Reviewer's report

Title: Using an onset-anchored Bayesian hierarchical model to improve predictions for amyotrophic lateral sclerosis disease progression

Version: 0 Date: 21 Oct 2017

Reviewer: Brian Smith

Reviewer's report:

The authors present an empirical study to develop and compare the performances of different Bayesian hierarchical linear regression models for the prediction of disease progression in Amyotrophic Lateral Sclerosis (ALS) subjects. Subject data come from an open-access ALS clinical trials database. Disease progression is measured by the ALS Functional Rating Scale (ALSFRS) and is predicted 9 or 12 months into the future. Of the 4838 subjects in the ALS database, 1300 were considered for the analysis based on having received placebo and having at least one available ALSFRS score. Three classes of Bayesian hierarchical models were assessed: (1) linear, (2) linear mixture, and (3) onset-anchored. Primary comparisons of model performance were based on cross-validated estimates of mean squared error (MSE). The onset-anchored model was found to perform best. Addition of covariates to the onset-anchored model was also considered but not found to improve performance substantially. Takeaways from the results are that the noted simplistic approach of adding a leverage point in the onset-anchored method outperforms the other two commonly used regression methods; prediction can be performed after a single baseline measure of ALSFRS; the better prediction may lead to more efficient clinical trials design; and the approach may be of benefit in other neuromuscular disorders.

Comments on the manuscript are as follows.

Model Comparison Section

1) p.8, line 39: The choice of T3 distribution is justified with a Q-Q plot of residuals from a simple linear regression analysis. Posterior predictive p-values, based on appropriate test statistics, would be a standard approach for assessing distributional assumptions in a Bayesian analysis. The authors are encouraged to add such a posterior predictive check to their Q-Q plot analysis.

2) p.8, lines 50-52: The b_i parameter is specified in the formula here as having a normal prior distribution, but said in the text to be restricted to be non-positive. The BUGS pseudocode in Appendix A1 suggests that the prior is a truncated normal. The formulas for parameters with truncated distributional specifications should include the truncations
and not just the standard, untruncated distributional part of the specification. There may be affected formulas in other portions of the manuscript.

3) p.9, line 24: The priors for the remaining variables are said to be uninformative, but the hyperparameter specifications are not given. Explicitly state the specifications or indicate where they can be found (e.g. in an appendix). Note that a bounded uniform prior that restricts the range of a parameter to a subset of its support would more appropriately be called "vague" or "weakly informative" than "uninformative". These comments apply to other sections of the manuscript in which prior hyperparameters are not explicitly specified and/or are said to be uninformative.

4) p.11, lines 45-49: Linear regression was used to turn repeatedly measured lab features into (slope and intercept) covariates for the ALSFRS analysis. For the purpose of estimating predictive performance of the ALSFRS models, the linear regression step would ideally be performed within the cross-validation procedure. Indicate whether that was done and the implication for interpreting the results if not.

5) p.12, line 7: Expanding the descriptor used here to "100 replicates of 10-fold cross-validation" might make the approach more recognizable to some readers and also make the number of replicates clear at this point.

Discussion Section

6) p.19, lines 15-22: Clarify the context in which using an artificially created data-point as observed data is believed to be novel. In general, last value carried forward might fit that description, although it is applied in different settings (and typically not advisable).

Figure Legends

7) p.27, line 49: Clarify whether "single cross-validation analysis" refers to one replicate of the analysis, one fold, or something else.

Minor Edits

8) p.6, line 47: "data is" -> "data are"

9) p.9, line 21: "chosen reflect" -> "chosen to reflect"
10) p.14, line 32: "due disease" -> "due to disease"

11) p.15, Table 2: Indicate in the table caption that the presented results are for the onset-anchored model. Define IQR; e.g. "The inter-quartile range (IQR) ...".

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

Yes

**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/). 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