GEOSPATIAL MODELLING FOR POTENTIAL HIGH RISK TUBERCULOSIS AREAS IN SHAH ALAM, SELANGOR

ABDUL RAUF BIN ABDUL RASAM

UNIVERSITI SAINS MALAYSIA

2018

GEOSPATIAL MODELLING FOR POTENTIAL HIGH RISK TUBERCULOSIS AREAS IN SHAH ALAM, SELANGOR

by

ABDUL RAUF BIN ABDUL RASAM

**Thesis submitted in fulfilment of the requirements
for the degree of
Doctor of Philosophy**

March 2018

ACKNOWLEDGEMENT

Alhamdulillah, in the name of Allah, The Most Merciful. All praises to Allah, upon His Messenger for giving me strength and determination to successfully complete this PhD thesis. I would like to express my genuine gratefulness to a number of people who have contributed to the whole thesis process from its beginning track to the end of the journey. Firstly, countless thanks for my supervisors, Associate Proffesor Dr Noresah Mohd Shariff from Universiti Sains Malaysia (USM), and Dr Jiloris F. Dony, former of Sector Head of TB and Leprosy, Ministry of Health (MOH), Malaysia. They have been generous and supportive advisors especially in providing valuable guidelines and advices throughout this thesis expedition.

Secondly, I would also like to express my gratitude to Universiti Teknologi MARA (UiTM) and Ministry of Higher Education (MOHE), Malaysia for providing a staff scholarship (SLAI), enabling me to pursue my PhD in Geospatial Modelling and Geovisualistion. A warm appreciation also goes to my research colleagues from multidisplinary fields, especially from USM, UiTM, Unit of TB and Leprosy from MOH Malaysia, Selangor State Health Department (JKNS), District Health Office (PKD), Petaling for their ideas and practical contribution to this thesis. Cooperation and information from local organisations such as Selangor Town and Rural Department, Shah Alam City Council (MBSA), Selangor Statistic Department, and previous scholarly researchers are also very much appreciated.

Thirdly, I express my deep gratitude to my beloved family, especially my late parents, Abdul Rasam Jikiri and Indah Rujah Mad Abdul who continue to be my inspiration especially in achieving my ambition to further my studies. Besides them, I am indebted to may siblings, in-laws and friends, particularly Siti Falmah, Siti Asiah, Siti Salmah, Siti Jahala, Siti Kalsum, Nurizam and Siti Hanina who have given me their prayers, encouragement, and support throughout this challenging journey. Most importantly, a bundle of appreciation goes to my lovely wife, Junainah Askalani (Putrie) for her patience, moral support, love, encouragement and inspiration that greatly facilitated the accomplishment of this intellectual work.

Personally, I would also like to say that this thesis has given me a million memories and knowledge to further explore geospatial health in Malaysia. Thus, the thesis is believed could be beneficial to academicians and society, particularly people who deal with medical geography and health sciences. Based on the philosophy of spatial epidemiology that a disease can be controlled through a decent environmental management and the applications of a geospatial or geomatic technology, several of innovative prototypes in this study have been developed namely MyTB Risk Index© and GeoEpiMap© for TB control programme. Besides this, the success of community (USM, UiTM and JKNS) project conducted at the Shah Alam hospital on September 30, 2017 with the theme of "Education, Detection and TB Screening through Innovation: GeoEpiMap© of TB and MoSco 2.0", is hoped to open eyes of decision makers and the communities to adapt the technology in the National TB Control Programme and their daily lives, in general.

TABLE OF CONTENTS

CHAPTER 2: SPATIAL EPIDEMIOLOGY AND MODELLING OF TUBERCULOSIS

CHAPTER 3: RESEARCH METHODOLOGY AND MODELLING FRAMEWORK

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

bTB	Bovine Tuberculosis
CDC	Centres for Disease Control and Prevention
CDR	Cases Detection Rate
DC	Communicable Disease
DOTS	Directly Observed Treatment (for Short-Course of TB)
DV	Dependent Variable
EM	Epidemic Model
GeoEM	Geospatial Epidemic Model (Proposed Model)
GIS	Geographical Information System
HCW	Health Care Worker
HRTBAs	High Risk Tuberculosis Areas (Proposed Operational Term)
IDW	Inverse Distance Weighted
IV	Independent Variable
JKN	State Health Department (Jabatan Kesihan Negeri, Selangor)
KAP	Knowledge Attitude and Practice
MOH	Ministry of Health (Malaysia)
*M*TB	*Mycobaterium* Tuberculosis, *M.* Tuberculosis,
PKD	Petaling Health District Office (Pejabat Kesihatan Daerah, Petaling)
PTB	Pulmonary Tuberculosis
SE	Spatial Epidemiology
SES	Socio-economic Status
SEIR	Susceptibility, Exposure, Infection, Removal
TB	Tuberculosis
WHO	World Health Organization

PEMODELAN GEOSPATIAL DALAM MENENTUKAN KAWASAN BERPOTENSI BERISIKO TINGGI PENYAKIT TUBERKULOSIS DI SHAH ALAM, SELANGOR

ABSTRAK

Pertubuhan Kesihatan Sedunia (WHO) melaporkan Malaysia mempunyai kadar beban penyakit batuk kering atau tuberkulosis (TB) yang sederhana, tetapi kes semasa menunjukkan penyakit ini masih meningkat. Kementerian Kesihatan Malaysia (KKM) telah mengadakan beberapa garis panduan untuk mengawal penyakit ini, namun laporan teknikal TB kebangsaan pada tahun 2015 masih menekankan kelemahan kaedah sedia ada untuk mengesan kes TB di lapangan. Kajian ini dijalankan untuk membangunkan model geospatial dalam menentukan kawasan berpotensi berisiko tinggi TB sekali gus menyasarkan kumpulan berisiko terutamanya kes yang gagal dikesan dan tidak didiagnosis. Kajian ini mempunyai empat objektif utama iaitu: i) untuk mengkaji corak dan kelompok bagi taburan spatial TB, ii) untuk menentukan faktor-faktor risiko yang menyumbang kepada kes TB, iii) untuk membangunkan model epidemik geospatial (GeoEM) dalam menentukan kawasan yang berisiko tinggi TB, dan iv) untuk mengaplikasikan dan mengesahkan model GeoEM. Shah Alam di daerah Petaling dipilih sebagai kawasan kajian kerana ia mencatatkan kes TB yang konsisten dan mempunyai ciri-ciri persekitaran yang berkaitan dengan trasmisi jangkitan penyakit TB. Model epidemik geospatial (GeoEM) telah menggunakan kaedah spatial (SE), sistem maklumat geografi (GIS), GIS-membuat keputusan pelbagai kriteria (MCDM), regresi logistik dan geostatistik. Dapatan kajian telah menunjukkan corak taburan TB di Shah Alam adalah sederhana serta terdapat kelompok kes TB di sekitar zon utara, zon tengah

dan beberapa di zon selatan. Kelompok ini tertumpu di 10 seksyen terutamanya di U17, U18, U19, U20, S7, S17, S18, S20, S27 dan S28. Kawasan berpotensi berisiko ini mempunyai ciri-ciri geografi yang sama seperti yang berlaku di negara-negara yang mempunyai beban TB yang tinggi di dunia. Selain itu, dapatan ini telah menjana atribut baru terhadap faktor risiko dan taburan TB tempatan di mana penyakit ini telah tersebar ke zon utara-timur akibat pengaruh pembangunan bandar baru dan pergerakan manusia. Tujuh faktor risiko yang mempengaruhi TB tempatan telah dipilih berdasarkan turutan tahap risiko masing-masing, iaitu kawasan menjadi tumpuan pergerakan pesakit, keberadaan kumpulan berisiko, status sosio-ekonomi dalam sesebuah rumah, bilangan orang dalam rumah, jenis rumah, jarak dari kilang dan kawasan bandar. Walaupun setiap faktor ini memberi kadar risiko yang berbeza terhadap kejadian TB, namun gabungan faktor-faktor ini akan memberikan impak yang lebih besar lagi kepada keseluruhan penumpuan risiko TB. Kajian ini juga mendapati faktor berasaskan manusia lebih mempengaruhi risiko TB tempatan berbanding dengan faktor bio-fizik persekitaran seperti risiko TB bagi satu lokasi akan meningkat sebanyak hampir 4 kali jika seseorang itu tinggal bersama dengan individu yang mempunyai penyakit TB dalam jangka masa tertentu. Pemodelan GeoEM dapat menganggarkan 65 % kawasan adalah berisiko tinggi TB (HRTBAs) dan menyasarkan 102 lokaliti sebagai kelompok berisiko TB yang terletak di beberapa seksyen yang berisiko. Penemuan menarik juga mendapati kebanyakan kes TB berlaku di kawasan perumahan yang bertingkat, berhampiran dengan jarak kilang perindustrian, dipenegaruhi oleh corak mobiliti pesakit, diduduki oleh golongan berisiko dan status sosio-ekonomi rendah, tinggal di kawasan yang ramai dan di kawasan bandar. Kesimpulannya, model ini mempunyai keupayaan analisis yang munasabah dalam menentukan kawasan berpotensi untuk berisiko tinggi TB di Shah

Alam melalui pemetaan taburan, analisis dan pemodelan secara geospatial. Keunikan model ini terletak pada penggunaan teknik bersepadu dan penghasilan peta berisiko TB yang holistik di kawasan kajian. Model ini boleh digunakan untuk mencadangkan lokaliti berisiko dalam menjalankan program saringan TB. Selain itu, untuk meningkatkan pemahaman pekerja kesihatan terhadap situasi sebenar TB dan untuk menilai tahap faktor risiko terhadap penyakit ini mengikut persekitaran tempatan.

GEOSPATIAL MODELLING FOR POTENTIAL HIGH RISK TUBERCULOSIS AREAS IN SHAH ALAM, SELANGOR

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 increasing. The Ministry of Health (MOH), Malaysia has set up several guidelines to control the disease, however, the national TB technical report in 2015 highlighted that existing detection methods of TB on the site still need to be improved to strenghten the current TB control programme. A geospatial model is proposed in this study to identify potential high-risk areas of TB and targeted risk population especially for missing cases and undiagnosed people. This study has four main objectives: i) to examine the spatial pattern and clustering of TB distribution, ii) to determine the influential risk factors contributing to local TB cases, iii) to develop a geospatial eopidemic model (GeoEM) for potential high-risk TB areas, and iv) to validate the model of GeoEM for actual application. Shah Alam in the district of Petaling is selected as a study area since it has recorded constant TB cases and it also has a diversity of environment related TB risk factors. GeoEM is innovatively developed using spatial epidemiology (SE) approach, geographical information system (GIS), GIS-multicriteria decision making (MCDM) method, logistic regression and geostatistical method. The overall spatial pattern of TB in Shah Alam is a slight medium random that exists in certain clustered areas. TB clustering was concentrated around northern zone, central zone and a few areas around southern zone. There are 10 main sections from 47 sections which have significant relationships with the clustering. It comprises section U17, U18, U19, U20, S7, S17,

S18, S20, S27 and S28. These risk areas have similar geographical features as occurred to the high burden countries worldwide. The results stimulate new attribue of risk factor and interpretaion on the disease that is scattered towards north-eastern zone due to the new township development and human mobility. Seven influential risk factors with the local TB are selected and sequently ranked from human mobility, high risk group, 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 such as a location of TB risk will be increased by 4 times if individuals are living together with people who have TB disease for a particular time period. GeoEM has estimated a 65% of potential high risk TB areas (HRTBAs) and targeted 102 high risk localities in the risk sections. These risk localities have general similarities with other endemic areas worldwide. Interesting findings also revealed that most of these cases occurred in high-rise housings, close to industrial location, influencing mobility of human pattern, located at low socio-economic or high risk group, living in a crowded house and an urban city. In conclusion, this proposed GeoEM model has appropriate capabilities in estimating potential risk areas of TB in Shah Alam through spatial pattern, spatial analysis and spatial modelling. The unique part of the model is that it is an integrated technique for a holistic risk map of HRTBAs. The model could be applied to propose clustered localities of local TB for intervention programme in order to enhance the understanding of the local staff to explain the real situations of the TB, and to study a risk-factor assessment of the disease according to a particular area.

CHAPTER 1

INTRODUCTION

1.1 Introduction

This chapter provides an overview of central issues of the study to identify potential high risk tuberculosis areas in Malaysia using a geospatial model. Currently, estimating the risk areas is a crucial agenda of the Ministry of Health, Malaysia (MOH) since there are missing and undiagnosed TB cases. Geospatial model is a GIS model to simplify the representation of geographical reality of TB dynamics. A comprehensive background and research problem statement are discussed in this chapter. It includes the descriptions of research objectives, relevant research significance and scope of the study. In theoretical background, a global and local review in TB epidemiology is also specifically discussed to understand the current situation and its possible solution in the context of local TB surveillance plan and spatial insights.

1.2 Research Background

Malaysia has a medium burden of TB incidence based on WHO indicator, but the trend of TB cases is still generally increasing from 16 665 in 2006 to 25 739 in 2016 (MOH, 2017). Selangor is among the top-three highest TB cases in the country with more than 4000 cases reported from 2013 to 2016 (MOH, 2017). Therefore, the Ministry of Health (MOH), Malaysia has set up several guidelines to control the disease 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 (2016), 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 (2015).

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, Nishikiori, (2011) 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. Missing cases and undiagnosed patients of TB cannot be identified if real or actual cases are not reported in the official system.

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 (such as high-risk group and socio-economic status), but also are driven by environmental risk factors such as land use, human movement and housing condition. Consequently, this method needs to be combined with other techniques (Burgos and Pym, 2002; Mathema et al., 2006; Narayanan, 2004) in order to improve management of cases and analytical power. Narayanan, (2004) has also suggested that the combination between molecular techniques and geographical techniques can enhance the TB transmission analysis for identifying different geographical areas in a high-burden disease areas and subsequently enhancing targeted screening efforts. As such, this study proposes geospatial model as an alternative method to limit the spread of TB and to detect potential risk areas. This proposed model can identify potential high-risk geographical area of clustered TB and characterise the areas with TB potential risk areas.

Unfortunately, some health staff in the country expressed their doubts towards geospatial capabilities to control and prevent the disease because the model is quite a new tool to them. However, after giving a brief explanation on geospatial application in TB management, some health experts in Selangor favoured to have some confidence on the geospatial capabilities to manage the local TB.

Interviews with the local experts on TB from State Health Department (JKN), Selangor have given an insightful view into development such predictive model in Petaling using geospatial techniques. These experts believed that the proposed model may not only benefit for tuberculosis control but also for other high-burden diseases in the state as well as the country. Nurwanida (2017) as head unit of TB Surveillance in JKN, Selangor also agreed that the major challenges to efficiently conduct case detection rates (CDR) for TB is to locate the high burden TB areas. Besides that staff workload and lack of TB detection facilities (e.g. Mobile X-ray) can be reduced on the field. The proposed model of GeoEDM is able to reinforce the existing TB detection and cure rate performance in the state which consequently reduces the staff workload and financial constraints geographically.

In order to understand the behaviour of the disease, spatial epidemiology (SE) of TB needs to be clarified properly. Generally, SE is empirically applied to provide integrated spatial frameworks for enhancing current technical issues on case detection of a disease. SE is a fundamental study to thoroughly grasp the current scenario of a disease (such as TB) either in global or local perspective. The core study includes the spatial analysis (spatial pattern and spatial risk correlation) and spatial modelling as a local disease surveillance management tool.

Technical aspects in SE, GIS and TB epidemiology are combined to develop a geospatial model for identifying potential high-risk TB areas in the study area. For example, a main application of molecular techniques in the study of TB epidemiology is to study of *Mycobaterium tuberculosis* transmission dynamics through laboratory stray detection and geographic spread determination (Burgos and Pym, 2002). Meanwhile, a geospatial model can perform a better spatially decision-making system for public health management (Torre et al., 2012; Fuller et al., 2014) to estimate potential high-TB risk areas (Daley and Gani, 2005; Lin et al., 2012; Middelkoop et al., 2014; Moonan et al., 2006; Musenge et al., 2013; Wei et al., 2016)

1.3 Research Problem Statements

In Malaysia, the trends of TB cases are increasing. Currently, Selangor has been reported to have the second highest TB cases in the country. The other two states with a high number of TB cases are Sabah and Sarawak. Similar trends occur in Selangor, where the pattern of TB cases for five current years slightly increased in 2014. The district of Ulu Langat and Petaling recorded the highest notification rates and cases in the state with 16% and 32% of the occurrences correspondingly. Detecting missing cases and undiagnosed patients of TB are two challenges faced by the State Health Department against the increasing case detections rate (CDR) and reducing mortality of TB.

Currently, 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 example, TB Unit in the Petaling Health District Office has even been using a Google Map and M. Excel in identifying static clustered areas of local TB cases. Azhar Shah et al., (2002) and Norman et al., (2011) emphasised on the descriptive analysis on the geographical distribution and the trend of tuberculosis in Cheras and Shah Alam using secondary data of TB.

Besides these, global studies have also studied risk areas and vulnerable population of TB using various techniques (Lin et al., 2012; Middelkoop et al., 2014; Moonan et al., 2006; Musenge, Vounatsou, Collinson, Tollman, Kahn, et al., 2013; Wei et al., 2016: Sun et al., 2015; Patterson et al., 2017; Shaweno et al., 2017). Current geospatial-based model has been developed in estimating tuberculosis incidence, risk areas and to identify ‘at-risk’ population (Sun et al., 2015; Patterson et al., 2017; Shaweno et al., 2017). Although these studies have demonstrated the power of integrated method according to TB epidemic stages, two meticulous criteria need to be considered to produce an accurate epidemic modelling of the risk areas of the disease as recommended by scholarly experts in SE.

The criteria include understanding the biological transmission of diseases (Childs et al., 2015; Lloyd Smith et al., 2015; Pellis et al., 2014) and applying multidisciplinary approaches (Garner and Hamilton, 2011), including geospatial techniques. The first criterion is such a challenge to be achieved since there is no single standard on the risk factor preferences for all countries in the world even though the second criterion is always explored by experts by dealing with an interdisciplinary and integrative modelling for high risk TB areas. As suggested by scholars, geospatial model is expected to produce a better prediction in identifying risk areas which occur in real situations (Garne and Hamilton, 2011; Lena Sanders, 2007; Shaweno et al., 2017).

1.4 Research Questions

The research questions for this study are as follow:

i. What are the current spatial patterns and clustering areas of TB cases in the study area?

ii. What are the factors causing the current cases of TB in the study area?

iii. How much of the selected risk factors can substantially affect local TB cases and risk areas?

iv. What are the application capabilities of the proposed geospatial model in identifying the potential high-risk TB areas in the study area which can be used by local health organisation for TB prevention?

v. How good and realiable is the proposed geospatial epidemic model?

1.5 Research Aims and Objectives

This research is mainly aimed to model potential high-risk TB areas in Shah Alam using geospatial model in limiting the spread of TB and screening programme on the site. This study was conducted with the following four main objectives:

i. To examine the spatial patterns and clustering of TB distributions in Shah Alam, Selangor for three-year trends (2013 to 2015).

ii. To identify the risk map based on influential risk factors contributing to local TB cases in Shah Alam.

iii. To develop a geospatial epidemic model (GeoEM) for potential high-risk TB areas in Shah Alam.

iv. To validate the geospatial epidemic model (GeoEM) for actual application.

1.6 Research Significance

This study is mainly conducted according to the national and WHO's agenda in order to enhance the current TB control and prevention programme in handling cases, management of contacts and implementation of treatments (MOH, Malaysia, 2015; WHO, 2017). The purpose of the study is also to achieve the goal of Regional Strategy to Stop TB in the Western Pacific Region for 2011-2015 by reducing the prevalence and mortality due to all forms of TB in a country. Currently, one of the main agenda of TB unit in Selangor is to strengthen cases detection rate (CDR) activities.

Specifically, this study examines the existing pattern distribution and influential risk factors in the study areas, and to generate knowledge and understanding on local spatial epidemiology of TB. Quantifying environmental risk factors of disease risk pattern can be a challenging task in Shah Alam due to spatial environmental influences, transmission process, while each area may have its own unique risk factors and dynamics. The task is to identify the localised risk factors and to produce a risk map. First is to examine spatial characteristics of TB landscape ecology that is crucial for studying the relationships between TB ecological processes and environmental factors, in particular, the ecosystems. Second is to generate a risk level of TB risk factors in the study area. The model can also easily identify the possible factors influencing local TB cases.

The technical framework in the proposed model could become a flexible capacity prototype to support existing disease management information system, reducing staff workloads and technical cost. Public and health staff may use the model to indicate TB disease pattern and to share with others for better public health awareness and promotion platform. Geospatial capabilities have been theoretically proven in health and environmental applications, particularly in characterising risk factors of the disease (H. Lin et al., 2012; Middelkoop et al., 2014; Musenge et al., 2013; Oliveira et al., 2013; Wei et al., 2016). Lin et al. (2012) demonstrated the capabilities of spatial model to improve risk prediction for multidrug-resistant (MDR) TB in locations where accessibility is not available to provide drug-susceptibility testing. More importantly, the proposed model could facilitate the local decision makers in the health sectors to plan the strategy in combating the epidemics in the country as addressed in the National Tuberculosis Control Strategic Plan (2016-2020) towards a National Early Warning Surveillance System of TB.

1.7 Research Scope and Limitation

This study was conducted to develop a geospatial model for estimating potential high-risk TB areas in Shah Alam, Selangor. The aspects investigated were the influential risk factors and their concentration areas using geospatial approach. The scope of the study area was Shah Alam which included 47 sections as defined by District Health Office (PKD), Petaling for its local disease surveillance monitoring system only. There are Sections S1 to S24, Sections S26 to S28, and Section U1 to U20. This study covered all forms of TB cases as reported in the MyTB system from January 2013 to December 2015 as a population of the study.

All 923 cases were identified from the myTB system but only cases with household income were used in the development of proposed model. The cases selected included pulmonary TB as well as extra pulmonary TB. In this study, similar risk factors were estimated for both types. This study used only eight variables or risk factors that covered land use/urbanisation, type of house, proximity of healthcare facilities, proximity of industrial zone, number of people in the household, social economic status, concentration of risk group and TB patient mobility.

Meanwhile, spatial-statistical analysis (clustering, variation and interaction), GIS-multi-criteria decision method (MCDM), binary logistic regression method, and geostatistical method main techniques were applied in this study to achieve the objectives. Tobler's theory and spatial diffusion were also applied to define the spread pattern and risk point concentration of TB cases on the site. In addition, although this study utilised the concept of epidemiology, biomedical or clinical approaches were not applied except for extracting knowledge from health experts on TB risk factor selection and steps of epidemic modelling.

1.8 Operational Definition of Terms

The key terms that were used in this study are defined in the following sections.

1.8.1 Geospatial

Geospatial is related to data that have implicit or explicit association with a location or spatial relative to the Earth (ISO/TC 211, 2016). It also deals with technology and analysis for gathering, display, and manipulation of imagery, GPS, satellite photography and historical data. The data are described explicitly in terms of geographic coordinates or implicitly, in terms of street address, postal code, or forest stand identifiers as they are applied to geographic models (Smith et al., 2016)

1.8.2 Model

In conceptual view, a model is a diagrammatic process model that is usually presented in a graphic map (scale) or a flowchart, while in the mathematical perspective, it allows prediction either by using probabilistic technique (use probability or randomness theory) or deterministic technique (behaviour controlled by natural laws) (Usery, n.d).

1.8.3 Geospatial Model

Generally, a geospatial model is a simplified representation of geographic reality (Chang, 2011). It can be classified into two types; spatial and process. Spatial model is generally presented in static and model distributions such as maps, GIS databases, and cartographic models (based on map algebra), while process model can be either static or dynamic and model processes, including growth or accumulation (such as urban growth, climate change, sea level rise) and flows (spatial interaction, gravity model, location-allocation) (Usery, n.d). Since this study proposed a geospatial model related to dynamic and model process of TB transmission and risk factors, thus the geospatial model here is also developed based on process model.

1.8.4 Tuberculosis (TB)

TB is an airborne disease caused by the bacterium mycobacterium tuberculosis (M. tuberculosis). This study focuses on all forms of human TB from *M. tuberculosis* that is carried by airborne particles, called droplet nuclei. Infectious droplet nuclei are generated when a person who has pulmonary or laryngeal TB disease is coughing, sneezing, shouting or singing. (CDC, USA, 2016).

1.8.5 Components of Risk Factors of TB

There are four components of risks that determine the probability of transmission of M. tuberculosis as illustrated in Table 1.1. These components and the transmission are important to explain the TB epidemic stages from individual to population/area unit measurement (CDC USA, 2016). It can be said that risk factors can define resources reservoir, potential risks and then map a community for health risk assessment. This study combines the main risk factors of TB into two indicators which are human-population factors (susceptibility and infectiousness) and bio-physical factors (environment and exposure) to define high-risk areas in the study area.

Table 1.1: Components of Epidemic Transmission for TB Risk Factors

Components of TB Risk Factors	Possibility of Transmission
Susceptibility	Susceptibility or immune status of the exposed person
Infectiousness	Infectiousness of the person with TB disease is directly connected to the number of tubercle bacilli that he or she expels into the air.
Environment	Environmental factors that affect the concentration of M. tuberculosis organisms such as space, air and ventilation system.
Exposure	Proximity, frequency, and duration of exposure. For example, the longer the duration of exposure, the higher the risk for transmission

Source from CDC USA, 2016.

1.9 Organisation of Research Thesis

This thesis is presented in seven chapters. **Chapter 1** sets out the background of the study which include the main problem statements, the aim and objectives, scope of the research and the structure of the thesis. **Chapter 2** reviews the previous and current studies related to the central topic on spatial epidemiology and geospatial modelling of tuberculosis (TB). Spatial epidemiology of TB in Malaysia needs to be explored comprehensively to identify well the current problem and to enhance future plan of TB cases detection and surveillance system on the sites.

The research methodology of geospatial modelling is explained in **Chapter 3** which is divided into several sub-sections namely, theoretical and conceptual research framework, description of the study area, data collection, data processing and analysis, and modelling development. Ethical implementation and uncertainty are also considered in this study.

Chapter 4, 5, and 6 present the results and discussions obtained from the data analysis which was done according to the objectives of the study. The discussions about the spatial pattern analysis of TB distribution, environmental risk factors of TB and modelling development using a geospatial model for high risk TB areas. The final part is **Chapter 7** which summarises the main findings of the study and identifies the fields for further improvement. Significant implications are discussed in this last chapter towards local knowledge and society contribution of spatial TB epidemiology.

1.10 Summary of Chapter

Chapter one introduces the research background, research statements, research questions, research objectives, research significance to specifically explain the real situation and current concern of the study. Operational definition of significant research terms, research scope and thesis organisations are also included in this chapter.

Research background explains the current scenario of TB both of global and local perspectives, developing geospatial modelling for identifying potential high risk tuberculosis areas. Main issues and justifications of the local study are raised by several main questions, particularly in the aspects of the spatial pattern of latest TB distribution, spatial risk variation of local TB and the application capabilities of geospatial model for risk areas.

From these crucial phenomena, four objectives of the research and their significances are determined according to local requirements. The first two objectives emphasise on analysing spatial pattern and risk factors influencing the local TB cases, and the two last objectives focus on modelling and validation of the risk map of TB using geospatial approach. Each objective contributes to research novelties in terms of knowledge enhancement, an integrated control tool, and society implications.

Some terms used in the research are also defined operationally in parallel with the study context, such as geospatial model, TB, and risk factors. Besides these, limitation of study is explicitly described in the section of scope. The final section deals with the structure of thesis, from Chapter 1 to Chapter 7. By referring to this introduction of research, a reader can generally capture the central issues and solutions proposed in this study for identifying potential risk areas of TB in Shah Alam using a geospatial model.

CHAPTER 2

SPATIAL EPIDEMIOLOGY AND MODELLING OF TUBERCULOSIS

2.1 Introduction

This chapter reviews the well-established studies that relate to the central topic on geospatial modelling for high risk tuberculosis areas using a GIS and spatial epidemiological (SE) approach. The reviews include the major issues, solutions and theories deal with TB transmission, spread, spatial pattern, risk factors and plan for TB control programme. Generally, spatial epidemiology is a subfield of health geography which is commonly required to understand the actual scenario of an infectious disease in a particular area. This context of study, the knowledge and practical aspects of SE are utilised to examine the spatial pattern, spatial variations and geospatial modelling for local TB cases.

SE is applied to provide integrated spatial frameworks or models (such as mapping, modelling and information management) for enhancing current technical; issues on case detection of TB, contacts and intervention activities on the sites. The aim of this study is to develop a geospatial modelling for high risk TB areas by answering the research questions: i) What are current spatial pattern of TB cases? ii) What are factors contributing to the current TB cases? and iii) What are capabilities of the geospatial-based model to detect the potential high risk TB areas? Geospatial model provides an appropriate framework to consider the possible input intervention for identifying potential risk TB areas and towards formulating a health policy.

2.2 State of the Art of TB Challenges and Solutions in Developing a Geospatial Epidemic Model: Global to Local Perspective

Spatial epidemiology is a fundamental study to thoroughly grasp the current scenario of a disease (such as TB). The core of the study includes the spatial analysis (spatial pattern and spatial risk correlation) and spatial modelling as a local TB surveillance tool (Figure 2.1). This study revealed that the central issue of the TB scenario is lacking of integration and multidisciplinary approach to control and prevent the disease. Globally, World Health Organization (WHO) has also explained the issues and strategies to control and eradicate the TB which is stated in the Global TB Report 2016.

The strategies include a TB care and prevention, financing, universal health coverage, social protection and determinants, research and development, a new era of global TB monitoring and services, diagnosis and treatment, HIV-associated TB and drug-resistant. These strategies are also adapted in Malaysia to face the challenging tasks in fighting the local TB dynamincs. One of the current challenges of TB control is to enhance the existing method to detect missing cases and contacts of TB on the field (MOH, 2015). The missing cases can be defined as those people do not get legitimate care and they are "missed" by healthcare systems such as undiagnosed, untreated and unreported group. Therefore, if they fail to be treated properly, they can contribute to drug resistance, fatality or continue to be sick and transmit the disease to others.

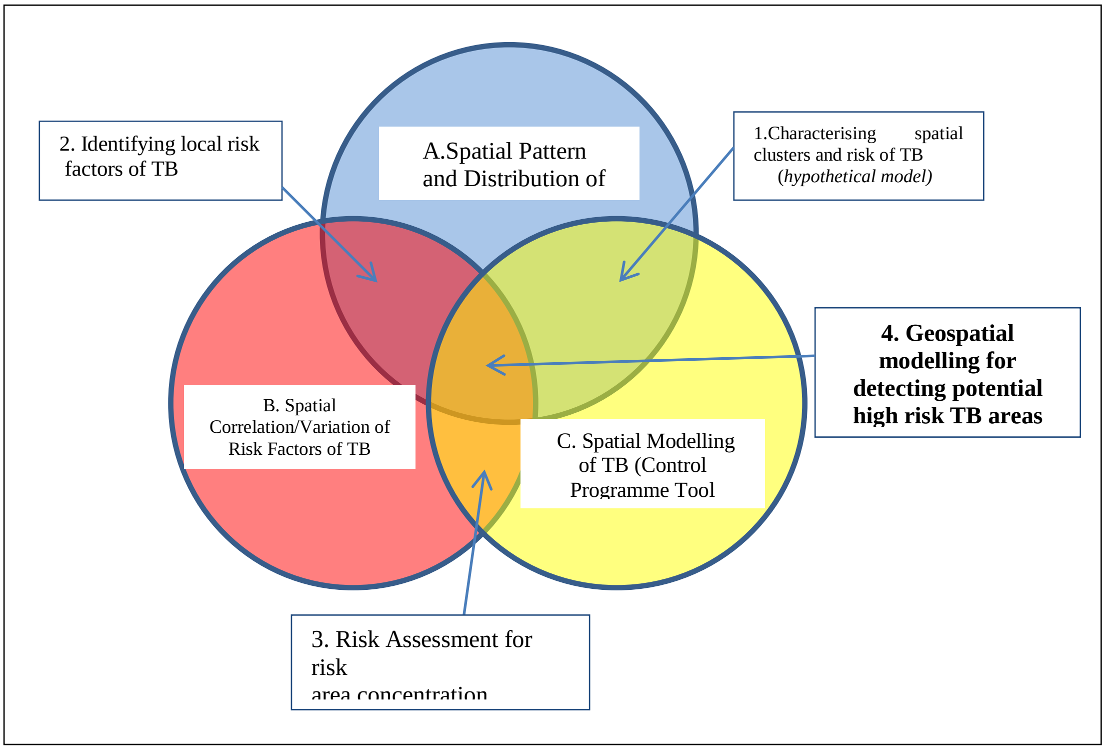

Figure 2.1: Investigating 3 Core Studies (A, B, C) in a Spatial Epidemiology (SE) to Define the Central Issues of Study (gap of field) and Develop a Proposed Geospatial Modeling (Gap of approach) through Main Activities (1, 2, 3, 4) for Detecting Potential High Risk TB Areas (Modified from Pfieffer et al., 2008)

Therefore, this study proposed a dynamic approach to solve the problem by developing a geospatial model for identifying potential high risk TB area in Shah Alam. The model can be used as a TB control tool for investigating, treating and preventing the disease in geographical perspective. According to global experts, two challenging aspects need to be considered to produce an accurate epidemic modelling of a disease; a proper understanding of biological risk transmission (Childs et al., 2015; Lloyd Smith et al., 2015; Pellis et al., 2014) and applying multi-disciplinary approaches (Garner and Hamilton ,2011). The examples of approches are spatial techniques, statistical or mathematical method, social network analysis and agent based model.

In the context of this study, the first challenge deals with determining a proper local transmission using a theory and real data, aetiology, pathogenesis and ecology of TB. The second challenge is to explore a suitable geospatial approach for detecting the TB cases according to local situation. Producing a precise disease model is important to study the mechanisms by which diseases spread, to predict the future course of an outbreak and to evaluate strategies to control an epidemic (Daley and Gani, 2005).

2.3 Challenge 1: Dynamic Transmission, Pattern Distribution, and Risk Factors of Tuberculosis

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 interaction process, related on clear changes or finding causality underlying the type and speed of the observed evolution. These conceptual contexts are important for simplifying or abstracting the complexities of structure system and phenomena. The concept can provide a simplified map, statistical information, explanatory model based on known fundamental mechanism, and towards the application of a multidisciplinary model (Lena Sanders, 2007; Garner and Hamilton, 2011).

For example, studying the transmission dynamics of TB is important to explore the factors underlying recent trends in incidence, particularly socio-economic status in Hong Kong (Wu, et al., 2010) and USA (Guzzett, et al., 2011). Guzzett, et al., (2011) assessed the impact of individually targeted control strategies, such as contact network investigation of index cases and treatment of latent TB infection (LTBI). The method used are the SEIR (Susceptibility, Exposure, Infection, Removal) compartmental structure and socio-demographic individual based model (IBM) with a realistic, time-evolving structure of preferential contacts in a population in households, schools and work-places transmission. The results show a prediction of TB transmission rates (first infection in household contacts and the rate of secondary cases in household and workplace contacts) and the simulation of alternative scenarios, particularly for TB eradication targets. Since in the context of transmission and pathogenesis of TB is across disciplines from a smaller field (molecularly) to a bigger field (epidemiology to geography or ecology) as shown in Figure 2.2, this study focuses on geospatial or population-level modelling by asssuming that infection of TB in areal or population level can control individual level of the cases.

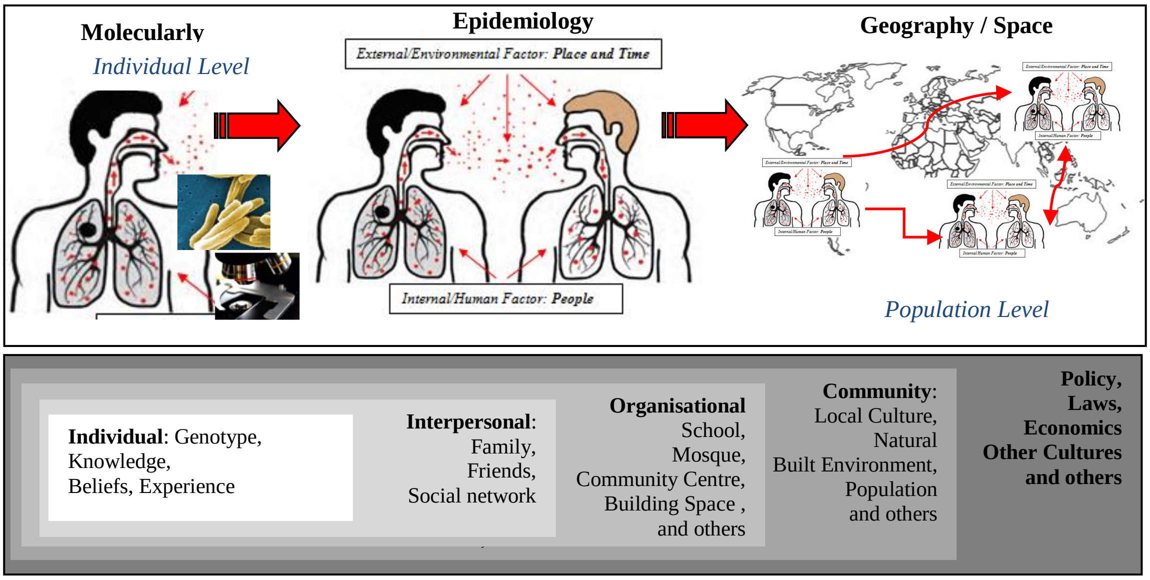

Figure 2.2: TB Transmission is Across Discipline from Biology to Geography (Idea generated from CDC, United States, 2011) and Affected Human from Individual Level to Population and Policy Level

Tuberculosis is caused by bacteria of *Mycobacterium* tuberculosis, *M*TB that most often affect the lungs, spreading from person to person through the air (WHO, 2015). This molecular aetiology is essential element in understanding the general facts about *M*TB (molecularity), causative agent of TB transmission (aetiology) and outbreak on human, animal and plant in an ecosystem (ecology). The transmission mechanism leads to determine influential risk factors, ecological or environmental characteristics, human activities and distribution patterns of TB incidences. The challeging part is that there are no a specific standard on risk factor guideline that can to be followed by all countries in the world to quantify the overall risk infection. Thus determining influential risk factors according to local environemnt is being crucial and challenging in particular area. The next sub section will explore on risk factors influencing the local TB cases in a general perspective.

2.3.1 Theory on TB Molecularity, Epidemiology and Geography

Molecule is a group of atoms bonded together, representing the smallest fundamental unit of a compound that can take part in a chemical reaction (Oxford Dictionary, 2015). In tuberculosis, it covers the major part of the disease particularly clinical and epidemiological features of the *Mycobacterium tuberculosis* (*MTB*) strains causing disease. The importance of the field of molecular epidemiology in public health generally aims to investigate whether naturally occurring strains differ in epidemiology. A main application of molecular techniques in the study of TB epidemiology is to study of *MTB* transmission dynamics from detection of laboratory stray in types to determination of geographic spread of strains in population, lastly to investigate the evaluation of *MTB* (Burgos and Pym, 2002). The functions of transmission dynamics includes:

i. Identification and confirmation of unsuspected/suspected outbreaks or transmission,
ii. Tracing of transmission chain. (Contact investigation)
iii. Evaluation of transmission in specific populations
iv. Identification of transmission in a given setting
v. Identification of risk factors and groups at risk of *M*TB infections

These classical molecular method has advantages to detect the *M*TB in individual level, but the use of genotyping method would increase the cost and labour required for analysis and complicate analysis and interpretation (Mathema et al., 2006). Since molecular epidemiology has related with population-based (transmitted within populations) and for advanced dimension, the method need to be combined with other techniques (Burgos and Pym, 2002). Mathema et al. (2006) stated there is a great need for the development of more methodological approaches (such as novel study design and mathematical models) to analyse molecular epidemiologic data that account for confounding, effect modification, and statistical power.

For example, capabilities between molecular techniques and geographical techniques can enhance to analyse the TB transmission as suggested by Narayanan, (2004) by identifying different geographical areas and populations in a high burden of disease and enhancing targeted screening efforts for interruption of disease transmission and ultimately incidence reduction. It could be said that providing a spatial molecularly understanding of TB transmission dynamics within a community is important to stimulate the implementation of control measures on a global scale.

2.3.2 Geographical Pattern of TB Distribution

Distribution of a disease pattern is a key function of health or medical geography and spatial epidemiology that link to time, place, and personal or population characteristics. This technical function is required in epidemiology for determining relationship or spatial variation analysis in disease risk to the social environment and the health of people in places, demographic, environmental, behavioural, socioeconomic, genetic, and infectious risk factors (Ostfeld et al., 2005). The functions include disease mapping, geographic correlation studies, identifying risk factor, transmission pattern, disease clusters, clustering and distribution of the disease.

Nowadays, TB is still a global concern. It can occur every part of the world. TB disease can also attack any people or regions in the world both in developed or developing countries (Figure 2.3). Mathema et al., (2006) estimated that one-third of the global community is infected with M. tuberculosis. In 1993, the World Health Organization (WHO) declared TB to be a global public health emergency, at a time when an estimated 7–8 million cases and 1.3–1.6 million deaths occurred each year. In 2000, an estimated 8 to 9 million incident cases and approximately 3 million deaths due to TB occurred worldwide. In 2010, WHO stated that there were approximately 8.5–9.2 million cases and 1.2–1.5 million deaths from TB worlwide.

TB is the second leading cause of death from an infectious disease (after HIV), and current trends suggest that TB will still be among the 10 leading causes of global disease burden in the year 2020.

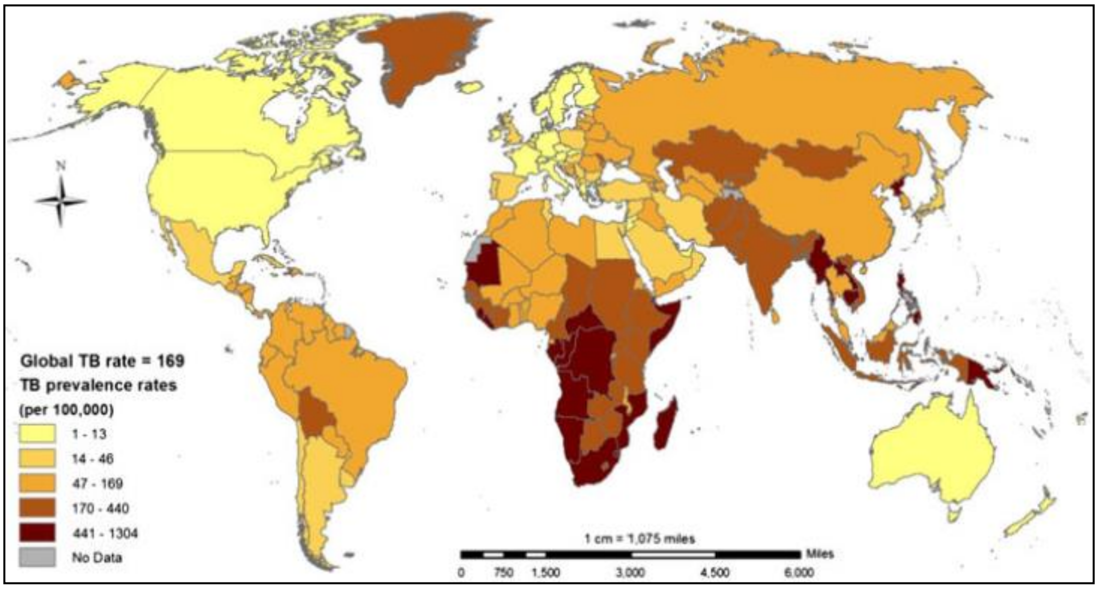

Figure 2.3: Spatial Distribution of TB Prevalence Rates, 2012 (WHO, 2013).

The global distribution of TB cases in 2000 is skewed heavily toward low-income and emerging economies. WHO (2014) reported that the majority of the TB cases come from African and Asian regions especially in South Africa and India. In 2013 (Figure 2.4), about 80% of reported TB cases occurred in 22 countries, where the largest number of new TB cases occurred in the South-East Asia and Western Pacific Regions, accounting for 56% of new cases globally (WHO, 2015). The highest prevalence of TB cases is concentrated in Asia such as China, India, Bangladesh, Indonesia, and Pakistan.

In Africa, or more specifically in sub-Saharan Africa, has the highest incidence rate of TB, with approximately 83 and 290 per 100,000, respectively. TB