Reviewer's report

Title: Bayesian random effects modelling with application to childhood anaemia in Malawi

Version: 3
Date: 24 September 2014

Reviewer: Ezra Gayawan

Reviewer's report:

The article maps residual spatial variation of childhood anaemia in Malawi and the findings could be useful in designing intervention strategies. However, there are a number of drawbacks that need to be addressed before the article can be recommended for publication. One vital issue is that the usage of English language throughout the manuscript was very poor, making it difficult to comprehend some of the claims. I provide a list of other concerns below:

There are a number of recent articles that use Bayesian models to examine anaemia among young children in sub-Saharan Africa that are necessary to be cited in the article. Among them are:


Since the article is not paged, I assumed the abstract page to be page 1.

Background

1. WHO 2008 cited looks outdated. There will be more recent citation.
2. The last statement in the second paragraph "In Malawi of all children ..." does not seem to connect to previous statement.
3. On page 3, the third line, there are more references including those cited above.
4. Pg. 3, the last paragraph, "Spatial mixed models ... to assess the geographical effect" of what??

Methods

5. Despite that the section is long, it is completely silence on the priors assigned to the different functions and parameters. Most of the details contained in the section are not necessary, especially for a journal such as this. I suggest the authors cut down on many of the details therein. For instance, the paragraph under Eqn. (3) might not be necessary, details about probit and log-log functions on pg 6 might not be necessary.
6. On pg. 7, the last sentence in the first paragraph listed "the problems to be addresses" which should not be mentioned here.

7. Pg. 7, Under the sub-heading "Empirical Bayes ...", the authors were meant to refer to equation 10 not 6. Likewise, most of the details under this sub-headings are not necessary.

8. Study area and data: This sub-heading should have come before the model sub-section.

9. One major issue I gathered from here is that the study excludes children below age 12 months. However, results were presented for children aged 6 months and above (see the nonlinear effect of child’s age). The authors need to clarify this issue. If some children were excluded from the analyses, I think there should be more robust justification for that than what is contained here. If it is based on that "... such children are known to experience physiological decrease of haemoglobin level" the source of this information need to be clearly stated.

10. The authors need to state the source of the cut-off point used in classifying the children into the three categories of anaemia.

Analysis

11. The authors used 4 models based of subset of the variables for the analyses. The general form of the models need be given to cover the binary and cumulative logistic instead of repeating almost the same thing over again. We should be told at the point where it was said that child anaemia ~ Bernoulli () what ij represent.

Results

12. In the descriptive results, Table 1 presents the data on region basis but in the comment of the authors, they were making mention of lake shore district which makes the presentation confusing to someone not familiar with Malawi. I think they need to be consistent with what is in the table.

13. In Table 2, there is a variable "Meat" what does it stand for?

14. On pg 15, Anaemia prevalence among male children is almost same with female (Table 2). So, to say that "Males have higher prevalence ..." could be misleading. Also, prevalence of anaemia does not decrease with higher level of education. In fact, from Table 2, it was highest for children whose mothers attained higher education. Again, I could not find breastfeeding in Table 2 as claimed in the last statement.

15. I do not think Table 3 and the explanations for it are necessary for this study.

16. Model selection: what is AIC and GCV? Also, there is no point mentioning BIC since the choice is based on AIC and GCV.

17. Pg 16, last paragraph, there is no point re-mentioning the cut-off point for the cumulative model here as it has already been said.

18. The statement in line 7 from below, "The sign of the covariate ..." is not clear. Is it referring to the alphas or what?

20. On Tables 4 & 5, the authors need not present the coefficient of all models. Presenting only results for the best models for the binary and cumulative models
side-by-side would be enough. This allows the reader to quickly compare the coefficients. It is cumbersome in the present form. The same should apply to other effects.

21. Nonlinear effects: Contrary to what is presented, breastfeeding shows a linear relationship. The world "steadily" used to describe the relationship between breast feeding and anaemia should be removed as the results did not show that.

22. On child age, there are some points where the CI contains 0. I suggest the authors relax discussions on significance for the nonlinear effect, CI containing 0 etc and just focus on the relationship in terms of the curve exhibited by the various effects.

23. Still under child's age, "The chance of having general anaemia ..." what is general anaemia?

24. Results on mother's age need to be re-written.

25. Discussion: The statement "Modelling of metrical ... could not be rejected if they were modelled linearly" is unclear. The statement that follow after has little meaning. It needs to be re-written.

26. Pg 24, second paragraph, the first statement needs to be referenced. In fact, and subsequent claims in the paragraph.

27. Pg. 25 in the third paragraph, it is not clear what antenatal anaemia is.

28. Pg 26 last paragraph, the limitation that the spatial estimate might not be exact as data was from survey seems to overlook the whole essence of sampling. If the survey was properly conducted, for which most DHS surveys are known for, estimates from the surveys are true representative of the population and hence, this limitation is not necessary.

29. The references should be properly checked for correctness. For instance, Kazembe LN and Isak, N (2013) is different from what is in the body of the paper.

Level of interest: An article of importance in its field

Quality of written English: Not suitable for publication unless extensively edited

Statistical review: Yes, and I have assessed the statistics in my report.

Declaration of competing interests:

I declare that I have no competing interests