Reviewer's report

**Title:** A Multiple Imputation Method based on Weighted Quantile Regression Models for Longitudinal Censored Biomarker data with Missing Values at Early Visits

**Version:** 0  **Date:** 04 Aug 2017

**Reviewer:** Guangyu Zhang

**Reviewer's report:**

Review of "A Multiple Imputation Method based on Weighted Quantile Regression Models for Longitudinal Censored Biomarker data with Missing Values at Early Visits"

The paper focuses on a multiple imputation procedure based on weighted censored quantile regression that accounts for censoring and the missing data at early visits. The basic procedures include: (1) modeling the missing data process; (2) fit weighted censored quantile regression; (3) impute missing and censored data. The authors conducted simulation studies and applied their method to the Prospective Study of Outcomes in Ankylosing spondylitis data. I have the following questions about this paper.

1. On page 5, step (1), the authors mentioned "we predicted the probability of missing CRP at early visits conditional on the observed CRP data at the rest of follow up visits (if censored, imputed by half of detection limit (DL), i.e., DL/2)". If a person has missing CRP at the first and the second time points, the probability of missing at time 1 would depend on CRP at time 3 and forward. And if a person has missing CRP at the first time point only, the probability of missing at time 1 would depend on CRP at time 2 and forward. Am I right?

2. On page 5, step (3), for the censored data, the authors generated v from a uniform distribution between 0 and ω, where ω was the estimated conditional probability of censoring, and imputed the censored value by v-th conditional quantile. Is it possible that the imputed values were greater than the detection limit? Have the authors checked that all the imputed values for the censored data were less than the detection limit?

3. The imputation were based on the parameter estimates from step (2), which used the estimated probability weights from step (1). And in step (1), the censored CRP data was naively imputed as the half of detection limit. I think a more proper imputation procedure is to iterate steps 1-3. After impute censored CRP level, used the imputed CRP
to re-estimate the probability weights at step (1), refit the regression model at step (2) and re-impute missing and censored CRP levels at step (3). Repeat this procedure until converge.

(4) One page 6, line 42, regression model (1), the error term was missing.

(5) Page 6, last line, "We simulated data for 75% of patients who had missing data for up to first 3 visits", what were the percentages of subjects missing one time point, missing the first two time points?

(6) Page 7, lines 24-26, it is not clear to me how the MCMC imputation was conducted.

(7) Page 7, lines 40-47, is this paragraph redundant?

(8) Page 10, Figure 2, why were the missing and censored data in the middle? I think the censored data should be on the left side (i.e., censored when less than the detection limit)?

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.

Yes
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/](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