Reviewer’s report

Title: Discovery of under immunized spatial clusters using network scan statistics

Version: 0 Date: 25 May 2018

Reviewer: Renato Assunção

Reviewer's report:

I approached this paper with great interest. The problem is relevant and the authors creatively use two sources of data to detect the under immunized geographical clusters in two large regions in US. After detecting the clusters with a different methodology than the usual scan statistics, they proceed to explain them with a logistic regression with some areas classified as under-immune (those falling inside the clusters) against the other ones. Although I think the paper has merits, I find some problems of concern.

One problem is the fact that the authors ignore the serious over-fitting suffered by the likelihood maximization when the cluster is unconstrained. When the cluster shape is arbitrary, there is a tendency to find clusters looking like octopus, with long arms. One could think that this is simply how the true clusters are and that finding them is an algorithm merit over the simpler circular scan statistics. However, unfortunately, this is not so. Simulate a situation where there is one single circular cluster with relative risk higher than the rest. Run the unconstrained algorithm. You will find statistically significant clusters that are not circular but shows long and thin arms spreading from the cluster center, much like those in the authors' paper.

There is a large published literature on the detection of spatial clusters with arbitrary shape that are not cited by the authors. I copy from Costa, Assuncao, and Kulldorff (2012) (see complete ref below): "Computational heuristics for irregular cluster detection using a scan-based algorithm were proposed by Duczmal and Assuncao (2004) using simulated annealing. Patil and Taillie (2004) introduced cluster detection using tessellation techniques, while Tango and Takahashi (2005) proposed an exhaustive search within pre-sized circular clusters. Assuncao et al. (2006) proposed a growth technique based on likelihood maximization in a graph structure. As a result of the cluster detection procedure, these methods are likely to generate the octopus effect (Duczmal and Assuncao, 2004), which indicates oversized clusters that are oddly shaped with multiple narrow branches, linking areas with excess cases. The problem lies in the use of the unconstrained likelihood maximization to provide the cluster estimate. Duczmal et al. (2006b) proposed the addition of a non-compactness penalty function to the log-likelihood. Different penalty levels can provide different significant clusters with distinct shapes. Duczmal et al. (2006a) use multi-objective optimization to decide among different cluster estimates.” Other more recent references such as Costa and Kulldorff (2014) are relevant here.
Therefore, the authors need to control somehow for this over-fitting with some kind of regularization. Otherwise, most of their oddly shaped detected clusters can not be trusted.

A second important issue with the paper is that it is not entirely clear how was calculated the variable unimm(b), which is the estimated number of under-immunized children in block b. Let v be a child and S(v) her school. Then, \( P(v) = \) probability of child v getting both doses of MMR, which is simply the immunization rate in the school S(v). Now, unimm(b) is the sum of P(v) for all v in b? Did you then round the final number? Did you simulate a Bernoulli random variable for each child with P(v) as the success probability and repeat this many times to create pseudo datasets? What exactly was done? My guess is that it was the first option. If so, what we have as data is simply the spread of each school body around its neighboring blocks. It look to me that whatever cluster we find in the end is ultimately a cluster of schools.

References:


Are the methods appropriate and well described?
If not, please specify what is required in your comments to the authors.

Does the work include the necessary controls?
If not, please specify which controls are required in your comments to the authors.

Are the conclusions drawn adequately supported by the data shown?
If not, please explain in your comments to the authors.

Are you able to assess any statistics in the manuscript or would you recommend an additional statistical review?
If an additional statistical review is recommended, please specify what aspects require further assessment in your comments to the editors.
Quality of written English
Please indicate the quality of language in the manuscript:

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.

Statement on potential review bias
Please complete a statement on potential review bias, considering the following questions:

1. Did you co-author any publication with an author of this manuscript in the last 5 years?

2. Are you currently or recently affiliated at the same institution as an author of this manuscript?

If you can answer no to all of the above, write 'I declare that I did not publish with these authors in the last 5 years and also meet the affiliation criteria”. If your reply is yes to any, please give details below.

I declare that I did not publish with these authors in the last 5 years and also meet the affiliation criteria.
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.