Geospatial-Based Model for Diagnosing Potential High-Risk Areas of Tuberculosis Disease in Malaysia

Abdul Rauf Abdul Rasam^{1,2*}, Noresah Mohd Shariff², and Jiloris F. Dony³

¹Faculty of Architecture Planning and Surveying, Universiti Teknologi MARA Selangor, Malaysia

²Geography Programme, School of Distance Education, Universiti Sains Malaysia, Penang, Malaysia

³Sabah State Health Department, Ministry of Health Malaysia

Abstract. Malaysia has a medium burden of tuberculosis (TB) incidence based on World Health Organization (WHO) indicator, but the current trend of TB cases is generally alarming. The Ministry of Health (MOH), Malaysia has set up several guidelines to control the disease, however, the national TB technical report in 2015 addressed that existing detection methods of TB on the site still need to be integrated with relevant alternatives. A geospatial based model is proposed to identify potential high-risk areas of TB especially for targeting missing cases and undiagnosed people. The model was developed with three core stages; framework construction, data collection, and risk analysis and modelling. Eight risk factors: urbanisation, distance to factory, socio-economic status (SES), risk group, human mobility, house type, distance to healthcare centres, and number of population were utilised to determine risk rate of TB modelling. This innovative model has successfully estimated a 65 % of potential high-risk TB areas and targeted 106 high-risk localities in the 10 risk sections of the study area. These risk localities have general similarities with other endemic areas worldwide, but there are some interesting findings revealed in this local study towards in the TB control programme. Most of these cases did not only occur in high rise housing areas, but they are concentrated at industrial location, mobility pattern and socio-economic status in urban city. Although, urban areas are favoured area for the local TB, the disease could also potentially occur in semi-urban or rural areas.

1 Introduction

Ministry of Health (MOH) Malaysia reported that the country has a slightly increasing trend of tuberculosis (TB) cases starts from 16 665 in 2006 to 25 739 in 2016. Selangor is among the top-three highest TB cases in the country with more than 4000 cases reported from 2013 to 2016. Therefore, the MOH, Malaysia has set up several guidelines to control the disease systematically. However, the findings from the national technical report on TB in 2015 asserted that the existing methods for TB screening among high-risk groups need to be strengthened in order to increase TB cases detection rate (CDR). [1] of WHO also agreed that the current method still fails to address inequitable distribution of disease and does not diagnose many TB patients in marginalised areas.

This situation is caused by several factors especially the inefficiency of the existing method or system to comprehensively detect the TB cases. For example, although the biomedical method has advantages in terms of TB diagnosis on human body, the method does not consider geographical or environmental factors. This is because TB cases are not only influenced by human-based factors, but also are driven by environmental risk

factors such as land use, human movement and housing condition. Consequently, this existing method needs to be combined with other techniques [2-4] in order to improve management of cases and analytical power.

In parallel with this issue, [4] has suggested that the combination between molecular techniques and geographical or geospatial techniques can enhance the TB transmission analysis for identifying different geographical areas in a high-burden disease area and subsequently enhancing targeted screening efforts. As such, this study proposes a geospatial model as an alternative method to limit the spread of TB and to detect potential risk areas.

Malaysian researchers have made some attempts to analyse and predict the incidence of TB in particular areas using geospatial or different techniques, but there are only a few attempts of geospatial applications to estimate TB risk areas at the local level. For examples, [5-7] emphasised on the spatial pattern on the geographical distribution and the risk analysis of tuberculosis using secondary data of TB. Therefore, technical aspects in this spatial epidemiology (SE) such as GIS multicriteria decision analysis (MCDA), geo-statistics and spatial statistics are creatively combined in this study to develop a geospatial-based model for

* Corresponding author: rauf@salam.uitm.edu.my

identifying high-risk TB areas. A geospatial model can perform a better spatially decision-making system for public health management [8] and to estimate potential high-risk TB areas [9-13].

2 Review on tuberculosis spatial epidemiology and modelling

2.1 Tuberculosis distribution and risk factors

Tuberculosis (TB) is produced by bacteria of *Mycobacterium tuberculosis* (MTB) that most frequently affect the lungs, spreading from person to person through the air [14]. This transmission mechanism helps to determine influential risk factors and distribution patterns of TB incidences that link to time, place, and personal or population characteristics. Risk factors and pattern functions are important element to determine spatial variation analysis in disease risk to the social environment and the health of people in places [15]. The specific functions include disease mapping, geographic correlation studies, identifying risk factor, transmission pattern, disease clusters, clustering and distribution of the disease.

[16] estimated that one-third of the global population is potentially infected with MTB. [17] reported that the majority of TB cases come from African and Asian regions especially in South Africa and India which are skewed vigorously toward the low-income and developing economies. There are distinct differences in the contributing risk factors between the developing and developed countries. TB cases in developing countries are clustered in populated density, low socioeconomic status and poor housing conditions or environment, whereas in the developed countries the cases are driven by foreign-born immigrants and animal-borne TB factors.

In Malaysia, the general trends of TB cases are slightly increasing [18]. Sabah, Selangor and Sarawak recorded the three-top highest prevalence of TB cases in the country. The local spatial pattern of TB cases in Peninsular Malaysia has revealed that the main concentration of TB is located at inner-city areas especially those with a high number of the population, a low socioeconomic status and high urbanisation [5].

2.2 Epidemic modelling of tuberculosis: a geospatial approach

Detection of TB cases, contacts and intervention programme at the site are still important to control and prevent the disease global as highlighted in WHO and Malaysia agenda. Unfortunately, the current TB cases in Malaysia still record a slight increase in the incidence trends even though there are many as set in the Millennium Development Goal (MDG), 2015. A geospatial-based modelling is proposed as an innovative approach in detecting the missing cases of TB.

Modelling is a multiple and analytical process to measure and understand a phenomenon, such as the mechanisms underlying the spread of infectious diseases.

Modelling technique has been applied in a certain field to quantify pattern, correlation and assessment of a situation. The combination of epidemic modelling concepts and geospatial modelling techniques is an ideal way in estimating TB risk areas and their correlation with influential risk factors such as socio-economic status, human and physical environment.

The modelling approach utilised will differ depending on the aim of the study, how well the study of disease transmission is comprehended, the amount and quality of information accessible, and the foundation and experience of the modellers [19]. Thus, choosing an appropriate formalism for an epidemiological modelling can pose a challenge. Geospatial or spatial-temporal model are the alternative approaches in understanding and prediction of the infection spread or incidence as suggested by previous experts [20-21].

Modelling in SE such as geospatial or GIS perspective is generally similar in terms of context to medical geography. [22] defined a model as a simplified representation of a phenomena or a system for certain application. These models according to [24-25] have advantages in analysing disease phenomena. [22] proposed a dynamic GIS or spatial model as a predictive model as having special characteristics such as prescriptive, stochastic and deductive or theoretical techniques. Geospatial dimension will also play a significant role in modelling social population and ecology physical phenomena of diseases.

3 Materials and methods

A conceptual framework of spatial epidemiological (SE) data analysis [24] and GIS-multi-criteria decision making or analysis (MCDM/MCDA) method [22-23] are mainly adapted in this geospatial model. It includes the description of spatial patterns, identification of disease cluster or risk factors and explanation or prediction of disease risk as illustrated in Fig. 1.

Fig. 1. Main framework in developing a geospatial-based modelling for potential high-risk TB areas.

3.1 Research framework of risk factors and level scale of Tuberculosis

First framework explains theoretical framework on the concept of a local TB transmission that is derived from well-established studies on TB spatial pattern and risk factors. The main components are insert/internal factors and contact/external factors. These eight risk factors include urbanisation, distance to factory, socio-economic status (SES), risk group, human mobility, house type, distance to healthcare centres, and number of population were utilised to determine risk rate of TB modelling. Scholarly review and exclusive reports from WHO and MOH were used in selecting local risk factors based on a MCDA.

Second framework emphasizes on determination of risk level or scale of the risk areas. High risk areas TB are operationally defined according to selected risk factors and a five-risk scale. Whereby 1 (dark green) and 5 (red) indicate the lowest and highest potential risk areas. The determination of the risk scale is combined from GIS-MCDA and logistic regression method. The concept is similar to the definition of high burden countries by WHO, and high-risk group of TB State Health Department (JKN), Selangor.

These frameworks are crucial to define risk concentration of TB areas and targeting the possible risk localities in the study area as exhibited in Fig.2. Concentration risk in this disease can be characterized as any single (direct and/or indirect) exposure or group exposures with the possibility to deliver infection sufficiently extensive towards undermining individuals' wellbeing or its capacity to maintain its area.

Fig. 2. Proposed a five-risk scale concentration of TB area

3.2 Study area, data collection and techniques

Shah Alam in Selangor as shown in Fig. 3 is selected as a study area since it currently has a diversity of environment related TB risk factors. Meanwhile, the data and instruments used in the study include the disease cases, spatial-environmental data for risk factors, and spatial-statistical packages (SPSS and ArcGIS).

Primary data obtained by the researchers especially from expert opinions (District Health Officers from Petaling) and site visit at high-risk areas of the study area. For the main secondary data used in this study are TB cases from 2013 to 2015 collected from MyTB system, JKNS, and spatial data related to eight risk factors, namely urbanisation (density of population and physical development), distance to factory (in meters),

income status (RM), risk group (young and senior citizen, PTB, non-Malaysian, no BCG, diabetics and others), human mobility (location/distance of red flag cases such as patient with a HIV/AIDS and other), house type (cost, size and condition), distance to healthcare centres (in metres), and number of population (in a house). Base maps (land use and Google Earth) were also collected for disease mapping.

Fig. 3. Map of Shah Alam for JKN Selangor's TB monitoring system.

3.3 Data process, analysis, modelling and evaluation

This stage involves the development of proposed geospatial modelling using GIS-MCDA or index model, logistic regression, geostatistical model and related theories of Tobler and spatial diffusion. This main analytical framework of research is innovatively combined to produce a geospatial model as suggested by previous experts [22],[24] and [26].

In general, GIS-MCDA method comprises three main stages as illustrated in equation (1): i) selection of risk factors, ii) calculation of risk factor weights and iii) eliciting local risk factors [22-23]. Each criterion or selected risk factor was directly ranked (from 1 to 8) by the selected experts (four local health staff) and then the values or weight were standardised using rank sum techniques as showed in equation (1) from 0 to 1.

$$W_j = \frac{n - r_j + 1}{\sum (n - r_k + 1)} \quad (1)$$

Where;

W_j = the normalized weight for the j^{th} criterion

n = the number of criteria under consideration ($k=1,2,3,\dots,n$) and

r_j = the rank position of the criterion each of the criterion is weighted ($n - r_k + 1$) and then normalized by the sum of all weights and that is $\sum (n - r_k + 1)$.

The final stage is eliciting the influential risk factors and risk weights of local TB according to expert's rank or MCDA method. The values of standardised weight (W_j) were multiplied with each risk factor risk value. Then, the total of risk values were determined (as Total Risk Rank [DV]) using a Likert scale by 1 to 5; Scale 1 and 5 illustrate the lowest and the highest risk of the TB risk respectively. The values of 1-5 were added to the TB geodatabase and were also placed into logistic outcome using 1 (Yes Risk) and 0 (No Risk) for the mapping of TB risk areas. In order to determine which model is suitable for the study area, three tables in block 1 reports of regression were mainly referred to as the statistical tests of goodness.

Geostatistical mapping (Inverse Distance Weights, IDW) was then used for deterministically displaying spatial data display and exploratory analysis of TB disease in Shah Alam. A developed five-risk scale of the existing TB cases or control points were interpolated to measure the unknown surface areas. The accuracy of the risk map of TB that were referred to error assessment and randomness found in the model, especially in the residual means standard error (RMSE).

4 Results and Discussion

4.1 Modelling of potential high-risk TB locations

The regression output of risk model indicates that the selected risk factors of TB patient were significantly associated with the probability of potential TB risk locations ($p=0.008$, Wald test). The output provides the coefficients for intercept = -46.807 and risk factors with their coefficients. These coefficients were then entered in the logistic regression equation (2) to estimate the probability of risk model by converting odds to probabilities ($Y = \text{Odds}/1 + \text{Odds}$ or $1/(1 + \exp^{-(\beta_0 + \beta_1 x_1 + \dots)})$).

The Log Outcome (0-1) was included in the database as probability of potential TB risk locations (Y) and served as control points to estimate risk map using IDW in a geostatistical environment. The values of the outcome in the equation are important to gain insightful view into the real TB situation and to identify risk TB areas in Shah Alam.

$$\text{TB Risk Location Model (Y)} = 1/(1 + \exp(-(-46.807 + 1.161 (\text{Urban}) + 2.061 (\text{Popul}) + 1.715 (\text{House}) + 2.395 (\text{SES}) + 2.841 (\text{Group}) + 1.598 (\text{Factory}) + 3.654 (\text{Mobility})))$$

(2)

4.2 Mapping of potential high-risk TB areas

Fig. 4 reveals the overall risk areas of TB in Shah Alam by combining a logistic regression technique (equation 2), GIS and geostatistical model. The estimated area has potential risk of 65 %, while 35% of non-potential risks areas. The estimated risk areas in geostatistical technique are also consistent with results found in the logistic

regression model where the study area has a medium-high risk of TB. As expected earlier, the risk areas are dispersed at northern zone (Section U20 and U19), central zone (Section 17) and southern zone (Section 26 and 27), with the possibility of having a high level of TB transmission.

Fig. 4. Map of high-risk TB areas in Shah Alam using the geospatial-based models.

Non-risk areas are mostly located at central zone and a few areas at southern and northern zone. Clustering in the zones is generally frequented by people who are converged at residential, social and economic areas as revealed from the qualitative findings. Anthropogenic or human factors are the main indicators of TB especially to people who have high-risk characteristics such as a weakened immune system, tobacco smoker, poor community [14].

The estimated potential risk sections have similarities of environmental risk factor related to TB occurrences such as crowded population and high-risk group especially from medium and low SES and foreign-born immigrants. These areas are also in close to the industrial zones where there are large numbers of foreign workers, besides being distant from healthcare centres that provide TB facilities (x-ray and other TB detection tools), except some sections in the central regions.

The contradictory factors are the type of house and number of population which are not clearly related to local TB because most of the sections are underdeveloped which means that they are still in the setting of a village except for Section U13, U5, U6, and U17 which are undergoing development. Another feature includes a disorganised settlement pattern or village condition specifically in northern zone (U5, U6, U13,

U17, U18, U19, U20), southern zone (S27 and 28) and few in central region (S17 and S19).

4.3 Model application for targeting potential risk areas of TB concentration

The imperative contribution of this study is to produce a risk map of TB using an integrative geospatial-based model for estimating the potential spots and screening programme in the study area. Therefore, this study proposed a geospatial model that has strengthened the power of risk estimation for potential risk TB areas. It covers aspects of local spatial knowledge elicitation (GIS-MCDA), spatial uncertainty (regression method), and spatial neighbouring (geostatistical method). The proposed model has also a holistic risk mapping for identifying high-risk TB areas. Specifically, the potential or existing high-risk TB area in Shah Alam have been suggested to assist the PKD, Petaling and JKN Selangor in TB screening on the site since there is no single and clear definition of local endemic or clustered areas.

5 Conclusion

One of the main agendas of Malaysia is to enhance existing methods in controlling the spread of the disease on the field. Developing a decent TB detection tool is not a straightforward matter while the accepted models of disease causation also need meticulous interaction of factors and settings before a disease occurs. Combining geospatial based model with epidemic-based approach to detect TB on a human body and investigating the environmental factors are required for holistic detection or prevention plan of TB spread. Therefore, a geospatial-based modelling is recommended in this study to evaluate spatial human and environmental risk factors for the detection of potential high-risk TB locations in Shah Alam as a case study area. The model could be applied in the country because it consists of decent characteristics in locally defining the potential risk TB areas as suggested by epidemic modelling experts.

The authors would like to thank the TB and Leprosy Sector of Ministry of Health Malaysia and Selangor States Health Department for providing TB datasets in this research paper. Universiti Teknologi MARA Selangor is also appreciated for providing a research fund (600-IRMI/PERDANA 5/3 BESTARI (052/2018)). This research has been registered in the National Medical Research Register, Malaysia (ID: NMR R - 15-2499-24207).

References

- 1 N. Nishikiori. Tuberculosis control among high risk and vulnerable populations. [Online]. Retrieved from [http://health.gov.fj/PDFs/Tuberculosis control among high risk.pdf](http://health.gov.fj/PDFs/Tuberculosis%20control%20among%20high%20risk.pdf) . (2011)
- 2 M. V. Burgos and A. S. Pym, *Molecular epidemiology of tuberculosis*. doi:10.1183/09031936.02.00400702," European Respiratory Journal,

- doi:10.1183/09031936.02.00400702, **20**, 54S–65s (2002).
- 3 B. Mathema, N. E. Kurepina, P. J. Bifani, and B. N. Kreiswirth, *Molecular epidemiology of tuberculosis: current insights*. Clinical Microbiology Reviews, vol. 19(4), no. doi:10.1128/CMR.00061-05, 658–85 (2006).
- 4 S. Narayanan, *Molecular epidemiology of tuberculosis*, Indian J Med Res, 233–247 (2004).
- 5 A. R. Abdul Rasam, N. M. Shariff, J. F. Dony, and A. Misni, Socio-environmental factors and tuberculosis: an exploratory spatial analysis in Peninsular Malaysia, *International Journal of Engineering & Technology*, **7** (3.11), 187-192, (2018).
- 6 S. Azhar Shah, M. I. Mohd Nor, A. H. Harman Shah, and T. Aris, *Penggunaan Aplikasi GIS dalam Penyakit Tuberkulosis di Cheras, Kuala Lumpur, Malaysia.*, Malaysian Journal Public Health Medicine, **2**, 15–26, (2002).
- 7 A. R. Abdul Rasam, N. M. Shariff, and J. F. Dony, *Identifying High-Risk Populations of Tuberculosis Using Environmental Factors and GIS Based Multi-Criteria Decision Making Method.*, ISPRS - International Archives of the Photogrammetry, Remote Sensing International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, **XLII-4/W1**, 9-13, (2016).
- 8 D. O. Fuller, A. Troyo, T. O. Alimi, and J. C. Beier, *Participatory Risk Mapping of Malaria Vector Exposure in Northern South America using Environmental and Population Data*, no. doi:10.1016/j.apgeo, **48**, 1-7, (2014).
- 9 D. J. Daley and J. Gani, *Epidemic Modeling: An Introduction*. NY: Cambridge University Press., (2005).
- 10 H. Lin, *Use of Spatial Information to Predict Multidrug Resistance in Emerging Infectious Diseases*, **18**, 5-7, (2012).
- 11 K. Middelkoop, L. Bekker, C. Morrow, N. Lee, and R. Wood, *Decreasing household contribution to TB transmission with age: a retrospective geographic analysis of young people in a South African township*. BMC Infectious Diseases , **14**, 221, 1-7, (2014).
- 12 P. K. Moonan, *What Is the Outcome of Targeted Tuberculosis Screening Based on Universal Genotyping and Location?*, Am J Respir Crit Care Med, **174**, 599–604, (2006).
- 13 E. Musenge, P. Vounatsou, M. Collinson, S. Tollman, and K. Kahn, *The contribution of spatial analysis to understanding HIV/TB mortality in children: a structural equation modelling approach.*, Glob Health Action, (2013).
- 14 WHO, World Health Organization Web site. [Online]. <http://www.who.int/mediacentre/factsheets/fs104/en>. (2015)

- 15 R. S. Ostfeld, G. E. Glass, and F. Keesing, *Spatial epidemiology: an emerging (or re-emerging) discipline*, Trends in Ecology & Evolution, **20**, 6, 328–36, (2005).
- 16 B. Mathema, N. E. Kurepina, P. J. Bifani, and B. N. Kreiswirth, *Molecular epidemiology of tuberculosis: current insights*. Clinical Microbiology Reviews, **19**, 4, 658–85 (2006).
- 17 WHO, World Health Organization Web site. [Online].
<http://www.who.int/mediacentre/factsheets/fs104/en>
(2014)
- 18 TB/Leprosy Sector, Ed., Ministry of Health; The National TB strategic plan (2010-2015). Putrajaya, Putrajaya: NTLP, (2015).
- 19 M. G. Garner and S. A. Hamilton, *Principles of Epidemiological Modelling*, Review Scientific et Technique, **30**, 2, 407-416, (2011).
- 20 C. S. Childs, S. Florida, P. Health, and M. College, *Evaluating the Spatial Relationship Between Remotely-Sensed Urban Vegetation and Tuberculosis in Guayaquil, Ecuador*, **5**, (2013).
- 21 M. Roberts, *Nine challenges for deterministic epidemic models*, **10**, 49-53, (2016).
- 22 K. S. Chang,. New York, USA: McGraw Hill, 389-399. (2011)
- 23 J. Malczewski, *On the use of weighted linear combination method in GIS: Common and best practice approaches*, Transactions in GIS, **4**, 50-22, (2000).
- 24 D. U. Pfeiffer, T. P. Robinson, M. Stevens and D. J. Rogers, *Spatial Analysis in Epidemiology*, Oxford University Press, USA, (2008).
- 25 A. B. Lawson and P. Leimich, *Approaches to the space time modelling of infectious disease behaviour*, IMA Journal of Mathematics Applied in Medicine and Biology, **17**, 1-13, (1999).
- 26 P. A. Longley, M. F. Goodchild, D. J. Maguire, and D. W. Rhind, *Geographic information systems and science*, 2nd ed., John Wiley & Sons Chichester, Ed., (2005).