Author’s response to reviews

Title: Spatial-temporal detection of risk factors for bacillary dysentery in Beijing, Tianjin and Hebei, China

Authors:

chengdong xu (xucd@lreis.ac.cn)
yuanyuan li (lyy@lreis.ac.cn)
jingfeng wang (wangjf@lreis.ac.cn)
gexin xiao (biocomputer@126.com)

Version: 1 Date: 19 Jul 2017

Author’s response to reviews:

July 19, 2017

Dear Editors of BMC Public Health,

Thank you very much for your time and reviewer’s comments. We have thoroughly revised the manuscript according to the comments in your decision letter on June 22, 2017.

We now resubmit our fully revised paper entitled “Spatial-temporal detection of risk factors for bacillary dysentery in Beijing, Tianjin and Hebei, China” that we would like to be considered for publication in BMC Public Health. The author responses’ to reviews are followed in this letter.

Thank you very much for your review. We look forward to hearing from you in due course.
With best regards

Chengdong Xu

State Key Laboratory of Resources and Environmental Information System,
Institute of Geographic Sciences and Natural Resources Research,
Chinese Academy of Sciences.
Beijing 100101, PR China
Tel: +86 10 64889055
Email: xucd@Lreis.ac.cn

Authors’ Responses to Reviewers

Lawrence Kazembe (Reviewer 1): The present manuscript considers a spatial and temporal analysis of risk factors of bacillary dysentery in Beijing-Tianjin-Hebei. The authors apply a GeoDetector methods.

The paper in my view is concerned with an important disease affecting many urban areas in Asia. This paper seems to be an extension of a previous work done by Xiao et al (2014), however, the review is silent or offer little comparison to what was done earlier- and what the present manuscript does or differs from or adds to it. I suggest that there be discussion to this effect.

R: Thank you for your comments. The setting for the previous study (Xiao et al, 2014) was only a fifth the area covered by the current study, and the former focused on the most developed urban area only, and was not included a wide rural area. In the previous study, the area at risk of
bacillary dysentery was identified using SaTScan method, and the linear effects in relation to the risk factors were analyzed. While the area at risk of this disease in the current study was identified using GeoDetector method, and the nonlinear effects pertaining to the risk factors were evaluated. A relevant discussion of the comparison has been added to the manuscript. Please see lines 290-297 in “Track Changes” version.

Other issues that need to be considered in the revision are:

1. Abstract: There is no objective in the background.

R: The study objective was to analyze the spatial-temporal pattern and to determine any contributory risk factors on the bacillary dysentery. These objectives have been added to the “Abstract”. Please see lines 17-18 in “Track Changes” version.

2. Abstract: Sentence on line 22-23 is incomplete.

R: This sentence has now been revised and reads as follows: “GeoDetector method was used to determine the impact of potential risk factors, and to identify regions and seasons at high risk of the disease”.

3. line 73: where GeoDetector is mentions please provide a reference and also explain briefly what it does. Or do mention that "this will be expanded in the methods section"

R: Revised, please see lines 78-79 in “Track Changes” version.

4. line 141: insert a reference for the GeoDetector.

R: A reference has been inserted, please see lines 172 in “Track Changes” version.
5. line 147: the R need to be explained or defined more explicitly. R is said to move in space and time yet this is not clear at this point.

R: R has been explained, please see line 180 in “Track Changes” version.

6. line 163: to aid better understanding of the results, it is important that as you mention the counties, please include the code- which has been listed in Figures 4 and 5. Or alternatively just give full names of counties in Figures 4 and 5.

R: Revised, please see Figures 4 and 5.

7. line 165: write "value is about 29.75...".

R: Revised.

8. The issues of non-linearity need to be explained in Table 2, as well as in the methods.

R: This has been explained. Please see lines 183-186 in “Track Changes” version.

9. Figure 4: please provide the X and Y -axis labels

R: Revised.

10. Figure 5: please provide the X and Y -axis labels.

R: Revised.
Carolina Perez-Heydrich (Reviewer 2): Authors presented an application of the GeoDetector method to classify the degree to which certain exposure variables help explain the outcome of interest. Specifically, in this paper, authors evaluated risk factors associated with bacillary dysentery. Overall, the paper lacked clarity in describing (1) the merit of this approach over other commonly used spatial regression methods, (2) spatial attributes of the data used in the application, and (3) how parameters for the GeoDetector model were obtained. More detailed comments are provided below for consideration.

R: A comparison of GeoDetector and other commonly used regression methods has been included in the revised manuscript. Please see lines 358-373 in “Track Changes” version. In addition, the applicability of the spatial data attributes, as well as the preparation of the parameters and the calculation process, have been clarified in the “Method” section. Please see lines 167-172 in “Track Changes” version.

*Line 88: When describing variables associated with "Potential risk factors," authors state that data were obtained from the 2012 Statistical Yearbook. Authors should consider elaborating more on this statement. Namely, authors should mention the spatial resolution of the data and discuss how variables were spatially linked, since it did not seem like all variables were available at the same spatial scales (though this was unclear).

R: In the study, all the variables were linked by the name of each county on a map. The data on the socio-economic characteristics of the study subjects and the disease were aggregated for each county. This has now been explained. Please see lines 146-148 in “Track Changes” version.

*Line 100: When describing meteorological variables, I think the term "wind" should be changed to "wind speed" to better describe the variable (if indeed this is what authors meant by "wind").

R: Revised to “wind speed”.

*Additionally, authors should also justify how the variables chosen as proxies of certain risk factors are adequate proxies for these. I'm unconvinced of this as the manuscript currently stands.

R: A pathogenic bacterial environment, conducive to bacterial breeding and transmission, is one of the most significant and influential factors for a high incidence of bacillary dysentery. Variables were used to explain the risk of disease. However, they could not be measured directly, so corresponding proxy variables were selected, as has occurred in previous studies. Those corresponding references have been added in the revised manuscript, please see lines 109-110 in “Track Changes” version. Usually, it is not possible to collect the data as direct variables
affecting disease transmission in an epidemiological study, and corresponding proxy variables are used for this reason. Regrettably, this does introduce a degree of uncertainty when interpreting the results. This issue has been covered in the “Discussion” section in lines 387-389 in “Track Changes” version.

*Authors may also consider using population age structure as a potential risk factor in their analysis.

R: Thank you for this suggestion. The age structure of the population was included as a potential risk factor for bacillary dysentery in the revised manuscript as we determined that certain age groups have a strong association with the disease. The implication of this is that greater attention should be paid to vulnerable groups identified as being at high risk of bacillary dysentery.

Please see lines 221-225, Figure 3, Table 2, and lines 341-343 in the “Discussion” section in “Track Changes” version.

*Line 110: Authors should define the terms "primary," "secondary," and "tertiary" industry.

R: Definitions of the relevant terms have been added to the revised manuscript. Please see lines 113-118 in “Track Changes” version.

*Line 112: Authors should describe the data source used to describe medical and technical personnel.

R: The socio-economic data for the medical and technical personnel in each county were obtained from the 2012 Statistical Yearbook. This information has been added to the revised manuscript. Please see lines 97-98 in “Track Changes” version.

*Lines 132-136 are extremely poorly written, and authors should rewrite this section entirely to clarify their points.

R: The relevant lines were revised, please see lines 152-154 in “Track Changes” version.

*Line 142: Authors should define how the size of the target spatio-temporal field was determined for the equation, along with describe how σ was estimated. Additionally, authors should explicitly define the equation for q(x_1∩x_2).
R: x1 ∩ x2 can be implemented by using the overlay operation for the two layers, x1 and x2, in Geographic Information System (GIS) software, e.g., ArcGIS®. Please see lines 193-195 in “Track Changes” version.

*Line 148: It is unclear what the risk factors are stratified by. Authors should clarify this.

R: This has been clarified, please see lines 179-180 in “Track Changes” version.

*Overall, it is unclear under what scenarios this approach would be better to use over an adjusted odds ratio. As it is currently depicted, the GeoDetector method seems to be a form of a coefficient of determination, which can be estimated via regression methods. Authors should discuss more thoroughly why their proposed approach is better than other existing approaches.

R: This has been discussed, please see lines 358-373 in “Track Changes” version.

*Line 153: Remove "(univariate)" next to "Weaken" to keep it consistent with the "Enhance" terminology stated above.

R: Done.

*Line 157: Citation #14 is not relevant to cite here.

R: The reference has been deleted.

*Line 160: Authors mention incidence estimates, but never defined how these were obtained in the Methods section. This description should be added somewhere in the Methods section.

R: Disease incidence was calculated by determining the ratio between the number of cases with bacillary dysentery and the population size of a given county or age group. This has been added in the Methods section.

*Line 162: It is more appropriate to state the incidence rate in the form of 365 per 100,000 population. This also applies hereafter.

R: Revised.
*Line 167: Authors should clarify their use of the term "cluster regions." Were these cluster regions formally defined using some statistical method (e.g. Kulldorff's scan statistic, LISA statistics)

R: The term “cluster regions” simply refers to regions with a high incidence of disease. This has been clarified within the manuscript. Please see lines 209-210 in “Track Changes” version.

*Line 168: The sentence beginning with "However" needs to be rewritten. The message being communicated here is not clear, specifically as it relates to the role that spatial distribution plays in this observation. This seems to be a comment more related to the age distribution of cases, not their spatial distribution.

R: This has been rewritten, please see lines 209-212 in “Track Changes” version.

*Line 174: It is unclear whether q values represent overall or county-specific contributions to the outcome. This needs to be clarified here and in the Methods section when the equation is introduced.

R: This has been clarified, please see line 164 in “Track Changes” version.

*Line 176: Since authors note that GeoDetector does not assume normality of errors (and consider this to be a strength of the approach), why do authors report on Pearson correlation coefficients here? This doesn't seem appropriate.

R: Pearson correlation coefficient is usually used to depict a linear association between two variables. In the current study, we only used the information needed to portray the general direction of the association, e.g., to determine a positive or negative relationship between two variables. We did not make use of the correlation coefficient values. This has now been clarified in the manuscript. Please see lines 216-217 in “Track Changes” version.

*Lines 179-80: The finding stated here could be explained by the lower population density of rural areas. The two variables seem to be addressing the same mechanism. With a regression approach, one would be able to determine the relative contribution of population density after accounting for urbanity (or vice versa). This univariate nature of associations addressed via GeoDetector appears to be a significant weakness of the approach. An approach that would consider partial correlations would seemingly be more appropriate to address epidemiologic questions.
R: This has been clarified, please see lines 374-381 in “Track Changes” version.

*Line 210: Similar to the above issue, high wind speed is directly related to lower temperature. Thus, the mechanism driving the association between high wind speed and dysentery incidence is more likely to be via temperature (based on the biology of these pathogens). It is unclear how wind speed would affect the bacterial environment in a way that's not already captured by temperature.

R: This has been clarified, please see lines 374-381 in “Track Changes” version.

*Line 217: Elaborate on the interpretation of the determinant power of the interaction between temperature and wind speed. This should be explained more clearly with respect to epidemiologic context.

R: The interpretation has been added in the revised manuscript, please see lines 332-338 in “Track Changes” version.