Reviewer's report

Title: The Spatial Epidemiology, Clinical Presentation, and Environmental Risk Factors Influencing the Transmission of La Crosse Virus in West Virginia

Version: 1 Date: 29 April 2010

Reviewer: Kees C. van den Wijngaard

Reviewer's report:

This is a study on the spatial epidemiology, clinical presentation and environmental risk factors associated with La Crosse virus (LACV) cases in West Virginia in 2003-2007.

Major compulsory revisions

1. In the methods section, page 6 line 116-117, the authors should explicitly describe the difference between confirmed and probable cases and also present how many cases were regarded confirmed and possible. They might also want to evaluate the chance that a patient with meningitis or encephalitis due to causes other than LACV has LACV antibodies due to earlier asymptomatic LACV infection, and thus could be wrongly included in the study.

2. Also in the methods section, the authors should add how many reported LACV cases during the study period were excluded due to missing data on clinical symptoms and location, or due to the fact that they were not <=15 yrs of age. They should also assess (in the discussion section) the possibility of variation in reporting clinical symptoms and/or location data between different health departments, leading to bias in the analyses results.

3. For spatial analyses of the incidence risk per region (either county or census tract), the authors use spatial empirical Bayesian smoothing, and present maps with both the unsmoothed and smoothed incidence risks, both showing higher risk in the south of the state. Then the authors use Global Moran’s I test to confirm that there is significant spatial autocorrelation in the data, and the Moran Local Indicators of Spatial Association (LISA) to confirm the location of significant high-risk clusters. The statement of the authors on page 14 line 304-306 that for this LISA analyses they did not correct for multiple comparisons seems controversial. Especially for the census tracts (n=466) it seems likely that statistical significance will occur by coincidence. Although they cite a publication that proposes not to correct for multiple testing, other publications do propose correction. The authors could discuss the effect of adjusting on their analyses results: how many significant clusters are still detected after adjusting for multiple comparisons? Another option could be to consider methods for spatial cluster detection that automatically adjust for multiple testing like the spatial scan statistic published by Kulldorff. An additional benefit of this and possibly other alternative methods is identification of clusters of various sizes, so also clusters
that contain several census tracts, instead of one at the time. This prevents that such “multiple-census-tract-clusters” are missed if the census tracts are separately not significant.

4. In the results section of the abstract the authors indicate the number of detected high-risk clusters as “n=4” and “n=30” (on county and census level respectively). This is confusing, as some readers might think that “n” indicates the total number of counties or census tracts. The authors should therefore rephrase this.

Minor essential revisions

5. Regarding the environmental risk factors the authors acknowledge in the discussion section that they did not perform a case-control study, but only gathered information of (known) risk factors for reported cases. For this reason they should rephrase line 39-41 on page 2 (abstract), as some readers might think that they investigated what the environmental risk factors are for LACV cases, whereas the authors actually describe the results of gathering data on apparently already known environmental risk factors for reported cases.

6. In addition to comment 5 above: Page 15, line 323-324 “The presence…. LACV”. Some readers might think that this sentence at the end of the article also suggests that this study proofs that wooded areas and containers are a risk factor for LACV, whereas this is not a case-control study, as mentioned before. Please rephrase.

7. In the discussion section, page 10 line 213-216, the authors mention an increase in incidence, whereas in this paper they describe spatial variation in the incidence of LACV in 2003-2007, and not an increase in time. Please rephrase or clarify.

Discretionary revisions

8. The authors already described in an earlier study (reference 37) that spatial analysis on data with higher spatial resolution (census tract level vs county level) leads to more precise results on the spatial distribution of disease incidence. It would therefore greatly improve the readability of the current paper, if only the census tract level results would be extensively described. The comparison between results on county and census tract level could then very shortly be described in one paragraph, just to confirm that the census tract level is the preferred spatial resolution for these analyses.

9. Page 4 line 79-80 The authors might want to rephrase this sentence. Probably they intend to say that LACV meningitis and/or encephalitis has been emerging in West Virginia, and not that LACV has been found to be the cause of an emerging disease of formerly unknown cause?

10. Some minor details:
Page 7 line 141, shouldn’t it be “the number of cases per 100,000 persons per
What is the “block level”? Same as county level? Please clarify.

Page 11 line 237-238 repeats statement made in line 227-228.

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

**Quality of written English:** Acceptable

**Statistical review:** Yes, but I do not feel adequately qualified to assess the statistics.

**Declaration of competing interests:**

I declare that I have no competing interests