Author's response to reviews

Title: Spatiotemporal analysis of indigenous and imported dengue fever cases in Guangdong province, China

Authors:

Zhongjie Li (lizhongjiecdc@163.com)
Wenwu Yin (yinww@chinacdc.cn)
Archie Clements (a.clements@sph.uq.edu.au)
Gail Williams (g.williams@sph.uq.edu.au)
Shengjie Lai (laishengjie@foxmail.com)
Hang Zhou (zhouhang@chinacdc.cn)
Dan Zhao (janedano@163.com)
Yansha Guo (guoys@lreis.ac.cn)
Yonghui Zhang (zyh@cdcp.org.cn)
Jinfeng Wang (wangjf@lreis.ac.cn)
Wenbiao Hu (w.hu@sph.uq.edu.au)
Weizhong Yang (ywz126@vip.sina.com)

Version: 3 Date: 25 February 2012

Author's response to reviews: see over
Dear Prof. Mockenhaupt,

Thank you very much for your letter of the 11\textsuperscript{th} January 2012 including the referees’ comments on our paper. Please find enclosed a copy of our revision with the following changes, according to the suggestions/comments:

**In response to referee 1 (Mathuros Tipayamongkholgul )**

**Concern 1.1:** The authors should explain in detail how to distinguish the imported case from indigenous. What are the significant criteria?

1. More details about how to distinguish imported and indigenous cases, including the criteria, are provided on page 7.

   “An imported case of dengue fever means the case came from outside of the mainland of China with a history of being bitten by mosquitoes overseas within 15 days before the onset of illness, or the gene sequence of the virus isolated from the case is highly homologous with that reported by the country to/from which the patient had travelled. Indigenous cases are defined according to the absence of evidence for the case being imported.”

**Concern 1.2:** Using the linear regression model to examine the expansion of dengue affected area may be incorrect, since the pattern of an expansion of dengue affected area is generally in Poisson distribution.

2. We have decided to remove this analysis in the revised manuscript and will perform a comprehensive analysis to explore the spatial dispersion of dengue in another follow-up study.

**Concern 1.3:** The paper would be much more interesting if the authors could examine whether the cluster of imported cases relate the cluster of indigenous cases in county or its vicinity

3. To explore the relationship of clusters between the imported and indigenous cases would be interesting. However, it would be better to incorporate social and environmental factors to examine the relationship. Currently we don’t have such detailed data. We will fully explore the relationship between the occurrence of indigenous cases and the imported cases in a future study.

**Concern 1.4:** Why did the imported case present seasonal pattern? Where were the imported cases from? Was the infected area of those imported case relevant to the seasonal in the original infected area?
4. Some reasons for the seasonal pattern of imported cases were added on page 13

“Most of the imported cases in Guangdong province came from Southeast Asian countries with endemic dengue fever, including Singapore, Indonesia, Cambodia and the Philippines. Several studies have also reported that dengue fever cases from South-east Asian countries have triggered local outbreaks in Guangdong province. The seasonal pattern of epidemics in these Southeast Asian countries and the seasonality of travel to Guangdong were possible reasons for imported cases presenting a seasonal pattern.”

Response to the concerns from referee 2 (Benjamin M. Althouse)

Concern 2.1: The large problem I have with the manuscript is that it is too terse. It is short on justification for the statistical methods employed, details regarding the statistics and the interpretation of the results of analyses. I have trouble seeing the utility of the Kulldorf scan statistic here when a simple plot of cumulative incidence identifies the relevant clusters. The authors need to justify carefully this choice of statistic. They also need to do a better job of interpreting the results. What does a RR of 7.52 actually mean for the “most likely cluster of imported cases”? Confidence intervals need to be added throughout the manuscript. A statement of “P<0.001” is not informative.

5. A justification for use of the Kulldorf scan statistic in this study was added on page 8. More detailed interpretation of these results has been added on pages 12-14.

“Space-time cluster analysis was performed on the imported and indigenous dengue fever cases to test whether the cases were distributed randomly over space and time and, if not, to locate space-time clusters and determine their statistical significance (inference which cannot be made from simple visualisation of the raw data). The software SaTScan version 9.0 was employed to conduct the retrospective space-time scan statistic test. In this study, cases were assumed to be Poisson-distributed in each location. The program applies a likelihood function to scanning windows, which move over areas and centre on the centroid of each county. The statistic compares observed and expected case numbers inside and outside the scan window to detect clusters that are least likely to have occurred by chance. The statistical significance for each cluster is obtained through Monte Carlo hypothesis testing, with the number of Monte Carlo replications set to 999. The relative risk (RR) is given by the SaTScan software to present the risk of disease within the scanning window compared to that outside of the scanning window. SaTScan also identifies secondary clusters in the data set in addition to the most likely cluster, and orders them according to their likelihood ratio test statistic. In this study, only the most likely cluster and the secondary clusters with statistical significance are reported when the p-value of the clusters below 0.05. For the space-time scan statistic testing in our study, the spatial data element was the county of case’s residence and the temporal data element was the month of case’s onset of the illness, which were obtained from the individual information of the recorded cases.”
Concern 2.2: Additionally, I’m worried about potential surveillance biases. Are the majority of cases from hospitals or doctors offices? Are you more likely to detect cases in areas of high population density merely due to the fact that there are more hospitals or doctor’s offices? This should be acknowledged in the manuscript.

6. More details about the dengue surveillance system are provided on page 7.

“A standard form was adopted by physicians and epidemiologist to collect individual information on each dengue fever case, including age, address, date of onset, diagnosis, laboratory test result and travel history. Routine case reporting is done by hospitals. However, because dengue fever is an emerging disease in China, possible dengue fever cases are also traced by active field investigation when outbreaks occur in the community. Diagnosis in this case involves laboratory testing by the local public health institute. Thus, dengue surveillance involves both passive and active case detection.”

The potential limitations of the surveillance data were discussed on page 14.

“A potential limitation to the study could be underreporting and misreporting of both imported and indigenous cases due to asymptomatic infection, clinical misdiagnosis, and the capacity of some laboratories to test samples. However, because dengue fever is a notifiable disease and a high priority in Guangdong, surveillance data quality supervision was regularly performed in clinics and hospitals. More importantly the capacity of laboratories to do dengue testing was relatively high in this province. Therefore, we have reasonable confidence in the quality and completeness of reporting. “

Concern 2.3: The conclusion that control measures should be targeted at the clusters, while correct, is obvious based on the cumulative incidence and population density. If the authors wish to use the Kulldorf statistic, they should say what about it justifies the conclusion above and beyond cumulative incidence and population density.

7. More discussion about the implementation of Kulldorf statistic are provided on page 8

Refer to item 5 for response to concern 2.1.

- Major Compulsory Revisions

No. 1: Add justification for the use of the Kulldorf scan statistic as well as more details about the assumptions involved and the actual analysis done. What statistical software was used?

8. Refer to item 5 for response to concern 2.1.

No. 2: Add more detail about the linear regression model used to assess geographic spread. What was the model form?
9. The linear regression model has been removed (please refer to item 2 for response to concern 1.2)

**No. 3:** Add interpretation of the RR from the Kulldorf analysis.

10. We have provided further interpretation of the RR from the Kulldorf analysis in the revised manuscript on page 8.

“The relative risk (RR) is given by the SaTScan software to present the risk of disease within the scanning window compared to that outside of the scanning window. “

**No. 4:** Add 95% confidence intervals to all reported statistics.

11. While the SaTScan software employed in this study does not provide 95% confidence intervals for reported statistics it does provide a P-value for the reported clusters, which we have provided in the paper.

--- **Minor Essential Revisions**

**No. 1:** P. 7: Wrong citation? “with a rainy season from April to September [1]”

12. A correct citation was used on page 6 as the referee suggested. Thanks.

**No. 2:** There are spelling and grammatical errors throughout the manuscript.

13. Yes, we further carefully checked the spelling and grammar of the manuscript, and revised all the errors found.

**No. 3:** P.13: Where is the evidence for this sentence: “A large number of people have moved from poor areas of inland China to this region to seek job opportunities and better living conditions, and this group has increased the susceptible population in this region.”? How do we know those moving into the area are susceptible? Are they a large enough group compared to the existing population to change disease dynamics?

14. A relevant reference for the population migration was cited on page 12. More discussion of the susceptible population and disease dynamics changing are provided on page 12.

“As reported by the census in 2010, a mobile population of more than 30 million people came from inland areas of China to live temporarily in Guangdong province, to seek job opportunities. The mobile population is extremely large, accounting for about 1/3 of the total provincial population. It is concentrated in the metropolises especially the cities of Guangdong, Shenzhen and Zhuhai. The mobile population may increase the susceptible population in this region because they originate from inland areas where dengue is absent.”

**No. 4:** There’s no reason on a full-page figure (number 4) for the labels to be “0”, “1”, “2” and force the reader to look at the caption. They should be something along the lines of “Non-affected”, “Ever affected” and “Newly affected”.

15. Figure 4 is revised as the referee suggested.

--- Discretionary Revisions

No. 1: P. 13: I would rearrange this paragraph to have the sentence “Traditionally, residents have had the habit … benefitting mosquito breeding.” Come just after the sentence ending “...having a favorable natural climate for mosquito breeding [3, 4, 22].”


Response to the concerns from referee 3 (Aeilko Having Zwinderman)

Concern 3.1: 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?

17. More discussion about the implications of the study has been added on page 15.

Besides, the study also demonstrated that travel between endemic and non-endemic areas may influence the observed spatiotemporal distribution of dengue in non-endemic regions. For example, in Guangdong province, imported dengue cases have led to seasonal peaks in dengue risk and spatial clustering in areas of the province which receive a higher number of travellers from endemic regions in Southeast Asia. These findings could be used to inform risk-based surveillance aimed at identifying high-risk locations and time-periods during which more intensive control measures can be implemented.

Concern 3.2: 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.

18. The results on the number of cases with clinically and laboratory confirmed were presented on page 9.

“1658 dengue fever cases were reported in Guangdong province from 2005 to 2010, with 94 (5.7%) imported cases and 1564 (94.3%) indigenous cases. 87 (92.56%) imported cases and 1280 (81.84%) indigenous cases were laboratory confirmed.”
We decided to include both clinically confirmed and laboratory confirmed cases in this study, which would benefit to describe the overall space-time distribution and clustering of dengue fever case. Given that dengue is an emerging disease in China, clinically diagnosed dengue fever cases that do not have laboratory evidence to support the diagnosis are reviewed by a group of experienced physicians or epidemiologists. Therefore, we think the accuracy of clinical diagnosis of the dengue fever cases was acceptable, and the clinically confirmed and laboratory confirmed cases can be combined to analysis in this study.

The definition of the imported and indigenous cases has been added on page 7.

**Concern 3.3:** 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?

19. We revisited age-standardisation in our analysis and found that it did not make a difference in the identification of clusters or the observed spatial pattern of cumulative incidence. Therefore, we decided to delete age-standardisation from the study.

**Concern 3.4:** 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?

20. We decide to delete this analysis from the revised manuscript (also please see item 2 for response to concern 1.2).

**Concern 3.5.** 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?

21. 50% of the total population is commonly used as an upper limit of the geographic area of potential clusters (a larger cluster simply infers a small cluster of low risk in the excluded area). This permits detection of large clusters that tend to have a small relative risk but a high statistical significance. It is also common to set a smaller maximum cluster size, such as 10% of the total population, to detect potential sub-clusters. These tend to have a smaller size with a more extreme relative risk. Therefore, this did not mean the number of clusters must be between 10 and 2. Actually, the number of the cluster is decided by the spatial and temporal distribution of the cases. Because the population is large (about 104M) an upper limit of 10% of the total population provides ample power to detect the smaller clusters.

More explanation of the cluster analysis is provided on page 9.
Concern 3.6: 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?

22. The cumulative incidence is a measure of disease frequency during a period of time, which is calculated by the cumulative number of new cases during a period divided by the number of people in the population at the beginning of the study. In our paper, we adopted the cumulative cases by county from 2005-2010 as the numerator and the population of the start of study (in 2005) in the corresponding county as the denominator.

The definition of cumulative incidences is provided on page 8. The age-standard method of calculating cumulative incidence was deleted (please see item 19 for response to concern 3.3).

Concern 3.7. Page 11. What do the regression weights of 355.6 and 261.8 mean? Is this kilometers per month?

23. The linear regression model has been removed (please see item 2 for response to concern 1.2)

Response to the concerns from referee 4 (Marc Souris)

Concern 4.1: Is the question posed by the authors well defined? The purpose of spatio-temporal analysis is dengue diffusion pattern definition, understanding how the disease spread from index cases. The geographical scale at which this kind of analysis need to be developed is fundamental, but not addressed in the paper. As we can see in the presented maps, the location of reported cases (very few) are limited to very few county and no cluster analysis, at this scale, is needed. Spatio-temporal analysis (like space-time cluster detection) is not equivalent to disease mapping by year and county.

24. The objective of the study is to explore the spatiotemporal distribution (including clustering) of reported imported and indigenous dengue fever cases during recent years in Guangdong province, so as to identify high-risk areas of the province and thereby help plan resource allocation for dengue interventions. The geographic scale for spatio-temporal analysis was at the county level, as described in the methods, which is the level at which the data were available, and provides sufficient resolution to explore sub-provincial spatial variation in dengue risk. It is common practice to present a descriptive spatiotemporal analysis by presenting maps of the crude data (the cumulative incidence maps) and then test hypotheses about spatial patterns (such as spatiotemporal clustering) using exploratory statistical analyses (such as spatial cluster detection algorithms).
More description about the objective of the study and are provided on page 6. The spatial scale at the county level is addressed on page 8.

**Concern 4.2:** As mentioned in conclusion (too late), Dengue is asymptomatic in 80% of cases; Dengue cases are only severe fever cases; DHF is a strong marker of dengue cases; but no DHF cases are mentioned. The dataset description is too short. No data on percentage of PCR or IgM detection. No description of public health system and report quality (especially, no description of difference between urban and rural areas). No data on geographical accuracy of cases. Immune status of the population is essential to understand the epidemiology of the disease.

25. Dengue infection presents with a wide range of clinical manifestations, including asymptomatic infection, dengue fever (DF), dengue haemorrhagic fever (DHF) and dengue shock syndrome (DSS). The purpose of this study is to explore the spatiotemporal distribution (including clustering) of reported imported and indigenous dengue fever cases, including DF, DHF and DSS. We agree that information on both asymptomatic infections and more severe dengue infection (DHF, DSS) would be key to quantify the disease burden, but this is outside the scope of this study. More details about dengue data and its quality are provided on page 7.

**Concern 4.3:** Incidence is very very low, and we think that the data quality is not enough to conduct such kind of analysis. Maps (integration by counties) shows only few location of infection, and no cluster analysis is needed to analyze this situation. Space-time cluster analysis means cluster analysis by time windows. We did not see any result of this kind.

26. We disagree with the reviewers comments regarding data quality (see our previous response regarding how the data are collected). More explanation of the appropriateness of cluster analysis in this study is provided on page 8. The time windows of the space-time cluster analysis are presented in the table 2 as 'Duration', and the results on the time window of cluster were added on page 11.

“Many studies have successfully adopted the SaTScan to detect the cluster of disease with very low incidence, and even rare disease [1, 2, 3, 4].”

**In response to Editor**

Concerns 1: In revising, the authors should particularly provide evidence that the quality of the used database is sufficient for the analysis conducted.

27. More discussion about the quality of the used database is provided on pages 7 and 8.

“In this study, most of the dengue cases were confirmed by the laboratory (92.56% for imported cases, and 81.84% for indigenous cases) and clinically diagnosed cases that do not have laboratory evidence to support the diagnosis were carefully reviewed by a group of professional physicians and epidemiologists. Since 2004, detailed individual
information on dengue fever cases was recorded, including the age, address, date of onset, clinical diagnosis, laboratory test result and travel history of each case.

Quality control has been an important component of infectious disease surveillance in China. At the end of each year, the Chinese Center for Disease Control and Prevention Centre (China CDC) actively surveys hospitals to identify the proportion of notifiable diseases that went unreported. The data have been widely utilized by governments as well as overseas organizations to assess public health programs in China.”

Concerns 2: Also, in line with one reviewer’s remark the presence of actual space-time cluster analysis should be illustrated, or alternatively, the title changed.

28. As your advice, the title was changed to “Spatiotemporal analysis of indigenous and imported dengue fever cases in Guangdong province, China”.

Concerns 3: Please clarify in your manuscript whether all data used for your study is openly available.

29. The data used in the paper are not openly available. Only the accumulative data on the dengue fever for whole country is openly available at DengueNet, World Health Organization (http://www.who.int/csr/disease/dengue/denguenet/en/index.html)

With kind regards.

Yours sincerely,

Weizhong Yang
Professor, Deputy Director-general, Chinese Center for Disease Control and Prevention
Postal address: No. 155, Changbai Road, Changping District, Beijing, 102206
E-mail address: ywz126@vip.sina.com
Telephone & Fax numbers: 86 10 5890 0311
References:


