The University Of Sheffield.

Healthcare seeking behaviour as a link between tuberculosis and socioeconomic factors

Chu-Chang Ku

A thesis submitted to the University of Sheffield in partial fulfilment of the requirements for the degree of Doctor of Philosophy

Health Economics and Decision Science, School of Health and Related Research

October 2019

Acknowledgements

There are many people who have made this thesis possible. kám-siā

Firstly, I must thank my supervisors, Dr Peter J Dodd and Prof Simon Dixon, whose dedication to excellence and clarity of vision motivate me a lot. Pete: I am especially grateful for your patience that you always give me chances to clarify my ideas with my calamitous expression. Simon: I appreciate your generosity supporting me through some challenges along with the PhD study.

The research in Chapter 6 was done in collaboration with Dr Hsien-Ho Lin and Dr Chien-Chou Chen. I also thank Dr Sung-Ching Pan, Jenny Wu, and Dr Wei-Cheng Lo for supporting my data management, disease definition, and the understanding of the clinical practice. Also, thanks go to the members from the Centers for Disease Control and Prevention, Taiwan for motivating the interventions in Chapter 9.

I am also grateful to my mà-chih, Tung for struggling with the world together, Dong for hating the world together, and Wei for loving the world together. You are a great power sustaining my sanity during these years. With a special thank goes to friends from Xindian Taekwondo club for sharing your life from the other side of the Earth. Thanks go to ScHARR PGR, especially my office friends for enriching our office with humorous and miracle air, slowing speaking rate for me, and company at studying.

Very special gratitude goes out to my significant other, Han Fu. We went through the challenging periods of UK life together. You are my best friend, partner, and supporter (and rival?).

I am grateful to my sibling and mother, Sin-jû and Siok-bín, who have provided me with emotional support in my life. I also need to thank my father for encouraging me to do what I enjoy. A special thank goes to my granny for teaching me how to cook Taiwanese cuisine before I came to the UK and Jī-koo-pô for encouraging me at the beginning of my PhD. I am also grateful to my other family members and friends who have supported me along the way.

Abstract

Socioeconomic barriers to tuberculosis care-seeking and costs due to care-seeking lead to unfavourable treatment, epidemiological and economic outcomes. Especially in the post-2015 era, socioeconomic interventions for tuberculosis control are receiving increasing attention. In Taiwan, the National Health Insurance programme minimises out-of-pocket expenses for patients, but important delays to tuberculosis treatment still exist.

Based on the population and tuberculosis epidemiology in Taiwan, I develop an analysis for profiling the efficacy of tuberculosis care provision and patients' care-seeking pathways. The results highlight that the interrupted tuberculosis evaluation processes and low diagnostic capacity in small local hospitals stands as key causes of extended delays to treatment, unfavourable outcomes, and costs. I analyse socioeconomic status (SES) of employment, vulnerability, and residential contexts, to identify risk factors for different aspects of care-seeking.

To link the care-seeking pathways to the nationwide tuberculosis epidemiology, I develop a data-driven hybrid simulation model. The model integrates the advantages of agent-based approaches in representing detail, and equation-based approaches in simplicity and low computational cost. This approach makes feasible Monte-Carlo experiments for robust inferences without over-simplifying the care-seeking details of interest. By comparing the hybrid model simulations with a corresponding equation-based comparator, I confirm its validity.

I considered interventions to improve universal health coverage by decentralising tuberculosis diagnostic capacity. I modelled specific interventions increasing the coverage of tuberculosis diagnostic capacity using various SES-targeted scale-up strategies. These show potential benefits in terms of reducing dropouts and reducing the tuberculosis burden, without significant increases in the inequality of care-seeking costs.

I suggest considering additional SES variables such as education, health illiteracy, and social segregation to find other care-seeking barriers. Further investigations of SES-related interventions against tuberculosis, including formal impact and health economic evaluation, should be pursued in collaboration with policymakers able to advise on feasibility and patients able to advise on acceptability.

Declaration

The work presented in this thesis was carried out between September 2016 and August 2019 in the Health Economics and Decision Science, School of Health and Related Research at the University of Sheffield under the supervision of Dr Peter J Dodd and Prof Simon Dixon.

I, the author, confirm that the Thesis is my own work. I am aware of the University's Guidance on the Use of Unfair Means (www.sheffield.ac.uk/ssid/unfair-means). This work has not been previously been presented for an award at this, or any other, university.

The research in Chapter 5 is published in Ku and Dodd [1].

This thesis was typeset using LaTeX 2ε with the template by Andrew Hills [2].

[1] Ku CC, Dodd PJ. Forecasting the impact of population ageing on tuberculosis incidence. PLoS One [Internet]. 2019; Available from:
https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0222937

[2] https://ahills60.github.io/Thesis-Template/

Contents

List of Figures

List of Tables

Acronyms

SES Socioeconomic status

Health

LTBI latent TB infection

LTFU loss to follow-up

MTB *Mycobacterium tuberculosis*

TB tuberculosis

Statistics / Statistical models / Analyses

AIC Akaike information criterion

ARIMA autoregressive integrated moving average

ARMA autoregressive moving average model

BIC Bayesian information criterion

CI confidential interval (for frequentist estimators)

CrI credible interval (for Bayesian estimators)

DIC Deviance information criterion

IPPA Individual Patient Pathway Analysis

KL-divergance Kullback–Leibler divergence

LCM Lee-Carter model

MAP Maximum a posteriori estimation

MLE Maximum likelihood estimation

PI prediction interval (for forecasts)

PPA Patient Pathway Analysis

SVAR structural vector autoregression model

Simulation models

ABM agent-based model

DES discrete event simulation

EBM equation-based model

Model builds in this thesis

CSM Agent-Based model of care-seeking

HM Hybrid model of TB dynamics
MFM Mean-Field model of TB dynamics
StSp State-Space model of care-seeking

Institutions / Programmes / Databases

ACF active case-finding
BFP Bolsa Familia Programme
CTCH close to community/home health provision
NHI National Health Insurance
NHIRD National Health Insurance Research Database
NTP national tuberculosis control programme
SDGs United Nations Sustainable Development Goals
TCDC Centers for Disease Control and Prevention, Taiwan
UHC universal healthcare coverage

Chapter 1

Introduction

Tuberculosis (TB) is the top infectious killer globally, with an estimated 10 million new cases and 1.3 million deaths among the HIV negative in 2018 [1]. An estimated 40% of global incident TB cases were not notified to the authorities, implying that many of them may not have been diagnosed or treated. Comparing estimates of incidence and prevalence allows a rough estimation of the mean duration of active disease; using the data presented in the 2018 WHO TB report, this suggests a mean global duration of over one year. Taken together, the low case detection ratio and long duration of active disease highlight the importance of patient care seeking and access to care in improving TB control worldwide.

In 2015, it was highlighted by the global call for ending poverty in the United Nations Sustainable Development Goals [2, 3] and the End TB Strategy of the WHO [4] that poverty and its consequences need to be addressed when developing TB control policies. The End TB Strategy aims for three targets to be met in 2035, compared with the level of 2015: (1) a 95% reduction in TB mortality, (2) a 90% reduction in TB incidence, and (3) zero catastrophic costs of TB-affected families incurred as a result of TB. The third target directly concerns poverty due to TB, while the first two address the consequences of poverty. To respond to these targets, related evidence and their policy implications need to be systematically investigated.

1.1 Tuberculosis and poverty

TB has long been known as a disease of poverty [5]. In terms of observed trends, the rapid decline in TB epidemic in England and Wales during the first half of the 20th century was associated with social development [6]; the epidemics in Brazil, the Americas and worldwide from 1900 to 2010 suggest a negative association be-

tween TB incidence and social inequality [7]. In terms of spatial patterns, studies in Brazil [8], Ethiopia [9], and Spain [10] have shown an association between regional socioeconomic factors and the risk of TB. Therefore, poverty reduction is seen as a critical component in delivering the End TB Strategy targets of incidence and mortality reduction.

TB patients often suffer an enormous financial burden, especially in the absence of social support. Throughout an active TB episode, a financial burden is incurred by both TB-affected families and the wider population. This means TB can drive people into poverty and trap them in poverty. As TB spreads in communities, the poverty caused by TB also spreads. From a macroeconomic point of view, Grimard and Harling [11] dealt with this relation empirically and suggested that there was an annual loss of 1.4 to 2.8 billion US dollars correlated with TB worldwide. A review in low-and middle-income countries indicated that the cost of TB care (both direct and indirect) consumed 58% of individual income and 39% of household income [12].

Lönnroth et al. [13] conceptualised the causality of TB and socioeconomic status (SES) by reviewing related studies. In their model, poverty and low SES are upstream causes driving various proximal risk factors for TB [14]. These factors impact TB epidemics in several ways: overcrowding increases TB transmission rates; malnutrition weakens immunity against TB; severe indoor pollution impairs respiratory tracts and then increases the risk of TB; limited healthcare accessibility delays care-seeking; and so on. In the end, the consequences of TB result in a lower SES and form a reinforcing feedback loop.

However, awareness of the linkage between TB and SES has been rising. SES-based actions are required worldwide for TB control. These actions either specifically target TB in low SES populations or change the SES of TB patients through social support. To ensure the effectiveness of an intervention, I need to address the social determinants of TB and plan the resource allocation with ex-ante modelling. After the intervention, I need to evaluate the policy impacts with ex-post modelling and then apply the successful experiences.

In the following sections, I deconstruct the relationships between TB and SES with respect to three aspects: evidence, control strategy, and modelling. For each aspect, I further locate our discussion at three stages: prior to TB care-seeking (e.g. transmission and infection), during TB care-seeking (e.g. care-seeking and diagnosis), and after TB care-seeking (e.g. care, treatment, and death).

1.2 Evidence on linking SES and TB

1.2.1 Evidence on TB prior to care-seeking

Epidemic features of TB prior to care-seeking include transmission, infection, immunity and disease activation. An active TB case can spread the TB pathogen, *Mycobacterium tuberculosis*, in the community; people who inhale the pathogen without immunity against TB can become infected. About ten percent of the people infected develope the TB disease over their lifetime (around one half of these within 2 years). Naturally, during the pre-TB care phase, TB patients are not supervised by any health care system, so their data are not normally available prior to diagnosis. What evidence there is for this phase is therefore based on retrospective data, often as it is recalled by patients.

Incidence rate is a standard aggregated measurement capturing the overall effect of the epidemic features mentioned above. It is defined as the number of new cases per population at risk in a given time period. Since TB incidence is impossible to measure in the general population, notification rate is often used to approximate incidence rate with a correction for under-detection. Apart from the example mentioned in the previous section, Ploubidis et al. [15] analysed TB incidence in European countries and found the incidence trend was negatively associated with regional income per capita and positively associated with income inequality. Similarly, in Zimbabwe, a hospital-based study demonstrated that the notification rate of TB co-varied with the national business cycle; that is, there were more TB cases during the recession [16]. Geographical patterns have also been investigated: Maciel et al. [17] and Figueiredo Mafra Magalhaes and Medronho [18] utilised indices of the quality of urban municipality, a social deprivation index and regional variables of illiteracy, loneliness, and low income to investigate spatial inequalities in TB incidence in Brazil.

The coverage of vaccination with Bacillus Calmette–Guérin (BCG) is an index of the susceptibility or immunity to TB of children and young adults. Perez-Then et al. [19] used a school-based census of school children and found that lower vaccination coverage was associated with both malnutrition and the distance to city centres. Guthmann et al. [20] conducted a cross-sectional study on adults and identified that being taken into care during childhood and their educational level were risk factors for low vaccination coverage.

1.2.2 Evidence on TB care-seeking

TB care-seeking involves both a patient's behaviour and the efficiency of the healthcare system. During the period of care-seeking, although healthcare systems can collect data from TB patients, data availability is still limited. This is mainly because TB shares many symptoms with other respiratory diseases and is relatively uncommon among patients with similar symptoms. In practice, it is inefficient to collect and trace care-seeking pathways of all patients with respiratory symptoms to the point that each pathogen has been diagnosed. Therefore, most studies assessing care-seeking have used retrospective data.

The basic measurement for TB care-seeking is an assessment of the duration before proper treatment (i.e., delays). The delays for care-seeking include several stages of TB diagnosis. Machado et al. [21] analysed the patient delay (from disease onset to first healthcare visit) and system delay (from first healthcare visit to diagnosis) in Rio de Janeiro, Brazil and found the patient delay was associated with personal income and employment, while the system delay did not correlate with their SES of interest. Chen et al. [22] measured the system delay using longitudinal data based on the National Health Insurance system in Taiwan, and identified personal income as a risk factor.

As an assessment of healthcare systems, Patient Pathway Analysis (PPA) was used to understand access to TB diagnostic tools and treatments (see Hanson et al. [23] and the associated special issue). In their analysis framework, the primary focus is on patients that can be properly identified and treated in facilities at their initial care seeking. However, the use of PPA has ignored SES-related issues and only unlinked aggregate data (as opposed to individual-level data) has been used, thereby limiting its ability to explore patient and system delays.

1.2.3 Evidence on TB care and outcome

The epidemic features for the period after care-seeking include compliance with TB treatment, drug reactions and treatment outcomes. At this stage, longitudinal data on TB patients becomes more available, and TB mortality can be estimated through death registration. Hence, the evidence in this field is stronger.

There were systematic reviews that explored the risk factors of treatment outcomes. Di Gennaro et al. [24] reviewed the social determinants of the treatment failure of TB patients. Their summary of fifty studies emphasised that low income and low education were significant risk factors in general. Tola et al. [25] focused on the treatment outcome of loss to follow-up (LTFU), highlighting several SES factors, such as lack of a job, financial constraints, and lack of education. Munro

et al. [26] synthesised 44 qualitative studies and highlighted that household, and community support played critical roles in adherence to TB treatment.

Some observational studies looked at specific SES factors on TB patients and assessed their effect on treatment outcomes. For example, in Brazil, Ranzani et al. [27] found that being homeless was associated with low treatment success due to loss-to-follow-up. In London, Cegolon et al. [28] found that immigrant patients and patients that are the most deprived had a higher risk of treatment failure. In Georgia, a prospective cohort study observed that the poor treatment outcome was associated with low household income, unemployment, and low education [29]. Similar evidence is seen worldwide.

In addition to the studies that focussed on specific types of outcome, some studies summarised the quality of life or financial impacts for TB patients under treatment. Measurements by Aggarwal et al. [30], in north India, indicated that whether people had a higher quality of life was associated with less unfavourable consequences. Mauch et al. [31] and Foster et al. [32] assessed the financial burden of TB patients in Kenya and South Africa respectively. Both described financial burden as a barrier to completing TB treatment and a medical poverty trap.

In the international or global scope, Nagavci et al. [33] analysed the historical data of 11 European countries, and highlighted that populations with lower SES have higher TB mortalities. A study by Khazaei et al. [34] confirmed this relationship at a global scale with aggregated data. The relationship between the human development index and TB treatment success rate suggested the importance of social and environmental factors in policy making.

1.3 TB control actions based on SES

Moving from evidence to action presents challenges. With the available evidence, finding actionable risk factors to remove and then estimating their public benefit involves a complicated process.

1.3.1 Interventions on TB incidence and TB prior to care seeking

In the period before care-seeking, people can be classified in three groups: people who are not TB infected, people with latent TB infection (LTBI), i.e. those that do not have any symptoms or illness, and TB patients who have not sought care yet. Uninfected people are potential TB patients if there are infectious TB patients in their communities to infect them.

For patients with LTBI and for ill people who have not sought care, systematic screening is a potential intervention. Active case finding is the strategy to identify LTBI and TB patients at an early stage. Patients with early diagnosis and treatment usually benefit from a higher probability of treatment success, and preventive therapy for LTBI can reduce the chance of TB disease onset. A study, which investigated mobile radiographic screening of the homeless and drug users, showed a 45% prevalence decline within four years [35]. In remote areas of Ethiopia, a close to community/home health provision programme that included TB screening, demonstrated improvements in case detection and treatment outcome [36].

1.3.2 Interventions on TB care seeking

As for patients who have started care-seeking but have not been identified as TB patients or received anti-TB drugs yet, shortening the care-seeking process is one of the targets for interventions. These "missing" and "delayed" patients affect the epidemiology of TB and increase uncertainty for policy making because they are in a stage with a high probability of TB transmission. Care-seeking processes of patients with TB determines the duration of the infectious period and affects treatment outcomes after being diagnosed. Longer infectious periods lead to a higher contribution to force of infection in the communities and then generate more future TB patients. Lack of interventions on the care-seeking process is a gap to be closed.

Although some links between SES and care-seeking behaviours have been investigated (e.g. Chen et al. [22] and Machado et al. [21]), the SES-related interventions on the interaction of patients with healthcare systems remain underexplored. The critical task behind the issue is how to translate the available evidence into policy.

1.3.3 Interventions on TB care and death

According to the evidence, disadvantaged SES is an obstacle preventing patients from accessing proper care. Hence, removing this obstacle could be a potential breakthrough in TB control. The evidence on TB treatment outcomes considering SES was richer than the care-seeking prior to treatment, so the available examples of SES-based interventions are focussed on improving treatment process instead of targeting population with low SES.

Three projects have examined or are examining, on how social support can improve TB treatment, in response to the global awareness of TB and SES. In South Africa, a cluster randomised controlled trial attempted to increase treatment ad-

herence by offering food vouchers [37]. This exemplifies how to increase treatment success with social support. It suggests that only two programmes have been tried in Latin America. The Bolsa Familia Programme (BFP) in Brazil conducted a country-level cash transfer to TB patients to improve treatment outcomes [38]. The programme resulted in a declining TB incidence rate and an increased cure rate. In Peru, Wingfield et al. [39] conducted a household-randomised controlled trial. This trial is examining the cash transfer for both TB treatment in active TB patients and preventive treatments for latent TB infection. The study is attempting to evaluate the exact effect of cash transfers by controlling the external impact of exogenous variables.

Moreover, many studies have reported that the use of incentives with social protection for TB control work as feasible interventions. In an ecological study, Reeves et al. [40] showed general social protection correlated inversely with the burden of TB epidemics by analysing the data of 21 European countries. de Pee et al. [41] reviewed the effect of providing food support to encourage adherence of TB patients, with eight studies in different settings consistently favouring the intervention.

1.4 Modelling of SES and TB

Models have been used to glue data, knowledge, and theory together to provide frameworks for inference and projections of interest. Statistical models connect data and probability theory to explore empirical findings; mathematical models derive new theories by integrating knowledge; simulation models apply theories, absorb new evidence, and make projections. Over the past few decades, epidemiologists have been using mathematical and simulation models to overcome the difficulties such as time cost, ethical issues and limited budget seen in observational studies [42]. Functionally, these models provide a virtual environment to conduct policy experiments that are not feasible in reality.

Regarding TB and SES, Pedrazzoli et al. [43] reviewed mathematical models of TB with social determinants and identified eight studies in total. Four of them modelled low SES as a risk factor of TB transmission [44–47]. Three studies used SES as a risk factor for disease progression [48–50] and one study allowed SES to influence case detection [51]. Apart from these models, Nili et al. [52] modelled the interaction of TB epidemics and economic growth in the sense of macroeconomics. The model demonstrated a negative feedback loop trapping the poor with TB.

Most of the simulation modelling studies on SES covered the pre-care stages. In modelling the epidemic features of pre-TB care, incidence rate is usually affected

jointly by disease transmission and progression. The main focus of these studies was to explore the effect of wealth and urbanisation on TB. For example, Andrews et al. [44] and Rehkopf et al. [53] highlighted the importance of social segregation while implementing interventions. These models used surrogate variables (i.e. using low BMI and undernourishment as proxies for low SES) to parametrise SES in disease progression. However, none of those modelling studies discussed the roles of SES on the care-seeking processes and the delays to TB treatments.

In addition, to model TB care seeking and treatment outcomes, Reeves et al. [51] linked case detection and treatment success with macroeconomic fluctuations. In the other models, all events associated with healthcare such as screening, diagnosis, and treatment outcomes, are summarised as constant parameters without the SES strata. Without parameterising SES in healthcare seeking directly, Andrews et al. [44] composed a diagnosis improvement intervention favouring the poor and showed the importance of interventions that consider the aspect of SES. In addition, van den Boogarrd et al. [54] modelled the socioeconomic barriers that keep patients from good adherence. The study results recommended an increase in social support for TB patients.

1.5 Study scope

Following the announcement of the United Nation Sustainable Development Goals (SDGs) in 2015 TB control programmes have sought innovative SES-related TB interventions and their empirical evaluations [55]. Care-seeking behaviours and TB services provision are keys to connect SES with TB. There are a large number of studies related to the adherence to TB treatments. The evidence linking socioeconomic factors and treatment outcome is well developed, and a lot of SES-based interventions have been shown to improve treatment outcomes. However, less attention has been paid to other types of care-seeking.

The limited availability of data related to care-seeking that precede TB confirmation is a challenge. Before TB confirmation, patients with TB are not technically "TB patients" by definition. Therefore, most of the related studies collected data after individuals were identified as TB patients and the events before the diagnosis were based on retrospective assessments.

Routinely collected healthcare utilisation data typically contain details of healthcare provision since the first care-seeking for TB-suggestive symptoms. The currently available measurements for patient pathways are restricted in delays (e.g. Sreeramareddy et al. [56] and Chen et al. [22]), dropout cascades (e.g. Subbaraman et al. [57]), and the patient pathway analysis (PPA [23]). The delays indicate

the time spent care-seeking and the dropout shows the adherence of patients in different stages of care-seeking. Both these measurements are good at reflecting aspects of the care-seeking process. However, they do not consider the capability of a health system of providing accurate and efficient TB servicesd, which usually connects to interventions. On the other hand, the PPA extracts the information for healthcare provision and patient ends but it does not consider temporal information. Without temporal information, further intervention evaluation will be limited.

Assuming the details of the care-seeking process including TB-services provision and needs are available, how to use them to project the population-level epidemiology of TB is another challenge. As the review of Pedrazzoli et al. [43] summarised that mathematical modelling by equation-based models (EBMs) can incorporate socioeconomic determinants into the simulation of TB dynamics while EBMs aggregate individuals to population-level. Using agent-based models (ABMs), which can replicate empirical results in the individual level and include transmission process. However, the computational burden of ABMs in large scale population is a major drawback. Providing an efficient simulation model to compactly represent the system while supporting interventions on the individual care-seeking process are needed.

This thesis is focusing on TB and SES through the care-seeking process. There are twofold purposes. First, jointly considering the perspectives from both healthcare providers and demands, I aim to identify the obstacles delaying TB diagnostics and inducing unfavourable treatment outcomes using an individual, longitudinal data. Second, given the empirical results in individual-level, I want to build a simulation model, providing a platform for exploring SES-based interventions.

1.6 Overview

Chapter 2 reviews the evidence of SES and TB with the connections of care-seeking and empirical impacts of SES-based interventions on TB-related outcomes. The chapter includes three scoping reviews, aiming to summarise the existing literature in the field. The reviews identify gaps in evidence and use this to highlight potential SES-based interventions that are explored later in my thesis.

Chapter 3 reviews the hybrid modelling approach in the context of health. The review focuses on the transitions between agents and values in equations given different levels of needed details. Finally, this chapter discusses the suitability of this approach for my research questions with examples of hybrid modelling and technical details in the applications.

Chapter 4 reviews the applications of health behaviour simulation in agent-based modelling. The chapter investigates how previous models simulate actions with the influence of personal tracts, social networks, and overall contexture. The review results provide information on the parameterisation of ABMs from reality and a guideline for simulating health behaviour.

Chapter 5 introduces the demography and epidemiology of TB in Taiwan by presenting the statistical summary and modelling. This chapter analyses and forecasted the age-specific TB incidence rates using Lee-Carter models and time-series analysis. The chapter emphasises the impact of population ageing on TB burden in the future.

Chapter 6 develops a method, the individual patient pathway analysis (IPPA), for extracting and analysing care-seeking pathways of TB patients and SES. This analysis uses a health utilisation dataset based on the National Health Insurance programme in Taiwan.

Chapter 7 uses the results of Chapter 6 to examine the socioeconomic determinants associated with extended care-seeking processes and economic outcomes caused by TB care-seeking.

Chapter 8 develops a TB hybrid model which can simulate nationwide TB dynamics of Taiwan. The model has an agent-based submodel and an equation-based submodel. The agent-based part of this model instantises the patients with active TB who seek care and are under treating based, while the equation-based part captures uninfected population and LTBI who are at risk of TB. The methodology is supported by results of Chapter 4 and Chapter 3. The model is driven by empirical results of Chapter 6 and Chapter 7.

Chapter 9 uses the model developed in Chapter 8 to consider the SES-based intervention. The simulation focuses on increasing the TB diagnostics provision. This chapter evaluates the intervention responses on the epidemiological and economic aspects.

Finally, the last chapter summarises this thesis and discuss the limitations and future perspectives.

Bibliography

[1] Organization WH, Others. Global tuberculosis report 2018. World Health Organization; 2018.

[2] Assembly G. Sustainable development goals. SDGs), Transforming our world: the. 2015;2030.

[3] Griggs D, Stafford-Smith M, Gaffney O, Rockström J, Ohman MC, Shyamsundar P, et al. Policy: Sustainable development goals for people and planet. Nature. 2013 Mar;495(7441):305–307.

[4] Organization WH, Others. Documentation for World Health Assembly 67. Geneva: WHO. 2014;.

[5] Stevens P. Diseases of poverty and the 10/90 gap. International Policy Network, http://who int/intellectualproperty/submissions/InternationalPolicyNetwork pdf. 2004;.

[6] Bhatti N, Law MR, Morris JK, Halliday R, Moore-Gillon J. Increasing incidence of tuberculosis in England and Wales: a study of the likely causes. BMJ. 1995 Apr;310(6985):967–969.

[7] Guimarães RM, Lobo AdP, Siqueira EA, Borges TFF, Melo SCC. Tuberculosis, HIV, and poverty: temporal trends in Brazil, the Americas, and worldwide. J Bras Pneumol. 2012 Jul;38(4):511–517.

[8] Harling G, Lima Neto AS, Sousa GS, Machado MMT, Castro MC. Determinants of tuberculosis transmission and treatment abandonment in Fortaleza, Brazil. BMC Public Health. 2017 May;17(1):508.

[9] Alene KA, Viney K, McBryde ES, Clements ACA. Spatial patterns of multidrug resistant tuberculosis and relationships to socio-economic, demographic and household factors in northwest Ethiopia. PLoS One. 2017 Feb;12(2):e0171800.

[10] C\'ordoba-Doña JA, Novalbos-Ruiz JP, Su\'arez-Farfante J, And\'erica-Fr\'ıas G, Escolar-Pujolar A. Social inequalities in HIV-TB and non-HIV-TB patients in two urban areas in southern Spain: multilevel analysis. Int J Tuberc Lung Dis. 2012;16(3):342–347.

[11] Grimard F, Harling G. The impact of tuberculosis on economic growth. McGill University, Montréal. 2004;.

[12] Tanimura T, Jaramillo E, Weil D, Raviglione M, Lönnroth K. Financial burden for tuberculosis patients in low- and middle-income countries: a systematic review. Eur Respir J. 2014 Jun;43(6):1763–1775.

[13] Lönnroth K, Jaramillo E, Williams BG, Dye C, Raviglione M. Drivers of tuberculosis epidemics: The role of risk factors and social determinants; 2009.

[14] Hargreaves JR, Boccia D, Evans CA, Adato M, Petticrew M, Porter JDH. The social determinants of tuberculosis: from evidence to action. Am J Public Health. 2011 Apr;101(4):654–662.

[15] Ploubidis GB, Palmer MJ, Blackmore C, Lim TA, Manissero D, Sandgren A, et al. Social determinants of tuberculosis in Europe: a prospective ecological study. Eur Respir J. 2012 Oct;40(4):925–930.

[16] Burke SJ, Lass E, Thistle P, Katumbe L, Jetha A, Schwarz D, et al. Increased incidence of tuberculosis in zimbabwe, in association with food insecurity, and economic collapse: an ecological analysis. PLoS One. 2014 Feb;9(2):e83387.

[17] Maciel ELN, Pan W, Dietze R, Peres RL, Vinhas SA, Ribeiro FK, et al. Spatial patterns of pulmonary tuberculosis incidence and their relationship to socio-economic status in Vitoria, Brazil. Int J Tuberc Lung Dis. 2010 Nov;14(11):1395–1402.

[18] Figueiredo Mafra Magalhaes MdA, Medronho RdA. Spatial analysis of Tuberculosis in Rio de Janeiro in the period from 2005 to 2008 and associated socioeconomic factors using micro data and global spatial regression models. Cien Saude Colet. 2017;22(3):831–839.

[19] Perez-Then E, Shor-Posner G, Crandall L, Wilkinso J. The relationship between nutritional and sociodemographic factors and the likelihood of children in the Dominican Republic having a BCG scar. Am J Public Health. 2007;21(6):365–372.

[20] Guthmann JP, Chauvin P, Le Strat Y, Soler M, Fonteneau L, Lévy-Bruhl D. Family history of immigration from a tuberculosis endemic country and low family income are associated with a higher BCG vaccination coverage in Ile-de-France region, France. Vaccine. 2013 Nov;31(48):5666–5671.

[21] Machado ACdFT, Steffen RE, Oxlade O, Menzies D, Kritski A, Trajman A. Factors associated with delayed diagnosis of pulmonary tuberculosis in the state of Rio de Janeiro, Brazil. J Bras Pneumol. 2011 Jul;37(4):512–520.

[22] Chen CC, Chiang CY, Pan SC, Wang JY, Lin HH. Health system delay among patients with tuberculosis in Taiwan: 2003-2010. BMC Infect Dis. 2015 Nov;15(1):491.

[23] Hanson C, Chin D, Osberg M, Brown J, Durham G. Conducting patient pathway analysis to inform tuberculosis programming: methodology. Journal of Infectious Disease. 2017;216(Suppl 7):S679–85.

[24] Di Gennaro F, Pizzol D, Cebola B, Stubbs B, Monno L, Saracino A, et al. Social determinants of therapy failure and multi drug resistance among people with tuberculosis: A review. Tuberculosis. 2017 Mar;103:44–51.

[25] Tola HH, Tol A, Shojaeizadeh D, Garmaroudi G. Tuberculosis Treatment Non-Adherence and Lost to Follow Up among TB Patients with or without HIV in Developing Countries: A Systematic Review. Iran J Public Health. 2015 Jan;44(1):1–11.

[26] Munro SA, Lewin SA, Smith HJ, Engel ME, Fretheim A, Volmink J. Patient adherence to tuberculosis treatment: a systematic review of qualitative research. PLoS Med. 2007 Jul;4(7):e238.

[27] Ranzani OT, Carvalho CRR, Waldman EA, Rodrigues LC. The impact of being homeless on the unsuccessful outcome of treatment of pulmonary TB in São Paulo State, Brazil. BMC Med. 2016 Mar;14:41.

[28] Cegolon L, Maguire H, Mastrangelo G, Carless J, Kruijshaar ME, Verlander NQ. Predictors of failure to complete tuberculosis treatment in London, 2003-2006. Int J Tuberc Lung Dis. 2010 Nov;14(11):1411–1417.

[29] Djibuti M, Mirvelashvili E, Makharashvili N, Magee MJ. Household income and poor treatment outcome among patients with tuberculosis in Georgia: a cohort study. BMC Public Health. 2014 Jan;14:88.

[30] Aggarwal AN, Gupta D, Janmeja AK, Jindal SK. Assessment of health-related quality of life in patients with pulmonary tuberculosis under programme conditions. Int J Tuberc Lung Dis. 2013 Jul;17(7):947–953.

[31] Mauch V, Woods N, Kirubi B, Kipruto H, Sitienei J, Klinkenberg E. Assessing access barriers to tuberculosis care with the tool to Estimate Patients' Costs: pilot results from two districts in Kenya. BMC Public Health. 2011 Jan;11:43.

[32] Foster N, Vassall A, Cleary S, Cunnama L, Churchyard G, Sinanovic E. The economic burden of TB diagnosis and treatment in South Africa. Soc Sci Med. 2015 Apr;130:42–50.

[33] Nagavci BL, de Gelder R, Martikainen P, Deboosere P, Bopp M, Rychtaříková J, et al. Inequalities in tuberculosis mortality: long-term trends in 11 European countries. Int J Tuberc Lung Dis. 2016 May;20(5):574–581.

[34] Khazaei S, Rezaeian S, Baigi V, Saatchi M, Molaeipoor L, Khazaei Z, et al. Incidence and pattern of tuberculosis treatment success rates in different levels of the human development index: a global perspective. South Afr J Epidemiol Infect. 2017 Aug;32(3):100–104.

[35] de Vries G, van Hest RAH, Richardus JH. Impact of mobile radiographic screening on tuberculosis among drug users and homeless persons. Am J Respir Crit Care Med. 2007 Jul;176(2):201–207.

[36] Datiko DG, Lindtjørn B. Health extension workers improve tuberculosis case detection and treatment success in southern Ethiopia: a community randomized trial. PLoS One. 2009 May;4(5):e5443.

[37] Lutge E, Lewin S, Volmink J, Friedman I, Lombard C. Economic support to improve tuberculosis treatment outcomes in South Africa: a pragmatic cluster-randomized controlled trial. Trials. 2013 May;14:154.

[38] Nery JS, Rodrigues LC, Rasella D, Aquino R, Barreira D, Torrens AW, et al. Effect of Brazil's conditional cash transfer programme on tuberculosis incidence. Int J Tuberc Lung Dis. 2017 Jul;21(7):790–796.

[39] Wingfield T, Tovar MA, Huff D, Boccia D, Montoya R, Ramos E, et al. A randomized controlled study of socioeconomic support to enhance tuberculosis prevention and treatment, Peru. Bull World Health Organ. 2017 Apr;95(4):270–280.

[40] Reeves A, Basu S, McKee M, Stuckler D, Sandgren A, Semenza J. Social protection and tuberculosis control in 21 European countries, 1995-2012: a cross-national statistical modelling analysis. Lancet Infect Dis. 2014 Nov;14(11):1105–1112.

[41] de Pee S, Grede N, Mehra D, Bloem MW. The enabling effect of food assistance in improving adherence and/or treatment completion for antiretroviral therapy and tuberculosis treatment: a literature review. AIDS Behav. 2014 Oct;18 Suppl 5(5):S531–41.

[42] Grassly NC, Fraser C. Mathematical models of infectious disease transmission. Nat Rev Microbiol. 2008 Jun;6(6):477–487.

[43] Pedrazzoli D, Boccia D, Dodd PJ, Lönnroth K, Dowdy DW, Siroka A, et al. Modelling the social and structural determinants of tuberculosis: opportunities and challenges. Int J Tuberc Lung Dis. 2017 Sep;21(9):957–964.

[44] Andrews JR, Basu S, Dowdy DW, Murray MB. The epidemiological advantage of preferential targeting of tuberculosis control at the poor. Int J Tuberc Lung Dis. 2015 Apr;19(4):375–380.

[45] Bhunu CP, Mushayabasa S, Smith? RJ. Assessing the effects of poverty in tuberculosis transmission dynamics. Appl Math Model. 2012 Sep;36(9):4173–4185.

[46] Dye C, Bourdin Trunz B, Lönnroth K, Roglic G, Williams BG. Nutrition, diabetes and tuberculosis in the epidemiological transition. PLoS One. 2011 Jun;6(6):e21161.

[47] Odone A, Houben RMGJ, White RG, Lönnroth K. The effect of diabetes and undernutrition trends on reaching 2035 global tuberculosis targets. Lancet Diabetes Endocrinol. 2014 Sep;2(9):754–764.

[48] Oxlade O, Huang CC, Murray M. Estimating the Impact of Reducing Under-Nutrition on the Tuberculosis Epidemic in the Central Eastern States of India: A Dynamic Modeling Study. PLoS One. 2015 Jun;10(6):e0128187.

[49] Ackley SF, Liu F, Porco TC, Pepperell CS. Modeling historical tuberculosis epidemics among Canadian First Nations: effects of malnutrition and genetic variation. PeerJ. 2015 Sep;3:e1237.

[50] Lin HH, Murray M, Cohen T, Colijn C, Ezzati M. Effects of smoking and solid-fuel use on COPD, lung cancer, and tuberculosis in China: a time-based, multiple risk factor, modelling study. Lancet. 2008 Oct;372(9648):1473–1483.

[51] Reeves A, Basu S, McKee M, Sandgren A, Stuckler D, Semenza JC. Tuberculosis control and ecnomic recession: longitudinal study of data from 21 European countries, 1991-2012. Bull World Health Organ. 2015;93(6):369–379.

[52] Nili F, Delfino D, Simmons PJ. Infectious Disease and Economic Growth: the Case of Tuberculosis. University of York; 1999.

[53] Rehkopf D, Furumoto-Dawson A, Kiszewski A, Awerbuch-Friedlander T. Spatial Spread of Tuberculosis through Neighborhoods Segregated by Socioeconomic Position: A Stochastic Automata Model. Discrete Dyn Nat Soc. 2015 Oct;2015.

[54] van den Boogaard J, Msoka E, Homfray M, Kibiki GS, Heldens JJHM, Felling AJA, et al. An exploration of patient perceptions of adherence to tuberculosis treatment in Tanzania. Qual Health Res. 2012 Jun;22(6):835–845.

[55] Reid MJA, Arinaminpathy N, Bloom A, Bloom BR, Boehme C, Chaisson R, et al. Building a tuberculosis-free world: The Lancet Commission on tuberculosis. Lancet. 2019 Mar;393(10178):1331–1384.

[56] Sreeramareddy CT, Panduru KV, Menten J, Van den Ende J. Time delays in diagnosis of pulmonary tuberculosis: a systematic review of literature. BMC Infect Dis. 2009 Jun;9(1):91.

[57] Subbaraman R, Nathavitharana RR, Mayer KH, Satyanarayana S, Chadha VK, Arinaminpathy N, et al. Constructing care cascades for active tuberculosis: A strategy for program monitoring and identifying gaps in quality of care. PLoS Med. 2019 Feb;16(2):e1002754.

Chapter 2

Review: socioeconomic status and TB care-seeking

2.1 Introduction

Socioeconomic status (SES) impacts the transmission, diagnostics, treatment, and outcome of TB distally. As conceptualised by Lönnroth et al. [1], the influences of SES can pass down through multiple routes within multiple aspects of TB epidemiology. As more and more scientific facts have been revealed, the interest in controlling TB through structural changes related to SES and targeting TB issues for low socioeconomic groups has increased. Following the announcement of the United Nation Sustainable Development Goals (SDGs) in 2015, the TB control programmes have called for innovative SES-related TB interventions and the attendant empirical evaluations.

In terms of the care-seeking process of TB, SES plays a fairly straightforward role. From the healthcare provision end, higher level SES communities are able to support healthcare facilities with higher quality since providing high quality and sufficient healthcare services requires the requisite human and financial resources. Meanwhile, from the patient end, individuals who have fewer concerns about seeking care have a better chance of accessing appropriate care when they are unwell. These concerns can result from income loss, travelling costs, lack of information, or stigmatisation. Intuitively, if the barriers to accessing appropriate TB services can be ameliorated or removed, and the healthcare systems can provide appropriate diagnostics and treatments, the TB-related burden could be alleviated. However, the care-seeking of TB patients is a complex interaction of decision making by patients and healthcare providers (health system, healthcare facilities, and clinicians). To ensure the causal pathway is well-connected and SES-

based interventions are effective, a systematic arrangement of the evidence related to a number of aspects is required.

This chapter summarises the available studies on the relationship between SES and TB that involves the mediator of care-seeking dynamics before it goes on to systematically discuss the state-of-the-art SES-related TB control strategies. The chapter is divided into a number of sections. Section 2.2 reviews the evidence that supports SES-based TB control strategies. Section 2.3 investigates the SES-based TB control actions based on Universal Health Coverage (UHC) and social protection (SP), which target support deprived TB patients. The last section then discusses the challenges related to SES-based TB control strategies.

However, before proceeding, I must clarify the terminology I use when referring to specific individuals. Here, **ill people** refers to the individuals who have active TB, **patients** refers to ill individuals who have initiated their care-seeking, and **TB patients** refers to the individuals who have been diagnosed with TB or who have begun their TB treatment.

2.2 The influence of SES on care-seeking and treatment adherence

The aim of this section is to bring together the evidence that links SES, to care-seeking, and treatment adherence and examine its the policy implications. To this end, I conducted a scoping review to explore the evidence that indicates the existence of these links. Here, I conceptualised the connections between the evidence and patient behaviour by using the Theory of Planned Behaviour (TPB) [2], which is a psychological and conceptual model used to investigate an action as a result of attitudes, subjective norms, and perceived barriers. In this section, the term "care-seeking" encompasses care-seeking prior to treatment, the adherence to TB treatment, and the adherence to preventive TB treatments for TB latency.

2.2.1 Searching strategy

The search was conducted using the Web of Science with keywords related to TB, SES, and care-seeking behaviours. Specifically speaking, I searched **"Tuberculosis" AND (socioeconomic OR "socio-economic*" OR "social determinant*" OR "catastrophic cost*" OR "catastrophic health cost*" OR "social develop*") AND ("care seeking" OR "delay*" OR "patient pathway*" OR "compliance" OR "adherence" OR "default")**. I included two types of records: the available review articles with synthesised evidence and research articles with new evidence.

I reviewed the two article types separately.

I curated review articles published before July 2019 and research articles published between January 2015 and July 2019 (post-2015 era). I only included studies published in English and only those that considered any SES factors associated with care-seeking behaviours. I included qualitative, cross-sectional, observational, trial, and statistical modelling studies. I excluded any studies based on SES-based interventions (which will be analysed in the next section). I also excluded studies involving participants restricted to specific SES groups without a comparator group and any that involved assessing the TB treatment failures that were not due to the care-seeking behaviour of the patients. I also excluded any studies that had previously been collected in review studies in order to prevent any potential duplication. In terms of review articles, I excluded those studies that only provided narrative summaries.

A narrative synthesis was undertaken using the TPB to group findings into three categories. According to the TPB, the intention of action can be affected by three factorial dimensions: attitude, subjective norms (SN), and perceived behavioural control (PBC). Here, the attitude toward care-seeking is related to whether individuals who are experiencing potential TB-suggestive illness think they need to seek proper healthcare. Risk perceptions related to TB and the awareness of TB are examples of the attitude in the TPB. I also linked education to the attitude dimension. Meanwhile, the SN generally refers to the influences of peers or of significant others. I mapped SN to the stigmatisation of TB and to the discrimination toward TB patients. Finally, the PBC relates to whether people believe that they can access TB-related services and believe that the services could actually solve their health problems. I referred PBC to various costs of adherence, low SES, and the capacity of the assessed hospitals.

2.2.2 Results

As Figure 2.1 shows, I retrieved 20 review articles and 105 research articles. The title and abstract screening resulted in nine review articles and 41 research articles. Following the subsequent full text review, I retained three review articles and 19 research articles in the final collection. All the reviews were related to treatment adherence. Among the 19 research articles, six investigated care-seeking prior to treatment start; two studies for latent TB preventive therapy, ten looked at treatment adherence, and one of assessed pre-hospital care-seeking and treatment nonadherence simultaneously.

Care-seeking before treatment

I identified seven studies that assessed the care-seeking behaviours before the TB treatment had started. Of these one was qualitative, one used a mixed-methods

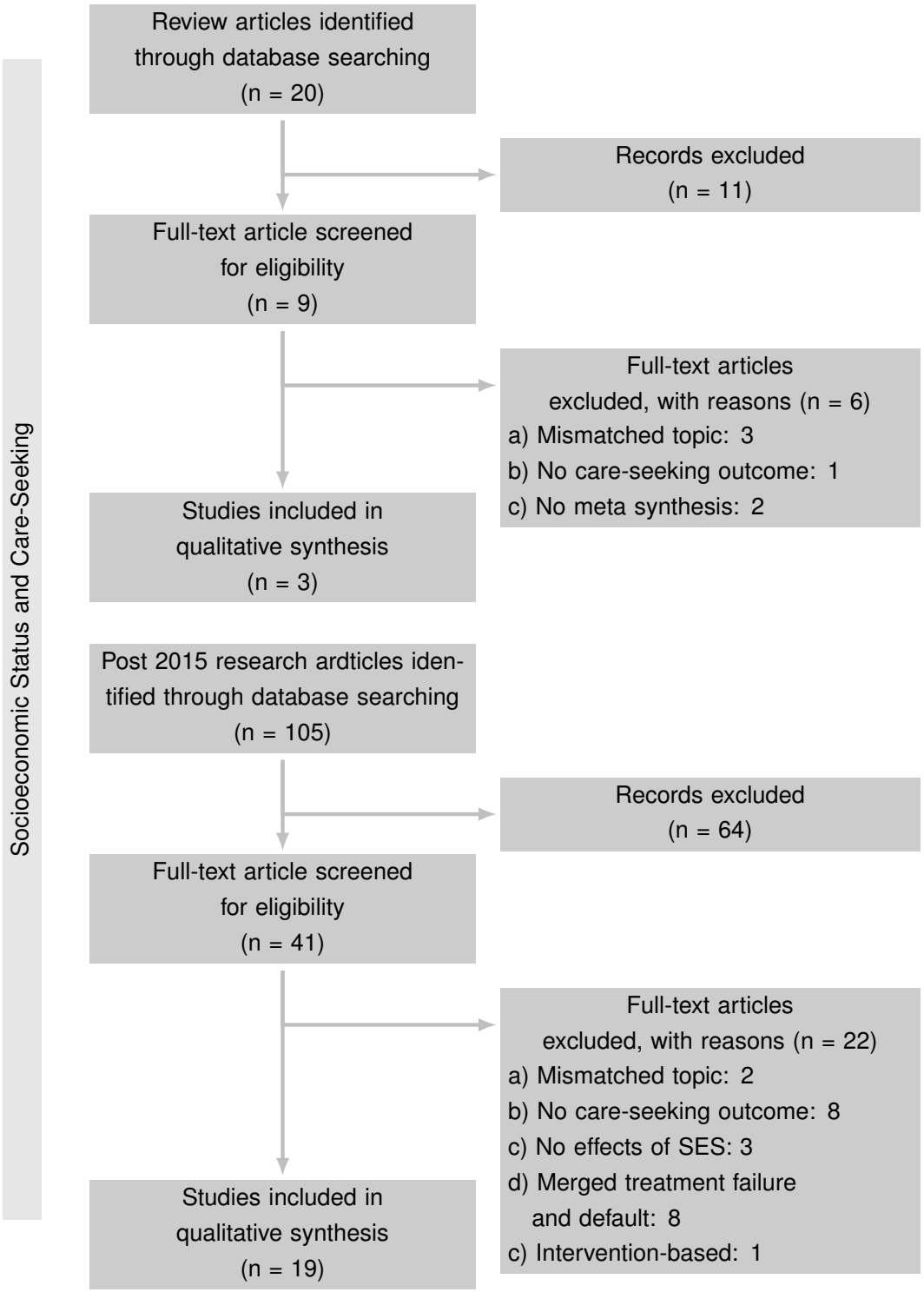


Figure 2.1: PRISMA flow diagram: Care-seeking and SES

approach, and five were quantitative studies. All of them were retrospective studies with case-control design or case-control nested in a cohort or trial study, and all surveyed the TB patients about their event history. Before the initial contact with the hospital was made, the intention of consultation for a cough that had persisted for more than two weeks in relation to the pre-hospital care-seeking was investigated [3, 4]. Here, delays between the onset of the symptoms [5] and the initial care-seeking [6–9] were found to exist.

Attitude Five out of the seven studies considered education level or risk perception [3–7]. However, the direction of the effects was diverse. Christian et al. [4] found that people with higher education levels were less likely to consult for coughs lasting for more than two weeks, while Laohasiriwong et al. [5] and Westerlund et al. [6] found negative correlations between lower education levels and longer delays, while the other two studies showed null effects [3, 7].

Subjective norms Four out of the seven studies considered stigmatisation and regional SES. While quantitatively testing the stigmatisation of TB was not mentioned in the quantitative studies, Cremers et al. [3] highlighted the concerns related to stigmatisation through interviewing TB patients in qualitative terms. Meanwhile, Saqib et al. [8] showed a regional disparity in that living in rural areas correlated to prolonging treatment delay.

Perceived behavioural control Financial and transportation barriers were addressed in all seven studies. With the except of Said et al. [7], the researchers found that the existence of these barriers correlated to longer delays in various terms.

Adherence to TB preventive treatment

One review article and two research articles investigated the dropout rate during Isoniazid preventive therapy [10–12]. A retrospective cohort study conducted in Brazil identified that a low Human Development Index [13] at a regional level was associated with a higher dropout rate [10]. Meanwhile, the other study, a prospective cohort study on HIV infected patients in Tanzania, found that the dropout rate was not associated with either an asset-based SES index or education levels [11].

Adherence to TB treatment

I found two review articles and 11 research articles that had investigated TB treatment adherence. Table 2.1 lists the extracted data of the 11 research articles. Here, I located ecological, cross-sectional, and nested case-control studies related to the field in question. Of these, three out of the 10studies were qualitative, two used a mix-methods approach and the remainers were quantitative studies. One ecological study [14] used districts as the basic units while the others were based on TB patients. Table 2.2 lists the identified factors from the 11 studies that related to the TPB framework.

Table 2.1: Studies for non-adherence to TB treatment.

Reference	Year	Setting	Subjects	Study type	LTFU measurement
[15]	2017	Lima and Callao, Peru	TB patients (18-65)	Cross-sectional, quantitative	Proportion of drugs taken
[16]	2016	South Africa	TB patients ($\geq$ 18)	Cross-sectional, mix methods	Self-reported missed visits and doses in proportion
[17]	2016	Masan and Soel, Korea	TB patients (20-65)	Nested case-control	Interrupted Treatment 60 days
[3]	2016	Urban Zambia	TB patients	Cross-sectional, mix methods	Interrupted Treatment $>$ 7 days
[18]	2017	Remote Region, Papua New Guinea,	TB patients ($\geq$ 18)	Cross-sectional, qualitative	Not applicable
[19]	2015	Buenos Aires, Argentina	TB patients ($\geq$ 18)	Cross-sectional, quantitative	Interrupted Treatment $>$ 60 days
[14]	2018	Rio De Janeiro, Brazil	Districts	Ecological	Refusal of completing treatment
[20]	2018	Taiz and Al Hudaydah, Yemen	TB patients ($\geq$ 16)	Cross-sectional, quantitative	Interrupted Treatment $>$ 60 days
[21]	2016	Cairo, Egypt	TB patients ($\geq$ 18)	Cross-sectional, qualitative	Not applicable
[22]	2015	Multiple countries in Southern Africa	TB patients ($\geq$ 18)	Cross-sectional, quantitative	Had a missed DOTS dose
[23]	2017	Brazil	TB-HIV Co-Infected patients	Nested case-control	Interrupted Treatment $>$ 30 days

LTFU: loss to follow-up

Table 2.2: SES factors of non-adherence to TB treatment.

Reference	Attitude/Information	Subjective Norms	Perceived Behavioural Control
[15]	TB knowledge, motivation	emotional support	self-efficacy, financial support
[16]			transport cost, financial cost
[17]	education		housing, occupation
[3]	education	stigma	employment, Financial constraint, travel time/distance
[18]	witchcraft belief	discrimination	income loss
[19]			subsistence capacity, housing, granted leave of absence
[14]	illiteracy, HDI (education)		asset, HDI (income), SDI
[20]	Illiteracy		travel cost, waiting time, travel time
[21]		social context, segregation, family feels	income loss
[22]	health illiteracy		household income
[23]	illiteracy		living shelter house

HDI: Human Development Index, SDI: Social Development Index

Attitude and information I identified that illiteracy, health-related illiteracy, and education levels were investigated in the selected body of work. According to the review from Tola et al. [24] the effects of education were diverse, that is, higher education does not necessarily imply better adherence to TB treatment. In my selected research articles, I found no instance where higher education was associated with lower adherence, while a number did suggest that the association was insignificant. Algeria-Flores et al. [15] used structural equation modelling to show how the education variables affect the treatment adherence only through increasing personal skills related to following the treatment, while no direct effects were found. Meanwhile, in Theron et al. [22], both education level and health-related illiteracy were investigated. Here, the treatment adherence was found to be affected by health-related illiteracy but not by education level. The qualitative study from Diefenbach-Elstob et al. [18] demonstrated an example where the patient did not accept the results of TB diagnostics nor TB treatment, because she regarded it is a sin to have TB. This example indicated that the stigma attached to TB may stop patients from adhering.

Subjective norm Stigmatisation and emotional support from the family were commonly found in the field in question. In fact, both review articles [24, 25] arouse the importance of these aspects. In the selected qualitative studies, stigmatisation was a common concern related to TB treatment, while any such association was not consistently identified in the quantitative analyses. A mixed-methods study from Cremers et al. [3] addressed how the stigma of TB was a factor in their qualitative analysis but not in their quantitative analysis. However, in their covariate analysis of suffering from the stigma and non-adherence revealed that this factor was insignificant. In addition to the stigma of TB, Lohiniva et al. [21] revealed an issue regarding social norms in terms of individuals who are incapable of working. Being regarded as an "useless" can be a risk factor for non-adherence.

Perceived behavioural control In terms of PBC, I found various measurements related to the barriers to completing treatments within the selected body of work. With regard to patient assets, living in shelters and a lack of water supplies were identified as risk factors. While many measurements based on income were commonly assessed, the attendant effects were found to be diverse. In Maciel et al. [14], higher regional income was correlated with higher adherence. However, at the individual level, the effect of personal income or employment was rarely identified. As for the treatment-related barriers, the effects of income loss, granted income for absence due to treatment compliance, financial costs, and travelling costs were consistently identified within the selected studies [3, 16, 18–21]. In terms of the incoming cost and disposable income, Herrero et al. [19] used the capacity of subsistence to measure the poverty level and found a positive correlation to treatment adherence. In addition to these measurements, Alegria-Flores et al. [15] surveyed an index of self-efficacy and concluded that the index affected TB treatment adherence.

2.3 SES-based TB interventions

The evidence connecting TB to SES through care-seeking is abundant. However, how to control TB in view of the SES factors remains debatable [26]. Universal Health Coverage (UHC) and Social Protection (SP) are the two main streams that echo the call for SES-based interventions [27]. UHC is aimed at ensuring every person with an illness can have access to appropriate healthcare services. The attendant programmes are generally aimed at a general improvement in care-seeking attempts and healthcare experiences. On the other hand, SP is literally aimed at protecting TB patients and TB-affected households from financial and health risks

when the patients are receiving care. To understand the implementation of SES-based interventions, I carried out two scoping reviews, one for UHC and one for SP. I restricted the search the Web of Science to the publications released before July 2019 that were compiled in English. Again, I included both research articles and review articles. I collected the evaluations of the interventions through trial-based, ecological, statistical modelling, and mathematical modelling studies.

2.3.1 Universal Health Care, UHC

The outcome of TB can be catastrophic not only in terms of TB fatality but also in terms of the financial burden, which will affect the survival of an individual or a household. Therefore, ending TB is firmly included in the roadmap of achieving UHC. Meanwhile, the implementation of UHC, in turn, encourages people to access the healthcare system when they have TB-suggestive illness. However, while all the measurements on catastrophic costs can inform the demand for UHC, but whether UHC can substantially empower TB-related care-seeking remains debatable.

Searching and screening

I reviewed the implementation of UHC in terms of empirical or modelling evaluations of the impacts on TB treatment and on epidemiological or economic outcomes. Since social health insurance (SHI) and close to community/home health provision (CTCH) are common practices of UHC, I specified the related terms as keywords. In view of relating TB and SES with the keywords relate to UHC, I searched using the following terms, **tuberculosis AND ("universal healthcare" OR "universal health coverage" OR "universal coverage" OR "universal care" OR "social insurance" OR "social health insurance" OR "compulsory insurance" OR "close to community" OR "close to home")**. Following this, I retrieved 59 initial search records published before July 2019 from the Web of Science.

Results: universal health coverage

As Figure 2.2 shows, I retained three studies after applying the following exclusion criteria: (1) did not highlight the impact of UHC on TB and (2) did not evaluate the impact of UHC. One of these studies were about SHI and two were related CTCH.

Social health insurance

SHI is provided to mitigate the financial burden of underprivileged people when they are seeking healthcare. SHI is a common approach to achieve UHC and secure the fundamental human right to good health. Ideally, a fully functioning SHI should be able to cover all the catastrophic costs related to TB diagnostics and treatments. However, my search only identified one empirical study that evaluated the effect of SHI implementation on TB outcomes. The focus of the study was on how Indonesia extended its social health insurance programme to private healthcare providers in 2014, which presented a joint SHI and private-public mix to relieve the financial burden of TB patients. Here, Fuady et al. [28] surveyed TB patients following the policy implementation and found that the catastrophic costs of TB diagnostics and treatment were lower for poor people than for the non-poor. However, the researchers asserted that the SHI with private-public mix could not ensure zero catastrophic costs.

Close to community/home health

SHI is a passive SES-based intervention because SHI it works when the ill people have the intention to seek care. However, as a further aspect of UHC, CTCH aims to provide healthcare in more active terms. Here, healthcare workers access

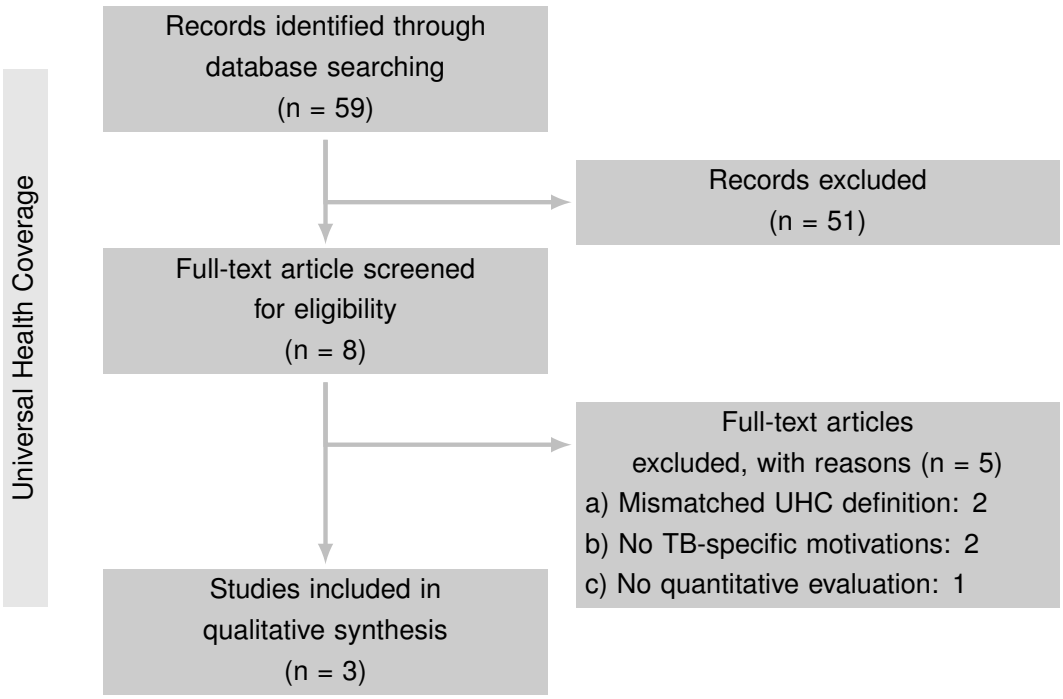


Figure 2.2: PRISMA flow diagram: Universal Health Coverage

community or homes to identify healthcare the needs or to eliminate the travelling costs. In fact, CTCH is directed at the areas where people face high travelling costs in terms of either money or time and where the severity of the illness is often misjudged. Overall, it provides an additional link between the underprivileged population and the healthcare system.

In Ethiopia, a CTCH-based, health extension program was launched in 2003. The health extension workers (HEWs) were deployed to provide essential preventive health services in the community. With this, the TB REACH project attempted to introduce TB-related services to the routine tasks of the HEWs. Looking at the provider end, Datiko et al. [29] qualitatively surveyed the reactions to the project from the HEWs, supervisors, and laboratory technicians. In general, the HEWs and supervisors had positive reactions in that they believed the new tasks could stimulate the working environments and could increase their connection to the communities. However, from the perspective of the laboratory technicians, the collection of samples with standard quality by the HEWs remained a challenge. In a follow-up study on the outcomes, Datiko et al. [30] confirmed that the project had a substantially improved on the case detection and the treatment outcome.

2.3.2 Social Protection, SP

SP programmes related to TB intend to provide supportive environments that encourage TB patients to complete their TB treatment. Unlike UHC, SP firmly targets people with specific conditions. In the TB programmes, the conditions generally refer to attending TB treatments regularly or participating in health education events. With the focus on TB-affected households, the practices tend to be more TB-specific than UHC practices, while the intervention effects are easier to address. In echoing the risk factors for non-adherence to TB treatment, SP presents an intuitive approach to improving TB patients' outcomes.

Searching and screening

I reviewed the SP implementation through empirical or modelling evaluations of the impacts on TB treatment and on epidemiological or economic outcomes. In view of relating TB and SES to the keywords related to SP, I searched using the following terms, **tuberculosis AND ("social protection" OR "social support" OR "cash transfer" OR "financial support")**. Following this I retrieved 244 studies among the initial search records published before July 2019 from the Web of Science.

Results: social protection

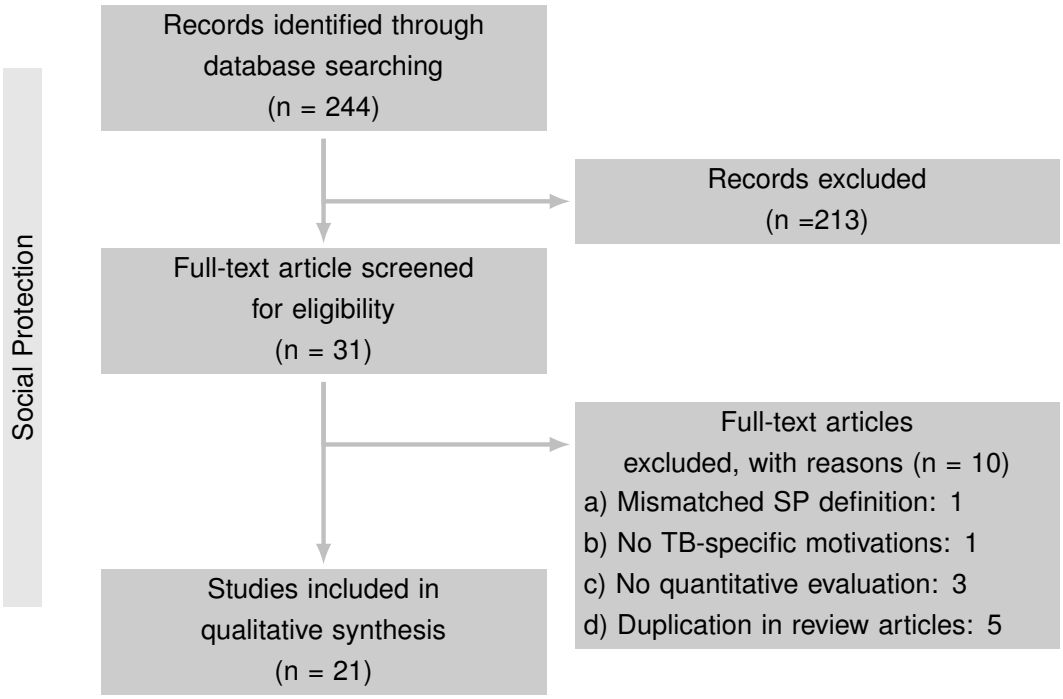


Figure 2.3: PRISMA flow diagram: Social Protection

As Figure 2.3 shows, I retrieved 21 studies after excluding the studies that: (1) did not highlight the impact of SP on TB outcome, (2) did not evaluate the impacts of SP programmes, and (3) had previously been considered in collected review articles. Here, I identified two large-scale social protection programmes in the selected studies: the Bolsa Familia Programme (BFP) in Brazil (eight studies) and the Community Randomized Evaluation of a Socioeconomic Intervention to Prevent TB (CRESIPT) in Peru (two studies). To ensure that I formed a comprehensive summary of these two programmes, I separated them from the summary of the others. With regard to the other studies, I identified four review articles and two research articles that assessed TB treatment outcomes under SP, two modelling studies that evaluated the economic outcomes and three modelling studies that evaluated the epidemiological outcomes.

Treatment outcomes

One direct impact of SP is the improvement of treatment outcomes. All the reviews [31–34] in my collection investigated the treatment adherence and outcomes in relation to SP programmes. Although in practice SP highly depends on the regional or cultural norms, the reviews consistently confirmed that the SPs were positively

associated with treatment success or adherence. The SPs are either general or are specifically related to food assistance [31], or to economic [34], socioeconomic, or psycho-emotional [32] aspects, particularly in low- or middle-income settings [33]. Two studies [35, 36] published after the reviewed articles confirmed the consistency of the overall effectiveness.

Economic outcomes

The SP that is focused on financial support is firmly aimed at mitigating the catastrophic costs. Based on the aggregated national data of seven countries, Rudgard et al. [37] evaluated and compared two financial support schemes for TB control. One provided financial support to TB patients in low SES groups while the other focused on low SES groups that involved high TB risks. The results suggested that the former will be overwhelmingly effective and will require lower costs in eliminating the burden of the catastrophic costs due to TB. Fuady et al. [38] built a simulation model based on a statistical analysis of TB patient data from Indonesia to study tested various combinations of financial support in terms of income loss, food assistance, and transportation support. Their results suggested that financial support alone cannot reduce the catastrophic costs for TB-affected households to zero.

Epidemiological outcome

In addition to the direct effects on individual treatment and economic outcomes, SP programmes can indirectly influence the dynamics of TB. For example, the health education provided to TB patients may spread to their household or to other affected households and even to the community as a whole, meaning the overall awareness of TB within a community will increase. Evaluating these externalities requires looking at the outcomes at a population level. In view of this, I found three modelling studies that had evaluated the total effect of SP implementation based on aggregated datasets. Here, Reeves et al. [39] investigated the relationship between the expenditure on SP policies and the different epidemiological outcomes in 21 European countries. The model's predictions show strong preventive effects on incremental SP spending in terms of TB notification, incidence, and mortality after adjusting for gross domestic product and general public health spending. Following Reeves et al. [39], Siroka et al. [40], applied the analysis to the global scale. Here, the authors addressed how SP spending that accounts for 1% of GDP could significantly reduce the incidence and prevalence of TB as well as the attendant mortality, while diminishing returns to increasing SP

expenditure were also found in the study. Elsewhere, Carter et al. [41] developed a statistical model to address how the change in poverty levels correlated with the global incidence of TB, in responding to the poverty reduction in SDG. Through widely extending SP coverage, a three-quarter reduction in the global incidence of TB could be expected by 2035.

BFP and CRESIPT

BFP is a national programme in Brazil that was launched in 2003 with the aiming of using various conditional cash transfers to eliminate poverty across the country [42]. I found eight studies related to BFP that fitted my criteria. Here, Durovni et al. [43] and Oliosi et al. [44] found that being covered by BFP correlated to a higher chance of treatment success in relation to dropout rate and death during treatment. More specifically, Reis-Santos et al. [45] indicated that BFP leads to higher cure rates for the patients who received the cash transfers. Also, I identified several studies addressing the epidemiological effects. Lower TB incidence and lower mortality rates were found in populations participating in BFP after adjusting for social groups [46–49]. I also found a simulation modelling study from, Boccia et al. [50], who used the findings from the aforementioned BFP studies in various terms.

Focusing on a specific SES group, CRESIPT was a randomised clustering control project that was conducted in the shantytowns of Lima, Peru between 2014 and 2015 [51]. Similar to BFP, CRESIPT provided TB patients with financial support. In addition, CRESIPT employed monthly community meetings, provided TB-related health education and focus group discussions to the patients and their close contacts. If they attended the meeting, they would receive extra assistance with food vouchers. I identified two studies that had evaluated the effect of CRESIPT. Here, the general consensus was that the intervention served the purpose of reducing the catastrophic costs of the TB-affected households [52] and improved the clinical treatment outcomes [53]. However, to the best of my knowledge, the epidemiological outcome of CRESIPT has not yet been revealed.

2.4 Discussion

Utilising three scoping reviews, I have identified the studies that have assessed the links between SES and TB-related care-seeking and the potential impact of SES-related interventions on TB control. In summary,

- the effects of SES on care-seeking prior to treatment are different depending

on the setting, while, in general, higher SES usually implies better adherence to TB treatment.

- Meanwhile, the influence of UHC programmes on TB have been poorly evaluated. While increasing the coverage may improve the accessibility of TB-related services may also raise concerns regarding service quality.
- The implementation of SP programmes is varied, but the preventive effects on different aspects of TB were consistently confirmed by the existing literature.

Notably, the studies related to BFP demonstrated the existence of an ecological system in this field, which includes evidence, interventions, and evidence of interventions. First, I identified the studies that had assessed the effects of SES on various care-seeking behaviours in Brazil. This included the compliance to TB treatment, to DOT, and to preventive latent TB treatment by TB patients in general or by those who had specific attributes. Secondly, after considering the implementation of BFP, the intervention effects on the cure rate, treatment adherence, and the epidemiological outcomes were evaluated. Lastly, a modelling study [50] identified that had developed a data-driven modelling scheme to utilise the evidence for intervention projections. However, one crucial drawback of the BFP studies related to the nature of the programme. In short, BFP is a national programme, meaning the effects of the programme were based on ecological studies, which implied that the effects of BFP could not be explicitly isolated from the other TB control approaches. In addition, one conditionality for receiving a cash transfer from BFP is that health checkups and health and nutrition seminars are regularly attended. These approaches imply that BFP is aimed at increasing healthcare coverage, the effects of which cannot be identified solely through the cash transfers. Similarly, in terms of the CRESIPT project in Lima, Peru, a great proportion of the empirical studies I found were conducted in Peru. In fact, five out of the 15 studies collected byThomas et al. [25] were related to the Peruvian capital, Lima.

The literature that addressed care-seeking before and after treatment revealed a certain imbalance. In fact, most of the studies were related to SES in terms of risk factors for TB treatment non-adherence or in terms of improving the adherence through SES-related interventions. Furthermore, I did not find any review articles that summarised the relationship between SES and treatment delay. The reason for this imbalance can be attributed to data availability. Before TB is diagnosed or before the treatment is initiated, the patients are not technically “TB patients” yet, which means the TB-specific programmes do not cover them and thus their interactions with the healthcare system would not be recorded. Therefore, routinely

collected data on more generally healthcare purposes or retrospective surveys are required for assessing the care-seeking behaviours that precede TB confirmation. Similarly, the evaluations of UHC programmes which do not have to consider TB-specific issues were scarce compared to those related to SP programmes.

In this chapter, I aimed at providing an elevated view of the context of interest rather than gaining an in-depth understanding of a specific outcome. Therefore, I did not search in broader terms, merely restricting my search to the Web of Science and did not search more broadly. Also, I used a narrative technique to summarise the selected studies rather than meta-synthesis in terms of measurement quality. In short, I merely required a clarity of definitions and a good fit to my area of interest.

To bring the non-intervention evidence into contact with behaviour change motivations, I introduced the notion of TPB. Prior to that, I had considered the Health Belief Model (HBM) [54, 55], which related to TB diagnosis in its early development. However, under the framework of the HBM, I could not appropriately locate peer influence, stigma, or health education in communities, so I opted to apply the TPB, which involves the components of subjective/social norms as a component. Also, the information-motivation-behavioural skills (IMB) model, which has applications in treatment adherence in the sense of psychology, was a potential behavioural theory [56, 57]. Here, the "motivation" term can be regarded as the integration of attitude and subjective norms into the TPB, while the "information" aspect summarises the knowledge related to the behaviour of interest, which I classified as the attitude term in the TPB. The behavioural skills, however, focus on self-efficacy rather than physical barriers such as income loss and travel cost.

Spatial inequalities were related to TB incidence or risk of TB. For example, a subnational spatial analysis, Pereira et al. [58], in Brazil found positive effects of the social development index and per capita income on lower TB incidence rates. In this review, I attempted to use the attitude or social norms in the TBP to summarise the spatial or contexture effects. However, the retrieved studies were usually conducted on specific sites or specific vulnerable population. They shared homogenous background features, such as culture and beliefs. I suggested this is a reason why the cultural effects were generally addressed in the qualitative studies, e.g. Cremers et al. [3] and Saqib et al. [8], but hardly evaluated in the quantitative analysis. Therefore, to address the spatial effects on care-seeking behaviours, more subnational or international studies are required.

In a number of selected studies, I found the influence of risk perception and the stigma of TB. These two factors can be regarded as socioeconomic factors or medi-

ators of as the influence of the other socioeconomic factors. To simplify my search, I did not include studies that assessed their upstream factors. However, as Craig et al.'s [59] review revealed, highlighting the actions to reduce the stigmatisation caused by SES has rarely been addressed.

In the Western Pacific region, the adult TB incidence rate tends to increase with age. This is especially the case in China, Japan, Singapore, South Korea, and Taiwan that the TB suffers account for the major part of TB-related burden. The missing patients and the non-adherent to treatments in this region remains a huge concern. However, the available studies related to SES-based TB interventions only targeted children and young adults. In short, the income-based SES indices are not applicable for the retired. Thus, shifting the focus to regional SES, saving-based SES, asset-based SES, and psycho-socio-economic indices would be more appropriate in these settings. Further research on care-seeking behaviours should perhaps consider savings, investment income, living with children, or social event engagement.

Bibliography

[1] Lönnroth K, Jaramillo E, Williams BG, Dye C, Raviglione M. Drivers of tuberculosis epidemics: The role of risk factors and social determinants; 2009.

[2] Ajzen I. The theory of planned behavior. Organ Behav Hum Decis Process. 1991 Dec;50(2):179–211.

[3] Cremers AL, Gerrets R, Kapata N, Kabika A, Birnie E, Klipstein-Grobusch K, et al. Tuberculosis patients' pre-hospital delay and non-compliance with a longstanding DOT programme: a mixed methods study in urban Zambia. BMC Public Health. 2016 Oct;16(1):1130.

[4] Christian C, Burger C, Claassens M, Bond V, Burger R. Patient predictors of health-seeking behaviour for persons coughing for more than two weeks in high-burden tuberculosis communities: the case of the Western Cape, South Africa. BMC Health Serv Res. 2019 Mar;19(1):160.

[5] Laohasiriwong W, Mahato RK, Koju R. Health System Delay among the Pulmonary Tuberculosis Patients Presenting in the DOTS Centers of Nepal. J Clin Diagn Res. 2016 Jun;10(6):LM01–3.

[6] Westerlund EE, Tovar MA, Lönnermark E, Montoya R, Evans CA. Tuberculosis-related knowledge is associated with patient outcomes in shantytown residents; results from a cohort study, Peru. J Infect. 2015 Sep;71(3):347–357.

[7] Said K, Hella J, Mhalu G, Chiryankubi M, Masika E, Maroa T, et al. Diagnostic delay and associated factors among patients with pulmonary tuberculosis in Dar es Salaam, Tanzania. Infect Dis Poverty. 2017 Mar;6(1):64.

[8] Saqib SE, Ahmad MM, Amezcua-Prieto C, Virginia MR. Treatment Delay among Pulmonary Tuberculosis Patients within the Pakistan National Tuberculosis Control Program. Am J Trop Med Hyg. 2018 Jul;99(1):143–149.

[9] Bonadonna LV, Saunders MJ, Guio H, Zegarra R, Evans CA. Socioeconomic and Behavioral Factors Associated with Tuberculosis Diagnostic Delay in Lima, Peru. Am J Trop Med Hyg. 2018 Jun;98(6):1614–1623.

[10] Mendonça AMC, Kritski AL, Land MGP, Sant'Anna CC. Abandonment of Treatment for Latent Tuberculosis Infection and Socioeconomic Factors in Children and Adolescents: Rio De Janeiro, Brazil. PLoS One. 2016 May;11(5):e0154843.

[11] Shayo GA, Moshiro C, Aboud S, Bakari M, Mugusi FM. Acceptability and adherence to Isoniazid preventive therapy in HIV-infected patients clinically screened for latent tuberculosis in Dar es Salaam, Tanzania. BMC Infect Dis. 2015 Aug;15:368.

[12] Makanjuola T, Taddese HB, Booth A. Factors associated with adherence to treatment with isoniazid for the prevention of tuberculosis amongst people living with HIV/AIDS: a systematic review of qualitative data. PLoS One. 2014 Feb;9(2):e87166.

[13] Stanton EA. The human development index: A history. PERI Working Papers. 2007;p. 85.

[14] Maciel EMGdS, Amancio JdS, Castro DBd, Braga JU. Social determinants of pulmonary tuberculosis treatment non-adherence in Rio de Janeiro, Brazil. PLoS One. 2018 Jan;13(1):e0190578.

[15] Alegria-Flores K, Weiner BJ, Wiesen CA, Lich KLH, Van Rie A, Paul JE, et al. Innovative approach to the design and evaluation of treatment adherence interventions for drug-resistant TB. Int J Tuberc Lung Dis. 2017 Nov;21(11):1160–1166.

[16] Birch S, Govender V, Fried J, Eyles J, Daries V, Moshabela M, et al. Does treatment collection and observation each day keep the patient away? An analysis of the determinants of adherence among patients with Tuberculosis in South Africa. Health Policy Plan. 2016 May;31(4):454–461.

[17] Choi H, Chung H, Muntaner C, Lee M, Kim Y, Barry CE, et al. The impact of social conditions on patient adherence to pulmonary tuberculosis treatment. Int J Tuberc Lung Dis. 2016 Jul;20(7):948–954.

[18] Diefenbach-Elstob T, Plummer D, Dowi R, Wamagi S, Gula B, Siwaeya K, et al. The social determinants of tuberculosis treatment adherence in a remote region of Papua New Guinea. BMC Public Health. 2017 Jan;17(1):70.

[19] Herrero MB, Arrossi S, Ramos S, Braga JU. Social determinants of nonadherence to tuberculosis treatment in Buenos Aires, Argentina. Cad Saude Publica. 2015 Sep;31(9):1983–1994.

[20] Saleh Jaber AA, Khan AH, Syed Sulaiman SA. Evaluation of tuberculosis defaulters in Yemen from the perspective of health care service. J Pharm Health Serv Res. 2018 Dec;9(4):381–392.

[21] Lohiniva AL, Mokhtar A, Azer A, Elmoghazy E, Kamal E, Benkirane M, et al. Qualitative interviews with non-national tuberculosis patients in Cairo, Egypt: understanding the financial and social cost of treatment adherence. Health Soc Care Community. 2016 Nov;24(6):e164–e172.

[22] Theron G, Peter J, Zijenah L, Chanda D, Mangu C, Clowes P, et al. Psychological distress and its relationship with non-adherence to TB treatment: a multicentre study. BMC Infect Dis. 2015 Jul;15:253.

[23] Prado TNd, Rajan JV, Miranda AE, Dias EDS, Cosme LB, Possuelo LG, et al. Clinical and epidemiological characteristics associated with unfavorable tuberculosis treatment outcomes in TB-HIV co-infected patients in Brazil: a hierarchical polytomous analysis. Braz J Infect Dis. 2017 Mar;21(2):162–170.

[24] Tola HH, Tol A, Shojaeizadeh D, Garmaroudi G. Tuberculosis Treatment Non-Adherence and Lost to Follow Up among TB Patients with or without HIV in Developing Countries: A Systematic Review. Iran J Public Health. 2015 Jan;44(1):1–11.

[25] Thomas BE, Shanmugam P, Malaisamy M, Ovung S, Suresh C, Subbaraman R, et al. Psycho-Socio-Economic Issues Challenging Multidrug Resistant Tuberculosis Patients: A Systematic Review. PLoS One. 2016 Jan;11(1):e0147397.

[26] Hargreaves JR, Boccia D, Evans CA, Adato M, Petticrew M, Porter JDH. The social determinants of tuberculosis: from evidence to action. Am J Public Health. 2011 Apr;101(4):654–662.

[27] Lönnroth K, Glaziou P, Weil D, Floyd K, Uplekar M, Raviglione M. Beyond UHC: monitoring health and social protection coverage in the context of tuberculosis care and prevention. PLoS Med. 2014 Sep;11(9):e1001693.

[28] Fuady A, Houweling TAJ, Mansyur M, Richardus JH. Catastrophic total costs in tuberculosis-affected households and their determinants since Indonesia's implementation of universal health coverage. Infect Dis Poverty. 2018 Jan;7(1):3.

[29] Datiko DG, Lindtjørn B. Cost and cost-effectiveness of smear-positive tuberculosis treatment by Health Extension Workers in Southern Ethiopia: a community randomized trial. PLoS One. 2010 Feb;5(2):e9158.

[30] Datiko DG, Lindtjørn B. Health Extension Workers Improve Tuberculosis Case Detection and Treatment Success in Southern Ethiopia: A Community Randomized Trial; 2009.

[31] de Pee S, Grede N, Mehra D, Bloem MW. The enabling effect of food assistance in improving adherence and/or treatment completion for antiretroviral therapy and tuberculosis treatment: a literature review. AIDS Behav. 2014 Oct;18 Suppl 5(5):S531–41.

[32] van Hoorn R, Jaramillo E, Collins D, Gebhard A, van den Hof S. The Effects of Psycho-Emotional and Socio-Economic Support for Tuberculosis Patients on Treatment Adherence and Treatment Outcomes - A Systematic Review and Meta-Analysis. PLoS One. 2016 Apr;11(4):e0154095.

[33] Andrade KVFd, Nery JS, Souza RAd, Pereira SM. Effects of social protection on tuberculosis treatment outcomes in low or middle-income and in high-burden countries: systematic review and meta-analysis. Cad Saude Publica. 2018 Feb;34(1):e00153116.

[34] Richterman A, Steer-Massaro J, Jarolimova J, Luong Nguyen LB, Werdenberg J, Ivers LC. Cash interventions to improve clinical outcomes for pulmonary tuberculosis: systematic review and meta-analysis. Bull World Health Organ. 2018 Jul;96(7):471–483.

[35] Ukwaja KN, Alobu I, Gidado M, Onazi O, Oshi DC. Economic support intervention improves tuberculosis treatment outcomes in rural Nigeria. Int J Tuberc Lung Dis. 2017 May;21(5):564–570.

[36] Priedeman Skiles M, Curtis SL, Angeles G, Mullen S, Senik T. Evaluating the impact of social support services on tuberculosis treatment default in Ukraine. PLoS One. 2018 Aug;13(8):e0199513.

[37] Rudgard WE, Evans CA, Sweeney S, Wingfield T, Lönnroth K, Barreira D, et al. Comparison of two cash transfer strategies to prevent catastrophic costs for poor tuberculosis-affected households in low- and middle-income countries: An economic modelling study. PLoS Med. 2017 Nov;14(11):e1002418.

[38] Fuady A, Houweling TAJ, Mansyur M, Burhan E, Richardus JH. Effect of financial support on reducing the incidence of catastrophic costs among

tuberculosis-affected households in Indonesia: eight simulated scenarios. Infect Dis Poverty. 2019 Feb;8(1):10.

[39] Reeves A, Basu S, McKee M, Stuckler D, Sandgren A, Semenza J. Social protection and tuberculosis control in 21 European countries, 1995-2012: a cross-national statistical modelling analysis. Lancet Infect Dis. 2014 Nov;14(11):1105–1112.

[40] Siroka A, Ponce NA, Lönnroth K. Association between spending on social protection and tuberculosis burden: a global analysis. Lancet Infect Dis. 2016 Apr;16(4):473–479.

[41] Carter DJ, Glaziou P, Lönnroth K, Siroka A, Floyd K, Weil D, et al. The impact of social protection and poverty elimination on global tuberculosis incidence: a statistical modelling analysis of Sustainable Development Goal 1. Lancet Glob Health. 2018 May;6(5):e514–e522.

[42] Campello T, Neri MC. Bolsa Família Program-a decade of social inclusion in Brazil. Ipea; 2013.

[43] Durovni B, Saraceni V, Puppin MS, Tassinari W, Cruz OG, Cavalcante S, et al. The impact of the Brazilian Family Health Strategy and the conditional cash transfer on tuberculosis treatment outcomes in Rio de Janeiro: an individual-level analysis of secondary data. J Public Health. 2018 Sep;40(3):e359–e366.

[44] Oliosi JGN, Reis-Santos B, Locatelli RL, Sales CMM, da Silva Filho WG, da Silva KC, et al. Effect of the Bolsa Familia Programme on the outcome of tuberculosis treatment: a prospective cohort study. The Lancet Global Health. 2019 Feb;7(2):e219–e226.

[45] Reis-Santos B, Shete P, Bertolde A, Sales CM, Sanchez MN, Arakaki-Sanchez D, et al. Tuberculosis in Brazil and cash transfer programs: A longitudinal database study of the effect of cash transfer on cure rates. PLoS One. 2019 Feb;14(2):e0212617.

[46] Nery JS, Rodrigues LC, Rasella D, Aquino R, Barreira D, Torrens AW, et al. Effect of Brazil's conditional cash transfer programme on tuberculosis incidence. Int J Tuberc Lung Dis. 2017 Jul;21(7):790–796.

[47] de Souza RA, Nery JS, Rasella D, Guimarães Pereira RA, Barreto ML, Rodrigues L, et al. Family health and conditional cash transfer in Brazil and its effect on tuberculosis mortality. Int J Tuberc Lung Dis. 2018 Nov;22(11):1300–1306.

[48] Malacarne J, Kolte IV, Freitas LP, Orellana JDY, Souza MLPd, Souza-Santos R, et al. Factors associated with TB in an indigenous population in Brazil: the effect of a cash transfer program. Rev Inst Med Trop Sao Paulo. 2018 Oct;60:e63.

[49] Pelissari DM, Rocha MS, Bartholomay P, Sanchez MN, Duarte EC, Arakaki-Sanchez D, et al. Identifying socioeconomic, epidemiological and operational scenarios for tuberculosis control in Brazil: an ecological study. BMJ Open. 2018 Jun;8(6):e018545.

[50] Boccia D, Rudgard W, Shrestha S, Lönnroth K, Eckhoff P, Golub J, et al. Modelling the impact of social protection on tuberculosis: the S-PROTECT project. BMC Public Health. 2018 Jun;18(1):786.

[51] Wingfield T, Boccia D, Tovar MA, Huff D, Montoya R, Lewis JJ, et al. Designing and implementing a socioeconomic intervention to enhance TB control: operational evidence from the CRESIPT project in Peru. BMC Public Health. 2015 Aug;15:810.

[52] Wingfield T, Tovar MA, Huff D, Boccia D, Montoya R, Ramos E, et al. The economic effects of supporting tuberculosis-affected households in Peru. Eur Respir J. 2016 Nov;48(5):1396–1410.

[53] Wingfield T, Tovar MA, Huff D, Boccia D, Montoya R, Ramos E, et al. A randomized controlled study of socioeconomic support to enhance tuberculosis prevention and treatment, Peru. Bull World Health Organ. 2017 Apr;95(4):270–280.

[54] Becker MH. The health belief model and personal health behavior. vol. 2. Slack; 1974.

[55] Rosenstock IM. Historical Origins of the Health Belief Model. Health Educ Monogr. 1974 Dec;2(4):328–335.

[56] Fisher WA, Fisher JD, Harman J. The information-motivation-behavioral skills model: A general social psychological approach to understanding and promoting health behavior. Social psychological foundations of health and illness. 2003;82:106.

[57] Fisher JD, Fisher WA, Amico KR, Harman JJ. An information-motivation-behavioral skills model of adherence to antiretroviral therapy. Health Psychol. 2006 Jul;25(4):462–473.

[58] Pereira AGL, Medronho RdA, Escosteguy CC, Valencia LIO, Magalhães MdAFM. Spatial distribution and socioeconomic context of tuberculosis in Rio de Janeiro, Brazil. Rev Saude Publica. 2015 Aug;49:48.

[59] Craig GM, Daftary A, Engel N, O'Driscoll S, Ioannaki A. Tuberculosis stigma as a social determinant of health: a systematic mapping review of research in low incidence countries. Int J Infect Dis. 2017 Mar;56(SI):90–100.

Chapter 3

Review: health-related hybrid modelling

3.1 Introduction

Agent-based models (ABMs) and equation-based models (EBMs) span a spectrum in the field of simulation modelling. At the ABM or discrete event simulation (DES) end, all the individuals are modelled as entities, agents, individuals, building up a bottom-up system. Agents have their own attributes and rules for interacting with an environment. Placing them in an environment of interest, their rules of actions and interactions can jointly describe scenarios of interest. I can then observe the system at the individual level, population level, or a more aggregated view. At the EBM or system dynamics (SD) end, a system is described by a set of equations. Equations capture the behaviours which drive the dynamics of a system. In an EBM, I calculate variables of interest with mathematical functions. The rules-of-thumb, advantages, and disadvantages of choosing between ABMs and EBMs have been discussed in the literature [1–3]. In general, ABMs can deal with a wider variety in non-linear behaviour, stochastic terms, and data inputs with heterogeneity in individual-levels, and EBMs simplify systems and keep clarity in model specification and requires less computation resources. However, both of them have apparent drawbacks: ABMs are often computationally expensive and data-hungry; EBMs may require strong assumptions which depart from complex reality.

In fact, the paradigms of ABMs and EBMs are not incompatible with each other. Coupled simulation including both ABMs and EBMs has the potential to exploit the benefits of both. Brailsford et al. [4] showed two case studies of hybrid modelling in healthcare, anticipating that hybridising SD and DES can bring

healthcare modelling to the "holy grail". Heath et al. [5] started from a technical perspective to discuss ideas for mixing aspects of ABMs, DES, and SD in depth. Although the methodology was not precisely described, they concluded that introducing one approach within another is possible with proper adaptations. In practice, multi-scale and metapopulation modelling illustrate applications coupling ABMs and EBMs. Multi-scale modelling either uses EBMs to describe response to external shocks of many sub-models of an ABM, or constructs an ABM where every agent has an EBM embedded. For example, Banos et al. [6] used an EBM to control transportation among many ABM cities; Caudill and Lawson [7] modelled epidemics in hospital wards while every patient has an EBM of human-pathogen interaction embedded. Metapopulation models can define different populations using different modelling paradigms. For example, Manore et al. [8] modelled a mosquito-borne disease, with an equation-based mosquito layer and an agent-based human layer, the two layers could transmit the disease to each other.

Excluding multi-scale and metapopulation modelling, this chapter investigates a closer hybridisation between ABMs and EBMs in the health context. The hybrid modelling in this chapter indicates not only having both paradigms of ABMs and EBMs but also allowing a single individual switch between an agent and a count number in equations. This chapter is interested in examples that instantiated agents as more details were required and reduced them to variable values for improved computational feasibility.

3.2 Search strategy and data extraction

3.2.1 Query

I searched entries systematically in the electronic database Web of Science. I required only clarity in methodology and implementation. I constrained the search for publications written in English and published before July 2019. Both journal articles and conference proceedings were accepted. The query was formulated with the terms from the following three domains:

Set 1: Equation-based models (EBMs) An EBM describes the behaviour of a single entity through a dynamic system of equations. In infectious disease modelling, terms like compartmental models, differential equations and their derivatives are commonly used. In health services research or operational research, the terms "System dynamics" dominates the discourse. The keywords were combined as follows: **("equation based" OR "system dynamic*" OR "differential equation*" OR "Compartment*")**

Set 2: Agent-based models (ABMs) An agent-based model, alias multi-agent system, describes behaviours of a collection of multiple entities. In infectious disease modelling, since "agent" means pathogen as well, ABMs are sometimes called individual-based models. Again, considering potential articles in operational research, I regarded microsimulation and discrete event simulation as types of ABMs although in some contexts they are differentiated. The related keywords were composed as follows: **("agent based" OR "individual based" OR "multi agent" OR microsimulation OR "micro simulation" OR "discrete event simulation")**.

Set 3: Topics related to disease dynamics I targeted models related to disease dynamics but not restricted to any lifestyle factors, specific diseases, or medications. The terms were as follows: **("disease*" OR "medic*" OR "epidemiolog*" OR "health" OR "infect*")**

3.2.2 Selection criteria

In title and abstract screening, I included articles based on the following considerations:

Topics: I focused on behaviours related to disease dynamics. These could involve any process, such as the exposure process, self-protection, behaviour, and care-seeking. I did not consider models at microbiological or immunological scales. It is to be noted that, since I was not focusing on human-specific behaviour, studies for non-human populations were eligible if they met the other criteria.

Having an ABM and EBM: I required studies have both an ABM and EBM in a single simulation. I included them only if there was communication between the two paradigms.

Individuals switching between agents and values for equations: Although my search might fetch multi-scale and metapopulation models in the initial hits, I included them only if they had individuals transferring between the two paradigms.

Inclusion/exclusion based on the nature of methodology required an assessment of the detailed descriptions at the full-text stage

Model implementation: I was interested in how hybrid models work in practice, so I excluded the studies without a real example of implementation.

3.2.3 Data extraction

Focusing on technical details, I extracted three sets of methodological information from the selected studies. The first and the second were for ABMs and EBMs respectively: the terminology used for them, the types of modelled entities, stochasticity, discrete time or continuous time, the use of data. The last one captured the relation between the agent-based and equation-based components: what and how they communicated with each other, how they synchronised.

3.3 Results

3.3.1 Search results

Figure 3.1 shows the results of my search and selection process. Identifying from 319 initial searched result, I included 7 studies: 2 journal articles and 5 conference proceedings. The publication dates ranged from 2007 to 2016. The studies can be grouped into three fields: 2 disease outbreak models, 1 disease onset models and 4 care-seeking models.

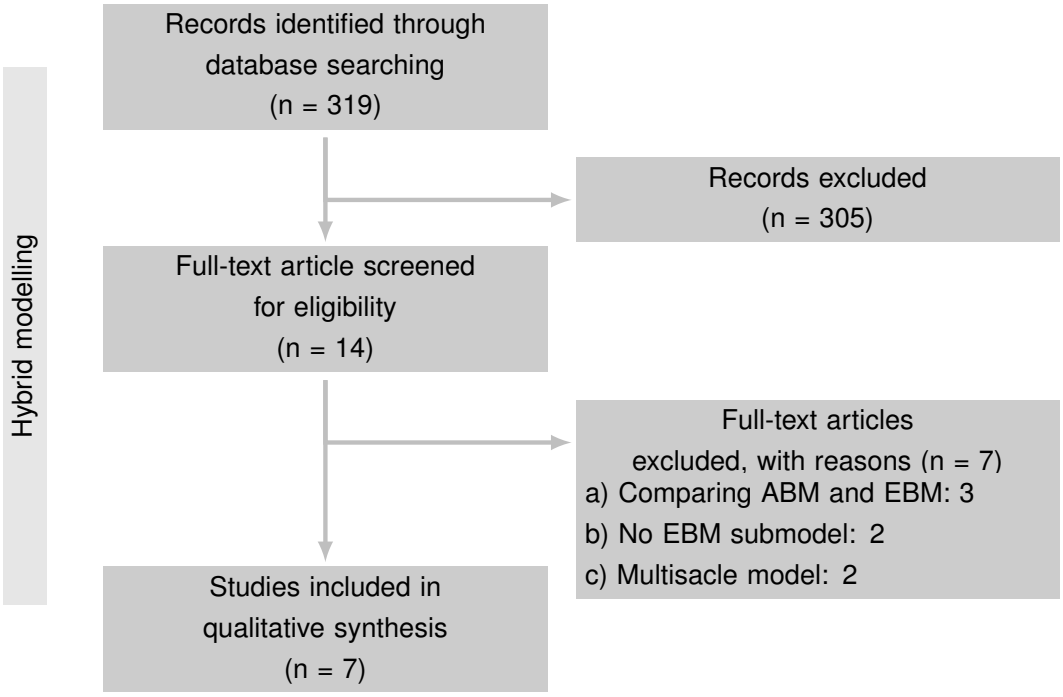


Figure 3.1: PRISMA Flow Diagram

3.3.2 Topics

Disease outbreak models

I found two studies using ABMs when outbreaks are initialising and EBMs when the infected populations reach threshold count values [9, 10]. Bobashev et al. [9] built a sequential modelling scheme: as the number of the infected agented reaches a threshold, the agents are summarised to determine the variables values of the EBM and the simulation switches to an EBM. The study claimed the ABM had better capability to capture the nature of an epidemic, while the EBM saves simulation resources and can be a good approximation if the epidemic is established. Using a similar concept about EBMs and ABMs, Alam et al. deployed a parallel hybrid scheme modelling HIV outbreaks in specific clusters alongside an endemic environment [10]. While the study focused on an intervention that is based on clusters defined by virus genetics, the model of HIV outside the target cluster did not require individual-level resolution.

Disease onset models

Leonenko et al. [11] demonstrated a hybrid modelling scheme utilising EBMs and ABMs according to enhance the detail of disease status represented. The framework was exemplified with TB and HIV epidemiology. The ABM part is in charge of the people with active diseases because the agents have complex health states and transition rules. The EBM part deals with the demographic dynamics and exports newly infected people to the ABM part. Then, the ABM part calculates and reports the force of infection in the system. The study aimed to speed up the ABM simulation by transferring the agents that are less interested to values in the EBM part.

Care seeking models

Focusing on the healthcare process, five of the collected studies modelled people in healthcare settings by ABMs, and otherwise by EBMs [12–15]. Prospective Health Technology Assessment (ProHTA) is a new approach evaluating innovative health technology with synthesised interdisciplinary knowledge before deployment into practice. To introduce hybrid modelling to ProHTA, Djanatilev et al. [12] developed a modularised framework. The framework standardised population dynamics, disease-specific changes, healthcare, and healthcare finance. Within them, only the healthcare part is agent-based simulation. The study showed the potential to use hybrid modelling in ProHTA. Extending from Djanatilev et al. [12], Djanatliev

and German [13] switched the implementations from directly to indirectly coupling (to be discussed later). The study improved the clarity of the job of each sub-model and improved the simulation efficacy. Mielczarek and Zabawa [14] assessed the impact of population dynamics on future healthcare demand. The model used an ABM to project the disease process of people with cardiovascular diseases and their respective healthcare needs. Lastly, Viana et al. [15] hybridised an infectious disease EBM and a clinic patient flow ABM. The ABM imported the patients from the EBM and adapted the healthcare capacity according to the epidemiology.

3.3.3 Coupling methods

I found two types of coupling protocol in the literature. Some allowed their submodels to listen to events emitted from each other directly (direct coupling). The others employ controllers to collect, summarise, and distribute information to submodels (indirect coupling).

Direct coupling

Some models allowed their submodels to send information directly to the other. Bobashev et al [9] did not simulate ABMs and EBMs simultaneously. The EBM summarised the information of the ABM after the epidemic reached the threshold to populate initial values, i.e., they used a one shot coupling. Mielczarek and Zabawa [14] used a unidirectional communication where the ABM fed the patients to the EBM every timestep without feedback. For the models with bidirectional communication Djanatliev et al. [12] and Leonenko et al. [11] applied a turn-based update where submodels used the information from the other models at the end of the previous timestep. Between two time points, the submodels were simulated independently.

Indirect coupling

The other studies employed central controllers to collect, summarise, and distribute information to submodels. The controllers might also validate the information to maintain the functionality of the submodels. Alam et al. [10] used a phylogenetic tree simulation model to connect the two submodels. The tree simulation directed the ABM according to the epidemiology of the EBM part. Viana et al. [15] provided a coupling method with lower technical requirements. The model had EBM and ABM parts built in different software and they communicated to each other via an external spreadsheet. In each time step, the two submodels

read and wrote the spreadsheet serially. Lastly, Djanatliev and German [13] provided an implementation for hybrid modelling based on a process-oriented simulation scheme. In the approach, the agents did not actively schedule events to the simulator but only respond to the events assigned by the central controller. The central controller scheduled every event for agents given the status of the EBM part. Therefore, the agents only take the events approved by the EBM part, and the ABM part does not require the details of the EBM part.

Information passing between EBMs and ABMs

Classically, EBMs are executed in discrete-time, and ABMs can be both executed in discrete-time (turn-based) or continuous-time (event-driven). For hybrid modelling, the information in discrete-time and continuous-time parts are always misaligned. Therefore, synchronising the information between the modelling paradigms plays an important role in the validity of hybrid modelling. For example, an action of an agent might be triggered if a variable x in an EBM reaches 4. At time 0, x is 3. After a timestep, x reaches 5 at time 1. Technically, x was in 4 at a certain timing between 0 and 1. However, the action of the agent is triggered until time 1. On the other hand, an EBM would require an extremely fine timestep to ensure events in an ABM could be included within a tolerable lag time.

All the studies found, apart from Bobashev et al. [9] as well as Mielczarek and Zabawa [14], required no bidirectional synchronisation. The others discretised the continuous-time components and applied a turn-based simulation. In particular, for the models using indirect coupling, the central controllers summarised the information to pass to each other in every timestep. Djanatliev et al. [12] discussed the issue in depth. The results compared the theoretical and the simulated population dynamics, showing a system fluctuation due to the synchronisation issue but the overall trend of the population was not biased.

3.4 Summary

Although hybrid modelling was encouraged by literature, applications in the health-related domain are still scarce. The applications showed that hybrid modelling allows high complexity while keeping the computational cost as low as possible. Studies ranged from theoretical work to applications with explicit healthcare settings, indicating the potential of hybrid modelling related to health. However, the implementation methodology was typically poorly described. Synchronising the information between EBMs and ABMs requires careful assessment. Information loss from the hybridisation procedure needs to be quantified.

Linking back to this thesis, I focus on the care-seeking of TB patients while I need the disease transmission process to investigate the influence of the care-seeking on population level transmission. The examples I found demonstrated the use of EBMs for population dynamics and ABMs for healthcare and disease progression processes. My population of interest without ongoing TB-related care-seeking can be modelled by an EBM and while the patients undergoing complex care-seeking processes and changes in their disease status could be detailed using an ABM.

Bibliography

[1] Parunak HVD, Savit R, Riolo RL. Agent-based modeling vs. equation-based modeling: A case study and users' guide. In: International Workshop on Multi-Agent Systems and Agent-Based Simulation. Springer; 1998. p. 10–25.

[2] Brailsford SC, Hilton NA. A comparison of discrete event simulation and system dynamics for modelling health care systems. 2001;.

[3] Brennan A, Chick SE, Davies R. A taxonomy of model structures for economic evaluation of health technologies. Health Econ. 2006 Dec;15(12):1295–1310.

[4] Brailsford SC, Desai SM, Viana J. Towards the Holy Grail: Combining System Dynamics and Discrete-event Simulation in Healthcare. In: Proceedings of the Winter Simulation Conference. WSC '10. Baltimore, Maryland: Winter Simulation Conference; 2010. p. 2293–2303.

[5] Heath SK, Buss A, Brailsford SC, Macal CM. Cross-paradigm Simulation Modeling: Challenges and Successes. In: Proceedings of the Winter Simulation Conference. WSC '11. Phoenix, Arizona: Winter Simulation Conference; 2011. p. 2788–2802.

[6] Banos A, Corson N, Gaudou B, Laperrière V, Coyrehourcq SR. The Importance of Being Hybrid for Spatial Epidemic Models:A Multi-Scale Approach. Systems. 2015 Nov;3(4):309–329.

[7] Caudill L, Lawson B. A hybrid agent-based and differential equations model for simulating antibiotic resistance in a hospital ward. In: 2013 Winter Simulations Conference (WSC); 2013. p. 1419–1430.

[8] Manore CA, Hickmann KS, Hyman JM, Foppa IM, Davis JK, Wesson DM, et al. A network-patch methodology for adapting agent-based models for directly transmitted disease to mosquito-borne disease. J Biol Dyn. 2015;9:52–72.

[9] Bobashev GV, Goedecke DM, Feng Yu, Epstein JM. A Hybrid Epidemic Model: Combining The Advantages Of Agent-Based And Equation-Based Approaches. In: 2007 Winter Simulation Conference; 2007. p. 1532–1537.

[10] Alam SJ, Zhang X, Romero-Severson EO, Henry C, Zhong L, Volz EM, et al. Detectable signals of episodic risk effects on acute HIV transmission: strategies for analyzing transmission systems using genetic data. Epidemics. 2013 Mar;5(1):44–55.

[11] Leonenko VN, Pertsev NV, Artzrouni M. Using High Performance Algorithms for the Hybrid Simulation of Disease Dynamics on CPU and GPU. In: Koziel S, Leifsson L, Lees M, Krzhizhanovskaya VV, Dongarra J, Sloot PMA, editors. International Conference on Computational Science, Iccs 2015 Computational Science at the Gates of Nature. vol. 51 of Procedia Computer Science; 2015. p. 150–159.

[12] Djanatliev A, German R, Kolominsky-Rabas P, Hofmann BM. Hybrid Simulation with Loosely Coupled System Dynamics and Agent-based Models for Prospective Health Technology Assessments. In: Laroque C, Himmelspach J, Pasupathy R, editors. Proceedings of the Winter Simulation Conference. WSC '12. Berlin, Germany: Winter Simulation Conference; 2012. p. 69:1–69:12.

[13] Djanatliev A, German R. Prospective Healthcare Decision-making by Combined System Dynamics, Discrete-event and Agent-based Simulation. In: Proceedings of the 2013 Winter Simulation Conference: Simulation: Making Decisions in a Complex World. WSC '13. Piscataway, NJ, USA: IEEE Press; 2013. p. 270–281.

[14] Mielczarek B, Zabawa J. Modeling healthcare demand using a hybrid simulation approach. In: 2016 Winter Simulation Conference (WSC). Winter Simulation Conference Proceedings; 2016. p. 1535–1546.

[15] Viana J, Brailsford SC, Harindra V, Harper PR. Combining discrete-event simulation and system dynamics in a healthcare setting: A composite model for Chlamydia infection. Eur J Oper Res. 2014 Aug;237(1):196–206.

Chapter 4

Review: health behavioral modelling in agent-based models

4.1 Introduction

Behaviour is central to many questions related to human health. Risk factors making substantial contributions to disease burden (e.g. smoking, obesity) involve lifestyle choices. Many of these behaviours may be influenced by peers through social networks of family and friends. Even after the disease has begun, behaviours such as healthcare seeking or adherence to medication often play a key role in determining outcomes. As such, health-related behaviour includes aspects that are relevant to determining both disease incidence and outcomes.

Modelling is increasingly used in public health as a means to bring together evidence, understand epidemiology and estimate the impact of interventions [1]. Agent-based models (ABMs) in particular have advantages in representing complex individual behaviours [2]. ABMs have great flexibility in the way they represent behaviour, incorporate heterogeneity, and can make model state spaces more manageable and less complex (e.g. in cases of history-dependence or multiple, overlapping conditions). It is not surprising, therefore, that there is widespread use of ABMs to investigate emergent behaviour among individuals who influence one another. ABMs have become a key tool in behavioural economics [3], and Robertson [4] has argued that they are a natural methodology for many questions in behavioural operational research. [1]

However, many different approaches to modelling health-related behaviour within ABMs have been attempted, with studies being published in a wide range of journals representing very different disciplines. I therefore sought to systematically review this literature to summarise the approaches used, with particular

focus on the way behavioural rules for agents have been implemented, the determinants modelled as influencing behaviour, and the conceptual behavioural models employed. Based on this analysis, I identify gaps in utilisation of certain methods and suggest areas for future work.

4.2 Search strategy and data extraction

4.2.1 Search strategy and query

I searched entries systematically in the electronic databases Pubmed, Web of Science and Scopus in order to cover studies from the medical domain and a wide range of other disciplines. I did not require studies with well-validated data or results but only clarity in methodology, so journal articles and conference proceedings were both admissible.

I formulated my search terms around three aspects: ABM, health, and decision making. Specifically speaking, my query was: **“agent based" OR “individual based" OR “multi agent" OR microsimulation OR “discrete event simulation" OR “computational economics"), (health OR disease* OR epidemi* OR medic* OR hospital OR illness OR well-being) and (Behavio* OR Decision)**. All studies before 1 January 2019 published in English with full text available were screened.

4.2.2 Selection criteria

At title and abstract screening, I included articles based on the following eligibility criteria:

Model type: I only included models constructed from individual-level units and with temporal dynamics. In the first clause in my query above, microsimulation, agent-based models, and their relatives were included.

Human agent: The individuals whose behaviour is modelled should be human. For example, models of influenza transmission from birds to humans that focused on bird behaviours were excluded.

External environment: I only included studies that described models involving human interaction with the external environment, for example human to human or human to health care system interaction. Models only considering within-individual processes (e.g. immune response to pathogens and pharmacodynamics) were excluded.

Implementation: Review and conceptual articles which do not include implementation of the model or example simulations were excluded.

Following title and abstract screening, full articles were accessed and the following inclusion/exclusion based on the nature of behavioural modelling were applied:

Type of health-related behaviour: The health-related behaviours of interest in this chapter were defined *a priori* as: 1) behaviours that are known factors influencing the risk of ill health, positively or negatively (e.g. smoking, physical activity and diet); 2) behaviours specifically intended to prevent diseases, such as vaccination and voluntary screening; and 3) decisions made by patients (e.g. care seeking and returning to work during recovery). Articles not related to these categories of behaviour were excluded.

Agent-oriented decision making: This review is interested in agent-oriented behaviours so I excluded the behaviours of decision makers as opposed to individuals. For example, the design of emergency department workflows by hospital managers and disease diagnosis by clinicians would both be excluded. Decisions of parents to vaccinate their children were included because the parents are affected directly if their children get infected.

Contact behaviour: Models of social contact patterns or sexual partnership formation (relevant to infectious disease transmission) were explicitly excluded unless the contact behaviours depend on health status or perceived health status. For instance, behaviour around sexual partner formation that considers risks of sexual transmitted diseases would be included; daily commuting patterns, which may influence the transmission of infectious disease would be excluded.

4.2.3 Data extraction

Through iterative piloting, I developed an appraisal form for extracting information from collected studies. The form had three sets of questions. The first set was about general information: type of behaviour, terminology used for ABM, purpose of modelling, and conceptual behavioural model (conceptual driver for behaviour). The second set examined the determinants of behaviour considered. The third set examined implementation details: variable type of behaviour, method for generating behaviour, method for concluding determinants and data use.

4.3 Results

4.3.1 Search results

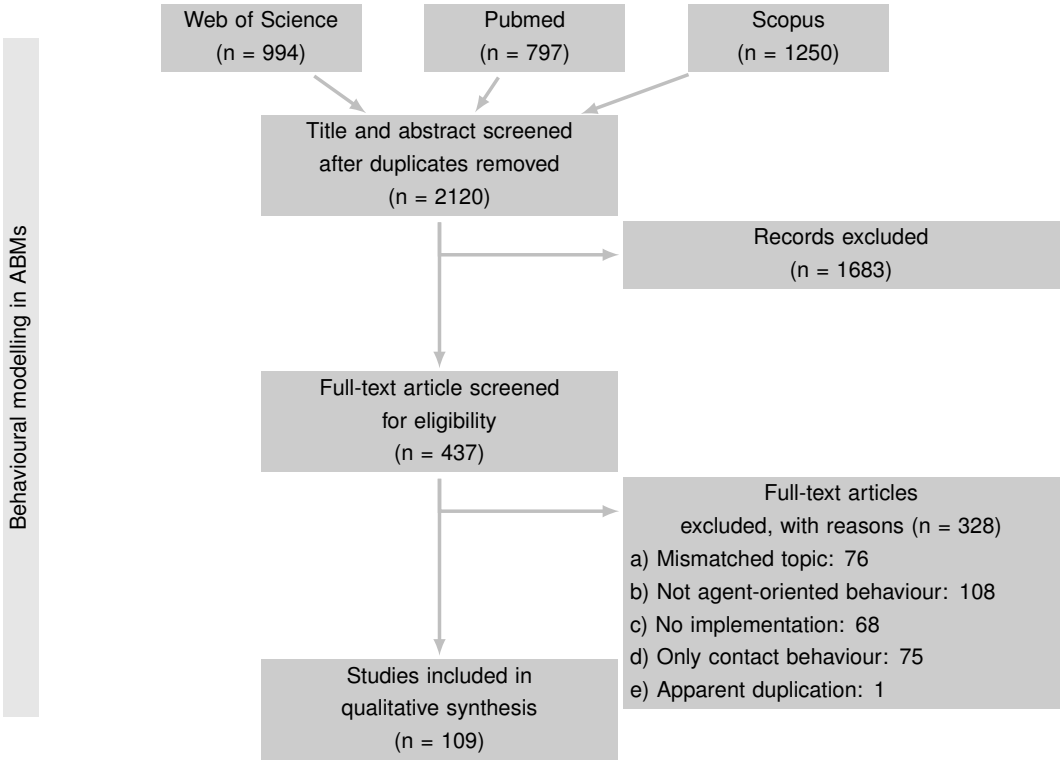


Figure 4.1: PRISMA flow diagram

Figure 4.1 shows the results of my search and selection process. The main reasons for exclusions were: 1) not agent-oriented behaviour, 2) no model implementation, and 3) contact behaviour only. As a result, I identified 109 studies for inclusion: 94 journal articles and 15 conference proceedings. The earliest paper I identified was published in 1999 with the number of studies sharply increasing from around 2010 (see Figure 4.2). I divided the type of health related behaviour modelled into one of four domains: 1) lifestyle choice (24 studies), such as diet and body weight control; 2) addictive behaviour (32 studies), such as smoking and drug use; 3) preventive action (35 studies), such as vaccination and insurance purchase; and 4) Healthcare access (18 studies), such as care-seeking, sick leave taking, and leaving healthcare after recovery. The majority of the studies were classified as either addictive behaviour or preventive action.

Table 4.1 compares modelling purpose, terminology used for ABMs, conceptual behaviour model, and determinants across these domains of health related behaviour. I firstly investigated the descriptions of ABMs in the retrieved studies.

"Agent-based model" was the most commonly used term (79 studies, 72.5%) and "microsimulation" was the second (14 studies, 12.8%). "Individual-based model" appeared to be a special term only applied in infectious disease modelling. I did not identify any studies matching the keywords "computation economics".

I classified the purposes of modelling into four groups from theoretical to practical: theory demonstration (demonstrating a theoretical statement), model demonstration (exemplifying a modelling framework with a realistic scenario), phenomena explanation (explaining empirical phenomena), and policy projection (forecasting and evaluating the outcome of policies). In preventive action, I found 60% of the models were for theory demonstration. In addictive behaviour, more than half of the models were for policy projection.

In the rest of this section I consider the typology of approaches used by these behavioural models for: the conceptual behavioural models; including the determinants of behaviour; their technical implementation; and, finally, the usage of data.

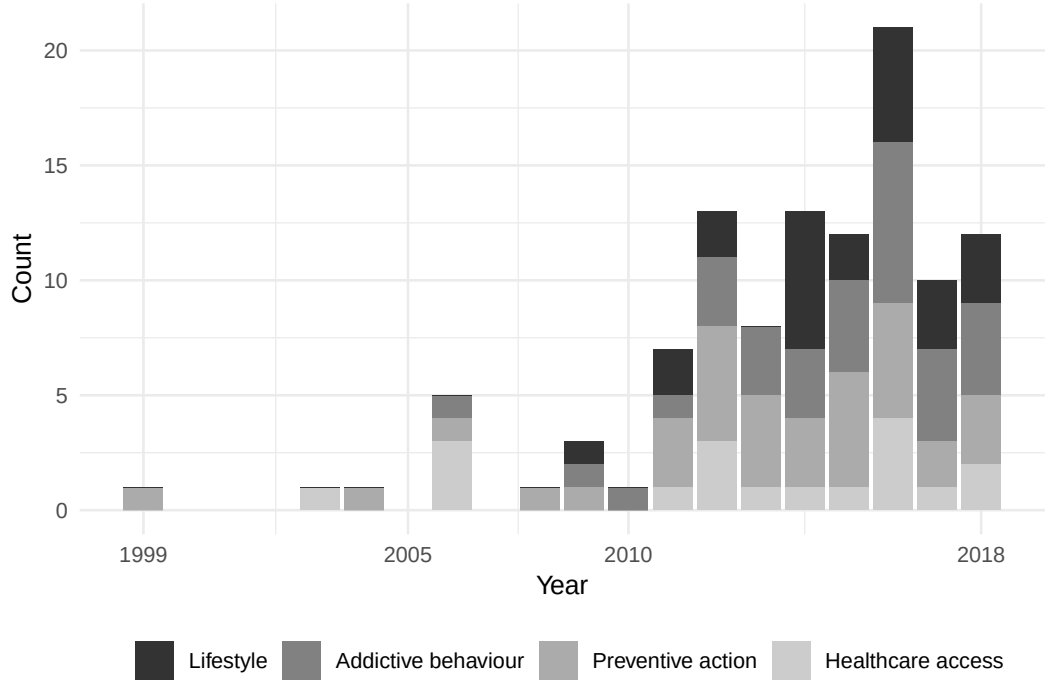


Figure 4.2: Number of studies by year and groups of behaviours. *Lifestyle*: diet, BMI control, and physical activity; *Addictive behaviour*: alcohol consumption, smoking, drug use, and problem gambling; *Preventive action*: self-protection and vaccination; *Hospital access*: care seeking, compliance to health care, taking sick leave, and returning to work during recovery.

Table 4.1: Summary of included studies by characteristics and health domain.

	Lifestyle	Addictive behaviour	Preventive action	Healthcare access	Total
N	24	32	35	18	109
Terminology of ABM					
Agent-based model	21 (87.5%)	19 (59.4%)	26 (74.3%)	13 (72.2%)	79 (72.5%)
Multi-agent model	0 (0%)	1 (3.1%)	1 (2.9%)	0 (0%)	2 (1.8%)
Individual-based model	0 (0%)	1 (3.1%)	5 (14.3%)	1 (5.6%)	7 (6.4%)
Discrete event simulation	0 (0%)	3 (9.4%)	0 (0%)	4 (22.2%)	7 (6.4%)
Microsimulation	3 (12.5%)	8 (25%)	3 (8.6%)	0 (0%)	14 (12.8%)
Behavioural driver					
Utility (Economics)	8 (33.3%)	2 (6.2%)	11 (31.4%)	3 (16.7%)	24 (22%)
Perception	0 (0%)	0 (0%)	10 (28.6%)	1 (5.6%)	11 (10.1%)
Intention/motivation	3 (12.5%)	4 (12.5%)	1 (2.9%)	2 (11.1%)	10 (9.2%)
Imitation/accessibility	3 (12.5%)	4 (12.5%)	2 (5.7%)	0 (0%)	9 (8.3%)
Mixture	0 (0%)	0 (0%)	1 (2.9%)	0 (0%)	1 (0.9%)
Not specified	10 (41.7%)	22 (68.8%)	10 (28.6%)	12 (66.7%)	54 (49.5%)
Purpose of model					
Model demonstration	4 (16.7%)	4 (12.5%)	2 (5.7%)	2 (11.1%)	12 (11%)
Theory demonstration	3 (12.5%)	3 (9.4%)	21 (60%)	4 (22.2%)	31 (28.4%)
Phenomena explanation	9 (37.5%)	6 (18.8%)	6 (17.1%)	5 (27.8%)	26 (23.9%)
Policy projection	8 (33.3%)	19 (59.4%)	6 (17.1%)	7 (38.9%)	40 (36.7%)
Determinants					
Personal characteristics					
Age, sex, location	17 (70.8%)	16 (50%)	4 (11.4%)	4 (22.2%)	41 (37.6%)
Socioeconomic status	19 (79.2%)	9 (28.1%)	4 (11.4%)	6 (33.3%)	38 (34.9%)
Health status	7 (29.2%)	11 (34.4%)	9 (25.7%)	10 (55.6%)	37 (33.9%)
Preference	10 (41.7%)	16 (50%)	12 (34.3%)	4 (22.2%)	42 (38.5%)
Financial cost	9 (37.5%)	9 (28.1%)	4 (11.4%)	2 (11.1%)	24 (22%)
Non-financial cost	8 (33.3%)	2 (6.2%)	7 (20%)	0 (0%)	17 (15.6%)
Neighbour					
Health status	2 (8.3%)	0 (0%)	20 (57.1%)	0 (0%)	22 (20.2%)
Behaviour	9 (37.5%)	14 (43.8%)	9 (25.7%)	1 (5.6%)	33 (30.3%)
Global variable					
Health status	2 (8.3%)	0 (0%)	5 (14.3%)	0 (0%)	7 (6.4%)
Behaviour	3 (12.5%)	2 (6.2%)	2 (5.7%)	1 (5.6%)	8 (7.3%)
Environment	9 (37.5%)	8 (25%)	3 (8.6%)	5 (27.8%)	25 (22.9%)

4.3.2 Conceptual models driving behaviours

Fifty four out of 109 included studies did not employ a conceptual framework for their behavioural modelling, with direct use of data on behaviour predominating. Among those that did, I identified four main types of behavioural drivers: utility from the concept of economics (24 studies), perception (11 studies, e.g. health belief model, HBM [5]), intention (10 studies, e.g. theory of planned behaviour, TPB [6]), imitation or diffusing by accessibility (9 studies), while one study [7] jointly used TPB and utility. The classification was based on both directly indicating by studies or mentioning relevant keywords.

Decision driven by utility

These studies were identified by their use of economic terms such as I extracted the models with economic concepts by identifying the keywords of "economics", "utility", "pay off", "game", and "demand elasticity".

The first variant is the game theoretical approach, which conceptualises the economic system as interactions between agents or interactions between agent and the system. The agents in such models are specified as game players; game players maximise their expected utility given available information from the environment under assumptions relating to the likely strategies of the other players. In vaccination studies, I found a preponderance of simulations based on game theory, in particular the "vaccination game". Vaccination games have been used in discussing voluntary vaccination since this approach was introduced by Bauch and Earn [8]. There are 7 studies included that implemented the vaccination game theoretical framework as an agent-based simulation[9–15]. Within this framework, modellers can also consider incomplete or imperfect information, e.g. Iwamura et al. [11], in order to model effects such as information passing.

The second variant is the econometric approach, where utility maximisation is implicit and determinants alter some behavioural response variable (e.g. consumption) according to some regression model [16]. For instance, uses price elasticities of sugar-sweetened beverages (i.e. the percentage change in demand for a given percentage change in price) to determine agent's calorie intake, which in turn, influences body weight. 29% of economic behavioural models were based on regression analyses of empirical data.

Decision driven by perception or cognition processes

The HBM is directly adapted from cognitive theory in psychology, which studies the cognitive process during decision making. An HBM agent summarises different aspects of perceptions about outcomes (susceptibility, benefits, severity and barriers) that, in conjunction with its motivations, trigger a relevant behavioural action. "Perception" is at the heart of HBM. Identifying perceptions requires cognitive data which is difficult to obtain from observational studies or trials; surveys specific to the relevant issues are a potential data source to inform HBMs. Karimi et al. [17] provided an example of how to obtain the perception components by questionnaire. Alternatively, modellers can set mathematically-formulated forecasting rules for these perceptions, e.g. Brailsford and Schmidt [18]. Without detailed perception data, Mei et al. [19] applied fuzzy cognitive maps to estimate the effects of available variables that might potentially inform perceptions. This ap-

proach amounts to using machine learning for a model that mimics the cognitive process to match specific observed behavioural patterns and make predictions.

Five studies relating to disease prevention [20–24], did not mention HBM directly, yet focused on forming a perceived risk of infection during epidemics. As such, these models did not fully explore the potential use of the HBM in predicting behaviour.

Decision driven by the intensity of intention or motivation

The TPB is the most influential behavioural model of health behaviours in the psychology literature. The TPB regards an action as an outcome of an intention which is a combination of personal attitude, subjective norms and perceived behaviour control. Attitude refers to the knowledge and belief of an behaviour; subjective norm is the influence from significant others, neighbours, and social contexts, which is frequently aliased by peer influence, social influence or social norm; the last part is a measurement of personal ability to fulfil the plan. I also considered the social norm approach adopted by Perkins and Berkowitz [25], which focuses on social influence, to be a special case of the TPB. In addition, the theory of reasoned action [26], the predecessor of the TPB, was considered.

In total, by using the terms "intention", "attitude", "social norms", "subjective norms", I identified 10 studies basing their model on TPB [27–35]. Continuing the work of Brailsford and Schmidt [18], Brailsford et al. [28] analysed the features of the HBM and the TPB, and highlighted the advantages of the TPB in dealing with empirical data. In practice, they asserted that although HBM is feasible and comprehensible, some important components, such as clues to action and health motivation, are abstractly defined, resulting in difficulty in collecting data. By contrast, TPB components are more aligned with data. Purshouse et al. [33] demonstrated how to combine the TPB and data using a probit regression model to map variables in data onto the TPB components. They used survey data to initialise the baseline population and applied an evolutionary algorithm to estimate the effect of each variable, instead of using empirical covariates. Zhang et al. [34] combined the TPB framework with economic features in order to determine TPB components from aggregated data. For example, price elasticity of foods were a surrogate of behaviour control in their models.

Four studies [29–32] highlighted peer influence on subjective norms, but neglected the other aspects of planned behaviour. Starting from an *a priori* specification of conceptual components, Fitzpatrick et al. [30] modelled college drinking mathematically and emphasised what kind of data can support policy making in the field. Fitzpatrick et al. [31] extended this work using elicited data and policy

options in their simulations. Collinson et al. [29] modelled the mass media as primary determinant of social norms influencing protective behaviours during an epidemic.

Decision driven by imitation and diffusion

The last group of the collected conceptual models has behaviour driven by imitation and diffusion by accessibility. This approach replaces the direct examination of cognitive process or the belief of behaviour with an indirect link via the process of behaviour diffusion. I excluded here studies which had been grouped under other conceptual models with strong peer influence, such game theoretical and social norm approaches. Hammond and Ornstein [36] employed the "Follow the Average" (FTA) model [37] to modelling Body Mass Index (BMI) control behaviour. Their agents adjusted their BMI in order to match the average BMI of their neighbours. The other studies described the health behaviours as communicable diseases. The behaviours can be transmitted through social networks. Mao [38] modelled spreading word of mouth as an infectious process; agents discuss with each other and adapt their self-protection against an infectious disease. Gorman et al. [39] regarded alcohol consumption as an infectious disease. The model labelled every agent as susceptible drinker, current drinker (the infected), or former drinker (the recovered). In addition, I identified models used Huff's gravity model [40] to compose the intensity of a behaviour or the accessibility of having a behaviour. For example, Li et al. [41] weighted food shops by preference and distance with the Huff model, determining the food choice.

4.3.3 Determinants of behaviour

Investigating heterogeneity is an important use of ABMs. Each agent can behave differently according to their individual features. The lower part of Table 4.1 shows the determinants at a personal-, neighbour-, and global-level that were included in studies on different topics.

Personal features and costs

There was considerable variation in the types of personal traits incorporated into the ABMs across the different types of behaviour. The use of age, sex, and location and socioeconomic status was common in studies of lifestyle choices, less so in addictive behaviour and healthcare access studies, and rarely used in illness prevention models. Interestingly, socioeconomic status were considered in 79.2% of lifestyle models but only 28.1% of addictive behaviour models. This suggests

that most of addictive behaviour models assumed a less important role for socioeconomic status in decisions. Individual preferences or attitudes, summarised the personal incentives which were not measured by the other variables, were used in 50% of addictive behaviour studies and 34.3% of preventive action studies.

Neighbour influence

Modelling the interactions among agents is a central feature of ABMs. The role of neighbour traits was included most frequently in the disease prevention models, and in particular, the infectious disease studies. The main route of influence in these studies was through individuals using neighbour's health status to compose their own risk perceptions of a disease. However, neighbour behaviour was also widely used in lifestyle and addictive behaviour models (40% of each), where neighbours' behaviours influenced an agent's preference for diet, alcoholic drinking, smoking, or drug abuse.

Incorporating neighbour traits, embedding a social network or a grid map modelling is required so the neighbours can be distinguished from the others. In terms of conceptual model, neighbour effects can be modelled through features of opponents of a game in economic model; social norms in the TPB; perceived threat or cue to action in the HBM.

Global variables

Global variables are overall indices or background societal features in the models. Few studies included a role for aggregate health status or behaviour at the system level (6.4% of all studies). I identified some studies where agents adapted to policy change. In Li et al. [32], intervening on health education changed food consumption decisions. Likewise, Zhang et al. [34] utilised economic drivers, such as tax, to influence agents' food choices. Dray et al. [42] modelled community use of drug-detecting dogs; drug-using agents would feel threatened by their presence and reduce drug trading.

4.3.4 Implementation details

I analysed the anatomy of the implementations in the retrieved studies around: how determinants were summarised and what kind of behaviour was modelled. The implementation details of each study are detailed in Appendix A.

Behaviours defined as binary or categorical variables

Forty eight of included studies used binary representations of behaviour, 24 of them drawing from a Bernoulli distribution. Of these 24, the trigger probability was often specified from regression analysis of survey data and the behaviours studied were often around smoking and alcohol drinking behaviour (e.g. Adams and Schaefer [43], Rahhali et al. [44]). Where empirical data were lacking, theory was often used to motivate a formula or algorithm for the trigger probability. Some models of behaviour during epidemics used a modelled perception of risk to specify the trigger probability (e.g. Zhang Fa et al. [24]).

Triggering behaviour based on exceeding some threshold was also common here, often using a cumulative exposure. For example, Carley et al. [45] triggered care seeking behaviour based on the level of illness reached. Barrett et al. [20] modelled preventive behaviours that were triggered when a disease prevalence exceeded a threshold.

Game-theoretic approaches have been used to model protective behaviour against infectious diseases (9 studies). For example, in vaccination games (e.g. Fu et al. [10] and Iwamura et al. [11]), susceptible agents can decide to be vaccinated or not depended on their perceived risk of infection and their neighbours' vaccination status. This literature tended to focus on theory and use less empirical data.

More than two categories of behaviour have also been considered; using multinomial logistic regression, as in Riva and Smith [46], and using multiple thresholds their own comparators (see Barrett et al. [21]).

Behaviours defined as continuous variables

Using continuous variables as actions were found in the subjects of lifestyle and addictive behaviours. In body weight controlling, continuous actions were usually formulated with mathematical functions. In this approach, Orr et al. [47] built a body weight control process with hierarchical functions of related behaviours. They showed the potential of connecting many behaviours in a simulation model. Dealing with the amount of alcohol drinking and smoking, which were in rich data settings, tended to employ regression models. While the regression models required the assumption of linearity, it is difficult to incorporate theoretical components of interest but lacked data. Representing behaviour by continuous variables was common for studies of lifestyle and addictive behaviours. In obesity studies, actions were often specified via mathematical functions. For example, Orr et al. [47] defined a body mass control process using hierarchical functions of related behaviours, to illustrate this approach to modelling multiple linked be-

haviours. Studies of alcohol and smoking often had more relevant data available, and tended to use regression approaches.

Triggering behaviours through optimisation

Triggering an event using optimisation, or more specifically utility maximisation, is a standard approach in neoclassical economics. I retrieved 17 studies using this approach in their models , mainly focussing on lifestyle and patient choice. Behaviours were represented by categorical variables, continuous variables, and vectors of decisions. Some approaches long these lines use both theory and empirical data. For example, Trogdon and Allaire [48] formulated a utility including food consumption as a positive term and high body mass as a negative term, in order to model food choices under a constrained budget. Optimal strategies have also been considered for game-theoretical behavioural models, where individuals' utility is dependent on the actions of others.

Triggering actions through imitation or diffusion

Finally, some models directly mapped to the conceptual model of imitation and diffusion. Instead of generating a variable to represent a decision or a behaviour, five models used transmission models to propagate decisions between agents through neighbour imitation. Five models included behaviours changed by imitation or transmission. Hammond and Ornstein [36] modelled a community which people always try to adjust their Body Mass Index to the average of their neighbours. Perez et al. [49] described a drug user network which drug users change drug use levels according to their peers. These agents in these models were identical, however heterogeneity in levels of imitation can be modelled. For example, Hammond and Ornstein [36] provided a realistic application with more personal attributes, such as age and sex, showing the feasibility of greater realism.

4.3.5 Roles of data

Table 4.2: The role of data in modelling by purpose of model.

	Model demonstration	Theory demonstration	Phenomena explanation	Policy projection	Total
N	12	31	26	40	109
Data usage					
Drive model by data	4 (33.3%)	3 (9.7%)	14 (53.8%)	25 (62.5%)	46 (42.2%)
Fit model to data	0 (0%)	2 (6.5%)	4 (15.4%)	3 (7.5%)	9 (8.3%)
Both	2 (16.7%)	0 (0%)	8 (30.8%)	10 (25%)	20 (18.3%)
No data involved	6 (50%)	26 (83.9%)	0 (0%)	2 (5%)	34 (31.2%)

Data plays an important role in linking models to reality. I identified two roles

of data in the studies I included: 1)driving model by the data of the values of determinants (e.g. age) or effects of determinants (evidence from literature to inform the effect of age on smoking behaviour), 2) and calibration (fitting the behavioural model to some target data as part of the study, including fitting regression models). Table 4.2 compares role of data and model purpose.

In theory demonstration, real data were rarely involved. Only five studies tried to inform their models with empirical elements. Phenomena explanation and policy projection, included diverse use of data. In the included studies, a third used regression models to summarise determinants. Often determinant data for the modelled population was taken from one source, and the regression applied to these features derived from data from another population (and study). Empirical covariates for these studies were usually taken from survey data. Where data was not used, most models assigned determinants either assumed values or used values related to some simplifying criterion (e.g. yielding an equilibrium).

4.4 Discussion

In this article, I have systematically reviewed the approaches used in agent-based models to model individual health-related behaviours. I found that use of ABMs to model health-related behaviour is expanding rapidly, and work is being undertaken across a range of disciplines and with a focus on a number of aspects of behaviour. The purpose of the modelling varied, with exploration of theoretical models predominating for disease prevention (e.g. vaccination behaviours) and applied policy evaluation predominating in studies on substance abuse. I focussed on describing the approaches used for the conceptual frameworks underpinning behavioural models, the determinants of behaviour, and approaches to implementing behaviours in models; as well as how these varied by topic of interest.

4.4.1 Conceptual model, determinant, and model implementation

Introducing a conceptual behavioural model before implementation provides the modeller with a convenient way to clarify and summarise underlying assumptions. The major quality of using conceptual models in ABMs is be an implementation blueprint that allows a modeller to pay more attention to the question itself and a terminology guideline to fit audiences. For instance, when a model mentions economic utility, it implies agents are utility maximisers, considering balance between benefit and cost; modellers should define a utility function or a pay-off matrix, clarify approaches to projection and time preference, and then

optimise utility with respect to possible choices. These information are commonly acquired by people in economics. In terms of implementation, utilising a conceptual model does not strictly frame an ABM. Some components of the identified conceptual models can map to each other by simply carrying different meanings. For instance, cost in economic approach can be interpreted as a perceived barrier in the HBM, and a measure of perceived behaviour control in the TPB. Due to the quantitative nature of economics, economic models are relatively well formulated. The steps of implementation across my collection are homogeneous but the details of how to implement were diverse. In addition, there are extensions that can be utilised for specific details. For example, dynamic programming (policy iteration, Howard [50]) assists the modeller to introduce future outcomes in the long run; defining constant variables allow heterogeneity among population. Applying the concept requires judging the benefits and cost of variables or identifying variables related to heterogeneity. Once the steps and utility function defined, economic model can be executed in ABMs. Nevertheless, data from empirical model does not include benefit/cost judgements.

The TPB and the HBM are fundamentally psychological approaches for health behaviour. They have lesser linked with model implementation than economic approach. This has the advantage that various types of implementation are possible, but the disadvantage that modellers cannot easily find guidelines for implementation, and the description takes time to be understood by readers. For this reason, the HBM and the TPB are more typical in health behaviour discussions, but appear less frequently than economic models in my review. The motivation for applying HBM and TPB is not in laying out an implementation plan, but in providing reasonable explanation and realistic policy implications. The main difference between HBM and TPB is that HBM focuses on future outcomes while TPB focuses on current status and experience. A secondary difference is that HBM drives a behaviour mainly by self-oriented variables, while TPB allows influence directly from social networks and environments. For example, in terms of peer influence in my review, HBM allowed a neighbour's health status to influence risk perception and TPB utilised neighbour's behaviours directly to change agent's behaviour. Imitation and diffusion models focus on neighbour effect mainly. It can be seen as a reduced form of the other conceptual models considering neighbour influence. As the relevant models in my search minimised other determinants and less data were involved, all of them were theoretical works.

Furthermore, as for the half of identified models which did not specify conceptual models, the appropriateness of using an underlying model or not should be clarified first. I note that the specification may be less appropriate under the

following conditions. First, when individual level data are available and manageable, a conceptual model might be less important. Second, when the behaviours involved follow hierarchical decision making steps or case-dependent scenarios, conceptual models cannot capture the behaviour straightforwardly. For instance, the body weight control models of Orr et al. [47] and Wang et al. [51] applied diet behaviour as an upstream process; they needed to balance complexity of each component for clarity. Last, when the size of upstream parameters is small, a conceptual model may introduce unnecessary complexity and over-parametrisation.

4.4.2 Roles of data

While studies focussed on smoking and drinking made heavy use of survey data to empirically inform parameters, studies on illegal drug use were more reliant on theoretical approaches (e.g. theoretical drug markets, see Hoffer et al. [52]) to model agent behaviours. Data are less available for these behaviours and subject to recall and social desirability biases, which may account for this.

Using conceptual models seemed favoured when data were lacking. The models based on empirical studies (e.g. Xenakis et al. [53]) were closely tied to the behavioural outcomes and determinants data collected. Therefore there was less flexibility to include further assumptions and sub-models around behaviour. Only a few papers used conceptual frameworks with components directly linked to data [17, 28, 33].

4.4.3 A reporting framework for health behavioural modelling?

Our analysis of the methodological and implementation approaches showed high heterogeneity. Flexibility is an important feature of ABMs, however, identifying common features and developing a reporting protocol for models of this type would be useful. Grimm et al. [54] provided a widely-used reporting protocol to aid communication among modellers, but this was developed by modellers in ecology, focusing on overall model building process. Smajgl and Barreteau [55] developed an ABM reporting framework based on work modelling human behaviour for environmental studies. In the field of health behaviour modelling, a specialised protocol that was better matched to the models used and accessible to those without any computer science background would be more easily applied and also facilitate working in interdisciplinary teams. Instead of whole model, I focused on a specific modelled behaviour so I zoomed in by introducing conceptual models. As part of my analysis of model implementation, I categorized various aspects of their approach (see also Appendix A); this could be used as a

basis for reporting implementation. The diversity of modelling approaches and domains remains a challenge for standardized reporting.

4.4.4 Strength & Limitations

My work does have some limitations. The review did not include double data extraction or sifting, and did not pre-register a protocol. The very diversity in topic, emphasis and extent of reporting made the job of identifying approaches used by these studies more difficult. I did a search on broad topic around health behaviour so the second set of keywords were not strongly defined for specific behaviours. However, most studies found were within public health. Indeed, I excluded many studies in drug market that focussed on whole market dynamics without details of drug use, since these studies did not report on implementation of behavioural models. Excluding decision support required some judgement: decision support is usually the ultimate purpose of modelling in public health, and without some assessment at title/abstract stage, a huge number of studies would be included for full-text review.

However, this work is the most comprehensive and systematic effort to identify and classify approaches to modelling behaviour in ABMs applied to questions around health. I had a transparent and reproducible search strategy and have gone to some lengths to identify qualitative features and their quantitative patterns within this literature. Since use of ABM is still emerging in health-related fields, the relatively new idea is often criticised as a "black box" approach. My work attempted to "unbox" the models by detailing their components in a systematic way. I anticipate that my analysis framework may aid translation between agent-based modelling and health-related research. However, I stopped short of trying to identify a unified framework to cover all aspects of the studies I found because of their multifaceted diversity. I aimed to balance breadth and depth of synthesis.

4.5 Summary

In this chapter, I reviewed how to model health-related behaviours in agent-based modelling. Referring to the purpose of this thesis, the review highlighted 1) conceptual behavioural model can guide the parametrisation as well as data collection and 2) using data is a key to bring a theoretical work to policy making. Meanwhile, the selected studies also informed this thesis how to include the influences from upstream determinants and how they fitted to behavioural model.

Bibliography

[1] McDaid D, Needle J. Economic evaluation and public health: mapping the literature. 2006;.

[2] Maglio PP, Mabry PL. Agent-based models and systems science approaches to public health. Am J Prev Med. 2011 Mar;40(3):392–394.

[3] Epstein JM. Agent-based computational models and generative social science. Complexity. 1999;4(5):41–60.

[4] Robertson DA. Agent-based models and behavioral operational research. In: Behavioral Operational Research. Springer; 2016. p. 137–159.

[5] Becker MH. The health belief model and personal health behavior. vol. 2. Slack; 1974.

[6] Ajzen I. The theory of planned behavior. Organ Behav Hum Decis Process. 1991 Dec;50(2):179–211.

[7] Xue J, Terano T, Deguchi H, Ichikawa M. Simulation analysis of immunization policy diffusion in social network with ABM approach. In: 2016 International Conference on Behavioral, Economic and Socio-cultural Computing (BESC). IEEE; 2016. p. 1–6.

[8] Bauch CT, Earn DJD. Vaccination and the theory of games. Proc Natl Acad Sci U S A. 2004 Sep;101(36):13391–13394.

[9] Brown GDA, Fincher CL, Walasek L. Personality, Parasites, Political Attitudes, and Cooperation: A Model of How Infection Prevalence Influences Openness and Social Group Formation. Top Cogn Sci. 2016 Jan;8(1):98–117.

[10] Fu F, Rosenbloom DI, Wang L, Nowak MA. Imitation dynamics of vaccination behaviour on social networks. Proc Biol Sci. 2011 Jan;278(1702):42–49.

[11] Iwamura Y, Tanimoto J, Fukuda E. Effect of intermediate defense measures in voluntary vaccination games. J Stat Mech. 2016 Sep;2016(9):093501.

[12] Keane CR. Resilience, tipping, and hydra effects in public health: emergent collective behavior in two agent-based models. BMC Public Health. 2016 Mar;16:265.

[13] Keane C. Chaos in collective health: Fractal dynamics of social learning. J Theor Biol. 2016 Nov;409:47–59.

[14] Teoh Shian Li, Labadin J, Phang Piau, Rahman SA, Ling Yeong Tyng. The effect of vaccination decision in disease modelling through simulation. In: Tiong, WK and Ujir, H and Mit, E and Iskandar, DNA, editor. 2015 9th International Conference on IT in Asia (CITA); 2015. p. 1–6.

[15] Liu XT, Wu ZX, Zhang L. Impact of committed individuals on vaccination behavior. Phys Rev E Stat Nonlin Soft Matter Phys. 2012 Nov;86(5 Pt 1):051132.

[16] Basu S, Vellakkal S, Agrawal S, Stuckler D, Popkin B, Ebrahim S. Averting obesity and type 2 diabetes in India through sugar-sweetened beverage taxation: an economic-epidemiologic modeling study. PLoS Med. 2014 Jan;11(1):e1001582.

[17] Karimi E, Schmitt K, Akgunduz A. Effect of individual protective behaviors on influenza transmission: an agent-based model. Health Care Manag Sci. 2015 Sep;18(3):318–333.

[18] Brailsford S, Schmidt B. Towards incorporating human behaviour in models of health care systems: An approach using discrete event simulation. Eur J Oper Res. 2003 Oct;150(1):19–31.

[19] Mei S, Zhu Y, Qiu X, Zhou X, Zu Z, Boukhanovsky AV, et al. Individual Decision Making Can Drive Epidemics: A Fuzzy Cognitive Map Study. IEEE Trans Fuzzy Syst. 2014 Apr;22(2):264–273.

[20] Barrett C, Bisset K, Leidig J, Marathe A, Marathe M. Economic and social impact of influenza mitigation strategies by demographic class. Epidemics. 2011 Mar;3(1):19–31.

[21] Barrett CL, Bisset K, Leidig J, Marathe A, Marathe M. Estimating the Impact of public and private strategies for controlling and epidemic: a multi-agent approach; 2009. p. 34–39.

[22] Delaney L, Kleczkowski A, Maharaj S, Rasmussen S, Williams L. Reflections on a Virtual Experiment Addressing Human Behavior During Epidemics. In: Proceedings of the 2013 Summer Computer Simulation Conference. vol. 45 of

SCSC '13. Vista, CA: Society for Modeling & Simulation International; 2013. p. 33:1–33:8.

[23] Durham DP, Casman EA. Incorporating individual health-protective decisions into disease transmission models: a mathematical framework. J R Soc Interface. 2012 Mar;9(68):562–570.

[24] Zhang Fa, Zhao Qiao-xia, Li Lu. Intervention for contagious disease: Agent-based modeling and simulation. In: 2011 2nd IEEE International Conference on Emergency Management and Management Sciences. IEEE; 2011. p. 155–158.

[25] Perkins HW, Berkowitz AD. Perceiving the community norms of alcohol use among students: some research implications for campus alcohol education programming. Int J Addict. 1986 Sep;21(9-10):961–976.

[26] Sheppard BH, Hartwick J, Warshaw PR. The Theory of Reasoned Action: A Meta-Analysis of Past Research with Recommendations for Modifications and Future Research. J Consum Res. 1988 Dec;15(3):325–343.

[27] Beheshti R, Jalalpour M, Glass TA. Comparing methods of targeting obesity interventions in populations: An agent-based simulation. SSM Popul Health. 2017 Dec;3:211–218.

[28] Brailsford SC, Harper PR, Sykes J. Incorporating human behaviour in simulation models of screening for breast cancer. Eur J Oper Res. 2012 Jun;219(3):491–507.

[29] Collinson S, Khan K, Heffernan JM. The Effects of Media Reports on Disease Spread and Important Public Health Measurements. PLoS One. 2015 Nov;10(11):e0141423.

[30] Fitzpatrick B, Martinez J, Polidan E, Angelis E. The Big Impact of Small Groups on College Drinking. J Artif Soc Soc Simul. 2015 Jun;18(3).

[31] Fitzpatrick BG, Martinez J, Polidan E, Angelis E. On the Effectiveness of Social Norms Intervention in College Drinking: The Roles of Identity Verification and Peer Influence. Alcohol Clin Exp Res. 2016 Jan;40(1):141–151.

[32] Li Y, Zhang D, Pagán JA. Social Norms and the Consumption of Fruits and Vegetables across New York City Neighborhoods. J Urban Health. 2016 Apr;93(2):244–255.

[33] Purshouse RC, Ally AK, Brennan A, Moyo D, Norman P. Evolutionary parameter estimation for a theory of planned behaviour microsimulation of alcohol consumption dynamics in an English birth cohort 2003 to 2010. In: Igel, C, editor. Proceedings of the 2014 Annual Conference on Genetic and Evolutionary Computation. ACM; 2014. p. 1159–1166.

[34] Zhang D, Giabbanelli PJ, Arah OA, Zimmerman FJ. Impact of different policies on unhealthy dietary behaviors in an urban adult population: an agent-based simulation model. Am J Public Health. 2014 Jul;104(7):1217–1222.

[35] Zhang Q, Northridge ME, Jin Z, Metcalf SS. Modeling Accessibility of Screening and Treatment Facilities for Older Adults using Transportation Networks. Appl Geogr. 2018 Apr;93:64–75.

[36] Hammond RA, Ornstein JT. A model of social influence on body mass index. In: Dube, L and Webb, P and Arora, NK and Pingalid, P, editor. PATHS OF CONVERGENCE FOR AGRICULTURE, HEALTH, AND WEALTH. vol. 1331 of Annals of the New York Academy of Sciences; 2014. p. 34–42.

[37] Hammond RA, Epstein JM. Exploring Price-Independent Mechanisms in the Obesity Epidemic; 2007.

[38] Mao L. Predicting self-initiated preventive behavior against epidemics with an agent-based relative agreement model. Journal of Artificial Societies and Social Simulation. 2015;18(4):6.

[39] Gorman DM, Mezic J, Mezic I, Gruenewald PJ. Agent-based modeling of drinking behavior: a preliminary model and potential applications to theory and practice. Am J Public Health. 2006 Nov;96(11):2055–2060.

[40] Huff DL. A probabilistic analysis of shopping center trade areas. Land Econ. 1963;39(1):81–90.

[41] Li Y, Du T, Peng J. Understanding Out-of-Home Food Environment, Family Restaurant Choices, and Childhood Obesity with an Agent-Based Huff Model. Sustain Sci Pract Policy. 2018 May;10(5):1575.

[42] Dray A, Perez P, Moore D, Dietze P, Bammer G, Jenkinson R, et al. Are drug detection dogs and mass-media campaigns likely to be effective policy responses to psychostimulant use and related harm? Results from an agent-based simulation model. Int J Drug Policy. 2012 Mar;23(2):148–153.

[43] Adams J, Schaefer DR. How Initial Prevalence Moderates Network-based Smoking Change: Estimating Contextual Effects with Stochastic Actor-based Models. J Health Soc Behav. 2016 Mar;57(1):22–38.

[44] Rahhali N, Millier A, Briquet B, Laramée P, Aballéa S, Toumi M, et al. Modelling the consequences of a reduction in alcohol consumption among patients with alcohol dependence based on real-life observational data. BMC Public Health. 2015 Dec;15:1271.

[45] Carley KM, Fridsma DB, Casman E, Yahja A, Altman N, Li-Chiou Chen, et al. BioWar: scalable agent-based model of bioattacks. IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans. 2006 Mar;36(2):252–265.

[46] Riva M, Smith DM. Generating small-area prevalence of psychological distress and alcohol consumption: validation of a spatial microsimulation method. Soc Psychiatry Psychiatr Epidemiol. 2012 May;47(5):745–755.

[47] Orr MG, Galea S, Riddle M, Kaplan GA. Reducing racial disparities in obesity: simulating the effects of improved education and social network influence on diet behavior. Ann Epidemiol. 2014 Aug;24(8):563–569.

[48] Trogdon JG, Allaire BT. The effect of friend selection on social influences in obesity. Econ Hum Biol. 2014 Dec;15:153–164.

[49] Perez P, Dray A, Moore D, Dietze P, Bammer G, Jenkinson R, et al. SimAmph: an agent-based simulation model for exploring the use of psychostimulants and related harm amongst young Australians. Int J Drug Policy. 2012 Jan;23(1):62–71.

[50] Howard RA. Dynamic programming and markov processes. 1960;.

[51] Wang Y, Xue H, Chen HJ, Igusa T. Examining social norm impacts on obesity and eating behaviors among US school children based on agent-based model. BMC Public Health. 2014 Sep;14:923.

[52] Hoffer LD, Bobashev G, Morris RJ. Researching a local heroin market as a complex adaptive system. Am J Community Psychol. 2009 Dec;44(3-4):273–286.

[53] Xenakis JG, Kinter ET, Ishak KJ, Ward AJ, Marton JP, Willke RJ, et al. A discrete-event simulation of smoking-cessation strategies based on varenicline pivotal trial data. Pharmacoeconomics. 2011 Jun;29(6):497–510.

[54] Grimm V, Berger U, Bastiansen F, Eliassen S, Ginot V, Giske J, et al. A standard protocol for describing individual-based and agent-based models. Ecol Modell. 2006 Sep;198(1):115–126.

[55] Smajgl A, Barreteau O. Framing options for characterising and parameterising human agents in empirical ABM. Environmental Modelling & Software. 2017 Jul;93:29–41.

Chapter 5

Tuberculosis epidemiology and population dynamics, Taiwan

This thesis investigates the relationship between TB-related care-seeking and socio-economic status (SES). I used the setting of Taiwan throughout this thesis. I considered that Taiwan (1) has the National Health Insurance programme, which covered the healthcare system universally and collects claim data routinely, and (2) has an intermediate TB burden with a steady declining trend but without apparent TB-HIV coinfection. Before assessing the SES and TB in depth, this chapter introduced the population and TB epidemiology in Taiwan with a statistical analysis of notification data coupling with a synthetic population. I also revealed a difficulty of TB control in Taiwan due to the ageing population.

The research in this Chapter was published in Ku and Dodd [1].

5.1 Introduction

In 2018, tuberculosis (TB) was still the top infectious killer in the world [1]. The End TB strategy aims at a 90% reduction in TB incidence rate by 2035 compared with 2015, but the current global rate of decline of around 2% per year is not on track to achieve this [2]. Latent TB infection risk accumulates over lifetimes while TB transmission is ongoing. The prevalence of latent TB infection is highest

[1] Ku CC, Dodd PJ. Forecasting the impact of population ageing on tuberculosis incidence. PLoS One [Internet]. 2019; Available from: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0222937

in older age groups [3], who not only have had the longest exposure, but were often exposed to higher TB transmission rates in the past. Ageing, with associated higher rates of progression [4], thus acts as a demographic driver towards higher per capita TB incidence [5]. In the Western Pacific region, many countries have their highest [6, 7] per capita TB incidence rates among older age groups [1]. Among Western Pacific region countries, China, Hong Kong (China), Japan, South Korea, Singapore, and Taiwan are facing both high TB burden and population ageing [6, 7].

The age profile of future TB incidence is critical for forecasting public health needs and rational policy design [8]. First, older populations will have higher TB (and background) mortality rates [9, 10], which implies added difficulty in meeting treatment success targets. Secondly, older adults have more comorbidities and more complex health care needs, which may lead to a longer care-seeking process and higher healthcare expenditure per case. For instance, patients with chronic lung diseases may have signs or symptoms overlapping with TB, making correctly diagnosing their TB slower and more costly [11]. Thirdly, the proportion of TB cases in older age groups should inform policy making, for example suggesting integrating TB care entry points into long-term care programmes, or through clinician training highlighting older people as a TB risk group with their own diagnostic and management challenges [11].

Quantitatively forecasting the TB incidence age profile needs combined models forecasting demographic change and statistical forecasts of age-specific TB incidence. However, a time series analysis producing age-specific forecasts of the TB incidence has not been published to my knowledge. Use of autoregressive integrated moving average (ARIMA) models, often including the seasonality of TB incidence is more common,[12] and comoving time series analysis has been applied [13] without age-specific information. Age-specific TB incidence modelling, including the use of age-period-cohort (APC) models, has been undertaken but without producing epidemic forecasts (e.g. Iqbal et al. [14] and Wu et al. [15]). Mechanistic mathematical modelling, with age structure, also has the potential to generate forecasts [5, 16–18]. Among them, Arregui et al. [18] provided the age-specific incidence forecast and mentioned the effect of population ageing. Indeed, Arregui et al. [18] developed forecasts for the effects of demographic change on TB epidemics, focussing on four relatively young countries; our interests are in developing statistically rigorous time-series approaches and in focusing on an example of a much older population.

In many settings, the demographic transition and population ageing are outpacing declines in TB incidence, so methods to understand and forecast the im-

pact of changing demography on TB epidemics are needed. I, therefore, developed a statistical method capturing age-specific incidence trends and forecasting future epidemics while accounting for demographic change so more methods to understand and forecast the impact of changing demography on TB epidemics are needed.

5.2 Methods

Setting and data sources

TB incidence in Taiwan has steadily declined from 64 confirmed TB cases per 100,000 in 2007 to 41 per 100,000 in 2017. Since 2005, the proportion of TB cases in Taiwan over 65 years of age has been over 50% and increasing. Between 2007 and 2017, the average age in Taiwan increased from 36 to 40, and the proportions of adults above 65 rose from 10% to 14% [7].

Notification data of culture-confirmed TB cases, excluding foreigners, were obtained from the Taiwan Center for Disease Control (TCDC) surveillance system. Counts were reported by age group, sex, month, and county. Ages were reported as (0-4, 5-9, ..., 65-69, 70+) years. The demographic data were obtained from the Department of Statistics, the Ministry of the Interior, Taiwan. These data included the mid-year population estimators, deaths, migration in single-year ages, and fertility in five-year age groups (15-19, ..., 45-49). I used data in 2005-2018 as a training set. The demographic data from 2005 to 2017 were collected for the population demographic modelling (a shorter period because of the release schedule). All the training data in this article were published by the Taiwan officials and free access on the Open Government Data Platform [https://data.gov.tw]. Also, a processed dataset specified for this study can be retrieved from https://github.com/TimeWz667/AgeingTB. The usage is licensed by the Open Government Data License: [https://data.gov.tw/license].

Importantly, I assumed no case detection gaps existed during the time frame covered by this article. I, therefore, regarded "TB notification" and "TB incidence" as synonymous with the number culture-confirmed tuberculosis cases notified during a specific period.

Age-specific incidence modelling and forecasting

I considered annual incidence rates by age and sex. The incidence rates by age groups and sex were calculated as the yearly notification counts divided by corresponding mid-year population estimates. Females and males were analysed

separately with the same parameterisation. I modelled the incidence rates using Lee-Carter Models (LCMs) [19] formulated with age and time-varying terms. The LCMs were initially designed for mortality rate modelling, where they now predominate. Estimation, forecasting, bootstrapping methods for LCMs are well-developed.

I performed a likelihood-based LCM estimation, and also the comparable Poisson regression:

$$log(E(y_{age,t})) = \alpha_{age} + \beta_{age}\kappa_t + log(n_{age,t}) \tag{5.1}$$

where $E(.)$ is expectation function, $t \in \{2005, \ldots, 2018\}$ is the calendar year, $y_{age,t}$ is incident cases at (age, t), $n_{age,t}$ is population size at (age, t), α_{age} is age effect term, κ_t is period effect term, and β_{age} is coefficients adjusting period effects for different age groups, and $age \in \{0-4, 5-9, \ldots 70+\}$ represents the age categories. To maintain identifiability, I imposed the constraints $\Sigma_t \kappa_t = 0$ and $\Sigma_{age}\beta_{age} = 1$. Given the conditions, I followed the fitting procedure provided by Brouhns et al. [20] for likelihood maximisation.

Two nested Poisson models, one using an age-profile and a discrete period effect, $\alpha_{age} + \kappa_t$, and another using an age-profile and a linear effect, $\alpha_{age} + t\kappa$, were used as comparators. Akaike information criterion (AIC), Bayesian information criterion (BIC), and log-likelihood were considered as the goodness of fit metrics. The definitions of these metrics were identical to the ordinary Poisson regression model [21]. See Appendix B for the details of the LCM specification in our approach.

For forecasting, inspired by the Lee-Carter demographic forecasting, I used Autoregressive Integrated Moving Average (ARIMA) models with drift [22], constructed from the LCM period effects. In forecasting, the death and birth processes applied semiparametric bootstrap sampling [23].

Population modelling and forecasting

I constructed a synthetic population with birth, death and migration processes. The demographic methods adapted from those used in the Taiwan National Development Council's population projection report [7]. The demography was modelled by single age (0-100 years old) and sex. Mortality forecasting used the Lee-Carter model [19] below 84 years of age, and we used the Coale-Kisker method [24] for those aged over 85 years in death rate modelling as it was found to have better reliability for small sample sizes in inferring of death rates. The birth forecasting used the fertility rates of women in childbearing ages, from 15 to 49, with

a modified LCM [25]. For consistency with incidence forecasting, semiparametric bootstrap sampling was used for deaths and births [23]. The Migration process was modelled by linear regression with age effects and a linear trend; the forecasting applied residual bootstrap sampling with the age-specific parameters seen in 2017. The forecasts were used for the next step by aggregating to the age groups as that of the incidence data. See Appendix C for the detailed methodology of the synthetic population.

Forecasting overall TB incidence

The TB incidence model and the demographic model were built independently. Forecasts of age-specific TB incidence were weighted by forecasted population demography to obtain forecasts of per capita TB incidence for the whole population. TB incidence was calculated as per 100,000 rates by given strata. TB incidence rate reductions were calculated with respect to the incidence in 2015 and presented as percentages. For simplicity, some results were presented with age groups of 0-14, 15-34, 35-64, and above 65. In forecasting, the 95% prediction intervals and mean values were computed from 10,000 bootstrap samples. Uncertainty was propagated from every submodel. To compare with the global reduction target of the End TB strategy [2], I forecasted the incidence until 2035. The milestones of 2020, 2030 and 2035 of the End TB strategy of percentage reductions in per capita-year TB incidence from 2015 were used as intermediate outcomes.

Incidence attributable to demographic change

I performed a scenario analysis to clarify the potential impact of demographic change. While forecasting the age-specific TB incidence to 2035, I kept the population size and age structure fixed as it was in 2018. This TB incidence was compared against values including projected changes in population structure by computing the fraction of total TB incidence attributable to demographic change in each year as $(I_{1,t} - I_{0,t})/I_{1,t}$, where $I_{1,t}$ and $I_{0,t}$ are the incident cases with and without demographic change respectively and t is the calendar year. This corresponds to the definition of population attributable fraction [26].

All the analyses were performed using R 3.5.1 [27] and analysed/visualised by R package `StMoMo`, `TSA`, `ggplot2` [28–30]. See Appendix J for the links of codes and data used in this chapter.

5.3 Results

5.3.1 Incidence modelling

Figure 5.1 shows the estimators of the Lee-Carter models of the incidence data. The age effect estimators (α_{age}) suggested the baseline incidence rates increase with age. In both sexes, the higher levels in age groups older than fifteen years correspond to higher TB incidence rates. The point estimators of age-period adjustments (β_{age}) showed no specific trend. However, there are large uncertainties for all estimates pertaining to under 15 year age groups due to the small numbers of notifications observed. The period effect estimators (κ_t) had nearly constant slopes with calendar years. Figure 5.1 also demonstrates the forecasting of period effects with 95% prediction intervals: prediction intervals of both sexes grew at a constant rate with calendar time. Table 5.1 shows the goodness of fit of the LCMs, the nested age-period Poisson models, and age-trend Poisson models. In AIC, BIC and log-likelihood on the training data, the LCM result is preferred over the other two although it costs a higher degree of freedom. Figure 5.2 shows the two-way residual plots of females and males by calendar years and age-groups. The positive and negative values were presented in red and blue blocks, respectively. As the blocks are reasonably random coloured, I identified no specific pattern by calendar years and age-groups. See Appendix D for the details of the goodness of fit, and residuals plots.

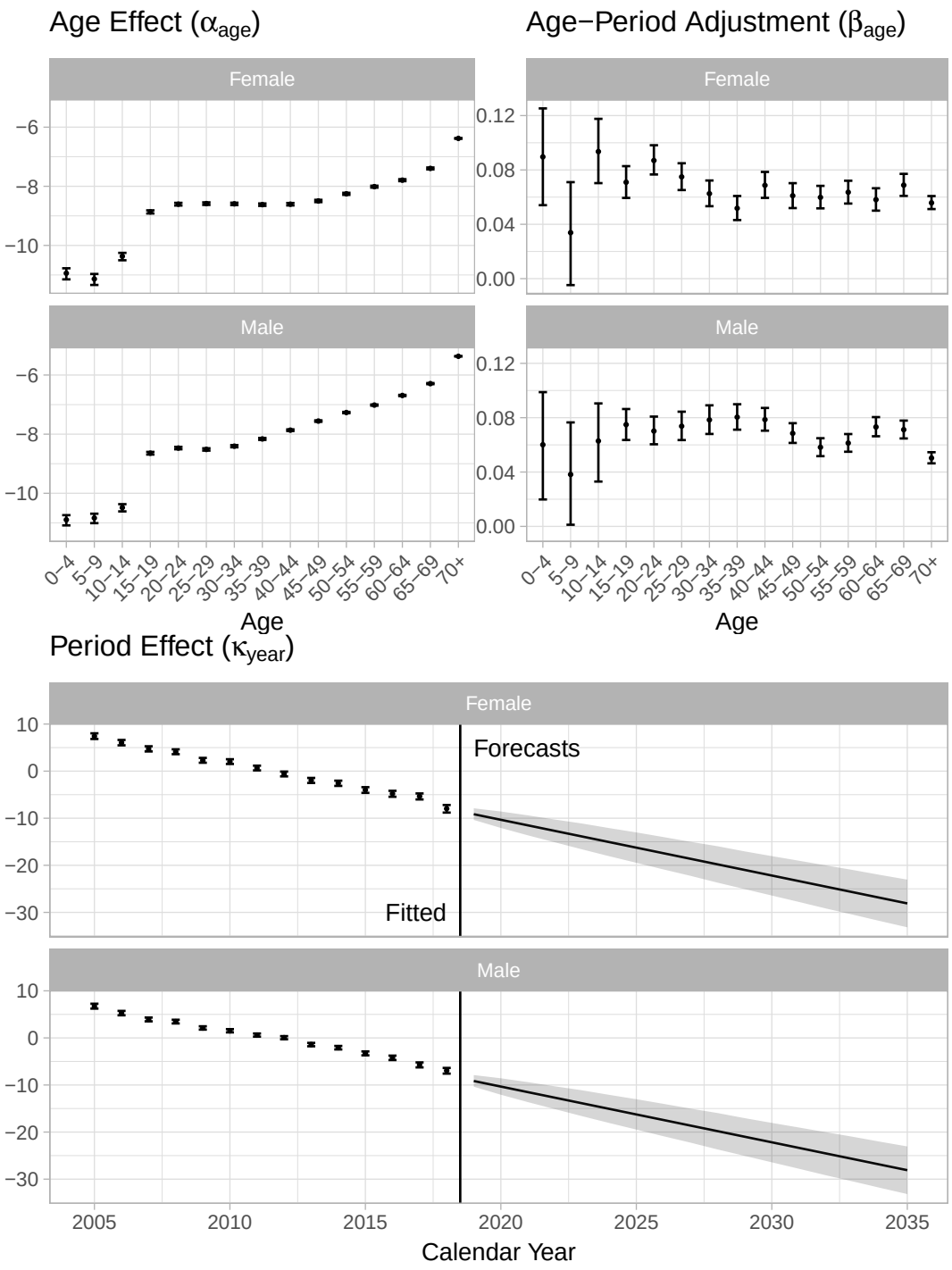


Figure 5.1: Lee-Carter model fitting and forecasting of the annual TB incidence. (Data: 2005-2018, Forecasting: 2019-2035). 95% confidence intervals of estimators and prediction intervals of forecasts were calculated through bootstrapping with 10,000 sample size.

Table 5.1: Summary of model comparison.

	Age-Trend	Age-Period	Lee-Carter Model
Model family		Poisson Regression	
Period effect	Linear	Discrete	Discrete
No. observations	420	420	420
No. parameters	32	56	86
Log(Likelihood)	-1855	-1819	-1682
AIC	3773	3751	3531
BIC	3902	3977	3871

AIC: Akaike information criterion, BIC: Bayesian information criterion

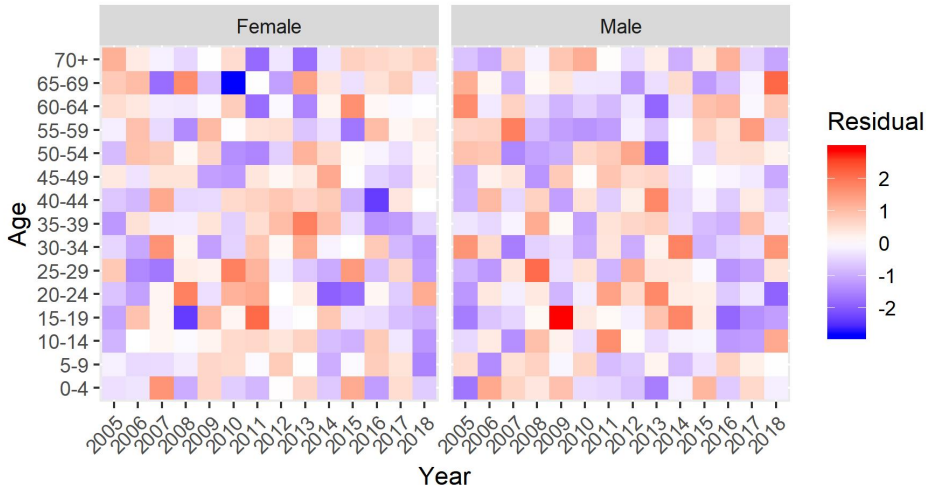


Figure 5.2: Residual plot by age and calender year.

5.3.2 Population forecasting

Figure 5.3 shows the demographic change from 2005 to 2035. In Figure 5.3-A, the population will reach a maximum of 23.6 million in 2023, and will start shrinking to 23.2 million in 2035. The proportion of the population aged over 65 is increasing across the period and will reach 27% in 2035. The proportion of the population aged under 15 is declining to around 11%. Figure 5.3-B compares the age structure of the Taiwanese population in 2018 and 2035, highlighting the population ageing.

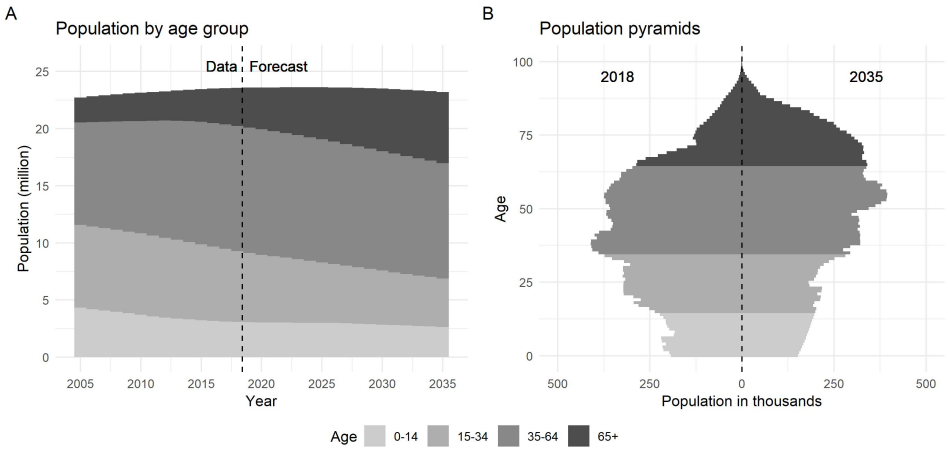


Figure 5.3: Demographic change. (Data: 2005-2017, Forecasting: 2018-2035)

Incidence forecasting and age structure

Figure 5.4 demonstrates the trends of the population TB incidence rate and TB incidence rates by age-group (< 15, $15-34$, $35-64$, $65+$). The forecast in Figure 5.4-A suggests that the TB incidence in 2035 will be 22 (95% Prediction Interval (PI): 19-25) per 100,000. The overall incidence reduction will reach 54% (95% PI: 45%-59%) in 2035, which is 37% short to reach the goal of the End TB Strategy. Figure 5.4-B shows the age-specific incidence rates will have 60% to 80% reductions from 2015 to 2035 apart from the 5-9 group. The rate reductions in most age groups will be higher than the forecast reduction of 44% in the whole population. Figure 5.4-C shows the overall incidence rates by age group as a stacked histogram. The TB incidence rates from age groups below 65 will be gradually decreasing whereas the above 65 will nearly stay constant from 2018 to 2035. Figure 5.4-D shows the proportion of TB incidence in each age group. The proportion among adults aged over 65 years will reach 68% (95% PI: 67%-69%) and 79% (95% PI: 78%-81%) in 2025 and 2035, respectively. In 2035, more than 97% of incident cases will occur among those aged 35 years or older, indeed the contribution from

cases under 15 years is nearly invisible in Figures 5.4-C and 5.4-D.

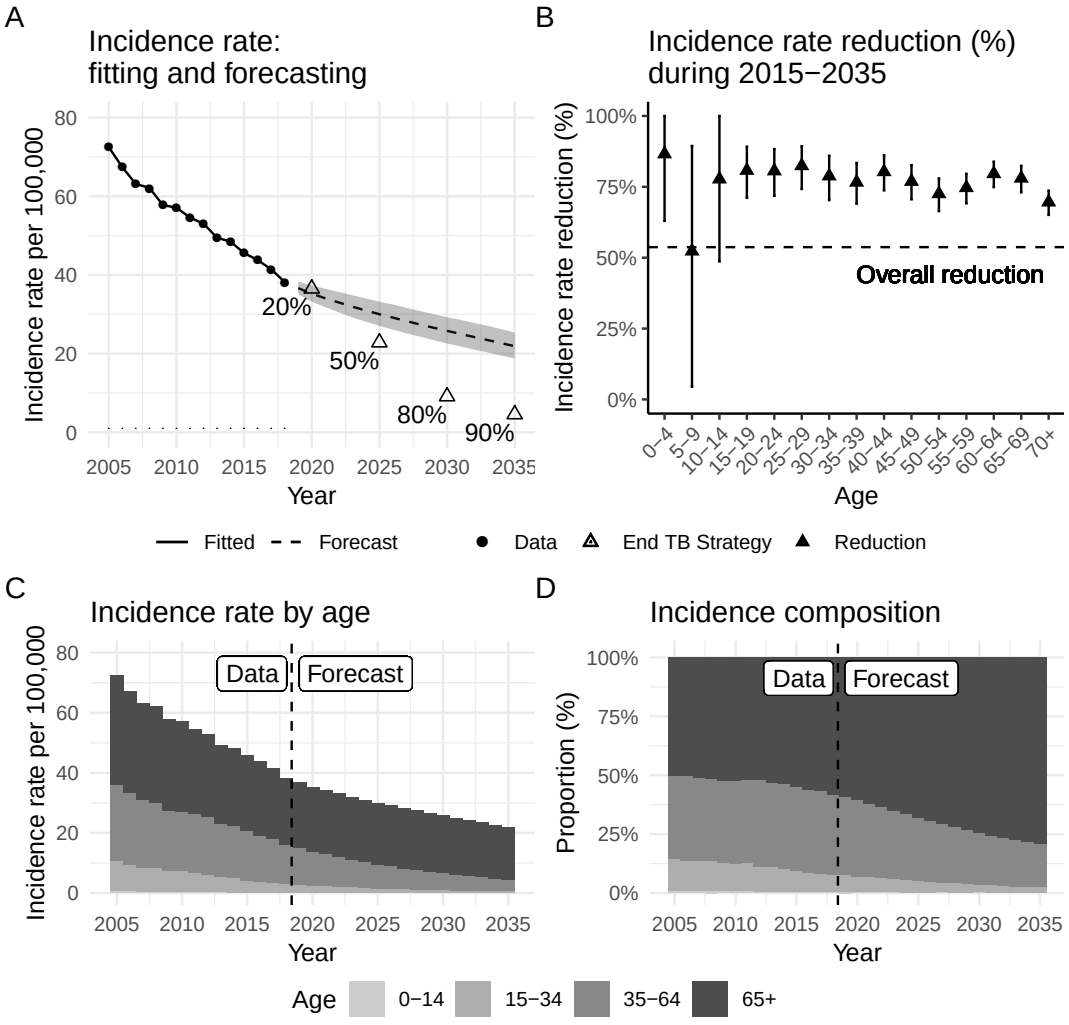


Figure 5.4: TB incidence rate forecasting. A. Overall incidence rate per 100,000-year. In the forecasting, dashed line features the mean values and the shaded area is 95% prediction interval. B. Incidence rate reductions by five-year age groups during 2015-2035 with 95% prediction interval. C. Incidence rates attributed to age groups. D. Proportions of age groups in Incidence cases.

5.3.3 Impact of demography on TB incidence

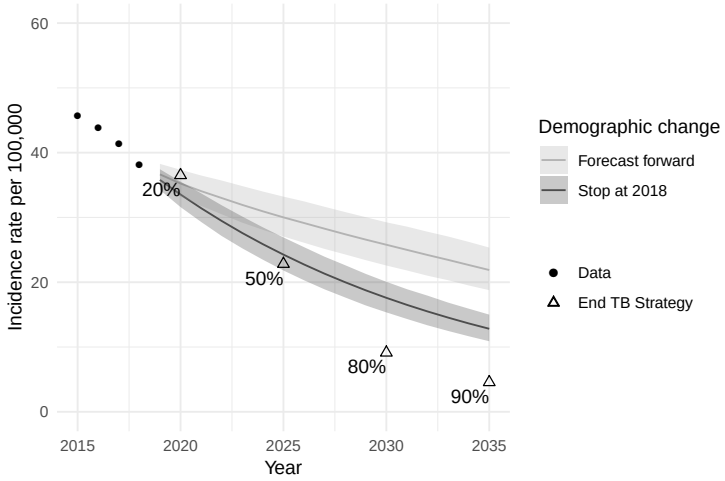


Figure 5.5: TB incidence with and without demographic change. (ribbons show 95% prediction intervals)

Table 5.2: Summary of reductions in TB incidence with and without demographic change.

Year	Percentage reduction in per capita TB incidence from 2015 (mean and 95% PI):		Percentage of total TB incidence attributable to demographic change (mean and 95% PI)
	with demographic change	without demographic change	
2020	23.1% (18.2%, 27.2%)	26.6% (22.2%, 30.8%)	4.6% (1.6%, 7.5%)
2025	35.2% (27.2%, 40.7%)	47.0% (41.0%, 52.2%)	18.3% (15.5%, 21.2%)
2030	45.0% (35.9%, 50.5%)	61.6% (56.0%, 66.4%)	30.0% (27.3%, 33.0%)
2035	53.7% (44.5%, 58.9%)	72.1% (67.1%, 76.1%)	38.8% (36.1%, 41.7%)

PI: prediction interval

Figure 5.5 shows the forecast incidence rates with and without demographic change. In the scenario without demographic change, the forecast suggests that the incidence in 2035 will be around 13 per 100,000 compared to 23 with demographic change and the 90% reduction target of 4.5 per 100,000. The 95% prediction intervals for forecasts with demographic change continuously expand year by year, whereas without demographic change they converge to a constant width within five years. Table 5.2 shows the impact of demographic change. Up to 2020, TB incidence rates will have 23% and 27% reductions with and without demographic change respectively. Considering the demographic change, the annual incidence rates are projected to reduce by 54% (95% PI: 45%-59%) from 2015 to

2035; without demographic change, the reduction will be 72% (95% PI: 67%-76%). In both scenarios, the trends of annual incidence rates showed diminishing reductions to the time scale. In 2035, the forecasts suggested that 39% (95% PI: 36%-42%) of incident TB cases can be attributed to demographic change.

5.4 Discussion

A substantial proportion of tuberculosis (TB) incidence in Taiwan is among people aged over 65 years. Social and economic development typically bring reductions in TB incidence but also reduced birth and death rates and population ageing. This study provides a novel investigation into the potential impact on TB incidence from population ageing using statistical modelling and forecasting. Current trends of TB incidence decline and demographic change suggest TB incidence rates in Taiwan will decrease to 25 per 100,000 by 2035. This represents a 45% reduction since 2015, missing the End TB goal of 90% reductions in TB incidence rates. I have shown that higher age-specific incidence rates in older age groups can mean that population ageing acts against reductions in TB rates, with TB incidence in 2035 projected to be 39% higher than without demographic change.

Previous studies have employed statistical methods either to forecast TB incidence, [12, 13, 31, 32] or to analyse patterns by age using age-period-cohort (APC) models, [14, 15] but I are the first study to statistically forecast age-specific TB incidence combining. Some transmission modelling studies [16, 18] have explored issues related to age-structure, and Arregui et al. [18] generated forecasts. However, the fitting in Arregui et al. was not likelihood-based and did not use age-specific TB data, and so could not evaluate age-specific goodness of fit for TB projections or compare alternative models with conventional metrics. I made novel use of Lee-Carter models (LCMs), [19, 22] which employ an elegant low-dimensional decomposition of age-specific rates to model trends and overall shape. LCMs were originally introduced for mortality rate modelling and are now the dominant approach, but have been applied elsewhere. Within demography, Hyndman [25] and Rueda-Sabater and Alvarez-Esteban [33] used LCMs to forecast the fertility rates, and Cowen [34] fitted LCMs to abortion rates. Kainz et al. [35] modelled chronic kidney disease prevalence as rate data, and Yue et al. [36] modelled cancer incidence and mortality. However, I am the first to apply LCMs to TB, finding they fitted better than Poisson Age-Period models. My approach offers a generalizable and easily-implemented method for forecasting age-specific TB incidence and the impact of demographic change on total TB incidence.

In my model fitting results, the age effects (α_{age}) demonstrated the TB inci-

dence rates positively correlated with age in both females and males. The period effect estimators (κ_t) were almost linear despite not assuming linearity in the LCM formulation. The declines may reflect improvements in infection control and case detection and the declining latent TB prevalence in each age group. Improvements in infection control and case detection both reduce the force of infection that will induce further TB incidence. For latent TB, which is accumulated during one's lifetime and depends on historical TB prevalent TB in history, different cohorts will have different prevalence. As TB incidence has been declining, the latent TB prevalence in recent cohorts will be lower than in historical cohorts at the same age. Lastly, the age-period interaction terms (β_{age}) were used to demonstrate how the incidence rate reduced differently in each age group, although no overall pattern was identified. The variance of the estimators in young people was larger because the only around 1% of incident TB (<100 cases every year in the recent decade) were from people below 15 years old.

For Taiwan and many other high-income settings in the Western Pacific region, TB notifications are thought to parallel TB incidence with only a small gap. [1, 37] Taiwan lacks survey data (e.g. capture-recapture studies) to directly inform on the magnitude of this gap. In settings where this gap is larger and changing over time, interpretation of TB notification data is more problematic, and notifications may not be a good proxy for incidence. Even in a declining TB epidemic with evolving case-mix, it is possible that case detection may change differently over time in different age-groups; we have not attempted to include such effects. Taiwan does not have United Nations Population Division demographic forecasts, hence my bespoke demographic modelling. For most nations, these forecasts could be used 'off the shelf'. I have presented results on percentage reductions in both per capita TB incidence rates and in total TB incidence (e.g. Table 2), which are similar because of Taiwan's small projected population change over the period considered; this may not be true in all settings. Later, in Section 8.7, I will measure the case-detection gaps in Taiwan with a modelling approach.

The decline in TB incidence in Taiwan probably has multiple contributory causes, including improvements in TB control, socio-economic development, and the reductions in the prevalence of latent TB as a result of declining transmission. For an infectious disease like TB, reduced transmission can amplify and sustain over time changes in underlying causative factors, complicating their analysis. The low TB rates in children aged under fifteen may reflect low exposure to TB in this group or potentially lower rates of case detection. My assessment of the impact of population ageing on TB incidence and case-mix has particular current relevance to many WHO Western Pacific region countries [1] and will be relevant

to many more countries and regions in the future. My analysis could provide a template for analysts who wish to explore issues related to future TB incidence and demography where age-specific data are available.

My analysis accounted for cohort propagation of latent tuberculosis infection (LTBI) in a phenomenological way. LTBI represents accumulated lifetime risk of infection by exposure to active tuberculosis disease. Older individuals in most settings have higher LTBI prevalence due both to longer cumulative exposure and (in declining epidemics) exposure to a higher mean infection rate over their lifetime. The ageing through of these LTBI positive cohorts thus generates a secular time trend in reactivation disease rates at a particular age. My approach does not explicitly model LTBI prevalence, because this would introduce additional parameters and, without LTBI data, identifiability issues. However, LTBI cohort effects are accounted for in my current approach indirectly by modelling the secular trends in age-specific incidence rates.

Another limitation of our approach is that it would fail to account for non-linear threshold behaviour, such as during outbreaks. However, in many high-income settings (including Taiwan), the steadily declining tuberculosis incidence implies the net reproduction number is below one. It is worth noting that according to Taiwan CDC surveillance, in 2005, 0.72% of TB cases in Taiwan were coded as HIV; neglecting HIV is unlikely to have impacted our results.

Extending the model by adding exogenous variables is possible. My analysis did not address the impact of other variables for simplicity and clarity. Important predictors could include socioeconomic status and comorbidities such as diabetes mellitus. [38] However, projections would require additional time-series analysis to forecast these explanatory variables.

Older age as a risk factor for TB disease has perhaps been under-explored since age is not a modifiable risk factor, and since in most current high-burden settings populations and the typical age of TB cases are fairly young. My result that population ageing will act to slow declines in TB incidence tallies with that of Arregui et al. [18], obtained for different settings using very different methods, and quantifies the magnitude of this effect However, the importance of older age groups to TB control is already evident in many Asian populations, [16] and this will be an increasingly widespread facet of global TB control if reductions in incidence continue and accelerate in the future. Older populations will also have their own particular challenges in terms of access, diagnosis and comorbidities complicating their care. Public health planning to develop adapted strategies for care and control to meet these changing population needs is essential.

5.5 Conclusion

The Lee-Carter model provides a tool to project age-specific tuberculosis incidence and hence forecast overall TB incidence while accounting for demographic change. In Taiwan, population ageing may slow the decline of TB incidence by 39% over the period 2015 - 2035. TB care and control programmes will increasingly need to address the needs of older adults, who will comprise a growing majority of the TB epidemic.

Bibliography

[1] World Health Organization. Global tuberculosis report 2018. 2018;.

[2] World Health Organization. Global strategy and targets for tuberculosis prevention, care and control after 2015. Geneva: World Health Organization. 2014;.

[3] Houben RMGJ, Dodd PJ. The Global Burden of Latent Tuberculosis Infection: A Re-estimation Using Mathematical Modelling. PLoS Med. 2016 Oct;13(10):e1002152.

[4] Schaaf HS, Collins A, Bekker A, Davies PDO. Tuberculosis at extremes of age. Respirology. 2010 Jul;15(5):747–763.

[5] Vynnycky E, Borgdorff MW, Leung CC, Tam CM, Fine PEM. Limited impact of tuberculosis control in Hong Kong: attributable to high risks of reactivation disease. Epidemiol Infect. 2008 Jul;136(7):943–952.

[6] United Nations Publications. World Population Ageing, 2015. UN; 2017.

[7] National Development Council, Taiwan. Population Projections for the R.O.C. (Taiwan): 2018-2065. National Development Council, Taiwan; 2018.

[8] Dye C, Williams BG. The population dynamics and control of tuberculosis. Science. 2010 May;328(5980):856–861.

[9] Pratt RH, Winston CA, Steve Kammerer J, Armstrong LR. Tuberculosis in Older Adults in the United States, 1993-2008; 2011.

[10] Hagiya H, Koyama T, Zamami Y, Minato Y, Tatebe Y, Mikami N, et al. Trends in incidence and mortality of tuberculosis in Japan: a population-based study, 1997-2016. Epidemiol Infect. 2018 Nov;p. 1–10.

[11] Negin J, Abimbola S, Marais BJ. Tuberculosis among older adults – time to take notice; 2015.

[12] Bras AL, Gomes D, Filipe PA, de Sousa B, Nunes C. Trends, seasonality and forecasts of pulmonary tuberculosis in Portugal. Int J Tuberc Lung Dis. 2014 Oct;18(10):1202–1210.

[13] van Aart C, Boshuizen H, Dekkers A, Altes HK. Time Lag Between Immigration and Tuberculosis Rates in Immigrants in the Netherlands: A Time-Series Analysis. International Journal of Tuberculosis and Lung Disease. 2017 May;21(5):486–492.

[14] Iqbal SA, Winston CA, Bardenheier BH, Armstrong LR, Navin TR. Age-Period-Cohort Analyses of Tuberculosis Incidence Rates by Nativity, United States, 1996-2016. Am J Public Health. 2018 Nov;108(S4):S315–S320.

[15] Wu P, Cowling BJ, Schooling CM, Wong IOL, Johnston JM, Leung CC, et al. Age-period-cohort analysis of tuberculosis notifications in Hong Kong from 1961 to 2005. Thorax. 2008 Apr;63(4):312–316.

[16] Harris RC, Sumner T, Knight GM, Evans T, Cardenas V, Chen C, et al. Age-targeted tuberculosis vaccination in China and implications for vaccine development: a modelling study. Lancet Glob Health. 2019 Feb;7(2):e209–e218.

[17] Brooks-Pollock E, Cohen T, Murray M. The impact of realistic age structure in simple models of tuberculosis transmission. PLoS One. 2010 Jan;5(1):e8479.

[18] Arregui S, Iglesias MJ, Samper S, Marinova D, Martin C, Sanz J, et al. Data-driven model for the assessment of Mycobacterium tuberculosis transmission in evolving demographic structures. Proc Natl Acad Sci U S A. 2018 Apr;115(14):E3238–E3245.

[19] Lee RD, Carter LR. Modeling and Forecasting U. S. Mortality. J Am Stat Assoc. 1992;87(419):659–671.

[20] Brouhns N, Denuit M, Vermunt JK. A Poisson log-bilinear regression approach to the construction of projected lifetables; 2002.

[21] Colin Cameron A, Trivedi PK. Regression Analysis of Count Data. Cambridge University Press; 2013.

[22] Lee R. The Lee-Carter Method for Forecasting Mortality, with Various Extensions and Applications. N Am Actuar J. 2000 Jan;4(1):80–91.

[23] Renshaw AE, Haberman S. On simulation-based approaches to risk measurement in mortality with specific reference to Poisson Lee–Carter modelling; 2008.

[24] Coale AJ, Kisker EE. Defects in data on old-age mortality in the United States: new procedures for calculating mortality schedules and life tables at the highest ages. Coale Kisker 1990 Asian and Pacifc Population Forum. 1990;.

[25] Hyndman RJ, Booth H. Stochastic population forecasts using functional data models for mortality, fertility and migration. Int J Forecast. 2008 Jul;24(3):323–342.

[26] Steenland K, Armstrong B. An overview of methods for calculating the burden of disease due to specific risk factors. Epidemiology. 2006 Sep;17(5):512–519.

[27] R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria; 2018.

[28] Wickham H. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York; 2016.

[29] Andres V, Millossovich P, Vladimir K. StMoMo: Stochastic Mortality Modeling in R. J Stat Softw. 2018;84(3):1–38.

[30] Chan KS, Ripley B. TSA: Time Series Analysis; 2018.

[31] Onozaki I, Law I, Sismanidis C, Zignol M, Glaziou P, Floyd K. National tuberculosis prevalence surveys in Asia, 1990-2012: an overview of results and lessons learned. Trop Med Int Health. 2015 Sep;20(9):1128–1145.

[32] Suarez PG, Watt CJ, Alarcon E, Portocarrero J, Zavala D, Canales R, et al. The Dynamics of Tuberculosis in Response to 10 Years of Intensive Control Effort in Peru. J Infect Dis. 2001 Aug;184(4):473–478.

[33] Rueda-Sabater C, Alvarez-Esteban PC. The analysis of age-specific fertility patterns via logistic models. J Appl Stat. 2008 Sep;35(9):1053–1070.

[34] Cowan SK. Cohort Abortion Measures for the United States. Popul Dev Rev. 2013 Jun;39(2):289–307.

[35] Kainz A, Hronsky M, Stel VS, Jager KJ, Geroldinger A, Dunkler D, et al. Prediction of prevalence of chronic kidney disease in diabetic patients in countries of the European Union up to 2025. Nephrol Dial Transplant. 2015 Aug;30 Suppl 4(4):iv113–8.

[36] Yue JC, Wang HC, Leong YY, Su WP. Using Taiwan National Health Insurance Database to model cancer incidence and mortality rates. Insur Math Econ. 2018 Jan;78:316–324.

[37] Lo HY, Yang SL, Chou P, Chuang JH, Chiang CY. Completeness and timeliness of tuberculosis notification in Taiwan. BMC Public Health. 2011 Dec;11:915.

[38] Marais BJ, Lönnroth K, Lawn SD, Migliori GB, Mwaba P, Glaziou P, et al. Tuberculosis comorbidity with communicable and non-communicable diseases: integrating health services and control efforts. Lancet Infect Dis. 2013 May;13(5):436–448.

Chapter 6

Individual patient pathway analysis

6.1 Introduction

Tuberculosis (TB) is a curable infectious disease but has become the world's leading infectious killer with unacceptably slow progress towards meeting global targets laid out in the End TB strategy [1]. These targets include the elimination of catastrophic costs due to TB [2], in recognition of the financial impact TB illness and care-seeking imposes on households [3]. One of the key gaps in TB control is the "missing millions": 4 in 10 people developing TB globally are either not diagnosed and treated, or are not notified to national TB programmes (NTPs) [1]. Understanding the health system pathways of patients who are seeking care for TB is key in identifying low-hanging fruit for TB control: programmatic shortcomings that mean patients are either missed, not notified, or incur ruinous expense in their efforts to obtain care.

Cascade of care analyses for TB have addressed the drop-offs at different stages from those needing treatment through to those completing treatment for active TB [4], and for latent TB infection [5]. This work has recently been extended with the introduction of patient pathway analysis (PPA) for TB [6], which provides a standardized analysis approach for drawing together data from different sources to identify misalignment between capacity for TB diagnosis and treatment, and patients' preferences in seeking care [7]. Initial work was undertaken in 5 countries [8–12], and notably found low coverage of diagnostic capacity for TB at facilities where patients typically first present [13]. This work has argued that more needs to be done to 'meet the patients where they are' [14], in line with the patient-centred care pillar of the End TB strategy.

However, work has so far relied on cross-sectional and aggregate data, meaning it has not been able to track the pathways of individuals as they interact with the health system. This makes it hard to investigate the key issue of delays prior to diagnosis. It also means it is not possible to understand patterns of referral between levels of facilities, the true complexity and heterogeneity of patient pathways, or the factors that influence these. Going beyond capacity alignment requires routine individual-level data with high coverage. Any data that includes individuals' interactions with the health system before a diagnosis of TB requires a methodology to interpret these events as related to TB care-seeking in light of their characteristics and timing. In this study, I develop a generalizable methodology for individual patient pathway analysis (IPPA) that makes use of health insurance data and apply it to Taiwan to construct and analyse the care pathways of patients who ultimately were diagnosed and treated for TB.

6.2 Methods

6.2.1 Setting

TB incidence in Taiwan has been declining: from 64 TB cases per 100,000 in 2007 to 39 per 100,000 in 2018. Taiwan is a high-income country and its healthcare system fully integrated with the National Health Insurance (NHI) programme. The NHI is compulsory (over 98% enrollment in 2017), and covers most essential healthcare services in both public and private sectors. The National Health Insurance Research Database (NHIRD) comprises records of all healthcare claims (e.g. diagnostics and medications) funded by the NHI.

6.2.2 Data

This chapter used a longitudinal dataset from the National Health Insurance Research Database (NHIRD). The dataset was released with a simple random sampled one million population based on the registration between 1996 and 2000, and its representativity had been validated with age and sex distributions in general population [15]. From the one million sampled individuals, this chapter used the individual characteristics (age, sex), healthcare records, and healthcare facilities visited from 2000 to 2010 for the analysis. The main reason for clinical visits and clinicians' diagnoses upon prescription are recorded in the NHIRD using the International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM). Appendix F lists the details of the dataset and describes the raw data.

The study was approved by the ethics committees of the University of Sheffield.

The NHIRD was provided by the National Health Insurance Administration, Ministry of Health and Welfare and managed by National Health Research Institutes (Registered number 100290). The interpretation and conclusions contained herein do not represent those of National Health Insurance Administration, Ministry of Health and Welfare or National Health Research Institutes.

6.2.3 Construction of individual patient pathways

To construct individual patient pathways, I restricted my sample to patients who between 2003 and 2010 met previously validated criteria for TB [16]. To include care-seeking prior to diagnosis, I used their records from 2000 to 2010 and followed a three-step process. First, I introduced a dynamic three-dimensional patient state with dimensions related to evaluation, treatment, and clinician consideration of related illnesses (all with default values of zero). Secondly, this state was used to parse sets of records into separate episodes of TB-related care (periods with ≥ 1 state non-zero; see Figure 6.1). Thirdly, the state values were used to interpret and label the stages of each episode, resulting in a pathway. The Taiwan CDC guidelines [17], WHO guidelines [18], and expert opinion on current clinical practice were used to define the algorithm for state transitions and stage labelling (see Table 6.1). In the rest of this section, I describe this approach in more detail.

The evaluation dimension represents the use of diagnostic procedures relevant to differential diagnosis of TB. Evaluation tools were ranked as probably for TB and possibly for TB according to how specific their use was to TB. Evaluation possibly for TB was triggered by various tools for examining the respiratory system and antibiotics for pneumonia. Evaluation probably for TB was triggered by the use of tools including chest computed tomography scan, chest x-ray, acid-fast bacteria culture, *Mycobacterium tuberculosis* culture, and tuberculin skin tests (see Appendix E for detailed definitions).

The treatment dimension represents the use of anti-TB drugs. In accordance with Taiwan CDC guidelines, anti-TB treatment requires more than two types of anti-TB drugs prescribed continuously for at least four weeks. Treatment was either first-line or second-line according to the drugs prescribed (see Appendix E).

The related illnesses dimension captures potential clinician consideration of alternative diagnoses. Increased values were triggered by features including ICD-9-CM codes corresponding to acute respiratory infections, chronic lung diseases, and nontuberculous mycobacterial infections.

Each state's value could return to zero after a 'time-out' period if no relevant trigger records occurred. A patient's treatment state also returned to zero after treatment completion. The choice of time-out sets a memory time-scale beyond

which records no longer influence a patient's state. The default time-out was 60 days, and sensitivity analyses used 30, 90 and 120 days.

Care-seeking episodes were defined as collections of records including anti-TB treatment (excluding exploratory treatment), during which a patient's state was continuously above zero in at least one dimension, separated by periods with

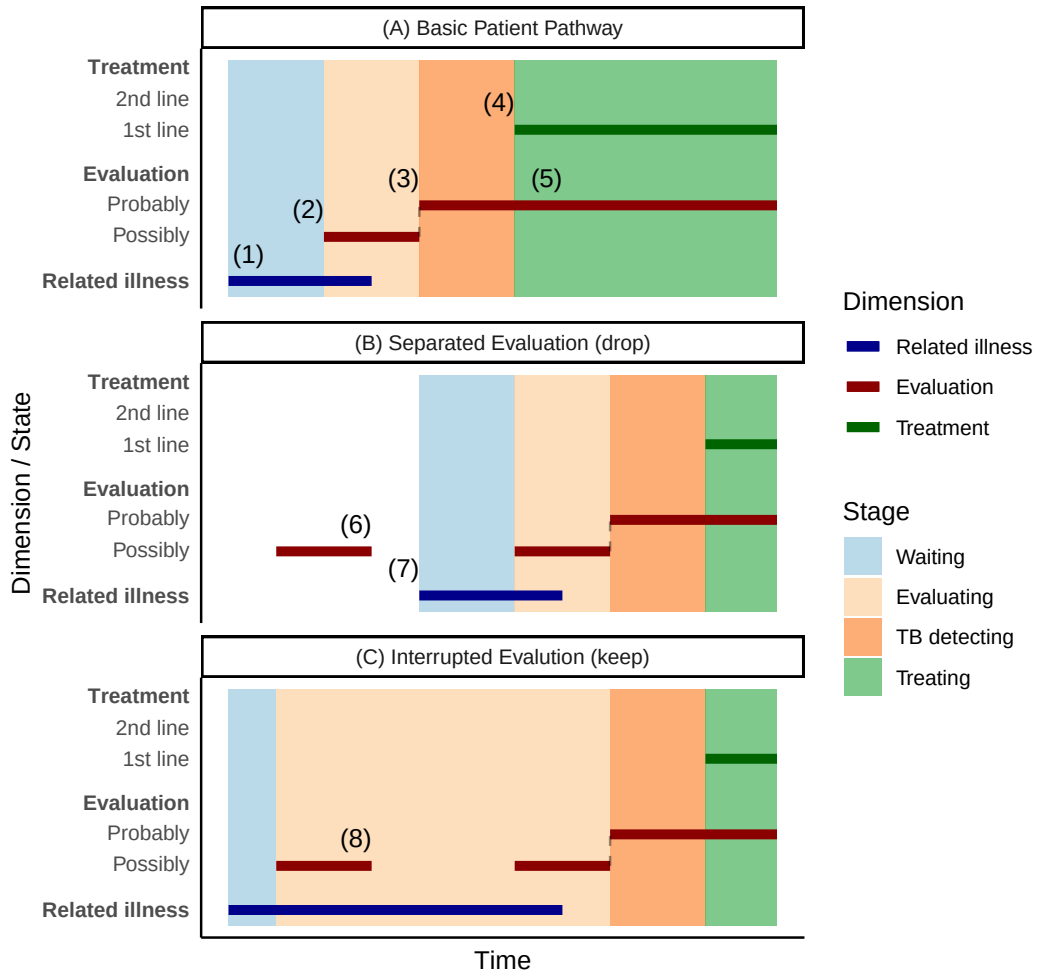


Figure 6.1: Patient pathway construction. Bars show patient states in three domains where non-zero. Background colours show the stage label within a pathway. **(A)** Basic patient pathway: (1) corresponds to initial care-seeking; (2) and (3) represent escalating clinical consideration of TB; (4) is treatment initiation; and (5) represents diagnostic evaluation while on anti-TB treatment. **(B)** Separated evaluation series. Considering there is no count in Related illness dimension the evaluation (6) would be dropped as separated by longer than 60 days and the pathway would begin at (7). **(C)** Interrupted evaluation. With evidence of clinician consideration of a related illness, the evaluation at (8) is included in the pathway.

all dimensions equal to zero (see Figure 6.1-A). Episodes could include records related to care-seeking prior to TB diagnosis. Without evidence that alternative diagnoses were being actively considered (i.e. related illness state and evaluation state both return to zero), prior evaluations are dropped from an episode (Figure 6.1-B). If there was evidence of ongoing consideration of alternative diagnoses to TB (non-zero related illness state), diagnostic procedures relevant to TB would be included in the episode even if the evaluation state returned to zero in the interim (Figure 6.1-C); this is termed an 'interrupted evaluation'.

To complete pathway construction, time within parsed episodes was labelled as one of four main stages according to the values of the state (Figure 6.1, background colors): Waiting Stage (before evaluation), Evaluating Stage (under evaluation for possible TB), TB Detecting Stage (under evaluation for probable TB), and Treating Stage (on TB treatment). Figure 6.2 shows all possible flows between states of my patient pathways.

Figure 6.2: Stages/states of patient pathways.

Table 6.1: Definitions and descriptions of stages/states

Stage	State	Description
Waiting	Waiting	Waiting for first TB-related evaluation or treatment
Evaluating	Evaluating	Being evaluated by procedures which can narrow the possibility down to TB (Evaluations possibly for TB)
	Interrupted Evaluation (IE)	Previous evaluations do not narrow the possibility down to TB because of (1) comorbidity, (2) false negative, or (3) self-referral.
	Re-Evaluating	Visiting Evaluating State after Interrupted Evaluation
TB-Detecting	TB-Detecting	Being evaluated by procedures which can identify TB (Evaluations probably for TB)
	Re-Detecting	Re-visiting TB-Detecting State after Interrupted Evaluation
Treating	First-line treatment	Being treated with first-line TB regimen
	Treatment change	Switching between two TB treatments or temporal treatment interruption due to health conditions.
	Second-line treatment / retreatment	Being treated with second-line TB regimen or any regimen after Treatment change

6.2.4 Analyses of individual patient pathways

Individual patient pathways were analysed to understand the coverage of services at different levels of the health system and referral patterns between them. Hospitals were classified into one of four levels following a standard classification (A for primary care/ general practice; B for regional hospitals with inpatient capacity; C for larger district hospitals; D for large hospitals with a range of specialists). I then calculated the coverage of each service at a given level of facilities (the proportion of facilities of the given level where the service was available), as well as the patient access to these services (proportions of levels of hospitals where patients initially sought care). I summarized these metrics with a modified version of the PPA visualization introduced by Hanson et al. [7] I also visualized the flow of patients between health service levels by pathway stage. I computed the distribution of delays from initial care-seeking to first reaching facility with anti-TB treatment capacity, to first reaching the facility where ultimately treated, and treatment initiation.

The heterogeneity in patient pathway complexity and time spent in different stages was visualized by sequence frequency plots, separately for sub-pathways leading up to, and following, treatment. To summarize time taken to reach a given stage of care, I also visualized the number of pathways reaching a given sub-stage by time after initial care-seeking. I calculated the proportions: of all clinical encounters among the pathways with the 10% least common sequence patterns; of these pathways for which treatment started over one year after initial care-seeking; of all pathways experiencing interrupted evaluation. Finally, I analysed risk factors, including age, sex, area, comorbidity and hospitals at initial care-seeking, for interrupted evaluation using logistic regression.

The pathway extraction was performed in Python 3.6; statistical analysis and visualisation in R 3.5.1 with `ggplot2`. Appendix G for implementation details, links in Appendix J for online documents and related code repositories.

6.3 Results

6,258 patients met my criteria for TB between 2003 and 2010. My algorithm generated 7,255 distinct pathways for these patients: 88% of patients had only one pathway. The use of longer timeouts resulted in fewer distinct pathways: 7,528 for 30 days, 6,963 for 90 days, and 6,831 for 120 days (See Appendix H, sensitivity analysis). Table 6.2 demonstrates the demographies of selected patient pathways and notification data. In the patient pathways, males contributed 70% of patient

pathways; 50.5%, 49%, and 0.5% of pathways were from those aged over 65, between 15 and 64, and below 14, respectively. The age and sex distributions of patient pathways were close to that of notification data in both 2005 and 2010.

Table 6.2: Demography of selected patient pathways compared with the notification data in 2005 and 2010.

	Patient pathways	Notification	
	from the NHIRD	2005	2010
All	7255 (100%)	16472 (100%)	13237 (100%)
Sex			
Female	2158 (29.7%)	5069 (30.8%)	4106 (31.0%)
Male	5097 (70.3%)	11403 (69.2%)	9131 (69.0%)
Age			
$<$15	36 (0.5%)	139 (0.8%)	97 (0.7%)
15-64	3558 (49.0%)	8064 (49.0%)	6192 (46.8%)
$\geq$65	3661 (50.5%)	8269 (50.2%)	6948 (52.5%)

6.3.1 TB-related services coverage and assess

Coverage and access by level of the health system are shown in Figure 6.3-A. This diagram was adapted from Hanson et al. [7], composed of TB service availabilities and accessibilities at first care-seeking attempts by different hospital levels. From Figure 6.3-A-1, the sizes of bars indicate the proportions of hospital levels at initial care-seeking. That is, there were 43%, 17%, 23% and 17% patient pathways initialised in level A to D, respectively. Figure 6.3-A-2 reveals the percentage of hospitals in each level that can provide the TB service of Evaluation possibly for TB. Taking Figure 6.3-A-1 and Figure 6.3-A-2.1 together, Figure 6.3-A-2.2 assesses how many patient pathways can access the TB service at their initial care-seeking. Namely, 7.3% of the patient pathways started with level A hospitals and can access the service at their initial care-seeking. In summary, initial care-seeking was most common at level A, but further evaluations and treatment rarely occurred at this level; only a minority of these facilities (around 15%) had the capacity for TB diagnosis and treatment. Coverage at facility levels higher than A exceeded 93%. Figure 6.3-B shows the associated flow of patients between levels of the health system, with an upwards flow across all stages. Around 50% of patient pathways began at levels A or B (36% where TB treatment was unavailable), but over 74% of pathways initiated treatment at levels C or D. Figure 6.4 shows the associated flow of patients between inpatients and outpatients, with an upwards flow across all stages. Around 85% of patient pathways were started as outpatients, while 40% of pathways initiated treatment as inpatients. There were 10% of pathways started

TB detection as in-patients but treated as out-patients.

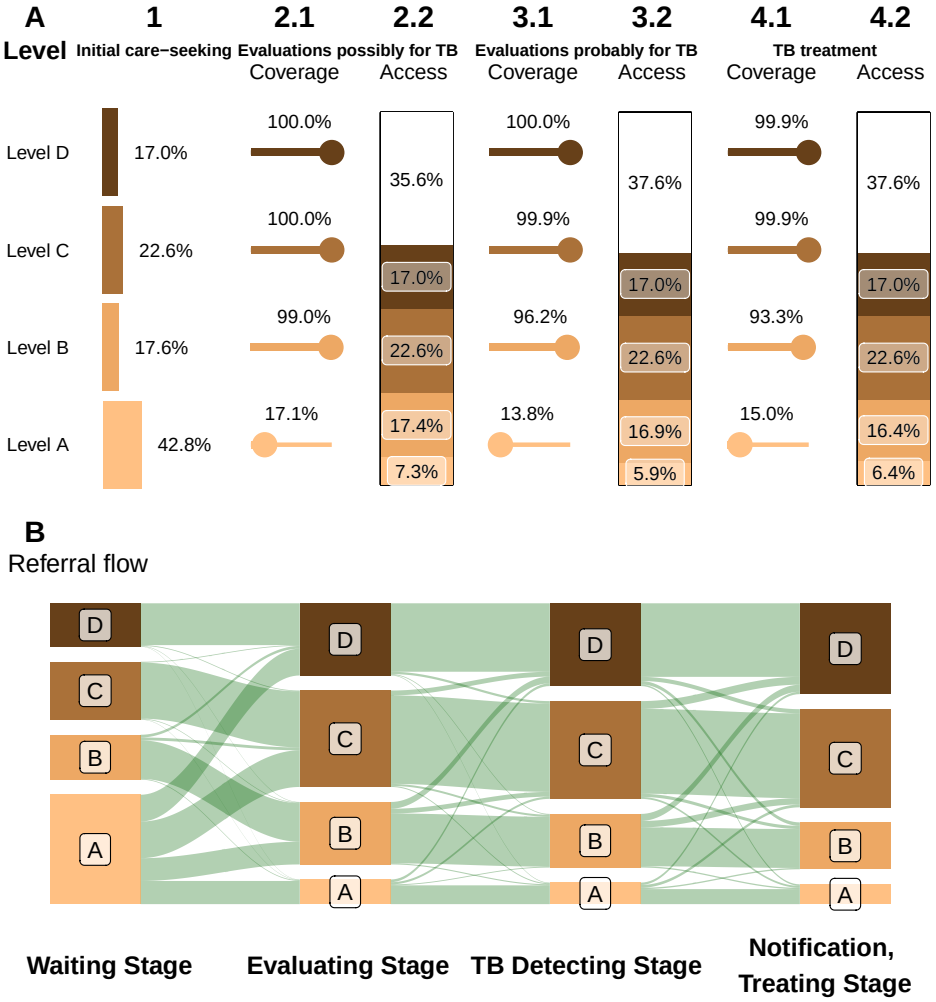


Figure 6.3: Alignment between the capacity for TB diagnosis and treatment, and patients' preferences in seeking care. Hospital levels: A for primary care/ general practice; B for regional hospitals with inpatient capacity; C for larger district hospitals; D for large hospitals with a range of specialists. **(A)** Coverage of and access to TB services at different levels of Taiwan's health system modified. This panel is a modified version of the visualization introduced by Hanson et al. [7] Coverage is the proportion of facilities at that level offering a service; access is the product of coverage and the fraction of patients seeking care at that level. **(B)** Patient flows between levels at each stage, including clinician- and self-referral (vertical heights of bands are proportional to numbers).

Delays from initial care-seeking to: a facility offering TB treatment; the facility ultimately providing treatment; and to treatment initiation are shown in Figure 6.5. 16% of pathways initiated treatment at their initial visit. By day 60, TB treatment was available for 90% of the pathways and 80% had arrived at the

facilities ultimately providing their treatments. By day 180, 84% of pathways had started treatment. The median system delay (from initial care-seeking to treatment initiation) was 41 days.

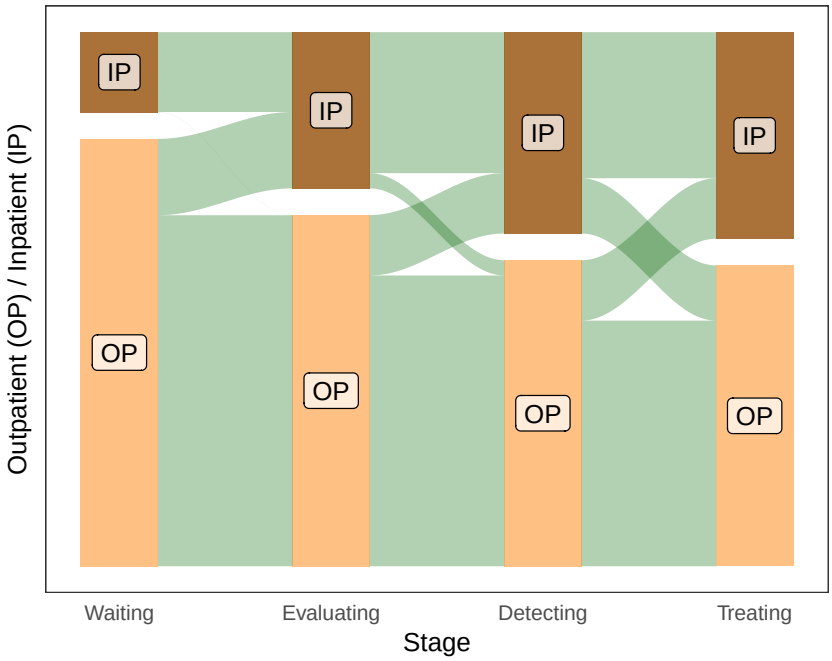


Figure 6.4: Patient flows between outpatients/inpatients at each stage, including clinician- and self-referral (vertical heights of bands are proportional to numbers).

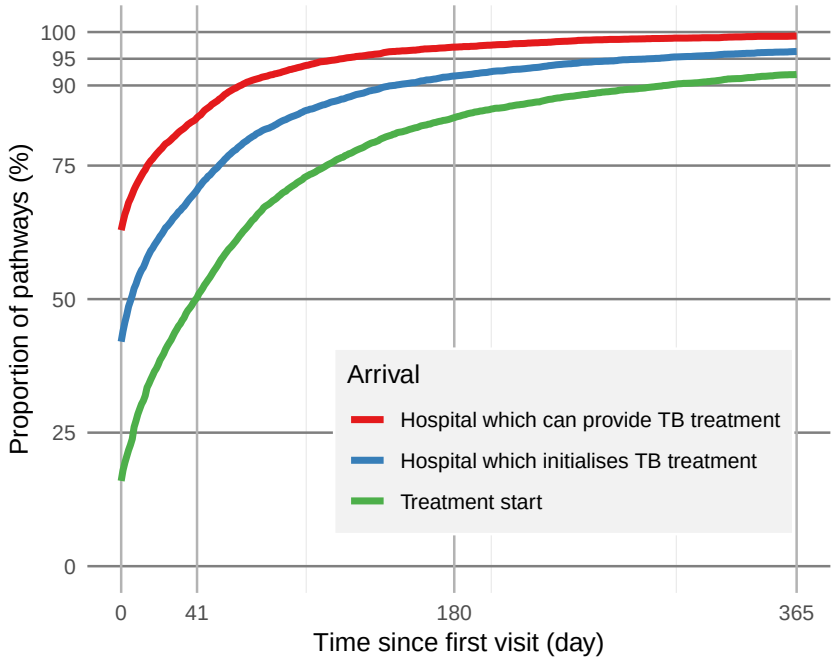


Figure 6.5: Delays from initial care seeking. The curves indicate the proportion of pathways by time that have: reached hospitals providing TB treatment (red); reached the hospital which ultimately initiates their TB treatment (blue); and started treatment (green). The vertical line on day 41 denotes the median system delay.

6.3.2 Heterogeneity among patient pathways

The typology of pathways before and after treatment start is shown in Figure 6.6. Before treatment starts, simpler pathways were most common. Second line anti-TB treatments (including fluoroquinolones) were used in 34% of pathways. However, the most complex pathways (i.e. pathways with the rarest 10% of sequence types, see Figure 6.6), exhibited intricate patterns of evaluation interruptions and re-evaluations; 100% of these pathways included interrupted evaluation.

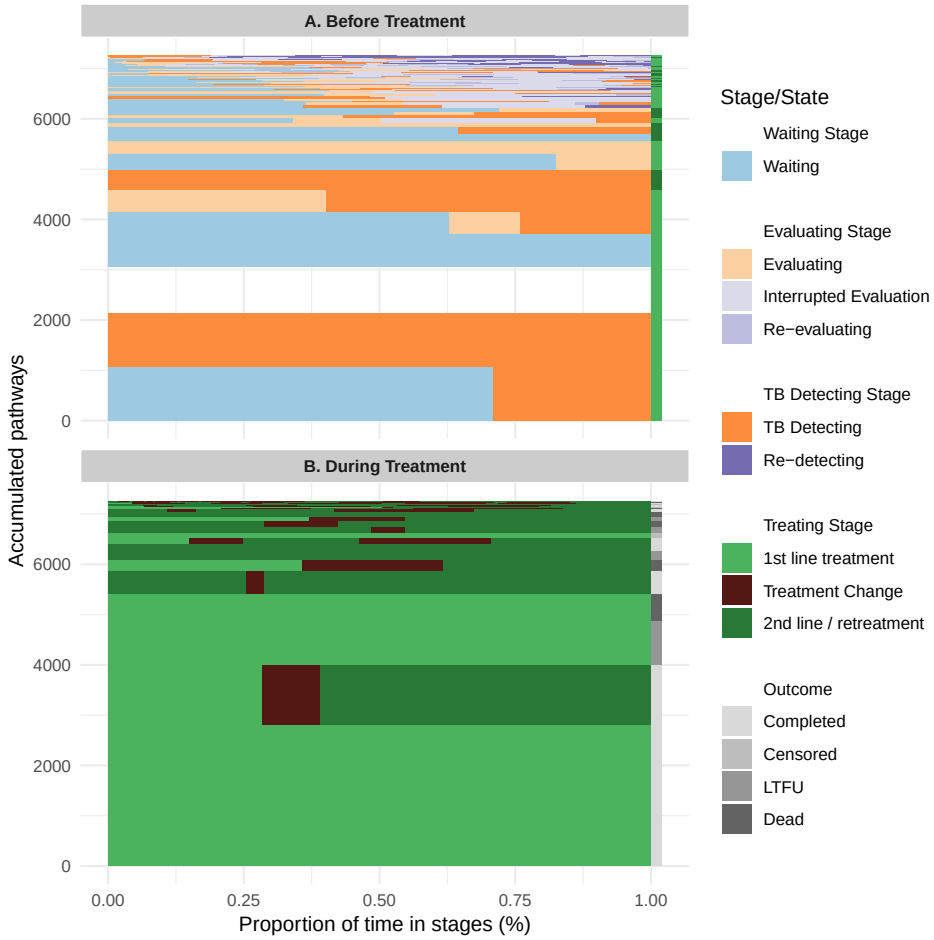


Figure 6.6: Cumulative pathway typology before and after treatment initiation. The aggregated number of pathways with a given sequence of stages (y axis) versus the proportion of their time in each stage (x axis). **(A)** Sequence patterns before treatment initiation **(B)** Sequence patterns from treatment initiation to treatment outcome. Completed = treatment period longer than 180 days. Censored = end with the end of data time frame. LTFU = lost to follow-up, incompleted treatment for no reason. See chapter text and Table 6.1 for stage definitions.

The pattern of delays from initial care-seeking to later stages of care is shown

in Figure 6.7, truncated at one year. The long tail of 7% of pathways yet to start treatment a year after initial care-seeking is clearly visible. One year after initial care-seeking, 21% of pathways were still on TB treatment (40% of these on second-line treatment), 41% had completed treatment, 18% were lost to follow-up, and 10% of patients had died. 55% of the 10% most complex pathways were yet to initiate treatment a year after initial care-seeking.

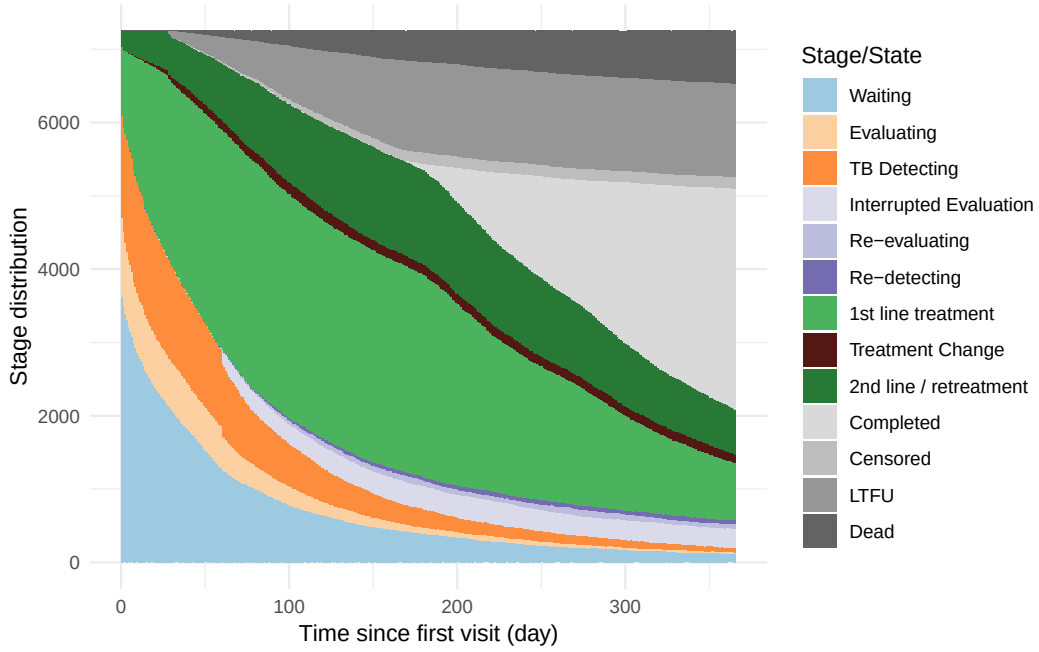


Figure 6.7: Time from initial care-seeking to given care stage. Completed: treatment period longer than 180 days. Censored: end with the end of data time frame. LTFU: lost to follow-up, incompleted treatment for no reason. See chapter text and Appendix E for stage definitions.

6.3.3 Interrupted evaluation

Interrupted evaluation occurred in 16% of all pathways, but these pathways included 48% of all visits/records and comprised a majority of the most complex and delayed pathways (see above). The median delay from initial care-seeking to treatment initiation in pathways with interrupted evaluation was 313 days (interquartile interval 164 - 607 days). Table 6.3 shows that pathways for patients aged 65 or older, odds ratio (OR) 2.4 (95% confidence interval (CI): 2.1 - 2.8) compared with patients aged 15-64, those with chronic lung conditions, OR 1.9 (95% CI: 1.6-2.2), and pathways, and which included 2003 (the year of the severe acute respiratory syndrome (SARS) epidemic), OR 2.3 (95% CI: 2.0-2.7), were signifi-

cantly more likely to experience interrupted evaluation. Pathways for patients with diabetes mellitus and who sought care at hospitals higher than Level A were significantly less likely to include interrupted evaluation, OR 0.64 (95% CI: 0.53-0.78).

Table 6.3: Logistic regression analysis of risk factors for interrupted evaluation.

	Number	IE (%)	Crude OR(95%CI)	Adjusted OR(95%CI)	P(Wald's test)	P(LR-test)
All	7255	1125(16%)				
Age						<0.001
< 15	36	4(11%)	1.18 (0.41,3.34)	0.97 (0.33,2.8)	0.952	
$15 - 64$	3558	342(10%)	reference			
≥ 65	3661	779(21%)	2.54 (2.22,2.91)	2.42 (2.1,2.8)	<0.001	
Sex						0.817
Female	2158	315(15%)	reference			
Male	5097	810(16%)	1.11 (0.96,1.27)	0.98 (0.85,1.14)	0.817	
Area						<0.001
Center	1332	234(18%)	reference			
East	419	76(18%)	1.04 (0.78,1.38)	1.12 (0.83,1.51)	0.458	
Kaohsiung city	1507	267(18%)	1.01 (0.83,1.23)	1.09 (0.89,1.33)	0.419	
North	740	98(13%)	0.72 (0.55,0.92)	0.77 (0.59,1)	0.054	
South	1087	179(16%)	0.93 (0.75,1.15)	0.95 (0.76,1.19)	0.672	
Taipei city	2170	271(12%)	0.67 (0.55,0.81)	0.75 (0.62,0.92)	0.005	
Comorbidity						
CLC	1192	277(23%)	1.86 (1.6,2.17)	1.91 (1.62,2.24)	<0.001	<0.001
DM	1191	136(11%)	0.66 (0.55,0.8)	0.64 (0.53,0.78)	<0.001	<0.001
HIV	30	4(13%)	0.84 (0.29,2.4)	1.52 (0.51,4.49)	0.451	0.472
Initial level at						<0.001
A	3104	539(17%)	reference			
B	1275	203(16%)	0.9 (0.76,1.08)	0.78 (0.65,0.93)	0.007	
C	1644	210(13%)	0.7 (0.59,0.83)	0.68 (0.57,0.82)	<0.001	
D	1232	173(14%)	0.78 (0.65,0.94)	0.75 (0.62,0.92)	0.005	
Pathway overlaps 2003	2110	493(23%)	2.18 (1.91,2.48)	2.32 (2.02,2.66)	<0.001	<0.001

OR: odds ratio, CI: confidence interval, LR-test: likelihood-ratio test, IE: Interrupted Evaluation, CLC: chronic lung condition, DM: diabetes mellitus, HIV: human immunodeficiency virus
Comorbidities and Pathway overlaps 2003 are binary variables, applying 'None' as the reference groups. Crude and adjusted ORs were calculated from the results of univariate and multivariate logistic regressions respectively.

Figure 6.8 reveals the distribution of healthcare visits before treatment initiation by the Lorenz curve [19] regarding chronic lung condition and IE. From left

to right patient pathways were sorted by the number of healthcare visits, and the y-axis cumulated the number of visits. Three-quarters of patient pathways shared only one-quarter of the visits. However, the 10% patient pathways with the most visits shared 50% of all healthcare visits. The two colour curves feature proportions of the attributes of CLC and IE by moving averages through every 10% of patient pathways. The proportion of IE shows that 77% of the right 10% of pathways experienced IE before treatment initiation. The curve of patients aged 65 showed the older patient needed more healthcare visits to treatment start. The curve of CLC shows a U-shape, which is relatively common in the pathways with very low and very high visits. However, the trend of CLC curve, framed between 10% and 25%, is much slighter than IE, from 0% to 77%.

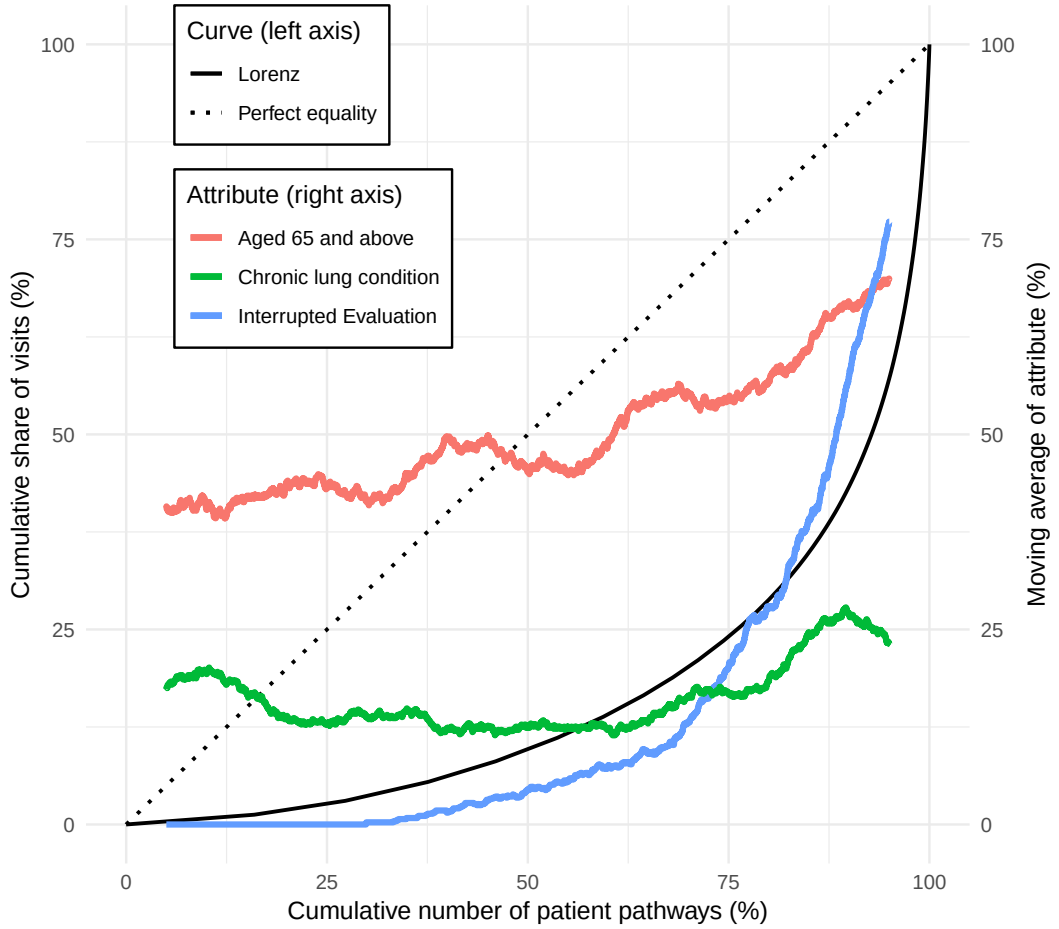


Figure 6.8: Lorenz curve of visits during evaluation

6.4 Discussion

In this study, I constructed individual patient pathways of tuberculosis care-seeking, diagnosis and treatment, using routine national health insurance data in Taiwan. To do this, I developed a generalizable method to algorithmically interpret insurance claim records in light of the context of previous and future events on the pathway. This approach was necessary in order to determine whether healthcare utilisation prior to eventual diagnosis was in fact related to a given episode of TB episode. Even in this well-resourced setting, there is low apparent coverage of TB care capacity at the lowest-level facilities where patients most commonly started seeking care, and substantial heterogeneity in the duration and complexity of TB-related care pathways.

The most complex patient experiences and the most prolonged delays prior to treatment were associated with what I have termed interrupted evaluations. These pathways are those that include unusually long delays between the first relevant diagnostic procedure and TB diagnosis, as well as evidence that alternative causes of disease are being considered. Interestingly, my regression analysis showed that patients with diabetes mellitus (DM) were significantly less likely to experience interrupted evaluation, perhaps indicating an awareness among clinicians of DM as a risk factor for TB. However, patients whose care-seeking overlapped with the 2003 SARS epidemic, patients who had chronic lung conditions, or patients aged 65 or above were all significantly more likely to experience interrupted evaluation. The SARS epidemic may have shifted attention towards alternative aetiologies of respiratory symptoms, as well as potentially generating increases in case-load that distorted usual practice. Comorbid chronic lung disease may have obscured TB symptoms, which would be unfortunate as chronic lung disease (e.g. chronic obstructive pulmonary disorder) shares risk factors with TB and may causally increase the likelihood of TB and worsen outcomes [20–22]. Lastly, differential diagnosis of TB may be more complex for older adults, and successful treatment is more challenging [23].

In addition to service coverage, I was also able to analyse the referral flows of patients between different levels of the healthcare system. I found a general trend for escalation in level from initial care-seeking, with most patients seeking care at lower levels and most being treated at higher levels. While coverage of TB services was very high except at the lowest level, 36% of first attempts to seek care were still at facilities without the capacity to diagnose or treat TB. Patients did rapidly reach facilities where TB treatment was available, but 34% initiated treatment at a facility other than the first one they visited offering TB treatment,

and the median delay between arriving at the facility where they were treated and treatment initiation was 6 days. These features may represent a lack of confidence or familiarity in diagnosing TB among clinicians at the primary care level. Further investigations to understand the reasons for this are warranted.

My work builds on the patient pathway analysis (PPA) introduced in a series of recent papers [7–13]. I was able to obtain similar outputs, including similar conclusions around coverage at the facility level typical of initial visits, but there are a number of differences which stem from my use of longitudinal individual-level data. My coverage statistics for the availability of diagnosis at facilities are based not on direct facility data, but on inferred availability from events in patient records. This will underestimate the actual coverage at facilities where tools are available but have not been used for patients in my data. This is likely to be particularly true at the lowest level of care. Importantly, my definition of 'initial care-seeking' differs from that used up to now in PPA, where initial care-seeking refers to patients' preferred location to first access care for symptoms related to TB, determined from retrospective surveys [7]. My initial care-seeking events need not involve the initiation of care; indeed multiple visits are typical between initial care-seeking and treatment initiation, including missed opportunities for earlier diagnosis. My approach to identifying care-seeking related to a particular TB episode also differs from that used in Chen et al. [16]. Due to the potential for interrupted evaluation. This is why my median delay from initial visit to treatment (41 days) is somewhat longer than theirs (29 days), and longer than the median health system delay reported by a systematic review [24].

The key strengths of this work lie in the use of individual-level data from a large population-representative sample, and the analyses of heterogeneity, referral patterns, and determinants that this allowed. Due to the compulsory National Health Insurance system, I had available data relating to essentially all clinical encounters over a long period of time in individuals diagnosed with TB. Through my algorithm, I was able to identify early TB care-seeking seeking events prior to diagnosis and analyse their patterns. It has been argued that the correct start-point for cohort-based analysis of patient care is the initial attempt to seek care rather than successful diagnosis or treatment initiation [6]. However, this has previously relied predominantly on experiences reported retrospectively by patients [24], which are subject to recall biases. I was also able to link patient-level covariates, including comorbidities for risk-factor analysis.

The main limitations of my analysis are around the nature of routine data, and the assumptions needed by any algorithm to interpret these records in terms of TB-related care-seeking. For example, my choice of 'time-out' period to determine

whether respiratory-related care-seeking events were part of a TB patient pathway affects how protracted and fragmented pathways are prior to diagnosis. The initial value of 60 days was based on a previous study evaluating health system delay of TB diagnosis in Taiwan [16]. However, sensitivity analyses varying the choice of time-out showed that my main conclusions were not substantially affected. I relied on clinician coded ICD-9-CM codes to characterise clinical encounters; there may have been miscoding or omission of relevant secondary diseases for multimorbid patients. The choices of timings characterising treatments were based on standard treatment, and may not have applied to extra-pulmonary TB, although rates of extra-pulmonary TB are low in Taiwan. In addition, my methods may work less well for TB patients with underlying chronic lung disease, for which they also interact with health services. Finally, my analysis cannot account for the delay between the start of infectiousness or symptoms and initial care-seeking. Prevalence-to-notification ratios in high TB burden settings typically correspond to a mean disease duration over a year [25, 26], whereas the predominant approaches to measuring delays by asking patients typically result in shorter delays of the order 6 months [24]. My analysis sheds light on the health system delay contribution in this setting (from care-seeking to diagnosis) and suggests mean durations are strongly influenced by the few individuals with very long delays.

I developed an approach to analysing patient-level routine data using health insurance data for Taiwan, but this analysis framework could be adapted and applied to other settings, as well as forming the basis of further analyses for Taiwan. The prerequisite for this analysis is individual-level high-coverage routine data on healthcare utilisation, ideally beyond just TB services. While such data are still relatively rare in many high TB burden settings, there is increased development of case-based notification systems for TB, and there are subnational health informatics and linkage efforts that may allow such an analysis [27]. Use of different diagnostic codes, different locally relevant comorbidities (e.g. HIV) and clinical practices could all be included by adapting the open source code for my analysis. Likewise, consideration should be given to different 'time-out' values as these are inevitably related to a system and patient characteristics. Future work building on my analysis for Taiwan will include analysis considering health system costs and patient prescription charges, analysis with respect to proxy measures of socioeconomic status, and developing simulation models replicating observed patterns of patient care-seeking.

6.5 Conclusion

Patient-level routine healthcare data can be analysed to generate individual patient TB care-seeking pathways. Training to improve clinician awareness of TB in key risk groups at the lowest level of the health system may improve patient experience of accessing TB care in Taiwan.

Bibliography

[1] World Health Organisation. Global tuberculosis report 2017; 2017.

[2] Uplekar M, Weil D, Lonnroth K, Jaramillo E, Lienhardt C, Dias HM, et al. WHO's new End TB Strategy. Lancet. 2015;385(9979):1799–1801.

[3] Tanimura T, Jaramillo E, Weil D, Raviglione M, Lönnroth K. Financial burden for tuberculosis patients in low-and middle-income countries: a systematic review. Eur Respir J. 2014;43(6):1763–1775.

[4] Subbaraman R, Nathavitharana RR, Satyanarayana S, Pai M, Thomas BE, Chadha VK, et al. The Tuberculosis Cascade of Care in India's Public Sector: A Systematic Review and Meta-analysis. PLoS Med. 2016 Oct;13(10):e1002149.

[5] Alsdurf H, Hill PC, Matteelli A, Getahun H, Menzies D. The cascade of care in diagnosis and treatment of latent tuberculosis infection: a systematic review and meta-analysis. Lancet Infect Dis. 2016 Nov;16(11):1269–1278.

[6] Chin DP, Hanson CL. Finding the Missing Tuberculosis Patients. J Infect Dis. 2017 Nov;216(suppl_7):S675–S678.

[7] Hanson CL, Osberg M, Brown J, Durham G, Chin DP. Conducting Patient-Pathway Analysis to Inform Programming of Tuberculosis Services: Methods. J Infect Dis. 2017 Nov;216(suppl_7):S679–S685.

[8] Naidoo P, Theron G, Rangaka MX, Chihota VN, Vaughan L, Brey ZO, et al. The South African Tuberculosis Care Cascade: Estimated Losses and Methodological Challenges. J Infect Dis. 2017 Nov;216(suppl_7):S702–S713.

[9] Fekadu L, Hanson C, Osberg M, Makayova J, Mingkwan P, Chin D. Increasing Access to Tuberculosis Services in Ethiopia: Findings From a Patient-Pathway Analysis. J Infect Dis. 2017 Nov;216(suppl_7):S696–S701.

[10] Masini E, Hanson C, Ogoro J, Brown J, Ngari F, Mingkwan P, et al. Using Patient-Pathway Analysis to Inform a Differentiated Program Response to Tuberculosis: The Case of Kenya. J Infect Dis. 2017 Nov;216(suppl_7):S714–S723.

[11] Fatima R, Haq MU, Yaqoob A, Mahmood N, Ahmad KL, Osberg M, et al. Delivering Patient-Centered Care in a Fragile State: Using Patient-Pathway Analysis to Understand Tuberculosis-Related Care Seeking in Pakistan. J Infect Dis. 2017 Nov;216(suppl_7):S733–S739.

[12] Garfin C, Mantala M, Yadav R, Hanson CL, Osberg M, Hymoff A, et al. Using Patient Pathway Analysis to Design Patient-centered Referral Networks for Diagnosis and Treatment of Tuberculosis: The Case of the Philippines. J Infect Dis. 2017 Nov;216(suppl_7):S740–S747.

[13] Hanson C, Osberg M, Brown J, Durham G, Chin DP. Finding the Missing Patients With Tuberculosis: Lessons Learned From Patient-Pathway Analyses in 5 Countries. J Infect Dis. 2017 Nov;216(suppl_7):S686–S695.

[14] Reid MJA, Goosby E. Patient-Centered Tuberculosis Programs Are Necessary to End the Epidemic. J Infect Dis. 2017 Nov;216(suppl_7):S673–S674.

[15] National Health Insurance Research Database, Taiwan;.

[16] Chen CC, Chiang CY, Pan SC, Wang JY, Lin HH. Health system delay among patients with tuberculosis in Taiwan: 2003-2010. BMC Infect Dis. 2015 Nov;15:491.

[17] Centers for Disease Control, Ministry of Health and Welfare, R O C (Taiwan). Taiwan Guidelines for TB Diagnosis & Treatment (6E). Centers for Disease Control, Ministry of Health and Welfare, R.O.C.(Taiwan); 2017.

[18] World Health Organization. Implementing tuberculosis diagnostics: policy framework. World Health Organization; 2015.

[19] Lorenz MO. Methods of Measuring the Concentration of Wealth. Publications of the American Statistical Association. 1905 Jun;9(70):209–219.

[20] Inghammar M, Ekbom A, Engström G, Ljungberg B, Romanus V, Löfdahl CG, et al. COPD and the risk of tuberculosis–a population-based cohort study. PLoS One. 2010 Apr;5(4):e10138.

[21] Byrne AL, Marais BJ, Mitnick CD, Lecca L, Marks GB. Tuberculosis and chronic respiratory disease: a systematic review. Int J Infect Dis. 2015 Mar;32:138–146.

[22] Yakar HI, Gunen H, Pehlivan E, Aydogan S. The role of tuberculosis in COPD. Int J Chron Obstruct Pulmon Dis. 2017 Jan;12:323–329.

[23] Lee JJ, Wu RL, Lee YS, Wu YC, Chiang CY. Treatment Outcome of Pulmonary Tuberculosis in Eastern Taiwan—Experience at a Medical Center. J Formos Med Assoc. 2007 Jan;106(1):25–30.

[24] Sreeramareddy CT, Panduru KV, Menten J, Van den Ende J. Time delays in diagnosis of pulmonary tuberculosis: a systematic review of literature. BMC Infect Dis. 2009 Jun;9:91.

[25] Horton KC, MacPherson P, Houben RMGJ, White RG, Corbett EL. Sex Differences in Tuberculosis Burden and Notifications in Low- and Middle-Income Countries: A Systematic Review and Meta-analysis. PLoS Med. 2016 Sep;13(9):e1002119.

[26] Borgdorff MW. New measurable indicator for tuberculosis case detection. Emerg Infect Dis. 2004 Sep;10(9):1523–1528.

[27] Davies MA, Tsondai P, Tiffin N, Eley B, Rabie H, Euvrard J, et al. Where do HIV-infected adolescents go after transfer? - Tracking transition/transfer of HIV-infected adolescents using linkage of cohort data to a health information system platform. J Int AIDS Soc. 2017 May;20(Suppl 3):21668.

Chapter 7

Socioeconomic determinants, Economic outcomes, and care-seeking pathways

7.1 Introduction

Continuing the analysis of the care-seeking pathways from Chapter 6, this chapter explores their linkages to upstream socioeconomic determinants at individual and regional levels and downstream economic burdens on the patients during care-seeking. Explicitly, for socioeconomic determinants, I disassemble the care-seeking pathways and link them to various socioeconomic status (SES) variables through regression analyses matched to the different data types. For economic burdens, I used the Lorenz curves and Gini coefficients to measure the inequality of costs incurred. The messages from this chapter will feed into the modelling parametrisation in the following chapters.

I structure this chapter as following sections. Section 7.2 introduces the data and definitions used in this chapter. Section 7.3 includes three subsections: Subsection 7.3.1 investigates the care-seeking before the first TB-related procedure prescribed; Subsection 7.3.2 investigates the care-seeking during evaluation (before treatment start); Subsection 7.3.3 investigates treatment outcomes. The last section is the discussion.

7.2 Methods

7.2.1 Data

I used the 7,255 care-seeking pathways extracted in Chapter 6 as the core data. In the following sections, I extracted the features mapping to different stages of care-seeking. Apart from the care-seeking pathways, I used the basic information and socioeconomic attributes in the National Health Insurance Research Database (NHIRD) [1] as covariates. Also, I linked to the national statistics of Taiwan for socioeconomic factors at regional level. The national statistics were published by the Taiwan officials and free to access on the internet.

7.2.2 Covariates

Individual SES The Nation Health Insurance programme (NHI) determines the insurance fee by considering salary income, working status, family structure, and vulnerability. For people with a paid job, they pay the insurance fee directly; for an unemployed with working family members, one of their working family members is in charge of paying the insurance fee; the fee of those who in low-income households, living in welfare institutions, or the retired without working family members are funded by the government. For employed people (including self-employed), the income level affects the insurance fee as well. Since the income from the NHIRD might have different validities for different occupations, I adjusted the systematic bias addressed by Lien 2011 [2] to sustain the validity.

Explicitly, the individual SES variables are as follows.

Employment: categorical variable with levels of unemployed (reference), employed with income lower than the median among the TB patients, and otherwise.

Subsidised: receiving a subsidy for low-income household, binary variable.

Vulnerable: unemployed adults without dependants and living in a welfare institution, binary variable.

Regional SES I used the residential area from the NHIRD to link to the regional statistics. I included: population density (crowding), ageing (regional burden), hospital density (healthcare accessibility), and proportion of population deprived (regional poverty). I transformed fractional indices into natural logarithmic scales before analysis. For each variable, I used the finest scale for the statistic (from

district to county). I interpolated or extrapolated to nearest data points if data missed at any time point. Table 7.1 lists the details of the regional SES indices.

Table 7.1: Regional SES indices

Variable	Data type	Level	Rescale	Definition
Offshore	Boolean	District (Town)	In an offshore island of Taiwan	
PopDen	Ratio	City (County)	standardised	Population density, residents per square kilometre
HospDen	Ratio	City (County)	*log*, standardised	Hospital density, hospitals per square kilometre
Ageing	Ratio	District (Town)	*log*, standardised	The ratio of people above 65 and under 15 years old
Poverty	Proportion	District (Town)	*log*, standardised	Low-income households among all households

log: taking natural logarithm Standardisation: centering the means to zero and rescaling the variances to 1

Other covariates I also considering the comorbidity of chronic lung conditions (CLCs), diabetes mellitus (DM), and human immunodeficiency viruses infected (HIV). The definitions of these comorbidity can be found in Appendix E. I considered if a pathways overlapped with 2003 as a covariate because there was a pandemic of Severe acute respiratory syndrome (SARS), which affected care-seeking related to all respiratory illness.

7.2.3 Regression models

Throughout this chapter, the features of care-seeking pathways include binary, categorical, and time-to-event (delay) data. I employed logistic regression for binary variables, multinomial logistic regression for categorical variables, and the Cox Proportional Hazard models for time-to-event variables. As for the missing values in the fields, I assumed they were missing completely at random and dealt with them by deleting the pathways.

7.2.4 Equality measurements

I used the Lorenz curve [3] and Gini coefficient [4] to investigate the inequality in the costs during care-seeking before treatment. The costs considered included the healthcare costs that occurred to the healthcare system, the out-of-pocket payment (a copayment term to the NIH) from patients, and the number of healthcare visits as an index of time cost.

The Lorenz curve is a common approach in economics to measure inequality in income and cost. The curve graphs the cumulative percentage of overall costs, given a percentage of individuals from whom with the lowest cost to higher and is formulated as follows. First, sort all care-seeking pathway increasingly by the

value of a cost. Second, calculate the cumulative cost for every care-seeking pathway from the first. Last, rescale the cumulative cost to percentages of the overall cost and plot them against the respective percentages of the pathways. A Lorenz curve is with a diagonal line from (zero percentage of pathways, zero cost) if it represents perfect equality. Within the triangle area under the perfect equality line, the Gini coefficient is defined as the area between the perfect equality line and the Lorenz curve. The Gini coefficient ranges from zero to one, and a higher Gini coefficient indicates high inequality.

In addition, I added measures of SES on the Lorenz curves to enhance the information conveyed. I summarised SES factors by moving average with a width equal to 10% of pathways. For example, the value for the 50th percentile of the care-seeking pathway is the average of the pathways from 45th to 55th percentiles.

7.3 Results

There were 7,255 care-seeking pathways extracted in Chapter 6. However, 111 care-seeking had missing residential area, so only 7,144 were eligible for the following covariate analysis. Tables 7.2 summarised the number of pathways for the analysis in each stage. For the analyses of initial care-seeking (Table 7.3) and treatment outcomes (Table 7.6), the 7,144 care-seeking pathways were included. 3,572 care-seeking pathways with Evaluation Delay were used in the analysis before evaluation and 5,241 care-seeking pathways with Diagnosis Delay were used in the analysis during evaluation (before treatment start).

Table 7.2: Summary of sample size in each stage

		Excluded pathways with missing regions		
	All pathways	All pathways	Pathways with Evaluation Delay	Pathways with Diagnosis Delay
Variable				
All	7255(100%)	7144(100%)	3572(100%)	5241(100%)
Age				
<15	36(0.5%)	36(0.5%)	23(0.6%)	20(0.4%)
$15-34$	875(12.1%)	857(12%)	423(11.8%)	554(10.6%)
$35-64$	2683(37%)	2638(36.9%)	1253(35.1%)	1831(34.9%)
≥ 65	3661(50.5%)	3613(50.6%)	1873(52.4%)	2836(54.1%)
Sex				
Female	2158(29.7%)	2124(29.7%)	1168(32.7%)	1612(30.8%)
Male	5097(70.3%)	5020(70.3%)	2404(67.3%)	3629(69.2%)
Employment				
Unemployed	3696(50.9%)	3631(50.8%)	1719(48.1%)	2644(50.4%)
Lower income	2772(38.2%)	2743(38.4%)	1458(40.8%)	2024(38.6%)
Higher income	787(10.8%)	770(10.8%)	395(11.1%)	573(10.9%)
Vulnerability				
Subsidised	217(3%)	217(3%)	99(2.8%)	130(2.5%)
Vulnerable	1773(24.4%)	1742(24.4%)	771(21.6%)	1249(23.8%)
Regional Det.				
Offshore	29(0.4%)	29(0.4%)	14(0.4%)	26(0.5%)
Hospital Level				
A	1192(16.4%)	1180(16.5%)	558(15.6%)	871(16.6%)
B	1191(16.4%)	1170(16.4%)	547(15.3%)	850(16.2%)
C	30(0.4%)	29(0.4%)	9(0.3%)	19(0.4%)
D				
Comorbidity	3104(42.8%)	3065(42.9%)	2732(76.5%)	2280(43.5%)
CLC	1275(17.6%)	1247(17.5%)	412(11.5%)	919(17.5%)
DM	1644(22.7%)	1610(22.5%)	253(7.1%)	1115(21.3%)
HIV	1232(17%)	1222(17.1%)	175(4.9%)	927(17.7%)
Pathway overlapps 2003	2110(29.1%)	2090(29.3%)	1179(33%)	1486(28.4%)

Lower income is the employment with income below the median among patients, otherwise, higher income

HospDen: hospital density, PopDen: population density, Poverty: percentage of unprivileged population, CLC: chronic lung condition, DM: diabetes mellitus, HIV: human immunodeficiency virus

Hospital Level is level at initial care-seeking

Comorbidities and Pathway overlaps 2003 are binary variables

7.3.1 From initial care-seeking to initial evaluation

I first analysed the hospitals at initial care-seeking by levels: A for primary care/ general practice; B for regional hospitals with inpatient capacity; C for larger district hospitals; D for large hospitals with a range of specialists. Table 7.3 demonstrates the results of the covariates analysis with a multinomial logistic regression model of the hospital accesses at initial care-seeking. Level A was the reference for Level B, C, and D. The odds ratio were based on the multinomial odds. For example, the OR in the column of Level B and the row of males was the odds of starting care-seeking at Level B against Level A for males versus odds of starting care-seeking at Level B against Level A for females while holding the other predictors at fixed values. In the binary covariate by definition, the OR was the ratio of $pr(level = B|sex = males, x = X)/pr(level = A|sex = males, x = X)$ and $pr(level = B|sex = females, x = X)/pr(level = A|sex = females, x = X)$, where the x is the set of predictors other than sex fixed at X and pr(.) is probability function. The care-seeking pathways for patients males, and with chronic lung conditions (CLCs) were more likely to start at Level B and C hospitals compared with Level A. The pathways for patients with paid jobs were less likely to initiate care-seeking at Level B and C hospitals compared, or more likely to start at Level A and D alternatively. The vulnerable and subsidised patients have higher probabilities of starting with Level B. However, the subsidised hardly initiated at Level D hospitals (odds ratio (OR) 0.55 (95% confidence interval (CI): 0.32 - 0.94).

After initial care-seeking, I extracted 1) whether the patients received TB evaluations at initial care-seeking (no evaluation delay), and 2) the duration between initial care-seeking and initial evaluation (evaluation delay). Tables 7.4 shows their associations with the covariates. In general, half of the care-seeking did not have evaluation delay and the diagnosis delay was 40 days in median. Conditioning on the variables except for the SES factors, care-pathways started initial care-seeking at Level A were more likely to have no evaluation delays. The employment and vulnerability did not show significant associations with evaluation delays and its existence. The patients who were living in areas with higher population density were more likely to receive TB evaluations at initial care-seeking (OR: 0.93, 95% CI: 0.88-0.98).

Figure 7.1 layouts the distributions of costs by the Lorenz curves and Gini coefficients. Among the healthcare system, out-of-pocket, and time costs, the time cost was the most homogenous (Gini: 0.63). In both the healthcare system and out-of-pocket costs, about 90% of overall costs occurred among the top quartile of care-seeking pathways. Considering the individual SES, out-of-pocket expenses were rarely incurred by the subsidised (green line) and the vulnerable (blue line).

Table 7.3: Multinomial Logistic regression analysis, hospital levels at initial care-seeking

	Hospital Level at Initial Care-Seeking (Reference = A. primary care/GP)		
	B. Regional	C. District	D. Medical Centre
Baseline	Mean	Mean	Mean
Total	18%	23%	17%
Variable	OR (95% CI)	OR (95% CI)	OR (95% CI)
Age			
<15	0.000 (0.000, 0.000)	0.631 (0.265, 1.502)	0.476 (0.138, 1.642)
$15-34$	Reference	Reference	Reference
$35-64$	1.148 (0.905, 1.457)	1.166 (0.945, 1.440)	1.096 (0.873, 1.376)
≥ 65	1.242 (0.981, 1.572)	1.128 (0.914, 1.392)	1.271 (1.007, 1.603)
Sex			
Female	Reference	Reference	Reference
Male	1.305 (1.117, 1.524)	1.258 (1.096, 1.444)	1.041 (0.895, 1.210)
Employment			
Unemployed	Reference	Reference	Reference
Lower income	0.797 (0.665, 0.956)	0.835 (0.712, 0.980)	0.923 (0.768, 1.109)
Higher income	0.738 (0.559, 0.973)	0.638 (0.499, 0.815)	1.292 (1.015, 1.645)
Vulnerability			
Subsidised	1.732 (1.204, 2.493)	1.202 (0.814, 1.776)	0.547 (0.319, 0.936)
Vulnerable	1.630 (1.339, 1.984)	1.159 (0.963, 1.394)	1.536 (1.258, 1.876)
Regional Det.			
Offshore	1.812 (0.752, 4.366)	0.517 (0.146, 1.836)	1.106 (0.356, 3.435)
Ageing	1.001 (0.998, 1.003)	1.001 (0.999, 1.004)	1.001 (0.999, 1.004)
Log(HospDen)	0.555 (0.360, 0.855)	1.189 (0.810, 1.746)	1.612 (1.058, 2.456)
Log(PopDen)	1.013 (0.961, 1.067)	0.972 (0.928, 1.019)	1.223 (1.157, 1.292)
Log(Poverty)	0.914 (0.806, 1.037)	0.804 (0.716, 0.903)	1.029 (0.898, 1.178)
Comorbidity			
CLC	1.651 (1.376, 1.981)	1.509 (1.273, 1.789)	1.694 (1.407, 2.039)
DM	1.041 (0.867, 1.250)	0.950 (0.803, 1.125)	1.061 (0.883, 1.276)
HIV	0.672 (0.136, 3.316)	1.330 (0.415, 4.263)	5.707 (2.238, 14.557)
Pathway overlapps 2003	1.153 (0.998, 1.332)	0.693 (0.602, 0.797)	0.827 (0.710, 0.964)

OR: odds ratio, CI: confidence interval, LR-test: likelihood-ratio test

Lower income is the employment with income below the median among patients, otherwise, higher income

HospDen: hospital density, PopDen: population density, Poverty: percentage of unprivileged population, CLC: chronic lung condition, DM: diabetes mellitus, HIV: human immunodeficiency virus

Comorbidities and Pathway overlaps 2003 are binary variables, applying 'None' as the reference groups.

ORs were adjusted ORs calculated from the results of **multinomial logistic regressions**.

Table 7.4: Regression analyses, before evaluation

	From Initial Care-Seeking to Initial Evaluation	
	Zero Evauation Delay	Evauation Delay
Baseline	Mean (95% CI)	Median (IQR)
Total	50.1% (49.0%, 51.3%)	40 (13, 88)
Variable	OR (95% CI)	HR (95% CI)
Age		
<15	1.102 (0.421, 2.889)	1.072 (0.701, 1.638)
$15-34$	Reference	Reference
$35-64$	0.938 (0.746, 1.180)	0.941 (0.840, 1.055)
≥ 65	0.526 (0.418, 0.662)	0.849 (0.758, 0.950)
Sex		
Female	Reference	Reference
Male	1.367 (1.180, 1.585)	1.038 (0.966, 1.116)
Employment		
Unemployed	Reference	Reference
Lower income	0.906 (0.761, 1.079)	0.947 (0.868, 1.033)
Higher income	0.789 (0.610, 1.020)	1.047 (0.923, 1.189)
Vulnerability		
Subsidised	0.863 (0.592, 1.259)	1.147 (0.925, 1.421)
Vulnerable	0.973 (0.804, 1.179)	0.996 (0.898, 1.104)
Regional Det.		
Offshore	1.338 (0.483, 3.704)	1.156 (0.680, 1.963)
Ageing	1.000 (0.997, 1.002)	1.001 (0.999, 1.002)
Log(HospDen)	0.957 (0.635, 1.444)	0.874 (0.706, 1.083)
Log(PopDen)	0.929 (0.883, 0.978)	1.019 (0.992, 1.046)
Log(Poverty)	1.074 (0.948, 1.216)	1.034 (0.970, 1.101)
Comorbidity		
CLC	0.677 (0.569, 0.806)	0.989 (0.899, 1.089)
DM	1.292 (1.079, 1.546)	1.138 (1.036, 1.249)
HIV	0.825 (0.267, 2.549)	1.321 (0.684, 2.551)
Hospital Level		
A	Reference	Reference
B	19.172 (16.126, 22.793)	1.161 (1.044, 1.291)
C	49.485 (41.188, 59.454)	1.134 (0.994, 1.293)
D	59.576 (48.407, 73.322)	0.870 (0.743, 1.018)
Pathway overlapps 2003	0.662 (0.573, 0.766)	0.695 (0.646, 0.747)

IQR: interquartile range, OR: odds ratio, CI: confidence interval, LR-test: likelihood-ratio test
Lower income is the employment with income below the median among patients, otherwise, higher income
HospDen: hospital density, PopDen: population density, Poverty: unprivileged population, CLC: chronic lung condition, DM: diabetes mellitus, HIV: human immunodeficiency virus
Comorbidities and Pathway overlaps 2003 are binary variables, applying 'None' as the reference groups.

The pathways for patients with paid jobs were more likely to need higher numbers of visits to receive their first evaluations.

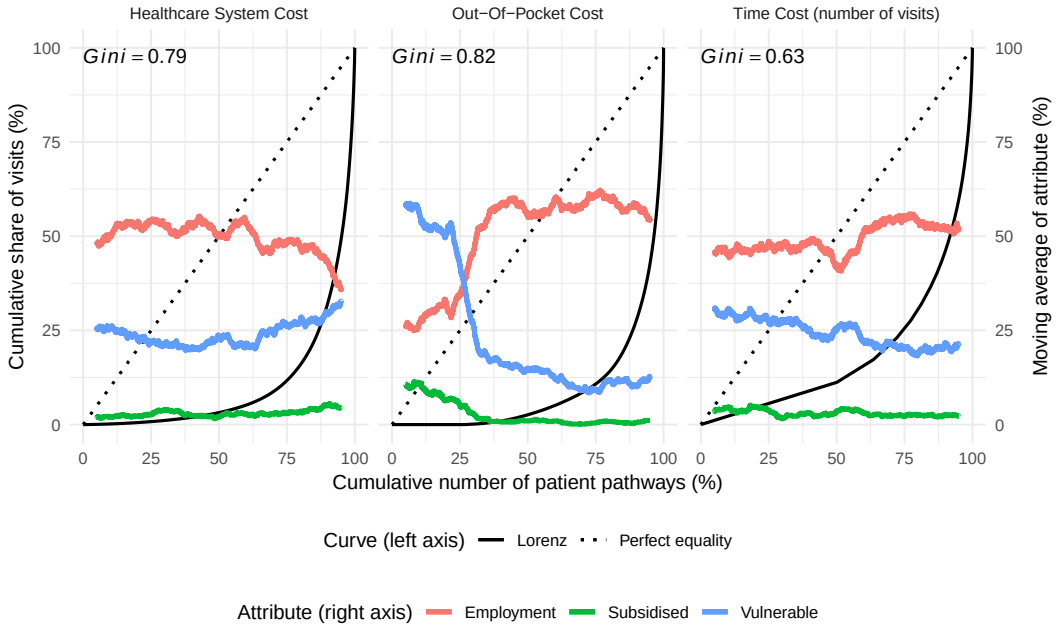


Figure 7.1: Lorenz curves and Gini coefficients of costs before evaluation

7.3.2 From initial evaluation to treatment start

Between initial evaluation and treatment start, I again extracted whether the patients initiated TB treatment at initial evaluation (no diagnosis delay), and the duration between initial evaluation and treatment start (diagnosis delay). According to the analysis in Chapter 6 I extracted if the evaluation process being interrupted (IE) and used IE to stratify the diagnosis delay. Apart from the covariates mentioned in the previous part, I included the hospital levels where the patients received first TB evaluations.

Tables 7.5 shows the results of covariate analysis. Overall, 26% of the care-seeking did not have diagnosis delay, 21% with IE, and the diagnosis delays with and without IE were 238 and 21 days in median respectively. The pathways for older patients were more likely to have longer evaluation periods: for the aged 65 and above, lower probability of zero diagnosis delay (OR: 0.43, 95% CI: 0.36-0.52), higher IE (OR: 3.87, 95% CI: 2.76-5.43). and lower rates to treatment start with and without IE (HR: 0.58, 95% CI: 0.42-0.79, and HR: 0.64, 95% CI: 0.58-0.71 respectively). Starting TB evaluations at Level C and Level D hospitals were associated with lower probabilities of zero diagnosis delay and longer delays without IE.

Table 7.5: Regression analyses, during evaluation

		From Initial Evaluation to Treatment Start		
	Zero Diagnosis Delay	IE	Diagnosis Delay without IE	Diagnosis Delay with IE
Baseline	Mean (95% CI)	Mean (95% CI)	Median (IQR)	Median (IQR)
Total	26.6% (25.6%, 27.7%)	21.1% (20.0%, 22.3%)	21 (7, 47)	238 (128, 483)
Variable	OR (95% CI)	OR (95% CI)	HR (95% CI)	HR (95% CI)
Age				
<15	1.233 (0.615, 2.471)	2.246 (0.690, 7.315)	1.436 (0.869, 2.372)	1.048 (0.359, 3.059)
$15-34$	Reference	Reference	Reference	Reference
$35-64$	0.727 (0.612, 0.862)	2.259 (1.600, 3.188)	0.780 (0.704, 0.865)	0.689 (0.495, 0.960)
≥ 65	0.433 (0.363, 0.516)	3.874 (2.764, 5.429)	0.646 (0.582, 0.718)	0.578 (0.420, 0.794)
Sex				
Female	Reference	Reference	Reference	Reference
Male	1.291 (1.141, 1.461)	0.966 (0.825, 1.131)	1.009 (0.942, 1.080)	1.013 (0.882, 1.164)
Employment				
Unemployed	Reference	Reference	Reference	Reference
Lower income	0.935 (0.811, 1.079)	0.975 (0.811, 1.173)	1.011 (0.931, 1.097)	1.006 (0.859, 1.178)
Higher income	0.746 (0.608, 0.916)	0.694 (0.501, 0.961)	0.934 (0.835, 1.046)	0.992 (0.733, 1.342)
Vulnerability				
Subsidised	1.625 (1.198, 2.203)	1.486 (0.980, 2.252)	0.827 (0.661, 1.035)	0.965 (0.691, 1.348)
Vulnerable	0.973 (0.829, 1.144)	1.017 (0.831, 1.244)	0.996 (0.908, 1.093)	0.852 (0.716, 1.013)
Regional Det.				
Offshore	0.352 (0.105, 1.185)	0.776 (0.253, 2.384)	1.182 (0.772, 1.809)	0.691 (0.256, 1.864)
Ageing	0.999 (0.997, 1.001)	0.998 (0.995, 1.000)	1.001 (1.000, 1.002)	1.002 (0.999, 1.004)
Log(HospDen)	0.972 (0.691, 1.365)	0.918 (0.591, 1.428)	1.028 (0.849, 1.245)	1.144 (0.772, 1.696)
Log(PopDen)	1.003 (0.963, 1.045)	0.968 (0.916, 1.023)	1.011 (0.986, 1.037)	1.028 (0.979, 1.079)
Log(Poverty)	1.179 (1.065, 1.305)	1.084 (0.950, 1.236)	1.012 (0.954, 1.075)	1.069 (0.950, 1.203)
Comorbidity				
CLC	1.152 (0.992, 1.338)	1.991 (1.673, 2.369)	0.740 (0.676, 0.809)	0.758 (0.654, 0.877)
DM	1.187 (1.025, 1.375)	0.639 (0.520, 0.784)	1.001 (0.921, 1.087)	1.311 (1.088, 1.580)
HIV	1.115 (0.510, 2.439)	2.113 (0.658, 6.783)	0.766 (0.459, 1.277)	1.029 (0.378, 2.802)
Hospital Level				
A	Reference	Reference	Reference	Reference
B	0.452 (0.374, 0.548)	0.966 (0.733, 1.272)	0.932 (0.820, 1.060)	0.973 (0.772, 1.227)
C	0.595 (0.499, 0.710)	0.879 (0.672, 1.149)	0.788 (0.698, 0.891)	1.058 (0.841, 1.331)
D	0.430 (0.355, 0.519)	0.773 (0.585, 1.021)	0.799 (0.705, 0.906)	0.932 (0.734, 1.184)
Pathway overlapps 2003	1.182 (1.050, 1.331)	2.620 (2.263, 3.034)	0.993 (0.924, 1.068)	0.476 (0.416, 0.544)

IE: Interrupted Evaluation, IQR: interquartile range, OR: odds ratio, CI: confidence interval, LR-test: likelihood-ratio test

Lower income is the employment with income below the median among patients, otherwise, higher income

HospDen: hospital density, PopDen: population density, Poverty: unprivileged population, CLC: chronic lung condition, DM: diabetes mellitus, HIV: human immunodeficiency virus

Comorbidities and Pathway overlaps 2003 are binary variables, applying 'None' as the reference groups.

ORs were adjusted ORs calculated from the results of **multivariate logistic regressions**.

Figure 7.2 demonstrates the cost distributions before treatment after initial evaluation. The general patterns were similar to Figure 7.1, but the equalities were higher in the healthcare system cost (Gini: 0.72) but lower in the time cost (Gini: 0.70). Along with the individual SES, the proportions the vulnerable patients were more frequently with higher ranks of the healthcare system cost. In the time cost, the individual SES factors were homogeneously distributed across the percentages of the ranked care-seeking pathways.

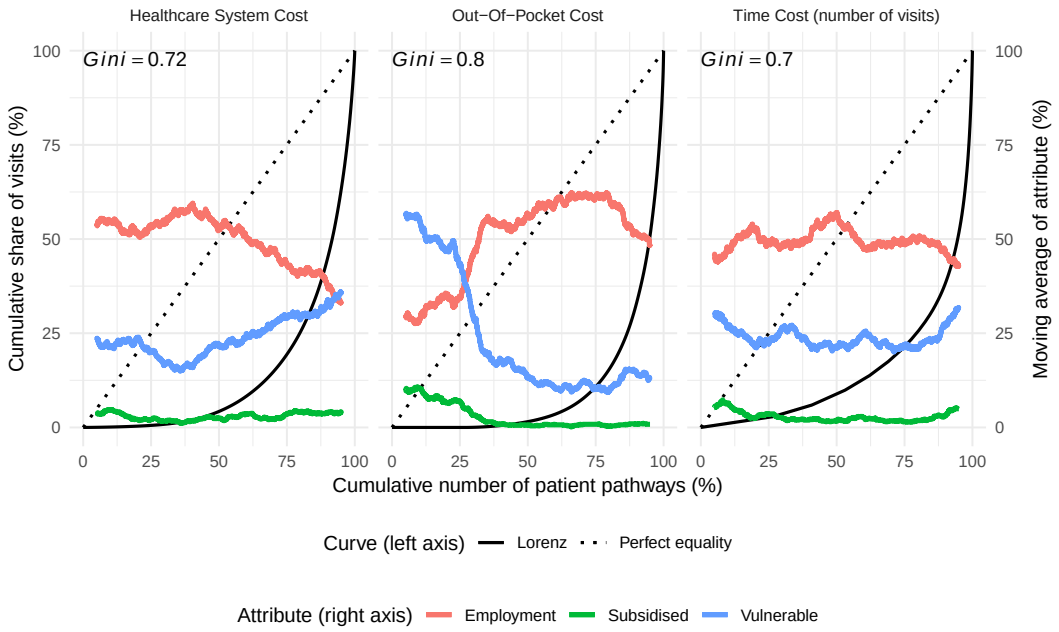


Figure 7.2: Lorenz curves and Gini coefficients of costs during evaluation

7.3.3 Treatment outcomes

I analysed the risk factors of unfavourable outcomes of all-cause death and loss to follow-up (LTFU). For the censored cases (ended with the end of the data cohort), I deleted the data entries by assuming they were missing completely at random. The definitions of treatment outcomes were the same as Chapter 6.

Tables 7.6 shows the results of multinomial logistic regression of treatment outcomes using treatment completed as the reference group. Except for the care-seeking pathway with censored treatment outcome, 68% of the pathways completed their treatment, 14% died, 18% were LTFU. Regarding all-cause death, the odds were higher for pathways for TB patients aged 35-64 (OR: 6.52, 95% CI: 3.37-12.63), aged 65 and above (OR: 21.96, 95% CI: 11.44-42.17), were male (OR: 1.25, 95% CI: 1.05-1.48), and with comorbidities. For SES factors, the employed with in-

come higher than the median was a protective factor (OR: 0.41, 95% CI: 0.26-0.64) while living in subsidised households associated with a higher risk of death (OR: 1.73, 95% CI: 1.18-2.54). Regarding LTFU, the odds were higher for pathways for TB patients aged 35-64 (OR: 1.34, 95% CI: 1.08-1.65), were the vulnerable group (OR: 1.34, 95% CI: 1.12-1.61), and with chronic lung condition (OR: 1.66, 95% CI: 1.40-1.95). The employed with income higher than the median was a protective factor (OR: 0.66, 95% CI: 0.51-0.84).

Table 7.6: Multinomial Logistic regression analysis, treatment outcome

	Treatment Outcome (Reference = Treatment completed)	
	Dead, all cause	LTFU
Baseline	Mean	Mean
Total	14%	18%
Variable	OR (95% CI)	OR (95% CI)
Age		
<15	N.A.	1.049 (0.446, 2.469)
$15-34$	Reference	Reference
$35-64$	6.521 (3.365, 12.639)	1.338 (1.084, 1.651)
≥ 65	21.960 (11.436, 42.169)	0.994 (0.802, 1.232)
Sex		
Female	Reference	Reference
Male	1.248 (1.053, 1.479)	1.071 (0.929, 1.234)
Employment		
Unemployed	Reference	Reference
Lower income	0.866 (0.716, 1.048)	0.956 (0.808, 1.131)
Higher income	0.412 (0.264, 0.644)	0.657 (0.513, 0.843)
Vulnerability		
Subsidised	1.728 (1.175, 2.542)	0.908 (0.622, 1.327)
Vulnerable	1.185 (0.974, 1.443)	1.343 (1.116, 1.615)
Regional Det.		
Offshore	0.636 (0.141, 2.861)	1.575 (0.652, 3.805)
Ageing	1.000 (0.997, 1.002)	0.999 (0.997, 1.002)
Log(HospDen)	0.727 (0.460, 1.148)	1.135 (0.768, 1.677)
Log(PopDen)	1.034 (0.975, 1.096)	0.989 (0.943, 1.037)
Log(Poverty)	1.038 (0.907, 1.189)	1.040 (0.924, 1.170)
Comorbidity		
CLC	1.667 (1.408, 1.974)	1.655 (1.403, 1.953)
DM	1.208 (1.010, 1.445)	1.059 (0.891, 1.259)
HIV	7.678 (2.540, 23.205)	1.542 (0.622, 3.825)

OR: odds ratio, CI: confidence interval, LR-test: likelihood-ratio test, N.A.: not available
Lower income is the employment with income below the median among patients, otherwise, higher income
HospDen: hospital density, PopDen: population density, Poverty: unprivileged population, CLC: chronic lung condition, DM: diabetes mellitus, HIV: human immunodeficiency virus
Comorbidities and Pathway overlaps 2003 are binary variables, applying 'None' as the reference groups.
ORs were adjusted ORs calculated from the results of **multinomial logistic regressions**.

7.4 Discussion

Care-seeking, including delays to treatment and treatment adherence, can be associated to socioeconomic determinants. These associations are the clues to innovating TB control strategies. However, in Chapter 2 my review found that different socioeconomic factors might act differently for different parts of care-seeking processes in different settings. Continuing on from Chapter 6, I investigated the relationships between SES and care-seeking pathways by decomposing these care-seeking pathways. From the hospital visited for initial care-seeking, delays before treatment start, to treatment outcomes, I found several links, which were relevant to supporting future strategic recommendations for TB services provision in Taiwan.

To the best of my knowledge, this study is the first study considering the system delay (the period from initial care-seeking to treatment start) as separate delays to evaluation start and treatment start. Also, I considered if the evaluation processes for pathways experienced interruptions as a critical property determining the length of system delay. The analysis of the initial hospital levels includes the care-seeking preference for patients with different characteristics. The pathways without evaluation delays or diagnosis delay correspond to the physician's immediate decision concerning the features or health status of patients. The delays represent the patients' care-seeking behaviours and health status, the time-spent for diagnostics, as well as when physicians aware of the possibility of TB during caring.

I found that social vulnerability did not always lead to the worst care-seeking processes. The low-income subsidised population preferred going to higher-level hospitals for their initial care-seeking and got TB treatment quicker. That may be attributable to the national policy in Taiwan that exempted some out-of-pocket cost to the eligible underprivileged population. Also, the results of the regional SES implies that the physicians working in more impoverished areas are more skilful at identifying TB. However, the social vulnerability associated with more unfavourable outcomes of death and LTFU. Two hypotheses need testing. First, that these patients initiated care-seeking after they were very ill. Second, that the poor living standard and other physical barriers affected the mortality or kept them from proper care. However, these hypotheses require qualitative data to address. On the other hand, the effects of employment on care-seeking act in two directions. The employed usually started care-seeking with lower-level hospitals and had a lower chance of zero diagnosis delay. On the other hand, their evaluation processes were hardly ever interrupted.

I used the Lorenz curve and the Gini coefficient for measuring the inequality of costs. Since the Lorenz curve did not consider the nature of heterogeneity among the population, I attached the moving average of SES variables to strengthen the interpretation. Thus, I observed that the social security net in Taiwan protected vulnerable groups from out-of-pocket cost, although they started care-seeking at higher-levels. However, in the number of visits before treatment initiation, the vulnerable patients suffered from time cost similar to other patients.

I found that age played a vital role in care-seeking pathways. The older adults with TB-suggestive symptoms sought care at higher-level hospitals. However, for most of the indices I assessed, they were more likely to have longer system delays even after controlling for comorbidities. The difficulty in diagnosing TB for older adults is expected as an increasing challenge in the future.

For the treatment outcome of death, this study did not consider the background death rates. For example, age and sex have been commonly linked to background mortality, older people and males usually have higher death rates. In order to understand better the TB-specific death rates, I need to characterise them by offsetting background mortality.

Furthermore, this study did not appropriately address socioeconomic factors for older adults. My SES variables from the NHIRD were mainly based on employment. For those who retired, they had no salary income by definition. An SES based on housing, the capacity of subsistence, and social segregation could be considered. Also, I suggest stratifying care-seeking behaviours by age or career stage to consider the transitioning of decision making corresponding to life stages.

Reflecting on the Theory of Planned Behaviour (TPB)[5] used to conceptualise the linkage with socioeconomic determinants, the SES factors in this study captured the components of the subjective norm (SN) and perceived behavioural control (PBC) to some extent (see Chapter 2). Employment, income, vulnerability, and the density of hospitals can be considered proxies of perceived behavioural control. They are related to apparent barriers to care-seeking. The regional poverty and regional ageing index can be considered as referring to the social context or the subjective norm. However, I did not assess the attitude toward seeking care for TB. The risk perception, self-stigma, and stigma require qualitative approaches to consider further.

The SES variables of interest might depend on each other and cause collinearity in the regression models. For example, the income level is associated with seniority in a job, which is increasing along with age increases. I used categorical variables to reduce the potential influence, such as combining income level and employment. Further analysis, using dimension reduction, could shed more

light on the effects of SES on care-seeking. Multiple Correspondence Analysis (MCA) [6], which is a dimension reduction method for categorical data, might be a solution. For example, Birch et al. [7] used MCA to summarise variables about assets before inputting them into covariate analysis of TB treatment non-adherence. However, referring back to the exact SES properties, identifying which affect care-seeking after dimension reduction is challenging. Alternatively, structural equation modelling (SEM) [8], which uses structural assumptions to summarise upstream variables with conceptual components (latent variables), could build firmer links between SES variables and care-seeking features. For example, Alegria-Flores et al. [9] used SEM to introduce a behavioural theory to investigate treatment adherence.

7.4.1 Recommendations for future studies

This chapter is a bridge connecting evidence to modelling and interventions. For modelling, the messages are 1) zero delay and lower delay do not have the same upstream determinants, and 2) the hospital level at initial care-seeking mediates the effects of employment and vulnerability on the delay to evaluation started. As for policy implications, the key messages are 1) hospital levels at initial care-seeing affects the time -to-evaluation started, and 2) those under subsidisation for low income household require support for treatment adherence.

Bibliography

[1] Institutes NHR. National Health Insurance Research Database, Taiwan; 2014.

[2] Lien HM. How to Construct Social-Economic Variables from National Health Insurance Data. Journal of Social Sciences and Philosophy. 2011;23(3):371–398.

[3] Lorenz MO. Methods of Measuring the Concentration of Wealth. Publications of the American Statistical Association. 1905 Jun;9(70):209–219.

[4] Gini C. Concentration and dependency ratios. Rivista di politica economica. 1997;87:769–792.

[5] Ajzen I. The theory of planned behavior. Organ Behav Hum Decis Process. 1991 Dec;50(2):179–211.

[6] Greenacre M, Blasius J. Multiple Correspondence Analysis and Related Methods. Chapman and Hall/CRC; 2006.

[7] Birch S, Govender V, Fried J, Eyles J, Daries V, Moshabela M, et al. Does treatment collection and observation each day keep the patient away? An analysis of the determinants of adherence among patients with Tuberculosis in South Africa. Health Policy Plan. 2016 May;31(4):454–461.

[8] Kline RB. Principles and Practice of Structural Equation Modeling, Fourth Edition. Guilford Publications; 2015.

[9] Alegria-Flores K, Weiner BJ, Wiesen CA, Lich KLH, Van Rie A, Paul JE, et al. Innovative approach to the design and evaluation of treatment adherence interventions for drug-resistant TB. Int J Tuberc Lung Dis. 2017 Nov;21(11):1160–1166.

Chapter 8

Hybrid modelling of TB dynamics with detailed care-seeking processes

8.1 Introduction

Care-seeking of TB patients plays an essential role in TB epidemiology. Especially in responding to the urgency of TB elimination and the maturity of TB treatment and diagnostics. The marginal returns on improving diagnostic sensitivity and treatment efficacy are diminishing. Instead, actions such as active case finding, that improve potential TB patients' access to TB diagnosis and treatment are receiving growing attention. In this thesis, I want to investigate the care-seeking and diagnostic processes of TB patients to find the possibility of introducing new SES-related interventions and evaluate their effectiveness. From the IPPA, I have detailed results describing care-seeking pathways. This chapter then intends to consider the transmission process, which is usually an essential component in evaluating the impact of an infection control strategy on a population level.

Agent-based modelling which models a system from every individual is suitable to link detailed individual data to population influence. However, in my setting, the overall Taiwanese population is approximately 23 million. The population size is too infeasible to perform Monte Carlo inferences, such as for statistical model fitting. Simplifying the IPPA results to aggregate levels and employing an EBM might be a solution to capture the TB dynamics in Taiwan, but it cannot implement TB control strategies concerning individual attributes and evaluate their impact on the individual level. Given 10,000 incident TB cases and an expected disease duration of half a year, the expected TB prevalent cases is 5,000 in

an endemic equilibrium. Constructing an ABM for this number has a relatively low computational cost. As I investigated in Chapter 3, hybrid modelling allowed us to switch some agents out of the main focus to EBMs and mitigate the computational burden. I therefore only need to construct an ABM of TB patients while modelling the others in an EBM.

Reviewing the previous chapters, Chapter 5 analysed the demography and TB epidemiology in Taiwan using public data. Chapter 6 did an empirical study to detail the care-seeking pathway by aligning the TB-service provision and demands in Taiwan using individual healthcare data. Chapter 4 reviewed the methods to model health behaviour, including care-seeking in ABMs. Chapter 3 reviewed the methods and applications to hybridise ABMs and EBMs in a simulation to improve performance. This chapter then, is to combine these results in order to build a simulation model.

8.1.1 Aim and objectives

The aim of this chapter is to developed and validated a hybrid model which fits the epidemiological features of TB in Taiwan and can compactly include the information extracted from the IPPA results. I will test various parametrisations as well as prior parameters and select one which is the best fit to data in the statistical sense. I will optimise the hybrid model until it can perform Monte Carlo experiments in a personal computer in manageable simulation time.

8.1.2 Overview

Figure 8.1 demonstrates an overview of the modelling scheme in this chapter. The first column layouts the data to be synthesised; the second includes the statistical model linking the data to parameters driving simulation models; the third are simulation models; the fourth column included the outputs I want to deliver using the simulation models. (1) Starting from the upper parts, I use the IPPA results to construct a state-space model of the care-seeking process, StSp. I tested various parametrisation and find a version which has the best performance in reconstructing the distribution of the system delays, from initial care-seeking to treatment start, in the IPPA summary. (2) I use a statistical model to estimate the TB-specific death rates after treatment initiated and other dropouts according to the IPPA results. The results will be an input to the following models. (3) I wrap StSp to agents and add dropout process to formulate an ABM of the care-seeking process, CSM. (4) Given the results of the age-sex distribution at notification from Chapter 5, I use CSM to construct the time-series of the initial care-seeking rates. (5) I

construct a mean-field model of TB dynamics, MFM, which is a compartmental EBM. I use the initial care-seeking rates from (4) and the notification rates by sex to calibrate MFM. (6) Start with MFM, I replace the compartments model those who seek care for active TB with CSM as the final hybrid model, HM. Also I compare the simulations of MFM and HM, ensuring the consistency. Since MFM stands the mean-fields of the system of interest, HM can render the same dynamics as MFM on average almost surely.

This chapter structures by following the modelling scheme. Section 8.2 constructs and calibrates StSp. Section 8.3 estimates the TB-specific mortality and loss to follow-up rates using the IPPA results. Section 8.4 wraps StSp into CSM and estimates the initial care-seeking rates. Section 8.5 constructs and calibrates MFM. Section 8.6 brings CSM and MFM together to build and valid HM. Section 8.7 simulates the TB epidemiological indices and economic burden of TB in Taiwan with CSM and HM. Lastly, Section 8.8 discusses the technical aspect of the modelling and the results.

I use the abbreviations of StSp, MFM, CSM, and HM to indicate the model builds with certain model parametrisations and calibrated parameters (or posterior parameter sets). Also, to be noted that "rates" in this thesis are annual rates, otherwise, stated.

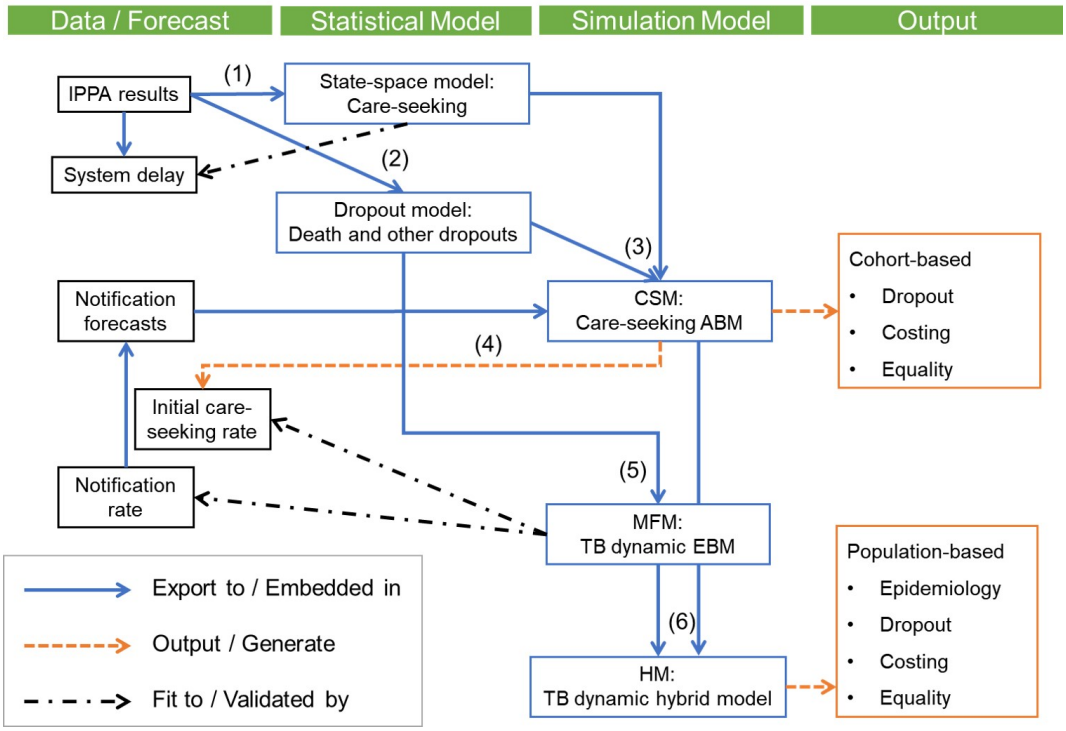


Figure 8.1: Modelling scheme

8.2 Care-seeking state-space modelling, StSp

In the IPPA, I described the care-seeking pathways with four stages according to the prescriptions of TB-related diagnostics and treatment, namely, Waiting, Evaluating, TB Detecting, and Treating Stages, from the viewpoint of patients. Also, I profiled the alignments of capacities and accessibilities of the medications from the viewpoint of healthcare provision. In this section, I intend to use state-space models to combine both to allow the model to be responsive to the potential interventions on patients' care-seeking obstacles and healthcare provision.

8.2.1 Methods

Table 8.1 lists the options for constructing StSp. I tested various state-spaces, different methods to model the skips of states, and different baseline hazard functions.

State-space

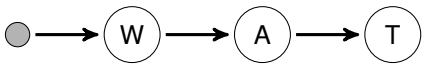
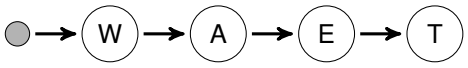
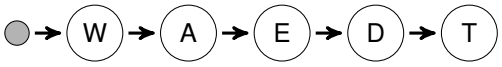
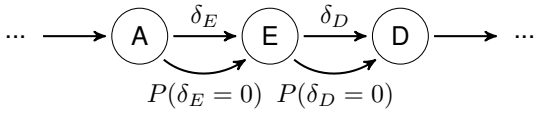
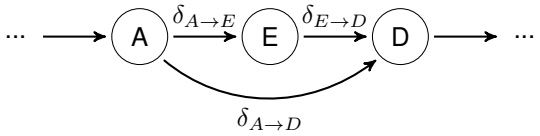
Following the four stages in the IPPA, after initiating care-seeking, a patient pathway starts with Waiting Stage; after prescribing a procedure which can possibly identify TB, the patient goes to Evaluating Stage; after a procedure which can probably identify TB, TB-Detecting Stage; After TB treatment starts, Treating Stage. Evaluating and TB Detecting Stages can be interrupted by misperception or the complexity of health status. In the state-space model, I firstly included the four stages as the basic four states. Apart from the four stages investigated in the IPPA, I introduced an additional state indicating the patients arrive at hospitals which are capable of providing TB diagnostics. Arriving at this state means a patient start having a chance to get diagnosed and to receive TB treatment.

The basic state-space in my options include all states mentioned, that is, Waiting (W), Diagnostics Available (A), Evaluating (E), TB Detecting (D), and Treating (T). Considering performance in simulation and the robustness of parameters, I explored two reduce state-spaces, WAET and WAT as Table 8.1 shows. I reduced the state-space by turning some states invisible but not alter the definitions of the states. For example, from WAEDT to WAET, I added up the durations of E-D and D-T to be the duration of E-T.

Zero delays

In Chapter 6, I found that a considerable proportion of the care-seeking pathways did not have all the four stages. That is, skipping some stage before reaching Treat-

Table 8.1: State-space modelling options

Group	Option	Diagram
State-space	WAT	
	WAET	
	WAEDT	
State skip	No skip	
	Zero inflated	
	Competing	

Stage: W, Waiting; A, Diagnostics available; E, Evaluating; D, TB Detecting; T, Treating
δ_x, duration in the state before x
$P(\delta_x = 0)$, probability of zero duration in the state before x
$\delta_{x \to y}$, duration of x before y

ing Stage was common in the setting. To capturing the skipping, I examined three parametrisations as the second part of Table 8.1. In the first option, I regarded the zero delays as extremely short delays and, hence, analysed them as zero in the response variables in the regression analysis. The second applies the Zero-Inflated approaches [1], which is a mixture distribution made of a Bernoulli process indicating zero value and a time-to-state distribution for the rest. The last option uses multi-state models to specify state transitions. In each state, the following states will compete for being the next. For example, in Waiting if Treating comes earlier than the others, the next will be Treating but not Diagnostics Available.

Predictors

I considered the predictors of age, sex, individual socioeconomic status (SES), regional SES, and comorbidities for simulating durations between transitions in my state-space models. These features applied the same definitions as in the SES analysis of Chapter 7. Some patients of care-seeking pathways had missing residential areas, and I used the areas of the hospitals which they initiated care-seeking instead. As for the missing values in the other field, I presume they missed completely at random and dealt with them by deleting the pathways.

Regression models

Throughout the state-space modelling, I employed the Cox Proportional Hazard models for linking the predictors to time-to-states and logistic regression for probabilities if the zero-inflated distribution applied. For the time-to-states, after estimating the coefficients of predictors, I extracted the baseline hazard functions and parametrised them. I considered the exponential and Weibull distributions to model the baseline hazard functions for each combination of state-space and skipping method.

It is to noted that, I considered the time-to-Diagnostics Available and the time-to-states after that separately. I firstly selected the optimal time-to-Diagnostics Available parametrisation according to the Akaike information criterion (AIC) before modelling the following time-to-states. For robustness of estimation, I trimmed the care-seeking pathways which had system delays over 95% quantiles.

Model selection

The main focus of this state-space modelling was to find a good parametrisation in simulating the system delays. Given the same distributions of the predictors, I simulated the care-seeking pathways using the candidate state-space models and calculated system delays. I then compared the distributions of system delays in the IPPA results and the simulated system delays by the Kullback-Leibler divergence (KL-divergence) [2]. Then, I selected a state-space model with the smallest KL-divergence.

The analysis was implemented by the procedure as follows. Given a candidate state-space model,

- Estimate the parameters of the state-space model.
- Based on the model and the parameters, use the predictors of care-seeking pathways to sample simulated care-seeking pathways (7,255 pathways).

- Calculate the 7,255 system delays of the 7,255 pathways.
- Summarise the system delays in the median value.
- Calculate the KL-divergence of the simulated system delays against empirical system delays.
- Repeat step 2 to 5 by 1,000 times, and fetch 1,000 median system delays and KL-divergences accordingly.

The analysis in this section was performed in R 3.5.1 with packages: `survival` [3] and `msm` [4].

8.2.2 Results

TB diagnostics availability

On average, 65% of patients who arrived at a hospital have TB diagnostics available at their initial care-seeking. For those who could not have TB diagnostics due to the capability, the time-to-Diagnostics Available was 31 days in the median with first and third quartiles of 10 and 68 days. Considering predictors with the Zero-inflated Cox regression, I found the Weibull distribution provided better goodness of fit than the exponential distribution.

Model selection

Table 8.2 shows the simulated system delays and KL-divergences. In general, the state-space WAET performed better than WAT and WAEDT. Most of the candidate models can simulate the mean of the median system delay around the empirical median system delay. The candidate models without state skipping usually rendered longer system delays; the simulations using the models with Zero-Inflated distributions were closer to the empirical given the same state-space and baseline hazard. Across the candidate models, the WAET state-space with the Zero-Inflated Exponential time-to-state distributions was the optimal choice, showing KL-divergence of 0.018 (95% CI: 0.011-0.025).

Table 8.2: Simulations of state-space models

State space	State skip	Baseline hazard	Med System Delay (Mean, 95% CI)	KL-divergence (Mean, 95% CI)
WAT	No skip	Exponential	53 (46, 62)	0.125 (0.115, 0.135)
	No skip	Weibull	36 (31, 42)	0.069 (0.061, 0.079)
	Zero inflated	Exponential	29 (25, 33)	0.113 (0.099, 0.126)
	Zero inflated	Weibull	30 (26, 34)	0.102 (0.088, 0.116)
WAET	No skip	Exponential	95 (81, 108)	0.519 (0.502, 0.537)
	No skip	Weibull	59 (50, 69)	0.254 (0.241, 0.269)
	Zero inflated	Exponential	49 (41, 59)	0.018 (0.011, 0.025)
	Zero inflated	Weibull	32 (27, 38)	0.045 (0.034, 0.055)
	Competing	Exponential	46 (39, 54)	0.028 (0.021, 0.035)
	Competing	Weibull	30 (25, 36)	0.091 (0.078, 0.104)
WAEDT	No skip	Exponential	126 (108, 148)	0.897 (0.875, 0.917)
	No skip	Weibull	67 (56, 81)	0.43 (0.412, 0.449)
	Zero inflated	Exponential	55 (46, 66)	0.025 (0.017, 0.033)
	Zero inflated	Weibull	31 (26, 37)	0.038 (0.028, 0.049)
	Competing	Exponential	40 (34, 47)	0.145 (0.135, 0.154)
	Competing	Weibull	24 (21, 28)	0.111 (0.101, 0.121)
			System Delay (Med, IQR)	
Data			41 (7, 111)	

Stage: W, Waiting; A, Diagnostics available; E, Evaluating; D, TB Detecting; T, Treating
Med: Median, IQR: interquartile range, OR: odds ratio, CI: confidence interval
KL-divergence: Kullback–Leibler divergence

8.3 Dropout estimation

Using the IPPA results, I seek to estimate the TB fatality rates in this section. Also, I want to estimate the rates of loss to follow-up (LTFU), which is compatible with the IPPA results, to support the care-seeking ABM in Section 8.4 as well as the mean-field model in Section 8.5.

8.3.1 Methods

Data and notations

I used the IPPA results and demographic data from the Department of Statistics, the Ministry of the Interior, Taiwan. I exacted the time-to treatment outcome and treatment outcome from the IPPA results and grouped the time-to treatment outcome by treatment outcomes. The demographic data included the mid-year population estimators and deaths by single-year ages from 2000 to 2010. The demographic data were published by the Taiwan officials and free access on the internet; the usage is licensed by the Open Government Data License:

[https://data.gov.tw/license]. The data descriptions are as follows:

agp: age-group $\in$ {<15, 15-34, 35-64, 65+}.

y: treatment outcome (completed: 0, death: 1, LTFU: 2) extracted from *IPPA* results.

t: time-to treatment outcome.

$dr_{age,sex,year}$: crude annual death rate of age, sex, $year$. Number of deaths over mid-year population.

fr_{agp}: TB-specific death rate per person-year of age group, agp

lr_{agp}: LTFU rate per person-year of age group, agp

Model

I derived the statistical model from a state-space model parametrisation. There are three types of treatment outcomes occurring: all-cause death, LTFU, and treatment completion. Treatment completion depends on the absence of the other two outcomes. I assumed death and LTFU follow Exponential distributions, which has a non time-varying hazard. The rates for death and LTFU are $(fr_{agp} + dr_{age,sex,year})$ and lr_{agp} respectively. The transition intensity matrix, R, before six month for a patient pathway is

$$\mathbf{R} = \begin{array}{c} \begin{array}{ccc} \textit{Treating} & \textit{Death} & \textit{LTFU} \end{array} \\ \begin{bmatrix} -(fr_{agp} + dr_{age,sex,year} + lr_{agp}) & fr_{agp} + dr_{age,sex,year} & lr_{agp} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \end{array} \begin{array}{c} \textit{Treating} \\ \textit{Death} \\ \textit{LTFU} \end{array} \tag{8.1}$$

Convert T to transition probability matrix P given time-to event t through $P = 1 - mexp(-tR)$, where $mexp(.)$ is matrix exponential function. Therefore, I have

$$\mathbf{P} = \begin{array}{c} \begin{matrix} \textit{Treating} & \textit{Death} & \textit{LTFU} \end{matrix} \\ \begin{bmatrix} 1-p & \frac{fr_{agp}+dr_{age,sex,year}}{fr_{agp}+dr_{age,sex,year}+lr_{agp}}p & \frac{lr_{agp}}{fr_{agp}+dr_{age,sex,year}+lr_{agp}}p \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \end{array} \begin{matrix} \textit{Treating} \\ \textit{Death} \\ \textit{LTFU} \end{matrix} \tag{8.2}$$

$$p = exp\big(-t(fr_{agp} + dr_{age,sex,year} + lr_{agp})\big)$$

and the values in the first row of P matches the probabilities of treatment outcomes at time t. Then, under a Bayesian framework, I used weak prior of uniform distributions, $U(0, 0.5)$ for fr_{agp} and lr_{agp}. The full model is

$$\begin{aligned}
&\textit{Foreach pathway } i \quad \{ \\
&\quad y_i \sim category(pr_i) \\
&\quad pr_i[0] = 1 - p_i \\
&\quad pr_i[1] = \frac{fr_{agp} + dr_{age,sex,year}}{fr_{agp} + dr_{age,sex,year} + lr_{agp}} p_i \\
&\quad pr_i[2] = \frac{lr_{agp}}{fr_{agp} + dr_{age,sex,year} + lr_{agp}} p_i \\
&\quad p_i = exp\big(-t_i(fr_{agp} + dr_{age,sex,year} + lr_{agp})\big) \\
&\} \\
&\textit{Foreach } agp \quad \{ \\
&\quad fr_{agp} \sim U(0, 0.5) \\
&\quad lr_{agp} \sim U(0, 0.5) \\
&\}
\end{aligned}$$

To be noted that, I ignored the index i for agp, age, sex, and $year$ for clarity, but they are specified for each pathway. By dropping the agp strata, I also used a simplified model to estimate the aggregated TB-specific death and LTFU rates.

I used Markov Chain Monte Carlo (MCMC) to sample to posterior distributions of fr_{agp} and lr_{agp}. I ran three chains of MCMC for the estimation procedure. The analysis in this section was performed in JAGS 4.3.0 [5] with the interface of `R2jags` [6] in R 3.5.1

Results

Considering the convergence of the posterior distributions, I discarded the first 2,000 iterations and collected every 10 iterations after that until 500 parameter sets retrieved for each chain. As Figure 8.2 shows, the traces of the posterior parameters converged and the samples drawn from the three chains were close. Table 8.3 shows the results of the estimation of TB-specific death rates and LTFU rates. Apart from the youngest group, the TB-specific death rates rose along with age. The TB patients aged 65 and above had highest TB-specific death rate of 0.19 (95% credible interval (CrI): 0.173-0.206). The TB patients aged below 15 shows a wider 95% CrI reflecting the relatively small sample size. Without age stratification, the TB-specific death rate was estimated at 0.116 (95% CrI: 0.106-0.126).

In terms of LTFU rates, apart from the oldest group, the rates were around 0.25 per person-year. The TB patients aged below 15 shows a wide CrI as well. I found that the oldest age group had the lowest LTFU rate of 0.208 (95% CrI: 0.16-0.189), showing an opposite direction to TB-specific death rate. Comparing the goodness of fit, the estimators of DIC prefers the model with age group strata.

Table 8.3: Estimators of TB-specific death rates and LTFU rates

Age	Number	TB-specific death rate Mean (95% CrI)	LTFU rate Mean (95% CrI)	DIC
<15	36	0.034 (0.002, 0.122)	0.265 (0.118, 0.449)	16904
15-34	875	0.019 (0.009, 0.031)	0.244 (0.209, 0.285)	
35-64	2683	0.071 (0.06, 0.083)	0.247 (0.228, 0.268)	
65+	3661	0.19 (0.173, 0.206)	0.174 (0.16, 0.189)	
Aggregated	7255	0.116 (0.106, 0.126)	0.208 (0.196, 0.22)	17145

LTFU: loss to follow-up, CrI: credible interval, DIC: deviance information criterion

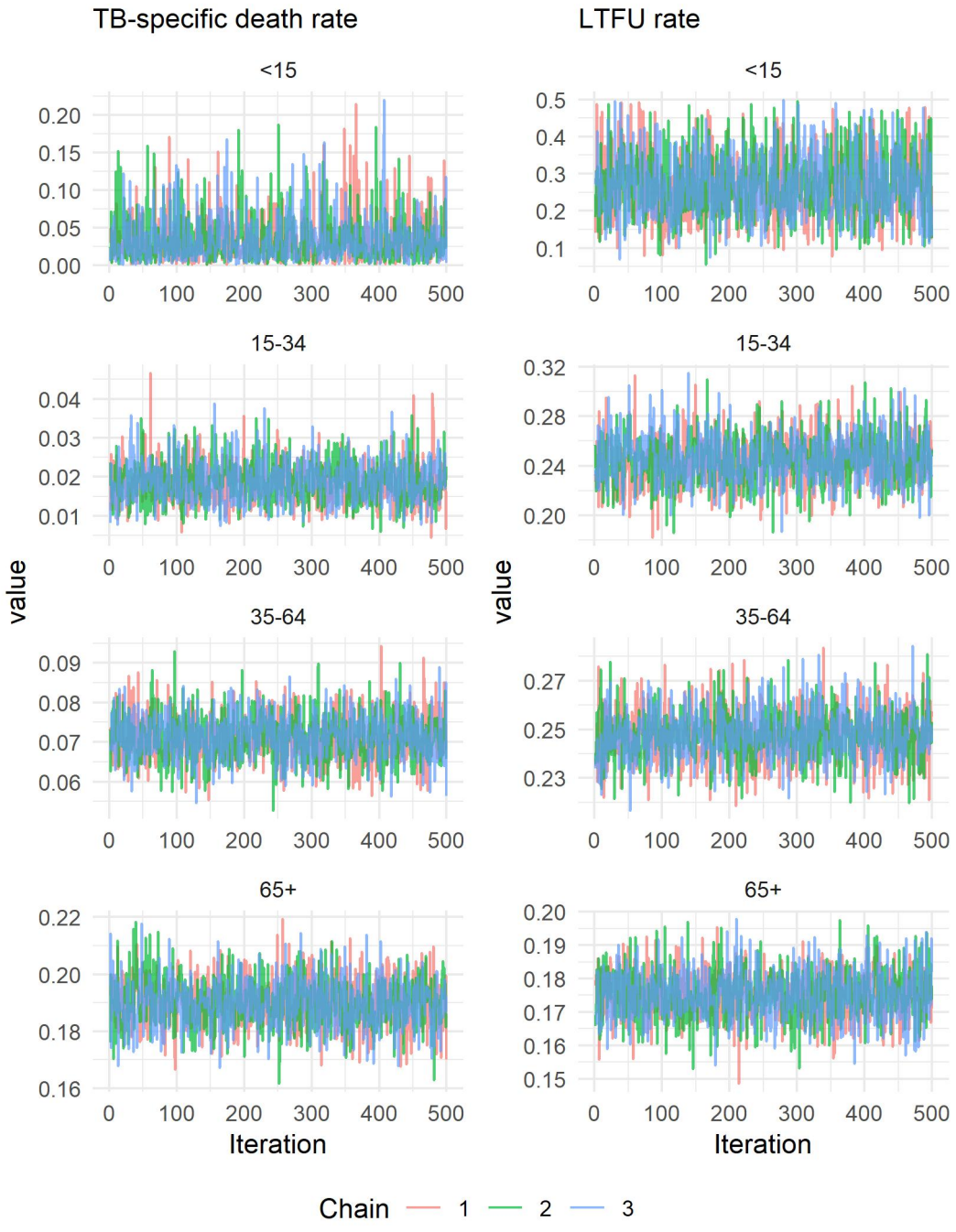


Figure 8.2: Traces of posterior TB-specific death rates and LTFU rates

8.4 Care-seeking agent-based modelling, CSM

From the previous steps, I have built a step-space model, StSp, for simulating the care-seeking pathways and estimators for dropout occurrences based on the IPPA results. Combining them, this step has the following three tasks.

- Constructing a self-contained agent-based model, CSM, which can assess the outcomes without transmission processes.
- Ensuring the relationships between the time-series of initial care-seeking (ICS) rates and notification rates and providing a likelihood function for calibrating the transmission models in the next step.
- Providing the age-sex distributions of patients at their initial care-seeking at each time point of interest in order to populate agents in the final hybrid model.

These tasks aimed to keep the uncertainty between the ICS and notification to minimal, making the model close to the empirical results and reality.

8.4.1 Methods

Using StSp from the previous section, I constructed an ABM. The agents are started with untreated active TB and at their ICS. Death, spontaneous recovery, and loss to follow-up might happen according to Section 8.3, and the agents will be deleted from the system after that. I considered ages ranged from 0 to 99 by single year, binary sexes, and years from 2005 to 2018.

In Taiwan, the TB notification recorded upon the prescription of anti-TB drugs. The notification rates should align with the rates of arriving at Treating Stage. The time difference between an initial care-seeking and its notification is the system delay, by definition. However, the initial care-seeking rates cannot be assessed by merely shifting the notification rates according to the system delays because of dropouts between the two rates were different in different ages. Death and spontaneous recovery without treatment might happen even if the patients have perfect attendance during care-seeking. To investigate the relationship the sex-age distribution at ICS and at the notification, I simulated cohorts started from ICS for different ages, sexes, and years. For each combination of age, sex, and year, I initialised 10,000 patient agents. I then calculated the proportion of the patients who had started TB treatment (notification) after one year of a calendar year. Given the probability of being notified, I inversely calculated the number

of initial care-seeking by dividing the notification by the probabilities in the same age-sex-year stratum.

After retrieving the four time-series of initial care-seeking and notification by sex, I then used them to prepare a likelihood function linking to the simulated data. Explicitly, I utilised the structural vector autoregression (SVAR) with one lag terms to the four time-series. I controlled linear temporal trends and intercepts before measuring the variance-covariance. Since the initial care-seeking rates were built from the notification, I structured the covariance matrix by assuming the dependency but kept the females and males independent. I used count data but not rates because the population dynamic was considered as an exogenous process in this thesis.

I used python 3.7 to perform the simulations and R 3.5.1 with packages `vars` [7] to perform the time-series analysis.

8.4.2 Results

Figure 8.3-A shows that the proportion of patients being notified over people with active TB who initiated care-seeking. The dropouts, combining death and self-cure, increase along with age in both sexes. 15% of the patients aged 65 or above cannot start treatment because of the dropout. I did not identify significant temporal trends in the dropouts. With the age-specific dropout proportions, I calculated the number of initial care-seeking for both sexes every year. Figure 8.3-B demonstrates the mean of initial care-seeking rates compared with the mean of notification rates across 2005-2018 by ages. The differences between them increase by ages. In the age groups above 70, the difference is 20 per 100,000 for females and is 100 per 100,000 for males.

Figure 8.4 shows the trends of and notification and inferred initial care-seeking rates by sexes. The differences in males are about twice as in females. Using SVAR, I estimated the variance of the initial cares-seeking and notification numbers were (140, 90) and (207, 135) for females and males respectively. The covariances were 98 and 146 for females and males. As a result, I formulated the likelihood function linking the results to TB dynamics simulation by multivariate normal distribution

as following:

$$p(x|\mu,\sigma) = \frac{1}{\sqrt{det(2\pi\Sigma)}} \exp\big(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\big) \tag{8.3}$$

$$\mu = \begin{bmatrix} ICS_f \\ Notification_f \\ ICS_m \\ Notification_m \end{bmatrix} \tag{8.4}$$

$$\Sigma = \begin{bmatrix} 91 & 98 & 0 & 0 \\ 98 & 140 & 0 & 0 \\ 0 & 0 & 138 & 146 \\ 0 & 0 & 146 & 207 \end{bmatrix} \tag{8.5}$$

where f for female, m for male; μ were time-series from this section while x denotes the respective values from simulation in Section 8.5.

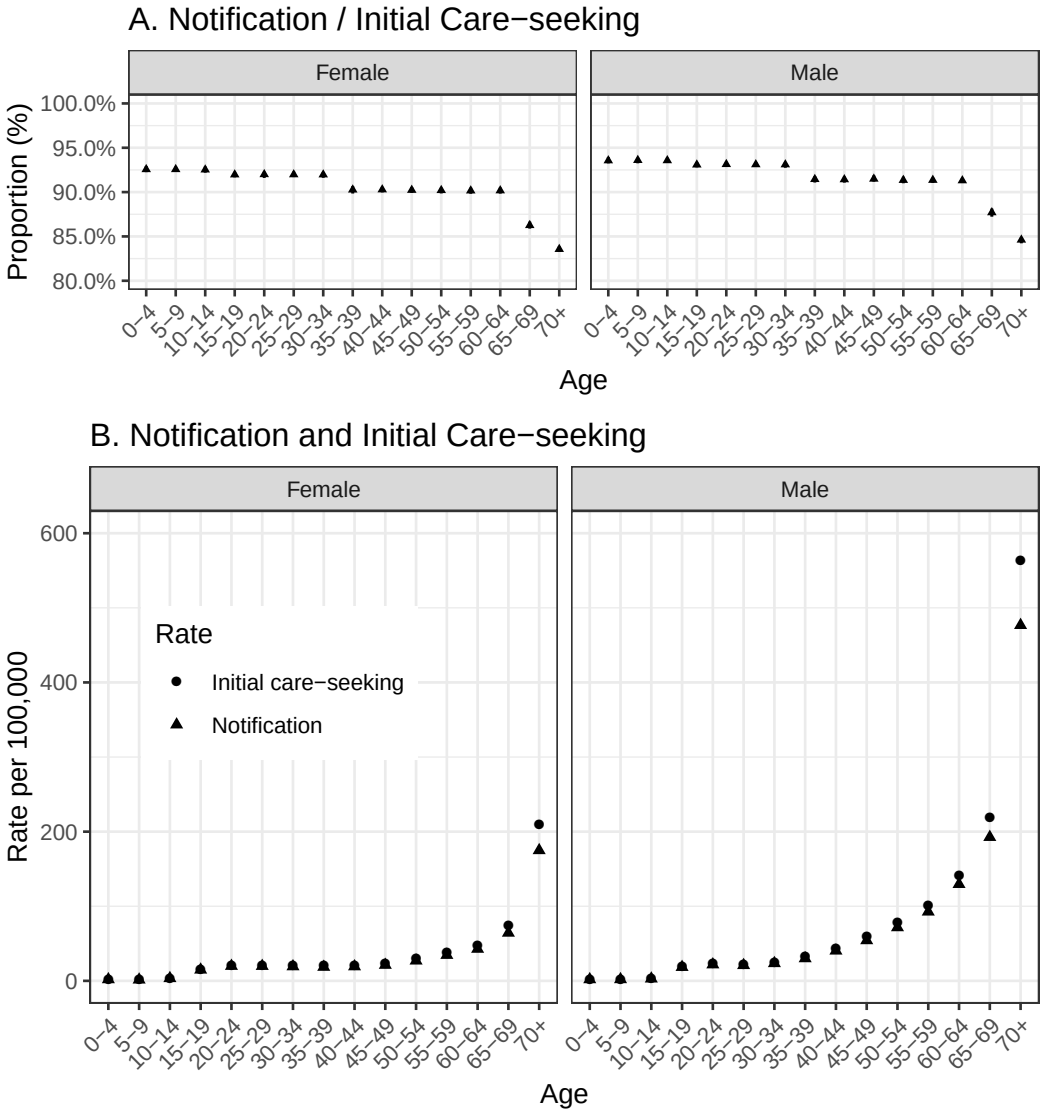


Figure 8.3: Dropouts before treatment by age

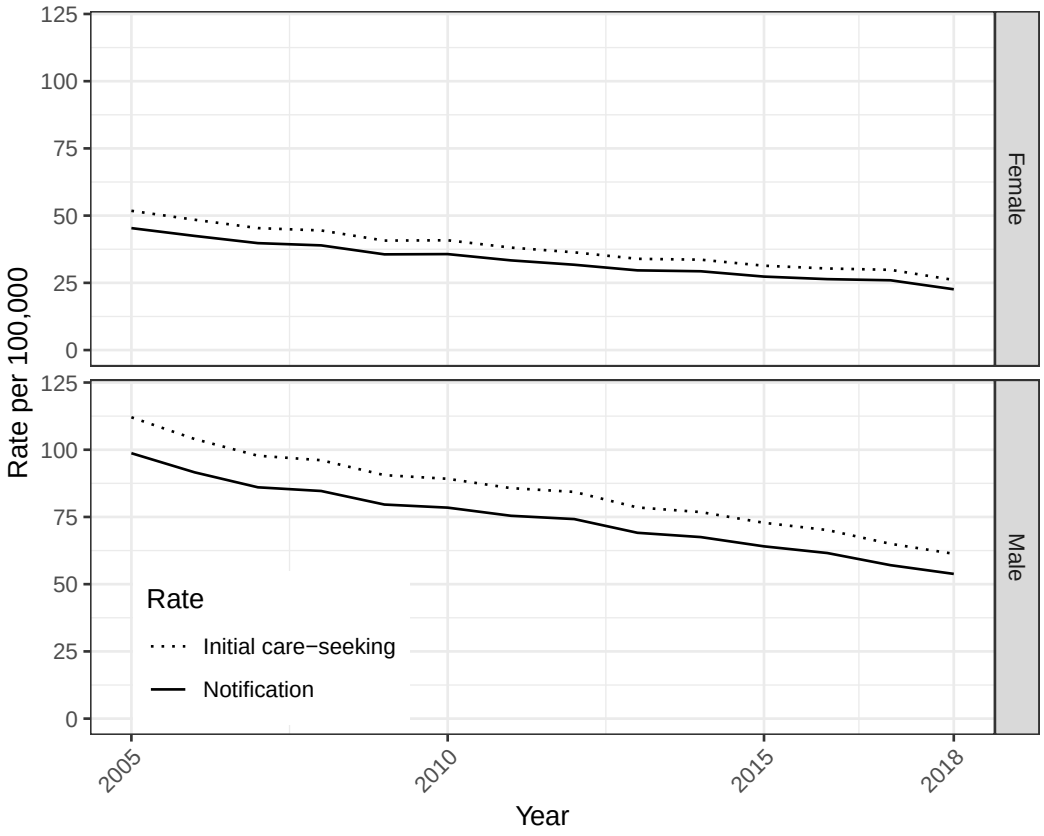


Figure 8.4: Notification and inferred initial care-seeking rates by sexes

8.5 The mean-field model for TB dynamics, MFM

One potential drawback of hybrid modelling is stochasticity in the simulation time. Given the same population size and dynamics, the simulation time of ABMs should be the upper bound of hybrid models. However, the time and memory costs of hybrid simulation highly depend on the number of agents in the system. Especially in model fitting, calibration procedures might input parameters which generate unrealistically large agent-populations in early stages and generate crashes due to insufficient memory. In purely ABMs and EBMs, I can usually expect the time and memory usages before model fitting but not in hybrid modelling.

Therefore, in this section, I first developed a compartmental EBM (MFM) of the system of interest. I tested a model structure for capturing TB dynamics in Taiwan and specified priors of parameters. Also, I regarded the constructed model in this section as the deterministic baseline model (or mean-field model). I expected the hybrid model of the next section behaves like MFM on average.

8.5.1 Model structure

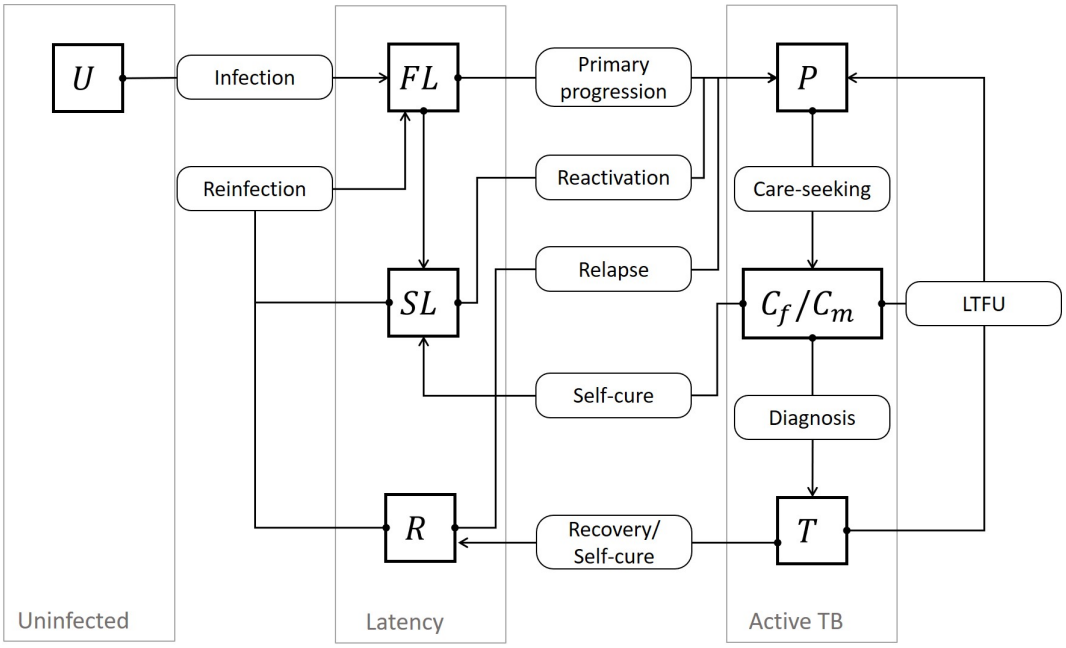


Figure 8.5: Modelling structure: mean-field model. U: uninfected, FL: fast latency, SL: slow latency, P: active TB before care-seeking, C_f: female active TB during care-seeking, C_m: male active TB during care-seeking, T: active TB under treatment, R: recovered

I developed a model based on Ordinary Differential Equations (ODEs). As the ODE system in Equations 8.6 and the conceptual model in Figure 8.5, there are eight compartments in MFM: uninfected (U), fast latency (FL), slow latency (SL), before care-seeking (P), females during care-seeking (C_f), males during care-seeking (C_m), under treatment (T) and recovered (R). From the beginning, people were born as U. After infection, U transit to FL which have a high primary progress rate r_{act} for being activated as P. If activation did not occur within five years, FL turn into SL, which can progress to P as well, but the rate is lower than the primary progression rate. P will seek care after a patient delay (δ_p) and transit to C_f and C_m by κ and $1-\kappa$ respectively. After a system delay (δ_s), C_f and C_m will flow to T and be notified at the same time. T go to R after treatment works. R might relapse to P according to relapse rate r_{rel}. The active TB, including P, C_m, C_f, and T are infectious with identical infectiousness, and they can die due to TB following the TB-specific death rates estimated by Section 8.3. They can also spontaneously cure (r_{cure}) without treatment. C_f, C_m, and T might flow back to P by the rate of r_{loss} if the evaluation interrupted or loss to follow-up. The reinfection of TB might incur to SL and R and bring them to FL after that, but the transmission rate is lower the infection because of a discount from partial immunity (π). Table 8.4 lists the values or prior distributions, and definitions of the parameters.

$$
\begin{aligned}
&flows = disease\ dynamics && +population\ dynamics \qquad (8.6)\\
\frac{\mathrm{d}U}{\mathrm{d}t} &= -\Lambda(t)U && +\Delta_U(t)\\
\frac{\mathrm{d}FL}{\mathrm{d}t} &= \Lambda(t)U + \Big(1-\pi\Big)\Lambda(t)\Big(SL+R\Big) - \Big(r_{prog}+r_{lat}\Big)FL && +\Delta_{FL}(t)\\
\frac{\mathrm{d}SL}{\mathrm{d}t} &= r_{lat}FL - \Big(r_{ract}+(1-\pi)\Lambda(t)\Big)SL && +\Delta_{SL}(t)\\
\frac{\mathrm{d}P}{\mathrm{d}t} &= r_{act}FL + r_{ract}SL + r_{rel}R + r_{loss}\Big(C_f+C_m+T\Big) - \Big(1/\delta_p+r_{cure}\Big)P && -\mu P+\Delta_P(t)\\
\frac{\mathrm{d}C_f}{\mathrm{d}t} &= \Big(\kappa/\delta_p\Big)P - \Big(1/\delta_s+r_{loss}+r_{cure}\Big)C_f && -\mu C_f+\Delta_{C_f}(t)\\
\frac{\mathrm{d}C_m}{\mathrm{d}t} &= \Big((1-\kappa)/\delta_p\Big)P - \Big(1/\delta_s+r_{loss}+r_{cure}\Big)C_m && -\mu C_m+\Delta_{C_m}(t)\\
\frac{\mathrm{d}T}{\mathrm{d}t} &= 1/\delta_s\Big(C_f+C_m\Big) - \Big(r_{rec}+r_{loss}+r_{cure}\Big)T && -\nu T+\Delta_T(t)\\
\frac{\mathrm{d}R}{\mathrm{d}t} &= r_{rec}T + r_{cure}\Big(P+C_f+C_m+T\Big) - \Big(r_{rel}+(1-\pi)\Lambda(t)\Big)R && +\Delta_R(t)
\end{aligned}
$$

The force of infection $\Lambda(t)$ at year t is defined as:

$$\Lambda(t) = \beta(t)\frac{P+C_f+C_m+T}{N} \qquad (8.7)$$

where $\beta(t)$ is the transmission parameter at year t and N is total population size. Then, I defined the transmission parameter as a time-varying function:

$$\beta(t) = max(\beta_0 \times rr_A^{A(t)-A(2000)}, \beta_{lower}) \tag{8.8}$$

where $A(t)$ is a population ageing index at t, β_0 is the initial transmission parameter, β_{lower} is the minimal of $\beta(t)$, and rr_A is the risk ratio of per unit $A(.)$ increment from 2000. I considered two population ageing indices, average age and the proportion of the population aged 65 and above. Also, I used the calendar year as an index and a scenario with constant *beta* as comparators.

The population dynamics (Δ for every state) applied the synthetic demography from Chapter 5. Though, I assumed no migration of those with active TB during care seeking and presumed a perfect border control. As for the other compartments, I supposed the distributions of the latency among the immigrants and emigrants are the same as the prevalence in the system. All newborns starts in the compartment U.

Table 8.4: Prior distributions and assumptions of parameters

Notation	Parameter	Value	Unit	Reference
Exogenous variables				
$1/r_{lat}$	duration of fast latency	5	$year$	By definition
r_{cure}	spontaneously cure rate	0.33	$year^{-1}$	Tiemersma et al. [8]
r_{rec}	recovery rate (with treatment)	1	$month^{-1}$	Assumed
r_{loss}	LTFU rate	0.208	$year^{-1}$	Section 8.3
π	partial immunity for reinfection	0.8	None	Vynnycky and Fine [9]
μ	TB fatality without treatment	0.122	$year^{-1}$	Tiemersma et al. [8]
ν	TB fatality after treatment	0.116	$year^{-1}$	Section 8.3

Notation	Parameter	Prior distribution	Unit	Reference
Parameters to fit and their dependants				
β_0	initial β	$U(1, 40)$	$year^{-1}$	Assumed
β_{lower}	minimal β	$U(1, \beta_0)$	$year^{-1}$	Assumed
r_{act}	primary progression rate	15%	$(5years)^{-1}$	Trauer et al. [10]
r_{ract}	reactivation rate	$0.01r_{act}$	$year^{-1}$	Assumed
r_{rel}	relapse rate	$0.01r_{act}$	$year^{-1}$	Assumed
κ	proportion of female patients at initial care-seeking	$U(0.2, 0.4)$	%	Chapter 5
rr_A	risk ratio of population ageing	$U(0.5, 1)$	None	Assumed
δ_p	patient delay	$U(7, 60)$	day	Chiang et al. [11]
δ_s	system delay	$U(7, 111)$	day	Chapter 6

8.5.2 Populating and equilibrating the model

I attempted to equilibrate the model by numerically solving the condition that all fractions of compartments are in a steady state. I dropped the parameter sets that led the model to disease-free equilibrium. I fixed all parameters and demography to values in 2000 while equilibrating period. In this model, I included population dynamics driven by real data. In reality, the population was growing in 2000. I made a dynamic adjustment to keep the equilibrium without distorting the original dynamics.

Explicitly, I transformed the model in numbers into fractions of compartments. Consider

$$x(t) = X(t)/N(t) \tag{8.9}$$

where $x(t)$ and $X(t)$ are the proportion and number of compartment x at t respectively, and $N(t)$ is the total population at t. The first-order derivative is

$$\frac{\mathrm{d}x(t)}{\mathrm{d}t} = \frac{\mathrm{d}X(t)/N(t)}{\mathrm{d}t} \tag{8.10}$$

$$= \frac{N(t)\mathrm{d}X(t)/\mathrm{d}t - X(t)\mathrm{d}N(t)/\mathrm{d}t}{N(t)^2} \tag{8.11}$$

$$= \frac{\mathrm{d}X(t)/\mathrm{d}t}{N(t)} - x(t)\frac{\mathrm{d}N(t)/\mathrm{d}t}{N(t)} \tag{8.12}$$

For keeping the nature of population dynamics, the population in each compartment need an additional outflow with the rate corresponding to the overall population growth in the initial year.

8.5.3 Calibration and model selection

I calibrated my model to fit the annual sex-specific TB notification and the incidence of initial care-seeking from 2005 to 2018 as Figure 8.4 with the variance-covariance estimated by Section 8.4. Explicitly, for each time point, the simulated notification and initial care-seeking follows a multivariate normal with a mean vector of the numbers of notification and initial care-seeking by sexes, and a variance-covariance matrix in Equation 8.3. I employed the Approximate Bayesian Computation with sequential Monte Carlo (ABC-SMC) procedure with a tolerance adaptation scheme to the calibration by the methods derived in Del Moral et al. [12]. Following the fitting algorithm, I set up 500 samples per iteration, and ran until the convergence in the value of tolerance and the effective sample size. Also, I solved the maximum likelihood estimator (MLE) and the maximum a posteriori

(MAP) by a genetic algorithm (GA). As ABC-SMC, 500 sample size per iteration were used and I ran until the convergence in the maximum values (likelihood for MLE, posterior probability for MAP).

I examined different rates of TB primary progression rates (10%, 15%, 20%, and 30% within 5 years after infection) and time-series indicating population ageing (average age, the proportion of the population aged 65 and above (PrO), calendar year, and constant). I used Deviance information criterion (DIC) to select the optimal model among the options for the forecasting and intervention analysis (Chapter 9). Also, I targeted the proportion of latency among the population in 2018 and 2035 at 20-40% roughly aligning the Global burden [13].

8.5.4 Results

Table 8.5 shows the DIC and the proportion of latency in 2018 and in 2035 given different model assumptions after calibration. Using the primary progression within 5 years at 15% and 20% rendered lower DICs in general. The comparator of constant transmission parameter shows higher DICs across the primary progression rates. Using calendar time as the index of population ageing shows similar DICs to that of average age. The cases with PrO had the lowest DIC across the range of the primary progression rates. I evaluated the models with primary progression within 5 years and ageing index of (15%, PrO) and (20%, PrO) were identically the best. However, considering the literatures, the primary progression probability with 5 years after infection, the estimation were ranged from 5% [14, 15] to 15% [10]. Taken it together with the stochastics of the simulation, I preferred (15% PrO) and its posterior parameters (see Table 8.6) as my final selection.

Figure 8.6 compares the selected model, (15% PrO), and the time-series of initial care-seeking rates and notification rates by sexes. The simulation captured the trends and the patterns of the four time-series. The estimators of MAP and MLE were close to data points without biases. The 95% credible intervals of the posterior simulations fully covered the data points. However, the median time-series of the posterior simulations were slightly above the data of initial care-seeking rates and below notification rates.

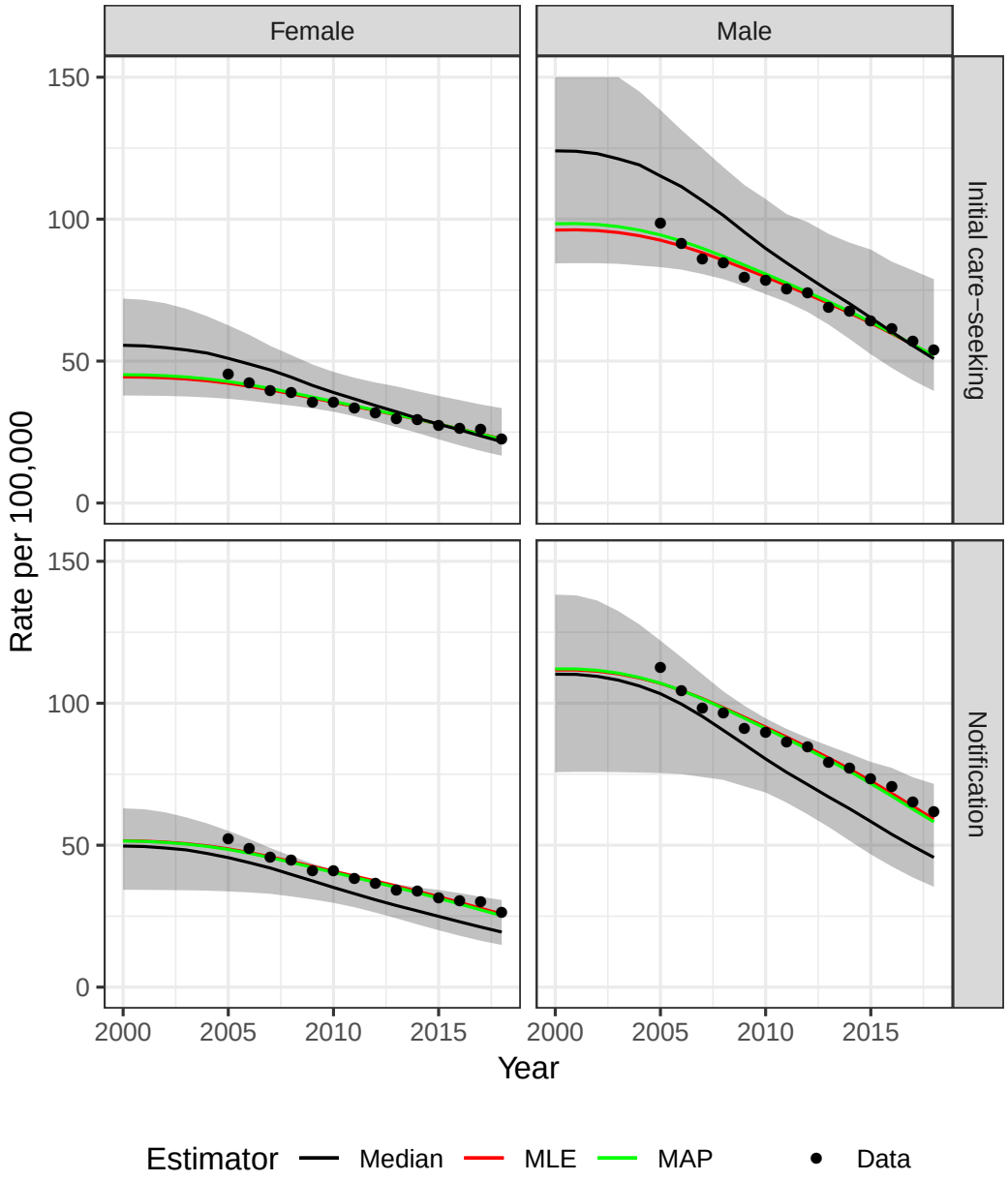


Figure 8.6: Posterior simulations with data points. The shaded areas are the 95% credible intervals of the posterior simulations

Table 8.5: Mean-field model, comparison

Primary progression within 5 years (%)	Ageing Index	DIC	Latency (%) in 2018	Latency (%) in 2035
10	Constant	6212	42 (38, 46)	46 (42, 50)
	Average age	4247	46 (39, 53)	44 (39, 51)
	Population $\geq$ 65	3202	47 (42, 53)	44 (40, 49)
	Calender year	5914	45 (38, 53)	44 (38, 52)
15	Constant	7031	28 (20, 36)	31 (22, 40)
	Average age	2696	44 (31, 56)	41 (29, 52)
	Population $\geq$ 65	2483	38 (29, 45)	35 (27, 41)
	Calender year	4214	42 (29, 56)	39 (28, 51)
20	Constant	7866	22 (11, 32)	24 (12, 36)
	Average age	6221	39 (25, 50)	36 (23, 46)
	Population $\geq$ 65	2361	31 (20, 41)	28 (19, 37)
	Calender year	6406	40 (27, 49)	37 (24, 45)
30	Constant	31846	14 (4, 26)	15 (4, 29)
	Average age	11921	27 (12, 43)	25 (12, 39)
	Population $\geq$ 65	16116	21 (8, 37)	20 (7, 34)
	Calender year	10485	28 (13, 43)	26 (12, 39)

DIC: deviance information criterion

Table 8.6: Posterior distributions for parameters

Parameter	Prior distribution	Unit	Posterior	
			Mean (95% CrI)	Median (Q1, Q3)
β_0	$U(1, 40)$	$year^{-1}$	28.63 (21.11, 37.81)	27.93 (24.99, 31.98)
β_{lower}	$U(1, \beta_0)$	$year^{-1}$	5.37 (1.28, 9.6)	5.5 (3.35, 7.27)
rr_A	$U(0.5, 1)$	None	0.14 (0.03, 0.26)	0.13 (0.08, 0.18)
κ	$U(0.2, 0.4)$	%	0.30 (0.29, 0.31)	0.30 (0.29, 0.30.46)
δ_p	$U(7, 60)$	day	37.42 (9.14, 59.57)	40.41 (23.84, 50.98)
δ_s	$U(7, 111)$	day	75.65 (14.02, 111.92)	83.76 (56.24, 99.47)

8.6 Hybrid modelling, HM

Combining the advantage of ABMs in describing details processes and the efficiency of EBMs, I developed a hybrid model for complex care-seeking processes

and overall TB epidemics. Starting from MFM I developed and calibrated in the last section, and I replaced the compartments after the active TB initiated care-seeking and before completing TB treatment with the agent-based submodel described in Section 8.4.

It should be noted that the EBM in this section always indicates the EBM submodel of HM while MFM indicates the EBM constructed in the previous section. Similarly, the ABM in this section indicates the ABM submodel of HM.

8.6.1 Coupling and synchronisation of the EBM and ABM

Transmission and infection processes were modelled in the EBM as MFM. For calculating the force of infection, the EBM requests the number of infectious people in the ABM by a turn-based schedule. Once the value updated, force of infection will change. The definition of transmission parameters $\beta(t)$ is the same as MFM, declining with population ageing until hitting the lower bound.

Tracing the number of infectious people is an important task of the ABM. In the ABM, when a new active TB being added or an active TB being inactivated, the number of infectious agents will be updated. However, the ABM calculates the moving average of the number of infectious agents and reports to the EBM each discrete time step. Before fetching the number of infectious in the ABM, the EBM will update to the time when the event happened in the ABM.

Initial care-seeking process is a transfer from the EBM to ABM. In every time step, I sampled a proportion of the active TB according to the rate of the invasion of patient delay (δ_p) and transferred them to the ABM as patients. When a patient is instantiated in the ABM, their details will be sampled according to age-sex distribution modelled in Chapter 5 and distributions of the other personal features fitted in Chapter 6.

Turn-based events might cause cycles of TB prevalence: ie more agents at the beginning of each turn, and the number declining until the beginning of next turn. To reduce the influence of this potential effect, I used a finer scale (10 per observation interval) on the time steps than the observational interval of one year to eliminate the information loss.

Death, recovery, loss to follow-up, and treatment success are events removing agents from the ABM. The agents, after recovered and treatment success, will be transferred back to the recovered of EBM. The patients dropped out because of IE/LTFU will go to the active TB prior to initial care-seeking and have a chance

to start care-seeking again. Those events applied a real-time update. The EBM updates to the event time and takes the recovered or lost patient whenever an event happens.

It is worth noting that T in MFM includes only the TB patients under treatment before TB inactivated. Once the TB has been inactivated patients will be counted as R. In the ABM part of HM, the agents will be transferred to R the EBM only if the treatment competed.

8.6.2 Simulation and validation

I used the posterior parameters directly from the MFM results. As I kept 500 posterior parameter sets in the previous section, I used them iteratively for simulating HM results. That is, for each parameter set, given the values, I simulated outputs using MFM and HM simultaneously and calculated the difference after that. I calculated the difference between MFM and HM by the percentage difference:

$$difference(hm_t - mfm_t) \equiv \frac{hm_t - mfm_t}{mfm_t} \times 100\% \tag{8.13}$$

$$\approx \Big(log(hm_t) - log(mfm_t)\Big) \times 100\% \tag{8.14}$$

where hm_t and mfm_t are variables at time t generated by HM and MFM respectively, The approximation was derived from the first order Taylor expansion of the logarithm function to prevent over-sensitive baseline value, i.e. mfm_t.

8.6.3 Results: Validation

Figure 8.7 compares the simulations of and differences between MFM and HM by sex-specific ICS and notification rates. The simulations of HM reproduced the patterns of MFM for all the four time-series, and HM rendered a wider prediction interval. In terms of the difference, the simulations of HM were unbiased, indicating the MFM can represent the mean-field of the system. HM sometimes generated sporadic outbreaks during simulation, but that did not cause HM crashed. Most of the differences between HM and MFM simulations were within 30%. However, some extreme outbreak cases of HM had 150% more than their MFM simulation.

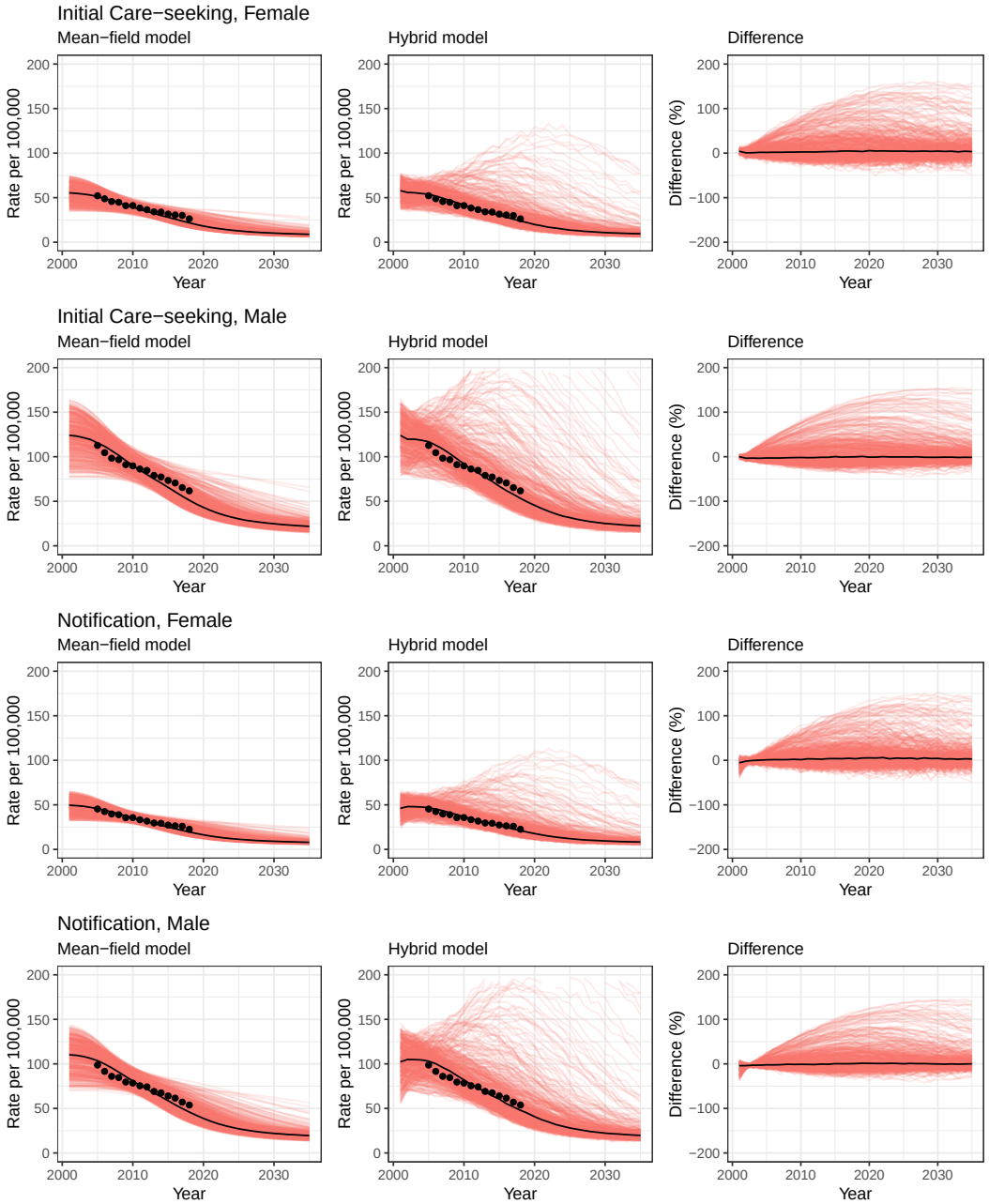


Figure 8.7: **Comparison of the results of MFM and HM.** Incidence and mortality are in annual rate. Red lines indicate simulations; black lines indicate the median values in each time stage; black dots are data points.

8.7 Simulation outputs and results

8.7.1 Epidemiology

Figure 8.8 demonstrates the TB incidence, prevalence, and mortality rates between 2001 to 2035. As Figure 8.7 three columns were results of MFM, HM, and their difference, respectively. The impacts of sporadic outbreaks showed in the prevalence and mortality as well. Also, the three indices by HM did not bias from MFM results. The results forecasted that the TB incidence will be 42 (95% prediction interval, PI: 22-136) per 100,000, and 19 (95% PI: 12-48) per 100,000 in 2020 and 2035 respectively.

Dropouts

Using CSM, Table 8.7 demonstrates cohort-based dropouts with 2019 and 2035 cohorts. The dropouts due to death and self-cure were categorised by stages: pre-hospital care-seeking, between initial care-seeking and treatment initiation, after treatment start. There is no LTFU because, in this cohort simulation, the patients who had LTFU went back to pre-hospital care-seeking (compartment P), they will recur to care-seeking until other dropouts or treatment completion happen. In both 2019 and 2035 cohorts at the 12th month, around 5% will drop before care-seeking, and around 10% will drop before TB treatment start. In total, in the first year after symptoms onset, 15% of active TB will not have a chance to receive TB treatment. The simulation forecasted that the distributions of dropouts in the two cohorts will be close. Considering the trends of dropouts, dropouts in all categories will happen in the first three months.

Figure 8.9 shows a population-based dropout cascades, which calculated from the annual rates of incidence, ICS, notification, and cured by TB treatment. Figure 8.9-A demonstrates that the four rates are declining at similar rates. From 2005, there were about ten incident cases per 100,000 lost between ICS and notification (treatment start). The gaps closed to 5 in 2015 and will be reducing within my forecast end. In terms of the relative numbers as Figure 8.9-B, the period between ICS and notification always contributed to around 10% of incidence rates; about 3% were during treatment.

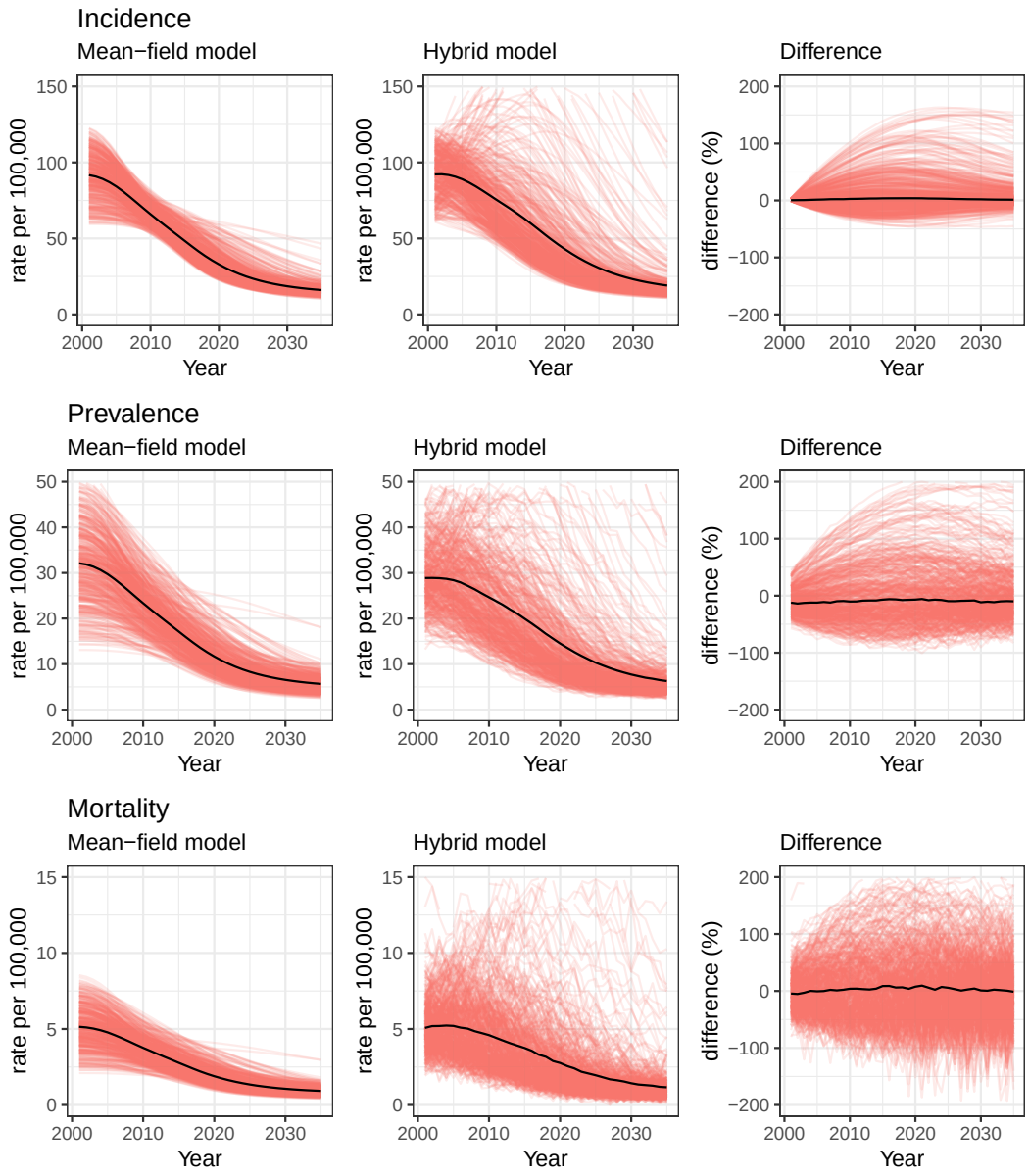


Figure 8.8: **TB epidemiology by MFM and HM.** Incidence and mortality are in annual rate. Red lines indicate simulations; black lines indicate the median values in each time stage.

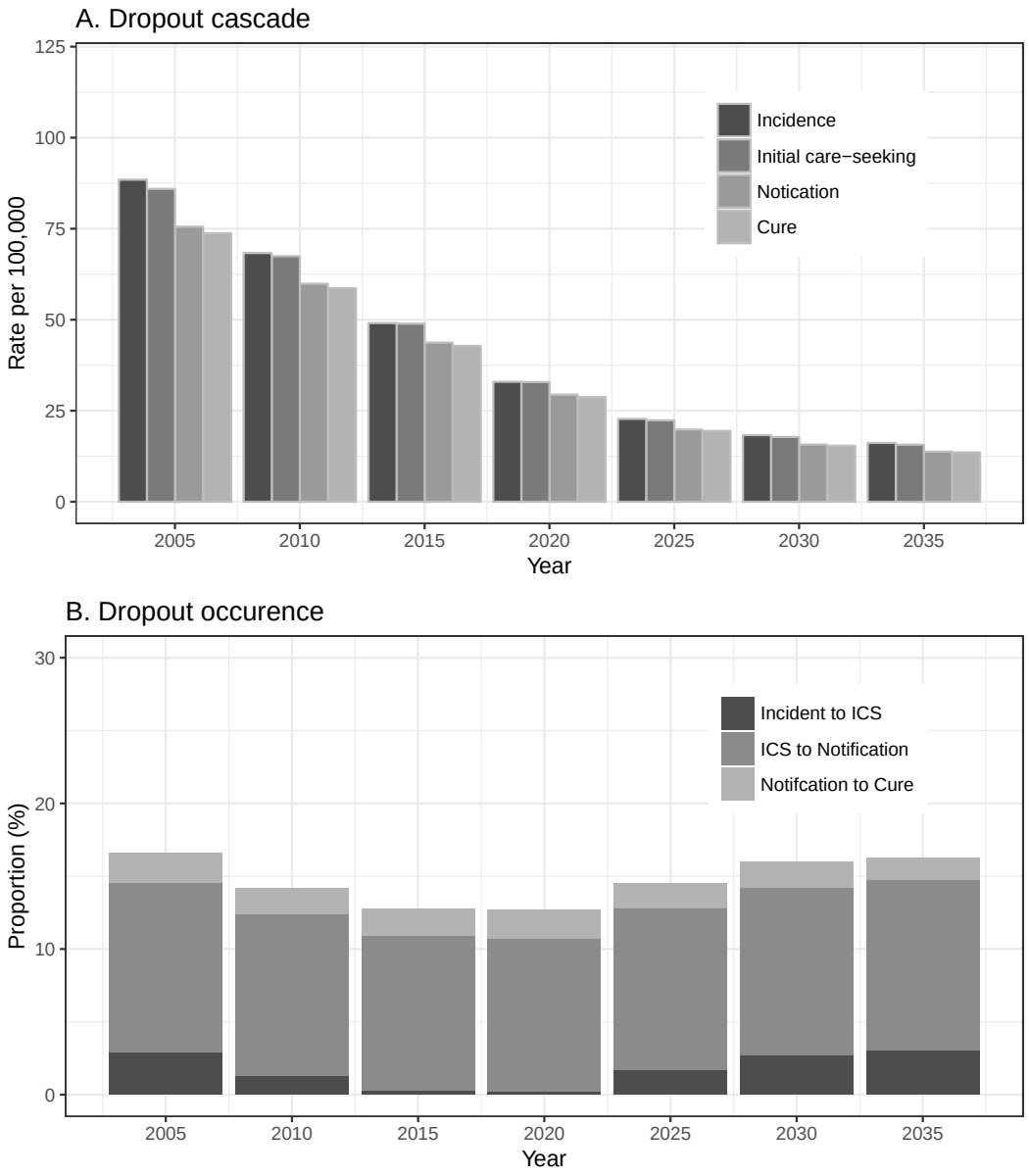


Figure 8.9: A population-based dropout cascades by hybrid modelling

Table 8.7: Cohort-based dropouts, 2019 and 2035 cohorts

Dropout	3rd month mean (95% PI)	6th month mean (95% PI)	9th month mean (95% PI)	12th month mean (95% PI)
In 2019				
		Pre-Hospital		
Self-cure	3% (0.9%, 4.2%)	3.5% (1%, 5.2%)	3.6% (1%, 5.5%)	3.7% (1%, 5.7%)
Death	1.4% (0.5%, 2%)	1.7% (0.5%, 2.5%)	1.8% (0.5%, 2.7%)	1.8% (0.5%, 2.8%)
		Before treatment		
Self-cure	3% (2.1%, 4.3%)	5.3% (4.4%, 6.4%)	6.3% (5.5%, 7.2%)	6.7% (5.9%, 7.5%)
Death	1.1% (0.7%, 1.6%)	2% (1.5%, 2.5%)	2.3% (1.9%, 2.8%)	2.5% (2.1%, 3%)
		During treating		
Death	0.3% (0.2%, 0.6%)	0.7% (0.4%, 1%)	0.8% (0.6%, 1.2%)	0.9% (0.6%, 1.2%)
In 2035				
		Pre-Hospital		
Self-cure	3% (0.9%, 4.2%)	3.4% (1%, 5.1%)	3.6% (1%, 5.4%)	3.7% (1%, 5.6%)
Death	1.5% (0.5%, 2%)	1.7% (0.5%, 2.5%)	1.8% (0.5%, 2.7%)	1.8% (0.5%, 2.8%)
		Before treatment		
Self-cure	3% (2.1%, 4.2%)	5.1% (4.3%, 6.2%)	6.1% (5.3%, 7%)	6.6% (5.8%, 7.4%)
Death	1.1% (0.7%, 1.7%)	1.9% (1.4%, 2.5%)	2.3% (1.8%, 2.8%)	2.4% (2%, 2.9%)
		During treating		
Death	0.3% (0.2%, 0.6%)	0.6% (0.4%, 0.9%)	0.8% (0.5%, 1.1%)	0.9% (0.6%, 1.2%)

PI: prediction interval

8.8 Discussion

This chapter forged a hybrid model of TB epidemiology, considering the care-seeking process in Taiwan using the results of the former chapters in this thesis. Chapter 6 empirically sketched the care-seeking pathways of TB patients with links to socioeconomic factors. Chapter 2 featured the intervenable factors or processes which can be considered when innovating TB control strategy. Chapter 4 summarised how to model health behaviours, including care-seeking behaviour in ABMs. Section 8.2 searched, examined, and selected from various state-space models, and used the best one, StSp, to simulate the patient pathways. Section 8.3 used a part of results from the pathways to estimate the parameters of TB-specific death and LTFU under treatment. Section 8.4 used StSp, and dropout parameters to develop an agent-based care-seeking model, CSM. Chapter 5 profiled the age-sex distributions of TB notification in the near future by analysing the notification data in Taiwan from 2005 to 2018. Section 8.5 used it to develop a mean-field model, MFM, to capture the epidemiology of TB in Taiwan. Lastly, Chapter 3 reviewed the examples of coupling agent-based and equation-based simulations in health fields. Section 8.6 combined the CSM and MFM part to develop a hybrid model, HM, which can be employed to assess the SES-based interventions on

population outcome.

8.8.1 Features of the final hybrid model

According to the strength of TB evaluation and treatment prescription, I identified Waiting, Evaluating, TB Detecting, and Treating Stages from the case-seeking pathways. In this section, I built state-space models to capture the dynamics of the transitioning among the stages. Then, among the options of state-spaces and parametrisation, I selected two candidate models which have the best capability of simulating the system delays with the lowest distance to that in the data.

I built and calibrated a hybrid model, i.e. HM, in this chapter. My models have the following features.

HM can consider the intervention through the quantified effects in the underlying state-space model, StSp. Once a substantial intervention target is available, I can propose impulses on the lengths of time-to-states or the probabilities of skipping a state and evaluate the responses of the model according to the changes.

HM can consider the intervention by changing the TB service capability distributions. The diagnostics availability in the constructed model were based on the IPPA results with individual hospital data. By assigning the capacity of providing some services to the hospitals in the IPPA results, I can assess their impulse-responses as well.

With the hybrid modelling approaches, HM facilitated the country-level agent-based simulations by collapsing the agents outside these who seek care for TB into a compartment model. For my case, in population 23 million with a TB endemic of 5,000 active TB cases, a simulation for 35 calendar years required only about three minutes in personal computers (see Section 8.6 for specification).

Along with HM, I provided an equation-based comparator, MFM. The concepts of MFM and HM are identical except for the details about the care-seeking and treatment processes. I used MFM to test the functionality and to calibrate the parameters before the hybrid modelling. Also, MFM provided a deterministic view of HM. Given the comparison of the simulations of HM and MFM, I assert that the HM results are compatible with the MFM results while HM revealed the issues around stochasticity.

As a new application of hybrid modelling, HM maximised the utility of data without adding uncertain modelling components or oversimplifying the components of interest. In the EBM usual, I used stratification to address the heterogeneous among the strata, and I used a single summary statistic to simplify a complex distribution to a term in an equation. These approaches introduce information losses, requiring further justification to address the biases. In the ABM

part, I modelled every component if it is involved in the process of interest. Sometimes, I do not have enough data or knowledge to inform all of them, so I needed to assume the mechanism driving them and conduct many sensitivity analyses to support these assumptions. For example, to model the transmission process, I might need a social network modelling or data from a contact survey. However, such surveys and modelling approaches come with high costs. Apart from the processes I had data to support, I parametrised them in the EBM sub-model as much as possible. If I expected a process would be relevant to the intervention analysis, I would consider them in the ABM sub-model.

8.8.2 State-space modelling

According to the model selection of StSp, it is preferable to model periods before and after the TB-related evaluation start. That is, partitioning system delays with serial distributions are better than using a single distribution to model. In terms of the parametrising stage skipping, zero-inflated distributions are better than multi-state models. This results confirmed the covariates analysis in Chapter 6 that the risk factors of non-zero delay and longer delay were diverse.

In my study, I used detailed data on the care-seeking pathways, which means I could parametrise the connection between stage with various options of parametrisation. Thinking the studies which can only collect data of the overall system delay or any delays to treatment start, my analysis strengthened the intuition of using phase-type distributions for capturing the distributions of delays. For example, The Coxian-Erlang distribution which has a zero-inflated term and few following hidden stages serially composed of Exponential distributions. Practically, the Coxian-Erlang distribution has been proved as a generalised distribution covering many distributions for time-to-event data if the first three statistical moments are fetched [16, 17].

8.8.3 Dropouts

I used different TB-specific death rates for before and after treatment start. The former used the external sources and later used the values from the IPPA results. For patients younger than 65, the fatality rates decrease after treatment while that of the patients aged 65 or above increases. Ideally, under proper TB care, a patient should be well off and reduce the mortality on average. However, the results were not contradictory. The death rates were not necessarily lower than the death rates before treatment starts concerning the health status decaying along with time.

I counted the spontaneous self-cure without treatment as a type of dropout

but not a favourable outcome. Although self-cured patients stop suffering from TB, they miss the opportunities of accessing proper diagnostics and receiving preventive treatment for latent TB infections. Moreover, unless reactivation, they will never realise that their risk of getting TB are higher than uninfected population.

I presented dropout measurements from both cohort-based and population-based perspectives. My population-based dropouts were calculated by the ratios of occurrence rates of critical events. This was compatible with the case-detection gaps which were commonly used for international comparison. However, since the approach does not consider time lag between occurrence rates, the estimators are biased by the annual decline rate of TB incidence. For example, given one year notification delay and TB incidence declines annually by 5%, the case-detection gap is less than 0% if dropout rates are smaller than 5%. My cohort-based dropouts differ from conventional approaches to measure dropout cascades, which depend on the proportional dropouts from cross-sectional survey data, e.g. Subbaraman et al. [18]. My approach complemented the drawback of using cross-sectional data, which does not consider temporal information. For example, in cross-sectional data, I cannot identify if a patient will not be or has not been diagnosed as a TB patient yet.

8.8.4 Mean-field model

The major tasks of mean-field modelling were capturing the transmission dynamics and TB latency. To reflect the active TB progression from being infected, there is a variety of parametrisations for capturing latency structure. The latency can be modelled by a single latent state or multiple states. For multiple latent states, the connections to them from the susceptible population can be parallel, the susceptible go to different latency by respective probabilities, or serial, the infected go through every latency than being active TB, the progression rates for different latent states can be identical or different. Some simulation methods allow progression rates to vary continuously by time-since-infection. Moreover, the flows among latent states provide more potentials for TB modelling. In line with the empirical observation that TB activation rate decreases after the incubation period, Menzies et al. [19] reviewed 312 models, suggesting that the model with a single latent state or model having a series of latent states ended with active TB performed poorly. Ragonnet et al. [20] reviewed and examined the structures for modelling TB in low TB burden settings. The study confirmed the structure using two latent states with different rates of disease progression as an optimal structure. Accordingly, I used two latent populations in my mean-field modelling. In my implementation, I separated the second population with lower disease progression rate into slow

latency (SL) and recovered (R) to differentiate the individuals treated before or not, while I did not parameterise them differently. This approach was to provide a potential for further investigations on the impact of previous TB treatment record on the diagnostic process.

8.8.5 Hybrid modelling

My hybrid modelling is a novel approach in addressing the interaction between TB patients and healthcare systems. To the best of my knowledge, it is the first application projection care-seeking pathways in individual level to overall TB epidemiology.

The model architecture was synthesised multiple studies from the collection of Chapter 3. I used the frameworks of Mielczarek and Zabawa [21] to populate agents and injecting population dynamics. I employed the idea from Djanatliev and German, 2013 [22], to design a coupling strategy, ensuring the information in the ABM and EBM parts were well-synchronised. I took Leonenko et al. [23] as an example to separate the infectious population and the others while keeping the value of force of infections unbiased.

In the results, by comparing the simulations of MFM and HM, MFM sufficiently represented the mean-field of HM, so I did not re-calibrated the HM but used the posterior parameters directly from MFM. However, HM sometimes rendered higher epidemics than MFM. I attributed this to a nature of stochasticity models. HM has a chance to simulate casual outbreaks, and the chance positively correlated to the prevalent cases.

8.8.6 Limitations

My model did not consider asymptomatic, extrapulmonary, and multiple drug-resistant (MDR) TB cases. The asymptomatic TB have infectiousness, but they have a lower chance of transmitting the pathogen through the air without coughing. Thus, this simplification might lead to potential underestimation on TB epidemiology. As for extrapulmonary and MDR TBs, I ignored them and regarded them as the other pulmonary TB. I decided due to its small share among all TB: about 3% [24], and since the features were less relevant to my primary interest. As extrapulmonary and MDR TB has more uncertainty around severity and infectiousness, my decision of ignoring them were also due to robustness reasons.

Apart from the transmission parameters and the population dynamics, I did not consider the other parameters as time-varying functions. The main reason was that I did not identify validated evidence to support the trends. I have thought

that the patient delay may reduce due to the implementation of active case finding (X-ray vans for mountainous areas and contact investigation). However, given the data I had (the IPPA results and the notification time-series), the trend of the transmission parameter and the patient delay was not identifiable. Regarding the trend of the transmission parameter implied the changes in social networks, the intention of self-protection, and transportation behaviour, I preferred to model the intrinsic temporal trend by the transmission parameter.

Although I used regional SES as predictors in parameter estimation for StSp, I did not consider geographical information in HM. The health disparity and TB clustering exist. Especially in the mountainous area in Eastern Taiwan, although the TB clustering is always a focus of TB control of Taiwan CDC, the healthcare accessibility is still below needed. Also the TB incidence and mortality were significantly over the national average. The southern end of Taiwan showed similar issues as well [24].

Bibliography

[1] Braekers R, Grouwels Y. A semi-parametric cox's regression model for zero-inflated left-censored time to event data. Communications in Statistics - Theory and Methods. 2016 Apr;45(7):1969–1988.

[2] Perez-Cruz F. Kullback-Leibler divergence estimation of continuous distributions. In: 2008 IEEE International Symposium on Information Theory; 2008. p. 1666–1670.

[3] Therneau TM, Grambsch PM. Modeling Survival Data: Extending the Cox Model. Springer Science & Business Media; 2013.

[4] Jackson CH, Others. Multi-state models for panel data: the msm package for R. J Stat Softw. 2011;38(8):1–29.

[5] Plummer M, Others. JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In: Proceedings of the 3rd international workshop on distributed statistical computing. vol. 124. ci.tuwien.ac.at; 2003. .

[6] Su YS, Yajima M. R2jags: A Package for Running jags from R. R package version 0 03-08, URL http://CRAN R-project org/package= R2jags. 2012;.

[7] Pfaff B, Others. VAR, SVAR and SVEC models: Implementation within R package vars. J Stat Softw. 2008;27(4):1–32.

[8] Tiemersma EW, van der Werf MJ, Borgdorff MW, Williams BG, Nagelkerke NJD. Natural history of tuberculosis: duration and fatality of untreated pulmonary tuberculosis in HIV negative patients: a systematic review. PLoS One. 2011 Apr;6(4):e17601.

[9] Vynnycky E, Fine PE. The natural history of tuberculosis: the implications of age-dependent risks of disease and the role of reinfection. Epidemiol Infect. 1997 Oct;119(2):183–201.

[10] Trauer JM, Moyo N, Tay EL, Dale K, Ragonnet R, McBryde ES, et al. Risk of Active Tuberculosis in the Five Years Following Infection . . . 15%? Chest. 2016 Feb;149(2):516–525.

[11] Chiang CY, Chang CT, Chang RE, Li CT, Huang RM. Patient and health system delays in the diagnosis and treatment of tuberculosis in Southern Taiwan. Int J Tuberc Lung Dis. 2005 Sep;9(9):1006–1012.

[12] Del Moral P, Doucet A, Jasra A. An adaptive sequential Monte Carlo method for approximate Bayesian computation. Stat Comput. 2012 Sep;22(5):1009–1020.

[13] Houben RMGJ, Dodd PJ. The Global Burden of Latent Tuberculosis Infection: A Re-estimation Using Mathematical Modelling. PLoS Med. 2016 Oct;13(10):e1002152.

[14] Vynnycky E, Fine PE. Lifetime risks, incubation period, and serial interval of tuberculosis. Am J Epidemiol. 2000 Aug;152(3):247–263.

[15] Behr MA, Edelstein PH, Ramakrishnan L. Revisiting the timetable of tuberculosis. BMJ. 2018 Aug;362:k2738.

[16] Osogami T, Harchol-Balter M. Necessary and Sufficient Conditions for Representing General Distributions by Coxians. In: Computer Performance Evaluation. Modelling Techniques and Tools. Springer Berlin Heidelberg; 2003. p. 182–199.

[17] Osogami T, Harchol-Balter M. Closed form solutions for mapping general distributions to quasi-minimal PH distributions. Performance Evaluation. 2006 Jun;63(6):524–552.

[18] Subbaraman R, Nathavitharana RR, Mayer KH, Satyanarayana S, Chadha VK, Arinaminpathy N, et al. Constructing care cascades for active tuberculosis: A strategy for program monitoring and identifying gaps in quality of care. PLoS Med. 2019 Feb;16(2):e1002754.

[19] Menzies NA, Wolf E, Connors D, Bellerose M, Sbarra AN, Cohen T, et al. Progression from latent infection to active disease in dynamic tuberculosis transmission models: a systematic review of the validity of modelling assumptions. Lancet Infect Dis. 2018 Aug;18(8):e228–e238.

[20] Ragonnet R, Trauer JM, Scott N, Meehan MT, Denholm JT, McBryde ES. Optimally capturing latency dynamics in models of tuberculosis transmission. Epidemics. 2017 Dec;21:39–47.

[21] Mielczarek B, Zabawa J. Simulation model for studying impact of demographic, temporal, and geographic factors on hospital demand. In: Chan V, DAmbrogio A, Zacharewicz G, Mustafee N, editors. 2017 Winter Simulation Conference (WSC). Winter Simulation Conference Proceedings; 2017. p. 4498–4500.

[22] Djanatliev A, German R. Prospective Healthcare Decision-making by Combined System Dynamics, Discrete-event and Agent-based Simulation. In: Proceedings of the 2013 Winter Simulation Conference: Simulation: Making Decisions in a Complex World. WSC '13. Piscataway, NJ, USA: IEEE Press; 2013. p. 270–281.

[23] Leonenko VN, Pertsev NV, Artzrouni M. Using High Performance Algorithms for the Hybrid Simulation of Disease Dynamics on CPU and GPU. In: Koziel S, Leifsson L, Lees M, Krzhizhanovskaya VV, Dongarra J, Sloot PMA, editors. International Conference on Computational Science, Iccs 2015 Computational Science at the Gates of Nature. vol. 51 of Procedia Computer Science; 2015. p. 150–159.

[24] Centers for Disease Control, Ministry of Health and Welfare, R O C (Taiwan). Taiwan Tuberculosis Control Report 2017. Centers for Disease Control, Ministry of Health and Welfare, R.O.C.(Taiwan); 2018.

Chapter 9

Analysis of interventions to improve TB diagnostics assess

9.1 Introduction

As I reviewed in Chapter 2, TB epidemics can be intervened on by two groups of SES-based approaches: Universal Health Coverage (UHC) and Social Protection. UHC ensures people with TB-suggestive symptoms can assess healthcare facilities which can treat TB. Social protection ensures that TB patients are able to complete the treatment without socioeconomic threatenings on them or their families. After searching for evidence on the impacts of Social protection and UHC on TB-related outcomes, I have identified that ex-ante and ex-post evaluations around UHC are both limited and that Social protection might act differently in different settings.

According to the results of Chapter 6 and Chapter 7, I identified that around 40% of patients started their care-seeking at the lowest level of hospitals (Level A, general practice/clinic). However, only about 15% of those hospitals usually cannot provide TB diagnostics. In Chapter 7, the results demonstrated that this mismatch was associated with longer system delays and some SES determinants.

Considering the implementation of UHC, this chapter explores the interventions that increase the coverage of TB diagnostics. Meanwhile, this chapter aims at evaluating the impacts of the interventions on TB-related outcomes using the model built, calibrated, and validated in Chapter 8.

9.2 Methods

9.2.1 Increasing capacity of TB diagnostic services

This chapter considers interventions targeted at the time-to-diagnostics-available (TTDA), the period from initial care-seeking to arriving at a hospital which can provide TB diagnostics. To improve the capacity of hospitals of providing TB diagnostics, I constructed interventions based on four rules. For each rule, I assigned a priority to each hospital that cannot provide TB-diagnostics and assigned capacity from the highest priority under a targeted scaling up. After assigning additional diagnostic capacity, I referred back the care-seeking pathways querying their history of hospital visits. If a care-seeking pathway met a TB-diagnostics hospital early, their TTDA and respective costs were shortened. Then, I parametrised the shortened care-seeking pathways as StSp model in Section 8.2.

Referring back to Chapter 7, including Baseline scenario, I designed three strategies to examine, Rule 1 and Rule 2 considered the results from Chapter 7 while Rule 3 is a nonstrategic approach, which is the theoretical less effective case.

Baseline, No intervention: this rule continues the parameters as before 2019. The index of population ageing, which determines the transmission parameters (β) follows the synthetic population built by Chapter 5. Except for β and population dynamics, none time-varying components are used.

Rule 1, assigning TB diagnostic capacity by regional SES: the analysis in Chapter 7 revealed that ill people with TB-suggestive symptoms living in the areas with low hospital densities were more likely to seek-care at Level A hospitals (general practice, clinics, usually without beds for in-patients). Therefore, this rule assigns the capacity from the hospitals visited in any care-seeking pathways sorted by their regional hospital densities from low to high, and to the other hospitals.

Rule 2, assigning TB diagnostic capacity by individual SES: the analysis in Chapter 7 revealed that the employed (and self-employed) patients usually started from lower-level hospitals and suffered from longer system delays. This rule sorts the hospitals by the numbers of pathways for employed patients that have ever visited the hospitals.

Rule 3, assigning TB diagnostic capacity without a strategy, shuffling: The last rule gives all hospitals random orders. All hospitals have equal weights for the randomisation.

Figure 9.1 demonstrates the flow of the interventions. Starting from the list of hospitals, I firstly filtered out the hospitals with TB diagnostic capacity already and kept the rest for the intervention analysis. Then, select a prioritisation rule and a scale of the intervention. It is worth noting that this chapter uses "Rule" to indicates a type of intervention and "scenario" to indicate a "Rule" with a certain scale. This chapter uses "X% scaling-up under Rule Y" to indicate that sorting hospitals which had no capacity of providing TB diagnostics by Rule Y, and assigning X% of them with the capacity according to the prioritisation. The baseline scenario is always at zero % scaling-up.

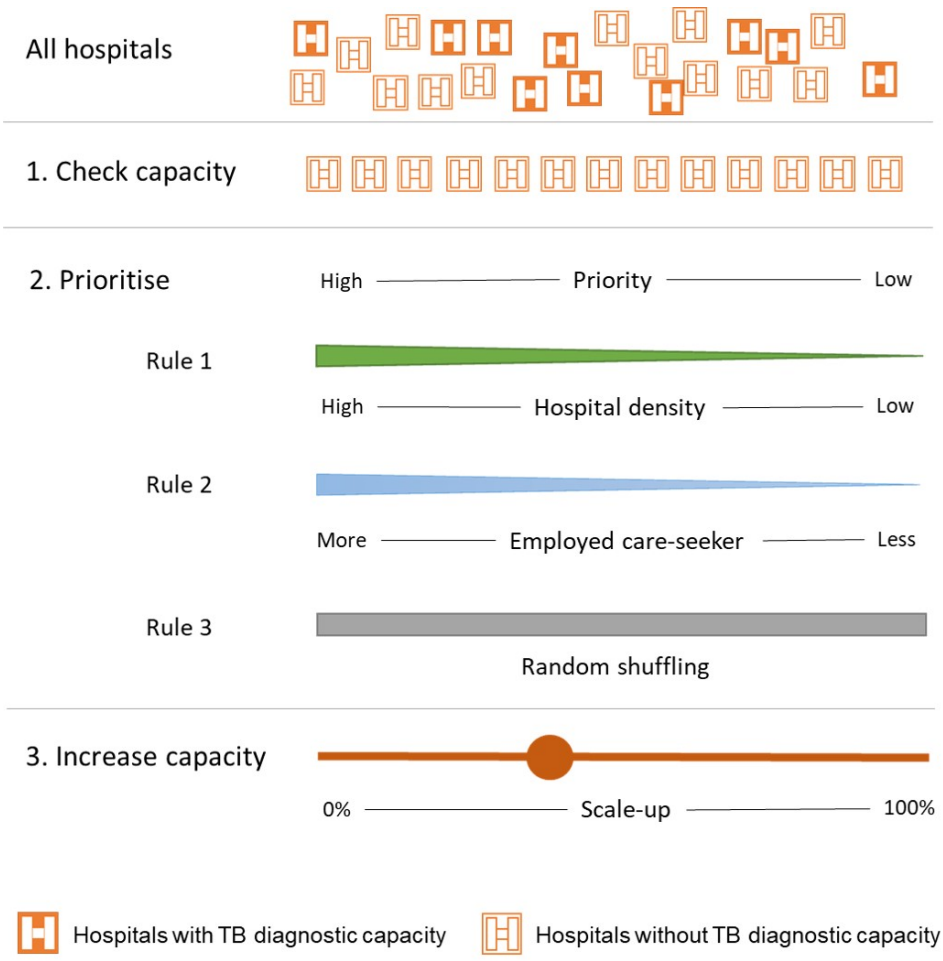


Figure 9.1: The interventions on TB diagnostic capacity. Start with the list of all hospitals. The initial step checks if the hospitals have the capacity for providing TB diagnostics, keeping the hospitals without the capacity. The second step selects a rule for prioritising the hospitals. Then, within the list of remained hospitals, the intervention is simulated by assigning the capacity with a given level of scale-up.

9.2.2 Measurements of effectiveness

Given a StSp with intervened agents, I inputted them to different model builds for different aspects of TB-related outcomes. I evaluated the effectiveness of the interventional scenarios from three perspectives.

Individual-level

From individual-level perspective, I evaluated the days of system delays averted by the interventions. As Chapter 7, I used the Gini coefficients to investigate the inequalities of costs of healthcare system cost, patient cost, and time cost (by the number of visits) before any TB evaluation starts (see Appendix E for definitions). I also considered the evaluation delay (the time difference between initial care-seeking and evaluation initiation) as a measurement of time cost. For rules 1-3, I increased the capacities of providing TB diagnostics, viewing the returns to scale-ups from 10% to 50%.

Population-level, cohort perspectives

Using CSM (Section 8.4), I initialised 500 cohorts (for 500 posterior parameters after calibration (see Section 8.6), simulated them for one year during 2035. Then, I evaluated the changes in the proportions of dropouts. The cohorts were started with 5,000 population at TB onset. I classified the dropouts as self-cure and death.

Population-level, population perspectives

Using the hybrid model constructed in Section 8.6, I simulated the full dynamics from 2000 to 2035 and introduced interventions at the beginning of 2019. I evaluated the epidemiology indices of TB prevalence, incidence, and mortality (TB-specific). I included an index of case detection gaps defined as the non-notified cases divided by the incident cases. Also, I calculated the incidence and mortality rates reduced by the interventions. I calculated the percentage reductions using the results of the baseline. Explicitly, for each parameter set, I inputted the simulations with different interventional scenarios. Then, I compared their results with their counterfactuals with the same parameter set and summarised the corresponding reductions.

9.3 Results

9.3.1 Individual effects on patients

Figure 9.2 demonstrates the costs and evaluation delay compared with the baseline. Generally, Rule 1 and Rule 2 showed similar patterns of effectiveness in the costs and evaluation delay, while Rule 3 usually showed two-thirds the effect of Rule 1 and Rule 2. Conditioning on 50% scaling up, the reductions of the healthcare system and out-of-pocket costs under Rule 1 and Rule 2 will be 25% and 30% respectively, while the visits and evaluation delays can reduce to 45% and 55% respectively. Across the four metrics, Rule 1 and Rule 2 showed diminishing returns to scale.

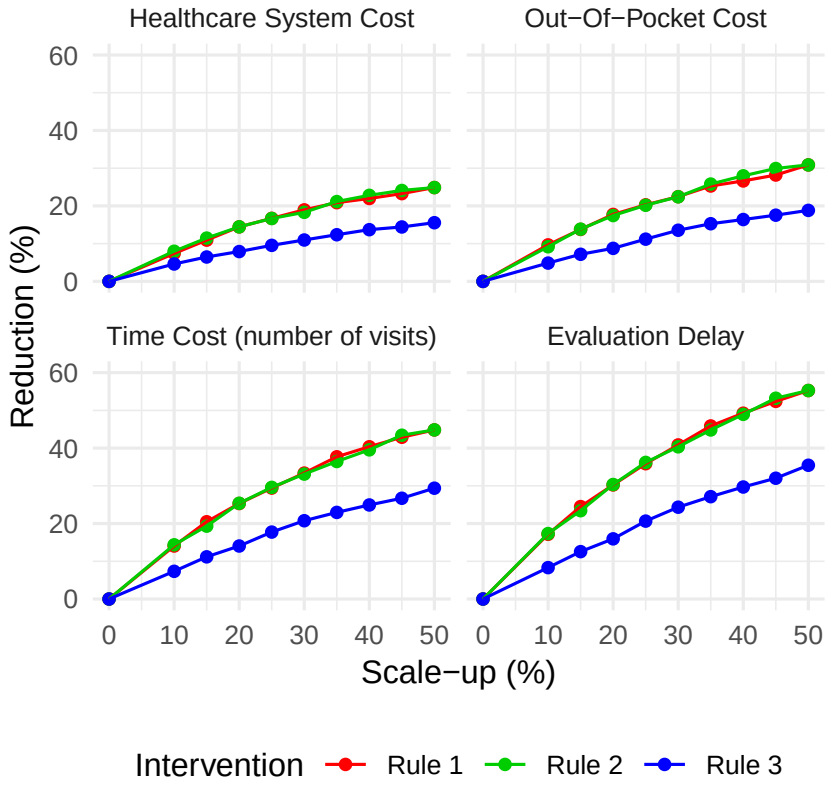


Figure 9.2: Impacts of interventions on the reductions in costs and the length of evaluation delay

Figure 9.3 shows the changes in inequality by Gini coefficients under interventions. For Rule 1-3, the values and trends in the Gini coefficients were close. With the interventions scaling up to 50%, the interventions decreased the inequalities in the time cost of visits but increased in the healthcare system cost, out-of-pocket

expense and evaluation delays.

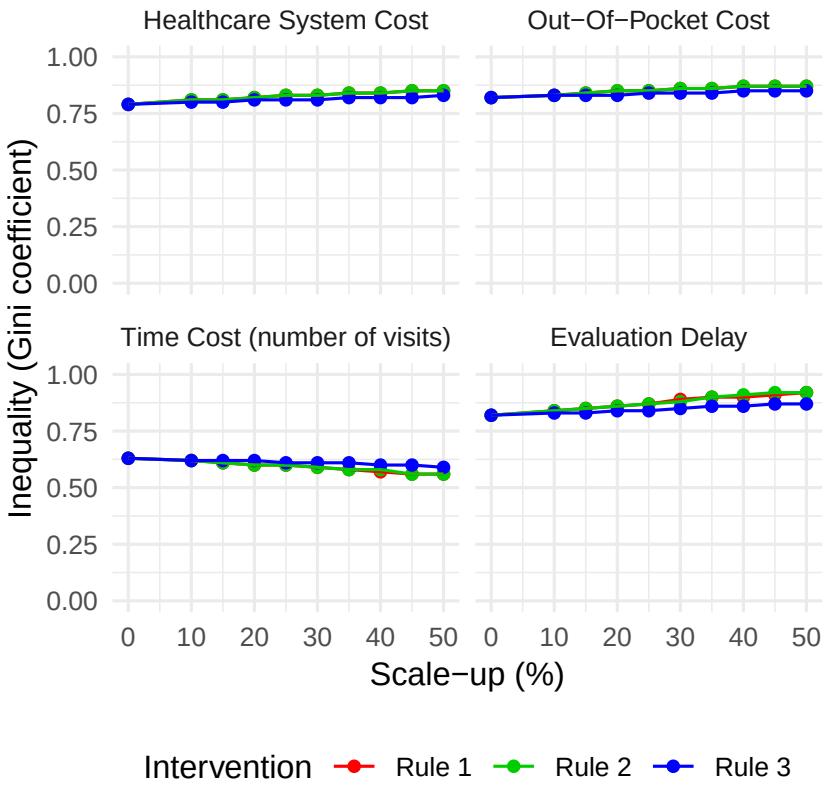


Figure 9.3: Impacts of intervention on the inequality in costs and the length of evaluation delays

9.3.2 Population effects of dropouts from cohort perspectives

Table 9.1 shows the dropouts, death and self-cure, between initial care-seeking and treatment initiation under interventions with 30% scale-up using the 2035 cohorts started with symptom onset. Under the baseline, the dropouts due to death will be 1.1% (95% PI: 0.7%-1.7%) and 2.4% (95% PI: 2%-2.9%) for the 3rd and 12th months respectively while the self-cured will be about three times as deaths. Compared with the baseline, Rule 1 will prevent about 0.7% self-cured and 0.2% before treatment initiation at the 12th month. The results of Rule 1 and Rule 2 will be alike. For Rule 3, the results did not show improvement in dropouts compared with the baseline.

Table 9.1: Dropouts under interventions with 30% scaling-up in 2035 cohort

Dropout	3rd month mean (95% PI)	6th month mean (95% PI)	9th month mean (95% PI)	12th month mean (95% PI)
Baseline				
Self-cure	3% (2.1%, 4.2%)	5.1% (4.3%, 6.2%)	6.1% (5.3%, 7%)	6.6% (5.8%, 7.4%)
Death	1.1% (0.7%, 1.7%)	1.9% (1.4%, 2.5%)	2.3% (1.8%, 2.8%)	2.4% (2%, 2.9%)
Rule 1 (30%)				
Self-cure	2.7% (1.9%, 3.8%)	4.6% (3.8%, 5.6%)	5.5% (4.7%, 6.3%)	5.9% (5.1%, 6.7%)
Death	1% (0.6%, 1.5%)	1.7% (1.3%, 2.2%)	2% (1.6%, 2.5%)	2.2% (1.8%, 2.6%)
Rule 2 (30%)				
Self-cure	2.7% (1.9%, 3.8%)	4.7% (3.8%, 5.7%)	5.5% (4.8%, 6.4%)	5.9% (5.2%, 6.8%)
Death	1% (0.6%, 1.5%)	1.7% (1.3%, 2.2%)	2% (1.6%, 2.5%)	2.2% (1.8%, 2.6%)
Rule 3 (30%)				
Self-cure	3% (2.1%, 4.2%)	5.2% (4.3%, 6.3%)	6.3% (5.4%, 7.3%)	6.8% (6%, 7.7%)
Death	1.1% (0.7%, 1.6%)	1.9% (1.5%, 2.5%)	2.3% (1.9%, 2.8%)	2.5% (2.1%, 3%)

PI: prediction interval

9.3.3 Population effects of TB epidemiology

From the viewpoint of population perspectives, Table 9.2 shows the trends of TB incidence, prevalence, mortality, case-detection gaps. Up to 2035, the TB incidence rates per 100,000 will be 18.8 (95% PI: 11.8-47.7), 18.3 (95% PI: 11.8-45.2), 18.4 (95% PI: 11.8-45.1), and 19.1 (95% PI: 11.9-50) by the baseline and Rule 1-3 respectively; the TB mortality rates per 100,000 will be 0.9 (95% PI: 0.4-2.4), 0.6 (95% PI: 0.2-1.4), 0.6 (95% PI: 0.3-1.6), and 0.7 (95% PI: 0.3-1.8) respectively. For the case detection gaps in 2020, Rule 1 and Rule 2 showed lower rates of 7.7% (95% PI: 2.4%-14%) and 7.8% (95% PI: 3.9%-12.9%) compared with the baseline, 10.8% (95% PI: 5.1%, 16.8%). However, the gaps will increase by 2035 to 13% (95% PI: 6.3%-20.1%) and 13% (95% PI: 9.3%-16.6%) since 2020, being close to the baseline, 14.1% (95% PI: 8.5%-20.3%).

Table 9.3 lays out the differences between interventions with 30% scale-up compared with the results of the baseline. For Rule 1 and Rule 2, the differences in TB incidence rates will increase with the calendar year from interval estimators overlapping with zero in 2020 and 2025 to 1.7% (95% PI: 0.25.6%) and 1.7% (95% PI: 0.2%-5.5%) for Rules 1 and 2 in 2035, respectively. However, Rule 3 will not cause significant changes in TB incidence across the time frame. As for TB mortality, Rule 1-3 will have reductions without time trends, which will be around 40% for Rule 1 and Rule 2, 30% for Rule 3. The interval estimators showed that uncertainty in the TB mortality reduction would be increasing with time for Rule 1 and Rule 2.

Table 9.2: TB epidemiology under interventions with 30% scaling-up

Year	Incidence per 100,000 mean (95% PI)	Prevalence per 100,000 mean (95% PI)	Mortality per 100,000 mean (95% PI)	Case detection gap per cent mean (95% PI)
Baseline				
2020	42.2 (22.5, 135.8)	14 (6.2, 48.6)	2.2 (0.9, 7.6)	10.8 (5.1, 16.8)
2025	30.1 (16.5, 103.9)	9.9 (4.6, 37.6)	1.6 (0.6, 5.6)	12 (6.2, 17.6)
2030	22.8 (13.6, 72.1)	7.5 (3.9, 24.7)	1.2 (0.5, 4)	13.1 (6.7, 19.1)
2035	18.8 (11.8, 47.7)	6.1 (3.3, 15.6)	0.9 (0.4, 2.4)	14.1 (8.5, 20.3)
Rule 1 (30%)				
2020	42.4 (22.6, 138)	13.4 (6.1, 47.4)	1.4 (0.5, 5)	7.7 (2.4, 14)
2025	29.7 (16.5, 101.1)	9.3 (4.3, 35.1)	0.9 (0.3, 3.8)	10.7 (4.6, 16.4)
2030	22.3 (13.5, 68.5)	6.9 (3.5, 22.4)	0.7 (0.2, 2.6)	12 (5.8, 18)
2035	18.3 (11.8, 45.2)	5.6 (3.2, 14.1)	0.6 (0.2, 1.4)	13 (6.3, 20.1)
Rule 2 (30%)				
2020	42.3 (22.6, 136.8)	13.3 (6.2, 47.3)	1.4 (0.6, 5.1)	7.9 (4, 13)
2025	29.7 (16.5, 101.2)	9.3 (4.3, 35.2)	1 (0.4, 3.8)	10.6 (6.9, 14.5)
2030	22.3 (13.5, 68.5)	6.9 (3.6, 23.1)	0.7 (0.3, 2.6)	11.9 (7.9, 15.4)
2035	18.4 (11.8, 45.1)	5.6 (3.3, 14.4)	0.6 (0.3, 1.6)	13 (9.3, 16.6)
Rule 3 (30%)				
2020	42.4 (22.6, 137.2)	14.3 (6.6, 49.9)	1.5 (0.6, 5.4)	11.9 (8.2, 16.9)
2025	30.5 (16.5, 106.1)	10.4 (4.8, 39.8)	1.1 (0.4, 4.3)	12.5 (8.7, 16.6)
2030	23.2 (13.6, 74.2)	7.8 (4, 26.4)	0.8 (0.4, 2.9)	14 (9.8, 17.7)
2035	19.1 (11.9, 50)	6.4 (3.6, 16.5)	0.7 (0.3, 1.8)	14.9 (11.1, 18.5)

PI: prediction interval

Table 9.3: TB incidence and mortality reductions under interventions with 30% scaling-up

Year	Reduction (%)	
	Incidence	Mortality
Rule 1 (30%) - Baseline		
2020	-0.3% (-2.1%, 1.5%)	36.6% (3.6%, 63.9%)
2025	1% (-0.4%, 3.5%)	38.9% (5.7%, 68.4%)
2030	1.6% (0.1%, 5%)	38.6% (-8.6%, 73.6%)
2035	1.7% (0.2%, 5.6%)	37.6% (-12%, 73.6%)
Rule 2 (30%) - Baseline		
2020	-0.2% (-1.5%, 1.2%)	37% (14.8%, 56.9%)
2025	1.1% (-0.1%, 3.2%)	38.1% (10%, 59.6%)
2030	1.6% (0.2%, 4.6%)	38.3% (7.7%, 62.8%)
2035	1.7% (0.2%, 5.5%)	37.4% (0.4%, 63.2%)
Rule 3 (30%) - Baseline		
2020	-0.2% (-1.6%, 1.1%)	32.8% (10.7%, 52.9%)
2025	-0.9% (-2.9%, 0.2%)	29.8% (1.9%, 53.4%)
2030	-1.1% (-3.6%, 0%)	28.5% (-8.5%, 52.6%)
2035	-1.1% (-3.8%, -0.1%)	28.4% (-7.8%, 55.9%)

PI: prediction interval,
reduction: $(intervened - baseline)/baseline \times 100\%$

9.4 Discussion

Chapter 2 reviewed that most of SES-based TB interventions were social protection for TB patients to improve their adherence to TB treatment. However, before TB confirmation or treatment start, patients with TB were not yet TB patients, so those interventions would not protect them for unfavourable outcomes. Therefore, this chapter focused on increasing the coverage of TB diagnostics provision. I examined three rules of empowering hospitals to provide TB diagnostics. In summary, based on the setting of Taiwan, the suggestions are as follows.

- Ranking the priority of hospitals to acquire TB diagnostics by TB patients' healthcare visits, weighted by either the SES of patients or the regional SES of hospitals, to effectively improve care-seeking processes.
- The interventions can reduce the costs before treatment, but might increase the inequality in costs of the healthcare system and out-of-pocket expenses. The reductions in costs will exhibit diminishing returns-to-scale for the interventions
- Increasing TB diagnostics coverage will not have effects in the short run on TB incidence but can have long term effects on it.

From a methodological viewpoint, this chapter demonstrated the use of hybrid modelling with individual-level details. Specifically, my interventions were built by matching of hospital details and aspects of the care-seeking history. It used strong connections between data and models in intervention analysis rather than just changing the values of parameters. Since the hybrid model preserved the details of care-seeking processes and was able to project population interventions, I can evaluate the impacts of interventions from multiple perspectives at different scales. Also, since hybrid modelling efficiently used computational resources (see Chapter 3), I could compute interval estimators using Monte Carlo experiments. My analysis confirms the usefulness of hybrid modelling and encourages more data-driven modelling approaches in intervention analysis.

I started the investigation of SES-based interventions with UHC approaches, while there were other potential interventions can be considered according to the results of Chapter 7. Considering the loss to follow-up (LTFU) during treatment, the underprivileged patients were more likely to be LTFU. In fact, my analytic framework can assess interventions to decrease LTFU by numerically switching the LTFU rates of the underprivileged to that of the other patients. However, I suggested investigating the intervention on adherence as future work after a study surveying the reasons for LTFU.

A modelling assessment highlighted the patient delay (pre-hospital care-seeking) is a key for TB control in high burden settings [1]. The active case-finding (ACF) and reducing the anxiety due to healthcare fee might improve the care-seeking. However, my modelling results (Chapter 8) showed the importance of periods between initial care-seeking and treatment start, i.e. system delay. The interventions on system delay remain underexplored. Based on the modelling, ill people were willing to seek care, but they cannot immediately receive TB diagnostics after initial care-seeking. Competing diseases lies a difficulty. As Chapter 5 showed, 55% of incident TB are aged 65 and above; one-third are male. The population features overlap with many illnesses have TB-suggestive symptoms such as chronic lung diseases and pneumonia. Considering other diseases may be more deadly, physicians might not put TB as the top concern. Lower awareness due to lower TB incidence is another difficulty in reducing system delay.

I considered the ACF which were routinely performed in Taiwan, such as community-based ACF for the impoverished or remote areas and workspaces as the background dynamics in my model. Therefore, it is difficult to consider changes to these interventions using my model. In order to explore a new SES-based ACF, further research could find an alternative parametrisation for patient delays and system delays, allowing some ill people or asymptomatic TB to access TB treatment under ACF.

I did not consider hospitals in relation to nearby hospitals. As I intervened on the hospitals by their regional SES or that of their patients, hospitals which were geographically near might receive similar priorities. Therefore, the demands of TB services for some pathways might be double-counted. I suggest that further investigation can introduce dynamic programming to sort the remaining hospitals after each hospital acquires the capacity of diagnosing TB. That is, considering the capacity among the nearby hospitals in the prioritisation.

In practice, to empower smaller-scale hospitals without expanding the scales, the system needs to incorporate institutions with laboratories or larger hospitals. The cost of this incorporation includes not only the cost of new tasks. As Datiko and Lindtjørn [2] discussed, expanding healthcare services, such as sputum collection, which were not the routine for a provider, the specimen might be in lower quality. Moreover, the transportation infrastructure and cost required further investigation. For the same scale of each scenario, the cost would grow differently due to their geographical distances. These extra-costs, which I cannot evaluate with the data in this study, are the main reason why I did not conduct the cost-effectiveness analysis for the interventions. Thankfully, the care-seeking processes in my model was an individual approach; my model has the potential to deal with

the spatial details if corresponding data are available.

Bibliography

[1] Vesga JF, Hallett TB, Reid MJA, Sachdeva KS, Rao R, Khaparde S, et al. Assessing tuberculosis control priorities in high-burden settings: a modelling approach. Lancet Glob Health. 2019 May;7(5):e585–e595.

[2] Datiko DG, Lindtjørn B. Health extension workers improve tuberculosis case detection and treatment success in southern Ethiopia: a community randomized trial. PLoS One. 2009 May;4(5):e5443.

Chapter 10

Discussion

Tuberculosis has long been identified as a disease of poverty [1]. In the post-2015 era, tuberculosis control programmes have been widely calling for socioeconomic actions and are targeting the elimination of catastrophic economic outcome [2–4]. Working within the context of Taiwan, this thesis departs from the empirical analysis of socioeconomic status (SES) and care-seeking, to bridge the evidence related to tuberculosis epidemiology with simulation models, and to provide an innovative approach for SES-based tuberculosis controls. In this chapter, I discuss the existing gaps in the knowledge, the empirical/modelling inference I addressed, and my methodological contributions as follows.

10.1 Knowledge and technical gaps before this thesis

While the association between TB and SES has been established in the existing literature, the evidence of the relationship between care-seeking of TB patients and socioeconomic determinants is somewhat disparate, as was seen in Chapter 2. My review of assessments of two prominent forms of SES-based interventions for TB control, UHC and social protection, found that most of the approaches were focused on TB treatment adherence. Meanwhile, interventions for care-seeking prior to treatment initiation remained largely unexplored.

Patient pathway analysis (PPA) [5] is an approach used to gain a better understanding of the alignment of TB services provision and healthcare visits. PPA has been used to explore in numerous studies to assess the treatment of TB and has been used to support efforts to improve treatment adherence. To the best of my knowledge, PPA mainly utilises cross-sectional data, which does not include temporal information. However, to fully understand the duration of untreated stages and their impacts on TB epidemiology, a method that incorporates temporal pat-

terns in PPA is much needed.

Care-seeking pathways are a combination of the decision making of patients and healthcare providers. To enrich the inference on behaviour changes, I reviewed the approaches to the simulation of health-related behaviour in agent-based models (ABMs) in Chapter 4. While ABMs that incorporate health behaviour demonstrated methods for driving agents using empirical data and behaviour theories, applications of care-seeking behaviours are limited.

Hybrid modelling presents a parsimonious way to model a system by using a combination of agent-based models for the most complex features and equation-based models for other supportive processes. Chapter 3 searched for health-related applications of hybrid modelling and, while only a few examples were found, they were largely heterogeneous. In terms of the technical aspects, synchronising the equation-based parts and the agent-based parts as well as describing the communication among them remained a challenge.

10.2 Empirical findings and modelling inference

10.2.1 Demographic changes and TB burden (Chapter 5)

Based on the notification rates of TB in Taiwan, I forecasted that the age-specific notification rates will have decreased by around 70% by 2035 for all age groups apart from children. However, the higher rates of TB among the elderly may have a detrimental effect on this decline. It is estimated that 76% of new TB cases will be aged 65 and above in 2035. Furthermore, the results suggested that 39% of the TB notifications in Taiwan in 2035 could be attributed to demographic change, which highlights how the attendant policies should be more age-specific in terms of the older age groups and their care needs.

10.2.2 Alignment between TB care provision and demand (Chapter 6)

Using the routine health insurance claim data in Taiwan, I found a general trend for escalation in level of care for initial care-seeking, with 43% of patient pathways started at the lowest level hospitals (clinics or general practices) while only a quarter of them (about 10% of all pathways) received treatments without being referred to higher levels. Unfortunately, only around 15% of hospitals at that level could provide TB diagnostics or treatment. The median delay between arriving at the facility where they were treated, and treatment initiation was 6 days. As for the delay to treatment, initiation was 41 days after initial care-seeking in the median. These results highlighted two care-seeking barriers before TB treatment:

availability of TB services at low-level hospitals and low attention to TB in the hospitals which can handle TB.

I found huge heterogeneity in pathways driven by the interruption of evaluation and re-evaluation: 10% patient pathways had system delays longer than 200 days, and about 75% of them encountered interrupted TB evaluations. Understanding the causes of interrupted evaluation should be a focus of further investigation in Taiwan.

10.2.3 SES factors as risks of inefficient care-seeking (Chapter 7)

I analysed the associations of individual/regional SES and features of the TB care-seeking pathways in Taiwan. Apart from age, which commonly led to extended system delays, I identified some SES factors that affected the lengths of delays. I found that unemployment and living in a more impoverished area or in a low-income household were associated with initial care-seeking in higher-level hospital levels and quicker evaluation processes. Although the underprivileged had shortened the system delays, they had higher chances of unfavourable treatment outcomes. I suggested two hypotheses. First, their shorter delays were due to patients being "fast-tracked" through the care-seeking pathway due their presentation in poorer states of health at their initial care-seeking, which is a result of longer delays before seeking care. The second hypothesis was that they were suffered from extreme costs other than healthcare fee, such as travelling costs or income loss due to care-seeking.

there was less support available, other than healthcare provision and access, for example removing financial barriers by granted income due to care-seeking or create supportive communities through health education.

10.2.4 Inequality in care-seeking (Chapter 7)

Using the Lorenz curve [6] and the Gini coefficient [7], I assessed the inequality in costs to the healthcare system, the out-of-pocket expenses of patients and the time costs, which were measured by the number of visits. While the costs were diverse among the patients; a small group of patients were faced with huge costs. For the out-of-pocket expenses, 90% of the costs were attributed to 20% of the patients with the highest out-of-pocket expenses. Meanwhile, I confirmed that the underprivileged patients were protected from out-of-pocket expenses by the national policy. However, this group were not protected from the time costs. Indeed, the number of visits before treatment initiation was homogeneous across the social groups, which implies that the costs of healthcare visits, such as travelling and

time costs, of the underprivileged patients may not differ from the others.

10.2.5 Age-specific mortality estimation during treatment (Chapter 8)

According to the extracted patient pathways, I estimated the TB-specific death rates by age-group. After adjusting for the crude death rates by age and sex for every individual, the TB patients aged 65 and above were found to have a higher TB-specific death rate (0.18 per person-year), with the rate potentially being higher than that of the untreated. Following the onset of symptoms of TB, the chance of being diagnosed increases due to care-seeking while the health status worsens. According to the patient flows from outpatient to inpatients (Figure 6.4), about 40% of patients initiated TB treatment as inpatients, but there were about 15% of inpatients at initial care-seeking. This suggested that the health status of patients were decaying during care-seeking. This is also a vital issue to reduce delay to TB treatment as early as possible.

10.2.6 Dropout cascade from cohort and population perspectives (Chapter 8)

This thesis explored dropouts due to death and self-cure before treatment in Taiwan. Roughly speaking, 10% of the incident TB remained undiagnosed due to death or self-curation. Especially for the old adults, the dropouts can reach as high as 17% or over due to higher background death rates and the extended delays to treatments. From the cohort perspective, I expect that the dropout profiles of 2019 and 2035 would not reveal significant differences. Meanwhile, from a population perspective, the case-detection gap, which may be affected by the annual decline in incidence rates, was found to be decreasing since 2005 but is expected to start increasing again after 2020. This mismatch should be considered in setting up the goals of NTPs, which generally use the population approach to identify case-detection gaps.

10.2.7 The impact of providing TB-service capacity (Chapter 9)

I examined the impact of the interventions related to providing the capacity for diagnosing TB for the hospitals that lacked it (most of them were lower level hospitals). Increasing service provision at the hospitals where TB patients with extended delays were likely to initiate their care-seeking could decrease the dropout rates before treatment and could have long-term impacts on the TB incidence and mortality. Here, I illustrated the impact of specific decentralisation schemes related to the provision of TB diagnostics.

10.3 Methodological contributions

10.3.1 Lee-Carter models in age-specific incidence forecasting (Chapter 5)

I introduced LCMs [8], which were initially designed for mortality modelling, in the modelling and forecasting of the age-specific TB notification rates. I argued that the death rates and notification rates have similarities in terms of age-specific patterns and time trends. The goodness of fit and the residual analysis also confirmed these arguments. In fact, while LCMs have been used for decades in population modelling, to the best of my knowledge, this is the first time they had been applying in infectious disease modelling. Moreover, I combined the notification and population forecasts to analyse the age profile of future TB notification. This method helped us to explore the future disease burden and the difficulty of TB control due to the population ageing issue.

10.3.2 Individual patient pathway analysis (Chapter 6)

One key contribution of this thesis lies in developing the framework of the individual patient pathway analysis (IPPA). The IPPA is a patient-centred approach that centres on the event history during care-seeking. The IPPA combines the concept of patient pathway analysis and delays analysis in addressing the care-seeking process of TB patients. Above all, it has the potential to the impacts of upstream determinants, especially SES, and provide a rich source of new information for locating the weaknesses of a given healthcare system.

The results I presented focussed on the healthcare system in Taiwan. However, the IPPA spans three dimensions of events: pre-evaluation, evaluation, and treatment. This provides a systematic procedure for defining events from a complicated space to generate specific patient pathways in the IPPA, which can be adapted to different definitions in accordance with different settings and data availability. I released the analysis package for general cases on a GitHub public repository (see Appendix J).

10.3.3 Hybrid modelling with individual care-seeking processes and population TB dynamics (Chapter 8)

I developed a hybrid simulation model that can capture nationwide TB epidemiology in Taiwan with the details of the care-seeking process. The hybrid simulation model synthesises the data related to TB service provision, demands and population dynamics. Overall, it provides a platform for performing the intervention

evaluation as a practice of prospective health technology assessment. The model incorporates the advantages of data integration pertaining to ABMs and the reductionism pertaining to EBMs. The simulation can be executed using a personal computer or laptop with a tolerable computation time for each run.

10.3.4 Model-based dropout profile (Chapter 8)

This thesis provides a model-based dropout assessment. Going beyond the dropout cascade that is entirely based on empirical data (e.g. Subbaraman et al. [9]), my approach complemented the components that do not have proper supportive data. For example, the period between the onset of TB symptoms and the initial care-seeking can be surveyed in terms of TB patients' pre-hospital care-seeking. However, starting a cohort with those who have not initiated their care-seeking is not possible due to ethical issues. Given the care-seeking pathways, I used a dynamic transmission model calibrated to the epidemiological data to infer the data-scarce components. My model-based approach subsequently captured interrupted evaluations and LTFUs as well as their relative recurrence. Moreover, the modelling approach can project the dropout cascade used in both population-based results (e.g. Table 8.7) and cohort-based results (e.g. Figure 8.9). In short, it provided a method for comparing dropout cascades in other settings with different types of available data.

10.3.5 Analysis codes and packages

Along with compiling this thesis, I created four packages for the analyses and the simulation modelling in addition to numerous related code repositories for visualisation and presentations. I published the codes on Github. The full list of codes and link to Github can be found in Appendix J.

For conducting an appropriate IPPA (see Chapter 6), I provided the requisite full codes for it. Also, I published pseudo-data for testing the codes and for guiding the data collection and management of further studies.

For profiling the age distribution of TB (see Chapter 5), I released the code for statistically modelling and forecasting the age-specific TB incidence data under the LCM [8]. Along with this package, I published a processed data set as an example for investigating the impacts of population ageing on TB epidemiology.

For managing the parameter structure and the Bayesian modelling, I developed "PyEpiDAG", which is a probabilistic modelling toolkit based on causal diagrams [10]. This package can incorporate probability distributions, empirical models from prior studies, and user-defined mathematical functions to support

the parameterisation of simulation models. Furthermore, this package provides the functions of modelling fitting in both Frequentist and Bayesian approaches. The package uses the Python programming language, while a Java equivalent is also available.

For the simulation modelling, I developed a package, "PyCx", which is a simulation modelling toolkit. This package can deal with agent-based, equation-based, and hybrid modelling. The package is compatible with PyEpiDAG and can apply the model-fitting functions provided by PyEpiDAG. All the simulation models in this thesis were constructed using this package.

10.4 Limitations

10.4.1 Loss to follow up

Given the fact that the incidence of drug-resistant TB among all incident TB in Taiwan accounted for approximately 1% or less, the 0.2 LTFU rate per person-year from the IPPA results would appear not to make sense. In an unreported result, I found that only 10% of the LTFUs were pathways completely lost in the tracking, while the remaining 90% still had healthcare events with on-going TB evaluation. I suggested three hypotheses for this mismatch. The first is related to the competing priority in relation to other health conditions. As more than half of TB patients in Taiwan were aged 65 and above, their health conditions were likely to be more complex. Here, I expected that many treatments were temporarily interrupted due to physiological reasons. Meanwhile, my second hypothesis is that their treatments were started due to false positive results in the previous assessments or because the patients were latent TB positive but were considered as active TB patients (the NHI does not cover latent TB treatments). Finally, for patients with special conditions, their treatments might be provided by the TCDC but not the NHI. Thus, these cases were not necessarily LTFU; rather the NHI database simply did not have them on record. As the TCDC supervises all the TB patients once it is notified, the actual treatment outcome is always tracked. However, since my analysis did not link the TCDC database to the validated treatment outcome data, I could not test these three hypotheses.

In the simulation modelling of this thesis, I did not pay a great deal of attention to the outcomes of LTFU. In my dropout analyses, the TB patients were either cured or had ultimately died, while the patients of LTFU will always have a chance to seek care again. For simplicity, I did not assume that the patients with treatment outcomes of LTFU have a rate of care-seeking that is different from those who have not yet sought care. Further study is required for addressing the influence of LTFU

with data of explicit health status.

In the reality, irregular use of anti-TB drugs may cause drug-resistance. I argue that the modelling did not significantly distort the reality because the drug-resistant TB in Taiwan only shared around 1% of overall TB incidence.

10.4.2 Immigration

In my modelling, while I did consider the migration aspect, I only aimed to simulate the age-sex distributions compatible with the real data of population dynamics. I assumed the distributions of latency among the immigrants and the emigrants were the same as the nationals of Taiwan. However, according to the last-decade national statistics pertaining to Taiwan, around 60% of immigrants were from China, while 20% were from Vet Nam, and 5% were from Indonesia. Meanwhile, the latest TB incidence rates of these countries were estimated to be approximately 1.5, 3, and 9 times that of Taiwan (38 per 100,000 as of 2018), respectively.

10.4.3 Generalisability

In this thesis, I used the setting of Taiwan as an example for profiling care-seeking and for constructing a hybrid model. I considered the models pertaining to individual-level data and the epidemiology of TB. However, it may be challenging to apply the study frameworks of this thesis to other settings.

In Taiwan, the public and private healthcare systems are fully integrated by the NHI, which provide identical coverage of health services for both sectors. The data of the NHI are also collected and encoded with the same standard for public and private sectors. For the settings with data qualities and healthcare coverages of private and public sectors were significantly different, standardising data before conducting an IPPA is needed.

Moreover, the TB notifications system always synchronises with the anti-TB drug prescriptions under the NHI. That is, the notification of TB happens immediately after prescribing anti-TB drugs. According to that, throughout this thesis, I assumed TB treatment initiation and TB notification were identical in timing. However, for scarce resource setting, availability of the data as well as any notification inefficiency should be considered.

10.5 Future work

10.5.1 Combining patient cost surveys with IPPA

This thesis highlighted that the out-of-pocket expenses for patients was prevented by the social welfare policy, the number of visits before treatment initiation for the underprivileged was found not to be lower than that of the other groups. To clarify the sources of costs incurred by patients, one could combine cost surveys with my individual patient pathway analysis. In fact, a key target of the "End TB strategy" [2] is to reduce the catastrophic costs related to TB care to zero on TB-affected households.

10.5.2 Exploring innovative SES-based interventions

This thesis built a hybrid simulation model as a flexible environment for testing intervention on care-seeking behaviours and for measuring the impacts from the individual level to the population level. However, I only performed two types of actions in my interventions analysis, while many other intervention implementations could have been considered. For example, with more detailed spatial information, I could experiment with community-based approaches such as Datiko and Lindtjørn [11] to improve early diagnosis and to increase the risk perceptions about TB. However, before evaluating more interventions, it is essential to have further discussions with stakeholders in order to prioritise those most acceptable and those most relevant to the policy.

10.5.3 Enhance the hybrid modelling and employ in healthcare modelling beyond TB

It should be noted that my hybrid modelling extends a detailed view of the care-seeking processes while incorporating transmission dynamics for projecting the epidemiology at a national level. My approach combined the advantages of ABMs and EBMs, and the simulation results stand somewhere between the reductionism of EBMs and the resolution of ABMs. However, in going from a complex to a simplified system, some information loss must be expected. I did not address precisely how much information and stochasticity was lost during the hybrid modelling. If I consider the full ABM model to be on the right side of the spectrum and the mean-field EBM to be on the left, the next step is to generate measurements that will allow us to locate the hybrid model on the spectrum. This will help inform the decision making based on hybrid models.

Furthermore, although the hybrid modelling was developed for investigating care-seeking related to TB, the frameworks were not actually TB-specific. For example, pneumonia patients, who may share similar features of pulmonary TB, may have care-seeking processes similar to those of patients with TB. In fact, this is especially the case in Taiwan, where risk groups often tend to overlap as well. Therefore, my model could be applied in terms of pneumonia, such that I could address similar questions without major adaptations. This statement is also true for some chronic lung diseases. Indeed, considering the difficulty in diagnosis and the chronic features of TB (e.g. latent progression), it is worth noting that the modelling scheme is applicable to other diseases that have complex progression and diagnostic processes.

Bibliography

[1] Stevens P. Diseases of poverty and the 10/90 gap. International Policy Network, http://who int/intellectualproperty/submissions/InternationalPolicyNetwork pdf. 2004;.

[2] World Health Organization. The End TB Strategy. Global strategy and targets for tuberculosis prevention, care and control after 2015; 2014.

[3] Assembly G. Sustainable development goals. SDGs), Transforming our world: the. 2015;2030.

[4] Reid MJA, Arinaminpathy N, Bloom A, Bloom BR, Boehme C, Chaisson R, et al. Building a tuberculosis-free world: The Lancet Commission on tuberculosis. Lancet. 2019 Mar;393(10178):1331–1384.

[5] Hanson C, Chin D, Osberg M, Brown J, Durham G. Conducting patient pathway analysis to inform tuberculosis programming: methodology. Journal of Infectious Disease. 2017;216(Suppl 7):S679–85.

[6] Lorenz MO. Methods of Measuring the Concentration of Wealth. Publications of the American Statistical Association. 1905 Jun;9(70):209–219.

[7] Gini C. Concentration and dependency ratios. Rivista di politica economica. 1997;87:769–792.

[8] Lee RD, Carter LR. Modeling and Forecasting U. S. Mortality. J Am Stat Assoc. 1992;87(419):659–671.

[9] Subbaraman R, Nathavitharana RR, Mayer KH, Satyanarayana S, Chadha VK, Arinaminpathy N, et al. Constructing care cascades for active tuberculosis: A strategy for program monitoring and identifying gaps in quality of care. PLoS Med. 2019 Feb;16(2):e1002754.

[10] Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. Epidemiology. 1999 Jan;10(1):37–48.

[11] Datiko DG, Lindtjørn B. Health extension workers improve tuberculosis case detection and treatment success in southern Ethiopia: a community randomized trial. PLoS One. 2009 May;4(5):e5443.

Appendix A

Reviews: behavioural modelling in agent-based models, paper list

A.1 Definitions and notations

In this appendix, the papers were grouped in terms of type of behaviour, and were summarised with explicit behaviours they modelled, data type of behaviour variable in model, critical function triggers final action, and function for collecting determinants. The notations and definitions used for the critical functions and collection functions are listed as follows.

Critical function: a function which triggers final action. Y indicates a final action (behaviour) of a model

Variable, Y $\leftarrow \mu$**:** an action is assigned directly by output of determinant collection.

Random variable, Y $\sim$ distribution(...)**:** an action is sampled from a probability density (mass) function and the parameters are sourced from determinant collection. The distribution will be a specific distribution if possible. Otherwise, distribution indicates a mixture distribution or other sampling rule.

Threshold, Y $\leftarrow$ if $\mu > \tau$**:** an action is triggered when a variable reaches threshold. The process can be replicated for vectorial cases.

Optimisation, Y $\leftarrow$ argmax(u)**:** an action is a result of optimisation.

Game, Y $\leftarrow$ game(u)**:** an action is the optimal option in a game with a pay-off matrix. Pay-off matrix are summary of determinants.

Intermediate variable: a variable of summary by a collection function and it will be the main input of a critical function.

Continuous value, μ**:** a continuous variable, usually the mean of an action.

Utility, u**:** a value of utility of a pay-off matrix.

Probability, p**:** a probability or a conditional probability table.

Collection function: a function which collects determinants of targeted behaviour. (X indicates a set of upstream variables)

Regression or linear combination, u = Xβ**,** p = logit(Xβ)**,** p = probit(Xβ)**, ...:** determinants are summarised by a linear combination function with determinants (covariates) and effects (coefficients). Transformation can be applied (e.g. logit and probit functions)

Mathematical function, μ = f(X)**:** determinants are summarised with a or a series of mathematical function except for regression function.

Algorithmic function, $\mu = \lambda(X)$**:** determinants are summarised with an algorithm, including query within a parameter table (e.g. reading probability in a conditional probability table).

Diffusion, imitation, or Huff function, $\mu = \Delta(X)$**:** behaviour is transmitted from neighbours or agents imitate neighbour behaviour. Determinants are mainly neighbour attributes.

A.2 List of the retrieved studies

A.2.1 Models of lifestyle

For lifestyle (Table A.1), we identified behaviours of food intake, Body mass index (BMI) control, physical activity. 24 eligible models were selected.

Table A.1: Summary of models of lifestyle

Article	Behaviour	Action Type	Action rule	Determinant collection
	Life style choosing			
Auchincloss et al. [6]	Food choice	Categorical	Y $\leftarrow$ argmax(u)	u = Xβ
Langellier [58]	Food choice	Categorical	Y $\sim$ Multinomial(p)	p = logit(Xβ)
Li et al. [62]	Food choice	Categorical	Y $\sim$ Multinomial(p)	p = Δ(x)
Li et al. [61]	Food choice	Categorical	Y $\sim$ Multinomial(p)	p = Δ(x)
Zhang et al. [105]	Food choice	Categorical	Y $\sim$ Multinomial(p)	p = logit(Xβ)
Choi et al. [28]	Food consumption	Continuous	Y $\leftarrow \mu$	μ = Xβ
Li et al. [60]	Food consumption	Binary	Y $\sim$ Bernoulli(p)	p = logit(Xβ)
Beheshti et al. [11]	Food planning	Vectorial	Y $\leftarrow$ if $\mu > \tau$	μ = λ(X)
Blok et al. [14]	Food planning	Vectorial	Y $\leftarrow$ argmax(u)	u = Xβ
Basu et al. [10]	SSB intake	Continuous	Y $\sim$ Norm(μ, ϵ)	μ = Xβ
Chen et al. [26]	SSB intake	Continuous	Y $\leftarrow$ if $\mu > \tau$	μ = Δ(X)
Beheshti et al. [12]	Weight control	Binary	Y $\leftarrow$ if $\mu > \tau$	μ = f(X)
Bruzzone et al. [21]	Weight control	Continuous	Y $\sim$ Norm(μ, ϵ)	μ = Xβ
Edwards and Clarke [34]	Weight control	Continuous	Y $\sim$ Norm(μ, ϵ)	μ = Xβ
Hammond and Ornstein [46]	Weight control	Continuous	Y $\leftarrow \mu$	μ = Δ(X)
Hennessy et al. [48]	Weight control	Continuous	Y $\leftarrow \mu$	μ = Xβ
Trogdon and Allaire [91]	Weight control	Continuous	Y $\leftarrow$ argmax(u)	u = f(X)
Wang et al. [96]	Weight control	Continuous	Y $\leftarrow \mu$	μ = f(X)
Orr et al. [75]	Dieting	Continuous	Y $\leftarrow \mu$	μ = f(X)
Orr et al. [76]	Dieting	Continuous	Y $\leftarrow \mu$	μ = f(X)
Blok et al. [13]	Physical Activity	TTE	Y $\sim$ Exponential(...)	μ = f(Xβ)
Yang et al. [103]	Physical Activity	Categorical	Y $\leftarrow$ argmax(u)	u = f(X)
Yang et al. [104]	Physical Activity	Categorical	Y $\sim$ Multinomial(p)	p = λ(X)
Yang et al. [102]	Physical Activity	Categorical	Y $\sim$ Multinomial(p)	p = logit(Xβ)

A.2.2 Models of substance use

For substance use (Table A.2), we identified behaviours of smoking, alcoholic drinking, addictive drug abuse, and problem gambling. 32 eligible models were selected.

Table A.2: Summary of models of substance use

Article	Behaviour	Action Type	Action rule	Determinant collection
Adams and Schaefer [1]	Smoking	Binary	Y $\sim$ Bernoulli(p)	p = logit(Xβ)
Chao et al. [25]	Smoking	Categorical	Y $\leftarrow$ argmax(u)	u = Xβ
Cherng et al. [27]	Smoking	Binary	Y $\leftarrow \mu$	μ = λ(X)
Schaefer et al. [84]	Smoking	Continuous	Y $\sim$ Norm(μ, ϵ)	μ = Xβ
Sun and Mendez [88]	Smoking	Binary	Y $\sim$ Bernoulli(p)	p = f(X)
Yang et al. [101]	Smoking	Binary	Y $\sim$ Bernoulli(p)	p = f(X)
Getsios et al. [42]	Smoking cessation	Binary	Y $\sim$ Bernoulli(p)	p = logit(Xβ)
Igarashi et al. [51]	Smoking cessation	Binary	Y $\sim$ Bernoulli(p)	p = logit(Xβ)
Xenakis et al. [98]	Smoking cessation	Binary	Y $\sim$ Bernoulli(p)	p = logit(Xβ)
Atkinson et al. [5]	Alcohol drinking	Categorical	Y $\leftarrow \mu$	μ = λ(X)
Fitzpatrick et al. [37]	Alcohol drinking	Continuous	Y $\sim$ Norm(μ, ϵ)	μ = f(X)β
Fitzpatrick et al. [38]	Alcohol drinking	Continuous	Y $\sim$ Norm(μ, ϵ)	μ = f(X)β
François et al. [39]	Alcohol drinking	Continuous	Y $\sim$ Norm(μ, ϵ)	μ = Xβ
French et al. [40]	Alcohol drinking	Continuous	Y $\leftarrow \mu$	μ = f(X)
Giabbanelli and Crutzen [43]	Alcohol drinking	Binary	Y $\leftarrow$ if $\mu > \tau$	μ = λ(X)
Gorman et al. [44]	Alcohol drinking	Binary	Y $\leftarrow \mu$	μ = Δ(X)
Millier et al. [71]	Alcohol drinking	Discrete	Y $\sim$ distribution(μ)	μ = Xβ
Purshouse et al. [78]	Alcohol drinking	Binary	Y $\sim$ Bernoulli(p)	p = probit(Xβ)
Rahhali et al. [79]	Alcohol drinking	Binary	Y $\sim$ Bernoulli(p)	p = logit(Xβ)
Riva and Smith [80]	Alcohol drinking	Categorical	Y $\sim$ Multinomial(p)	p = logit(Xβ)
Scott et al. [85]	Alcohol drinking	Categorical	Y $\leftarrow \mu$	μ = λ(X)
Scott et al. [86]	Alcohol drinking	Continuous	Y $\sim$ distribution(μ)	μ = λ(X)
Zur and Zaric [109]	Drinking level transition	Binary	Y $\sim$ Bernoulli(p)	p = logit(Xβ)
Bobashev et al. [15]	Drug abuse	Binary Y $\leftarrow$ if $\mu > \tau$	μ = λ(X)	
Dray et al. [32]	Drug abuse	Binary	Y $\sim$ Bernoulli(p)	p = f(X)
Gutfraind et al. [45]	Drug abuse	Binary	Y $\sim$ Bernoulli(p)	constant p
Heard et al. [47]	Drug abuse	Categorical	Y $\sim$ Multinomial(p)	p = logit(Xβ)
Hoffer et al. [49]	Drug abuse	Categorical	Y $\leftarrow \mu$	μ = λ(X)
Lamy et al. [57]	Drug abuse	Binary	Y $\leftarrow$ if $\mu > \tau$	μ = f(X)
Perez et al. [77]	Drug abuse	Binary	Y $\leftarrow \mu$	μ = Δ(X)
Xiong et al. [99]	Drug abuse	Binary	Y $\sim$ Bernoulli(p)	p = λ(X)
Markham et al. [67]	Problem gambling	Continuous	Y $\leftarrow \mu$	μ = Δ(X)

A.2.3 Models of preventive actions

For disease prevention (Table A.3), we identified behaviours of vaccination, self-protection, having risky contact, and accessing screening. 35 eligible models were selected.

Table A.3: Summary of models of preventive actions

Article	Behaviour	Action Type	Action rule	Determinant collection
Andrews and Bauch [4]	Self protection	Binary	Y $\leftarrow$ if $\mu > \tau$	μ = f(X)
Barrett et al. [8]	Self protection	Categorical	Y $\leftarrow$ if $\mu > \tau$	μ = f(X)
Barrett et al. [9]	Self protection	Categorical	Y $\leftarrow$ if $\mu > \tau$	μ = f(X)
Durham and Casman [33]	Self protection	Binary	Y $\sim$ Bernoulli(p)	p = logit(Xβ)
Karimi et al. [53]	Self protection	Binary	Y $\sim$ Bernoulli(p)	p = logit(Xβ)
Keane [55]	Self protection	Binary	Y $\leftarrow$ game(u)	u = f(X)
Keane [54]	Self protection	Binary	Y $\leftarrow$ game(u)	u = f(X)
Mao [65]	Self protection	Binary	Y $\leftarrow \mu$	μ = Δ(X)
Mao [66]	Self protection	Binary	Y $\leftarrow \mu$	μ = Δ(X)
Mei et al. [70]	Self protection	Categorical	Y $\leftarrow \mu$	μ = f(X)
Zhong [108]	Self protection	Binary	Y $\sim$ Bernoulli(p)	p = f(X)
Sahneh et al. [82]	Self protection	Binary	Y $\sim$ Bernoulli(p)	p = λ(X)
Buttenheim et al. [22]	Vaccination	Binary	Y $\sim$ Bernoulli(p)	p = f(X)
Collinson et al. [30]	Vaccination	Binary	Y $\leftarrow \mu$	μ = f(X)
Fu et al. [41]	Vaccination	Binary	Y $\leftarrow$ game(u)	u = f(X)
Iwamura et al. [52]	Vaccination	Binary	Y $\leftarrow$ game(u)	u = f(X)
Liu et al. [63]	Vaccination	Binary	Y $\leftarrow$ game(u)	u = f(X)
Salvarani and Turinici [83]	Vaccination	Binary	Y $\sim$ Bernoulli(p)	p = f(X)
Teoh Shian Li et al. [90]	Vaccination	Binary	Y $\leftarrow$ game(u)	u = f(X)
Xue et al. [100]	Vaccination	Binary	Y $\sim$ Bernoulli(p)	p = f(X)
Zhang Fa et al. [107]	Vaccination	Binary	Y $\sim$ Bernoulli(p)	p = f(X)
Nagoski et al. [74]	Sexual risk taking	Binary	Y $\sim$ Bernoulli(p)	p = λ(X)
Tully et al. [92]	Sexual risk taking	Continuous	Y $\leftarrow$ game(u)	u = f(X)
Tully et al. [93]	Sexual risk taking	Continuous	Y $\leftarrow$ game(u)	u = f(X)
White et al. [97]	Sexual risk taking	Binary	Y $\sim$ Bernoulli(p)	p = λ(X)
Bracher et al. [17]	Condom use	Binary	Y $\sim$ Bernoulli(p)	constant p
Hui et al. [50]	Condom use	Categorical	Y $\sim$ Multinomial(p)	p = λ(X)
Delaney et al. [31]	Social distance change	Binary	Y $\leftarrow \mu$	μ = f(X)
Epstein et al. [35]	Social distance change	Binary	Y $\sim$ Bernoulli(p)	p = f(X)
Brown et al. [20]	Social group choice	Binary	Y $\leftarrow$ game(u)	u = f(X)
Marquis and Buchanan [68]	Insurance planning	Categorical	Y $\leftarrow$ argmax(u)	u = f(X)
Sowa et al. [87]	Insurance planning	Continuous	Y $\leftarrow$ argmax(u)	u = f(X)
Suzuki et al. [89]	Insurance planning	Binary	Y $\sim$ Bernoulli(p)	p = logit(Xβ)
Marshall et al. [69]	Voluntary screening	Binary	Y $\sim$ Bernoulli(p)	p = λ(X)
Turner et al. [94]	Accepting screening	Binary	Y $\sim$ Bernoulli(p)	p = λ(X)

A.2.4 Models of hospital access

For patient choice (Table A.4), we identified behaviours of seeking, care attendance, treatment compliance, and taking seek leave. 18 eligible models were selected.

Table A.4: Summary of models of hospital access

Article	Behaviour	Action Type	Action rule	Determinant collection
Alibrahim and Wu [2]	Care-seeking	Binary	Y $\sim$ Bernoulli(p)	p = logit(Xβ)
Anderson et al. [3]	Care-seeking	TTE	Y $\sim$ Triangle(...)	constant p
Carley et al. [23]	Care-seeking	Binary	Y $\leftarrow$ if $\mu > \tau$	μ = f(X)
Macal et al. [64]	Care-seeking	Binary	Y $\sim$ Bernoulli(p)	constant p
Murata et al. [73]	Care-seeking	Categorical	Y $\sim$ Multinomial(p)	p = λ(X)
Turner et al. [95]	Care-seeking	Binary	Y $\sim$ Bernoulli(p)	constant p
Zhang et al. [106]	Transportation for care-seeking	Categorical Y $\sim$ Multinomial(p)	p = λ(X)	
Brailsford and Schmidt [19]	Attendance	Binary	Y $\sim$ Bernoulli(p)	p = f(X)
Brailsford et al. [18]	Attendance	Binary	Y $\sim$ Bernoulli(p)	p = logit(Xβ)
Ernecoff et al. [36]	Attendance	Binary	Y $\leftarrow$ if $\mu > \tau$	μ = f(X)
Bowers et al. [16]	Hospital choice	Categorical	Y $\leftarrow$ argmax(u)	u = f(Xβ)
Chang et al. [24]	Hospital choice	Categorical	Categorical Y $\sim$ Multinomial(p)	p = f(X)
Knight et al. [56]	Hospital choice	Categorical	Y $\leftarrow$ argmax(u)	u = f(X)
Ciavarella et al. [29]	Student absenteeism	Binary	Y $\leftarrow \mu$	μ = λ(X)
Sadeghipour et al. [81]	Dental visit	Categorical	Y $\leftarrow \mu$	μ = λ(X)
Lewis et al. [59]	Emergency calling	Binary	Y $\leftarrow$ if $\mu > \tau$	μ = f(X)
Balaban [7]	Returning to work	Binary	Y $\leftarrow \mu$	μ = λ(X)
Moro and Pellizzari [72]	Sick leave taking	Binary	Y $\leftarrow \mu$	μ = λ(X)

Bibliography

[1] Adams J, Schaefer DR. How Initial Prevalence Moderates Network-based Smoking Change: Estimating Contextual Effects with Stochastic Actor-based Models. J Health Soc Behav 2016 Mar;57(1):22–38.

[2] Alibrahim A, Wu S. An agent-based simulation model of patient choice of health care providers in accountable care organizations. Health Care Manag Sci 2018 Mar;21(1):131–143.

[3] Anderson J, Chaturvedi A, Lengacher D, Cibulskis M. Modeling the Health of Refugee Camps: An Agent-Based Computational Approach. In: Lee, DJ and Nutter, B and Antani, S and Mitra, S and Archibald, J, editor. 19th IEEE Symposium on Computer-Based Medical Systems (CBMS'06) COMPUTER-BASED MEDICAL SYSTEMS : PROCEEDINGS OF THE ANNUAL IEEE SYMPOSIUM; 2006. p. 641–645.

[4] Andrews MA, Bauch CT. The impacts of simultaneous disease intervention decisions on epidemic outcomes. J Theor Biol 2016 Apr;395:1–10.

[5] Atkinson JA, Knowles D, Wiggers J, Livingston M, Room R, Prodan A, et al. Harnessing advances in computer simulation to inform policy and planning to reduce alcohol-related harms. Int J Public Health 2018 May;63(4):537–546.

[6] Auchincloss AH, Riolo RL, Brown DG, Cook J, Diez Roux AV. An agent-based model of income inequalities in diet in the context of residential segregation. Am J Prev Med 2011 Mar;40(3):303–311.

[7] Balaban M. Return to work behavior of people with disabilities: A multi-method approach. In: Tolk, A and Yilmaz, L and Diallo, SY and Ryzhov, IO, editor. Proceedings of the Winter Simulation Conference 2014 Winter Simulation Conference Proceedings; 2014. p. 1561–1572.

[8] Barrett CL, Bisset K, Leidig J, Marathe A, Marathe M. Estimating the Impact of

public and private strategies for controlling and epidemic: a multi-agent approach; 2009. p. 34–39.

[9] Barrett C, Bisset K, Leidig J, Marathe A, Marathe M. Economic and social impact of influenza mitigation strategies by demographic class. Epidemics 2011 Mar;3(1):19–31.

[10] Basu S, Vellakkal S, Agrawal S, Stuckler D, Popkin B, Ebrahim S. Averting obesity and type 2 diabetes in India through sugar-sweetened beverage taxation: an economic-epidemiologic modeling study. PLoS Med 2014 Jan;11(1):e1001582.

[11] Beheshti R, Igusa T, Jones-Smith J. Simulated Models Suggest That Price per Calorie Is the Dominant Price Metric That Low-Income Individuals Use for Food Decision Making. J Nutr 2016 Nov;146(11):2304–2311.

[12] Beheshti R, Jalalpour M, Glass TA. Comparing methods of targeting obesity interventions in populations: An agent-based simulation. SSM Popul Health 2017 Dec;3:211–218.

[13] Blok DJ, van Lenthe FJ, de Vlas SJ. The impact of individual and environmental interventions on income inequalities in sports participation: explorations with an agent-based model. Int J Behav Nutr Phys Act 2018 Nov;15(1):107.

[14] Blok DJ, de Vlas SJ, Bakker R, van Lenthe FJ. Reducing Income Inequalities in Food Consumption: Explorations With an Agent-Based Model. Am J Prev Med 2015 Oct;49(4):605–613.

[15] Bobashev G, Goree S, Frank J, Zule W. Pain town, an agent-based model of opioid use trajectories in a small community. In: Bisgin H , Thomson R , Hyder A , Dancy C , editors. 11th International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction conference and Behavior Representation in Modeling and Simulation, SBP-BRiMS 2018, vol. 10899 LNCS Springer Verlag; 2018.p. 274–285.

[16] Bowers J, Mould G, Marshall C. Location of services and the impact on healthcare quality: insights from a simulation of a musculoskeletal physiotherapy service. J Oper Res Soc 2015 Jul;66(7):1212–1221.

[17] Bracher M, Santow G, Watkins SC. Assessing the potential of condom use to prevent the spread of HIV: a microsimulation study. Stud Fam Plann 2004 Mar;35(1):48–64.

[18] Brailsford SC, Harper PR, Sykes J. Incorporating human behaviour in simulation models of screening for breast cancer. Eur J Oper Res 2012 Jun;219(3):491–507.

[19] Brailsford S, Schmidt B. Towards incorporating human behaviour in models of health care systems: An approach using discrete event simulation. Eur J Oper Res 2003 Oct;150(1):19–31.

[20] Brown GDA, Fincher CL, Walasek L. Personality, Parasites, Political Attitudes, and Cooperation: A Model of How Infection Prevalence Influences Openness and Social Group Formation. Top Cogn Sci 2016 Jan;8(1):98–117.

[21] Bruzzone A, Novak V, Madeo F. Agent based simulation model for obesity epidemic analysis; 2012. p. 195–203.

[22] Buttenheim AM, Cherng ST, Asch DA. Provider dismissal policies and clustering of vaccine-hesitant families: an agent-based modeling approach. Hum Vaccin Immunother 2013 Aug;9(8):1819–1824.

[23] Carley KM, Fridsma DB, Casman E, Yahja A, Altman N, Li-Chiou Chen, et al. BioWar: scalable agent-based model of bioattacks. IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans 2006 Mar;36(2):252–265.

[24] Chang S, Yang W, Deguchi H. Assessing Long-Term Care Resource Distribution in China by Simulating Care-Seeking Behaviors. In: Kurahashi S, Ohta Y, Arai S, Satoh K, Bekki D, editors. New Frontiers in Artificial Intelligence, vol. 10247 of Lecture Notes in Artificial Intelligence Springer International Publishing; 2017. p. 204–219.

[25] Chao D, Hashimoto H, Kondo N. Dynamic impact of social stratification and social influence on smoking prevalence by gender: An agent-based model. Soc Sci Med 2015 Dec;147:280–287.

[26] Chen HJ, Xue H, Kumanyika S, Wang Y. School beverage environment and children's energy expenditure associated with physical education class: an agent-based model simulation. Pediatr Obes 2017 Jun;12(3):203–212.

[27] Cherng ST, Tam J, Christine PJ, Meza R. Modeling the Effects of E-cigarettes on Smoking Behavior: Implications for Future Adult Smoking Prevalence. Epidemiology 2016 Nov;27(6):819–826.

[28] Choi SE, Seligman H, Basu S. Cost Effectiveness of Subsidizing Fruit and Vegetable Purchases Through the Supplemental Nutrition Assistance Program. Am J Prev Med 2017 May;52(5):e147–e155.

[29] Ciavarella C, Fumanelli L, Merler S, Cattuto C, Ajelli M. School closure policies at municipality level for mitigating influenza spread: a model-based evaluation. BMC Infect Dis 2016 Oct;16(1):576.

[30] Collinson S, Khan K, Heffernan JM. The Effects of Media Reports on Disease Spread and Important Public Health Measurements. PLoS One 2015 Nov;10(11):e0141423.

[31] Delaney L, Kleczkowski A, Maharaj S, Rasmussen S, Williams L. Reflections on a Virtual Experiment Addressing Human Behavior During Epidemics. In: Proceedings of the 2013 Summer Computer Simulation Conference, vol. 45 of SCSC '13 Vista, CA: Society for Modeling & Simulation International; 2013. p. 33:1–33:8.

[32] Dray A, Perez P, Moore D, Dietze P, Bammer G, Jenkinson R, et al. Are drug detection dogs and mass-media campaigns likely to be effective policy responses to psychostimulant use and related harm? Results from an agent-based simulation model. Int J Drug Policy 2012 Mar;23(2):148–153.

[33] Durham DP, Casman EA. Incorporating individual health-protective decisions into disease transmission models: a mathematical framework. J R Soc Interface 2012 Mar;9(68):562–570.

[34] Edwards KL, Clarke GP. The design and validation of a spatial microsimulation model of obesogenic environments for children in Leeds, UK: SimObesity. Soc Sci Med 2009 Oct;69(7):1127–1134.

[35] Epstein JM, Parker J, Cummings D, Hammond RA. Coupled contagion dynamics of fear and disease: mathematical and computational explorations. PLoS One 2008 Dec;3(12):e3955.

[36] Ernecoff NC, Keane CR, Albert SM. Health behavior change in advance care planning: an agent-based model. BMC Public Health 2016 Feb;16:193.

[37] Fitzpatrick B, Martinez J, Polidan E, Angelis E. The Big Impact of Small Groups on College Drinking. J Artif Soc Soc Simul 2015 Jun;18(3).

[38] Fitzpatrick BG, Martinez J, Polidan E, Angelis E. On the Effectiveness of Social Norms Intervention in College Drinking: The Roles of Identity Verification and Peer Influence. Alcohol Clin Exp Res 2016 Jan;40(1):141–151.

[39] François C, Laramée P, Rahhali N, Chalem Y, Aballéa S, Millier A, et al. A predictive microsimulation model to estimate the clinical relevance of reducing alcohol consumption in alcohol dependence. Eur Addict Res 2014 Sep;20(6):269–284.

[40] French DA, Eisenberg M, Nance T, Teymuroglu Z. Analytical and computational study of an individual-based network model for the spread of heavy drinking. J Biol Dyn 2018 Dec;12(1):509–526.

[41] Fu F, Rosenbloom DI, Wang L, Nowak MA. Imitation dynamics of vaccination behaviour on social networks. Proc Biol Sci 2011 Jan;278(1702):42–49.

[42] Getsios D, Marton JP, Revankar N, Ward AJ, Willke RJ, Rublee D, et al. Smoking cessation treatment and outcomes patterns simulation: a new framework for evaluating the potential health and economic impact of smoking cessation interventions. Pharmacoeconomics 2013 Sep;31(9):767–780.

[43] Giabbanelli P, Crutzen R. An agent-based social network model of binge drinking among Dutch adults. Journal of Artificial Societies and Social Simulation 2013;16(2):10.

[44] Gorman DM, Mezic J, Mezic I, Gruenewald PJ. Agent-based modeling of drinking behavior: a preliminary model and potential applications to theory and practice. Am J Public Health 2006 Nov;96(11):2055–2060.

[45] Gutfraind A, Boodram B, Prachand N, Hailegiorgis A, Dahari H, Major ME. Agent-Based Model Forecasts Aging of the Population of People Who Inject Drugs in Metropolitan Chicago and Changing Prevalence of Hepatitis C Infections. PLoS One 2015 Sep;10(9):e0137993.

[46] Hammond RA, Ornstein JT. A model of social influence on body mass index. In: Dube, L and Webb, P and Arora, NK and Pingalid, P, editor. PATHS OF CONVERGENCE FOR AGRICULTURE, HEALTH, AND WEALTH, vol. 1331 of Annals of the New York Academy of Sciences; 2014.p. 34–42.

[47] Heard D, Bobashev GV, Morris RJ. Reducing the complexity of an agent-based local heroin market model. PLoS One 2014 Jul;9(7):e102263.

[48] Hennessy E, Ornstein JT, Economos CD, Herzog JB, Lynskey V, Coffield E, et al. Designing an Agent-Based Model for Childhood Obesity Interventions: A Case Study of ChildObesity180. Prev Chronic Dis 2016 Jan;13:E04.

[49] Hoffer LD, Bobashev G, Morris RJ. Researching a local heroin market as a complex adaptive system. Am J Community Psychol 2009 Dec;44(3-4):273–286.

[50] Hui BB, Gray RT, Wilson DP, Ward JS, Smith AMA, Philip DJ, et al. Population movement can sustain STI prevalence in remote Australian indigenous communities. BMC Infect Dis 2013 Apr;13:188.

[51] Igarashi A, Goto R, Suwa K, Yoshikawa R, Ward AJ, Moller J. Cost-Effectiveness Analysis of Smoking Cessation Interventions in Japan Using a Discrete-Event Simulation. Appl Health Econ Health Policy 2016 Feb;14(1):77–87.

[52] Iwamura Y, Tanimoto J, Fukuda E. Effect of intermediate defense measures in voluntary vaccination games. J Stat Mech 2016 Sep;2016(9):093501.

[53] Karimi E, Schmitt K, Akgunduz A. Effect of individual protective behaviors on influenza transmission: an agent-based model. Health Care Manag Sci 2015 Sep;18(3):318–333.

[54] Keane C. Chaos in collective health: Fractal dynamics of social learning. J Theor Biol 2016 Nov;409:47–59.

[55] Keane CR. Resilience, tipping, and hydra effects in public health: emergent collective behavior in two agent-based models. BMC Public Health 2016 Mar;16:265.

[56] Knight VA, Williams JE, Reynolds I. Modelling patient choice in healthcare systems: development and application of a discrete event simulation with agent-based decision making. Journal of Simulation 2012 May;6(2):92–102.

[57] Lamy F, Quinn B, Dwyer R, Thomson N, Moore D, Dietze P. TreatMethHarm: An Agent-Based Simulation of How People Who Use Methamphetamine Access Treatment. JASSS 2016;19(2).

[58] Langellier BA. An agent-based simulation of persistent inequalities in health behavior: Understanding the interdependent roles of segregation, clustering, and social influence. SSM Popul Health 2016 Dec;2:757–769.

[59] Lewis B, Swarup S, Bisset K, Eubank S, Marathe M, Barrett C. A simulation environment for the dynamic evaluation of disaster preparedness policies and interventions. J Public Health Manag Pract 2013 Sep;19 Suppl 2(2):S42–8.

[60] Li Y, Zhang D, Pagán JA. Social Norms and the Consumption of Fruits and Vegetables across New York City Neighborhoods. J Urban Health 2016 Apr;93(2):244–255.

[61] Li Y, Zhang D, Thapa JR, Madondo K, Yi S, Fisher E, et al. Assessing the role of access and price on the consumption of fruits and vegetables across New York City using agent-based modeling. Prev Med 2018 Jan;106:73–78.

[62] Li Y, Du T, Peng J. Understanding Out-of-Home Food Environment, Family Restaurant Choices, and Childhood Obesity with an Agent-Based Huff Model. Sustain Sci Pract Policy 2018 May;10(5):1575.

[63] Liu XT, Wu ZX, Zhang L. Impact of committed individuals on vaccination behavior. Phys Rev E Stat Nonlin Soft Matter Phys 2012 Nov;86(5 Pt 1):051132.

[64] Macal CM, North MJ, Collier N, Dukic VM, Lauderdale DS, David MZ, et al. Modeling the spread of community-associated MRSA. In: Laroque, C and Himmelspach, J and Pasupathy, R, editor. Proceedings of the 2012 Winter Simulation Conference (WSC) Winter Simulation Conference Proceedings; 2012. p. 1–12.

[65] Mao L. Modeling triple-diffusions of infectious diseases, information, and preventive behaviors through a metropolitan social network—An agent-based simulation. Appl Geogr 2014 Jun;50:31–39.

[66] Mao L. Predicting self-initiated preventive behavior against epidemics with an agent-based relative agreement model. Journal of Artificial Societies and Social Simulation 2015;18(4):6.

[67] Markham F, Young M, Doran B. Improving spatial microsimulation estimates of health outcomes by including geographic indicators of health behaviour: The example of problem gambling. Health Place 2017 Jul;46:29–36.

[68] Marquis MS, Buchanan JL. Simulating the effects of employer contributions on adverse selection and health plan choice. Health Serv Res 1999 Oct;34(4):813–837.

[69] Marshall BDL, Paczkowski MM, Seemann L, Tempalski B, Pouget ER, Galea S, et al. A complex systems approach to evaluate HIV prevention in metropolitan areas: preliminary implications for combination intervention strategies. PLoS One 2012 Sep;7(9):e44833.

[70] Mei S, Zhu Y, Qiu X, Zhou X, Zu Z, Boukhanovsky AV, et al. Individual Decision Making Can Drive Epidemics: A Fuzzy Cognitive Map Study. IEEE Trans Fuzzy Syst 2014 Apr;22(2):264–273.

[71] Millier A, Laramée P, Rahhali N, Aballéa S, Daeppen JB, Rehm J, et al. Cost-Effectiveness of Nalmefene Added to Psychosocial Support for the Reduction of Alcohol Consumption in Alcohol-Dependent Patients With High/Very High Drinking Risk Levels: A Microsimulation Model. J Stud Alcohol Drugs 2017 Nov;78(6):867–876.

[72] Moro A, Pellizzari P. An Agent-Based Model of Labor Market Participation with Health Shocks. In: Demazeau, Y and Ito, T and Bajo, J and Escalona, MJ, editor. Advances in Practical Applications of Scalable Multi-agent Systems. The PAAMS Collection, vol. 9662 of Lecture Notes in Artificial Intelligence Springer International Publishing; 2016. p. 157–168.

[73] Murata T, Morimoto G, Tokura S. Survey-based hospital closure estimation using agent-based simulation. In: 2011 IEEE International Conference on Systems, Man, and Cybernetics IEEE International Conference on Systems Man and Cybernetics Conference Proceedings; 2011. p. 233–238.

[74] Nagoski E, Janssen E, Lohrmann D, Nichols E. Risk, individual differences, and environment: an Agent-Based Modeling approach to sexual risk-taking. Arch Sex Behav 2012 Aug;41(4):849–860.

[75] Orr MG, Galea S, Riddle M, Kaplan GA. Reducing racial disparities in obesity: simulating the effects of improved education and social network influence on diet behavior. Ann Epidemiol 2014 Aug;24(8):563–569.

[76] Orr MG, Kaplan GA, Galea S. Neighbourhood food, physical activity, and educational environments and black/white disparities in obesity: a complex systems simulation analysis. J Epidemiol Community Health 2016 Sep;70(9):862–867.

[77] Perez P, Dray A, Moore D, Dietze P, Bammer G, Jenkinson R, et al. SimAmph: an agent-based simulation model for exploring the use of psychostimulants and related harm amongst young Australians. Int J Drug Policy 2012 Jan;23(1):62–71.

[78] Purshouse RC, Ally AK, Brennan A, Moyo D, Norman P. Evolutionary parameter estimation for a theory of planned behaviour microsimulation of alcohol consumption dynamics in an English birth cohort 2003 to 2010. In: Igel, C, editor. Proceedings of the 2014 Annual Conference on Genetic and Evolutionary Computation ACM; 2014. p. 1159–1166.

[79] Rahhali N, Millier A, Briquet B, Laramée P, Aballéa S, Toumi M, et al. Modelling the consequences of a reduction in alcohol consumption among patients with alcohol dependence based on real-life observational data. BMC Public Health 2015 Dec;15:1271.

[80] Riva M, Smith DM. Generating small-area prevalence of psychological distress and alcohol consumption: validation of a spatial microsimulation method. Soc Psychiatry Psychiatr Epidemiol 2012 May;47(5):745–755.

[81] Sadeghipour M, Shariatpanahi P, Jafari A, Khosnevisan MH, Ebn Ahmady A. Oscillatory Patterns in the Amount of Demand for Dental Visits: An Agent Based Modeling Approach. JASSS 2016;19(3).

[82] Sahneh FD, Chowdhury FN, Scoglio CM. On the existence of a threshold for preventive behavioral responses to suppress epidemic spreading. Sci Rep 2012 Sep;2:632.

[83] Salvarani F, Turinici G. Optimal individual strategies for influenza vaccines with imperfect efficacy and durability of protection. Math Biosci Eng 2018 Jun;15(3):629–652.

[84] Schaefer DR, Adams J, Haas SA. Social networks and smoking: exploring the effects of peer influence and smoker popularity through simulations. Health Educ Behav 2013 Oct;40(1 Suppl):24S–32S.

[85] Scott N, Livingston M, Hart A, Wilson J, Moore D, Dietze P. SimDrink: An Agent-Based NetLogo Model of Young, Heavy Drinkers for Conducting Alcohol Policy Experiments. JASSS 2016;19(1).

[86] Scott N, Livingston M, Reporter I, Dietze P. Using simulation modelling to examine the impact of venue lockout and last-drink policies on drinking-related harms and costs to licensees. Aust N Z J Public Health 2017 Jun;41(3):243–247.

[87] Sowa PM, Kault S, Byrnes J, Ng SK, Comans T, Scuffham PA. Private Health Insurance Incentives in Australia: In Search of Cost-Effective Adjustments. Appl Health Econ Health Policy 2018 Feb;16(1):31–41.

[88] Sun R, Mendez D. An application of the Continuous Opinions and Discrete Actions (CODA) model to adolescent smoking initiation. PLoS One 2017 Oct;12(10):e0186163.

[89] Suzuki R, Ishino Y, Takahashi S. Characterization of consumers' behavior in medical insurance market with agent parameters' Estimation process using Bayesian network. In: Mineshima K , Kojima K , Satoh K , Arai S , Bekki D , Ohta Y , editors. 9th JSAI International Symposium on Artificial Intelligence, JSAI-isAI 2017, vol. 10838 LNAI Springer Verlag; 2018.p. 112–128.

[90] Teoh Shian Li, Labadin J, Phang Piau, Rahman SA, Ling Yeong Tyng. The effect of vaccination decision in disease modelling through simulation. In: Tiong, WK and Ujir, H and Mit, E and Iskandar, DNA, editor. 2015 9th International Conference on IT in Asia (CITA); 2015. p. 1–6.

[91] Trogdon JG, Allaire BT. The effect of friend selection on social influences in obesity. Econ Hum Biol 2014 Dec;15:153–164.

[92] Tully S, Cojocaru M, Bauch CT. Coevolution of risk perception, sexual behaviour, and HIV transmission in an agent-based model. J Theor Biol 2013 Nov;337:125–132.

[93] Tully S, Cojocaru M, Bauch CT. Sexual behavior, risk perception, and HIV transmission can respond to HIV antiviral drugs and vaccines through multiple pathways. Sci Rep 2015 Oct;5:15411.

[94] Turner KME, Adams EJ, Lamontagne DS, Emmett L, Baster K, Edmunds WJ. Modelling the effectiveness of chlamydia screening in England. Sex Transm Infect 2006 Dec;82(6):496–502.

[95] Turner KME, Adams EJ, Gay N, Ghani AC, Mercer C, Edmunds WJ. Developing a realistic sexual network model of chlamydia transmission in Britain. Theor Biol Med Model 2006 Jan;3:3.

[96] Wang Y, Xue H, Chen HJ, Igusa T. Examining social norm impacts on obesity and eating behaviors among US school children based on agent-based model. BMC Public Health 2014 Sep;14:923.

[97] White PJ, Fox J, Weber J, Fidler S, Ward H. How Many HIV infections may be averted by targeting primary infection in men who have sex with men? Quantification of changes in transmission-risk behavior, using an individual-based model. J Infect Dis 2014 Dec;210 Suppl 2:S594–9.

[98] Xenakis JG, Kinter ET, Ishak KJ, Ward AJ, Marton JP, Willke RJ, et al. A discrete-event simulation of smoking-cessation strategies based on varenicline pivotal trial data. Pharmacoeconomics 2011 Jun;29(6):497–510.

[99] Xiong J, Wang J, Yang K, Peng S, Xu Q. Multiagent-based simulation of the HIV/AIDS spatial and temporal transmission among injection drug users. In: Liu, Y and Chen, A, editor. 2010 18th International Conference on Geoinformatics; 2010. p. 1–6.

[100] Xue J, Terano T, Deguchi H, Ichikawa M. Simulation analysis of immunization policy diffusion in social network with ABM approach. In: 2016 International Conference on Behavioral, Economic and Socio-cultural Computing (BESC) IEEE; 2016. p. 1–6.

[101] Yang W, Zou Q, Tan E, Watkins L, Beronja K, Hogan PF, et al. Future Health and Economic Impact of Comprehensive Tobacco Control in DoD: A Microsimulation Approach. Mil Med 2018 Jan;183(1-2):e104–e112.

[102] Yang Y, Auchincloss AH, Rodriguez DA, Brown DG, Riolo R, Diez-Roux AV. Modeling spatial segregation and travel cost influences on utilitarian walking: Towards policy intervention. Comput Environ Urban Syst 2015 May;51:59–69.

[103] Yang Y, Diez Roux AV, Auchincloss AH, Rodriguez DA, Brown DG. A spatial agent-based model for the simulation of adults' daily walking within a city. Am J Prev Med 2011 Mar;40(3):353–361.

[104] Yang Y, Diez Roux AV, Auchincloss AH, Rodriguez DA, Brown DG. Exploring walking differences by socioeconomic status using a spatial agent-based model. Health Place 2012 Jan;18(1):96–99.

[105] Zhang D, Giabbanelli PJ, Arah OA, Zimmerman FJ. Impact of different policies on unhealthy dietary behaviors in an urban adult population: an agent-based simulation model. Am J Public Health 2014 Jul;104(7):1217–1222.

[106] Zhang Q, Northridge ME, Jin Z, Metcalf SS. Modeling Accessibility of Screening and Treatment Facilities for Older Adults using Transportation Networks. Appl Geogr 2018 Apr;93:64–75.

[107] Zhang Fa, Zhao Qiao-xia, Li Lu. Intervention for contagious disease: Agent-based modeling and simulation. In: 2011 2nd IEEE International Conference on Emergency Management and Management Sciences IEEE; 2011. p. 155–158.

[108] Zhong W. Simulating influenza pandemic dynamics with public risk communication and individual responsive behavior. Comput Math Organ Theory 2017 Dec;23(4):475–495.

[109] Zur RM, Zaric GS. A microsimulation cost-utility analysis of alcohol screening and brief intervention to reduce heavy alcohol consumption in Canada. Addiction 2016 May;111(5):817–831.

Appendix B

Lee Carter model specifications

Specifications of the Lee-Carter approach

B.1 Notations and abbreviations

$y_{age,year}$: number of events of *age* in *year*

$n_{age,year}$: population size of *age* in *year*

α_{age}: age-specific effect of *age*

β_{age}: age-period adjustment term of *age*

κ_{year}: period effect at *year*

$E(.)$: expectation function

$sign(.)$: sign function of a value, 1 for positive and -1 for negative.

$\ell(.)$: log-likelihood function

$ARIMA(p, d, q)$: Autoregressive integrated moving average with p number of time lags of the autoregressive term, d order of differencing term, and q number of time lags of the moving-average term.

B.2 Model structure and assumptions

Lee-Carter model [1] was defined as

$$log(E(y_{year,age})) = \alpha_{age} + \beta_{age}\kappa_{year} + log(n_{year,age}) \tag{B.1}$$

Two constraints are needed for ensuring identifiability:

$$\sum_{year} \kappa_{year} = 0 \tag{B.2}$$

$$\sum_{age} \beta_{age} = 1 \tag{B.3}$$

In implementation, we used a Poisson-regression-based approach [2] for the Lee-Carter model

$$y_{year,age} \sim Poisson(\mu_{year,age}) \tag{B.4}$$

$$\mu_{year,age} = n_{year,age} exp(\alpha_{age} + \beta_{age}\kappa_{year}) \tag{B.5}$$

Thus, the log-likelihood function is

$$\ell(\alpha_{age}, \beta_{age}, \kappa_{year} | y_{year,age}, n_{year,age}) = -\mu_{year,age} + y_{year,age} log(\mu_{year,age}) - log(y_{year,age}!) \tag{B.6}$$

$$= y_{year,age}(\alpha_{age} + \beta_{age}\kappa_{year}) - n_{year,age} exp(\alpha_{age} + \beta_{age}\kappa_{year}) + constant \tag{B.7}$$

B.2.1 Comparator models

Two reduced models were considered as comparators:

Age-Period model assumed β_{age} are the same for all age groups, so the model reduced to

$$y_{year,age} \sim Poisson(\mu_{year,age}) \tag{B.8}$$

$$\mu_{year,age} = n_{year,age} exp(\alpha_{age} + \kappa_{year}) \tag{B.9}$$

Age-Trend model assumed β_{age} are the same for all age groups and a linear period effect, so the model reduced to

$$y_{year,age} \sim Poisson(\mu_{year,age}) \tag{B.10}$$

$$\mu_{year,age} = n_{year,age} exp(\alpha_{age} + year \times \kappa) \tag{B.11}$$

B.3 Model fitting

The maximum likelihood estimation of this model is to solve

$$\underset{\alpha_{age},\beta_{age},\kappa_{year}}{\operatorname{argmax}} \ \ell(\alpha_{age}, \beta_{age}, \kappa_{year} | y_{year,age}, n_{year,age}) \tag{B.12}$$

subject to

$$\sum_{year} \kappa_{year} = 0 \tag{B.13}$$

$$\sum_{age} \beta_{age} = 1 \tag{B.14}$$

We employed Newton methods as described in Brouhns et al. [2] to this task. Our implementation used *StMoMo::fit* function from package StMoMo [3].

B.4 Modelling and forecasting

For period effects, κ_{year}, we employed the Box-Jenkins method, which uses auto-correlation function (ACF), partial autocorrelation function (PACF), and extended ACF if necessary to specified a time-series model of κ_{year} as a $ARIMA(p,d,q)$ model with drift. That is,

$$\big(1 - \sum_{i=1}^{p} \phi_i L^i\big)(1-L)^d \kappa_{year} = c + \big(1 + \sum_{i=1}^{q} \theta_i L^i\big)\varepsilon_{year} \tag{B.15}$$

where ϕ_i and θ_i are coefficients of lag terms of κ_{year} and previous random errors ε_i, c is the constant drift term, and L is the lag operator, shifting a variable to its lag term, i.e. $L^i \kappa_{year} = \kappa_{year-i}$

This specification and modelling were implemented using functions,*TSA::acf*, *TSA::pacf*, and *TSA::eacf* from package *TSA* [4].

B.5 Bootstrap

The bootstrap simulation employed the semi-parametric bootstrap by Renshaw and Haberman [5]. Our implementation used *StMoMo::simulation* function from package StMoMo [3]. In general, we generated 10,000 simulations for each presented result.

B.6 Measurements of goodness of fit

Since the likelihood-based LCM we applied is a special case of the ordinary Poisson regression [6], the measurements of goodness of fit of Poisson regression can be directly applied.

Akaike information criterion (AIC): by definition,

$$2k - 2\hat{\ell}(\hat{\alpha}_{age}, \hat{\beta}_{age}, \hat{\kappa}_{year} | y_{year,age}, n_{year,age}) \tag{B.16}$$

where $\hat{\ell}(.)$ is the log-likelihood function given estimated parameters, k is the number of parameters, which equals *the sum of the numbers of* $\hat{\alpha}_{age}$, $\hat{\beta}_{age}$, $\hat{\kappa}_{year}$ *minus two constraints*.

Bayesian information criterion (BIC): by definition,

$$log(o)k - 2\hat{\ell}(\hat{\alpha}_{age}, \hat{\beta}_{age}, \hat{\kappa}_{year} | y_{year,age}, n_{year,age}) \tag{B.17}$$

where k is the number of parameters as above and o is number of observations.

Deviance residuals: We used the deviance residuals defined in Colin Cameron and Trivedi [6] to assess the goodness of fit.

$$sign(y_{age,year} - E(y_{age,year}))\sqrt{2\left[y_{age,year} log\frac{y_{age,year}}{E(y_{age,year})} - (y_{age,year} - E(y_{age,year}))\right]} \tag{B.18}$$

Bibliography

[1] Lee RD, Carter LR. Modeling and Forecasting U. S. Mortality. J Am Stat Assoc. 1992;87(419):659–671.

[2] Brouhns N, Denuit M, Vermunt JK. A Poisson log-bilinear regression approach to the construction of projected lifetables. Insur Math Econ. 2002 Dec;31(3):373–393.

[3] Andres V, Millossovich P, Vladimir K. StMoMo: Stochastic Mortality Modeling in R. J Stat Softw. 2018;84(3):1–38.

[4] Chan KS, Ripley B. TSA: Time Series Analysis; 2018.

[5] Renshaw AE, Haberman S. On simulation-based approaches to risk measurement in mortality with specific reference to Poisson Lee–Carter modelling; 2008.

[6] Colin Cameron A, Trivedi PK. Regression Analysis of Count Data. Cambridge University Press; 2013.

Appendix C

Synthetic population

Synthetic population methods

C.1 Background

The aim of this document is to describe the construction of a synthetic population of Taiwanese nationals as well as the basic population features based on the simulation of the model. The synthetic population will be used as an input of the TB modelling.

The National Development Council, Taiwan (NDC), developed a synthetic population for policy-making purposes and release a population projection report annually [1]. However, the model did not capture the stochasticity in its subprocesses and we required the lifespan between 0 and 100. Based on the assumptions and model choices in the NDC model, we constructed a stochastic synthetic population model for use in the TB incidence modelling.

The main features of our synthetic population were as follows.

- Modelling lifespan from 0 to 100 (assumed deterministic deaths when reaching aged 101).
- Being capable of generating stochastic population forecasts.

C.2 Notations and abbreviations

Data and demographic features

$year$: calender year

age: singe-year age

agp: five-year age group

sex: biological sex at birth

$PF_{year,age,sex}$: population size on 1st January of $year$

$PM_{year,age,sex}$: population size on 1st July of $year$

$PE_{year,age,sex}$: population size on 31st December of $year$

$D_{year,age,sex}$: deaths during $year$

$d_{year,age,sex}$: death rate at $year$

$B_{year,sex}$: number of newborns during $year$

$F_{year,agp,sex}$: number of newborns of sex during $year$, given agp of mothers

$f_{year,agp,sex}$: fertality rate of sex at $year$ given agp of mothers

$M_{year,age,sex}$: number of net migrations during $year$

$m_{year,age,sex}$: net migration rate at $year$

Lee-Carter components:

$\alpha_{age,sex}$: age-specific effect of (age, sex)

$\beta_{age,sex}$: age-period adjustment term of (age, sex)

$\kappa_{year,sex}$: period effect at $year$

$E(.)$: the expectation of a random variable

$ARIMA(p,d,q)$: Autoregressive integrated moving average with time lag p for the autoregressive term, order d for the differencing term, and lag q for the moving-average term.

C.3 Data

The demographic data were obtained from the Department of Statistics, the Ministry of the Interior, Taiwan. They were released publicly and available on the internet. All the training data in this article were published by the Taiwan officials and free access on the internet; the usage is licensed by the Open Government Data License: [https://data.gov.tw/license].

End-year population estimates. The estimated population size on 31st December of a year. These data were from 2004 to 2017 by sex, single-year age. We take them as the equivalent of the start-year population estimates from 2005 to 2018.

Death counts. The number of deaths on registration of a year. These data were from 2005 to 2017 by sex and single-year age.

Fertility counts. The number of births on registration of a year. These data were from 2005 to 2017 by sex of newborns and five-year-age groups (15-19, ..., 45-49) of mothers. We considered the childbearing ages of females from 15 to 49.

C.4 Modelling and forecasting

The synthetic population considers birth, death, ageing, and migration processes. As the NDC model, we used cohort-component methods [2] to combine them and drive the population dynamics. The essential model is:

$$PE_{year,0,sex} = B_{year,sex} + M_{year,0,sex} \qquad age = 0 \qquad \text{(C.1)}$$

$$PE_{year,age,sex} = PF_{year,age,sex} - D_{year,age,sex} + M_{year,age,sex} \qquad age \in [1, 100] \qquad \text{(C.2)}$$

$$PF_{year,0,sex} = 0 \qquad age = 0 \qquad \text{(C.3)}$$

$$PF_{year,age,sex} = PE_{year-1,age-1,sex} \qquad age \in [1, 100] \qquad \text{(C.4)}$$

For all submodels, we generated 10,000 bootstrap simulations for considering stochasticity with the time range from 2000 to 2017 for fitting and from 2018 to 2035 for forecasting.

C.4.1 Death process

We modelled the death rates of ages below 85 with the Lee-Carter model [3] and above 85 with the Coale-Kisker method [4].

Aged 0-84 The Lee-Carter model (LCM) is a statistical model decomposing the death rates into an age-specific baseline ($\alpha_{age,sex}$), age-specific trend terms with resect to time effect ($\beta_{age,sex}$), and an intrinsic time trend ($\kappa_{year,sex}$), which is modelled and forecasted using time-series methods. We applied the likelihood-based implementation derived by Brouhns et al. [5] to model the death rates ($d_{year,age,sex}$), where $d_{year,age,sex} = D_{year,age,sex} / PM_{year,age,sex}$. Therefore,

$$log(E(D_{year,age,sex})) = \alpha_{age,sex} + \beta_{age,sex}\kappa_{year,sex} + log(PM_{year,age,sex}) \tag{C.5}$$

We modelled $\kappa_{year,sex}$ using an ARIMA time-series model. Applying the Box–Jenkins method [6], we identified $\kappa_{year,male}$ as a $ARIMA(1,0,0)$ with drift and $\kappa_{year,female}$ as a $ARIMA(0,1,0)$ with drift. We then used the bootstrap method for LCM by Renshaw and Haberman 2008 [7] to generate 10,000 simulations of deaths

Aged 85-100 In the extremely old population, the sample sizes of mortalities might be small which usually leads to high uncertainty in statistical models. The Coale-Kisker method [4] fits exponentially increasing mortalities with age.

Explicitly, our implementation follows the steps: 1) set up a final death rate ($d_{year,100,sex}$) for each sex, 2) rescale $d_{year,84,sex}$ and $d_{year,100,sex}$ to a logarithmic scale, 3) linearly interpolate the $log(d_{year,age,sex})$ for $age \in [85,99]$, and 4) transform back to the original scale.

We assumed final death rates at age 100 of males and females as 1 and 0.8 as the NDC model to prevent the rates crossing over.

C.4.2 Birth process

As the LCM can can also be applied on fertility rates (e.g. Simpach [8] and Hyndman and Booth [9]), we modelled birth processes using the LCM. Considering the number of newborns and availability of data, we aggregated the population size of females to five-year age-groups (15-19, ... , 45-49). Using the same specification of the method of the death process, we modelled fertility rates as

$$log(E(f_{year,agp,sex})) = \alpha_{agp,sex} + \beta_{agp,sex}\kappa_{year,sex} + log(PM_{year,agp,female}) \tag{C.6}$$

For deaths, we then used the bootstrap method to generate the numbers of newborns ($F_{year,agp,sex}$). The total numbers of newborns by sex were

$$B_{year,sex} = \sum_{agp} F_{year,agp,sex} \tag{C.7}$$

C.4.3 Migration process

We did not use migration data directly, but used the residual method [10] to derive the net migration instead. Under this method, the net migrations were calculated

from the difference between net flows in the population at an age and flows due to birth and death:

$$M_{year,0,sex} = PE_{year,0,sex} - B_{year,sex} \tag{C.8}$$

$$M_{year,age,sex} = PE_{year,age,sex} - (PF_{year,age,sex} - D_{year,age,sex}) \tag{C.9}$$

Then, the net migration rates are

$$m_{year,age,sex} = M_{year,age,sex} / PM_{year,age,sex} \tag{C.10}$$

In terms of forecasting, we took the arithmetic mean of net migration rate for each single-year age, assuming the future migrations will be constant at a given age. We sampled the number of net migration, $M_{year,age,sex}$, by the Poisson distribution with mean $m_{year,age,sex}$.

C.4.4 Simulation

After the parameter estimation procedures above, we simulated the population dynamics as follows.

Step 1, Initialisation The first step starts with setting up the initial population sizes of each (age, sex). For the population aged above one, the initial size is from the end-year population estimates of the previous year. For the age zero population, the initial size is zero.

Step 2, birth, death and migration For the zero aged populations by sex, add the number of birth to the values. For each single-year age and sex, add the number of net migration and subtract the number of deaths to the population.

Step 3, Ageing Remove the oldest population group ($age = 100$) and count them as deaths. Then, shift the population upward by one year.

Step 4, Iteration Iterate through Step 1 to Step 3 until 2035.

Last, apply the procedure on every bootstrap sample to complete the simulation.

C.5 Notes

Software: The data manipulation, modelling, and forecasting in this document were conducted using R 3.5.1 with packages of StMoMo and TSA [11–13].

Time inconsistency due to ageing process We did not use migration data directly because there is a time inconsistency due to ageing. The time inconsistency of ageing and year progress commonly showed up in the dataset. For example, if a newborn baby who has their birthday in August died in the March of the next year, he/she would be counted in the zero-age death. However, if he/she died in September of the next year, that would be a count in the one age death. The same issue can be found in population and migration data. Therefore, the numbers are always unmatched by age structure. In order to fix this issue, we used the residual method for the migration process, ensuring the other aspects of data were balanced in every year.

Bibliography

[1] National Development Council, Taiwan. Population Projections for the R.O.C. (Taiwan): 2018-2065. National Development Council, Taiwan; 2018.

[2] Alho JM. Stochastic methods in population forecasting. Int J Forecast. 1990 Dec;6(4):521–530.

[3] Lee RD, Carter LR. Modeling and Forecasting U. S. Mortality. J Am Stat Assoc. 1992;87(419):659–671.

[4] Coale AJ, Kisker EE. Defects in data on old-age mortality in the United States: new procedures for calculating mortality schedules and life tables at the highest ages. Coale Kisker 1990 Asian and Pacifc Population Forum. 1990;.

[5] Brouhns N, Denuit M, Vermunt JK. A Poisson log-bilinear regression approach to the construction of projected lifetables. Insur Math Econ. 2002 Dec;31(3):373–393.

[6] Box GEP, Jenkins GM, Reinsel GC, Ljung GM. Time Series Analysis: Forecasting and Control. John Wiley & Sons; 2015.

[7] Renshaw AE, Haberman S. On simulation-based approaches to risk measurement in mortality with specific reference to Poisson Lee–Carter modelling; 2008.

[8] Simpach O. Fertility of Czech Females Could Be Lower than Expected: Trends in Future Development of Age-specific Fertility Rates up to the Year 2050. STATISTIKA-STATISTICS AND ECONOMY JOURNAL. 2015;95(1):19–37.

[9] Hyndman RJ, Booth H. Stochastic population forecasts using functional data models for mortality, fertility and migration. Int J Forecast. 2008 Jul;24(3):323–342.

[10] Hamilton CH. Effect of census errors on the measurement of net migration. Demography. 1966 Jun;3(2):393–415.

[11] R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria; 2018.

[12] Andres V, Millossovich P, Vladimir K. StMoMo: Stochastic Mortality Modeling in R. J Stat Softw. 2018;84(3):1–38.

[13] Chan KS, Ripley B. TSA: Time Series Analysis; 2018.

Appendix D

Tuberculosis epidemiology and population dynamics, Goodness of fit

Residuals

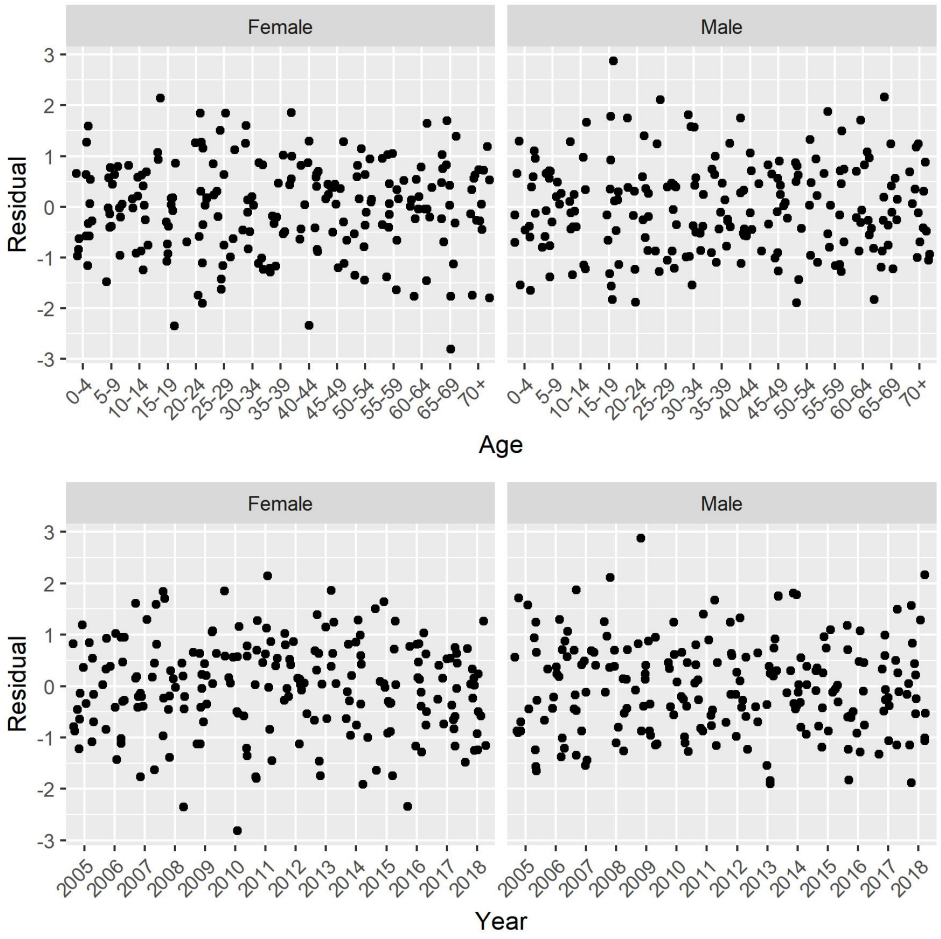


Figure D.1: Residuals by age and by calender year.

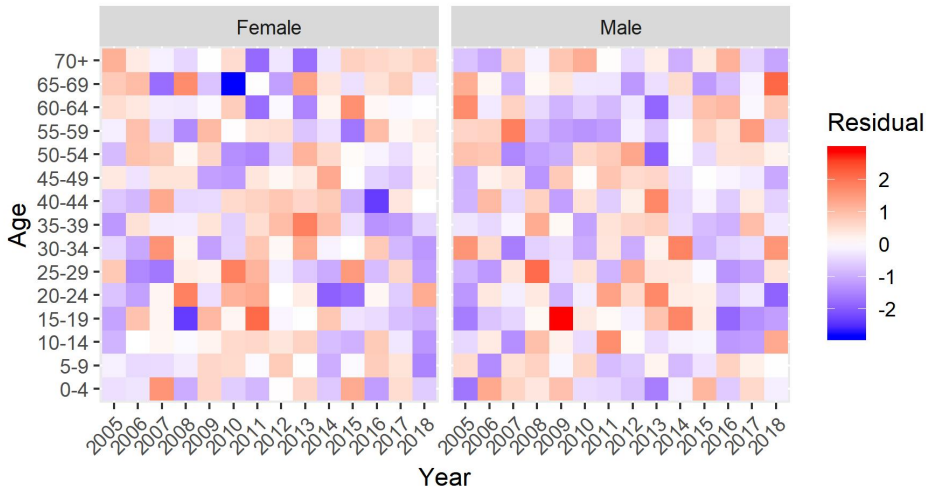


Figure D.2: Residuals by age and calender year.

Fitted TB incidence rates

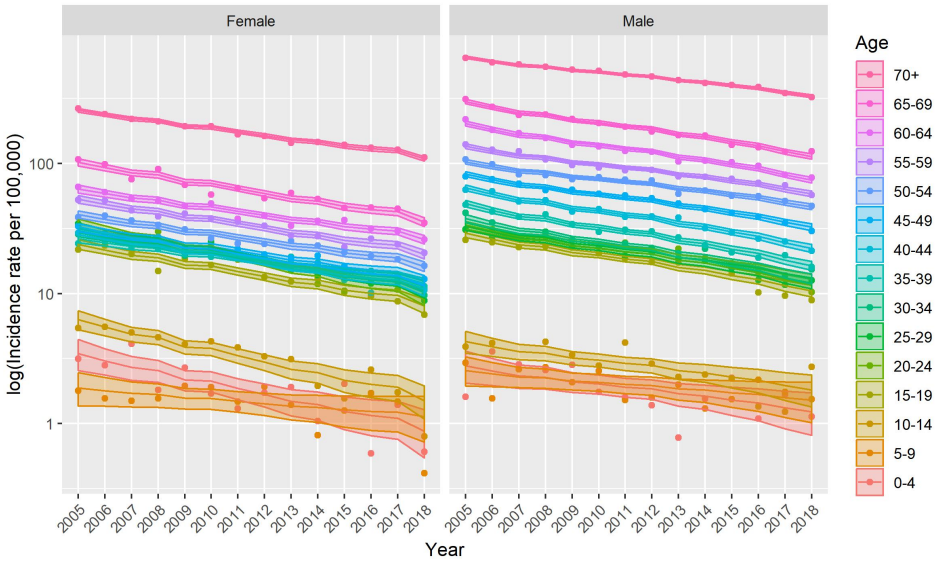


Figure D.3: Fitted TB incidence rates in logarithm to the base 10

Model comparison

We modelled female and male TB incidence separately with the same parameterisation. Apart from the likelihood-based Lee-Carter Model, two models were used as comparators. Both models can be seen as reduced Lee-Carter Models:

Age-Trend: A Poisson regression model with categorical age effects and a linear term of the calender year.

Age-Period: A Poisson regression model with categorical age effects and a discretised calender year.

Table D.1: Model comparison, female

	Age-Trend	Age-Period	Lee Carter Model
Family		Poisson Regression	
Period.effect	Linear	Discrete	
No. observations		210	
No. parameters	16	28	42
Log(Likelihood)	-852	-840	-811
AIC[1]	1737	1737	1706
BIC[2]	1790	1830	1847

Table D.2: Model comparison, male

	Age-Trend	Age-Period	Lee Carter Model
Family		Poisson Regression	
Period.effect	Linear	Discrete	
No. observations		210	
No. parameters	16	28	42
Log(Likelihood)	-1002	-979	-870
AIC	2036	2014	1825
BIC	2090	2108	1965

Appendix E

Individual Pathway Analysis, definitions

TB diagnosis and medications

TB case: An individual having 1) TB-related diagnosis codes (ICD-9-CM: 010-018) and 2) more than two anti-TB drugs prescribed for longer than 28 days or two consecutive 14 days prescriptions.

Regular treatment: An individual having 1) TB-related diagnosis codes (ICD-9-CM: 010-018) and 2) more than two anti-TB drugs prescribed for longer than 28 days (or two consecutive 14 days prescriptions).

First-line TB treatment: An individual having regular with a mixture of more than two first-line drugs: Isoniazid, Rifampicin, Pyrazinamide, Ethambutol, and Streptomycin.

Second-line TB treatment: An individual having regular with a more than two types of anti-TB drugs apart from above first-line drugs. To be noted that, a mixture of two first-line drugs and one second-line drug is a first-line TB treatment be the definition.

Treatment completion: An individual having taken regular treatments for more than 180 days.

Treatment initialisation, or empirical treatment: An individual having prescriptions of more than any two anti-TB drugs for shorter than 14 days.

Evaluation probably for TB: An individual having diagnostic or screening procedures that are specific to TB. The procedures are:

- Sputum culture test (Include Crystal, Parasite)
- Acid-fast culture
- Tuberculosis culture test
- Tuberculin Skin Test
- Chest X ray
- Chest computered tomography

Evaluation possibly for TB: An individual having medical procedures that can are highly relevant to TB suspicions. The procedures are:

- Bronchoscopy
- Mycoplasma pneumonia antigen test
- Streptococcus pneumonia antigen-latex agglutination
- Chlamydia pneumoniae antigen
- Pneumococcus antigen (urine)
- Legionella pneumophila antigen (urine)
- Cryptococcus antigen
- Aspergillus antigen
- Influenza A cystic fibrosis Antibody
- Influenza B cystic fibrosis Antibody
- Pleural fluid analysis

TB-related illness:

Chronic lung condition, CLC: An individual having 1) respiratory diagnosis codes (ICD-9-CM: 460-510) and 2) drugs for respiratory syndrome prescribed for longer than 28 days.

Acute respiratory disease, ARD: An individual having 1) respiratory diagnosis codes (ICD-9-CM: 460-510) and 2) all drugs for respiratory syndrome prescribed shorter than 28 days.

Nontuberculous mycobacteria, NTM: An individual having NTM diagnosis codes (ICD-9-CM: 031)

Comorbidities

Chronic lung condition, CLC: An individual having 1) respiratory diagnosis codes (ICD-9-CM: 460-510) and 2) drugs for respiratory syndrome prescribed for longer than 28 days.

Diabetes Mellitus, DM: An individual having 1) DM diagnosis codes (ICD-9-CM: 250, A181) and 2) DM control drugs prescribed for longer than 28 days. We cannot differentiate type 1 and type 2 diabetes in the National Health Insurance Research Database, NHIRD.

Human immunodeficiency virus, HIV: An individual having HIV related diagnosis codes (ICD-9-CM: 042, V08). We cannot differentiate HIV infection and acquired immune deficiency syndrome from the NHIRD.

Appendix F

Data descriptions, NHIRD

This appendix lists the variables and their descriptions in the National Health Insurance Research Database (NHIRD). The appendix only includes the tables and columns used in this study. The full description can be downloaded from Data description, Code book (Traditional Chinese only)

Healthcare facility (HOSP in the NHIRD)

The dataset stores all basic information the registered healthcare facilities, including clinics, hospital, pharmacy, etc.

Table F.1: Hospital data

Variable	Data type	Description
HOSP_ID	Character	Primary key; Hospital ID; encoded
HOSP_CONT_TYPE	Categorical	Facility type, level, and public or private
AREA_NO_H	Character, zip code	Registered area

Insured individuals (ID)

The dataset stores the sampled one million individuals.

- Age can be inferred from ID_BIRTHDAY
- INS_ID_TYPE indicates unemployment if the insured a dependant of another
- UNIT_INS_TYPE implies job type and social group

Table F.2: List of variables, insured individuals

Variable	Data type	Description
ID	Character	Primary key; Hospital ID; encoded
ID_SEX	Categorical	Facility type, level, and public or private
ID_BIRTHDAY	Date, YYYYMMDD	Birthday
INS_ID_TYPE	Categorical	Dependent type
INS_ID_AMT	Categorical	Insurance amount∼Income
REG_ZIP_CODE	Character, zip code	Registered area
UNIT_INS_TYPE	Categorical	Insuring institution
ID_OUT_DATE	Date, YYYYMMDD	Leaving date
AREA_NO_H	Character	Area ID

Health facility visits (CD for outpatients, DD for inpatients)

The dataset stores the records of the visits happened.

Table F.3: List of variables

Variable	Data type	Description
SEQ_NO	Character	Primary key; sequence number
ID	Character	Foreign key, referring to insured individuals
HOSP_ID	Character	Foreign key, referring to hospitals
FUNC_TYPE	Categorical	Specialist type
T_AMT	Number	Total cost
PART_AMT	Number	Cost from out of pocket money of patients
Outpatient only		
FUNC_DATE	Date, YYYYMMDD	Visit date
ACODE_ICD9_1	Character, ICD-9CM code	Main diagnosis
ACODE_ICD9_2 ACODE_ICD9_3	Character, ICD-9CM code	Minor diagnosis; optional
Inpatient only		
IN_DATE	Date, YYYYMMDD	Entering date
OUT_DATE	Date, YYYYMMDD	Leaving date
ICD9CM_CODE	Character, ICD-9CM code	Main diagnosis
ICD9CM_CODE_1 ICD9CM_CODE_2 ICD9CM_CODE_3 ICD9CM_CODE_4	Character, ICD-9CM code	Minor diagnosis; optional

Prescribed orders (OO for outpatients, DO for inpatients)

The dataset stores the records of the medications prescribed to patients.

- Age can be inferred from ID_BIRTHDAY
- For outpatients, total drug days = TOTAL_QTY ÷ DRUG_FRE
- For inpatients, total drug days is total hospitalised days

Table F.4: List of variables

Variable	Data type	Description
ORDER_SEQ_NO	Character	Primary key; sequence number of order
SEQ_NO	Character	Foreign key, referring to visit data
Outpatient only		
DRUG_NO	Character	Order code
DRUG_FRE	Number	Drug use frequency
TOTAL_QTY	Number	Total drug quantity
Inpatient only		
ORDER_CODE	Character	Order code

Appendix G

Individual Patient Pathway Analysis: Technical Guide

Overview

The objective of this appendix is to explain how to conduct an Individual Patient Pathway Analysis (IPPA) from individual healthcare data. This guidance is targeted for analysts to reason about and to implement the IPPA in their intended setting. The IPPA was developed using the information on TB patients in the National Health Insurance Research Database, Taiwan, though this guidance will indicate options to adapt to different data.

IPPA was designed to profile the efficacy of a healthcare system in responding to TB patients as well as to identify the obstacles of TB patients obtaining successful TB diagnosis and treatment. The IPPA approach was intended for use with passively and routinely collected healthcare utilisation data to inform the case seeking before patients being labelled as TB patients. Unfortunately, TB has many features overlapping with other respiratory diseases, such as pneumonia, and chronic lung disease. The presence of many non-TB related data entries is to be expected. In addition, healthcare utilisation data includes the use of diagnostic procedures and treatment, but the respective outcomes are not always available. These issues are dealt with during the IPPA implementation by inferring the results from subsequent events.

In general, the IPPA has two stages: patient pathways extraction and subsequent analysis of pathways. From the beginning, the data are usually collected with the perspective of healthcare providers in mind. The first stage of the IPPA translates the healthcare data to patient pathways, shifting information from provider-centred to patient-centred viewpoints. The extracted patient pathways are the first product of the IPPA. The second stage summarises and visualises the collected patient pathways.

The guidance is formulated as follows: Section 1 lists the terminology used during the IPPA. Section 2 prepares the IPPA ready data; Section 3 reads the data to a set of state-space dimensions; Section 4 trims the unnecessary information out and identifies care seeking episodes; Section 5 augments patient pathways by labelling stages of care seeking; Section 6 computes statistics from the patient pathways and Section 7 visualises the patient pathways step by step.

1. Terminology

<table>
<tr><th>Term</th><th>Definition</th><th>Example</th></tr>
<tr><td>Record</td><td>A healthcare record with variables of diagnosis, prescriptions, a visit time.

The diagnosis in the IPPA is the suggestion but not the actual health status.</td><td>
<table>
<tr><td>Time</td><td>Diagnosis</td><td>Medication</td></tr>
<tr><td>5</td><td>ARD*</td><td>None</td></tr>
<tr><td>10</td><td>ARD</td><td>Fast-acid test</td></tr>
<tr><td>20</td><td>TB</td><td>anti-TB drugs</td></tr>
</table>
ARD: acute respiratory disease
</td></tr>
<tr><td>Dimension</td><td>A state-space time-series linking a group of relevant records to inform the transition of state.</td><td>Related illness, capturing illness which can be initial consideration of a TB patient, or comorbidities with overlapped features of TB.

Evaluation, capturing the prescriptions of screening or diagnostic tools to identify TB.

Treatment, capturing the prescriptions of TB treatment</td></tr>
<tr><td>Value in dimension</td><td>A value reflects the information in the record given a dimension.

Zero value indicates there is no relevant record during a period of time.</td><td>In the Evaluation dimension,
Zero: no ongoing evaluation
Possible: diagnostic tools which might suggest TB
Probable: diagnostic tools which can identify TB</td></tr>
</table>

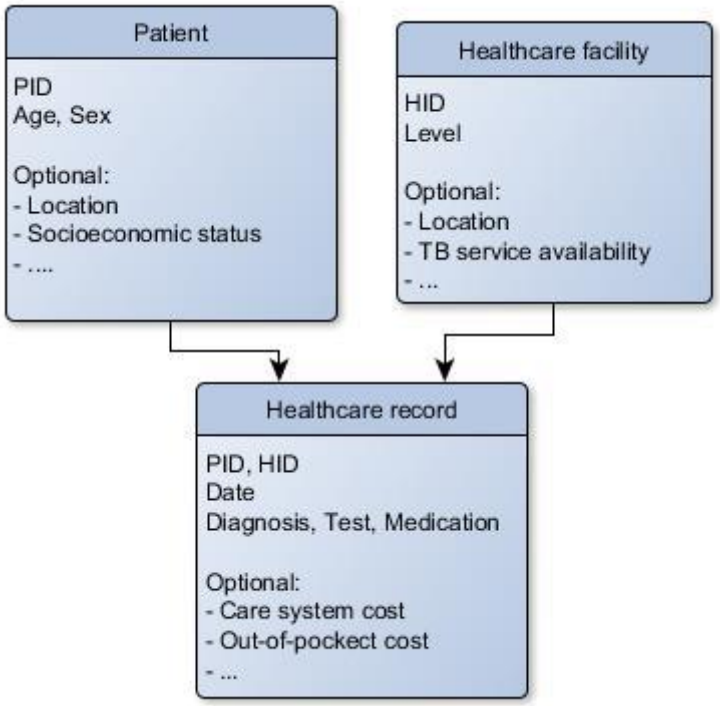
<table>
<tr><td>Time-out</td><td>The persistence of a healthcare record. The period will be extended if the next records contain similar information.</td><td></td></tr>
<tr><td>Episode</td><td>A collection of values of dimensions separated by periods with all dimensions equal to zero.

Non-TB episode: episodes with any of dimensions is not zero while none of the records during the episodes met the definition of confirmed TB.</td><td><table><tr><th>Time</th><th>Related illness</th><th>Evaluation</th><th>Treatment</th></tr><tr><td>105</td><td>ARD*</td><td>Zero</td><td>Zero</td></tr><tr><td>110</td><td>ARD</td><td>Possible</td><td>Zero</td></tr><tr><td>120</td><td>ARD</td><td>Possible</td><td>First line</td></tr><tr><td>...</td><td>...</td><td>...</td><td>...</td></tr></table>*ARD: acute respiratory disease</td></tr>
<tr><td>Patient pathway</td><td>An episode with a series of stages and the timing of progression.

Stage, indicating the progress of a pathway of a patient
Waiting, the patient not received and TB related medication yet after initial care seeking.
Evaluating, the patient is under evaluation but possibly for TB
Detecting, the patient is being considered as a potential TB patient.
Treating, the patient has been identified as a TB patient and initialised TB treatment.
State, specifying details of a stage. For example, the TB drug regimen</td><td><table><tr><th>Time</th><th>Stage</th><th>State</th></tr><tr><td>0</td><td>Waiting</td><td>Waiting</td></tr><tr><td>10</td><td>Evaluating</td><td>Evaluating</td></tr><tr><td>70</td><td>Evaluating</td><td>Interrupted Evaluation</td></tr><tr><td>90</td><td>Evaluating</td><td>Re-Evaluating</td></tr><tr><td>100</td><td>Detecting</td><td>Re-Detecting</td></tr><tr><td>120</td><td>Treating</td><td>First line treatment</td></tr><tr><td>300</td><td>End</td><td>Treatment completion</td></tr></table></td></tr>
</table>

2. Preparing IPPA ready data

The section guides how to manage data to which are ready to input to the IPPA. Three datasets (see figure), are required for the IPPA: patient information, healthcare facility information, and healthcare records. The patient information contains the basic variables of patients which are relevant to the decision making of either patients or healthcare providers. The healthcare facility information contains the type or the level of the facilities presented in the care-seeking records. The healthcare records, the most important dataset, contains the suggested clinical status (or diagnosis), prescriptions of diagnostic tools and medications, and keys referred to the patient and healthcare facility information.



Patient information

Every patient entry requires a key for their identity (PID), linking to the healthcare records of the patient. Considering further analysis with the related variables, basic information, such as age and sex, are suggested. Other variables, such as socioeconomic status and comorbidities, can be included if of interest and available.

Healthcare facility (or hospital) information

Every hospital entry requires a key for its identity (HID), linking to the healthcare records issued in the hospital. Level of hospitals is a necessary variable in this dataset. The capacity of TB services can be included if available as an external hospital dataset. Otherwise, the IPPA considers if a service has been used in as hospital to indicate capacity of the service. Area-specific information can be considered if spatial diversity/disparity is of interest.

Healthcare records

Healthcare records are the central elements of the IPPA. Apart from the foreign keys linking to the patient and hospital data, the records should contain timing, diagnosis, evaluations, and treatments. The diagnosis includes TB diagnosis and that of related illness, comorbidities and diseases which could potentially be TB. The evaluations of TB should be ordered data based on their capabilities to identify TB. Note that the sensitivity and specificity of diagnostic tools might be different from setting to setting, as may the order of use. The previous appendix of definitions, Appendix A, lists the related diseases and medications considered in our application to the Taiwanese setting.

3. Transforming healthcare records into a three dimension system

A patient pathway considers three dimension of information: related illness, evaluation, and treatment. First, Related illness captures the illness which can be initial consideration of a TB patient, or comorbidities with overlapped features of TB. Second, Evaluation captures the prescriptions of screening or diagnostic tools to identify TB. Last, Treatment captures the prescriptions of TB treatment. Every record contains information in different domains, and some records can inform multiple domains. Therefore, this step is designed to pass the information in each record to the intended dimensions.

The IPPA builds three state-space dimensions based on the three domains. However, the data of healthcare utilisation might not contain the duration of state supported by each record. For example, TB culture can take several weeks to complete. The time to having the test result and following actions can be affected by the workload of the clinical laboratory as well as the schedule of the patient. As for anti-TB drug prescriptions, we can technically infer the duration through the number of drugs. However, how regularly the drugs are being taken depends on the implementation of DOTS and the schedule of patients to revisit hospitals. To correctly address the durations requires very detailed data. In such cases, the IPPA introduces a “Time-out” limit to presume the durations. In each domain, a specific value is set. As a default, we put 60 days for Related illness and Evaluation after visits and 30 days for Treatment after the drugs apparently ran out.

Starting from Zero state of each dimension, IPPA scans records iteratively. For every record, each dimension checks whether the record is relevant to it or not. If it is relevant, the dimension will transition the respective state. Before that, if the record occurs later than the Time-out period since the last relevant record, the dimension will transition to Zero. After scanning all records, the IPPA will check the end of the dimensions, considering the end of the data timespan and death time if present.

4. Splitting Episodes and trimming unnecessary records off

In the data used - all the healthcare records over given a period - many non-TB care-seeking records are to be expected. However, they are necessary for completing the IPPA because some of them may be early events in the diagnosis of TB. As the previous step has joined the records in the same domain to formulate the system of three dimensions, this step is going to join the non-zero states in different domains.

Once the system of three dimensions aligned, we can locate the periods with all dimensions equal to zero. The step splits the three dimensions by these periods to specify episodes. For the separated episodes, we identify the episodes have confirmed TB (in our demonstration, ICD-9-CM codes for TB and more than two types of anti-TB drugs prescribed over 28 days), and drop the others. Each of the kept episodes will be formulated as patient pathways thereafter.

This step allows construction of episodes that included records related to care-seeking prior to TB diagnosis as possible. Also, it trims the healthcare records which are not relevant to TB care off. To be noted that, whether an episode is TB related or not is affected by the length of the “Time-out” period. For a short “Time-out”, an episode will be fragmented as several episodes. The care-seeking before TB diagnosis will be ill-addressed. For a long “Time-out”, TB episodes will be mixed with irrelevant episodes. (See the appendix of sensitivity analysis).

5. Finishing patient pathway construction

A TB episode includes the dynamics of the three state space dimensions but not the meaning of them. This step concludes the three dimensions into a patient pathway. For every separated TB episodes, this step. The stages and states in patient pathways should summarise by not only the current states of the three dimensions but also the states before and after. For example, the “Re-Evaluating” state indicates (1) evaluations are going on, (2) previous evaluations were interrupted, and (3) TB treatment has not started yet. As in the following definition, this step labels care-seeking stages considering the states of the three dimensions.

Waiting Stage

In Waiting Stage, the patients have started their care-seeking at hospitals but are not considered as potential TB patients. Namely, they are “waiting” for TB-related evaluation.

State	Dimensions*	Contexture	Note
Waiting	R: non-zero E: Zero T: Zero	Start with initial care-seeking	Waiting for the first TB-related evaluation or treatment.

* R: related illness, E: evaluation, T: treatment

Evaluating Stage

In the Evaluating Stage, physicians start to use evaluation techniques for the patients. TB might not be a consideration in the evaluation, but the techniques should be able to narrow the possibilities down until TB-specific evaluations prescribed.

State	Dimensions*	Contexture	Note
Evaluating	E: Possibly T: Zero	No IE before	Under evaluations which can narrow the possibility down to TB
Interrupted Evaluation (IE)	R: non-zero E: Zero T: Zero	After a period with the evaluation dimension in a non-zero state	The previous evaluation does not narrow the possibility down to TB which might be because of (1) comorbidity, (2) false negative, or (3) self-referral.
Re-Evaluating	E: Possibly T: Zero	Have IE before	

* R: related illness, E: evaluation, T: treatment

TB Detecting Stage

In TB Detecting Stage, the evaluation techniques which can identify TB if clinicians are well-trained. Apart from the techniques, anti-TB drugs with doses below regular regimens can be considered in this stage as well, namely, empirical treatment or treatment initialisation.

State	Dimensions*	Contexture	Note
TB-Detecting	E: Probably T: not meet 1st line or 2nd line	No IE before	Being evaluated by procedures which can identify TB (Evaluations probably for TB)
Interrupted Evaluation (IE)	R: non-zero E: Zero T: Zero	After period with the evaluation dimension in a non-zero state	IE can be a state in Evaluating Stage or TB-Detecting Stage. It depends on the most TB specific evaluation used before.
Re-Detecting	E: Probably T: Zero	Have IE before	Re-visiting TB-Detecting State after Interrupted Evaluation

* R: related illness, E: evaluation, T: treatment

Treating Stage

Once TB confirmed, Treating Stage will start. The stage captures the intensity of treatment used.

State	Dimensions*	Contexture	Note
First-line treatment	T: 1st line	Between two periods with the treatment dimension in non-zero states	Being treated with first-line TB regimen.
Treatment change	R or E: non zero T: see contexture	A zero treatment dimension between two periods of treatment dimension in non-zero states; or the timing when treatment intensity increases.	Switching between two TB treatments or temporal treatment interruption. Could be zero duration.
Second-line treatment / retreatment	see contexture	T: 1st line after Treatment change or T: 2nd line	Being treated with second-line TB regimen or any regimen after Treatment change.

Treatment Outcome

Treatment Outcomes are the end of patient pathways. The labels are totally customizable for different settings Datasets of treatment outcome and death registration outside healthcare records can be linked if available.

In our demonstration setting, registering in the National Health Insurance programme in Taiwan is compulsory. We used leaving the programme as the indicator of death.

Outcome	Definition
Treatment completed	> 180 days of treatment period
Censored	Reach the end of data
Dead	Leaving the National Health Insurance programme.
Lost to follow-up (LTFU)	Other

6. Statistics and indices of the Individual Patient Pathway Analysis

Hospital-level of initial care-seeking features the places to start care-seeking for the patients with TB symptoms. The levels of hospitals or facilities are defined differently setting by setting. Our demonstration only presents four levels while it can be disaggregated by private and public sectors, by hospital divisions, or by counties. In Taiwan, the National Health Insurance covers both private and public sectors with the same services, so we did not separate them.

Coverage of TB services denotes the percentage of hospitals have TB services available. The TB services are the evaluation tools with respect to the medications in Evaluating, TB Detecting, and Treating stage of patient pathways. The denominator is the number of hospitals has ever been visited in the collected patient pathways; the numerator is the number of hospitals has ever provided respective TB services during the data timespan. In the original IPPA, the index is measured by whether the hospitals can provide respective TB services. That might overestimate the values if the services are not considered in practice. On the contrary, our approach might underestimate the values in the case of the hospitals can provide the services but have not met the situations to use. We suggest reporting values from both approaches if possible as lower and upper bounds.

Hospital-level of stages denotes the distributions of hospital levels at the starts of stages. The options are the same as **Hospital-level of initial care-seeking.**

Hospital-level of notification features the places which notify TB cases to authorities. In the setting of Taiwan, and so our demonstration, the notification process and treatment prescribing are synchronised. The distribution was the same as **Hospital-level of Treatment Stage**, so we did not show them separately.

Accessibility to TB services at initial care seeking denotes the percentage of patient pathways which have initial care seekings at a hospital which has TB services available (evaluations or treatments). With the individual hospital data identifiable, it is calculated by (1) linking initial care-seeking records to the availability of the corresponding hospitals, and (2) finding records can access to TB services within all initial care-seeking records. However, if the data does not support individual hospitals, applying the original approach in the PPA [1] is suggested.

Time to arriving at a hospital with TB treatment denotes when is the first time seeking care at a hospital which can provide TB treatment. The period starts with the time of initial care-seeking. Again, the availability of TB treatment in our demonstration was based on whether the hospital has provided TB treatments or not during the timespan.

Time to arriving at the hospital initialise TB treatment denotes when is the first time seeking care at the hospital which ultimately provides TB treatment to the patient.

System Delay, time to treatment start, denotes the length of the period between the initial care-seeking and the start of regular TB treatment. This is the common output of care-seeking delay studies. However, the definitions may be different for different data sources.

No System Delay indicates a patient pathway starts with regular TB treatment prescription at the initial care-seeking. We suggested reporting this and System Delay and a proportion

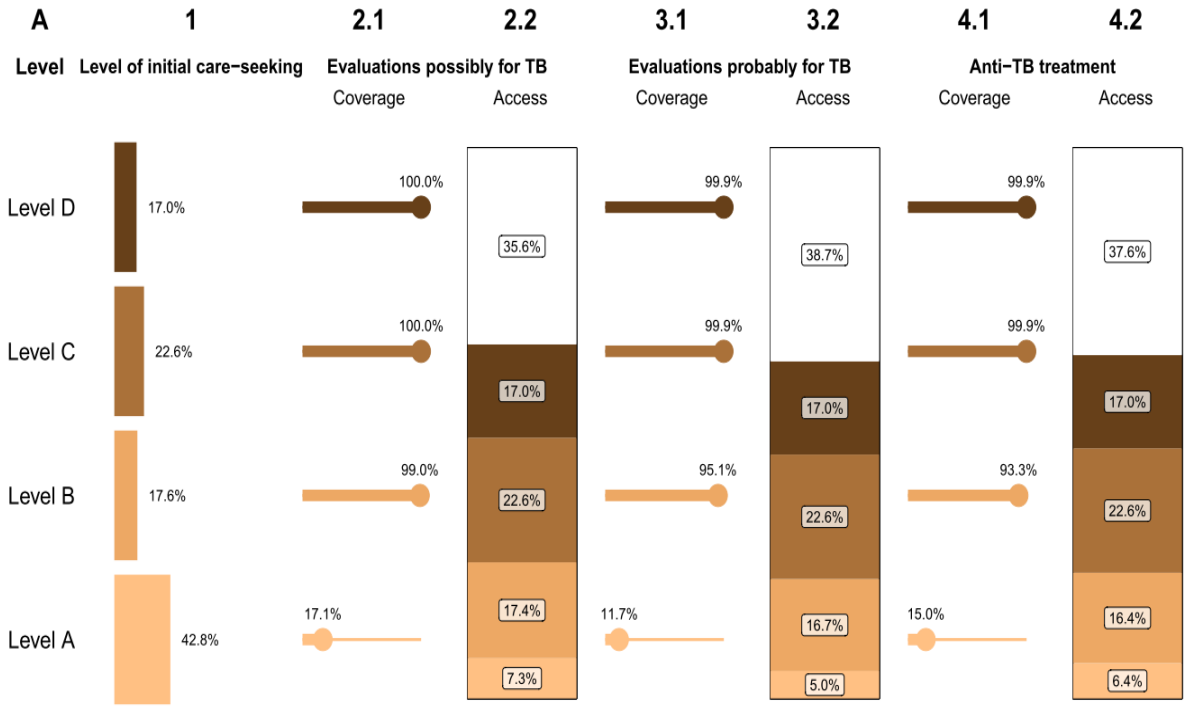
Interrupted Evaluation indicates at least one Interrupted Evaluation in a patient pathway. The presence shows if the diagnosis has been put on hold, potentially in favour of other investigations. The percentage of Interrupted Evaluation quantifies the complexity of TB diagnosis in a healthcare system. Applying covariate analysis can identify the risk factors related to interruption. The presence of Interrupted Evaluation has multiple meanings. Firstly, comorbidities, especially chronic lung diseases, can mask the features of TB. As the respiratory symptoms of TB are not unique, the clinicians might not aware active TB from patients who have had those symptoms already. Secondly, there are competing diseases to be considered before diagnosing a patient as a TB patient. For example, in the Western Pacific region, TB is prevalent in the old population and is sometimes similar to pneumonia. Considering the higher fatality rate of pneumonia than TB, TB could be less urgent. Thirdly, the false negative of any TB test results and false positive of non-TB test results might happen. Either of them can cause interrupted evaluation.

System cost, out-pocket cost, and healthcare contacts (optional) in each stage measure the burden for the perspectives of a healthcare system and patients. Using the records grouped by the stage when they occurred, these three indices are summarised. The number of healthcare contacts is a proxy of the time cost of a patient pathway during a stage. This should be available since it corresponds to the input of the IPPA, but the other two depend on availability. Additional costing survey is suggested to provide more precise measures and to broaden the IPPA.

7. Visualisation

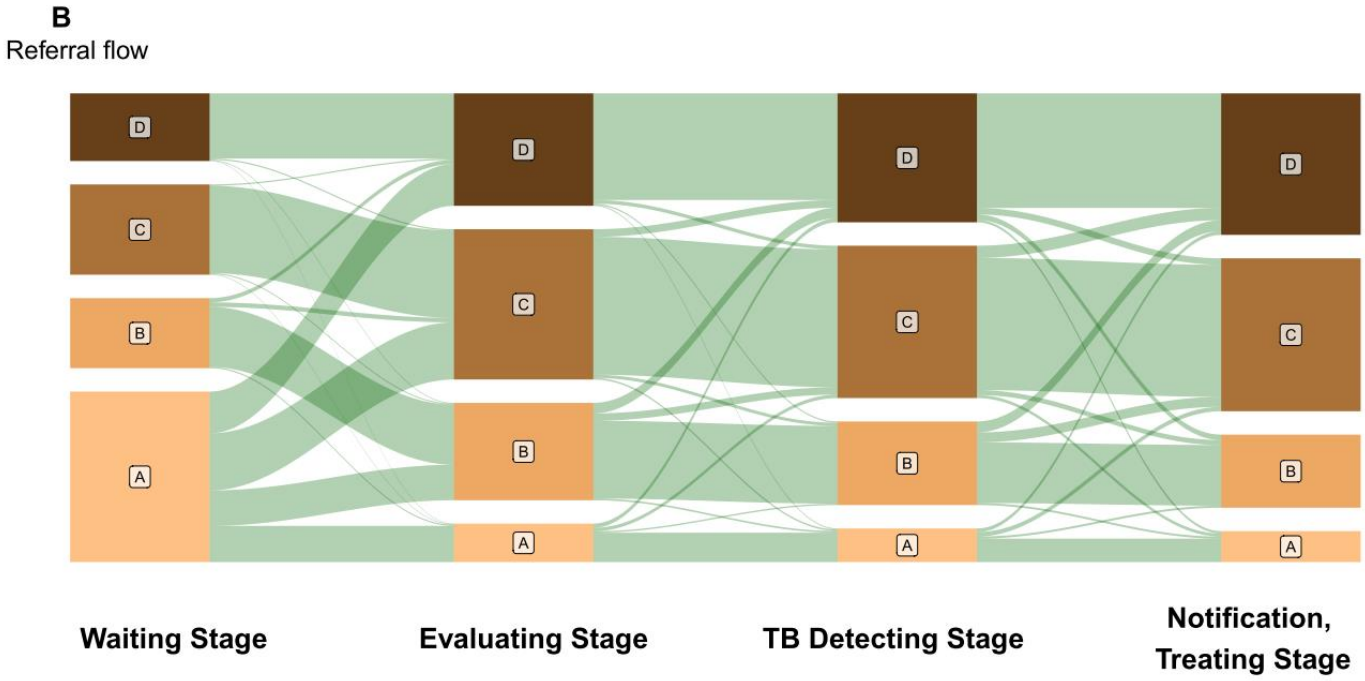
Accessibility and Coverage

The figure (as Fig 2A in the paper) is adapted from the original PPA. Seven columns are included in the diagram. The A-1 shows the places of initial care-seeking by hospital level. A-2.1 shows the coverages of evaluation services in proportions by hospital level. This is from the perspective of healthcare providers. A-2.2 shows the accessibility at initial care-seeking. This is from the perspective of patients, measuring if the patients seek care at proper places (See Statistics and Indices). A-3 and A-4 replicate A-2 but targeted at Evaluations probably for TB and treatment.



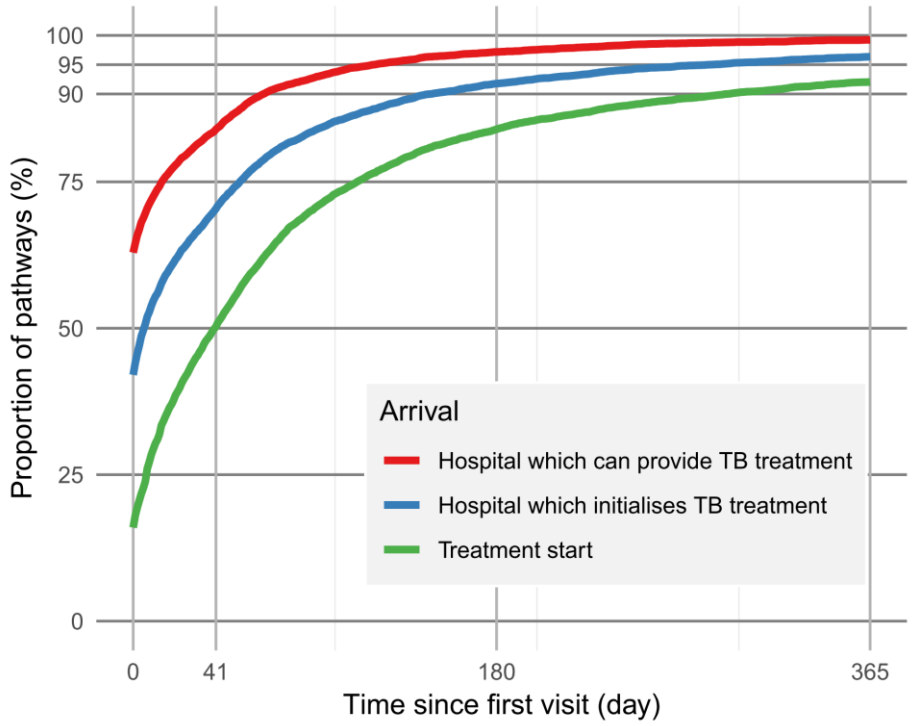
Referral Flow

Figure 2 (as Fig 2B in the paper) shows an example of the referral flow diagram. The diagram is a Sankey diagram describing the progress in hospital levels during care-seeking. Four columns match the four stages of patient pathways: Waiting, Evaluating, TB Detecting, and Treating. The blocks in each column are the hospital levels in each stage. For a pathway, the hospital level in a stage is based on the hospital of the first care-seeking in the stage. While some pathways do not have all four stages, the diagram uses the start of the next stage for the current stage. For example, a pathway started with TB-Detecting Stage at a Level B hospital. The hospital levels in Waiting and Evaluating Stages are Level B as well. The heights of the blocks are determined by the counts of the pathways. The ribbons between columns link the current hospital levels to the hospitals triggering the next stages.



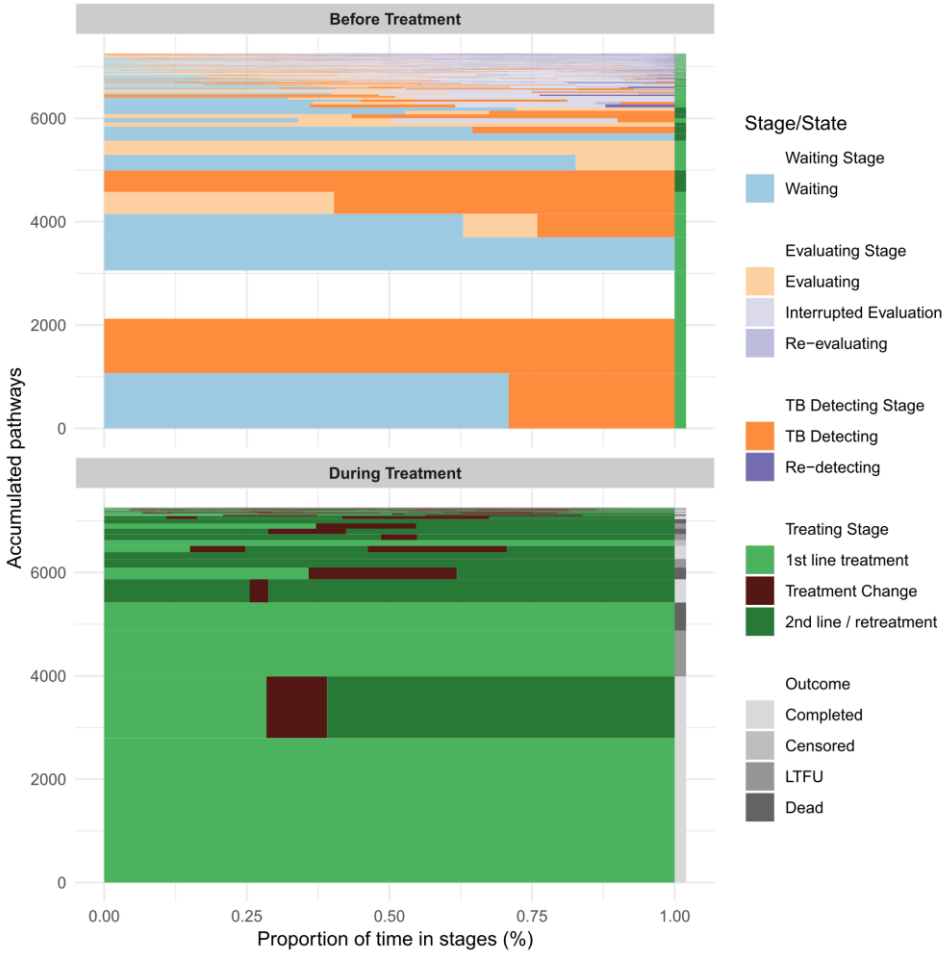
Care-seeking Gap

Figure 3 (as Fig 3 in the paper) features the gaps between treatment accessible and treatment start. Three checkpoints are considered. The first is when a patient arrived at a hospital with TB treatment. From this checkpoint, the patient starts having a chance to be under TB treatment. The second is when a patient arrived at the hospital which provides TB treatment for them ultimately. From this checkpoint, the inter-hospital referral has been completed in the pathway. The last is when TB treatment starts. The curves are the cumulative percentage of pathways that have reached the respective checkpoints. The difference between the first and the second checkpoints (and their curves) can be regarded as a care-seeking gap due to inter-hospital referrals. The difference between the second and the third checkpoints can be regarded as a gap due to intra-hospital referrals and diagnostic dimensions.



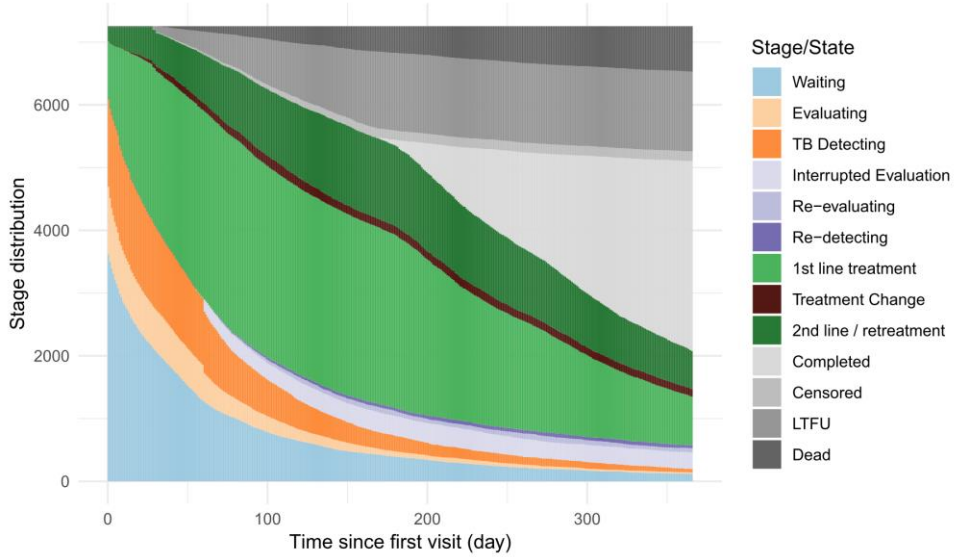
Pattern Frequency

Figure 4 visualise the patterns of patient pathways before and after treatment start. The pattern means the series of states in a pathway without considering their timings in the stages. For example, a pathway starting Waiting state at day 0, Evaluating state at day 30, Treating state at day 40 has the pre-treatment pattern of “Waiting-Evaluating-Treating”. In the figures, each horizontal bar represents a pattern; the height indicates the number of pathways having this pattern; the width indicates the proportion of time spent in each state on average. The tiny rectangles on the right indicate the last states of patterns while the widths are meaningless. The blink bar in the pre-treatment figure quantifies the pathways having treatment at initial care-seeking. The bars are sorted by the numbers of pathways in the respective patterns. The figures are used to highlight the heterogeneity of the pathways.



Stage Distribution

Figure 5 visualises the distribution of states of patient pathways. Aligning patient pathways by the time of initial care-seeking, every column in the figure indicates state distribution in one day; the height in each block indicates the number of pathways in the state at the day; for the pathways ended before the day, the treatment outcomes are extended to the day. The figure is used to understand the time spent in each state and to know the accumulation of treatment outcomes. For our example, the Interrupted Evaluation formed a flat tail, which dying out slowly, in the figure.



Reference

1. Hanson CL, Osberg M, Brown J, Durham G, Chin DP. Conducting Patient-Pathway Analysis to Inform Programming of Tuberculosis Services: Methods. J Infect Dis. 2017;216: S679–S685.

Appendix H

Individual Patient Pathway Analysis: Sensitivity analysis

In the individual patient pathway analysis (IPPA), a "time-out" system is applied to define whether two healthcare records is related or not. Here, we tested the sensitivities of Time-out for Related illness domain (TOR) and Time-out for Evaluation domain (TOE). We varied them from 30 days to 120 days by 30 days and (TOR, TOE) = (60 days, 60 days) were used as a comparator.

Throughout the sensitivity analysis, we will use two types of diagrams to demonstrate how the selected indices reflect on the time-out changes. The first uses heatmaps, mapping the indices given different time-out values (i.e. Figure H.1). The second is barplot (or unordered tornado chart) which shows the marginal changes of the indices while increasing or decreasing by 30 days (i.e. Figure H.2). The changes were measured by

Difference $N_{tor,toe} - N_{60,60}$ or

Change rate $\frac{N_{tor,toe} - N_{60,60}}{N_{60,60}} \times 100\%$,

depending on variable types.

H.1 Number of pathways

This section assesses the number of pathways rendered by the IPPA with different TOR and TOE. The changes were measured by the change rate. Figure H.1 shows that the number of pathways is negative corrected with both TOR and TOE. The changes were within 5%. Figure H.2 shows that increasing TOR and TOE caused higher change rates than decreasing while the number of pathways is more sensitive to TOE.

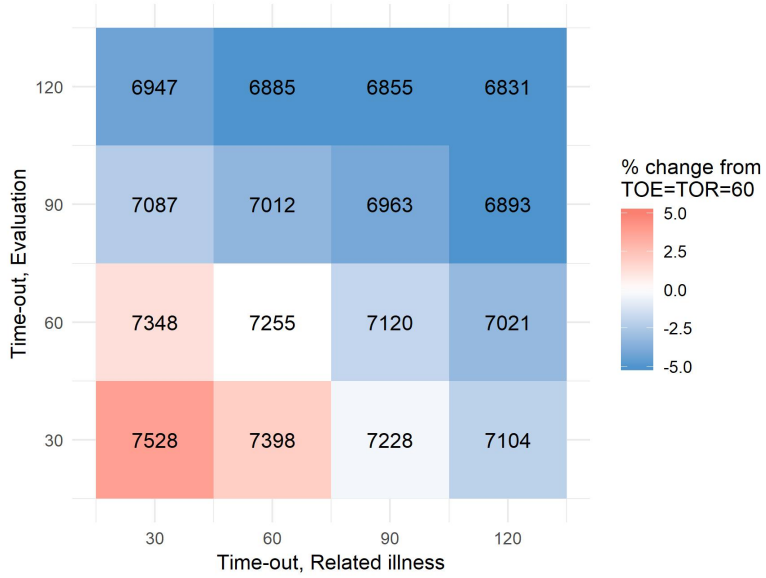


Figure H.1: Heatmap: Number of Pathways

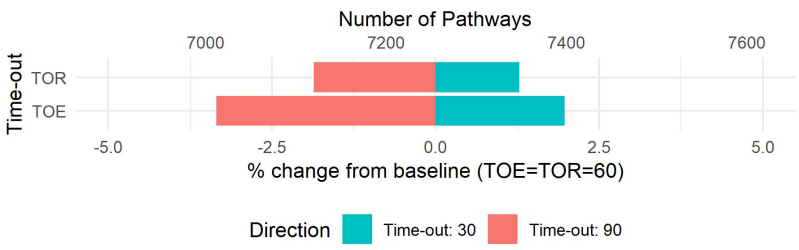


Figure H.2: Marginal changes: Number of Pathways

H.2 Length

This section assesses the length of patient pathways.

- **Number of Contacts, median**: the length considering how many healthcare contacts happened during the patient pathways started from initial care-seeking to treatment end.
- **System Delay, median**: the duration from initial care-seeking to treatment start.
- **Pathway Length, median**: the duration from initial care-seeking to treatment end.

These indices were summarised by median, and the changes were measured by change rate.

As Figure H.3 shows, the numbers of contact ranged from 16 to 25 while the values were positively correlated to TOR and TOE. Figure H.4 addresses the system delays, showing the values were sensitive to both TOR and TOE. When TOR and TOE were both 120 days, the change is more than being doubled compared with 60 days. Figure H.5 focuses on the pathway lengths. Figure H.6 summarises the sensitivity of the length of pathways. The system delay was the most sensitive to TOR and TOE, while the total pathway length was the last. TOR and TOE had equal impacts on the number of contacts in this analysis. Increasing TOR and TOE brought more changes than decreasing across these three indices. The high sensitivity of the median system delay suggested an external validation with future interview data in the same setting. A previous study, Chen et al. [1], estimated the system delay in Taiwan was 29 days (interquartile range 5–73) with the same database and TB definition as my study. Although their assessment did not consider interrupted evaluations and patients having chronic lung conditions, and so their estimates constitute a lower-bound for our approach. Comparing with other settings, Sreeramareddy et al. [2] summarised 52 studies, finding the system delays to TB treatments ranged from 2 to 87 days, finding that the low-income and high-income settings did not have a significant difference. However, the retrieved studies in their review showed an imbalance in that the studies with longitudinal data were conducted in specific hospitals or sub-populations, while the studies that covered the general population were cross-sectional. My study, which used longitudinal data on the general population, therefore, cannot be compared with them directly. Therefore, I suggest using a retrospective design with interviewing patients embedded in the longitudinal data. This approach can validate the

lengths of patient pathways from the IPPA and highlight the difference from perspectives of patients and the health system.

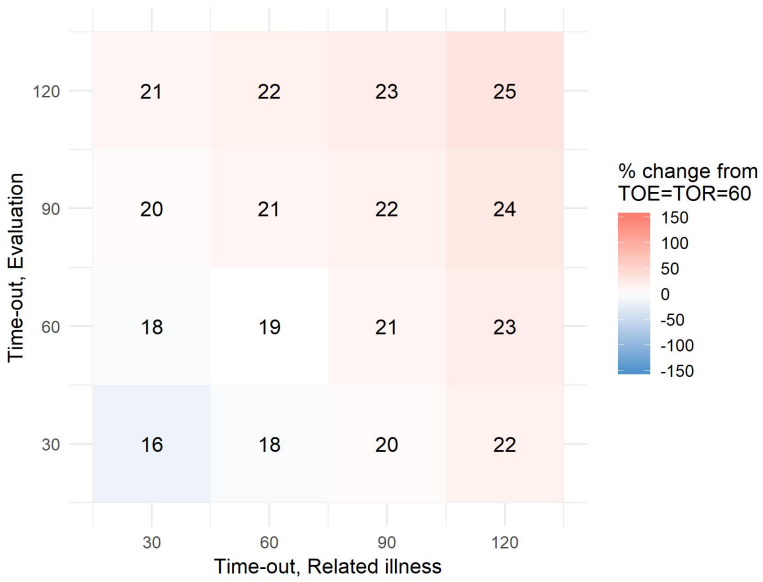


Figure H.3: Heatmap: Number of contacts in median

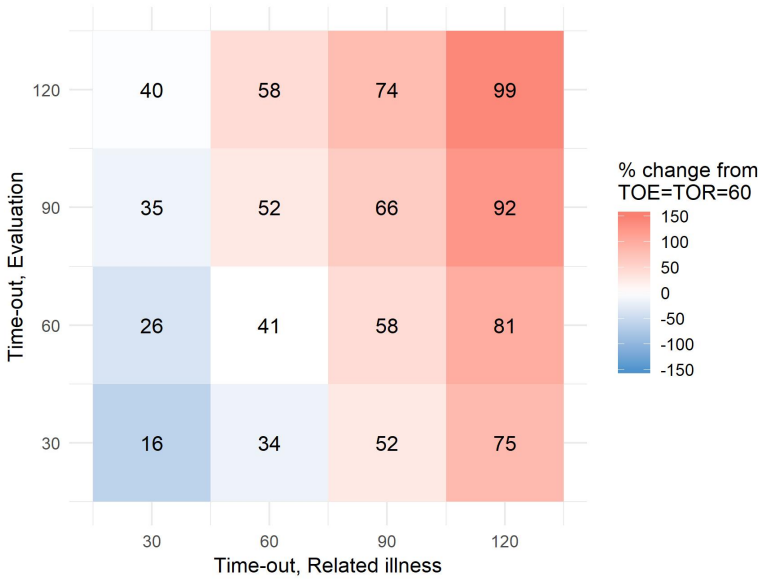


Figure H.4: Heatmap: System delay in median

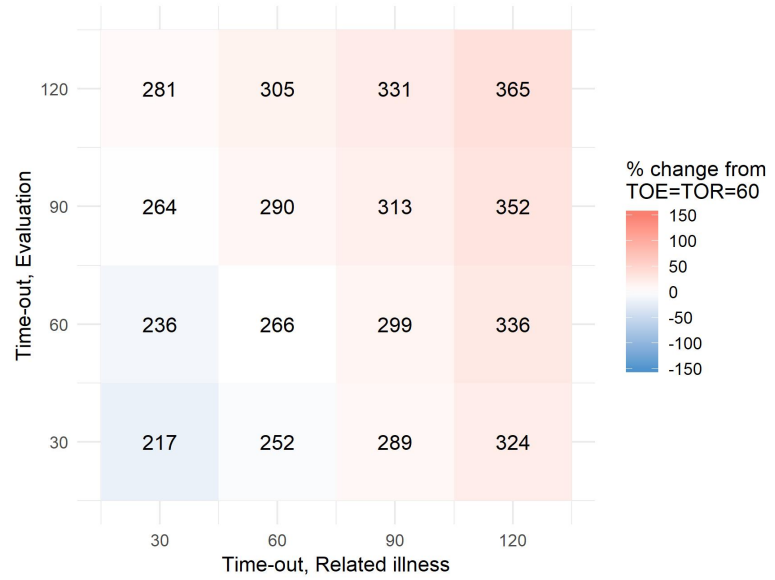


Figure H.5: Heatmap: Length of Pathways in duration

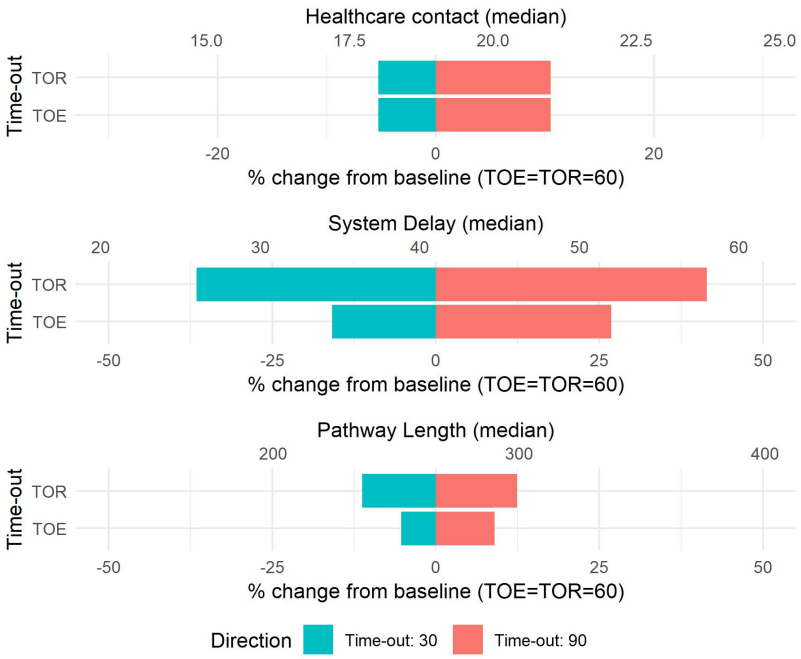


Figure H.6: Marginal changes: Length of Pathways

H.3 Typology

This section assesses the typology of patient pathways.

- **Initialised at Level A Hospital**: whether the initial care-seeking of pathways were in Level A hospitals.
- **Interrupted Evaluation**: if the pathways experienced interrupted evaluation.
- **Zero Delay**: indicates if the pathways started their treatment at the day of initial care-seeking.

These indices were summarised by proportion, and the changes were measured by difference.

As Figure H.7 shows, **Initialised at Level A Hospital** were no sensitive to TOE while that and TOR were positively correlated. Figure H.8 highlights higher TOR led more **Interrupted Evaluation** but TOE had negative influence. **Zero Delay** in Figure H.9 shows negative correlations with TOR and TOE while the two Time-outs were equally contributed. Figure H.10 summarises the sensitivity of the typology of pathways. The marginal changes were usually smaller than 5%. However, **Interrupted Evaluation** was very sensitive to TOR and TOE compared with the other two indices.

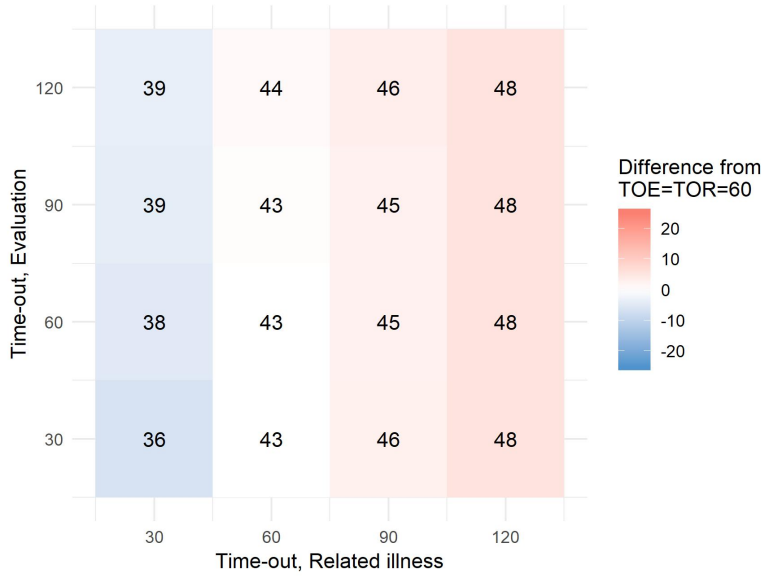


Figure H.7: Heatmap: Initialised at Level A hospital (%)

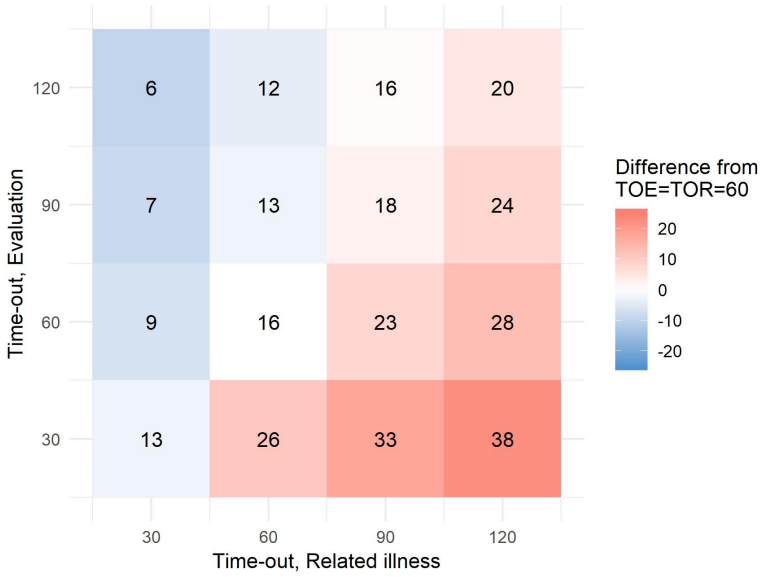


Figure H.8: Heatmap: Interrupted Evaluation (%)

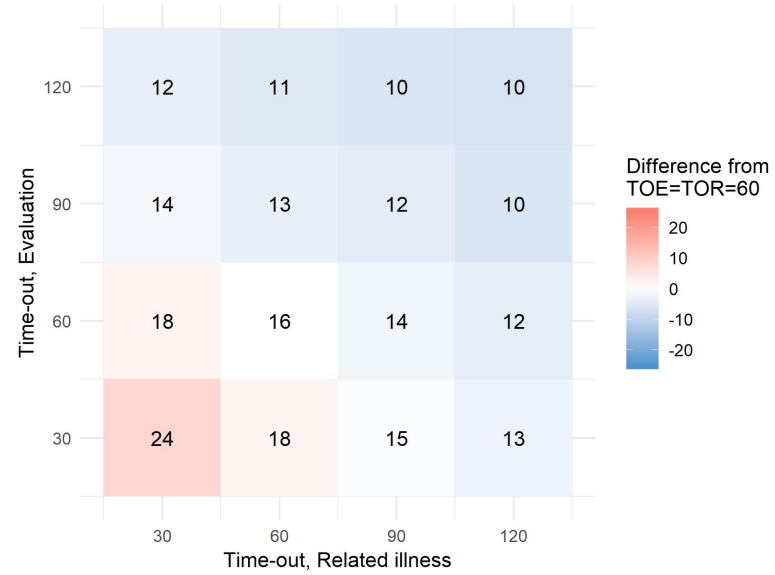


Figure H.9: Heatmap: Zero Delay (%)

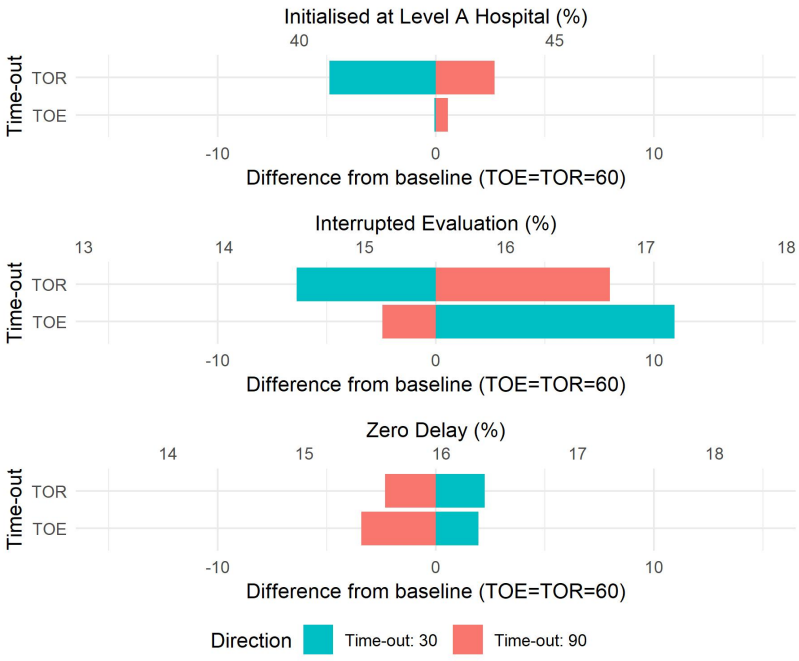


Figure H.10: Marginal changes: Typology

Bibliography

[1] Chen CC, Chiang CY, Pan SC, Wang JY, Lin HH. Health system delay among patients with tuberculosis in Taiwan: 2003-2010. BMC Infect Dis. 2015 Nov;15(1):491.

[2] Sreeramareddy CT, Panduru KV, Menten J, Van den Ende J. Time delays in diagnosis of pulmonary tuberculosis: a systematic review of literature. BMC Infect Dis. 2009 Jun;9(1):91.

Appendix I

Individual Patient Pathway Analysis, Algorithms

Algorithm I.1 Patient pathway extraction: overview

Data: $records$: an array of healthcare records of an individual
TOR: Time-out of *Related illness*
TOE: Time-out of *Evaluation*
TOT: Time-out of *Treatment*
Result: Patient pathways of a TB patient

```
1  /* Initialise dimensions */;
2  rel ← New Related illness dimension;
3  eva ← New Evaluation dimension;
4  tre ← New Treatment dimension;
5  /* Read all records in each dimension */;
6  rel.readRecords(TOR, records);
7  eva.readRecords(TOE, records);
8  tre.readRecords(TOT, records);
9  /* Collect dimensions and cut them into episodes */;
10 episodes ← cutEpisodes(rel, eva, tre);
11 pathways ← ∅ /* Initialise a empty set collecting pathways*/;
12 foreach episode ∈ episodes do
13     pathway ← formulatePathway(episode);
14     pathways.append(pathway)
15 return pathways;
```

See *dim*.readRecords(...) in Algorithm I.2
See cutEpisodes(...) in Algorithm I.3
See formulatePathway(...) in Algorithm I.4

Algorithm I.2 Read records in each dimension (dim.readRecords(...))

Data: dim: dimension, rel, eva, or tre
$timeout$: timeout for the selected dimension
$records$: an array of healthcare records of an individual
Result: the transition history in the dimension

```
/* Initialise the dimension with a Null event */;
time_curr ← 0;
time_wait ← ∞ /* the end of waiting time */;
state ← Null;
Initialise dim with state at time_curr;
foreach record ∈ records do
    time_curr ← time of record;
    if time_curr > time_wait then
        /* Reset state */;
        time_wait ← ∞ /* the end of waiting time */;
        state ← Null;
        dim transits to Null at time_wait;
    end
    if record is relevant to dim then
        time_wait ← time_curr + timeout;
        if record can progress state then
            /* Progress */;
            state ← the matched state of record;
            dim transits to state at time_curr;
        end
    end
end
/* Close record reading */;
dim transits to Null at time_wait;
```

Algorithm I.3 Collect dimensions and cut them by periods without events ($cutEpisodes(...)$)

Data: rel: state history in related illness dimension
eva: state history in evaluation dimension
tre: state history in treatment dimension
Result: A set of care seeking episodes

$ts \leftarrow \emptyset$ /* A collection storing state transition times */;
foreach $dim \in [rel, eva, tre]$ **do**
  forall *State transition time* t *of* dim **do** ts.add(t);
end
Remove duplicated time points in ts Sort ts (ascending);
$episodes = \emptyset$ /* A collection for storing episodes */;
foreach $t \in ts$ **do**
  if *all[*rel, eva, tre*] are* $Null$ *at* t **then**
    /* Separate state-transition history*/;
    $x \leftarrow$ state history before t split from rel;
    $y \leftarrow$ state history before t split from eva;
    $z \leftarrow$ state history before t split from tre;
    /* Join dimensions */;
    $episode \leftarrow [x, y, z]$;
    $episodes$.append($episode$);
  end
end
return $episode$;

Algorithm I.4 Patient pathway formulation ($formulatePathway(...)$)

Data: $episode$: an episode with state transition history in rel, eva, and tre
Result: A patient pathway

/* Identify key information */;
$t_{eva} \leftarrow$ time of first evaluation possibly for TB;
$t_{det} \leftarrow$ time of first evaluation probably for TB;
$t_{tre} \leftarrow$ time of the start of first regular TB treatment;
/* Group state transition history */;
$history_{eva} \leftarrow$ history between t_{eva} and t_{det};
$history_{det} \leftarrow$ history between t_{det} and t_{tre};
$history_{tre} \leftarrow$ history after t_{tre};
/* Start to construct pathway */;
$pathway \leftarrow \emptyset$ /* Initialise patient pathway */;
$state \leftarrow$ initial state of $episode$;
put $state$ into $pathway$;
$ie \leftarrow False$ /* indicating had interrupted evaluation or not */;
/* Read state series in **Evaluating Stage** */;
foreach $dimensions \in history_{eva}$ **do**
    $state \leftarrow$ find an state in **Evaluating Stage** matched $dimensions$ and ie;
    if $state$ *is* ***Interrupted Evaluation*** **then** $ie \leftarrow True$;
    put $state$ into $pathway$;
end
/* Read state series in **TB Detecting Stage** */;
foreach $dimensions \in history_{det}$ **do**
    $state \leftarrow$ find an state in **TB Detecting Stage** matched $dimensions$ and ie;
    if $state$ *is* ***Interrupted Evaluation*** **then** $ie \leftarrow True$;
    put $state$ into $pathway$;
end
/* Read state series in **Treating Stage** */;
$state \leftarrow$ find the initial treatment level in $history_{tre}[0]$;
put $state$ into $pathway$;
foreach $dimensions \in history_{tre}[0:]$ **do**
    $state \leftarrow$ find the treatment level in $dimensions$;
    if *treatment level increased* **then**
        put **Treatment Change** into $pathway$;
    end
    put $state$ into $pathway$;
end
/* Finalise patient patient formulation*/;
$state \leftarrow$ find the treatment outcome ;
put $state$ with the time of treatment end into $pathway$;
return $pathway$;

Appendix J

List of analysis codes and data on GitHub

Age-profile of incident TB cases (Chapter 5)

- **AgeingTB:**

  Description: Source codes/ original data/ results for applying Lee-Carter models in TB incidence forecasting

  Language: R >= 3.5

  URL: https://patientpathwayanalysis.github.io/

Individual patient pathway analysis (Chapter 6)

- **Online demonstration of the methods:**

  URL: https://patientpathwayanalysis.github.io/

- **Example data and output files:**

  URL: https://github.com/PatientPathwayAnalysis/IPPA-data

- **Pathway extraction:**

  Description: Extracting patient pathways from individual data

  Language: Python >= 3.6

  URL: https://github.com/PatientPathwayAnalysis/IPPA-py

- **Result visualisation (R):**

  Language: R >= 3.5

URL: https://github.com/PatientPathwayAnalysis/IPPA-vis

- **Result visualisation (d3):**

  Language: D3.js v3

  URL: https://github.com/PatientPathwayAnalysis/IPPA-d3

Simulation modelling (Chapter 8 & Chapter 9)

- **PyEpiDAG:**

  Description: Parameter modelling and statistical model fitting based on Direct Acyclic Graphs

  Language: Python >= 3.6

  URL: https://github.com/TimeWz667/PyEpiDAG

- **PyComplexism:**

  Description: An dynamic modelling toolkit for agent-based models, equation-based models, and hybrid models. The package is essentially designed for epidemiologists and health economists.

  Language: Python >= 3.6

  URL: https://github.com/TimeWz667/PyComplexism