**Reviewer’s report**

**Title:** Zika virus disease, microcephaly and Guillain-Barré syndrome in Colombia: epidemiological situation during 21 months of the Zika virus outbreak, 2015-2017

**Version:** 0  **Date:** 31 May 2017

**Reviewer:** T Papadopoulos

**Reviewer's report:**

Archives of Public Health

"Zika virus outbreak in Colombia: Situation of Microcephaly and Guillain Barre Syndrome, 2015-2017".

Research article

This study aims to present and analyse the epidemiological situation of ZVD, ZVD associated Microcephaly and ZVD associated GBS in Colombia during 19 months of outbreak. According to the objectives of the study the authors aimed to (1) update the epidemiological situation of ZVD in Colombia (2) evaluate the numerical relationship between ZVD indicators and GBS cases, and (3) comment on the geographic pattern of spread of ZVD through northern Colombia.

**General comments:**

The authors presented the overall situation during the 19 month outbreak in Colombia by plotting: (1) total number of cases of ZVD, (2) total pregnant women with ZVD (3) laboratory-confirmed pregnant women with ZVD, (4) suspected Zika virus-associated microcephaly cases and others abnormalities of central nervous system (5) laboratory-confirmed Zika virus-associated microcephaly cases and others abnormalities of central nervous system, and (6) GBS cases suspected due to Zika virus. They also presented the distribution of ZVD cases in Colombia by mapping across time using Kernel density estimator. Finally, they adopted kernel ridge regression and the Gaussian Kernel to predict the number of Guillain Barre cases given the number of cases of ZVD. The authors aimed to answer all the 3 objectives of the study, but they should probably focus only to one, as the structure is a bit confusing for the readers. Regarding the 3 objectives:
The authors used the public and freely available weekly epidemiological datasets from the National Health Institute of Colombia. To our knowledge, (and also as the authors state) these data have been already published in Institute's reports, but also in Pan-American Health Organization reports. According to these datasets the situation in Colombia has been already described in detail. The authors are using only descriptive epidemiology for answering the first objective of the study without any new analysis that could contribute interesting new results to the scientific community. In general presenting already published data should be avoided, with the exception of showing interesting results and conclusions after analysis.

The authors presented the distribution of ZVD cases by applying "the density kernel technique". In order to depict the intensity of ZIKV transmission in each map, the authors established five categories of transmission intensity. The authors do not explain how these categories were defined. It seems that different categories were used for each particular map and defining the "intensity". This was performed in a way that the reader can be confused as different scales are corresponding to different colors in each map in a way that the maps are not comparable. Analysis is missing and also the results and the conclusion of this, as they use only descriptive epidemiology.

The authors also aimed to predict the Guillain-Barre number of cases in Colombia given the ZVD cases. This estimation was performed by implementation of different models. They suggested that Gaussian kernel is the best choice for prediction as there is no proportional relationship between ZVD and GBS cases. I feel a little bit confused but, by reading Figure3 I cannot see any prediction of the Guillain-Barre number of cases in Colombia given the ZVD cases. To my knowledge this specific objective, to predict these cases, was not successful in the current study, but this is also an interesting result and should be clearly mentioned and discussed. Probably someone more familiar to these techniques can provide some help, however I would suggest that not all the readers of this journal have these specific skills.

Conclusions: I would suggest that the manuscript needs major revision to be considered for publication. The authors are presenting some updated valuable data concerning the ZIKA virus distribution in Colombia which is interesting of epidemiological view. However, they should clarify the objectives of the study, if they want to present an updated situation for ZVD in Colombia or suggest a model to predict GBS cases given ZVD cases. When the objectives are clear they should structure the manuscript (introduction, methods, results, discussion and conclusions) in a way that they will provide original research to the journal's readers and the scientific community.
Some specific comments

Abstract: The abstract is only a summary of some points of the manuscript. The numerical relationship (if any) between ZVD with microcephaly cases and GBS which is one of the three objectives is missing from the results.

Detailed comments are shown inside the manuscript.

Manuscript

Lines 82-85 "Because Brazil and Colombia have much bigger population than Honduras, Belize, Venezuela and El Salvador, the impact in terms of absolute number is more intense. As a consequence, the infection has had a huge impact on public health in Colombia and Latin American and its effects continue to be felt".

I am not sure for what the authors are trying to mention here. As a consequence to what? As a consequence to the bigger population? This suggests that in countries with rather smaller populations the impact in public health is not so huge. Please rephrase

Line 88 In the main document we haven't seen yet what is the period that the study refers to. I would suggest rephrase: So far April 8, 2017 and nearly 19 months later….

And replace also months rather than "monthly"

Line 89 I would suggest use "recorded" than "observed"

Line 100 during a 19 month period, from…

Line 101 we aim to evaluate? It is a little bit unclear

Line 107 (INS) is not needed. As you already defined an abbreviation.

Line 108 "used" rather than "analysed"
Line 108 replace "these" and "datasets" rather than "this" and "dataset" as more than one datasets were used

Line 110 "estimated" rather "subtracted"

Lines 114 to 117. "total and laboratory confirmed pregnant women"?

The variables were (1) total number of ZVD cases, (2) total pregnant women, (3) total pregnant women with a confirmation of ZVD, (4) Laboratory confirmed Zika-associated microcephaly and other CNS abnormalities cases and (5) GBS cases

Lines 119-129 These definitions were from the surveillance programmes. The definition of the clinical case…is referring to total ZIKA cases in the figure 1? I would also suggest that "clinically-confirmed case definition" does not apply to the current study. However, it would be useful to define the variables that were used only for your analysis, i.e. ZVD case, confirmed ZVD case in pregnant women, suspected ZIKV-induced microcephaly, confirmed ZIKV-induced microcephaly and finally suspected ZIKV- induced GBS syndrome. These variables are shown in the figure 1 and the reader should understand clearly the variables presented.

Lines 129-130 "ZIKV antibody testing was not done in Colombia due to high cross-reactivity with other endemic arboviruses" This is not to be included in methods, this is part of the discussion, I would suggest to omit this sentence.

Line 138 the goal "was" to, rather the goal "is" to …

Line 139 "identify the clusters of epidemic transmission"? please rephrase

Line 144-145 Why these are different for each map? It is rather confusing for the reader.

Line 151 total ZVD cases? ZVD among pregnant women?

Lines 153-163 parts of the discussion?

Line 174 "was" rather than "is"
Lines 175 "was" rather than "is"

Lines 177 "was" rather than "is"

Lines 183 "females" rather than "Females"

Lines 182-193 Are these results coming from your study? I suppose that these are existing, published already results. If these are already published, these should be part of the discussion.

Lines 195-199 Discussion points

Lines 207-208 "it is possible that people moving throughout the borders increased the mobility of the virus" Discussion points

Lines 214, 221, 223 what is a cluster? Please define in the Methods section

Line 218 "the incidence of ZIKV has been mainly established" please rephrase

Lines 225-226 "Figure 1 reveals that there is a relationship between GBS and ZVD cases in the country" How is this relationship occurs? Please explain.

Lines 232-243 Some of these are part of the discussion.

Line 235 use "cannot" rather than "can not"

Line 249 "and also to evaluate the numerical relationship between ZVD indicators and GBS cases" This is one of the objectives also. What about the prediction model?

Lines 251-255 reference?

Lines 256-257 "however, it can be expected that it will be higher than dengue because of arthritis-related and other disabilities" I would say it is irrelevant and there is no reference supporting this observation. I would omit this sentence.
Line 260 please use "described" than "suggested"

Line 260 "By 8 April" rather "In 8 April".

Line 308 i.e. rather than "ie."

Line 334 delete "the"

I haven't commented further the discussion as more work should be done to the methods and results, subsequently the discussion should be structured according to the results.

Figure 1: Why the authors are presenting 2 tables one per week and one per month? Why this is important according to the objectives of the study?

Use "suspected" microcephaly cases rather than "suspicious".

Figure 2: Why you used different scales for defining case density for each map. It is rather confusing as the maps are not comparable. What are the clusters?

The maps are also different according to time periods. One is for 20 days of October, 6 are for quarters of years and one for the whole period, however April (11 days) are missing.

Figure 3: On a first view prediction seems not to work, is this correct?

**Level of interest**
Please indicate how interesting you found the manuscript:

An article of importance in its field

**Quality of written English**
Please indicate the quality of language in the manuscript:

Needs some language corrections before being published
Declaration of competing interests

Please complete a declaration of competing interests, considering the following questions:

1. Have you in the past five years received reimbursements, fees, funding, or salary from an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

2. Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

3. Do you hold or are you currently applying for any patents relating to the content of the manuscript?

4. Have you received reimbursements, fees, funding, or salary from an organization that holds or has applied for patents relating to the content of the manuscript?

5. Do you have any other financial competing interests?

6. Do you have any non-financial competing interests in relation to this paper?

If you can answer no to all of the above, write 'I declare that I have no competing interests' below. If your reply is yes to any, please give details below.

I declare that I have no competing interests

I agree to the open peer review policy of the journal. I understand that my name will be included on my report to the authors and, if the manuscript is accepted for publication, my named report including any attachments I upload will be posted on the website along with the authors' responses. I agree for my report to be made available under an Open Access Creative Commons CC-BY license (http://creativecommons.org/licenses/by/4.0/). I understand that any comments which I do not wish to be included in my named report can be included as confidential comments to the editors, which will not be published.

I agree to the open peer review policy of the journal