Spatial and Statistics for Profiling Risk Factors of Diseases: A Case Study of Tuberculosis in Malaysia

A R Abdul Rasam1*, N Mohd Shariff2, J F Dony3 F Othman1
1Faculty of Architecture Planning and Surveying, Universiti Teknologi MARA Shah Alam, Selangor, Malaysia
2Geography Programme, School of Distance Education, Universiti Sains Malaysia, Penang, Malaysia
3Sabah State Health Department, Ministry of Health Malaysia

Email: rauf@uitm.edu.my

Abstract. Understanding concepts of a proper disease transmission risk is not a straightforward process. In the context of tuberculosis (TB) dynamics, the concepts require the exploration of two meticulous criteria to produce an accurate epidemic modelling of the risk areas of the disease. The criteria include interpreting the biological transmission of the disease and applying multidisciplinary approaches. Spatial statistics were used to evaluate the preferences of risk factors in Shah Alam, Malaysia. GIS-multicriteria decision making (MCDM) method and logistic regression method were specifically integrated to select the local risk factors and seven influential factors were ranked accordingly i.e. human mobility, high risk group, socio-economic status (SES), population, type of house, distance of factory and urbanisation. Each has relative risk rate that affects the cases and the combination of them will even impact more on the overall risk concentration of TB. Human–based factors are identified as dominant effects to the risk than biophysical factors, for example, a location of TB risk will be increased by four times if individuals are living together with people who have TB disease for a particular time period. This geospatial method is expected to predict a better factor prediction in identifying hotspot areas of the disease.

1. Introduction
The Ministry of Health (MOH), Malaysia has established several guidelines to control tuberculosis systematically. Although the guidelines seem to be focusing on human or biomedical approach (TB screening and x-ray methods) to control the disease as stated by WHO in 2014, the main aim of the guidelines or agendas is still the same which is to emphasise on strengthening control, prevention and elimination of the disease as suggested by the TB and Leprosy Sector, MOH Malaysia [15].

However, the findings from the national technical report on TB in Malaysia (TB and Leprosy Sector, MOH, Malaysia, 2015) asserted that the existing methods for TB screening among high-risk groups (such as missing and undiagnosed TB cases) need to be strengthened in order to increase TB cases detection rate (CDR). Similarly, [22] of WHO also agreed that the current method still fails to address inequitable distribution of disease and does not diagnose many TB patients especially among poor and vulnerable communities, and marginalized people.

This situation is caused by several factors especially the inefficiency of the existing method to comprehensively detect TB cases. For example, the risk factors influencing the TB cases is multi-sources from geographical or environmental factors human-based factors (such as high-risk group and socio-economic status). Consequently, this method needs to be combined with other techniques [2][14][21] in order to improve management of cases and analytical power. As such, this study examines the capabilities of spatial statistics to determine the risk factors contributing to TB cases in Shah Alam. This
Spatial epidemiology (SE) is empirically applied to provide integrated spatial frameworks for enhancing current technical issues on case detection of a disease (such as TB). The core framework includes the spatial analysis (spatial pattern and spatial risk correlation) and spatial modelling as a local disease surveillance management tool. A spatial statistics method can perform a better spatial decision-making system for public health management \[6\]|\[28\] to estimate potential high-TB risk areas \[12\]|\[16\]|\[18\]|\[29\] especially in using spatial statistical approach \[30\]|\[31\]|\[32\].

2. A review of spatial and statistics for risk disease assessment

2.1. Spatial epidemiology of tuberculosis transmission risk

At present, TB is still a global concern and it can occur in any part of the world. \[4\] estimated that one-third of the global community is infected with M. tuberculosis. In 2010, the World Health Organization (WHO) stated that there were approximately 8.5–9.2 million cases and 1.2–1.5 million deaths from TB worldwide. The global distribution of TB cases in 2000 is skewed heavily toward the low-income and emerging economies. \[33\] reported that the majority of the TB cases come from African and Asian regions especially in South Africa and India. Other countries of high TB prevalence in Asia include China, Bangladesh, Indonesia, and Pakistan.

There are notable 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 environments, whereas in the developed countries the cases are driven by foreign-born immigrants and animal-borne TB factors. However, an interesting aspect is that most of the foreign-born immigrants in developed countries have similar characteristics found in developing countries that are mainly crowded and poor environments.

In Malaysia, the trends of TB cases are slightly increasing \[17\]. Sabah, Sarawak and Selangor recorded the highest prevalence of TB cases in 2011 in the country. The general spatial clustering of the recent TB distribution patterns in Peninsular Malaysia showed that there is no specific clustering in the study area, but the overall spatial pattern of TB cases concentrated in urban areas especially those with a high number of the population, a low socioeconomic status, and high urbanisation. However, it is insignificantly related with and non-forested area of all the states \[1\]. Specific analysis at smaller areas needs to be conducted in Malaysia to determine the precise risk factors contributing to particular regions.

2.2. Spatial statistics for tuberculosis dynamics

An appropriate understanding on local TB transmission and its epidemic stages are fundamental steps to develop a better predictive model of diseases. A sufficient concept of TB transmission can simplify the dynamic representation of real TB phenomena or system according to a local actual situation and dynamics. Dynamics is the interaction process which is related to clear changes or finding the causality underlying the type and speed of the observed evolution. The concept of dynamics can provide a simplified map, statistical information, explanatory model based on a known fundamental mechanism, and towards the application of a multidisciplinary model \[7\]|\[10\].

Tuberculosis causation and transmission process is not a straightforward matter and the accepted models of disease causation require the precise interaction of factors and conditions before a disease occurs. However, the traditional model (epidemiologic triangle, epi-triad) or cycle of infection illustrates that infectious diseases result from the interaction of three components including external agent resources (bacterium in human, environments), a susceptible host (droplet, airborne) and an environment that brings the host and agent together. Specifically, transmission occurs when the agent or pathogen leaves its reservoir or host through a portal of exit, and is conveyed by a mode of transmission and enters through an appropriate portal of entry to infect a susceptible host.

Thus, the factors that determine the probability of TB transmission need to be studied in-depth and they can be divided into several epidemic stages from human body (intrinsic or exogenous factor) to human environment (extrinsic or endogenous factor). According to CDC USA (2011) and Narasimhan
et al. (2013), internal or human level conditions are dominant or well-established factors of susceptibility and infectiousness, while external factors include environment and exposure. Determining risk factors is important to identify the characteristics of high-risk population and areas according to global perspective. Thus, efforts should also be made to collect risk factors data in routine surveillance for TB disease [20].

Techniques in spatial epidemiology (SE) are applied in certain fields to quantify pattern, correlation and assessment of a situation. For example, in the context of health geography, the effect of a certain risk factor on disease outbreaks might be determined by estimating its correlation and risk areas using a certain model. In this study, the combination of statistical method and spatial method is applied in estimating TB risk areas and their correlation with influential risk factors such as socio-economic status, human and physical environments in Shah Alam. These methods can help in explaining complex and often difficult aspects of system behaviour to ‘non-experts’, especially if the models can produce graphical visual, even animated, outputs [27]. Besides this, [7] also stated that the methods can be a powerful tool to assist animal health policy development and disease prevention and control. SE introduced and applied a spatial framework for epidemiological studies [8][9][23][25] for spatial pattern, cluster, and prediction of TB disease risk.

GIS-MCMD and statistical methods are combined to produce a risk map of the TB incidence and population as conducted in local studies [34]. These methods also emphasise on the population or community health assessment rather than individual decisions (clinicians are concerned with the health of an individual). Lena Sanders (2007) stated that spatial dimension plays a key role in modelling for phenomena including social and human population and ecology physic environments phenomena besides knowledge from various fields and statistician model, while statistical framework is a common method to set out a hypothesis in the thematic model using dependant/to be explained (dependent variable), exploratory of independent variables (IV) both of quantitative and/or qualitative data. This combination would be enhancing the technical aspects in spatial epidemic models and draw attention to the lack of models that explore the relationship between communicable [5][26].

3. Data and materials

The main secondary data used in this study are TB cases (all types) from 2013 to 2015 obtained from MyTB, Selangor State Health Departments. Spatial and statistical methods were integrated to determine the potential risk factors of TB in Shah Alam. This spatial-statistical method is relevant to cover a larger scope of the epidemic model. The technical framework used in this geospatial-based method is referred to as GIS modelling and spatial epidemiological (SE). A binary logistic statistical method has been selected because the data did not meet the assumption of the normality, while spatial method has the capabilities in data mapping and analysis of disease risk assessment. Shah Alam is a suitable location to be selected as a study area due to its geographical condition, dynamic population growth and human movement. Four main steps applied in the study were i) risk factor selection, ii) spatial data collection, iii) spatial data processing, and iv) spatial statistical analysis.

3.1. Selection of risk factors

Establishing a spatial concept of local TB transmission and data collection of risk factors are essential tasks to identify the transmission components involved in the study and. Scholarly sources (a systematic review) from global TB studies, exclusive reports from WHO and MOH were used in selecting local risk factors based on a multi-criteria analysis, including human-factors and biophysical indicators.

Biophysical environment is defined as the biological environment as well as the physical activity within it, including land use/urbanisation, type of house, proximity of healthcare facilities and industrial zone. Human-population based environments or human risk factors refer to human and individual characteristics, which influence behaviour at work in ways which affect health such as density of population, social economic status, concentration of risk group and population mobility. The
The combination of these indicators of risk factors is important to develop a standard and holistic guideline for local levels of external risk factors and risk areas priorities in Malaysia.

Several technical methods are then used such as GIS-MCDM, and regression model. This main analytical framework of research is innovatively combined from GIS operations [4], spatial epidemiological methodology [25]. Table 1 presents a list of selected risk factors and their variables proposed in this study. These predictors were included in the MyTB databases for risk assessment as individual-level spatial information.

| Indicators                                    | Independent variables/Factors (Code) (Scale /No_Class)                                      | Data Model                      |
|----------------------------------------------|---------------------------------------------------------------------------------------------|---------------------------------|
| Natural environment of landscape             | Urbanisation-land use (Score_Urb) (Continuous scale from 1 to 5)                           | Surface data from point inputs  |
| (Biophysical and ecological environment)     | Types of house-settlement (Score_Hou) (Continuous scale from 1 to 5)                      | Surface data from point inputs  |
|                                              | Distance to healthcare facilities (Score_Hea) (Continuous scale from 1 to 5)               | Surface data from point inputs  |
|                                              | Distance to industrial factory locations (Score_Fac) (Continuous scale from 1 to 5)       | Surface data from point inputs  |
|                                              | Number of people in a house (Score_Peo) (Continuous scale from 1 to 5)                    | Surface data from point inputs  |
| Anthropogenic (Human and population)         | High risk groups (Score_Gro) (Continuous scale from 1 to 5)                              | Surface data from point inputs  |
|                                              | Household income-Socio economic status, SES (Score_Soc) (Continuous scale from 1 to 5)    | Surface data from point inputs  |
|                                              | Concentration of Patient Mobility (Score_Mob) (Continuous scale from 1 to 5)             | Surface data from point inputs  |

Data and instruments used in the study are the disease cases, spatial-environmental data for risk factors, and spatial-statistical packages. Two approaches were applied to collect the data, which are primary and secondary data collection as displayed in Table 2. Primary research consists of a collection of original primary data collected by the researcher especially from expert opinions (Health staff from TB Unit of Pejabat Kesihatan Daerah, PKD, Petaling) and site visit at high risk areas of the study area. Systematic reviews and questionnaires were conducted to define local risk factors and their risk rates. Secondary data refers to the data that was collected by others such as TB cases from MyTB System.

| Data                                           | Data structures and model                                                                 |
|------------------------------------------------|------------------------------------------------------------------------------------------|
| TB Cases (Secondary data)                       | Non-spatial (attribute Data) of all types of TB cases from 2013 to 2015.                  |
| Risk Factors of TB (Primary and Secondary data) | Combination of spatial and non-spatial data. They include 8 influential risk factors of local TB in Shah Alam as shown in Table 1. |
| Base maps of TB mapping (Primary and Secondary data) | Spatial data.                                                                              |
3.2. Spatial data processing and weightage calculation

Data processing and analysis were employed with ArcGIS and SPSS Statistics 21 for descriptive and scientific analysis. All non-spatial data required in this study were converted into spatial data using ArcGIS to become a geodatabase of TB. After selecting the risk factors, standardisation and classification of TB Risk Scale were also conducted. The attributes/tables were added with eight risk factors (as independent variables, IV of risk factor of TB), and one for total score of risk factors (as a dependent variable, DV) with their risk values. Each table has a standardised value of risk ranking from 1 to 5 since some original data were formed as nominal values.

A ranking procedure is based on local expertise and government standard guidelines. For household income, the Economic Planning Unit of the Prime Minister’s Department Malaysia (2012) has defined poverty as referring to people who have average monthly household income at or below RM 830.00. Thus, this study assumed that people who earn average incomes of less than RM 830.00 to be included as risk part of groups (value 5), while people earning more than RM 2490.00 or low-risk groups were categorised as value 1 and value 2. The values of 1-5 were added in the database for each IV and then the average was calculated to obtain a Total_Risk Score (DV). If a classification scale of risk factor could not be clearly decided using the judgement from the expert opinion or theoretical context, the statistical evaluation (mean value) is also alternatively used in ranking the risk.

Then, these scaling values were added into the existing databases of risk factors (IV). Total_Risk Score (DV) can be replaced with a new value as 1 (rank 4 and 5) and 0 (rank 1 to 3) for logistic regression modelling (Logistic_Outcome, DV).

Furthermore, Total_Risk Rank (DV) was created to develop a GIS–MCDM model according to expert opinion or knowledge driven values. These DV values can be used in calculating the risk level of TB factors. In general, MCDM method comprises of three main stages: i) selection of risk factors, ii) calculation of risk factor weights and iii) eliciting local risk factors. MCDM method [4][13] analysis was implemented to rank the TB infection risk rate or weight. Each criterion or risk factor was directly ranked (from 1 to 8) by the selected experts and then the values or weight were standardised using rank sum techniques as showed in equation (1) from 0 to 1. The final stage is eliciting the influential risk factors and risk weights of local TB according to expert’s rank or MCDM method. The values of 1-5 were added to the TB geodatabase. The scale values were also placed into logistic outcome using 1 (Yes Risk) and 0 (No Risk) for calculating the risk level of TB.

\[ W_j = \frac{n - r_j + 1}{\sum (n - r_k + 1)} \]  

(1)

Where:
- \( W_j \) = the normalized weight for the \( j \)th criterion
- \( n \) = the number of criteria under consideration (k=1,2,3…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 \(\Sigma (n - r_k + 1)\).

3.3. Spatial statistics analysis

A logistic regression analysis was used in characterising and prioritising the risks of TB after the data passes data assumptions. The model is an inferential statistical technique used to predict the presence or absence of a characteristic or outcome values (DV) of a set of predictor variables. The regression
coefficients can be used to estimate the odds ratios (OR) for each of the independent variables in the model. A logistic regression model used to predict the potential location of TB disease by estimating the value of $\beta_0$, $\beta_1 \ldots \beta_m$ (individual characteristics) and outcome probability of disease ($Y = \text{a chance or ratio of an outcome}$) is shown in equation (2). The result can be used to understand whether TB risk area/locality can be predicted based on environmental risk factors (i.e., "Yes Risk" or "No Risk". The following prediction equation (2) was used to calculate a predicted probability of having disease or to estimate the risk of disease [11][15]. In selecting the suitable modelling of risk assessment, a comparative study was conducted using 3 different assessments entering the risk factors. To determine which model is suitable for the study area, the tables in block report were mainly referred to as the statistical tests to determine the model goodness.

$$Y = \frac{1}{1 + \exp^{-\left(\beta_0 + \beta_1 X_1 + \ldots + \beta_m X_m\right)}} \quad (2)$$

For the results of MCDM, logistic regression model was used to estimate the unmeasured points of TB risk in the study area using GIS interpolation or geostatistical method. These points have all the characteristics and scale of TB risk factor (IV) and total TB risk factor (DV). In Geostatistical Analyst Wizard in ArcGIS, the input data, method and attribute must be correctly selected. The desired parameters should also be set on the IDW dialog box to provide the risk points, including Optimize Power value=2, Symbol size=standard, Neighbour including 15 with at least 10 points, Sector types=standard ellipse, Anisotropy factors=1, and coordinate estimate points. In addition, the box of weight, value menu and preview (neighbourhood) were selected.

4. Result and discussion on profiling spatial effect of TB risk factors
Spatial statistics is quantitative assessment is applied to support the spatial qualitative result in identifying the extent of the risk factors that have affected the incidence of TB, and ultimately use them to control and prevent the spread of TB. It was found that 7 out of 8 common risk factors are influential factors in the incidence according to multi-criteria and regression analysis. It includes factory distance, urban areas, house type, high-risk group, income status, mobility and population. The number of populations in a house is not a statistically significant factor, but according to local experts, this factor should be considered in defining the real effect of TB on the local study. While the distance of healthcare facilities is not statistically significant even though it is a common indicator of TB cases based on global and local perspective.

| Table 3. Justification of risk rate of selected TB risk factors in Shah Alam |
|-----------------------------------------------------|
| **Risk Factors of TB (Risk Value/Rate)** | **Theory and Previous Studies** |
| Mobility (3.654) | Theory yes, especially for patient with an active TB. They will be dangerous and infect others through closed space, short distance and exposure in long duration of time (8 hours). |
| Risk group (2.841) | This is a well-established factor as occurred in the world TB. |
| SES (2.395) | This is a well-established factor as occurred in the world TB, especially in Malaysia. |
| Population (2.061) | This is a well-established factor as occurred in the world TB, either in the house or outside. |
| House (1.715) | This is a potential risk factor of the local TB. The types of houses reflect the condition and space of the building. A poor, low and small/limited house will influence TB occurrences. |
Factory (1.598)  | This is a potential risk factor of the local TB. There is still no strong evidence to show its correlation with local TB occurrences, but it can be related with air pollution and the risk group in the industrial factories.
---|---
Urban (1.161)  | The factor has been proven from previous studies. Most of the cases were located at urban areas with poor housing areas.

The selection of the factors is the basis for the formation of the proposed model of the impact risk assessment and mapping potential areas using GIS-based logistics regression technique. Risk assessment using a logistic technique shows that statistical test indicates a strong relationship between prediction and grouping \((p=0.797)\), and the classification was generally and correctly completed at 93.2%. In addition, human-based indicators, especially human mobility, made a significant contribution to prediction instead of biophysical indicators.

Overall risk factors used in the proposed geospatial model is as shown in Table 3. According to the influential risk ranking and values, human indicators are more dominant factors than biophysical indicators as found in previous studies in Malaysia [1][17] and local expert opinion. Regression was even clearer by illustrating that the Coefficient (\(\beta\)) of risk factors according to ranking includes Score_Mob (3.654), Score_Gro (2.841), Score_Soc (2.395) and Score_Peo (2.061). These factors are included as human indicators. While biophysical factors only recorded minimum Coefficient (\(\beta\)) such as Score_Hou (1.715), Score_Fac (1.598), and Score_Urb (1.161), urbanisation is evidently not significant, but in terms of expert opinion it should be one of the major risk factors in local condition. The distance of healthcare facilities is not included as a factor due to the fact that it is statistically insignificant, as well as supported by experts’ opinion.

It can, therefore, be said that the combination of these seven factors can produce relevant TB risk factors in Shah Alam. Four risk factors based on human indicators are riskier than three biophysical indicators, but the combined results of these indicators can make the area more significant to TB risk. The results are also not only important for local health staff to fully understand the risk factor but also to use them in developing the proposed geospatial epidemic model [1].

In summary, spatial-logistic regression is based on real data and predictive analysis. The results suggest that the logistics may have minimal risk, but according to existing data and local experts’ view, this area should only have a medium burden of TB. This study suggests several recommendations to further strengthen the findings. It includes the logistic regression technique used as a direct (binary) technique that provides outcome to two choices Yes (1) or No (0) only and therefore, it is suggested that a comparison is made with other techniques such as spatial multivariate and ordinal regression models.

The techniques used will also take into consideration the possible variation effects on the environment and different distribution patterns of local TB because the variation can affect the accuracy of the results. This proposed spatial logistic regression model could help the health staff to predict the prevalence of TB in Selangor according to a population or areal level and combined with a GIS mapping for early identification of TB progression.

5. Conclusion
In Malaysia, the presence of spatial heterogeneousness of tuberculosis (TB) incidence may lead to the complexities of risk factors identification because each area has its own unique risk factors and dynamics. This paper is aimed to quantitatively assess the selected eight risk factors in a locality by employing an integration of GIS-MCDM and logistic regression method. Findings from MCDM suggests that 8 risk factors have contributed to the local TB cases, but quantitative risk assessment with a GIS-logistic method found that all seven risks are significant \((p < 0.05)\), including urbanisation, distance to factory, SES, high risk group, human mobility, house type, distance to healthcare centres, and the number of populations in a house. Each relative risk rate that affects TB occurrences and their combination will give more impact on the overall occurrences. Overall, the combination of spatial/GIS and statistical method can enhance the classical method to determine the risk factors contributing to
local TB occurrences. Furthermore, the proposed risk factors in this study can provide a holistic platform in identifying targeted spots for TB risk concentration in the study area.

Acknowledgement
The authors would like to express appreciation to the TB and Leprosy Sector of Ministry of Health Malaysia and Selangor States Health Department for providing TB datasets in this research paper. The research has been registered in the National Medical Research Register, Malaysia (ID: NMR R -15-2499-24207). The authors would also like to sincerely thank Universiti Teknologi MARA Selangor for providing LESTARI Research Fund (600-IRMI5/3/LESTARI (005/2018).

References
[1] Abdul Rasam, A.R., Shariff, N.M., Dony, J.F and Misni, A. (2018). Socio-environmental factors and tuberculosis: an exploratory spatial analysis in Peninsular Malaysia, International Journal of Engineering & Technology, 7 (3.11), 187-192.
[2] Burgos, M. V., & Pym, A. S. (2002). Molecular epidemiology of tuberculosis. European Respiratory Journal, 20(Supplement 36), 54S–65s. doi:10.1183/09031936.02.00400702
[3] CDC USA. (2013). To View or Order the Core Curriculum on Tuberculosis. USA. Retrieved from https://www.cdc.gov/tb/education/corecurr/pdf/corecurr_all.pdf
[4] Chang, K. (2011). Introduction to Geographic Information Systems, New York, USA McGraw Hill, 389-399.
[5] Childs, C. S., Florida, S., Health, P., & College, M. (2013). Evaluating the Spatial Relationship Between Remotely-Sensed Urban Vegetation and Tuberculosis in Guayaquil, Ecuador (Vol. 5).
[6] Dhanaraj B, Papanna MK, Adinarayanan S, Vedachalam C, Sundaram V, Shanmugam S, et al. (2015). Prevalence and Risk Factors for Adult Pulmonary Tuberculosis in a Metropolitan City of South India. PLoS ONE 10(4): e0124260. https://doi.org/10.1371/journal.pone.0124260
[7] Fuller, D. O., Troyo, A., Alimi, T. O., & Beier, J. C. (2014). Participatory Risk Mapping of Malaria Vector Exposure in Northern South America using Environmental and Population Data. Applied Geography (Sevenoaks, England), 48, 1–7. doi:10.1016/j.apgeog.2014.01.002
[8] Garner M.G. and Hamilton S.A. (2011). Principles of Epidemiological Modelling. Review Scientific et Technique, 30(2), 407-416.
[9] Lai, P.; Wong, C.M; Hedley, A.J; Leung, G.M. (2009). Hong Kong Medical Journal, 15 Suppl 9, 17-19.
[10] Lawson, A.B & Leimich, P. (1999). Approaches to the space time modelling of infectious disease behaviour. IMA Journal of Mathematics Applied in Medicine and Biology, 17, 1-13.
[11] Lena Sanders (2007). Models in Spatial Analysis. (L. Sanders, Ed.) (First., pp. 1–26). Great Britain: ISTE Ltd.
[12] Leung, P., Tran, L. , & Fast, A. . (2000). A logistic regression of risk factors for disease occurrence on Asian shrimp farms. Dis Aquat Org, 41, 65–76.
[13] Lin, H., Shin, S. S., Contreras, C., Asencios, L., Paciorek, C. J., & Cohen, T. (2012). Use of Spatial Information to Predict Multidrug Resistance in. Emerging Infectious Diseases, 18(5), 5–7. doi:DOI: http://dx.doi.org/10.3201/eid1805.111467
[14] Maleczewski, J. (2000). On the use of weighted linear combination method in GIS:Common and best practice approaches.Transactions in GIS 4:50-22.
[15] Mathema, B., Kurepina, N. E., Bifani, P. J., & Kreiswirth, B. N. (2006). Molecular epidemiology of tuberculosis: current insights. Clinical Microbiology Reviews, 19(4), 658–85. doi:10.1128/CMR.00061-05
[16] Mcnutt, L., Holcomb, J. , & Carlson, B. . (2000). 088626000015010003 #.pdf. Journal of Interpersonal Violence, 15(10), 1050–1059.
[17] Middelkoop, K., Bekker, L., Morrow, C., Lee, N., & Wood, R. (2014). 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. doi:10.1186/1471-2334-14-
221

[18] Ministry of Health Malaysia [MOH]. (2015). *The National TB strategic plan (2010-2015) and NTLP. TB/Leprosy Sector*. Putrajaya

[19] Moonan, P. K., Oppong, J., Sahbazian, B., Singh, K. P., Sandhu, R., Drewyer, G., … Weis, S. E. (2006). What Is the Outcome of Targeted Tuberculosis Screening Based on Universal Genotyping and Location? *Am J Respir Crit Care Med*, 174, 599–604.

[20] Musenge, E., Vounatsou, P., Collinson, M., Tollman, S., & Kahn, K. (2013). The contribution of spatial analysis to understanding HIV/TB mortality in children: a structural equation modelling approach. *Glob Health Action 2013*, 19266(1).

[21] Narasimhan, P., Wood, J., Macintyre, C. R., & Mathai, D. (2013). Risk factors for tuberculosis. *Pulmonary Medicine, 2013*, 1–11. doi:10.1155/2013/828939

[22] Nishikiori, N. (2011). *Tuberculosis control among high risk and vulnerable populations*. Retrieved from http://health.gov.fj/PDFs/Tuberculosis control among high risk.pdf

[23] Peterson. (2014). *Mapping Disease Transmission Risk. Enriching Models Using Biogeography and Ecology*. USA, John Hopkins University Press.

[24] Pfeiffer, D. U., & Stevens, K. (2011). *Using Data - and Knowledge - driven Approaches for Spatial Modelling* (pp. 1–44).

[25] Pfeiffer, D. U., Robinson, T. P., Stevenson, M., Stevens, K. B., & Rogers, D. J. (Techniques). (2008). *Spatial Analysis in Epidemiology United Kingdom*. Oxford University Press, USA.

[26] Ribeiro, F.K.C., Pan, W., Bertolde, A., Vinhas, S.A., Peres, R.L., Riley, L., Palaci, M., Maciel, E.L.(2015). Genotypic and Spatial Analysis of *Mycobacterium tuberculosis* Transmission in a High-Incidence Urban Setting, *Clinical Infectious Diseases*, 61, 5, 1 September 2015, Pages 758–766.

[27] Roberts, M., Andreasen, V., Lloyd, A., Pellis, L., Sciences, M., Mail, N. S., … Epidemiology, D. (2016). Nine challenges for deterministic epidemic models. *Epidemics*, 10, 49–53.

[28] Sun, W., Gong, J., Zhou, J., Zhao, Y., Tan, J., Ibrahim, A.N., Zhou, Y. A (2015). Spatial, Social and Environmental Study of Tuberculosis in China Using Statistical and GIS Technology. *Int. J. Environ. Res. Public Health*, 12, 1425-1448

[29] Taylor, N. (2003). *Review of the use of models in informing disease control policy development and adjustment* (pp. 1–80). Reading, United Kingdom. Retrieved from http://www.veeru.rdg.ac.uk/documents/UseofModelsinDiseaseControlPolicy.pdf

[30] WHO, W. H. O. (2014). Drug-Resistant Tb Surveillance & Response; Supplement Global Tuberculosis Report 2014 (pp. 1–20). 20 Avenue Appia, 1211 Geneva 27, Switzerland.

[31] Abdul Rasam, A.R., Mohd Shariff, N., Dony, J.F, & Maheswaran, P (2017). Mapping Risk Areas of Tuberculosis Using Knowledge-Driven GIS Model in Shah Alam, Malaysia, *Pertanika Journal of Social Sciences &Humanities*, 25 (S),135-144.