Reviewer's report

Title: Spatial and temporal variation and hotspot detection of kala-azar disease in Vaishali district (Bihar), India

Version: 2 Date: 14 September 2012

Reviewer: Guilherme Loureiro Werneck

Reviewer's report:

Major Compulsory Revisions

As I said in my previous report, the manuscript is interesting but has many problems. Although authors made a huge effort to revise the manuscript as required, I think it still needs further revision.

In my previous report I've raised a point of major concern that was not considered by the authors. Their paper is on using spatial statistics to help improve their knowledge about VL epidemiology and to inform control actions. However, they sometimes invert the objective and try to make it a more of a "methods paper" by giving too much emphasis on the techniques than on the subject. Spatial statistics is just a set of tools for getting better information from data and many papers have already shown the usefulness of these tools in studying the epidemiology of visceral leishmaniasis and other vector-borne diseases. So, for this paper to be relevant, it should focuses on the possible explanations for the findings and not on the techniques used. Authors used too many techniques that are, in a sense, redundant and make the manuscript too complex for the readers interested in the epidemiology of VL. I strongly suggest that the authors try to review the paper in that direction, putting more focus on the epidemiological aspects that arise from the use of spatial data analysis than on the techniques themselves.

Other major points:

- Lines 48-51: This phrase is, in a certain way, redundant considering the first phrase of conclusion. You may want to simply delete it and substitute it by another that says something concerning your findings and not to the utility of GIS and spatial statistics.

- Line 136: still have to include information about how “spatial weights” were defined. Also, I do not think “annualized” is a correct term, use just “annual average incidence”.

- Line 141: what does it mean “higher geographic scale”? Aren’t the rates just depicted at the same scale (that is, villages)?

- Lines 152-154: IDW was not used to “map”, but to produce predictions of incidence rates across the whole area. As a matter of fact, phrases the two very
- Lines 161-165: I do not think you should say here that kriging could be used. It seems ok just to describe IDW here and in the discussion section say that kriging would be an alternative for IDW. Actually your justification for IDW is debatable; many authors feel that kriging perform better than IDW (see, for instance, Meng et al. 2010, International Journal of Health Geographics and Yasrebi et al. 2009, Research Journal of Biological Sciences).

- Line 173: you say that “Spatial autocorrelation analysis was performed in the confirmed cases...”. Was that on “confirmed cases” or on “incidence rates”?

- Line 177: I think you are not testing “how villages were clustered”, but whether villages with high (or low) incidence were clustered.

- Lines 220-222: I am not sure what you mean by saying that “However, all the statistical methods used in this study do not depend on the absolute positions of the village. The results depend only on the relative position of cases compared to non-cases among all the villages.” As far as I understand you do not have information on location of cases within villages, that's the answer I’ve got from you! In your answer to my previous question you said “The study has been carried out up to village level, not at the household level.” Please make it clear.

Minor Essential Revisions

- Line 47-48: I do not think that you can conclude about the “necessity” of spatial statistics, maybe be you can say “usefulness” of spatial statistics

- Lines 58-60: this sentence on clustering of VL seems out of place and could be deleted since similar information is given in other parts of the text.

- Lines 73-75: Do not think you need it here, similar information is given in other parts of the text, I suggest just to delete.

- Line 78: I do not think the word “natural” is correctly used here; actually most of the transmission cycles of vector-borne diseases of public health interest are not “natural”. I suggest removing the term.

- Lines 78-79: the phrase “Although the progression in computer technology has been ample, the analysis, particularly in spatial and temporal terms is a very complex process” is not relevant and could be deleted.

- The sub-section “Measures of disease of occurrence” should come before the “spatial analysis” section.

- Lines 154-155: This phrase is a limitation of the IDW method and should be relocated in the Discussion section when considering limitations of the approach used.

- Lines 159-160: The phrase “The IDW method is to explore the condition of influence estimated value more by nearby points...” is awkward, please rewrite.
- Lines 224-225: I think the phrase “Mapping adjusted incidence per 10,000 inhabitants and per year allow us to analyze and compare visually the spatial pattern of the disease.” should be deleted, it is not a Result, just a motif for using maps.

- Lines 262-264: I do not think you need to describe the expected value and z-scores, they are meaningless for epidemiologic purposes and they are already on the table!

- Lines 283-284: same, do not need to describe the Values of G and Z-score.

- Tables 1 and 2: p-values cannot be zero or <0. Also strange is the value zero for variance.

**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.