Author's response to reviews

Title: Comparison of the Adult Strabismus Quality of Life Questionnaire (AS-20) with the Amblyopia and Strabismus Questionnaire (ASQE) among adults with strabismus in China

Authors:

Zong-Hua Wang (whuanhuan910@gmail.com)
Hui Ren (renhui_tmmu@163.com)
Rosemary Frey (.frey@auckland.ac.nz)
Yang Liu (magicalyangliu@yahoo.com)
Deborah Raphael (d.raphael@auckland.ac.nz)
Xian-Yuan Wang (xywang_tmmu@sina.cn)
Wei Bian (tianwei198466@sohu.com)

Version: 6 Date: 20 July 2014

Author's response to reviews: see over
Dear Editors and Reviewers,

We are honoured to have the opportunity to revise our manuscript, and highly appreciate those insightful comments and constructive suggestions. We have carefully considered those comments and accordingly modified the manuscript.

As below, on behalf of all co-authors, I (Xian-yuan Wang) have responded to the comments point by point (the reviewer’s comments are in italics). Hope the editors and the reviewers will be satisfied with our responses and the revisions so that the revised manuscript can still be considered worthy for publication by BMC Ophthalmology. Again, thanks very much for your time and effort to provide such helpful guidance.

Referee 1

1. **Comment:** Abstract- Sentence 1 needs re-wording. Not all subjects with strabismus experience decreased HRQoL as is implied here.
   **Response:** We changed the word “can” to “might”, and re-worded the sentence to: “The impact of strabismus on visual function, self-image, self-esteem, and social interactions might decrease health-related quality of life (HRQoL).”

2. **Comment:** Introduction- Needs a space after “prejudice” in the sentence “In addition, the appearance of misaligned eyes can result in social prejudice – associating strabismus....”
   **Response:** It has been re-edited as suggested.

3. **Comment:** It does not state when data was collected – over what time period?
   **Response:** The time period was “between March and August 2013”, having been added into the method part.

4. **Comment:** You should state why your patients were there in the first place. Were they presenting to the clinic as they were experiencing symptoms or wanting surgery? More detail is required.
   **Response:** They were patients there for wanting surgery. So we added “strabismic adults was invited on the day of being added to the waiting list for a strabismus surgery”.

5. **Comment:** More information is needed as to the collection of the data. Was this pre- or post- clinical examination? Who issued the questionnaires – clinician or researcher? Were the questionnaires issued in the same way for each participant (i.e. always AS-20 first) or was this purposely varied? Or were no specific instructions given? Where were
the questionnaires completed? At home? Waiting area? How can you be sure as to the level of quality of the responses given? Were completion times recorded?

**Response:** All information about the above inquiries has been detailed.

1) This was a pre-surgery survey.
2) Questionnaires were issued by researchers (W.B. & Y.L.), who were also clinical nurses.
3) For the purposes of standardization, the AS-20 was given first and then followed by the ASQE for each participant.
4) Written information and simple verbal instructions of this survey has been given by the researchers.
5) Questionnaires were completed in a reception room in the hospital.
6) We failed to record the completion time; this is a great alarm for us to consider “completion time” when conducting a survey next time. But by estimate, most of the participants completed the questionnaires within 15 mins.

**6. Comment:** You issued the AS-20 and the ASQE, however recent publications have suggested that shortened forms of these instruments should be used. (Leske et al. Evaluation of the Adult Strabismus-20 (AS-20) questionnaire using Rasch analysis. Invest Ophthalmol Vis Sci 2012;53(6):2630-9. Vianya-Estopa et al. An evaluation of the Amblyopia and Strabismus Questionnaire using Rasch analysis. Invest Ophthalmol Vis Sci 2010;51(5):2496-503). Some of the analysis in the papers suggest a change in domain structure of the questionnaires, which could influence the results you have displayed. There should be some acknowledgement that you issued the full questionnaires, and rationale for this (perhaps in the Discussion section).

**Response:** As suggested, we discussed about the Rasch analysis and the consideration of using the full questionnaires.

“Rasch analysis is one of the methods based on modern psychometric theory, which has been increasingly used to amend and improve existing HRQOL instruments [31]. Unlike classical test theory which explores the difference between raw scores, and assumes equal contribution of each item to the overall assessment of latent trait; Rasch model explores the probability of individual response to items, which is defined by the latent trait (e.g., ability) of individual subjects and difficulty of each item on the instrument being used [31]. Leske et al. [31] and Vianya-Estopa et al. [32] have conducted a Rasch analysis of the AS-20 and ASQE, respectively; and the results suggested that some items should be
removed. Nevertheless, in this study, we still issued the full form of both questionnaires considering the cultural difference. The items that removed in Rasch analysis might fit the adult strabismus patients in China. Although Leske et al. reported no significant differential item functioning (DIF) for sex or age, they failed to explore the DIF for the characteristic of ethnicity [31]. But it is admitted that a Rasch analysis of these instruments among Chinese strabismus patients is necessary and worthy to explore in future study.”

7. **Comment:** In the Methods section you stated a Cronbach’s $\alpha$ of 0.70 was satisfactory. In the Results section you state that the reliability scores for the AS-20 and ASQE were satisfactory and yet these ranged from below 0.70.

**Response:** We have pointed out the subscale with Cronbach’s $\alpha$ value below 0.70 as follows: Internal consistency reliability for both the AS-20 and ASQE were satisfactory (Cronbach’s $\alpha$ ≥ 0.70) except the subscale of the ASQE “double vision” ($\alpha$= 0.68).

8. **Comment:** It may be better to describe the internal consistency of the whole questionnaire first, and then the domain scores. Internal consistency values relate to the number of items within the questionnaire.

**Response:** As suggested, we re-organized the paragraph regarding internal consistency in “results” part. The overall internal consistency came first, and then followed by domain scores.

“The overall Cronbach’s $\alpha$ were 0.91 for the AS-20 and 0.89 for the ASQE. Specifically, the Cronbach’s $\alpha$ was 0.90 and 0.85 for the “psychosocial” (10 items) and “function” (10 items) subscale, respectively; and for the ASQE subscales, the $\alpha$ value were 0.89 for “fear of losing the better eye” (2 items), 0.86 for the “distance estimation” (10 items), 0.88 for the “visual disorientation” (3 items), and 0.76 for the “social contact and appearance” (4 items).”

9. **Comment:** You assess the validity of the AS-20 and the ASQE by comparing scores with the VFQ-25. The amount of correlation between the scores obtained from the instruments can be assessed using Spearman rank correlation coefficients. A strong correlation is defined as >0.70; moderate 0.30 to 0.70; and weak <0.30 (Tabachnick BG, Fidell LS. Using Multivariate Statistics. 4th ed. Boston: Allyn and Bacon; 2001). The correlations you display in Table 2 whilst some of which show statistical significance, the size of the correlations are only weak to moderate. The reader may not be aware of this and either in the Methods or Results section there should be some explanation as to accepted values
for “strength” of correlation. It is misleading to go by statistically significant p values alone. (You allude to this in the Discussion section).

Response: We looked through the reference provided above as well as some other material related to Spearman rank correlation. We have added in the “Methods” part about the range of weak/moderate/strong correlation as follows.

“The correlation coefficient is defined as follows: < 0.30 stands for a weak correlation and little clinical applicability, even when statistically significant; between 0.30 and 0.50, moderate and > 0.50 (> 0.70 is even better), strong correlation”.

10. Comment: You do not appear to discuss whether your participants are functional or cosmetic cases of strabismus. Surely this could influence the responses they give on some of the items/domains. I struggle to see how you can argue that the ASQE is perhaps a superior instrument, or that it provides further information to the AS-20. The authors of the ASQE did define more domains, but does that really make it more accurate? If you look at the items within the AS-20 you could hypothesise sub-domains also, using the same categories.

Response: Actually we failed to understand what did the reviewer mean by ‘functional or cosmetic cases of strabismus’. We tried to explain here; but if the reviewer has more concerns, we are very happy to consult and discuss with the reviewer. As you said, function and cosmesis are two main concerns for strabismus patients. In this study, I think we have a mix of functional or cosmetic cases of strabismus. Just as the results showed, one might predict that patients with diplopia often present functional concerns, but we found they also have cosmetic needs; for the patients with amblyopia, one might expect that they have more functional difficulties, but instead, more psychosocial complaints were revealed.

We might have a superficial discussion about the sub-domains between the AS-20 and ASQE. We have modified the discussion to make it sound more reasonable and convincing.

“It should be noted that the ASQE distributes the HRQoL of strabismus adults into more dimensions in comparison to the AS-20 which only summarizes two subscales as “psychological” and “function”. But if one looks at the items within the AS-20, the same categories as those in ASQE could be summarized. For example, in the AS-20 the item 9 (People react differently to me because of my eyes) could be categorized into the ‘social
contact’ and the item 14 (I have problems with depth perception) be into the ‘distance estimation’.

11. Comment: The size of the deviation and HRQoL scores could be explored. It could be postulated that the larger the deviation, the more likely the individual is to have lower psychosocial HRQoL scores. This is another way of assessing the validity of an instrument (known-group differences). However, it should be acknowledged that deviation size (as measured by prism cover test) is not always a reliable indicator of poor cosmesis. This may be linked to your comments regarding the younger study cohort (compared to ASQE).

Response: We did explore the influence of deviation size on HRQoL, yet no significant differences were found in this study. The reason we assume is that the participants of this study are patients who came to the hospital for a surgery. They might be more concerned about the deviation size and its influence on appearance. Therefore, even patients with a smaller deviation size (for example, \( \leq 25 \) pd), they still feel very sensitive and concerned to the angle compared to those with lager deviation size. Moreover, as you mentioned above, our study cohort were much younger; although deviation size is not always a reliable indicator of poor cosmesis, this might be obvious for those young patients.

12. Comment: The issuing of the questionnaires is important – and you make no comment about this. There are no “right or wrong” ways in collecting patient-reported outcome data, but the methods you choose do need to be described and their potential weaknesses acknowledged. The response burden to the individual of answering over 50 questions (for some individuals) needs to be acknowledged. You make no comment regarding the quality of your data – how do you know that individuals were not just ticking the same response for each question due to boredom, rather than a true expression of their health status? One of the ways in which this could be explored is through examining the floor/ceiling effects. I think this could be a useful addition to the manuscript.

Response: As you suggested, we added the examination of the floor/ceiling effects. There data was added into the results part and the table 6.

“A floor effect (20% of scores at the minimum score) was only found in the ‘social contact & appearance’ subscale. In contrast, all subscales except only the subscale ‘social contact & appearance’ showed a ceiling effect (20% at the maximum score), with a percentage of patients with the highest scores between 24.1% for the ‘psychosocial’ subscale and 62.7% for the “distance estimation” subscale (Table 6).”
Referee 2

Main Comments:

1. Comment1 (i) Methodology: Much greater detail about methodology is required. Are the subjects in the study patients or participants? Order of completion of the questionnaires? Time taken? Setting? [i.e. were they completed at home or there in the hospital]. Were some of the strabismus participants in the hospital owing to some other ocular condition or was it always for treatment/management of their strabismus? The authors needs to discuss how the results obtained in this hospital sample differs from other investigations in strabismics, and how the way in which participants were recruited (e.g. hospital sample versus recruitment in the community) might explain/account their results.

Response: As suggested, more details about methodology has been added.

1) “Subjects” has been changed to “Participants”.
2) For the purposes of standardization, the AS-20 was given first and then followed by the ASQE for each participant.
3) The questionnaires were completed on the day of being added to the waiting list for a strabismus surgery and in a reception room in the hospital.
4) The strabismus participants were all there for a strabismus surgery.

2. Comment1 (ii) Amblyopia does not seem to have been considered as a factor in explaining the results: is this correct? For example, it does not appear in Table 3. This seems like an obvious omission because the presence/absence of amblyopia might have predictable consequences; for example, amblyopia could conceivably add to functional difficulties or it might heighten fears associated with losing the better eye. Can the authors add data relating to amblyopia and make an amblyopia-present versus amblyopia-absent comparison in the same way that they have done for other factors (e.g. exotropia versus esotropia)

Response: We have added the data about amblyopia into table 2 and added content regarding a comparison between amblyopia-present versus amblyopia-absent.

“Compared to strabismus patients without amblyopia, significantly lower HRQoL score was showed among those with amblyopia in total AS-20 ($p = 0.012$) and the subscale of ‘psychosocial’ ($p = 0.011$). Same trend were found in overall ASQE ($p = 0.005$), and the
subscales of ‘fear of losing the better eye’ \((p = 0.026)\), ‘diplopia’ \((p = 0.009)\) and ‘social contact and appearance’ \((p = 0.012)\).”

3. **Comment**: (iii) The results section would be much clearer and more structured if it had subheadings. For example, one sub-heading might be “Comparison with VFQ-25 scores”, another might be “Diplopia versus no Diplopia”. This would I think make a substantial difference to the overall readability of the paper.

**Response**: Subheadings of “Demographic characteristics” “Clinical features and HRQoL” “Proportions of subnormal scores” “Comparison with control groups” “Correlation with NEI-VFQ-25 scores” and “Internal Consistency Reliability” have been added.

4. **Comment**: (iv) The discussion seems unnecessarily long because it restates many of the results. For example the opening sentences of the discussion as well as the 4th paragraph (“To the best of our knowledge..”) seem merely to restate much of what was said earlier in ‘results’. Can results duplication in the discussion be addressed? This will allow more space for a more in-depth comparison/integration with existing literature.

**Response**: This is a good advice. We deleted the content which unnecessarily repeat the results, and added more in-depth discussion about diplopia, amblyopia, Rasch analysis (from the reviewer 1’s comment), etc. As you can see in the discussion part, we have made major revision.

**Other Comments:**

1. **Comment**: (i) Abstract: 1st sentence of abstract needs to be altered …“.. negative impact of strabismus on visual dysfunction, self image disorders, low self esteem…” This would make more sense as “The impact of strabismus on visual function, self-image, self esteem…”

**Response**: This sentence has been re-worded to “The impact of strabismus on visual function, self-image, self-esteem, and social interactions might decrease health-related quality of life (HRQoL).”

2. **Comment**: (ii) Don’t know what you mean by “clinical applications” (2nd sentence abstract). Do you mean clinical utility?
Response: Yes. As you said, “clinical applications” means “clinical utility”, applying the questionnaires in clinical settings.

3. Comment: (iii) Is “internal consistency reliability” a recognised term in questionnaire research description or one you have created? It seems at best awkward…

Response: Like “test-retest reliability”, “internal consistency” is another index of reliability. Internal consistency reliability is usually measured with Cronbach's alpha coefficient, and the term of “internal consistency reliability” has been widely used in many academic articles [1-3].

4. Comment: (iv) Abstract, results: “.. have been showed..” [shown]

Response: It has been corrected.

5. Comment: (v) Abstract, results: I don’t understand “(r=0.204 – 0.476, P<0.05, P<0.01)”. Does this mean at the low end, it's P<0.05 but for the higher value it's P<0.01? This needs to be made much clearer here and throughout the manuscript where multiple r values and p values are reported together.

Response: We meant that the r values range from 0.204 to 0.476 (table 3), some of the significance was P<0.05 while others was P<0.01. Since the details were clear in the table 3, here we just put very brief information. This format might lead to misunderstanding, so we added ‘or’ between. As shown below: “(r = 0.21 ~ 0.34, P < 0.05 or P < 0.01)”.

6. Comment: (vi) Abstract, results: Why say “Over 60% of strabismus patients..” [are they patients or participants; why were they attending the hospital?] when the two numbers that follow are both over 70%?! Say “Over 70% of strabismus patients..” instead?

Response: The 60% is a percentage, while the two numbers that follow ‘71.31 & 75.50’ were threshold scores (table 4). We have changed a little bit, hoping to make it more clearer.

“More than 60% of strabismus patients scored below normal in overall scores (threshold scores: 71.31 for AS-20 and 75.50 for ASQE).”

7. Comment: (vii) Abstract conclusion: I don’t understand the basis for the final sentence [“However,…]. To me, the distinction being made between the two questionnaires is difficult to follow. The conclusion at the end of the paper is easier to follow. Can the authors please look again at the validity, clarity & consistency of their conclusions in the abstract and at the end of the discussion?
Response: As suggested, we added the consistency reliability in the abstract and in the discussion.

8. Comment1 (viii) Figure 1: title for Figure 1 says “median” but y-axis for each bar plot says “mean”?

Response: Thanks for pointing it out. We actually wanted to say “mean score” instead of “median score”. The title for the figure has been amended.

9. Comment1 (ix) Since this journal has a general readership, Cronbach’s alpha should be explicitly explained. Is a threshold value of >0.70 for Cronbach’s alpha standard practice? If so, please support this with citation. If not, please elaborate so that the reader understands if/how the results pattern might change if a lower or higher threshold was employed.

Response: We have explained the Cronbach’s alpha and threshold values in more details and also supported with citation.

“Additionally, the internal consistency reliability indicating the degree to which a set of items measure the same construct (inter-correlations among items) was assessed by Cronbach’s $\alpha$ coefficient. George and Mallery [4] defined a Cronbach’s $\alpha$ value $\geq .90$ as excellent internal consistency; value $\geq .80$ as good and $\geq .70$ as acceptable, while a value $< .70$ indicates questionable internal consistency, poor ($< .60$) and unacceptable ($< .50$) (p. 231). When $\alpha$ value $< .70$ was obtained, one might doubt whether all items measure the same construct. If $\alpha$ value increases when one item is removed, then this item might be considered to be amended or deleted. Otherwise, if removing one item results in no increase in $\alpha$ value, implying that this item correlates well with other items to test the same construct [1].”

10. Comment: (x) There are other terms relating to questionnaire research that are used but not explained (e.g. 2nd paragraph in Introduction: “item pool development”; research objects”; methods section: “adaptation procedures”)

Response: Probably because we are not English-native speaker and due to the language difference, we failed to use these terms properly. Here in the revised manuscript, we tried to find some other expressions. “a standard translation and adaptation procedures” was changed to “standard processes for translation and adaptation”[5, 6]. “item pool development” was changed to “the development of items”, and “research objects” to “subjects”.
11. **Comment**: (xi) Introduction, 2nd paragraph: “both measures” is used repeatedly. I assume this refers to the ASQE and AS-20 scores? Overall scores..? “Median scores on both measures” needs to be reworded since ‘scores’ and ‘measures’ are confusing when used together. Could just replace ’measures’ with ‘questionnaires’?

**Response**: The word “measures” has been changed to “questionnaires”.

12. **Comment**: (xii) Sometimes A&SQ is used but elsewhere its ASQE.

**Response**: As illustrated in Introduction- 3rd paragraph, the Amblyopia and Strabismus Questionnaire (A&SQ) [7] was originally designed in Dutch, and subsequently translated into English (the English-language version of the Amblyopia and Strabismus Questionnaire, ASQE) [2]. And we translated and validated the Chinese version from the English-version ASQE. Here in this manuscript, we mainly used the term ASQE unless sometimes we compare the date of A&SQ reliability and validity with the Chinese version (as you can see in Discussion- 3rd paragraph).

13. **Comment**: (xiii) Methods: minimum strabismus angle was 15 prism dioptres. Was this measured at distance or near? Were any cases of intermittent strabismus included in the sample?

**Response**: The strabismus angle was measured by prism at distance (6 m) and at near (33 cm) fixation. One of our inclusion criteria is strabismus patients with deviation angle at distance was no less than 15pd. That’s why you saw the minimum strabismus angle was 15pd. Patients with intermittent exotropia were included in this study.

14. **Comment**: (xiv) Results are sometimes given to one decimal place (d.p.) and sometimes to two (e.g. 45.07%). Suggest only one d.p. is needed.

**Response**: Thanks for this good advice. We have looked through all the numbers again. And appropriate adjustment has been made. The mean and percentage has been revised as suggested that only one d.p. is adopted. But we kept two d.p. somewhere since we think it might be better (e.g. Methods- with Cronbach’s $\alpha$ ranging from 0.73 to 0.87).

15. **Comment**: (xv) Results: end of first paragraph in results: …”…yet not in ASQE”.

*Please quote all $p$-values whether they or not they are less than 0.05. For example, if $p$ was =0.06 for ASQE this would be of interest. Thus as a general principle throughout the manuscript, please provide actual $p$ values, irrespective of whether or not they are <0.05.*

**Response**: Revision has been made as suggested. For the $p$ value, we adopted three digits instead of two, since some of the values were < 0.01 (e.g., $p=0.002$).
“The group with exotropia recorded better HRQoL scores than patients with esotropia in the overall AS-20 (p = .003) and the two subscales of psychosocial (p = .002) and function (p = .035), yet not in the ASQE. Except for the subscales of ‘psychosocial’ (p = .390) and ‘fear of loss’ (p = .146), strabismus patients with diplopia than those without diplopia reported significantly lower HRQoL scores in overall AS-20 (p = .001) and ASQE (p < .001) and all other subscales (all p < .001 except for the ‘cosmetic’ subscale, p = .021).”

16. Comment: (xvi) Results, 2nd paragraph, 1st sentence: esotropic rather than esotropia.
Response: “esotropia” has been changed to “esotropic”.

17. Comment: (xvii) Results, 3rd paragraph: “No significant differences were found in the proportion of subnormal composite scores....” Do you mean ‘proportion’ or ‘number’?
Response: In Collins dictionary, the “proportion” is defined as “the number of people of that kind compared to the total number of people in the group”. Here in this manuscript, we did use the exact “number” of patients who scored below threshold value to conduct the $\chi^2$ test (AS-20: 203/ 101 v.s. ASQE: 205/ 99, the number of patients scored below threshold value/ the number scored above). But in other words, this is also a comparison of “proportion”.

18. Comment: (xviii) Discussion: 1st sentence.. “...in the proportion of subnormal patients..”. I don’t understand what ‘subnormal patients’ means. Again there appears to be confusion between ‘proportion’ and ‘number’ [e.g. ‘number’ appears in the next sentence].
Response: ‘subnormal patients’ means patients with below normal scores or patients scored below normal. We have changed the expression of “subnormal patients”.

19. Comment: (xix) Discussion, 3rd paragraph: “Therefore one possible explanation for the relatively low internal consistency reliability...” This seems to contradict the statement at the beginning of the paragraph where it is stated that both instruments were satisfactory in this respect.
Response: There might be some misunderstanding due to improper language expression. When we said “Therefore one possible explanation for the relatively low internal consistency reliability”, we actually intended to express “the relatively low internal consistency reliability of the domain of ‘double vision’”. We have reorganized this part as follows.
“In the previous studies, the Cronbach’s $\alpha$ of AS-20 and its subscales were reported ranging from 0.94 to 0.95 by Hatt and colleagues [20]; …In this study, the internal consistency reliability of both instruments were satisfactory except the domain “double vision” (Cronbach’s $\alpha=0.68$). … Therefore one possible explanation for the relatively low Cronbach’s $\alpha$ of the domain “double vision” is that the proposed function descriptions found in these three items were not exclusively features of diplopia”.

20. **Comment**: (xx) Table 2: ASQE: 26 questions but the total in the “Domains or subscales (items)” for the ASQE is only 23! Please explain/amend.

**Response**: Thanks for your careful findings! We did ask the same question when we read the article of ASQE by Felius et al. [2]. We also put Dr. Felius Joost’s email response regarding this issue below. In their article, as you can see in the table 1, they stated that “the items 1 (I can see equally well with two eyes), 17 (I see double), and 24 (I have a squint) are filters for the contingency items that follow”. Also, in their table 2 “Subscales of ASQE”, Felius they did not count the item 1, 17, and 24 in. The total number showing in the table 2 is 23.

Dear Wang,

Thank you for your interest in the ASQE questionnaire.

Those three items (1, 17, and 24) “filter out” those subjects for whom the items that follow are not relevant:
- If someone chooses the response option “yes” to item 1, then items 2 and 3 become irrelevant (meaningless; not applicable.)
- If someone chooses the response option “none of the time” to item 17, then item 18 becomes irrelevant.
- If someone chooses the response option “none of the time” to item 24, then items 25 and 26 become irrelevant.

So, if you plan on analyzing ASQE data, then you may want to ignore items 25 and 26. *F* the value of the response option on item 24 is “none of the time”. (And similar for the other two filter items.)

Does this answer your question?

Joost

21. **Comment**: (xxi) Table 3: in the version I received, the symbol “>” does not appear for large angle strabismus (instead of a ‘>’, a square symbol appears)

**Response**: I think the square symbol was a format error. We have looked back to our original manuscript, the symbol is absolutely “>”.

22. **Comment**: (xxii) Table 4: what is the significance of italicised test in this table?

**Response**: Just like the symbol “†, ††”, the italic stands for: $\chi^2$ tests, the proportions of adults scored below normal threshold between with and without diplopia. We originally intended to highlight the significant results by using the italic. But since we have already utilized the symbol “†, ††”, we removed the italic format to avoid misunderstanding.
23. **Comment:** (xxiii) Figure 1 legend: no need to add that “* means p<0.05” since all cases of P<0.05 are also P<0.01 so only ‘**’ appears not ‘*’.

**Response:** “* means p<0.05” has been removed.

24. **Comment:** (xxiv) There are various other typographical errors and style issues (e.g. need for parentheses, use of hyphenation) throughout the manuscript but these can be rectified at a later stage.

**Response:** We have tried our best to correct the issues mentioned above. But we are also happy if more issues are required to be corrected.

---