PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Spatial clustering of drug-resistant tuberculosis in Hunan Province, China: an ecological study
AUTHORS	Alene, Kefyalew Addis; Xu, Zuhui; Bai, Liqiong; Yi, Hengzhong; Tan, Yunhong; Gray, Darren; Viney, Kerri; Clements, Archie

VERSION 1 – REVIEW

REVIEWER	Professor Virasakdi Chongsuvivatwong Prince of Songkla University
REVIEW RETURNED	24-Oct-2020

GENERAL COMMENTS	This paper addressed spatial distribution of drug-resistant TB in a province in China based on the surveillance data over 7-year period. As expected and reported in many other diseases, there are hot spots and cold spots. The authors recommended targeting interventinon in highest risk area baesd on strong unstructured random effects. While the title and the main texts emphasized on spatial distribution, the abstract has only two lines of results being so. I don't get a strong favour that this paper would be on TB spaital analysis. The research methods are fine but there is no novelty or high public health relevance in the research questions. All diseases have heterogeneity in spatial distribution. Some areas must have high and some must have low incidence. Most of this information (in supplement information about DR-TB distribution) is probably known to the program officers. What would this research add to local public health for their control activities and to international readers for their better understanding of TB spatial distribution. The main weakness was in the nature of the data. Mixing all drug-resistant TB together to increase statistical power had a strong penalty in examination of TB with different degrees of public health concern. XDR-TB for example, is much more serious to Publich Health security than TB with single drug resistance. DR-TB reports increased with time in the data set. The authors however did not analyse whether the problem increased in a homogenous or heterogeneous spatial pattern. How did the hot and cold spot move over time? Comparing these with those of drug-sensitive TB notification will enable the authors to see whether the low notification rate at the early period was likely due to poor coverage of the services or due to pure increase in drug resistant proportions. Local researchers should give discussion on this.

	The conclusion that the program should focus on relatively high incidence of DR-TB is based purely on variance of the unstructured random effects, which is difficult to translate into action because 'random effects' denotes that the small areas are randomly arranged. Large number of these random unit makes it difficult to pin-point where the problem actually is. The supplement information is easier to interprete. It reveals that the difference among prefectures are not very wide. The lowest prefecture, Shaoyang has nearly 3 DR-TB per 100,000 population, a magnitude that should be of concern. Targeted program is therefore not well justified. Both the simple information and the sophisticated regression ignored population size of the localities. Like many public heatlh problem, population size may have stronger population attributable risk than the high incidence area with low population. The analysis should take population attributable fraction into account before drawing the conclusion. Additonally, the analysis gave useful information on DR-TB treatment outcome, which is well known to be poor in China (due to financial barriers). The data showed no significant relationship between locality socio-economic status and DR-TB incidence. The authors should further examine whether the relationship is true foo DR-TB treatment outcome. Finally, either intellectual input or recognition of importance of local researchers in this paper is relatively low for such an important Public Health problems. Major authors are all western people. The paper does not show enough efforts on local capacity building. It is doubtful what will be left behind in the locality after this China funded project is finished and the foreign researchers have left. In summary, I have checked all "Yes" in the assessment indicating that the paper. If this paper is sent to a low imact journal, based on my above narrative comments, I would vote 'accept with major revision". I believe that the paper can be acceptd. However, for a journal with impactor factor above 5 like the currentl one, the quality of the data, the implication for international public health and the level of the novelty of the paper are all inadequate. My current vote is therefore "reject".

REVIEWER	Alex Cook National University of Singapore
REVIEW RETURNED	06-Nov-2020

GENERAL COMMENTS	I have very few comments on this geospatial analysis of clustering of drug resistant TB in Hunan, China. Using the authors' line numbering: 44 (and later): there is no point presenting RRs or other effect sizes

	without the units. A reader isn't going to know what this means, and potentially interpret it to mean that urban communities have a risk that is 2% higher than rural communities, which is not correct. (Also, check the CI, which is grossly asymmetric and may be in error.) 146 (and later): the notation is quite sloppy. I like that you indicate the districts, but writing Yi makes it look like Y times i. Better to use subscripts, and in the text to write Y_i rather than just Y. 158: Winbugs is so old, isn't there some newer software that can do this! 170: again, sloppy notation. What is j here? Why did Y have a superscript in the equation but now has (two) subscript(s)? 172: more sloppy notation. The log function shouldn't be capitalised. \beta k looks like the product of beta and k, and in the text you write beta_k so can you be consistent in the equation? Surely there should be a sum over k. 174: please provide distributions for U_i and V_i. 181: were covariates standardised? If not, are you sure this precision is appropriate for all covariates? 185: number of iterations and length of burn in. 210: can give the number as well as the proportion? 234: again, units please. 307: you may prefer to remove the note about the grant number not existing. T1: might be nice to have a comparison with DS TB in this table? T2: units. If GDP is (USD per person per month) then it would be better to transform it to (1000USD per person per month) or similar to make it interpretable, and the RRs and Cis respectively. F1: Chin*a* F3: Any explanation why Zixing is burning red in F2 but not in F3? The three maps in the main manuscript could feasibly be merged into one figure to allow comparison easier (perhaps trimmed a little to remove some of the neighbouring provinces). (Similarly for the supplementary maps)

VERSION 1 – AUTHOR RESPONSE

Reviewer 1: Reviewer Name: Professor Virasakdi Chongsuvivatwong;
Institution and Country: Prince of Songkla University; Please state any competing interests or state 'None declared': None declared

Comments to the Author
This paper addressed spatial distribution of drug-resistant TB in a province in China based on the

surveillance data over 7-year period. As expected and reported in many other diseases, there are hot spots and cold spots. The authors recommended targeting interventinon in highest risk area baesd on strong unstructured random effects.
Response: we thank you the reviewer for the thoughtful comments.
While the title and the main texts emphasized on spatial distribution, the abstract has only two lines of results being so. I don't get a strong favour that this paper would be on TB spaital analysis.
Response: we have now provided additional information on the result sections of the abstract. However, we are not able provide detailed information in the abstract section due to the journal requirement in word limits.
The research methods are fine but there is no novelty or high public health relevance in the research questions. All diseases have heterogeneity in spatial distribution. Some areas must have high and some must have low incidence. Most of this information (in supplement information about DR-TB distribution) is probably known to the program officers. What would this research add to local public health for their control activities and to international readers for their better understanding of TB spatial distribution.
Response: TB continues to have a significant impact on global health. The emergence of DR-TB poses additional health threats at global and national levels. Knowledge of the geographical distribution of the disease and factors that determine where and when DR-TB occur is crucial for local public health and policy makers to implement targeted interventions for the control and prevention of TB. This might be particularly important in resource-constrained settings and high TB burden settings such as Hunan Province. We therefore believe that this study was important to understand spatial distribution of DR-TB and identify areas of potentially high risk (i.e., hot spot areas) across small geographical areas, an important information for both international readers and local public health experts. There are actually relatively few publications exploring the spatial patterns of tuberculosis, particularly in China.
The main weakness was in the nature of the data.
Response: We agree that using routinely collected secondary data has its own limitation. This limitation has been mentioned in the discussion sections of the manuscript. However, this study has also generated evidence using existing TB surveillance data. Such types of study can maximize the usefulness of TB surveillance data for decision making.
Mixing all drug-resistant TB together to increase statistical power had a strong penalty in examination of TB with different degrees of public health concern. XDR-TB for example, is much more serious to Publich Health security than TB with single drug resistance.
Response: We agree that a separate analysis for XDR-TB would provide additional information. However, as we had few cases of XDR-TB case (n=89), we were not able to fit a separate model.
DR-TB reports increased with time in the data set. The authors however did not analyse whether the problem increased in a homogenous or heterogeneous spatial pattern. How did the hot and cold spot move over time? Comparing these with those of drug-sensitive TB notification will enable the authors to see whether the low notification rate at the early period was likely due to poor coverage of the services or due to pure increase in drug resistant proportions. Local researchers should give discussion on this.
Response: Thank you for these comments. Spatial maps showing the distribution of DR-TB for each year have been now produced and presented as supplementary information. See Figure S4. Spatial clustering of drug-resistant tuberculosis in Hunan Province from 2012 to 2018, based on the Getis-Ord Gi* statistic. To address the reviewer’s comment, we have also presented additional figures showing the prevalence of drug-resistant testing coverage over time in Hunan province (Figure S1). Local researchers who are also the authors of this paper have provided comments on this finding.
The conclusion that the program should focus on relatively high incidence of DR-TB is based purely on variance of the unstructured random effects, which is difficult to translate into action because 'random effects' denotes that the small areas are randomly arranged. Large number of these random unit makes it difficult to pin-point where the problem actually is.
Response: The spatial clustering analysis using Getis-Ord Gi* statistic identified hot-spot counties

where DR-TB were clustered - targeted interventions might be important to reduce the burden of the disease in these hot-spot counties. The model with the unstructured random effects represents residual variation after accounting for the model covariates – the fact that the unstructured random effects model provided the best fit does not mean that there is not clustering of the disease – it just means that the model covariates were able to explain part of the spatial structure in the data. Therefore, the random effects model provides additional information about the underlying drivers of the spatial pattern of TB in the province – information which is useful for local policymakers. This information is now included on page 6 line 154-156.

The supplement information is easier to interprete. It reveals that the difference among prefectures are not very wide. The lowest prefecture, Shaoyang has nearly 3 DR-TB per 100,000 population, a magnitude that should be of concern. Targeted program is therefore not well justified.

Response: Thank you for this comment. A substantial variation of DR-TB was not observed between prefectures but within prefectures at a county level. Even in prefectures with a relatively low or moderate incidence of DR-TB, we identified considerable geographical variability in DR-TB incidence at county level that far exceeded their respective prefectures averages. Identifying clustering of DR-TB at small geographical areas where policy implementation and administrative decision-making typically occur are important for health departments to develop tailored interventions.

Both the simple information and the sophisticated regression ignored population size of the localities. Like many public heatlh problem, population size may have stronger population attributable risk than the high incidence area with low population. The analysis should take population attributable fraction into account before drawing the conclusion.

Response: Thank you for this comment. We have included population size in two of our analyses. For example, the overall crude incidence rate of DR-TB was calculated by taking the total number of DR-TB cases reported during the study period as the numerator and the mid-point total population during the same period as the denominator. The standardized morbidity ratio (SMR) was then calculated for each district using the formula: Yi = [Oi/Ei]; where Y is the SMR in district i, O is the observed number of DR-TB cases in the district and E is the expected number of MDR-TB cases in the district across the study period. The expected number of DR-TB cases for each district was calculated by multiplying the mid-point population of each district by the overall crude DR-TB incidence rate for the study area and period. We have also included the proportion of males in a county and percentage of urban residence in the counties in our models. However, calculating population attributable fraction was not the focus of this study.

Additonally, the analysis gave useful information on DR-TB treatment outcome, which is well known to be poor in China (due to financial barriers). The data showed no significant relationship between locality socio-economic status and DR-TB incidence. The authors should further examine whether the relationship is true foo DR-TB treatment outcome.

Response: Thank you for this comment. We agree that this is an important area the needs further investigation. However, the relationship between socio-economic status and DR-TB treatment outcomes has not been the focus of this study.

Finally, either intellectual input or recognition of importance of local researchers in this paper is relatively low for such an important Public Health problems. Major authors are all western people. The paper does not show enough efforts on local capacity building. It is doubtful what will be left behind in the locality after this China funded project is finished and the foreign researchers have left.

Response: Thank you for raising this concern, although we were rather surprised to receive this as the study was a true collaboration between Chinese and Australian researchers. Half of the authors of this paper (4/8) were from Hunan Province, where this study was conducted. These authors contributed significantly to this work. They conceived and designed the study (ZX), supervised the data translation process (HY, LB, YT and ZX), and improved the drafted manuscript (ZX). They were also involved with critical revision of the manuscript and approved the final version of the manuscript (HY, LB, YT and ZX).

In summary, I have checked all "Yes" in the assessment indicating that the paper. If this paper is sent to a low imact journal, based on my above narrative comments, I would vote 'accept with major

revision". I believe that the paper can be acceptd. However, for a journal with impactor factor above 5 like the currentl one, the quality of the data, the implication for international public health and the level of the novelty of the paper are all inadequate. My current vote is therefore "reject".
Response: We are grateful for the thoughtful comments of the reviewer. We have carefully reviewed the comments and have revised the manuscript accordingly. We hope the revised version is now suitable for publication.

Reviewer 2: Reviewer Name: Alex Cook
Institution and Country: National University of Singapore Please state any competing interests or state 'None declared': no
Comments to the Author
I have very few comments on this geospatial analysis of clustering of drug resistant TB in Hunan, China. Using the authors' line numbering:
44 (and later): there is no point presenting RRs or other effect sizes without the units. A reader isn't going to know what this means, and potentially interpret it to mean that urban communities have a risk that is 2% higher than rural communities, which is not correct. (Also, check the CI, which is grossly asymmetric and may be in error.)
Response: Thank you for these comments. We have corrected this in the revised version of the manumit on page 2 line 45.
146 (and later): the notation is quite sloppy. I like that you indicate the districts, but writing Yi makes it look like Y times i. Better to use subscripts, and in the text to write Y_i rather than just Y.
Response: Thank you for these suggestions. These are now corrected in the revised version of the manuscript.
158: Winbugs is so old, isn't there some newer software that can do this!
Response: We have also OpenBUGS and R but for this analysis, we just used WinBUGS as it is more than capable of implementing the analytical requirements of this study.
170: again, sloppy notation. What is j here? Why did Y have a superscript in the equation but now has (two) subscript(s)?
Response: Thank you for this comment. It is now corrected in the revised version of the manuscript on page 7 line 170.
172: more sloppy notation. The log function shouldn't be capitalised. \beta k looks like the product of beta and k, and in the text you write beta_k so can you be consistent in the equation? Surely there should be a sum over k.
Response: Thank you for this comment. It is also now corrected in the revised version of the manuscript on page 7 line 172.
174: please provide distributions for U_i and V_i.
Response: we have provided the distribution for U_i and V_i in the revised version of the manuscript on page 7 line 180.
181: were covariates standardised? If not, are you sure this precision is appropriate for all covariates?
Response: Since the covariates have different units and scales of measurement, we standardized all the variables to a z-scale on the basis of their mean and SD,but converted the relative risks back to the original scale for presentation of the results.
185: number of iterations and length of burn in.
Response: The models were run for 100,000 iterations and convergence was successfully achieved after 80,000 iterations for each of the models.
210: can give the number as well as the proportion?
Response: We have now included the number in the revised version of the manuscript on page 9 line 211.
234: again, units please.
Response: We have now included the units in the revised version of the manuscript on page 10 line 239.
307: you may prefer to remove the note about the grant number not existing.

Response: Thank you for this comment. However, this is the requirement of the journal: “Please ensure to provide an award/grant number for your funder/s in the main document file and in ScholarOne. If there is none, please indicate 'Award/Grant number is not applicable' both in main document and in ScholarOne.”
T1: might be nice to have a comparison with DS TB in this table?
Response: Thank you for raising this important comment. We agree that DS-TB data might give additional information. However, we do not have these data to include in Table 1.
T2: units. If GDP is (USD per person per month) then it would be better to transform it to (1000USD per person per month) or similar to make it interpretable, and the RRs and Cis respectively.
Response: Thank you for this comment. We have now included additional information as a footnote in Table 2 to make it clear for readers.
F1: Chin*a*
Response: Thank you for this comment. We have now corrected this in the revised version of the manuscript on page 17 line 391.
F3: Any explanation why Zixing is burning red in F2 but not in F3? The three maps in the main manuscript could feasibly be merged into one figure to allow comparison easier (perhaps trimmed a little to remove some of the neighbouring provinces). (Similarly for the supplementary maps)
Response: Thank you for this comment. Figure 2 presents spatial distribution of DR-TB standardized morbidity ratios (SMR) and Figure 3 presents spatial clustering of DR-TB incidence using the Getis-Ord Gi* statistic. Since we used different methods for different purpose the color on the maps are also different. The Getis-ord Gi* statistic was used to determine clustering patterns of DR-TB to identify cold-spots and hot-spots. We have included the neighboring province to assist with navigation of the maps for readers and local health professionals.

VERSION 2 – REVIEW

REVIEWER	Professor Virasakdi Chongsuvivatwong Prince of Songkla University
REVIEW RETURNED	17-Jan-2021

GENERAL COMMENTS	Thank the authors for careful revision of the paper. My all points are basically well responded. The only concern that blocked my from accepting this paper is the weakness of the data. We know that there are plenty of drug resistant TB. They are different in level of public health important and need different approaches but were treated as single group in this paper. I am sorry to say that the paper should be transferred to a lower impact factor journal.

REVIEWER	Alex Cook National University of Singapore, Singapore
REVIEW RETURNED	26-Dec-2020

GENERAL COMMENTS	Many of the points I raised in my first review have not been addressed. Please address these. The notation is still sloppy despite the authors’ saying that they’ve revised it. If Y_i [L171] is the observed number of DRTB cases then what is Y^I [L170]? If \beta_k [L174] is the coefficient then what is the product of \beta and k [L173]? If the product of U and i [L175] is an unstructured random effect then what is U^i [L173]? If V_i [L175] is a spatially structured random effect then what is V^i [L173]? If the covariate is X_k [L175] then what is X^k [L173]? If the mean of the Poisson distribution for Y_i is \mu_i [L172] then what are the \mu^i [L170 and L173] terms? If E_i is the expected number of DRTB

	cases [L174] then what is E^i [L173]? There are hardly any places in the methods where the notation is actually internally consistent, which is especially disappointing given it's already been highlighted as a problem and the errors are literally one to two lines apart. If when fixing this the authors could add subscripts to L146 and L147 O and E respectively that would also be desirous. Terminology: 'the posterior parameter' [L185] is not a thing. Please replace by 'the posterior distribution of the parameters' I apologise that I could not identify the point in the text where the standardisation of all variables was described, which is mentioned in the rebuttal. Table 2 is still lacking quite a few units. Monthly average temperature for instance.

VERSION 2 – AUTHOR RESPONSE

Reviewer 2: Dr. Alex R Cook, National University Singapore Saw Swee Hock School of Public Health

Comments to the Author:
Many of the points I raised in my first review have not been addressed. Please address these.
The notation is still sloppy despite the authors' saying that they've revised it. If Y_i [L171] is the observed number of DRTB cases then what is Y^I [L170]? If \beta_k [L174] is the coefficient then what is the product of \beta and k [L173]? If the product of U and i [L175] is an unstructured random effect then what is U^i [L173]? If V_i [L175] is a spatially structured random effect then what is V^i [L173]? If the covariate is X_k [L175] then what is X^k [L173]? If the mean of the Poisson distribution for Y_i is \mu_i [L172] then what are the \mu^i [L170 and L173] terms? If E_i is the expected number of DRTB cases [L174] then what is E^i [L173]? There are hardly any places in the methods where the notation is actually internally consistent, which is especially disappointing given it's already been highlighted as a problem and the errors are literally one to two lines apart. If when fixing this the authors could add subscripts to L146 and L147 O and E respectively that would also be desirous.
Response: We have double-checked the statistical notation and it is correct. We are confused by some of these comments – for example, there is no product of U and i. The notation in the comments is different to the notation in the submitted draft. We did add one sentence to clarify that k refers to the number of covariates on page 7 line 175.

Terminology: 'the posterior parameter' [L185] is not a thing. Please replace by 'the posterior distribution of the parameters'
Response: This is now corrected on page 8 line 185

I apologise that I could not identify the point in the text where the standardisation of all variables was described, which is mentioned in the rebuttal.
Response: We have now included this in the revised version of the manuscript on page 8 line 196-198

Table 2 is still lacking quite a few units. Monthly average temperature for instance.
Response: Thank you for this comment. We have now included units of measurements for climatic covariates in Table 2.

Reviewer 1: Prof. Virasakdi Chongsuvivatwong, Epidemiology Unit, Faculty of Medicine, Prince of Songkla University, Hatyai, Songkhla, 90110, Thailand
Thank the authors for careful revision of the paper. My all points are basically well responded. The only concern that blocked my from accepting this paper is the weakness of the data. We know that there are plenty of drug resistant TB. They are different in level of public health important and need different approaches but were treated as single group in this paper. I am sorry to say that the paper should be transferred to a lower impact factor journal.
Response: We agree that using routinely collected secondary data has its own limitation. This limitation has been mentioned in the discussion sections of the manuscript on page 11 and 12 line 288-296. However, we still believe that this study was important to understand the spatial distribution of DR-TB and identify areas of potentially high risk at lower administrative levels, important information for both international readers and local public health experts. This study includes RR-TB, MDR-TB, and XDR-TB, all are major global public health concerns. We applied a robust geospatial analysis using DR-TB (i.e., RR-TB, MDR-TB, and XDR-TB) surveillance data and several ecological level variables collected from multiple sources to maximize the usefulness of DR-TB surveillance data for decision making.

VERSION 3 – REVIEW

REVIEWER	Alex R Cook National University of Singapore, Singapore
REVIEW RETURNED	31-Jan-2021

GENERAL COMMENTS	Thanks for responding to my email. The edits you propose will help clarify your results. Thanks for the nice paper. Comment to editor: I've written to the first author with two remaining minor issues to address (to the text and one table), and am happy with the response. I recommend acceptance of the paper once these minor changes are made. To allow this in the system I would recommend a minor revision. I don't need to see the paper again.

VERSION 3 – AUTHOR RESPONSE

Reviewer 2: Dr. Alex R Cook, National University Singapore Saw Swee Hock School of Public Health

1. It should be capital K. You're using lower case k for the dummy variable (lower part of the summation) and a capital K for the number of variables (upper part of the summation)
Response: Thank you. We have accepted the comments.
2. You had written Ui not U_i here. Ui is the product of U and i
Response: We have also accepted this correction.
3. Include the unit of measurements.
Response: We have now included the unite of measurements for all variables in Table 2.