Author’s response to reviews

Title: Item response models for the longitudinal analysis of health-related quality of life in cancer clinical trials

Authors:

Antoine Barbieri (antoine.barbieri@gmail.com; antoine.barbieri@umontpellier.fr)
Jean Peyhardi (Jean.Peyhardi@umontpellier.fr)
Thierry Conroy (t.conroy@nancy.unicancer.fr)
Sophie Gourgou (Sophie.Gourgou@icm.unicancer.fr)
Christian Lavergne (Christian.Lavergne@umontpellier.fr)
Caroline Mollevi (Caroline.Mollevi@montpellier.unicancer.fr)

Version: 1 Date: 14 Jul 2017

Author’s response to reviews:

Dear Editor, Dear reviewers,

I am submitting the revised manuscript “Item response models for the longitudinal analysis of health-related quality of life in cancer clinical trials” for publication in BMC Medical Research Methodology. We are pleased to answer the reviewers’ comments, which were relevant and which pointed out interesting questions.

We answered all of comments as thoroughly and as clearly as possible, and the manuscript is now improved. Mainly, we have reviewed and corrected the largely stylistic and conceptual imprecisions highlighted by the reviewers. As expected, we indicated all the changes in the text by highlighting using LaTeX tools. Moreover, an English language revision has been performed by a native English speaker and helped us to improve the English language within our manuscript. Please see below for our answers to each comment.

We hope we have met your requirements to improve this paper, and that it will now be found suitable for publication.
Thank you for your interest in our work.

Best regards,

Antoine Barbieri

RESPONSES TO REVIEWERS COMMENTS

Tianmeng Lyu (Reviewer 1): This paper reviewed three families of IRT models and made recommendations about model selection after some conceptual discussions. It also included simulation studies to compare IRT models and classically used LMM models in terms of the ability to detect the random part of the mixed models. Please see the following comments.

1. Among the simulation scenarios presented in Table 5, \( \beta_1 = 1, 0.3 \) or -0.3, but more simulation scenarios could be added, such as \( \beta_1 = -1 \) or any value between 1 and 0.3 for a more complete comparison if the authors wanted to make conclusions about how the results changed when \( \beta_1 \) was closer to 0 as they did in line 35, page 11.

We added simulation scenarios with the following values: \( \beta_1 = 0.5, \beta_1 = 0, \beta_1 = -0.5 \) and \( \beta_1 = -1 \). These additional simulations allowed then to make conclusions about how the results changed when \( \beta_1 \) was closer to 0.

Besides, in Table 5, the numbers of \( \sigma_1^2 \) values and the values themselves were different for different \( \beta_1 \) values and this is not common when we design simulation studies. The authors should explain the reason why doing so.

We agree with the reviewer, this is not common to use different values of parameters in the design of simulation studies. However, it was not necessary to keep the same \( \sigma_1^2 \) values for different \( \beta_1 \) values in our simulation study. For example in Table 5, for \( \beta_1 = 1 \) and \( \sigma_1^2 = 0.01 \), the percentage of selection was already very low and the scenario \( \beta_1 = 1 \) and \( \sigma_1^2 = 0.005 \) would not bring more. Moreover, to explain this, we added the following sentence at the beginning of the paragraph regarding Table 5:

“Only the values of \( \sigma_1^2 \) for which the capacities varied between the three models are presented for each considered value of \( \beta_1 \).”
2. It may be helpful for better understanding if the last paragraph of the simulation results section could be reorganized and elaborated. For example, "the LMM detected the random slope for a greater variance of this one whatever the $\beta_1$ value" is confusing.

- According to the recommendations of the reviewer, we modified this paragraph of the simulation results. The changes appear in blue in the new version of the manuscript.

3. The simulation results section only listed the observed results without discussing the practical importance of the simulation results, for example, the authors may discuss the suggestions for data analysis that can be learned from the simulation studies. Adding such discussion may be helpful to make the contribution of this part clearer.

We added a conclusion paragraph at the end of the simulation results section to make the contribution of this part clearer:

“In conclusion, the closer the value of $\beta_1$ to zero (small signal), the easier it is for the models to detect the random slope with a low variance. The IRT models are more sensitive and stable than the LMM whatever the parameter values. This result was expected because the LMM is based on the HRQoL score, which is a summary variable with less information than the raw data. Comparing the IRT models, the one which was not used to generate the data tended to wrongly detect a random effect where there was none.”

4. Overall the paper is clearly organized. But the logic doesn't flow well in some parts of the paper. For example, in the last paragraph of page 2, it seems that the sentence "The use of the LMM for …." should closely follow the sentence which explains why LMM is not appropriate for HRQoL analysis instead of following the sentence "These models allow to take into…" which aims to present the advantages of the LMM models.

We modified and shortened the background following Sunita Ghosh’s recommendations (reviewer 2), and then we took into account this remark 4 in the new introduction.

5. In line 32 page 8, the first "r" stands for “ratio of probabilities” while the second "r" stands for "the number of random parts”. Usually we don't use the same letter to represent two different things.

Reviewer 1 is right. We have corrected this point accordingly. We replaced the index “r” (number of random effects) with the character “a”.
6. Language should be polished and grammar could be improved. For example, line 7 page 3, "... LMM is not appropriated" should be "... LMM is not appropriate"; line 54 page 3, "in the result section" should be deleted; line 36 page 6, "will be consider" should be "will be considered"; etc. And the authors should avoid using too many "indeed".

We corrected these points and an English fluent person provided editorial assistance to improve grammar and language.

Sunita Ghosh (Reviewer 2): The paper titled "Item response model for the longitudinal analysis of health-related quality of life in cancer clinical trials" is very interesting to read and an important topic indeed specially for the Oncology trials. However, having said that the introduction or the background of the paper was very lengthy. My recommendation would be make the background short. The method section describes the three structure of the model used in the paper, I was a bit confused with the flow of methods as described. Since this is a methodology paper, the simulation codes used in the paper should be accessible to the readers to reproduce such kind of simulation data.

According to the recommendations of the second reviewer, we shortened the background. For example, we removed the paragraph talking about time to event models for the longitudinal analysis of the HRQoL because we did not use this approach in our manuscript. Moreover, we also shortened the part regarding the questionnaires.

Concerning the flow of methods as described, we tried to clarify through adding and/or modifying some sentences. I would refer you to the reviewer 3’s comments. For example, we moved the paragraph concerning the expressions of each response probabilities given the linear predictor and the cumulative distribution function after the ratio presentation in “The probability ratio: structure of the models” subsection rather than in “The cumulative distribution function” subsection.

Concerning the availability of code, we added the codes using the SAS procedure PROC NLMIXED to estimate the IRT models in additional file.

Sangchoon Jeon (Reviewer 3):

#1. On Page 7, the equation (5) may show the probabilities based on reversed categorical order. Based on the function F defined on page 5 (i.e. r_m(pi)=F(eta_m)), the probabilities would be pi_0=F(eta_0), pi_m=F(eta_m)-F(eta_m-1), pi_M=1-F(eta_M-1). In the reversed categorical
order, that is, \( r_{m(pi)} = 1 - F(\eta_{m+1}) \), the probabilities in eq (3) might be correct. If it's correct, it needs to be stated that the eq(3) was calculated for reversed order because the CDF for the sequential model seems not being calculated for reversed order.

We understand the incomprehension of the reviewer 3. To clarify this, we added the ratios of probabilities in descending order for cumulative and adjacent models in new equations (5) and (6), respectively. That completes the ratios of probabilities given in equations (3) and (4) associated with the ascending order. Then, we moved the paragraph concerning the expressions of each response probabilities given the linear predictor and the cumulative distribution function after the ratio presentation in “The probability ratio: structure of the models” subsection rather than in “The cumulative distribution function” subsection.

#2. On page 8, in the four components \((r, F, Z_q, U_r)\), the index \(r\), which is the number of random effect, is confused with the ratio of probabilities \((r)\). I recommend to replace the index \(r\) with another character.

Reviewer 3 is right as well as the Reviewer 1. We have corrected this point accordingly. We replaced the index “r” (number of random effects) with the character “a”.

#3. Title of Table 4 and 5 state "Frequency (on N=500 datasets) of …". Seems Table 4 and 5 are presenting percentages of selecting M1 or M2 rather than frequency of the selections.

Reviewer 3 is right. We changed the caption of the Tables replacing “Frequency (on N = 500 datasets) of the M1 (or M2) selection according to the BIC …” with “Percentages of selecting M1 (or M2) according to the BIC on N=500 datasets …”.

#4. On page 10, the authors state "This could be explained by the fact that the difficulty parameters were not uniformly separated around zero and also because they were too close." This statement does not explain why the other IRT model had good performance even compared to LMM. Additional explanations are recommended.

The other IRT model is the one used to generate the datasets, then it is naturally more close to the data than LMM. It had good performance because this is the most appropriate model given the data are derived from it. To improve the understanding, we replaced the following sentence:

“On the contrary under M1, the simulated model M1 was correctly chosen in most cases.”

by
“As expected under M1, the simulated model M1 was correctly chosen in most cases, and in particular for the IRT model used to generate the datasets.”

Following the reviewer 2’s recommendation, we added a short conclusion in this subsection, in which we added that LMM is based on score which is a summary variable and not on the raw data. Moreover, we replaced the sentence cited by the reviewer with:

“This is caused by the relationship between the latent variable $\theta$ which changes over time and $\delta^{ne}$ which accounts for the observed ordinal responses over time. For these specific parameter values ($\beta_1 \approx -0.3$ and $\delta^{ne}$ given in Table 3, the linear predictors $\eta_{itm}^{(j)}=(j)=\theta_{it}-\delta_{jm}^{ne}$ were close between them for $m=2,3$ whatever $j=1,2$. These linear predictors being negative and different from zero value, the probability of selecting the upper categories was very small over time and under-represented in comparison to the lower categories.”

#5. On page 12, the authors stated "it referred to a difference between the two arms at day 15, day 30 (during treatment period) but no necessarily at baseline" for the significant group difference (i.e. $|\beta_1|>0$) on diarrhea at baseline. I am confused how group difference during treatment period can be referred from the significant beta1. Maybe more detail explanations about the data are necessary.

To clarify this part, we modified the following sentence:

“It referred to a difference between the two arms at day 15, day 30 (during treatment period) but no necessarily at baseline. Then, the perception of diarrhea symptom remained higher in the Folfirinox arm over time. This result is expected because the Folfirinox is more toxic than Gemcitabine and is known to cause more diarrhea symptom.”

by

“This is caused by a difference between the two arms of the study during the treatment period (at day 15 and day 30). This result was expected because Folfirinox is more toxic than Gemcitabine, and is also known to cause more diarrhea symptoms. Given our model does not take into account a possible difference between the two treatments during only this period, the fixed intercept was affected. The perception of diarrhea symptoms remained higher in the Folfirinox arm over time, particularly during the treatment period.”

#6. Based on simulation study, the IRT models have greater powers to detect random slopes but it's compensated long with greater type I errors compared to LMM. More guidelines in application of IRT are recommended.
Thank you for this remark. We did not particularly consider the type I errors in this manuscript, focusing only on random effects. However, a previous simulation study to compare these models in terms of type I errors and statistical powers were performed on the fixed effects in the Anota’s paper (2015) that we cited in our manuscript.

EDITOR COMMENTS:

In addition to the revisions requested by the reviewers we would be grateful if you could make the following editorial revisions:

1. Please add a conclusions section after the discussion which should state clearly the main conclusions and provide an explanation of the importance and relevance of the study reported.

As requested by the editor, we added the main conclusions in a conclusion section after the discussion.

2. In the ethics approval and consent to participate section please provide details of how the data-set was accessed and who provided permission for this access. Please also provide details of why ethics approval was not required for the secondary use of these data.

As requested by the editor, we provided details in the ethics approval and consent to participate section. We wrote:

“UNICANCER R&D, the sponsor of the PRODIGE4 / ACCORD11 trial (ClinicalTrials.gov Identifier: NCT00112658), provided permission for the data base access. All participants provided written informed consent for the PRODIGE4/ ACCORD11 trial. Patient consent was not required for this study as we performed a secondary analysis of existing data.”

3. In the authors' contributions section of the declarations please state the contribution of each individual author who should be referred to by their initials. Please retain the statement to confirm that all the authors read and approved the final version of the manuscript.

We detailed the authors' contributions section such as:

“AB performed the comparison studies, the statistical analyses, interpretation and wrote the manuscript. JP designed and drafted the method section. CL and CM supervised this work. TC was the principal investigator of the clinical study (NCT00112658) and participated to the
patient’s inclusion. SG and TC interpreted the clinical results. All authors read and approved the final manuscript.”

4. Please combine the panels for figures 1 and 2 into single composite files that contain all parts of the figures. Please see our submission guidelines for more details of how to format figures.

As requested by the editor, we combined the panels for figures 1 and 2 into single composite files.