aim was to develop culturally appropriate Singapore Chinese and Singapore English versions of the scale and to provide a preliminary evaluation of the construct validity and reliability of the translated scales in AxSpA patients. Likewise, concluding that the scales can be used in Singapore for clinical trials or longitudinal studies based on the current analysis and results is still overly optimistic. For both applications, it needs to be shown that the scale, besides reliable and having adequate construct validity, is responsive to real changes in the construct of interest, and this has not been examined in the current study.
The addition of Rasch analysis (and comparison of DIF between the Singapore Chinese and Singapore English only) that the authors use as an argument in their reply for having dealt with these two issues, does not in my opinion change these fundamental issues.

Reply: We thank the reviewer for these comments. We have modified our conclusion in the light of the comments. We have also stated that the absence of information about responsiveness is a limitation of the study.

Action: In Discussion, page 23:

Finally, given the cross-sectional study design, it was not possible to evaluate the responsiveness of the new ASQoL language versions.

In conclusion, the present study has demonstrated that the Singapore English and Chinese versions of the ASQoL are culturally relevant, valid and reliable instruments when used with combined samples of AxSpA patients who speak either Chinese or English.

Additionally, I have two concerns about the section that was added about the sample size considerations on page 11-12. This section includes a support for a “decision to include” 100 patients for the cross-sectional study (based on rules of thumb for the Rasch analysis only, which was not even part of the previous version of the manuscript) and a "consideration of" 30 patients for the test-retest study.

First of all, I'm not convinced that these sample size calculations that are now reported were really performed a-priori, while the authors did write these down as if this is the case. It seems highly likely that these are actually more pragmatic, "post-hoc" power analyses trying to support the adequacy of the relatively small number of patients they recruited for both studies. This is in itself not necessarily wrong, but if this is indeed the case, it is important that the authors make this explicit (i.e., that no formal a-priori power analyses were performed) and move this section for instance to the discussion where this can be discussed in relation to other potential strengths and/or limitations of the sample(s).

Reply: We have removed the sample size estimation section, as it was not conducted a priori. This has been added to the Discussion as a limitation of the study.

Action: Discussion, page 21

A limitation of the study was the small sample size available for the validation study. Particularly, a larger sample size is required for Rasch analysis. Recommendations suggest seven times the number of items of the PROM of interest, and >100 patients; and more than 200 patients per subgroup for proper DIF evaluation (44). However, resource constraints did not allow us to recruit this number of AS patients, particular for the Chinese Speaking subgroup.

Second, the sample size calculation added for the test-retest study does not make sense. The authors added a sample size calculation based on determining if a certain correlation (in this case an "expected" r=0.85) is significantly different from zero. Obviously, this is not appropriate in
this context (it would be very disturbing if any test-retest coefficient would NOT be significantly different from zero!) and results in an extremely low required sample size of \( n=8 \). This is clearly not an appropriate sample size for any reliability study. A more appropriate sample size calculation would be for instance based on a reasonable width of the confidence interval around a certain value. Such a more realistic calculation would result in a required sample size much closer to the minimum standards generally suggested for test-retest studies (usually \( n>50 \)).

Reply: We have commented on the small sample size in the test-retest reliability study in the Discussion.

Action: page 23:

A sample size of at least 50 is recommended for establishing test-retest reliability (44). The actual sample available for test-retest reliability in our study (\( n=42 \)) may risk producing an inaccurate estimate of this property.