Reviewer's report

Title: Spatial analysis of factors associated with HIV infection among young people in Uganda, 2011.

Version: 1 Date: 15 February 2014

Reviewer: Veronica V Escamilla

Reviewer's report:

Discretionary Revisions:
1. Use “geographic clusters” instead of the term “hotspots”
2. Use a different term for “young people” such as sample, population, or participants

Minor Essential Revisions:
1. The sample numbers in the abstract do not match – the sample size should be presented in the methods section of the abstract
2. The number of HIV positive participants is inconsistent throughout the paper – Figure 1 reports 533 HIV positives, however this is not consistent with the 7% HIV positive reported in Table 1
3. Posterior odds ratio should be written before using acronym in the results of the abstract
4. Abstract – present results in the same order as the methods
5. Add a footnote to table 1 explaining the *
6. Table 1 – The ‘n=7518’ does not match the sample size in the table

Major Compulsory Revisions:
1. The overarching goal for the study is not clearly defined, and it is unclear why both the cluster detection and spatial random effects models are used – Is it to identify geographic clusters of HIV in high disease burden areas?
2. Additional background information describing other HIV/STI studies using spatial methods would help frame the paper
3. Second paragraph in background – Add more detail explaining the spatial autocorrelation and its importance in HIV research
4. Methods – what do the “clusters” in the 1st paragraph represent?
5. The data analysis section is difficult to follow. The model specifications are missing and should be added (e.g. what is the Bayesian model measuring and how are posterior odds ratios estimated?).
6. The multiple logistic regression and spatial random effects models should be presented in separate paragraphs in the methods.
7. An explanation for the use of two methods detecting high incidence (scan statistic and Bayesian model) needs to be added to the methods section.
8. Overlaying results from the cluster analysis and Bayesian model would be useful.
9. The variables used in all models need to be described in the methods section.
10. Pg 5 para 2: What is “area-based data aggregated to cluster level”?
11. The model variables are introduced for the first time in the results section. A paragraph describing the variables at multiple levels would help clarify the methods section.
12. The text in Figure 2 is difficult to read; how are the high risk areas in figure 2 defined?
13. Secondary and tertiary clusters need to be defined in the methods section before reporting the results.
14. Results – multiple logistic regression and spatial regression model results should be presented separately. It would help to add sub-headings (e.g. multiple logistic regression and spatial random effects)
15. All of the variables are introduced for the first time in the results and need to be introduced in the methods section
16. Pg 8 – ‘young coital debut’ is included in regression model but is missing from table 3
17. Pg 8 – there is a reference to OLS estimates, but wasn’t a logistic regression model used?
18. Pg 10 – the scan statistic adjusting for confounding variables is listed as a strength, however this is not a strength if confounding variables were not included in the cluster analysis. Were additional variables included in the cluster analysis?
19. Pg 11 conclusion – why are fishing communities as high risk populations introduced in the conclusion?
20. What analysis are the p-values in table 1 referring to? This needs to be explained in methods and results.
21. The random and spatial effects in table 3 need to be explained.
22. Why are both a non-spatial and spatial regression model used? Is it to make a comparison of results from a non-spatial and spatial regression? If so, the importance of the comparison needs to be explained in the background and presented in the results and discussion.

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:** No, the manuscript does not need to be seen by a statistician.

**Declaration of competing interests:**

I declare that I have no competing interests.