Reviewer's report

Title: Structured Additive Regression Models to Estimate Under-five Mortality risk factors in Ethiopia

Version: 2 Date: 12 December 2014

Reviewer: Jin Rou New

Reviewer's report:

The authors used a structured additive logistic regression model to estimate under-five mortality risk factors in Ethiopia using data from the Ethiopia Demographic and Health Survey 2011. Their analysis found significant risk factors of child mortality, as well as significant nonlinear relationships between child mortality risk and variables such as age of respondent. They also mapped out predicted child mortality risks by region.

The work the authors have done to highlight significant risk factors of child mortality and map out predicted child mortality risks in Ethiopia have not, to my knowledge, been done. This is important and useful research to guide policy action on reducing child mortality in Ethiopia that could also be extended to other countries.

However, the idea of using STAR models to examine risk factors of child mortality and producing child mortality risk maps is not new, e.g. see “Childhood mortality in sub-Saharan Africa: cross-sectional insight into small-scale geographical inequalities from Census data” by Lawrence Kazembe, Aileen Clarke, Ngianga-Bakwin Kandala (http://bmjopen.bmj.com/content/2/5/e001421.short). The authors can be more specific about how their paper adds to the current state of literature. The paper will also benefit from greater clarity (as elaborated upon in subsequent comments)

- Major Compulsory Revisions

1. Abstract: The abstract would benefit from being more coherent (e.g. The last sentence in Background does not link well to the first sentence in Methods; the flow of the sentences in Methods is problematic.) and specific (e.g. How was the structured additive logistic regression model adapted?). The results section can also be more succinct, e.g. the second sentence is redundant as it is effectively a repeat of the first sentence. The Results part also do not adequately flesh out all the findings of the analysis, e.g. spatial effects alluded to in Methods are not mentioned in Results.

2. Results: Effect of categorical variables on child mortality: I am concerned about the large number of categorical variables used in the model and the possibility of high collinearity among these categorical variables. Have the authors checked the correlation between all these variables and/or considered
combining these variables into one variable if collinearity is high?

3. Results: Are the CIs used 95% CIs? The level is not specified. It is not clear why these CIs refer to confidence intervals and not credible intervals if the model is fitted under a Bayesian framework.

4. Results and Tables 2-4: I might be missing something but it's not entirely clear to me why p-values are given and hypothesis tests conducted, since these parameter estimates come from the Bayesian framework?

5. Statistical Methods: More detailed specification of the model and consistent notation needed:
   a. What do #i and vj correspond to?
   b. Please specify the form of the functions fj(xj).
   c. Does # represent coefficients of categorical fixed variables (as specified in Paragraph 2) or coefficients of continuous covariates (as specified in Paragraph 3)?
   d. Specify more clearly the priors used for the spatial effects.

6. Statistical Methods: Could the authors specify the number of iterations and samples used in the MCMC simulation?

7. Statistical Methods: No mention of model diagnostics was given; were any model checks conducted?

8. Statistical Methods: Please note the software used to conduct this analysis.

9. Discussion, Paragraph 1: “to estimate the changes in risk factors of child mortality in relation to ongoing interventions by the government of Ethiopia” – I do not see any analysis that supports this claim; the model does not take into account governmental interventions, nor does it examine changes in risk factors over time.

10. Discussion, Paragraph 1: “Models fitted without taking into account the spatial structure were found to be less adequate when compared with spatial models.” – Are there references or the authors’ own analysis to back this claim?

11. Discussion, Paragraph 2: Instead of repeating the results, the authors could elaborate on the significance of their results, e.g. What further research or policy actions could be taken based on these results? Are any of these results surprising and would merit further analysis?

12. Discussion, Paragraph 4: “effective control measures of under-five mortality at household level in Ethiopia should start with proper mapping of the risk” – Could authors explain how mapping of the risk at the region level help with implementing control measures at the household level?

13. Discussion: Could the authors elaborate on the strengths and limitations of their study?

14. Figure 3b: I do not think this figure adds much to the discussion. A more usual visual would be a plot of predicted child mortality risks at the 90% probability level for instance (as Figure 3a gives the (I assume) posterior means or medians).
- Minor Essential Revisions

1. Orthographic errors (non-exhaustive list)
   b. Page 2, Line 6: “was adopted”, not “was adapted”
   c. Page 2, Line 9: “Results”, not “Result”
   d. Page 2, Line 10: “approaches seven”, not “approaches to seven”
   e. Page 2, Line 17: “findings”, not “finding”
   f. Page 2, Lines 17-18: reword “child mortality is a complex situation with at the individual, household and community level” or remove; it does not add to the Conclusion
   g. Page 3, Line 3: “are”, not “is”
   h. Page 3, Line 3: add comma after “In 2012”
   i. Page 3, Line 3: update with 2013 figures
   j. Page 3, Line 10: “due to the lack”, not “due to lack”
   k. Page 3, Line 10: add comma after “However”
   m. Page 3, Line
   n. Page 3, Line 18: capitalize “health” in “Ethiopian Demographic and Health Survey”
   o. Page 3, Line 18: remove extra “)”
   q. Page 3, Line 22: “located in”, not “located at”
   r. Page 3, Line 25: “country’s”, not “countries”
   t. Page 3, Line 1: add citation and elaboration for the Growth and Transformation Plan
   u. Page 4, Line 4: add hyphen after “country”
   v. Page 4, Line 5: capitalize “demographic” and “health”
   w. Page 4, Line 5: “took place over”, not “took over”
   x. Page 4, Line 5: add “a” before “five”
   y. Page 4, Line 5: add hyphen after “five”

2. Table 1: Add a space before brackets
3. Table 2: Add a space before brackets
4. Figure 1: It is not clear to me what the dots on the map represent, what
“decimal degrees” mean, and how the scale of decimal degrees is used.

5. Figure 2: Label axes and indicate what the dashed lines represent (95% credible intervals?)

6. Figure 3a: What does the scale “feet” represent?

7. Figures 3a and 3b: On what scale are the child mortality risks relative to?

8. Figures 3a and 3b: The colour scheme could be improved, e.g. light to dark red.

9. References
   a. Note “Date accessed” in the first two references
   b. Fix typos in the 3rd reference

- Discretionary Revisions

1. It would be helpful to the reader to have a separate Model Specification section under the Appendix that gives full details on the model, including the priors.

2. To my understanding, STAR models also allow for time effects; it would be interesting if the authors could elaborate in the Discussion on how to incorporate data from earlier EDHS surveys into the analysis to obtain time trends and possibly do short-term projections as well, if that is possible.

   Level of interest: An article whose findings are important to those with closely related research interests

   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.