Reviewer's report

Title: Space-time cluster analysis of indigenous and imported dengue fever cases in Guangdong province, China

Version: 2 Date: 30 November 2011

Reviewer: Aeilko Having Zwinderman

Reviewer's report:

The paper considers the distribution of dengue fever in the Guangdong province of China both in time (2005 to 2010) as well as in a geographic sense. Large differences in incidence were observed between years as well as between different regions in Guangdong. This applied both for indigenous and for imported cases. It seemed as if the region with dengue fever cases grew with time from 2005 to 2010.

The paper is well written, but there are a number of points unclear. They are listed below. I have one major remark and that concerns the relevance of the present results for the wider readership. It seems to me that the paper reports results that are highly specific for the Guangdong area in China but it is very unclear what the general message is for other parts of the world. In addition it is also unclear what the current analysis implies for the Guangdong area. The dengue incidence increases and this seems to be the case for many sub-areas but what does this mean for the future healthcare organizations?

1. The study included both clinically and laboratory confirmed cases of dengue fever. First I would like to know how many were clinically confirmed and how many laboratory confirmed. Since the definition of imported of indigenous cases was based on the opinion of epidemiologists this may bias the results and I wonder secondly whether it is possible to do separate analyses for laboratory confirmed cases only.

2. What is meant by age-standardization with the direct method. Is this standardization already applied in the figures in Table 2 or is it done in the regression analysis and if so how?

3. The geographical expansion was modeled with a linear regression model. It is unclear however what the dependent and independent variables of this model are: number of affected counties and distance from the area of 2005? Why a linear regression model?

4. The scanning statistic compares the incidences within and outside the window with clusters defined as 10 to 50% of the total population. Does this mean that the number of clusters must be between 10 and 2? If so, is that realistic? With a total population of 104M there seems plenty of power to identify much smaller clusters?
5. On page 10 in the results section on the second line of the second paragraph “the age-standardized cumulative incidence” is given as between 0 and 70.6 in 100,000. Why are these statistics cumulative incidences? Over what is cumulated?

6. Page 11. What do the regression weights of 355.6 and 261.8 mean? Is this kilometers per month?

**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.