Author’s response to reviews

Title: A semi-parametric mixed models for longitudinally measured fasting blood sugar level of adult diabetic patients

Authors:
Tafere Tilahun Aniley (tafere.tilahun@ju.edu.et;tafere27@yahoo.com)
Legesse kassa Debusho (debuslk@unisa.ac.za)
Zelalem Mehari Nigusie (zelalem2011@gmail.com)
Wondwossen Kassahun Yimer (wyimer@umc.edu)
Belay Birlie Yimer (belaya.birlie@gmail.com)

Version: 1 Date: 05 Oct 2018

Author’s response to reviews:

Response to the of referees’ comments on the manuscript:

"A semi-parametric mixed models for longitudinally measured fasting blood sugar level of adult diabetic patients" by Tafere Tilahun Aniley, Legesse kassa Debuaho, Zelalem Mehari Nigusie, Wondwossen Kassahun Yimer, Belay Birlie Yimer

Manuscript ID: BMRM-D-18-00226

Dear Editor,

We are very grateful to the reviewers for the thorough review of our manuscript and in particular for the insightful and detailed comments. The changes made to the original manuscript are indicated in red color in the revised version. We hope we have addressed all the points that the referees have made and provide the following summary of our changes.

Reviewer 1 (Qing Gang Pan) Comments
1. What is the optimization criterion in estimating the penalized splines? How are the random coefficients for the splines estimated? Please give explicit formula instead of the names of the package (gamm/glimmix) which includes tons of options and we don't know exactly what you did.

Response:

In our application, estimation of the coefficients of penalized and unpenalized terms was done using a penalized iteratively reweighted least squares (P-IRLS) scheme while estimation of the smoothing parameter is integrated through outer iteration using REML. This information is added to the revised version of the manuscript on page 6 toward the end of Section 2.4.

2. Page 7, "Assuming the vector of contrasts have approximately multivariate distributions with mean vector 0 and variance-covariance matrix (C'R-1C+B)-1" Do you have proof for the multivariate normal distribution?

Response:

The asymptotic normality of the regression coefficient $\beta$ was shown by Ruppert et al (2003) and Wood (2006) and we cite these authors in the revised version of the manuscript on page 8, before equation (5). With the aim of not distracting the readers we skipped the detail.

3. Model Selection: The -$2\log L$ of M4 cannot be found in Table 4. More importantly, the decrease in -$2\log L$ from 50542.51 in M3 to 50538.54 in M5 is trivial given the sample size of 534. The choice of penalized spline model is not justifiable. Especially when parameter estimates are similar, it is natural to choose the simpler model. Not sure why when the spline coefficients are all zero or variance of $b_1$ is zero, the splines reduce to the quadratic term with random coefficients. Also the mixed chi-square distribution for $\sigma^2_b$ square is based on the normal assumption of $\sigma^2_b/\text{estimate}$, which is not proved.

Response:

The issues mentioned in this item or comment helped us to make most of the changes in the revised version of the manuscript. Our responses as follows:
- The parametric estimates of model M4 are provided in Table 4 on page 12 of the revised version and the value of \(-2\log L\) and other fit statistics for this model are included in Table 6 on page 13.

- Reason why we use the penalized spline model in this study is discussed in the first paragraph of Section 3.4 on page 12 on the revised version. Additional justification supported by statistical result is also discussed in Subsection 3.4.1 on page 13. The test on \(\text{var}(\sigma^2_b) = 0\) and its implication is discussed on page 14.

- The normality assumption for random effects was checked using the QQ plot (but results were not reported in the original manuscript for the sake of brevity). In the revised version of the manuscript we have used two other alternative methods, Crainiceanu and Ruppert (2004) and Greven et al. (2008), which are recommended in the literatures (we came across these literatures while we are working on the reviewers’ comments). The likelihood ratio test for testing zero random effect variance is based on the assumption of independent and identically distributed (IID) of response data. However, the IID assumption needed for the mixture chi-square distribution may not hold for our data because of the unbalanced nature of the FBS level data. These are discussed in Subsection 2.2.2 on page 5 in the revised version.

4. What kind of improvement do the terms with the random effects (random intercept, random slopes, random quadratic, random spline coefficients, et al) provide in terms of clinical interpretation or prediction? I totally don't see the value of the much more complicated models for clinicians.

Response:

- We have considered the three variance-covariance structures based on the variance profile plot of FBS level. Like the mean profile plot, this plot shows the variance changes overtime. Therefore, we wanted to allow for more flexibility to estimate the between subject variability. This information is now included in the revised version in Section 2.2.1 on page 4.
- Along with the above point, literature (e.g. Verbeke and Molenberghs, 2000) suggests that the evolution of the variance is important to build an appropriate longitudinal model. We provided this reference in Subsection 2.2.2, on page 5, of the revised version.

- Furthermore, the individual (subject-specific) plots in Figure 1(a) show that different patients may have different evolutions over time (some may be constant but different among patients, some may have linear trend or quadratic trend or nonlinear form). Therefore, we assumed the subject-specific profiles to be different only by intercepts; different with subject-specific intercepts as well as slopes; and quadratic over time with subject-specific intercepts as well as slopes for the linear as well as quadratic time effect in order to build a reasonable model for the analysis. The authors think that the response given in the second bullet point for Item (3) addresses the reason why the semi-parametric model considered in this study. We believe such approaches will help to inform a physician or clinician that a clinical intervention package(s) that they give to diabetic patients may results in different responses or reactions among patients and/or these responses may have different trend over time. We have discussed this in the conclusion section of the manuscript.

Reviewer 2 (Abbas Bahrampour) Comments

1. Why did you use FBS (If you had used A1CHg as a criterion, it could be shown the trend, better than FBS)?

Response:

We used FBS instead of A1CHg because the latter is not often measured in the Jimma University Specialized Hospital (JUSH) diabetic clinic.

2. Using the interaction of gender and time in the model doesn't have logical, theoretical and clinical support.

Response:
We test for the possible interaction effect because we suspect males and females might have a different level of adherence to the intervention package. However, the interaction effect is not significant and is not included in our final model.

3. Could you please explain the theoretical support of using semi-parametric linear-mixed model, because the FBS has the normal distribution?

Response:

We are motivated to use the semi-parametric linear mixed model in our application for two reasons. (1) From our EDA, it was evident that the average profile of FBS over time is nonlinear. (2) We are interested in estimating the rate of change of FBS level over time. The rate of change is a useful measure to assess the success of the intervention packages. By using the semiparametric mixed model, we manage to able to estimate the time it takes for a particular individual to reach a stable FBS level.

4. In linear mixed models, it is necessary to determine the structural or non-structural model is used and why?

Response:

In the linear mixed model, we considered a random intercept term because individual have different level FBS level at time zero and we also considered a Radom effect term for the slope and quadratic time effect because individuals have a different rate of change over time. The authors think that the responses in the first two bullet points for Item (4) in the Reviewer 1’s comment also apply to this item too.

5. Interpret the variance components.

Response:

This now included in Section 3.3 and Subsection 3.4.1 using the statistical tests’ results on zero random variance components.
6. All patients are type two diabetes how it can be a variable?

Response:
All patients are not Type II. The data consists of both Type I and II diabetic patients, this is discussed in Section 3.1.

7. As you see the semiparametric linear mixed model is very close to the linear mixed model (the difference of AIC), and using the linear mixed model is easier for clinicians, what do you reckon?

Response:
This comment also shaped our revised results. The comparison that we wanted was between the final linear mixed model M4 and the semi-parametric model M5. However, the results for M4 were not in the original manuscript. The results in the revised version now show that the semi-parametric model has a better AIC value.