Mapping Risk Areas of Tuberculosis Using Knowledge-Driven GIS Model in Shah Alam, Malaysia

Abdul Rauf Abdul Rasam1*, Noresah Mohd Shariff2, Jiloris F. Dony3 and Punitha Makeswaran3

1Geography Programme, School of Distance Education, Universiti Sains Malaysia, 11800 Pulau Pinang, Malaysia
2Centre of Studies for Surveying Science and Geomatics, Faculty of Architecture, Planning and Surveying, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia
3The Ministry of Health Malaysia

ABSTRACT

Developing a model to map tuberculosis (TB) cases in Malaysia for boosting early detection is vital. A knowledge-driven geographical information system (GIS) modelling is an alternative approach developed for assessing potential risk areas of TB at Section 17, Shah Alam, Selangor. It is a weight-rating score model and spatial multi-criteria decision making (MCDM) method for producing a ranked map based on the index values and risk indicators with a five-score scale. Results showed 34.85% of the study areas are potential TB high risk zones, ranging from medium to very high risk. This is consistent with the findings obtained from overlay comparison with the current cases in 2015. The TB risk map and validation indicated a reasonable match with areas considered as potential TB risk areas, particularly in urban and crowded environments. Thus, a GIS-based MCDM technique can be applied in the national TB screening and monitoring programme.

Keywords: Disease mapping, knowledge-driven GIS model, spatial MCDM, risk area, tuberculosis

INTRODUCTION

In 2015, Malaysia had a medium burden of tuberculosis (TB) incidence according to the World Health Organization (WHO) with Selangor ranked among top tree in terms of number of cases. The Ministry of Health reported that Selangor had more than 4000 cases in 2015, where 31 percent of the cases occurred in the district of Petaling
where Shah Alam is located. Although there have been numerous attempts to control the disease comprehensively, the development of a suitable model is still challenging, especially for determining disease risk factors, spatial resolution, and modelling methodology.

In Malaysia, a spatial based predictive model for disease occurrences is not widely used. Therefore, a knowledge-driven GIS model is recommended for understanding and predicting disease incidence. This study was aimed at producing a spatial based MCDM technique for identifying potential risk areas of local TB occurrences. It posed two main research questions: i) where are potential high risk areas of tuberculosis in the study area?; and ii) why are particular potential areas been affected by the disease?

### Spatial Risk Management and Assessment of Tuberculosis using Knowledge-Driven GIS Model

Previous studies have focused on practical capabilities of geospatial technologies for disease risk assessment and management in three key functions, namely disease pattern analysis, spatial correlation and spatial prediction. Table 1 shows the integration

<table>
<thead>
<tr>
<th>Author(s)</th>
<th>Objective(s)</th>
<th>GIS-MCDM technique(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linet al., 2016 (Taiwan)</td>
<td>A novel approach for identifying sites of soil contamination. It can consider spatial uncertainty for any monitoring and or remediation initiative.</td>
<td>A decision-making approach with a MCDA, multivariate and geostatistical simulation techniques.</td>
</tr>
<tr>
<td>Lahue et al., 2016 (Spain)</td>
<td>To develop a spatially explicit modelling of animal tuberculosis at the wildlife-livestock interface for identifying trigger areas.</td>
<td>Using a spatially explicit, ecological niche models and factors such as wildlife demographics, hunting management, land use, climatic, and environmental variables.</td>
</tr>
<tr>
<td>Shen et al., 2016 (China)</td>
<td>An urban ecosystem vulnerability assessment method for decision makers and environmental manager’s applications.</td>
<td>Combining spatial context of GIS tool, MCDA method, ordered weighted averaging (OWA) operators, and socio-economic elements.</td>
</tr>
<tr>
<td>Fuller et al., 2014 (Colombia)</td>
<td>To develop a participatory risk mapping of malaria vector exposure for providing an accurate spatial representation of risk of potential vector exposure.</td>
<td>Conducting a participatory GIS multi-criteria decision analysis with expert opinion, different fuzzy functions, environmental and population factor weights.</td>
</tr>
<tr>
<td>Alcorn et al., 2013 (Mexico)</td>
<td>To develop a GIS-based volcanic hazard and risk assessment of eruptions sourced for evaluating a possible future eruption.</td>
<td>Using a GIS-based volcanic hazard tool to simulate pyroclastic fallout and density currents</td>
</tr>
<tr>
<td>Torre et al., 2012 (Europe)</td>
<td>An approach for mapping the vulnerability of European Union soils to antibiotic contamination.</td>
<td>Performing a spatial assessment or GIS based MCDM techniques (Getis-Ord Gi statistic)</td>
</tr>
</tbody>
</table>
of knowledge-driven GIS and MCDM techniques can assist health related agencies to make a better decision for risk pollution and contamination (Torre et al., 2012; Liu, et al., 2012), healthcare services (Diaby et al., 2013), hazard assessment (Alcorn et al., 2013; Satta et al., 2016) and epidemiology (Fuller et al., 2014).

Although there have been attempts to develop a generic spatial MCDM framework for epidemiology, the original procedures in the older framework are still being used (Fuller et al., 2014; Lahue et al., 2016) with appropriate local modifications. In Malaysia, this technique has been used in studies on natural disaster and ecology, but its applicability in the context of disease or health, especially for TB and lung diseases, is still limited.

Malczewski (2006) discusses the benefit of integrating MCDA technique into GIS procedures by inserting value judgments with regard to assessment criteria in the spatial decision-making processes. However, experts have suggested ways to enhance the capabilities of the technique by utilising a public GIS participatory, big data and multidisciplinary methodologies (Comes et al., 2011; Stevens & Pfeiffer, 2011; Pfeiffer & Stevens, 2015). In technical aspect, the GIS based MCDM technique uses an index model that computes the index value for each unit area and produces a rank based on the selected environmental risk factors and local standard guidelines. Weighted linear combination (WLC) is a common method for calculating the index value of the model in MCDA (Chang, 2011), in particular for incorporation factors and constraints (Malczewski, 2006). The model involves multi-criteria evaluation and depends on overlay operations for data processing as shown in Figure 1.

![Index Model](image)

**Figure 1.** Three levels of steps to build an index using the WLC method for risk criteria selection (modified from Chang, 2011)

**METHODS**

The knowledge-driven GIS modelling approach is developed using weighted linear combination (WLC) in the GIS index model and MCDM technique. These techniques, adapted from Malczewski (2000) and Chang (2011), are data acquisition, data processing and model development.
Data Acquisition and Risk Factor Selection

Data collection was between 2013 and 2014. The primary source of data was local experts’ opinion regarding the risk factors for TB transmission. Secondary data was obtained from several agencies according to environmental risk factors and maps needed. Earlier studies on global TB factors were reviewed to identify the risk factors in Shah Alam, mainly biophysical environment, demography and population, and socio-economic status of the residents (Table 2). Base map and geographical information of Shah Alam were collected from the Department of Town and Planning Selangor, Google EarthTM, and related organisations. The information related to healthcare facilities and high risk groups were extracted from TB database of the official My TB system.

Table 2
Selected indicators and risk factors of TB in Shah Alam, Selangor

<table>
<thead>
<tr>
<th>Indicators</th>
<th>Variables/Factors</th>
<th>Description/References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biophysical environment</td>
<td>Urbanisation (land use)</td>
<td>Urban and crowded situation facilitates TB infection. (Liu et al., 2012; Harling &amp; Castro, 2014; Wardani, Lazuardi, Mahendradhata, &amp; Kusnanto, 2014)</td>
</tr>
<tr>
<td></td>
<td>Types of house</td>
<td>The lower the cost of houses, the higher the risk of TB will be due to small spaces and limited ventilation</td>
</tr>
<tr>
<td></td>
<td>Distance to healthcare facilities</td>
<td>If health care facilities are located far away from residential areas, the potential of people to contract the TB is higher due to limited accessibility (Tudor et al., 2014)</td>
</tr>
<tr>
<td></td>
<td>Distance to factory locations</td>
<td>Factory is a source of air pollution and can pose great health risks</td>
</tr>
<tr>
<td>Human and population</td>
<td>Number of Population</td>
<td>Spread of TB will be more rapid in a small and crowded environment. (Zaragoza Bastida et al., 2012; Erazo et al., 2014; Harling &amp; Castro, 2014)</td>
</tr>
<tr>
<td></td>
<td>High risk groups</td>
<td>High risk groups relate to low immune people and poor lifestyles such as diabetes, no BCG vaccine, non-citizen, young and senior citizen. (WHO, 2015)</td>
</tr>
<tr>
<td>Socio economic status (SES)</td>
<td>Household income</td>
<td>TB is linked to poverty, particularly among low-medium income households. (Wang et al., 2012; Yakam et al., 2014)</td>
</tr>
</tbody>
</table>

Data Processing and Model Development

Data is processed by (i) evaluating and ranking the criteria weights proposed by experts rank sum technique in MCDM; ii) calculating the score or weights of all the sub-criteria used based on guideline and reviews using the WLC method in GIS index model; (iii) model development; and (iv) generating TB risk maps using the
developed model (Figure 2). All datasets were processed and analysed using ArcGIS ESRI software, Google EarthTM and M. Excel platform. Each criterion or risk factor was ranked in a standardised value of 5 score scale from very low risk to very high risk. This score or weight was evaluated using rank sum techniques (1), expert opinion, and weighted linear combination (WLC) methods. Four experts from the Petaling District Health Office were interviewed to rank for each criterion used. This ranking technique is the simplest and most popular method employed to quantify the importance of weights by positioning it in rank order. Meanwhile, the local experts are chosen based on their actual work experience.

The lowest rank is 1 and the highest 7. In terms of socio economic status, the lowest criteria id rank 7 and the lowest weight is 1. Table 3 contains the list of risk factors, the weighted calculation and score standardisation. For example, type of house, which is ranked as 3 in the straight rank column, the weight value is calculated as 5 (i.e. 7–3+1). The weight value is divided by the total values of weight, which is 28 (and multiply with 1) to get the normalised weight value of types of house of 0.18. To classify the threshold value of the potential risk level of TB endemic into five scale rank, related guidelines from local agencies and statistical calculation were examined. The score and the weight of each criterion were multiplied and then summed up using an overlay process in ArcGIS to get the total score or the overall index (aggregate) value for each unit area of the risk map. This total score is standardised and classified to show their risk level of the map. Score 1 and score 5 indicate the lowest and the highest risk level of the TB endemic respectively. The risk maps are validated using the current TB cases.

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

Formula (1) of rank sum weight:

- \(W_j\) = the normalised weight for the \(j^{th}\) criterion
- \(n\) = the number of criteria under consideration (\(k=1,2,3,\ldots n\)) and
- \(r_j\) = the rank position of the criterion each of the criterion is weighted (\(n - r_k + 1\)) and then normalised by the sum of all weights and that is \(\sum - r_k + 1\).
RESULTS AND DISCUSSION

The risk factors of Tuberculosis

The ranking of TB risk factors is important in this procedure. Seven risk factors were identified in this study (Table 3). Figure 3 displays the overall result of risk level (from 0 to 100). Results showed population (25%) and concentration of high risk group (21%) are the most important risk factors. It is followed by biophysical environment and socio-economic status that consists of type of house (18%), distance of factory from the house (14%), land use (11%), distance of healthcare facilities from housing location (7%) and household income or SES (4%).

Population is the main indicator of TB endemic in Shah Alam. The pattern of local TB dynamics in Peninsular Malaysia is also likely driven by human risk factors than by ecological risk factors (Abdul Rasam et al., 2016). Population-based factors is a common risk variable of TB. This is similar to TB-prone areas such as Latin America, Asia, and Africa because overcrowding can trigger such diseases (Zaragoza Bastida et al., 2012; Erazo et al., 2014; Harling & Castro, 2014; WHO, 2015). The WHO (2015) explains the characteristics of a person who is most at risk of TB: young adult, located in developing countries, HIV infected people with impaired immune system, and smoking. Human mobility are connected with the population factor which triggers TB (Lynn Feske, 2011; Nava-Aguilera et al., 2011).

### Table 3

The score standardisation of the TB risk factors derived from the experts’ opinion

<table>
<thead>
<tr>
<th>No</th>
<th>Risk Factor/Criteria</th>
<th>S_Weight</th>
<th>S_Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>No_Population</td>
<td>0.25</td>
<td>25%</td>
</tr>
<tr>
<td>2</td>
<td>Concentration_High Risk Group</td>
<td>0.21</td>
<td>21%</td>
</tr>
<tr>
<td>3</td>
<td>Type_House</td>
<td>0.18</td>
<td>18%</td>
</tr>
<tr>
<td>4</td>
<td>Distance_Built Up (Factory)</td>
<td>0.14</td>
<td>14%</td>
</tr>
<tr>
<td>5</td>
<td>Type_Landuse</td>
<td>0.11</td>
<td>11%</td>
</tr>
<tr>
<td>6</td>
<td>Distance_Healthcare Facilities</td>
<td>0.07</td>
<td>7%</td>
</tr>
<tr>
<td>7</td>
<td>Status_Socio-Economic, SES</td>
<td>0.04</td>
<td>4%</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>1</td>
<td>100%</td>
</tr>
</tbody>
</table>

Figure 3. Risk level (percent) of the TB factors according to the selected experts’ opinion in Shah Alam
The Risk Areas of Tuberculosis

Determining the potential areas of the TB endemics in Section 17, Shah Alam is important identifying future endemics in a spatially explicit way. Figure 4 shows the potential high risk areas, 34.85%. The overall situation of TB in Section 17 is still considered as medium risk and under control. In terms of geographical characteristics, the risk areas are more populated, urbanised and crowded due to rapid human concentration and mobility between Shah Alam and crowded cities such as Klang. The combination of population and ecological risk factors are indicators of high risk areas as shown in Figure 5.

Most of the locations at risk in Section 17 share similar risk factors such as being located in a crowded urban area (Liu et al., 2012; Wardani et al., 2014; Harling & Castro, 2014; Zaragoza Bastida et al., 2012; Erazo et al., 2014; Harling & Castro, 2014) and their proximity to factories and commercial areas. They could also come from low and medium income levels (Wang et al., 2012; Yakam et al., 2014; WHO, 2015). Interestingly, the study found the distance of facilities and type of house are
likely not linked with to greater likelihood of TB incidents although these factors are the main contributors of spread of TB worldwide.

In terms of map validation, the model accuracy for Section 17 is relevant as most of the current cases are located in the potential risk areas. However, there are pockets of high risk areas within low risk areas due to spatial uncertainty. This study does not only show environmental landscapes in high risk areas, it has also demonstrated the capacities of a GIS-based MCDM technique to develop a predictive model for identifying TB prone areas as well as prevention programme in the study area.

**CONCLUSION**

A multidisciplinary approach to control and prevent TB comprehensively is needed. Apart from bio-medical approaches, a GIS-MCDM based model can be also used as a surveillance tool in order to assist the health staff to establish an intervention programme on the sites. The findings from the risk map showed the level of potential risk in Section 17 is a medium rate. The spread of TB in the urban and populous environments are mainly
due to human risk factors than by ecological risk factors. The mapping model derived from this GIS context could be used for targeting specific TB risk areas and having a screening programme. It is also suggested that other risk factors and GIS based MCDM techniques need to be explored for more dynamic result and prediction of localised disease risk transmission.

ACKNOWLEDGEMENT
The authors gratefully acknowledge the Ministry of Health of Malaysia and the Selangor State Health Department for providing TB data. Universiti Sains Malaysia (USM) is also acknowledged for providing funds to undertake this study. This paper is part of a PhD thesis and an outcome of from the Postgraduate Colloquium School of Distance Education, USM 2016.

REFERENCE


Lin, W.-C., Lin, Y.-P., & Wang, Y.-C. (2016). A decision-making approach for delineating sites which are potentially contaminated by heavy metals via joint simulation. Environmental Pollution (Barking, Essex: 1987), 211, 98–110. doi:10.1016/j.envpol.2015.12.030


