Author's response to reviews

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

Authors:

   Rodney G Tsiko (tsikoruz@yahoo.com)

Version: 4 Date: 22 January 2015

Author's response to reviews: see over
Comments from Reviewer Mikoko Terashima

Specific Comments

Abstract

<Line 28>
“proximate determinants” refers to “factors that directly affect”. As such, “proximate determinants” has been replace with “factors or situational events that directly”

<Line 30>
Has been changed to: “Understanding the geographic distribution of…”

<Lines 35-38>
The sentence was restructured as follows:

“Bayesian semi-parametric regression models were used to quantify the spatial risk of childhood diarrhoea, fever and cough, as well as associations between childhood diseases and a range of factors, after accounting for spatial correlation between neighbouring areas.”

<Line 44>
The line has been rephrased as suggested by the reviewer. In addition, alternative words have been used to replace the word “association” to avoid repetition.

Background

<Line 61>
Epidemiological variables are characteristics that can be observed and/or measured such as time and place of exposure to a disease, age, gender or occupation of the person exposed. As such, examples have been provided to help clarify the meaning of “…a range of epidemiological variables”

<Lines 63-64>
“…which are a great deterrent militating against the demographics of under-developed nations” refers to the fact that childhood diseases are having a negative effect on the population dynamics of under-developed nations.

<Lines 92-110>
The majority of the paragraph has been removed leaving only a general description.

Methodology

<Around Line 137>
The author has added a table (Table 1) showing the sample size and related population at the provincial level.
Anthropometric status refer to indicators of nutrition or malnutrition often used by the World Health Organization. These include stunting, wasting, and underweight, and mid-upper arm circumference in children under five years of age and Body Mass Index in adults. The author replaced the term with an appropriate phrase.

The author took into account the possible audience of the paper and so the sentence has been rephrased and written in a way hopefully everyone would be able to understand.

Results

It definitely is 95% credible interval since its 2.5 on either side of the Gaussian graph. Thanks for noticing that.

Discussion

The author has retained the biological point of view but has removed the economic point of view as he feels that preference of gender by a parent has no bearing on the probability that child will/will not be affected by a common disease such as cough or fever. If the reviewer knows otherwise, the author would welcome any suggestions.

Statement changed as suggested

The author has recommended further investigation and would like to thank the reviewer for the suggestion.

Examples of “mother’s resources…” have been provided to help understand the statement.

Sentences were swapped.
This is a factual statement and documentation has been cited

Poverty datum line represents the cost of a given standard of living that must be attained if a person is deemed not to be poor.

“However…” has been used

Statements have been rewritten and some rephrased to avoid iteration.

Sentence rephrased.

Other studies have been cited for further reading

General Comments

- The number of Figures has been reduced to 8. Most have been consolidated.
- Although there are too many tables, the author deems them all necessary in conveying the findings of this research. In addition, Table 1 has been corrected to make sure the number of children with and without diseases are equal to number of observations. The error in the first place was including the missing values in the total number of observations.
- The methodology section has been reduced in length and written in such a way that hopefully someone without a background in statistics can understand it be able to reproduce the results.
- Concerning variables with missing values, the models in the first place were run excluding the missing values hence no need to rerun them. The percentages of the missing values (no matter how high they are for some variables) have been included in the study to highlight the difficulties of obtaining data in less developed countries. The author hopes that other researchers can come up with new ways of dealing with this problem besides leaving the variables out.
- The results section has also been reduced in length. General trends based on the prevalence of diseases (1999 - 2011) were reported, then similarities and differences on prevalence, spatial trends and covariates were compared.
Comments from reviewer Odwa Atari

Specific comments

Abstract

<Line 29>

Why under-five morbidity and mortality have increased in recent decades: Because the number of live births and population of children under age five has grown rapidly in the last 20 years in Sub-Saharan Africa (due to a high total fertility rate of 5.1 and the fact that only 20 percent of married women use modern forms of family planning), meaning more children have been exposed to childhood diarrhoea, cough and/or fever.

<Lines 31-34>

The first sentence has been restructured to;

“Understanding the geographic distribution of such diseases and their relationships to potential risk factors can be invaluable for cost effective intervention.”

The second sentence has been removed as the author felt it was best to have it under Methods.

<Lines 35-38>

The sentence was restructured as follows:

“Bayesian semi-parametric regression models were used to quantify the spatial risk of childhood diarrhoea, fever and cough, as well as associations between childhood diseases and a range of factors, after accounting for spatial correlation between neighbouring areas.”

<Lines 43-54>

Alternative words have been used to replace the word “association” to avoid repetition.

Background

<Line 60>

The word “etiology..” had been regrettably used in the wrong context. The author has used another word and the sentence reads as follows:

Population surveys of health and fertility provide valuable insight into the prevalence of diseases in third-world countries

<Lines 63-64>

“…which are a great deterrent militating against the demographics of under-developed nations” refers to the fact that childhood diseases are having a negative effect on the population dynamics of under-developed nations.
“Geoadditive models, introduced by Kammann and Wand [7]…”, has been added to make the statement clearer.

Considering that the author intends to make this paper more oriented towards the particular journal’s audience this paragraph was deemed not suitable in the background section. Only a general description has been kept and the other material moved to the methodology section. In addition, the spatial, categorical and continuous variables have been identified in the methodology section.

Research into childhood disease has been aided by the three reasons given in the subsequent sentences.

The author also identified that this particular sentence was unclear because of the paragraph above it. As such the “above paragraph” was removed.

**Methodology**

“Consider the triple…” was meant to be “Consider the three childhood diseases (diarrhoea, fever and cough)…”.

However, the author took into account the possible audience of the paper and so the sentence has been rephrased and written in a way hopefully everyone would be able to understand.

“According to Besag, York and Mollie [13]…”

“taking (3)…” means using equation 3 into account. The author has rephrased that part of the sentence.

Yes, the diseases are modelled individually, that is for each dataset the author separated the observations for diarrhoea, cough and fever and then applied the set of predictors shown by equations 6-9 to the observations for each disease.

“the following predictors…”, which are the subsequent equations 6-9.
Version of BayesX used was included and cited.

**Results**

<Lines 346-348>

The variables child’s age and mother’s age at birth of first child are just like any other variable used in this study, so in other words they are not special in any way. But what the author sees as special is treating variables in a statistical model the exact way they are related to the data collected. For this study, all variables were plotted on graphs to see how they relate to the data on childhood diseases and only these two were found to have a non-linear relationship with childhood diseases. If it was in another dataset and they had been collected as age groups (e.g. 0-10 months, 11-20 months) then the author would have treated them as fixed variables with a linear effect of childhood diseases.

<Lines 380-383>

Statements restructured

<Line 631>

IG stands for “Inverse Gamma”. The author has expanded on this abbreviation.

<Lines 637-645>

The three tables have been consolidated into one table.

**Discussion**

<Line 659>

Sentence rephrased to include citation about who came up with the theory.

<Lines 684-686>

Citation has been provided

<Lines 775-785>

The author has given his own explanation on the negative association between childhood diseases and children of mothers who received antenatal care during their pregnancy, but has also recommended further investigation

<Lines 796-802>

Yes, facilities are ill equipped, but these are mainly government sponsored hospitals. In the private sector, equipment is available and service is better (which improves childbirth) but it comes at a higher costs. The author has rephrased the statements to highlight this.

<Lines 809-811>
The author understands the concern raised by the reviewer about variables with too many missing values. Unfortunately, this is the nature of the data available to the public from most less developed countries. For us researchers it is a difficult scenario as the results don’t give the true picture but we have to try to make sense of things with the little data that can be found. The author would also like to assure the reviewer that BayesX allows one to run models whilst excluding cases/observations with missing data, which was done for this study hence no need to rerun the models.

The DHS data provides data on women’s age and when they had their first child. From this the author was able to find how many women within a certain age group had their first child at 18 then convert it to a percentage of the total number of women sampled.

This was a typing error, the provinces were swapped. Mashonaland provinces had a higher prevalence of childhood diseases than Matebeleland provinces. See Figures 4, 6 and 8.

Because health facilities in rural Zimbabwe are sparsely distributed when compared to those in urban areas due to population densities, and are mostly located a long distance from most rural villages thus issue of accessibility.

Other studies have been cited to support claim

“provincial level…” changed to “EA level…”

Other general comments

- The number of Figures has been reduced to 8. Most have been consolidated.
- Although there are too many tables, the author deems them all necessary in conveying the findings of this research. In addition, Table 1 has been corrected to make sure the number of children with and without diseases are equal to number of observations. The error in the first place was including the missing values in the total number of observations.
- The methodology section has been has been reduced in length and written in such a way that hopefully someone without a background in statistics can understand it be able to reproduce the results.
- Concerning variables with missing values, the models in the first place were run excluding the missing values hence no need to rerun them. The percentages of the missing values (no matter how high they are for some variables) have been included in the study to highlight the difficulties of obtaining data in less developed countries. The author hopes that other researchers can come up with new ways of dealing with this problem besides leaving the variables out.
The results section has also been reduced in length. General trends based on the prevalence of diseases (1999 - 2011) were reported, then similarities and differences on prevalence, spatial trends and covariates were compared.