Reviewer's report

Title: Socioeconomic Inequities and Hepatitis A Virus Infection in Western Brazilian Amazonian Children: Spatial Distribution and Associated Factors.

Version: 1  Date: 4 June 2015

Reviewer: Robin Turner

Reviewer's report:

This manuscript presents an analysis of Hepatitis A virus infection in Western Brazilian Amazonian children. The analysis is presented well; however, I have a number of concerns about the statistical methods used which I outline below.

Major Compulsory Revisions

1) If I have understood the analysis correctly the spatial properties are investigated using a fixed effects model. The analysis does not investigate the spatial correlation, nor look at the associations between the outcome and predictors accounting for this spatial correlation. The analysis really needs to take this next step to explain how much of the spatial distribution is explained by the predictor variables. Spatial analysis techniques would allow this to be done correctly. This also means that the logistic regression results may have standard errors that do not account for the spatial nature of the data. A more comprehensive analysis would allow these two disparate analyses to be done simultaneously and correctly account for the spatial nature of the data.

2) The figures show the spatial distributions with the individual data points placed on top. This information is potentially identifiable. By knowing the map of the area the points could be aligned to work out individually the characteristics of each person and whether they have the disease or not.

3) Page 9 lines 194-196. This does not match what is presented in the table, most children actually have a flushed toilet (54.5%), do not have direct contact with sewage (54.2%) and do not flood on rainy days (54.7%)

4) For the variable “number of person per room” in Table 1, has this been fitted continuously? If so was the functional form assessed (i.e. linearity) and by what method? Presenting the 16.66% overall proportion is not particularly useful here. It would be more useful to have a categorical version also showing how the proportion varies as the number per people per room increases.

5) Age is presented categorically (though it may have been measured continuously). However in the adjusted analysis only the 4 to 5 year age group is used (versus all other ages). Why was this done? There looks to be some increasing trend across the age categories and collapsing of age groups in this way means that there isn’t full adjustment for the confounding effect or age, nor full assessment of the relationship between age and the outcome.

6) Figure 2 legend says that “No areas with significant higher positiveness
density were identified”, however page 11 line 241 says “A hotspot of anti-HAV-positive children was found in the new easternmost area..”. Is there a significant area or not?

7) Page 6 line 135 “distribution of the independent variables was identified using Student’s t-test …”. These tests do not investigate the distribution but assume distributions. The methods used to assess the distributions should be outlined separately to the statistical tests used to investigate differences and associations.

Minor Essential Revisions

8) Decimals need to be given consistently throughout. For example on page 5 line 110 has a percentage given to 2 decimal places and line 111 has a percentage to no decimal places. Please change them all to be consistent.

9) Figure 1 y-axis needs a more descriptive label than ‘%’.

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

Quality of written English: Acceptable

Statistical review: Yes, and I have assessed the statistics in my report.

Declaration of competing interests:

I declare that I have no competing interests