Reviewer's report

Title: Bayesian Spatial Analysis of Demographic Survey Data. An Application to Childhood Diseases in Zimbabwe.

Version: 2 Date: 30 September 2014

Reviewer: Odwa Atari

Reviewer's report:

General Comments

Generally, the paper is well written. However, there is need for substantial corrections before I will accept its publication.

1. There is need for a historical context
2. The methodology section is very long and some variables used have very many missing values. There is need to exclude these variables and rerun the models.
3. Need to avoid repetitions. Report the general trends based on the prevalence of diseases (1999 - 2011), compare similarities and differences on prevalence, spatial trends and covariates?
4. There is an over generalization in the discussions
5. There too many figures and tables.
6. More detail comments and suggestions can be found below:
7. There is need to address these concerns before I can decide whether to accept or reject the manuscript:

Abstract

Introduction
Line 29: why is it in recent decades?
Line 31: program instate of planning
Lines 31 - 34: confusing...need to restructure

Methods
Lines 35 - 38: Confusing. Need to restructure. For example: Bayesian semi-parametric regression models were used to explore the spatial variations of childhood diarrhoea, fever and cough, and their associations (....) with a range of determinants including....

Results and Conclusion
Lines 43 - 44: need to highlight key findings
Lines 44 - 54: Too much repetition; reduce or use alternative words: association
Background
Generally, good background. However, there might be need for a historical background to provide context to some of the problems faced in Zimbabwe.
How does the population survey of health and fertility provide valuable insight into the etiology of disease?
What does this statement mean: which are a great deterrent militating against the demographics of under-developed nations?
The reference to methods used is in future...meaning they are yet to be used.
Lines 92: ...models, introduced by whom?
Lines 100 - 101: what are the spatial, categorical, and continuous variables that will be used???
Lines 103: proposed by???
Lines 107: referring to which factors
Lines 111 - 112: such research into childhood diseases has been aided by significant improvement in what?

Method
Generally, the methodology section is well written but it too long. Need for simplicity. This is not a statistics paper. Other concerns are outlined below:
Lines 139 - 145: how do you compromise the spatial variations of data in your analysis?
Line 164: the triple what?
Line 203: according to whom?
Line 207: taking (3) means what?
Line 289 the diseases are not modelled together but individually???
Line 292: the following predictors: which are?
Line 311: which version of BayesX was used... need for citation

Results
Good results section but there might be need to avoid repetitions:
Report the general trends based on the prevalence of diseases (1999 - 2011), compare similarities and differences on prevalence, spatial trends and covariates???
There are too many figures, need to consolidate some of the figures or report them in the supplementary section....
Residual spatial district effects ... have no significant levels maps
Lines 330 - 339: need to explain chronologically explain the results of the different models developed and there differences and why one model is selected
over the other
Lines 346 -348: If either model M2 or M3 could be used what is the importance/significance of the non-linear effects of child's and mother's age in your results and modelling
Lines 380 - 383: confusing... need to restructure the statement
Lines 384 - 388: could go to the background
Lines 631: IG stands for????
Lines 637 - 645: just report one table: if need be the report the rest of the tables 8-9) on the supplementary section

Discussion
Lines 654-655: This statement is an over generalization???
Lines 659: who suggests????
Lines 684 -686: what is the evidence??? Citation may help?
Lines 775-785: need for evidence??
Lines 796-802: but you stated earlier that the facilities are ill equipped. How will this help improve childbirth?
Lines 809 - 811: why use variables with a lot of missing values????? Need to exclude variables with very many missing values (e.g. vaccination variable) and rerun the models
Lines 812 -814 some of the vaccinations for these diseases will reduce the rate of fever.
Lines 848 - 850: Which cohort are you referring to here, since you don't have overall percentages for your total study subjects?
Lines 862 - 865: no evidence from your data
Lines 873 874: how is this possible. Explain???
Lines 876 - 878: Need to support the claim
Lines 883-884: there might be need to provide a table that shows the prevalence of these diseases by provinces
lines 895-896: where???
Lines 924 - 925: You mentioned around the start that the data was collected at an EA level????

Other comments
There are too many figures. Need to consolidate and remove some
There are also too many tables and some have problems. For example in Table 1 the total numbers of children with and without diseases are not equal to the total number of observation recorded.
Need to combine some variables for example Secondary and higher (only 69 cases) educations. Exclude variables that have very low counts or too many
missing variables from the analysis

With the posterior mean figure, how do I know which areas (provinces) are significantly different????