Mathematical Models Of The Impact Of Host And Environmental Risk Factors On the Incidence Of Tuberculosis (TB) Within A National Cohort.

A Thesis Submitted To The University Of Dublin, Trinity College

for the degree of Doctor Of Philosophy

Aidan Hanway

April 30, 2018

Declaration

I hereby declare that the work described within this thesis is entirely my own except where otherwise stated. This thesis has not been submitted as an exercise for a degree at this of any other university.

I agree that Trinity College Library may lend or copy this thesis upon request.

Signed,

_____

Aidan Hanway

28^{th} September 2016

Acknowledgements

I would like to take this opportunity to thank my supervisors: Professor Catherine Comiskey and Assistant Professor Katy Tobin for their on-going support and guidance throughout the thesis and publications. I'd also like to thank Dr. Ronan O' Toole for lending his guidance with publications.

I'd like to give a special thanks to the School of Nursing and Midwifery for funding this project and the Health Protection Surveillance Centre for supplying national surveillance data. I'd like to thank my continuation examiner Professor Cathal Walsh for lending his guidance with statistical inference methods.

I'd like to thank my family and friends for providing support in a number of ways over the last three years. I'd like to thank my mother for correcting my grammar in my publications, my father for his guidance on effective communication, and my friend Dan Foley for helping introduce me to Bayesian statistics.

I'd like to dedicate this thesis to everyone above, without whom completion would not be possible.

MATHEMATICAL MODELS OF THE IMPACT OF HOST AND ENVIRONMENTAL RISK FACTORS ON THE INCIDENCE OF TUBERCULOSIS (TB) WITHIN A NATIONAL COHORT.

Author: Aidan Hanway

Tuberculosis (TB) is an infectious disease that can prove fatal if untreated. Despite a re-emergence of TB in Ireland, research has failed to provide insight to the causes of recent increases. The study acquires national surveillance data and systematically identifies a number of significant trends related to TB. From these findings, epidemiological models are constructed and simulated and put through various scenarios. The primary aim of the study is to develop, simulate, and forecast deterministic epidemic models for the spread of TB with application to an Irish setting.

The study utilises anonymised cross-sectional surveillance data acquired from the Health Protection Surveillance Centre (HPSC). Ethical approval was granted by the Adelaide and Meath Hospital, and ethical approval recognised by Trinity College. Two SEIR (Susceptible Exposed Infection Recovered) models consisting of systems of ordinary differential equations (ODEs) were developed and simulated. Approximate Bayesian Computation and Metropolis-Hastings inference algorithms were implemented to estimate the basic reproductive number, R0, and other model parameters in preparation for simulation and forecasting.

Statistically significant differences were calculated between native and foreign-born TB notifications, which is in line with previously published literature. Significant seasonality was discovered in Irish TB notifications, which has not been previously shown in published research. Migrant and seasonal SEIR models were presented for analysis. The models forecast a modest decline in notifications nationally up until 2023. Key parameters were identified in each model to help strategies that involve population management. A scenario analysis conducting numerical simulations calculated marginal increases in notifications (from one to three cases annually) when a change in vaccination procedure from universal vaccination to selective vaccination is considered. Numerical simulations of the seasonal epidemic model suggest that it would be more cost effective to implement an infection control strategy such as vaccination during the period from January to June, rather

than all year round.

Further research is required to investigate the causes and effects of seasonality in TB notifications and whether foreign-born and native-born populations interact with each other in an Irish setting. The epidemiological parameters estimated in this thesis form a basis for future surveillance and modelling to take place in Ireland and other settings. The top contributing countries of the foreign-born population should be surveyed to ensure these trends continue, as variance in notifications for this group is larger than that of the native-born population. Further research is required to model vulnerable populations in Ireland such as the homeless, refugee, and unemployed populations.

Contents

List of Figures

List of Tables

Chapter 1

Thesis Introduction, Objectives, and Outline

1.1 Introduction

Tuberculosis (TB) saw an increase in the count of notifications in Ireland in recent years [1] which gave rise to this study. The purpose of this study is to understand the increase in TB and to understand the disease itself through the use of epidemiological models and national data. As there exists a vast number of epidemic models in the literature, this study adopts a systematic and impartial approach to its derivation of a viable model. It achieves this by utilisation of systematic methodologies used in it's literature search and by way of analysis of national data.

The study conducted an analysis using data acquired from the Health Protection Surveillance Centre (HPSC), a sister organisation of the Health Service Executive (HSE) in Ireland. The details of acquirement are outlined in chapter 4 (§4). These data are used throughout the thesis and were crucial in the establishment of a viable model. While annual surveillance reports are generated by the HPSC [2], little work has been completed evaluating TB over a prolonged time period which this study now attempts. The aims and objectives of the study now follow along with a thesis outline, which describes the contents

of each chapter briefly.

1.2 Aims and Objectives

1.2.1 Aim

To develop, simulate, analyse, and forecast one or more deterministic epidemic models for the spread of tuberculosis and to apply these models within an Irish setting.

1.2.2 Objectives

The objectives of this study include:

- Describe and analyse existing cross-sectional TB data from a national data source.
- To review and refine one or more deterministic models of spread that accurately describes underlying TB dynamics and incorporates attributes of the aforementioned analysis.
- For each model derive R_0, the basic reproductive number, for calculation and for sensitivity analysis.
- Given data and statistical inference methods, estimate epidemiological parameters and initial conditions for each model.
- For each model, simulate and extrapolate the underlying dynamics and numerically calculate the basic reproductive number.
- Perform a sensitivity analysis on the parameters and provide a scenario analysis for each model.

The following thesis outline describes the content each chapter.

1.3 Thesis Outline

The first chapter (§1) gives an introduction, sets out the study aims and objectives, and provides an outline of the thesis. In §2, the study reviews research related to TB. It gives an historical context of the disease, highlighting its origin and the burden TB has had on humanity throughout time. The symptomatology of TB is then described to lend insight to the underlying disease dynamics and various disease management approaches are described. Lastly, the epidemiology of TB is then summarized on a global, European, and Irish scale.

In §3, the study reviews mathematical epidemiology. The underlying systematic process of modelling is described and the history of mathematical epidemiology is detailed mentioning various key discoveries as the section progresses. The chapter then describes a specific modelling type, referred to as compartmental modelling, and simple models are reviewed. Lastly, a systematic search strategy is implemented to seek out a viable model for analysis and simulation. In §4 an exploratory analysis is conducted evaluating a national TB data set. Descriptive analysis is conducted on a number of variables and two key characteristics are found from the data: seasonality in TB notifications and an increase in foreign-born TB.

In §5, the first of two models is analysed and simulated. The model simulates TB for the population as a whole and proposes that certain parameters have seasonal increases and decreases. The model is then altered made to it and the parameters are estimated in order to fit Irish data. Simulation and extrapolation is conducted along with the calculation of the basic reproductive number for the model. In §6, the thesis considers a migrant model that stratifies the population into local and migrant subgroups. Two models are considered in this chapter: one that assumes no interaction occurring between each sub-group, and another that assumes an interaction is occurring. Similar to §5, analysis, simulation, extrapolation, and calculation of the basic reproductive number is conducted. In §7 a sensitivity and scenario analysis is performed. The sensitivity of each parameter for each model is calculated with respect to various objective values. A scenario analysis is com-

pleted for the local population in the migrant model established in §6 and an intervention is staged and further analysis conducted on the seasonal TB model established in §5. The thesis then discusses the results in §8. A comparison of results to findings in the literature is offered and limitations of the modelling methods are discussed. The thesis concludes in §9, highlighting key results and proposing further work to be conducted in the future.

Chapter 2

Review of Tuberculosis

2.1 Introduction

Tuberculosis (TB) is an infectious disease caused by the bacterium *Mycobacterium Tuberculosis* (M.Tuberculosis). The pathogen can occur in any organ of the body but primarily affects the lung. The impact of TB has been profound and it is estimated that it has has responsible for killing more people than any other pathogen [3]. Different names for the disease throughout history include: *phthisis*, *consumption*, *the grave yard cough*, and *the white death*. Humans have been exposed to the disease since ancient history.

This chapter provides background information with respect to this history. The chapter details the symptoms and risk factors of the disease to and provides context on how the disease spreads. Various TB management and prevention methods are discussed and, lastly, the global, European, and Irish epidemiology of the disease are detailed.

2.2 A Brief History Of Tuberculosis

2.2.1 Origin

Tuberculosis has infected humans since the expansion of human populations from Africa, approximately 35,000 years ago. However, the bacterium is believed to have originated over 150 million years ago and is, therefore, thought to be one of the longest living pathogens known to mankind [4]. Its origins are not well defined in literature. A proportion of literature states that it originated in the Neolithic period when man first began living in close proximity to cattle [5], however this theory is under dispute as comparative genome analysis has shown that it was unlikely that M. tuberculosis arose directly from a bovine (cattle) strain [6]. A consensus has yet to be reached with regard to its origin. Analysis based on the mutation rate of the bacteria indicates much of the diversity among strains currently in circulation had origins between 250 and 1000 years ago [7].

2.2.2 Progression Throughout History

While relatively little is known about its prevalence before the 19^{th} century, the number of infections is thought to have peaked between the end of the 18^{th} century and the end of the 19^{th} century. Figure 2.1 demonstrates the impact TB has had on humanity over time relative to other diseases.

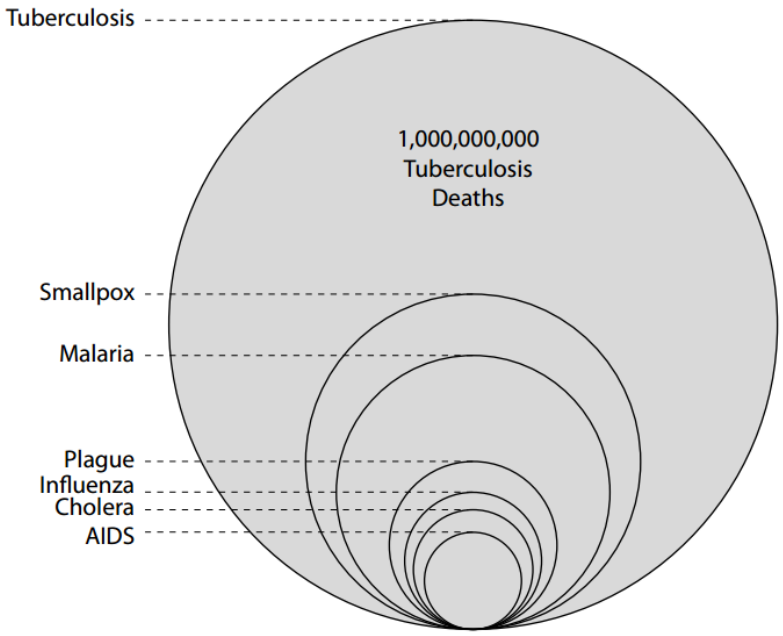

Figure 2.1: A Comparison Of Worldwide Deaths Caused By Tuberculosis With Other Infectious Diseases In The Past 200 Years. Source:[3]

TB was not identified as a unique disease until the 1820s. It was referred to as "tuberculosis" in 1839, by J. L. Schönlein [8]. M.Tuberculosis was later designated within an appropriate class of bacteria in 1882 by Robert Kock. Kock published his findings on tuberculosis, in which he reported the cause of the disease to be the slow-growing Mycobacterium tuberculosis. He received the Nobel Prize in physiology and medicine in 1905 for his research [9]. Koch announced a discovery of an extract called glycerine as a "remedy" for tuberculosis in 1890, calling it "tuberculin". While it was not effective, it was later successfully adapted as a screening test for the presence of pre-symptomatic tuberculosis [10].

Albert Calmette and Camille Guérin achieved the first success in immunization against tu-

berculosis in 1906, using a bovine-strain tuberculosis. It was called bacille Calmette–Guérin (BCG). The BCG vaccine was first used on humans in 1921 in France [11], but received widespread acceptance in the US, Great Britain, and Germany only after World War II [12].

An effective treatment for TB was developed in the UK by John Crofton who introduced Isoniazid, a newly discovered drug, as a treatment for patients. He was later knighted for his discovery. At the time it was possible to achieve a 100% cure rate with this treatment [13,14].

The new-found cure caused both the prevalence and incidence to decline globally. Hopes of completely eliminating TB (cf. smallpox) from the population were dashed, however, the disease began to infect individuals with AIDS for which treatment could not be effectively administered. The global AIDS pandemic in the 1980's combined with the rise of drug-resistant strains caused a resurgence globally. Because of the emergence of new strains, surgery has been re-introduced as an option within the generally accepted standard of care in treating TB infections[15].

2.3 Symptoms and Risk Factors

While the human body can harbour the bacteria that cause TB, the immune system can prevent the illness from developing. This is achieved by macrophages, a type of white blood cell, which engulfs the bacteria and may destroy it. Doctors make a distinction between various forms of TB, they are:

- ***Latent TB*** is defined as the condition whereby an individual may have the bacterium present but is asymptomatic (displaying no symptoms). An individual has Latent TB if either of the following scenarios occur:

    1. The bacteria gets absorbed and destroyed by macrophages.
    2. The bacteria get absorbed by macrophages, but are not destroyed.

The above two scenarios are difficult to distinguish clinically. While latent TB is not contagious, it can progress to active TB.

- ***Active TB*** is defined as the condition where the individual is displaying symptoms and can spread the disease to others. It can occur in the first few weeks after infection with the TB bacteria, or it may occur many years later.

If a tuberculosis infection does transition into an active case, it most commonly infects the lungs (in about 90% of cases) [16]. The infection is said to be a *Pulmonary Infection* when this occurs. Symptoms in this instance include: difficulty breathing, chest pain, coughing (occasionally with blood), night sweats, fatigue, and weight loss. Figure 2.2 illustrates the progression of TB inside the lungs.

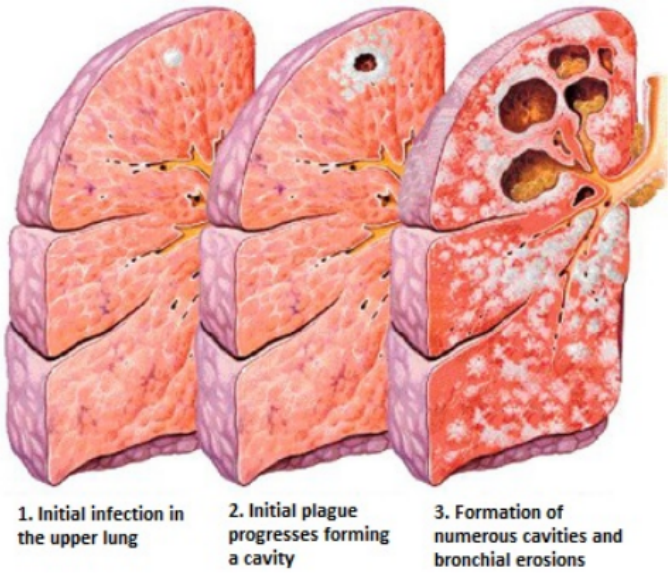

Figure 2.2: Stages Of Mycobacterium Tuberculosis Progressing Throughout The Lungs

In 15–20% of active cases, the infection occurs outside the lungs [17]. It can spread to almost any part of the human body including: other organs, bone, and the lymphatic

system. When this occurs the infection is defined as an *Extrapulmonary Infection*. It is possible for an individual to have both Extrapulmonary TB and Pulmonary TB simultaneously. Symptoms of Extrapulmonary TB depend on where the infection is occurring within the body, however, some common symptoms include: symptoms of pulmonary TB, blood in urine, headaches, back pain, and a sore throat.

Clinical presentations of Extrapulmonary TB include: Tuberculous meningitis, Skeletal TB, Gastrointestinal TB, and Genitourinary TB. Symptoms of each of these presentations follow:

- **Tuberculous meningitis:** Patients with tuberculous meningitis may present with a headache that has been either intermittent or persistent for a duration greater than 2 weeks. Possible progression includes the individual transitioning into a to coma over a period of days to weeks. Fever can be absent for this form of TB.

- **Skeletal TB:** This form of TB usually effects the spine, which is referred to as Pott disease. Back pain or stiffness can follow with the possibility of paralysis occurring in the lower-extremities (commonly only a single joint).

- **Gastrointestinal TB:** Occurs when any site along the gastrointestinal tract becomes infected. Symptoms of gastrointestinal TB are referable to the infected site and include the following: Ulcers of the mouth/anus, difficulty swallowing, abdominal pain, malabsorption (when infected in the small intestine), pain, diarrhoea, or hematochezia (when infected in the colon).

- **Genitourinary TB:** Symptoms of genitourinary TB may include frequent urination, flank pain, and dysuria. In males, genital TB may manifest as a painful scrotal mass, prostatitis, orchitis, or epididymitis. In females, genital TB may mimic pelvic inflammatory disease.

2.3.1 Risk Factors

The risk of developing tuberculosis is dependent on both the risk of being infected and the risk of latent infection becoming an active disease. The former will depend on the inci-

dence of tuberculosis in the community where the individual lives or works. The latter will depend on many factors including the genetic make up and environment of the individual.

As infection can be contingent on the strength of the individual's immune system, a number of other diseases can increase risk factors. The most common risk factors include:

- **HIV/AIDS:** The largest risk factor for developing active TB is concurrent HIV infection. The probability of latent TB progressing to active TB increases approximately one hundred-fold when the individual is HIV positive [18].
- **Immunosuppressive Treatment:** An individual being administered treatment that affects the immune response will have an increased risk of infection. Patients exposed to tuberculosis and being administered immunosuppressive treatment have very specific requirements for preventive therapy. An individual undertaking immunosuppressive treatment has approximately 12 times the likelihood of acquiring active-TB.
- **Malnutrition:** Studies have shown that malnutrition raises the risk of TB because of an diminished immune response [19,20]. TB itself can lead to malnourishment because of a loss of appetite and changes in the individuals metabolism [21].
- **Young Age:** Children are at higher risk of contracting both latent and active TB. Studies have shown that 60–80% of children exposed to an infectious case became infected [22]. The majority of the children less than two years of age get infected from a household source. For children greater than two years of age, the majority of them become infected from an external community source. Children with primary infection before 2 years or after 10 years of age are at increased risk for disease development [159]. The highest risk for TB-related mortality following primary infection tends to occur during infancy. The risk declines to 1% for children aged between 1 and 4 years. The risk then rises to more than 2% from 15 to 25 years of age [156].

- **Diabetes:** Diabetes has been shown to increase the risk of active TB disease developing. A systematic review conducting a meta-analysis of 13 studies found that diabetes was associated with an increased risk of TB (RR = 3.11, 95% CI 2.27–4.26). It is estimated that currently 70% of people with diabetes live in low- and middle-income countries [24].

- **Alcohol/Tobacco Smoke:** A systematic review concluded that the risk of active tuberculosis is higher (RR = 2.94, 95% CI = 1.89–4.59) among people who drink more than 40 grams of alcohol each day [25]. Reasons for increased risk include suppression of the immune system caused by the intake of alcohol. The association between smoking and TB has been studied by Bates and colleagues who conducted a systematic review [26]. Relative risk estimates ranged from 2.33 (95% CI, 1.97-2.75) to 2.66 (95% CI, 2.15-3.28) among individuals who are tobacco users.

Environmental and demographical factors are also important aspects that contribute to the progression of infection.

- **Socio-economic Status:** People with low socio-economic status are exposed to several risk factors discussed above (including malnutrition, indoor air pollution, alcohol intake, etc.) which increases their risk for TB. People with a lower socio-economic status have a higher likelihood of being exposed to crowded, less ventilated locations and often have limited access to healthcare facilities.

  In the recent work of Ortblad and colleagues [160], it was shown for low to middle income countries improved living conditions, health system access, and education have a negative relationship with tuberculosis case notification rates. It was also shown increased malnutrition, health expenditure, poverty, and inequality have a significant positive relationship with tuberculosis case notification rates.

  The STOP TB campaign [161], funded by the WHO state low income countries accounted for 65% of TB cases and 71% of TB related deaths in 2002. Approximately 42% of the worlds population lives in these countries implying the burden

of TB falls heavily on individuals living in poverty. As stated, low socio-economic status individuals are exposed to TB risk factors. This relationship is apparently a two way relationship. Poverty brings about TB notifications and TB itself produces poverty. It is estimated to have economic cost of $12 billion from the incomes of the worlds poorest communities annually.

- **Birthplace:** Studies conducted on high-income countries have shown the burden of TB falls disproportionately more on foreign-born communities [27]. The study reviews figures that show the proportion of individuals migrating with active TB is relatively small, however, the proportion of individuals with a latent TB infection can range from 5-72%. This in combination of the fact that reactivation occurs at a higher rate than the host population [202], causes an inflated notification rate for the foreign-born population relative to the native-born population. The increased latent infection rate is due to a number of factors, however, the duration of time the individual spends in their birth country and the TB burden within that country are believed to be two primary factors [28, 29].

- **Occupation:** Healthcare workers are at increased risk of exposure to TB. A review by Seidlerand and colleagues showed that, among healthcare workers in high-income countries, the overall incidence of TB in the general population was less than 10 per 100,000. Within native-born (individuals with birthplace of country they currently reside) healthcare workers incidence was observed to be 25 per 100,000 individuals [30].

2.4 Tuberculosis Management

2.4.1 Treatment

The underlying treatment pathway can be seen in figure 2.3.

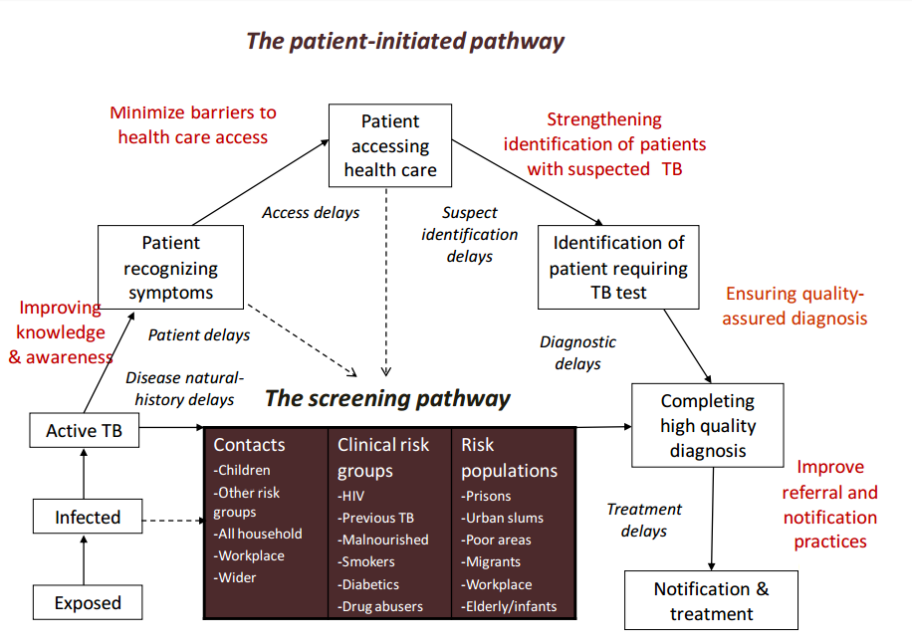

Figure 2.3: Pathways to TB diagnosis and treatment, and barriers or entry points for interventions Source:[31].

An accurate diagnosis of TB can often be difficult, particularly diagnosing active TB [32]. A diagnosis is often made when symptoms persist for more than two weeks. A chest X-ray and sputum cultures are typically part of the evaluation process. With respect to analysis of sputum cultures, treatment is often administered before a positive result is observed. This is due to the slow growth rate of the culture [33].

For latent TB the Mantoux skin test is often administered to screen individuals with TB risk factors. A standard dose of 5 tuberculin units is injected into the individual and the results are measured two to three days after administration [34]. A diagnosis is given by measuring the diameter of the hardened area in millimetres that appears (often appearing as a rounded hard bubble on the skin)[35].

An individual's risk factors determine the tolerable threshold for the diameter of the

hardened area. If the individual is HIV-positive, has come into recent contact with an infectious case, or has received immunosuppressive medication, a positive diagnosis is given when the hardened area has a diameter greater than 5*mm*. If the individual has arrived from a high-prevalence country, has a history of drug misuse, is a child, or works within a healthcare setting, a diameter of 10*mm* or more constitutes a positive diagnosis. Lastly, all remaining individuals with a diameter measurement of 15*mm* or more result in a positive diagnosis.

2.4.2 Control, Prevention, and Efficiency

Control guidelines set out by the Centers for Disease Control and Prevention [162] suggest a control program that incorporates the following: quick detection, airborne provisions, and rapid treatment.

The guidelines go on to detail that the best method of implementing the aforementional control mesures is by way of administrative (developing an infection-control plan, effective work practices, screening, etc.), enviornmental (exhaust ventilation, controlling airflow), and respiratory-protection (implementing a respiratory-protection program, education) methods.

With respect to population prevention and control, efforts primarily rely on vaccination of infants and appropriate treatment of active cases [36]. The only available vaccine today, which was discovered in 1921, is the bacilli Calmette-Guerin (BCG) vaccine. The majority of vaccinations are administered immediately after birth [37]. Booster vaccinations are be administered after birth in specific countries. The effectiveness of the BCG varies depending on geography, and appears to decrease the closer the underlying population is to the equator [38,39]. In a systematic review of randomized control trials, the vaccine was shown to have a relative risk ratio of 0.31 for inidivuals living in locations greater than 40 degrees latitude. This relative risk translates to a 69% efficacy rate. The duration the BCG offers protection is not clearly known. An absence of longitudinal data within studies contributes to the uncertainty. A study by Abubakar and colleagues [41] conducting a

systematic review concluded protection can last for over 10 years, however other studies have shown the protection the BCG offers can last a life-time [42].

With respect to Irish vaccination strategy, the BCG vaccine is given to protect babies against TB but can also be given to older children and adults who are considered to be at risk of developing TB. The vaccine is not administered when: an allergic reaction to a previous BCG vaccine has occurred, previous BCG vaccine has been administered, diagnoses with TB previously, pregnancy, positive tuberculin test, family history of problems with the immune system, HIV Positive individuals. Further information on Irish strategy can be found on through the Irish National Immunisation Organisation [163]

2.4.3 Self-Management

Infected individuals are usually prescribed medications to reduce the risk of transmission to others. Ensuring adherence to medication is important, as individuals who are not subject to direct monitoring may increase the risk of transmission and there is a high degree of reliance on self management and administration when confinement is not possible. Additional precautions taken include: confinement to home in order to reduce contact with other individuals; improved ventilation - TB spreads within closed confined spaces and it is recommended to have either a fan or open windows in the room in which the individual spends most of their time; wearing a mask to cover the mouth as TB primarily transmits through coughing. It is also recommended to safely contain or destroy the material used to cover the mouth.

2.5 Epidemiology of TB

2.5.1 Global Epidemiology

Tuberculosis, has seen a decline in incidence, prevalance, and mortality over the last decade. Despite declining rates, disease elimination is still not in sight. Figure 2.4 illustrates global trends between 1990 and 2015 of active cases.

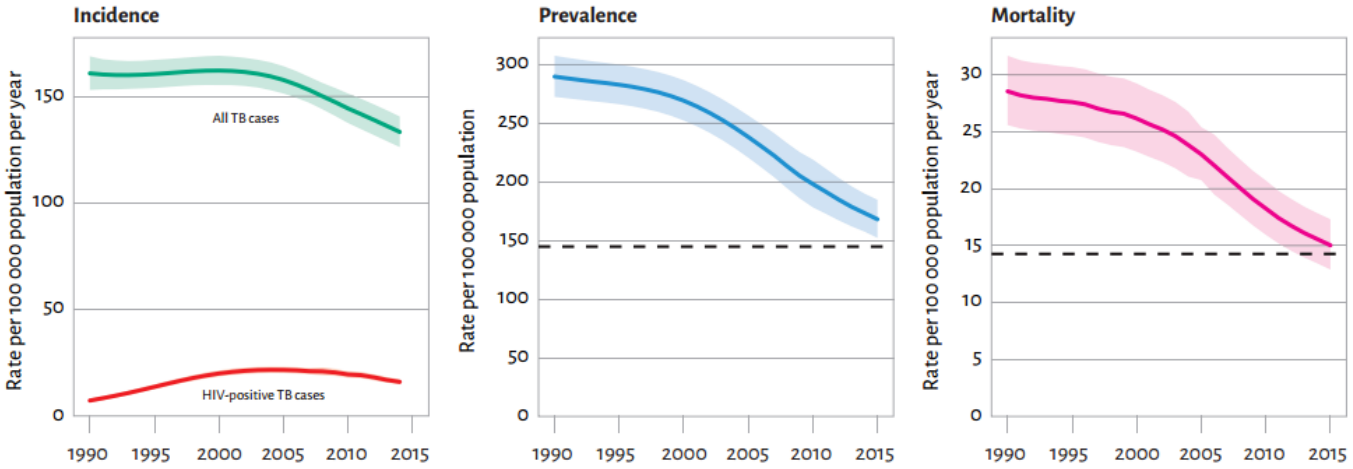

Figure 2.4: Global Trends In Estimated TB Incidence, Prevalence, And Mortality Rates 1990-2015. Dashed Line Representing STOP TB Control Strategy Target For 2015 Source: [43]

In 2010, it was estimated one-third of the global population was infected with either latent or active TB [44]. This amounts to approximately 2.3 billion individuals. The distribution of infections is not uniform and is heavily skewed towards developing countries. About 80% of the population in many Asian and African countries test positive in tuberculin tests, while only 5-10% of developed nations, such as the United States, test positive [45]. In 2014, there were approximately 1.5 million deaths attributed to TB [44]. India and China accounted for approximately 41% of total global notifications. India had an estimated 2.16 million new case notified during 2014 relative to a population of 1.29 billion, and China has the largest total count notifications with an estimated 12.7 million new cases notified during 2014, relative to a population of 1.37 billion. The largest incidence in a population was seen in South Africa, estimated at approximately 834 per 100,000 of the population in 2014. Figure 2.5 shows a heat map of TB incidence globally for 2014.

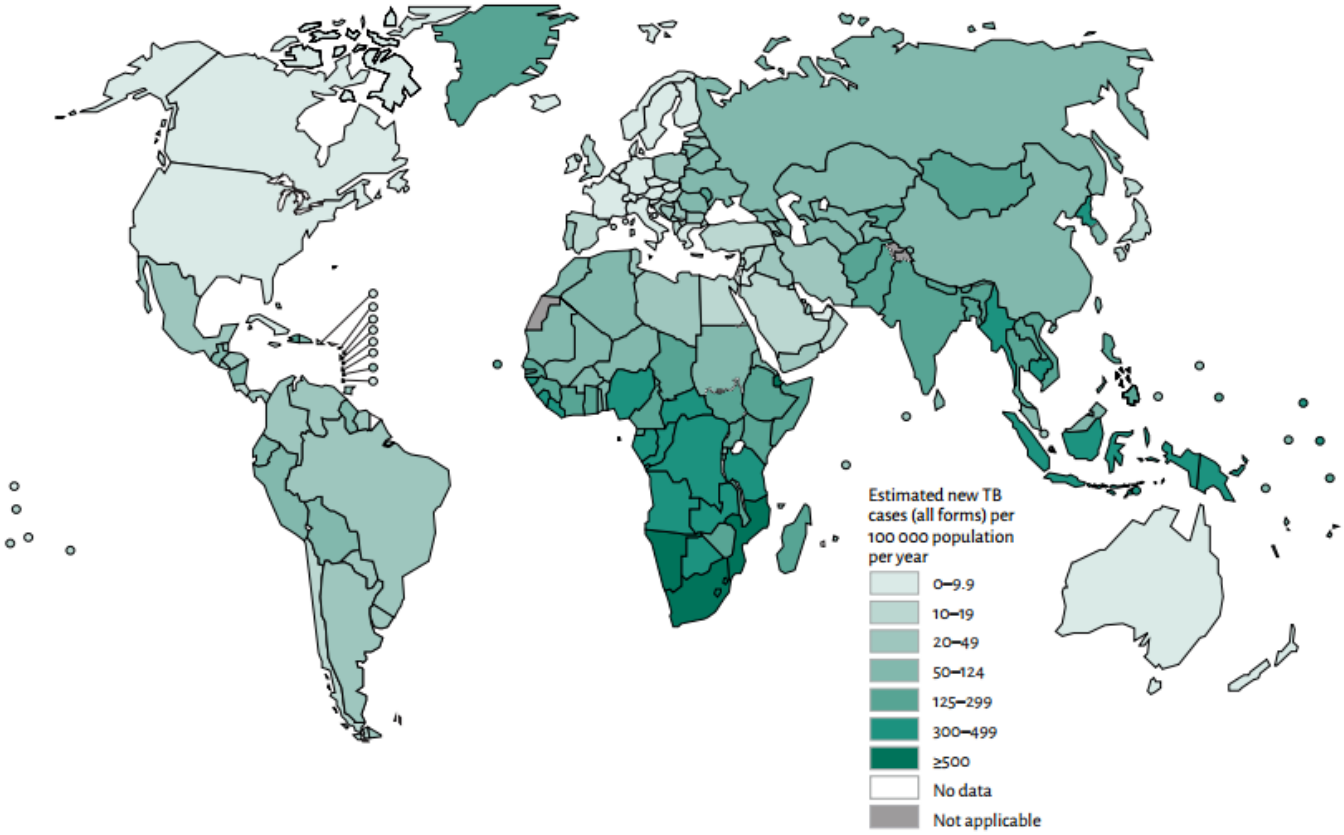

Figure 2.5: Estimated Distribution Of Global TB Incidence Rates 2014. Source: [43]

In the early 1990s, TB re-emerged in developed countries consciousness due to an increase in the number of outbreaks and has since become a significant public health agenda item. In 1993, concerned about the extent of the problem in most of the developing world, the World Health Organisation (WHO) declared TB a global emergency [48,49].

Over the last twenty years global strategies for TB control have been recommended in all countries. In May 2014, the World Health Assembly (WHA) approved a strategy aimed at ending global TB epidemics by 2035. Milestones for 2025 include a 75% reduction in mortality (compared to 2015 rates), and a 50% reduction in incidence rates. The end goal aims to reach a 95% reduction in mortality rates and a 90% reduction in incidence

rates[50].

Due to the burden of TB affecting certain countries more than others, the WHO categorised 30 countries as high burden countries (HBCs). These countries account for approximately 85–89% of the global burden. The categorisation was made to provide a focus and a galvanization for global action on TB.

2.5.2 European Epidemiology

The overall European TB case notifications reveal differences in the rates of TB between countries in Western Europe and those in Eastern Europe [51]. Figures for 2004 show the UK had notification rates less than 13 per 100,000 population. This is contrast with Romania and the Russian Federation whose rates were greater than 100 per 100,000 for their population. TB notifications in the countries of Central Europe fall somewhere in between these two extremes and while many of these countries have falling rates of TB, increases have occurred in countries such as Bulgaria and Herzegovina. TB notification rates in the European Union have been declining at a mean annual rate of 4.4% since 2006. In 2010, there were approximately 73,996 TB cases reported by the 27 EU Member States and the three additional countries (Iceland, Liechtenstein and Norway) [52]. This resulted in average notification rates below 100 per 100,000 population for all EU Member States for the first time in 2010.

Several publications have highlighted the higher notification and incidence rates evident among certain high-risk groups for TB which is overrepresented in most big cities. High risk groups include migrants from high-incidence countries, homeless people and drug and alcohol users [53,95].

Additional Remark: A publication can be found in Appendix D. The study conducts a cross-sectional analysis of nine European countries. The study examines the density of each country's HBC population and foreign-born incidence rates.

2.5.3 Irish Epidemiology

While a more detailed study of Irish epidemiological data is present in §4, a brief outline is given here. In 2015, Ireland experienced 318 TB notifications with a respective crude incidence rate of 6.9 per 100,000 individuals, this is a considerable decline compared to 1991 figures where notifications were 640 with an incidence of 18.2. The incidence rate did not decline linearly between those time periods. Incidence rates saw a modest increase from 2001 up until 2008. The cause of this increase has been detailed in literature as predominantly being due a rise in foreign-born notifications [1]. The current BCG vaccination strategy implemented within Ireland is one of universal vaccination from birth with no booster administered. All individuals born within Ireland receive this vaccination. Changes are currently being proposed to this policy and further discussion and analysis is presented in §7. Further analysis and discussion is completed on Irish epidemiological data within §4.

2.6 Conclusion

The chapter detailed the history of the origins, diagnosis, treatment, and epidemiology of TB and corresponding healthcare research. It examined the symptoms and risk factors currently associated with TB and looked at various methods used to manage and prevent the disease. Lastly, the global, European, and Irish epidemiology were described to give background for later chapters.

The following chapter will review the use of mathematical modelling within healthcare, and show how it has become an important tool within epidemiology. The application of such models are also introduced with respect to TB.

Chapter 3

Review of Epidemic Models

3.1 Introduction

Epidemiology is the study and analysis of the patterns, causes, and effects of health and disease conditions in defined populations [54]. The mathematical modelling of different diseases continues to be an area of active research. The aim of mathematical modelling within epidemiology is to understand and, if possible, to control the spread of the disease. To do this, epidemic modelling tries to relate disease dynamics at the population level to basic properties of the host and pathogen populations and of the infection process. For any given infectious disease, it's increase or sustainment over time can either be classified as epidemic, pandemic, or endemic.

Definition 3.1. *An **epidemic** is an outbreak of an infectious disease affecting a disproportionately large number of individuals in a population, community, or region within a short period of time. A disease is **pandemic** if the epidemic spreads to a large region (or worldwide). An infectious disease is **endemic** when it is maintained in a population without the need for external inputs.*

This chapter demonstrates the applicability of modelling to healthcare, then goes on to review the underlying process, methods, and gives an historical context to modelling. The chapter then finishes by reviewing two epidemic TB models. A search strategy is

implemented which is informed by an exploratory analysis which is detailed in chapter (§4). The methods and results of that search are reviewed at the end of this chapter.

3.2 The Mathematical Modelling Process, Healthcare Application, and Methods

3.2.1 The Modelling Process

Vynnycky and White [55] present a procedure illustrated in figure 3.1 that can be used to develop a model for the spread of infectious diseases.

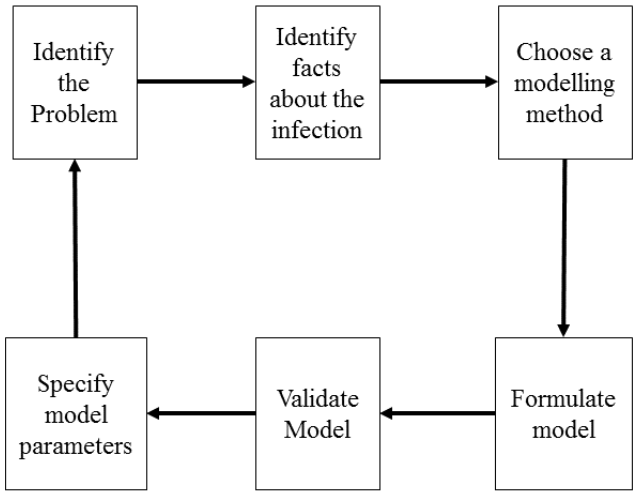

Figure 3.1: Schematic Diagram Of The Modelling Process. Source: Vynnycky and White [55]

With respect to this study, both the identification of the problem and of the facts about the infection were achieved through analysis of surveillance data. From the analysis a

mathematical modelling approach was deemed appropriate. Two key attributes of an appropriate mathematical model include:

- It should incorporates the main attributes of the phenomena it describes
- It is simple for the purpose of analysis and application.

An optimal model will find the correct balance of these two attributes. A model can incorporate many attributes but end up becoming too complex to analyse in any meaningful sense, and in contrast a model can be so simple that it does not actually represent the underlying process.

3.2.2 Application Of Modelling Within Healthcare

Mathematical models can contribute significantly to the understanding of diseases. They provide insights, improve intuitions, clarify assumptions for formal theory, allow for planning studies, estimating parameters, determining sensitivities, assessing conjectures, simulating simple and complex phenomena and providing future predictions, along with helping identify priorities and focus efforts [155]. Aside from communicable diseases, it should be noted that epidemiological modelling has been proven useful in modelling non-communicable diseases. The work of Briggs et al. [164] details discrete and continuous epidemic modelling as viable methodology in establishing economic evaluations of the disease. The work of Box [165] should also be mentioned as it discusses the intrinsic link between the more boarder concept of the scientific method and statistical modelling. His work highlight the importance of statistical models within a scientific framework which, in essence, is the basis from which this study is established.

Roberts and Heesterbeek [56] note that a model is a simplified representation of a complex system, usually designed to focus in on a specific question. Modelling is important in a range of areas such as:

- Preparing for an outbreak - modelling the impact of an epidemic.

- Predicting healthcare needs in the future, such as the long term health service resource requirements.
- Depicting what could happen with important public health issues if no interventions are undertaken.
- Understanding the impact of service redesign on different areas such as general practice waiting times, hospital bed occupancy.
- Estimating prevalence when detailed data are not available.
- Predicting demand on services from subgroups of the population, such as those at risk of emergency admissions or re-admissions

Roberts and Heesterbeek [56] also note that the analysis of mathematical models can lead to the discovery of concepts that play an important role epidemiology.

Some of the limitations noted include:

- The most complex models constructed are often still oversimplified.
- Our knowledge of key parameters in the underlying transmission process is often poor
- Making practical predictions, in the long term, can be especially difficult.

Having outlined some of the benefits and limitations of modelling within healthcare the following section details the mathematical modelling process.

3.2.3 Various Modelling Methods

While this study primarily focuses on deterministic modelling, some of the most widely used modelling methods detailed in literature [58] are summarized in this section together with their advantages and disadvantages. The models detailed below are given in order to provide the reader understanding of mathematical modelling, and do not necassarily

have application in healthcare but are well established in the sub-field of mathematical modelling. The models below are not necessarily mutually exclusive of each other, a modelling method can support another.

Empirical Modelling

An empirical model is one of the most elementary, but least insightful in an epidemiological sense. It typically is formed by constructing a time dependent function (referred to as a dynamic model), such as a polynomial or exponential function, with a set of unknown parameters (often coefficients). The discrepancy between the function and data is then minimized by altering the parameter set using data fitting techniques (such as maximum likelihood).
Advantages: Easy implementation, interpretation relatively simple.
Disadvantages: The fitted models parameter estimates do not apply to data outside the observable range. The interpretation of the parameter set has limited practical benefit.

Deterministic Modelling

This is the modelling method used in this study. Deterministic modelling can be defined as a modelling approach that does not consider random variation. Within epidemiology, when referring to deterministic modelling in recent times, it typically refers to a discrete or continuous system of dynamic, non-linear equations that attempt to describe basic relationships between variables of a problem. Output is solely determined through constructed relationships among states. A given set of initial conditions/parameters will always produce the same output.
Advantages: The model can be customised, offers reliable predictions on future states, capable of offering "what if" analysis. The work of Barlett [166] discusses the usefulness of the parameters of a deterministic model, as the parameters themselves can be interpreted on their own and can provide insight to the disease, in addition being used for surveillance purposes. Deterministic models often have a strong ability to predict disease dynamics as they incorporate theoretical representations of the underlying process.
Disadvantages: Empirical processes are not always completely deterministic. Interpreta-

tion or extraction of useful information from the system of equations can be analytically difficult and for small population sizes the model does poorly.

Stochastic Modelling

Similar to the deterministic model in its ability to be customised, this model also considers random variation. Probability statements can be formed such as the probability of an epidemic given a parameter set with predetermined distributions.
Advantages: Customisable, applicable to small populations, offers variation and probability statements.
Disadvantages: Typically not used for large populations as random variation is not as important in predicting future states, have the capability of becoming too complex to conduct a qualitative analyse on and can become computationally expensive (i.e. time/computer memory).

Simulation Modelling

Typically a model is constructed through a process of rules set between variables. It considers random variation and can output statistically interpretable data. Such processes include so called agent-based modelling system, detailed in the work of Silverman et al.[167]. This is a simulation method within which a mathematical function is given memory and autonomy and is defined as an agent. Multiple agents are then simulated to represent the underlying process.
Advantages: Highly customisable, Easy to build, not subject to a large set of assumptions, practical when easy to control models do not exist.
Disadvantages: Hard to identify an error within the model, does not guarantee an empirically optimal fit.

Statistical Modelling

A model is usually specified by mathematical equations that relate one or more random variables and possibly other non-random variables. A probability model consists of the

triplet (Ω, F, $\mathbb{P}$), where Ω is the sample space, F is a $\sigma-$algrbra of events, and $\mathbb{P}$ is a probability measure on F. A statistical model is a set S of probability models, this is, a set of probability distributions on the sample space Ω. Statistical models are non-deterministic.
Advantages: Hypothesis tests can be calculated, statistical inference can be conducted, Intervals can be calculated, descriptive analysis can be conducted.
Disadvantages: Can be difficult to extrapolate outside the observable range of data, hypothesis testing and inference subject to Type 1 and Type 2 errors.

Above I have reviewed a number of commonly used mathematical models. In this study, I have elected to proceed with a deterministic model because of their dynamic attributes, ability to be customized, and due to the complex nature of the modelling that will take place in §5 and §6. A simple empirical/statistical model is avoided in this instance as they both lack the ability to provide insight to the disease and do not consider the underlying disease dynamics. A stochastic model is avoided in this instance too as the nature of the TB modelling process is expected to be complex. Interpretation of a complex stochastic model may divert the studies ability to yield insights to healthcare stakeholders. In conjunction with the above reasons, a deterministic model will be selected for it's epidemiological interpretablity and for it's ability to consider underlying disease dynamics. The following section I will review the historical progression of mathematical models throughout time.

3.3 The Historical Progression of Modelling Within Epidemiology

3.3.1 General Epidemiology Progression

Mathematical and statistical modelling have an extensive history within population biology, dating as far back as the 17^{th} century. John Graunt was one of the first demographers. His book "*Natural and Political Observations Made upon the Bills of Mortality*", first implemented analysis of mortality data in 1662, consequentially leading to an endorsement from the current king gaining him entry to the Royal Society [59]

As noted by Bailey [60], progression of any form of observational modelling slowed drastically after the work of Graunt, the next record of epidemic modelling was in the 18^{th} century. In 1766, the mathematician Daniel Bernoulli began examining immunity to the smallpox disease. Trained as a physician, Bernoulli created a mathematical model to defend the practice of inoculating against smallpox. He concluded that life expectancy would increase from 26 years to 29 years should universal vaccination of the population occur. Other notable work in the 19^{th} century was the work of the epidemiologist Farr [61] in 1840. Farr is regarded as one of the founders of medical statistics. He implemented relatively simply empirical models analysing demographic and mortality data. Some of his notable work included his research on the cholera epidemic that occurred in London in 1849. Other work included detailing the impact occupation had on death rates.

Within the 20^{th} century epidemiological research evolved rapidly; growing availability of mortality statistics caused a growth in modelling, which concurrently grew with laboratory experiments in microbiology. The work of Ross [62] on malaria considered a *mass action principle*, a principle developed by Hamer in 1906 [63]. The principle states the progression of an epidemic depends of the rate of contact between susceptible and infectious individuals. Ross formulated a mathematical model for the dynamics of malaria that included vital dynamics (population birth and death rates).

The work of Ross was soon followed by the work of Kermack and Mc Kendrick [64], whose paper "*A Contribution to the Mathematical Theory of Epidemics*" was considered one of most important pieces of work in epidemiology. The construction of an *SIR* model (Susceptible, Infectious, Recovered) and the identification of a threshold within the underlying system of equations were the two pivotal contributions of the paper. The threshold states the introduction of infectious cases into a community of susceptible individuals would not result in an outbreak if the proportion of susceptible individuals was below a certain threshold value. Up to this point in time the vast majority of models were deterministic in nature, and a very large proportion were empirical. Mc Kendrick went on to publish one of the first stochastic models within epidemiology [60]. The evolution of the disease was modified to take into account the probabilistic aspect of the process.

After the work of Kermack and McKendrick, there were many extensions of the models that had been constructed thus far. Bailey [60] describes many of the extensions that had been made up to the time of its publication in 1957. Further updates during the next twenty years include the work of Hethcote [65], where models were modified to include temporary immunity, carriers, migration, and transmission by vectors. The work of Waltman [66] contributed further refinements to models such as inclusion of unique scenarios of transmission and a two population threshold model. The second edition of Bailey's book was published in 1975. Some of the refinements within this edition were made to give more realistic descriptions of micro-parasitic diseases by adding additional compartments. One such refinement was the incorporation of an exposed (latent) period, a time during which members of a population who have been infected but do not pass on the infection to others. Another refinement to models considered temporary immunity against reinfection or to the assumption of a sequence of removed stages which considered probabilistic aspects of the disease process.

One other recent notable and seminal work is that of Anderson & Mays' "*Infectious diseases of humans: dynamics and control*" published in 1991 [67]. This book provides a comprehensive starting point for modelling many communicable diseases such as malaria, measles, river blindness, sleeping sickness, schistosomiasis, and AIDS.

One main focus of mathematical epidemiology recently has been on the understanding and computation of the *basic reproduction number*. This value is typically denoted R_0 and sometimes referred to as the basic reproductive rate. The value can be interpreted as follows: if the basic reproduction number is less than one, the number of infectives will tend to zero and an epidemic is not imminent; if the basic reproduction number exceeds one, the infection will successfully spread and an outbreak is imminent within the population. Generally, the larger the value of R_0, the harder it is to control the epidemic. A formula for R_0 can be calculated from most epidemic models. This prediction value gives a criterion for whether a disease outbreak will develop into an epidemic or die out. One of the first appearances of this value came about through the work of Ross but had it's first modern

application within epidemiology through the work of Macdonald arising from his work on malaria in 1952 [68].

With the refinement and growing complexity of models due to the addition of more parameters and compartments, the interpretation and calculation of the basic reproduction number becomes more difficult. However, the basic reproductive number is undoubtedly the most central idea in mathematical epidemiology and is included in the vast majority of models today.

Although epidemic modelling has had a long and extensive history, moving forward there appears to be multiple challenges requiring attention. A discussion of such challenges is discussed in the work of Roberts et al.[168], within which the current state of modelling is detailed. Roberts and colleagues itemize the research required, which includes our current ability to understand the endemic equilibrium, incorporating heterogeneity processes, and modelling other complex dynamics such as: super infections, spatially explicit models, and non-communicable diseases.

3.3.2 Tuberculosis Modelling Progression

The first TB model constructed was the discrete recurrence equation developed in 1962 by Waaler [69]. The population was divided into three compartments: Susceptible, Latent, and Infectious. The inclusion of a latent class was necessary due to TB having a relatively extended latent period [70]. After estimating parameters within the model for south India, Waaler predicted that the time trend of TB was unlikely to increase (it may decrease, slowly). His model did not consider the mechanics of transmission. However, the parameters estimated from a specific area in India, provided a pivotal stepping stone for research to be conducted on the estimation of parameters in developing nations.

Sven Brogger developed a model [71] that refined Waaler's work. Brogger introduced heterogeneity (age) and altered the method used for calculating infection rates. His aim was to estimate the effects control strategies would have on infection rates. Using the

work of Brogger and Waaler, Charles ReVelle constructed the first non-linear system of ordinary differential equations directed at modelling tuberculosis [72 , 73]. In modelling the infection rate, he did not follow the typical mass action law, analytically detailed in the work of Kermack and McKendrick. It was ReVelle who first, in the context of TB dynamics, rigorously explained why the infection rate depends linearly on the prevalence using the probabilistic approach. ReVelle's main objective seemed to be associated with the evaluation and implementation of control polices and their cost, rather than establishing the trend of the infection over time. He developed an optimisation model and used it to select control strategies that could be carried out at a minimal cost.

The continuous decline of TB incidence in developed nations and the introduction of effective antibiotics suggested that elimination of active TB in developed nations was possible. This view may have been the main reason why there was almost no theoretical work on TB dynamics from the 1970s to the early 1990s. This has changed over the last decade because of the re-emergence of TB (due to new outbreaks in the U.S.A., Ireland, and in many developed nations).

Notable recent work on TB comes from Chavez and Song[74-76], both making important contributions to modelling. Such models were constructed to consider: a variable latent period, a slow and fast progression to the infectious compartment, multiple strain models, a model to include the possibility of reinfection once treated, household clustering models, age structured models, and more. The broad range of models provided in their work are often accompanied with relative qualitative analysis and simulation.

3.4 A Review of Compartmental Models

A predominant method of modelling the spread of infections is to stratify the population into compartments and define relationships between these compartments. The population will transfer from one compartment to another at some predefined rate, the rates are usually defined as the parameters within the system. Once one is able to compartmentalise

an infectious disease with a model, compartments can be extrapolated into future time periods to obtain an estimate of the expected number of infections. Several "what if" scenarios, such as what impact a mass vaccination strategy will have on the total number of infections over time can be conducted. The most common compartmental model analysed and simulated is the SIR model. Further details of the model follow.

3.4.1 The SIR model

The SIR (Susceptible, Infectious, Recovered) model originally constructed by Kermack and McKendrick [64], operates on the principle that individuals can be categorised into one of three compartments: Susceptible to the infection, Infected and therefore infectious, and Recovered and hence immune. The SIR model is fundamental to mathematical epidemiology. The model is represented in figure 3.2.

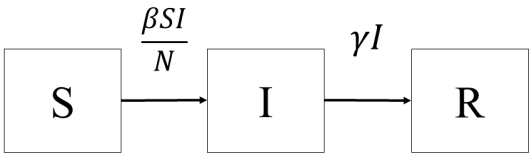

Figure 3.2: SIR compartment model

Some of the assumptions of the SIR model include:

1. The increase in the infective class is proportional to the total number of infectives and susceptibles within the population, this presumption originated through the work of Hethcote and colleagues [65]. The change is defined by the rate $\frac{\beta SI}{N}$.

2. The removal rate of infectives to the recovered class is proportional to that number of infectives and is defined by γI.

3. The incubation or latent period is short enough to be negligible, such that susceptibles instantaneously become infectious.

4. The total population, denoted N, remains constant over time. Birth rates and death rates are not considered.

5. Compartments are uniformly mixed; every individual has an equal probability of coming into contact with one another. This is referred to as an assumption of homogeneity.

The model is given by the system of ordinary differential equations (ODEs):

$$\frac{dS}{dt} = -\frac{\beta SI}{N} \tag{3.1}$$

$$\frac{dI}{dt} = \frac{\beta SI}{N} - \gamma I \tag{3.2}$$

$$\frac{dR}{dt} = \gamma I, \tag{3.3}$$

where the parameters $\beta > 0$ is the infection rate and $\gamma > 0$ the recovery rate. The constant population size is a characteristic of the system since

$$\frac{dS}{dt} + \frac{dI}{dt} + \frac{dR}{dt} = 0 \implies S(t) + I(t) + R(t) = Constant = N \tag{3.4}$$

As noted in Jones [77],the conditions for an epidemic occur when the number of infected individuals increases, i.e. $\frac{dI}{dt} > 0$,

$$\implies \frac{\beta SI}{N} - \gamma I > 0$$

$$\implies \frac{\beta S}{\gamma N} > 1$$

At the beginning of an epidemic it is stated $S(0) = S_0$,

$$R_0 = \frac{\beta S_0}{\gamma N} > 1$$

Making the assumption that the majority of the population is susceptible at the beginning of an epidemic, $S_0 = N$, the value R_0 becomes $\frac{\beta}{\gamma}$. A similar conclusion can be drawn for when infections decrease ($\frac{dI}{dt} < 0$), hence due to R_0 being a dimensionless quantity it can be viewed as the threshold ratio of the system. This number represents the expected number of secondary cases which one case would produce in a completely susceptible population. Once this value is greater than one, the disease is expected to spread throughout the population. In this instance, it is dependent on the transmission rate of the infection (β) and the rate at which individuals recover (γ). With respect to the SIR model, and theoretical deterministic models in general, the value can usually be constructed from the system of equations presented and can be represented by the models parameters.

3.4.2 The SEIR model

In many infectious diseases there is an exposed period after the transmission of infection from susceptibles to potentially infective members but before these potential infectives develop symptoms and can transmit infection [78,79]. Moreover, many parasitic diseases, such as mycobacterium tuberculosis, incubate inside a candidate for a period of time before they become infectious. Consequently, the development of SEIR models (Susceptible, Exposed, Infectious, Recovered) came about to investigate the role of an incubation period in disease transmission.

Within this model, it is assumed that a susceptible individual first goes through a latent period (and is said to become exposed or in the compartment E) after transmission of the infection. The individual stays in this compartment for a period of time before becoming actively infectious or dying of natural causes.

A number of mathematical models have been developed in the literature to gain in-

sights into the transmission dynamics of diseases with sub-population (compartments). Some of the research done on SEIR models can be found for instance in Chavez [74,76,132]; Zhang [80]; Li [81]; Yi [82]; and Shu [83].
The basic model can be illustrated in figure 3.3.

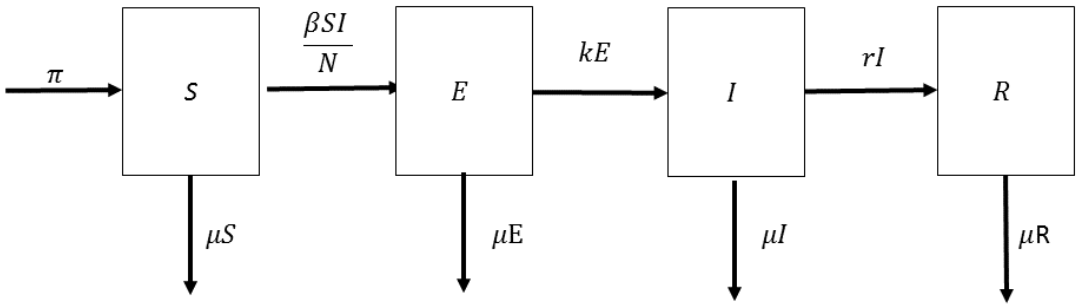

Figure 3.3: SEIR compartment model

The model has a birthrate or recruitment rate π which enter into the susceptible class. Each compartment has a death rate μ proportional to the number of individuals in each compartment, the rate is presumed constant for each compartment in the above model (e.g. if μ = 5%, this would imply 5% of the number of individuals within each compartment would die). Some models consider an increased death rate for the infectious compartment (μI in *figure 3.3* would become $\mu I + dI$, where d represents the proportion of individuals dying due the the infection). From the susceptible population a proportion of individuals enter into the Exposed compartment at a rate $\frac{\beta SI}{N}$. Once exposed, the population enters into the infectious compartment at a rate kE, and lastly from infectious the population becomes recovered at a rate rI. The model is represented by the following system of

differential equations

$$\frac{dS}{dt} = \pi - \beta SI - \mu S, \tag{3.5}$$
$$\frac{dE}{dt} = \beta SI - (k+\mu)E, \tag{3.6}$$
$$\frac{dI}{dt} = kE - (r+\mu)I, \tag{3.7}$$
$$\frac{dR}{dt} = rI - \mu R, \tag{3.8}$$
$$N(t) = S(t) + E(t) + I(t) + R(t). \tag{3.9}$$

The basic reproductive number of the above system being

$$R_0 = \frac{\beta \pi k}{\mu(k+\mu)(r+\mu)}.$$

In contrast to the SIR model the above model has different parameters, Hethcote [65] argues that when the mean latent period $\frac{1}{k} \to 0$ (or $k \to \infty$), the SEIR model becomes an SIR model with vital dynamics.

There exist clear similarities between the SIR and SEIR models. The SEIR model was adapted from the SIR model. This is a testoment to the flexibility of epidemic modelling. Various other SEIR models have been adapted and applied in other settings, the application of one such model to the SARS epidemic [169] proved vital in surveilling the disease and implementing a control strategy of isolating infectious individuals.

General Remarks On Compartmental Models

Some of the common attributes of compartmental models are listed below:

- Simulation of models is usually accomplished through numerical methods, as most ODE systems constructed for modelling purposes do not tend to have an analytic solution. In order to simulate these models numerical methods must be implemented.
- For numerical simulation, the fourth order Runge-Kutta method has been shown to be effective for simulation purposes when compared to the analytic solution of the

SIR model [84]. This numerical method has been selected for this study, however, a range of methods could be considered [170]. These methods operate under an iterative approach, namely an initial seed value is specified (Initial population conditions in this case), and from this value the next value one time step away is then calculated. From this value the next value one time step away is calculated, this continues in a similar fashion. The Runge-Kutta Fourth Order method of simulating ODE systems has been shown to have strong agreement with actual estimates, however, when applied to epidemic models with a smaller sample size, a larger error has been observed. Since this study will not consider a small population, this numerical method has been selected.

- A note to be made with varying models is that some models, when referring to a compartment, will be referring to the proportion of individuals within them, rather than the count. The underlying dynamics change by dividing terms within the model by the total population, N. A model is referring to the proportions of a population when the divisor N is not visible within the system. When the divisor N is visible it implies each compartment is referring to the count of individuals within them

- Compartmental models typically have two or more sets of equilibrium points[86], within which there is no change in the number of infectious individuals over time. As such the number of infectious individuals remains constant. The first equilibrium is often referred to as the disease free equilibrium; the number of infectives remains zero over time. The remaining equilibrium are sometimes referred to as the endemic equilibrium, which occur when the number of infectives remains at some constant greater than zero over time. These equilibrium states can be calculated by equating each of the equations within the system of ODEs to zero and then solving the resulting system with respect to each of the compartments present within the model. The equilibrium states are dependent on the parameters and initial conditions of the model.

The following section conducts a systematic search of the literature to obtain a viable TB model.

3.5 Systematic Literature Search

This section provides an additional review of research related to the mathematical modelling of TB. This search was informed by the exploratory analysis conducted on the national dataset presented in §4. Seasonality and foreign-born TB were concluded as being important factors contributing to national infection rates, hence a search strategy was implemented to ensure a body of relevant literature was available to account for those factors.

With regard selecting these factors over others, seasonality was selected as it has not received much research, particularly with respect to epidemiology and even more particularly in an Irish setting. The epidemiologists within the national body (the HSE) were unaware of this factor in Irish incidence. This thesis will be the first to examine this in an Irish setting. With regards selecting the foreign-born population, this has been cited in literature as being the primary cause for the recent rise in notifications [1].While statistical analysis has been conducted, no epidemiological modelling has taken place for this population.

3.5.1 Systematic Search Methods

The search methods used to review the literature that will accompany seminal texts, grey literature, and literature involving policy, are now described. The search methods consisted of inclusion and exclusion criteria for selecting studies, the search strategies, and the results of automated searches. The online search strategy for the literature review identified a large pool of potential records, and then obtained relevant records through the PRISMA Flow diagram presented in figures 3.4 and 3.5. This strategy is an evidence-based search strategy commonly used for systematic reviews and Meta analysis. It is widely used in health related research [87] and consists of four main stages; identification, screening, eligibility and inclusion. Automated searches were performed on six literature databases: MathSciNet, Zentralblatt Math, Web of Science, World Scientific, Science Direct, and Embase. Each database was searched for articles published up to and including December, 2014.

3.5.2 Study Selection

Two separate strings were constructed and used in each database. Each record's title, abstract, keyword phrases, and main body were searched. Search strings are search terms combined with logical AND and OR statements. Search terms used in each search are detailed in the table 3.1.

Denotation	Underlying Measure	Search Terms
A	Tuberculosis	"Tuberculosis" OR "Tuberculoses" OR "TB"
B	Mathematical Model	"Mathematical model" OR "Math model" OR "Math models" OR "Math modelling" OR"Maths modelling" OR "Maths models" OR "Maths model" OR "Mathematical models" OR "Mathematical modelling" OR "Experimental Model" OR "Experimental Models" OR "Experimental Modelling" OR "Compartmental model" OR "Compartmental models" OR"Compartmental modelling" OR "Epidemiological model" OR "Epidemiological models" OR "Epidemiological modelling" OR "Epidemic model" OR "Epidemic models" OR"Epidemic modelling" OR "Deterministic model" OR "Deterministic models" OR "Deterministic modelling" OR OR"Disease Dynamics" OR "Disease Dynamic"
C	Foreign-Born	"Foreign Born" OR "Foreign Birth" OR "Foreign Birthplace" OR "Not Native" OR "Non-native" OR "Overseas" OR "Immigrant" OR "Immigration"
D	Seasonality	"Seasonality" OR "Seasonally" OR "Seasonal" OR "Yearly Recurrence" OR "Annual Recurrence" OR "Monthly Variation"

Table 3.1: Composition of Search Terms

Using the denoted values for the search terms, each of the terms were combined with AND to form two search strings:

Search String One: (**A**) AND (**B**) AND (**C**)
Search String Two: (**A**) AND (**B**) AND (**D**)

For the pool of records acquired, inclusion criteria and exclusion criteria were applied. If the criteria were not met, this constituted a justification for exclusion.

The criteria comprised of the following for both searches:

- Analysis and/or simulation of an epidemiological mathematical model was conducted.
- The compartments were capable of modelling tuberculosis.
- The model considers: susceptible, exposed/latent, infectious, and recovered/treated/immune compartments.
- The underlying model was dynamic, continuous, and deterministic.
- The model was not constructed to exclusively consider an unrepresentative subpopulation.
- The underlying equations of the model were published

The criteria for the seasonal model included the following additional criteria point:

- The model considered seasonal fluctuation within the infectious compartment.

The criteria for the migration model included the following additional criteria point:

- The model was constructed with at least one compartment representing the foreign-born population and one representing the native-born population.

The PRISMA flow diagrams can be seen in figures 3.4 and 3.5.

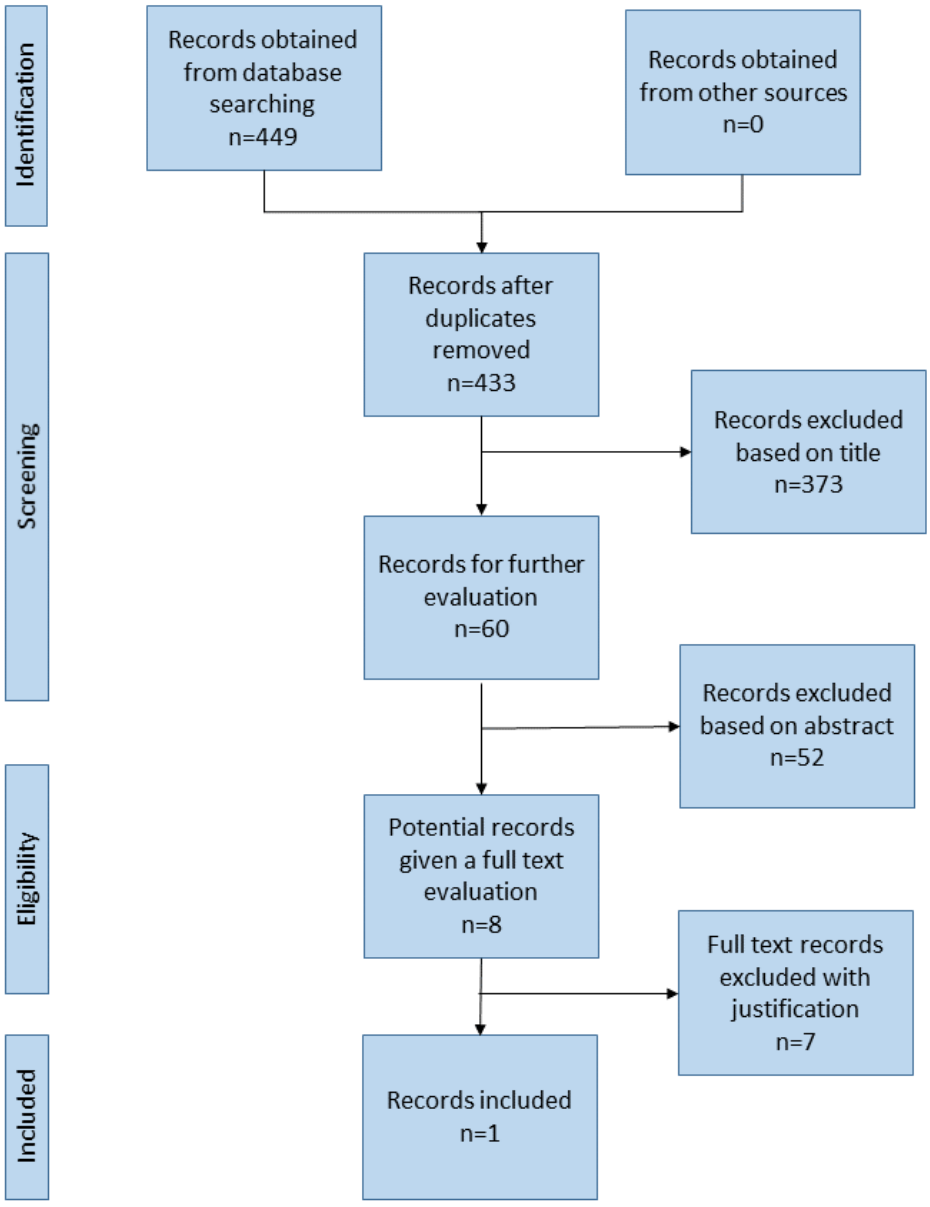

Figure 3.4: PRISMA Flow Diagram: Search Findings For A Seasonal Tuberculosis Model.

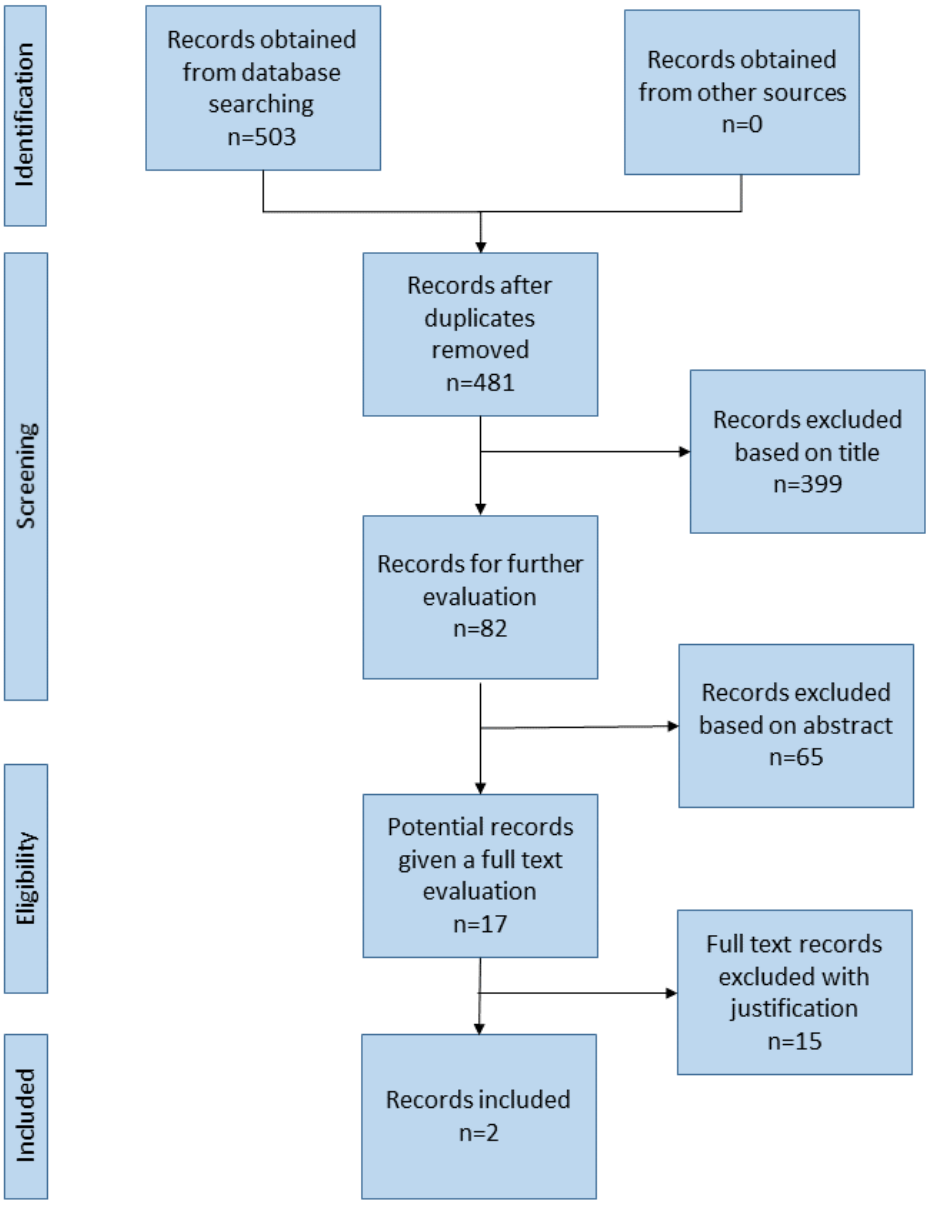

Figure 3.5: PRISMA Flow Diagram: Search Findings For A Tuberculosis Model Considering A Foreign-Born Population.

3.5.3 Systematic Search Findings

One study was found at the end of the systematic search for a TB model that analysed and simulated an SEIR model with a seasonal component, and two studies were considered at the end of the systematic search for a model that considers a foreign-born population.

Additional Exclusion

An additional exclusion occurred in one of the texts considering a seasonal variation within their model. Although passing all criteria set by this study, when analysis of the proposed model within the text began, an error was identified. Further explanation is given below.

Jia Z. et al. 2011 - *A mathematical model for evaluating tuberculosis screening strategies* [88]- After analysis of the system of differential equations, a fundamental problem was found that appeared to pass the screening process of publishing. The system of equations representing the dynamics for the migrant population were given to be

$$\begin{aligned}
\frac{dS_M}{dt} &= \pi - \beta_1 S_M I_M - \beta' S_M I_L - \mu S_M, \\
\frac{dE_M}{dt} &= \beta_1 S_M I_M + \beta' S_M I_L - (k_1 + \mu + \beta' I_M) E_M, \\
\frac{dI_M}{dt} &= k_1 E_M - (r_1 + \mu + \mu_{I_M}) I_M, \\
\frac{dR_M}{dt} &= r_1 I_M - \mu R_M, \\
N_M &= S_M + E_M + I_M + R_M
\end{aligned}$$

However, when analysis of this system began, calculating

$$\frac{dN_M}{dt} = \pi - \mu N - \mu_{I_M} I_M - \beta' I_M E_M,$$

results in the excess term $-\beta' I_M E_M$. To presume the change in the total migrant population will decrease at the rate exposed transition to infectious is not practical. As such it is

probable a $+\beta' I_M E_M$ term is missing within $\frac{dI_M}{dt}$. The paper additionally has other minor errors within it's referencing, such as its incorrect reference to the basic reproductive number. As of April 2016, the author and publication have been notified, however due to the apparent uncertainty surrounding the quality of the model and paper, this result has been excluded.

Discussion of Resulting Texts

The remaining two studies are reported in table 3.2 below and further discussed.

Study	Model Type	Types Of Analyses	Population Of Case Study
Liu et al. (2010)	Seasonal Model	Analysis of stability and equilibria, numerical simulation, case study simulation.	China
Jia et al. (2008)	Immigration Model	Analysis of stability and equilibria, numerical simulation, case study simulation.	Canada

Table 3.2: Included Studies

The two studies included in the literature review considered an SEIR model to measure the spread of TB. Both conducted a qualitative analysis on the system of ODEs, and acquired their respective basic reproductive numbers. Both studies also apply their models to population data and attempt to minimise residuals.

Liu et al. 2010 study

The study found by the systematic search with respect to seasonality was Liu and colleagues [89]. The system proposed is an alteration of a system considering slow and

fast progression rates constructed by Blower et al. and Ziv and colleagues [90-92] . The system is illustrated in figure 3.6.

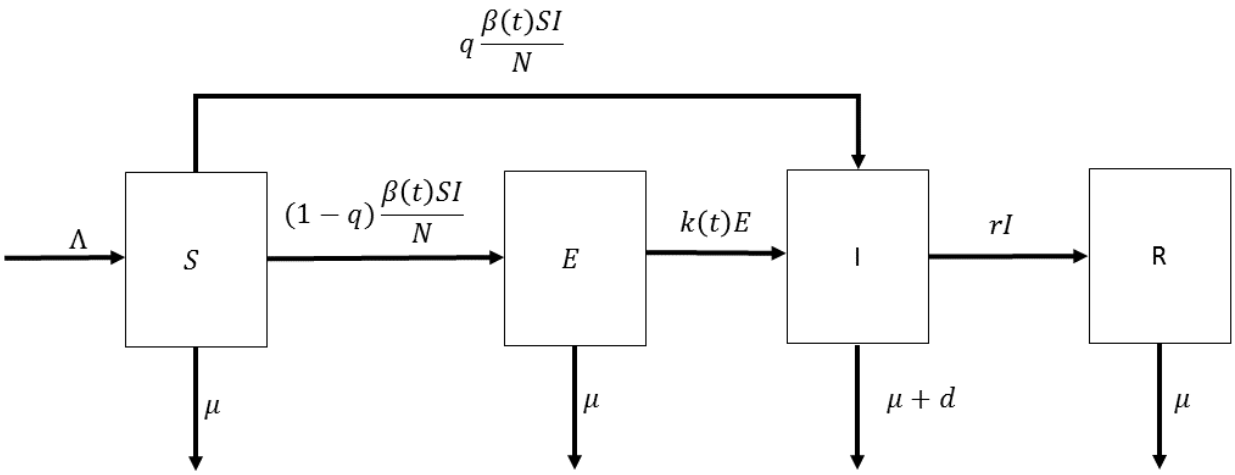

Figure 3.6: SEIR compartment model with seasonality

Slow progression is defined as entering into the exposed compartment before becoming infectious, and fast progression defined as transitioning directly from susceptible to infectious. The slow/fast progression rate is determined by proportional parameter $0 \leq q \leq 1$. The model considers seasonality through changing parameters: k - the reactivation rate, and β - the transmission rate, from constants in the original models to periodic functions of time.

The paper goes on to produce various qualitative analyses of the model and establishes an R_0. The functions $\beta(t)$ and $k(t)$ initially take on the assignment $a_0(1.1 + sin(\frac{\pi(t+1)}{6}))$, and $b_0(1.1 + sin(\frac{\pi(t+1)}{6}))$, respectively for analysis, however they take a different assignment later when applying the model to data.

The data fitting technique implemented was to initially fit the regression curve, denoted $f_{reg}(t)$, to the number of new infections over time. The number of new infectious is also

represented within the compartmental model and takes the following expression:

$$f_{mod}(t) = q\beta(t)\frac{S(t)I(t)}{N(t)} + k(t)E(t). \tag{3.10}$$

This can be seen as the rate at which the infectious compartment increases. The regression model took the form:

$$f_{reg}(t) = c_0 + \sum_{j=1}^{n} (c_j \cos(Ljt) + d_j \sin(Ljt)) + \varepsilon(t). \tag{3.11}$$

The assignment of L within the above trigonometric functions determines the period length of $f_{reg}(t)$, the study fixes $L = \frac{2\pi}{12}$ for a period length of 12. The value n is taken to be 5 and the parameters c_0, c_i and d_i are the coefficients being estimated for $i = 1,..,5$.

How $\beta(t)$ and $k(t)$ where acquired was by way of "*simulation and comparison*". It is presumed the study equates $f_{mod}(t)$ and $f_{reg}(t)$, then solves for $\beta(t)$ and $k(t)$. In doing so both $\beta(t)$ and $k(t)$ are fixed into being functions with similar properties to $f_{reg}(t)$, differing only by way of addition and multiplication of constants. As noted within the study the selection of the parameters n and L within the regression determine both the period and the domain of the transmission parameters within the model.

The author notes the model fits the data "*quite well*", and upon visual inspection that appears to be the case, however the study does not publish any error measurements, such as a *BIC* or an *AIC* estimate for the regression model, $f_{reg}(t)$. Cases appear to peak in late spring/early summer which corresponds to this study's findings.

Jia et al. 2008 study

The study by Jia and colleagues in 2008 [93] formulates a model in which the total population is divided into two sub-populations: the immigrant population, and the local population. The SEIR model proposed is divided into two SEIR models, one for the immigrant population and one for the local population. A directional interaction occurs from the immigrant infectious population to the local population. The study neglects to include a

directional interaction from the local population to the immigrant population. Figure 3.7 constructed below illustrates interactions of the model.

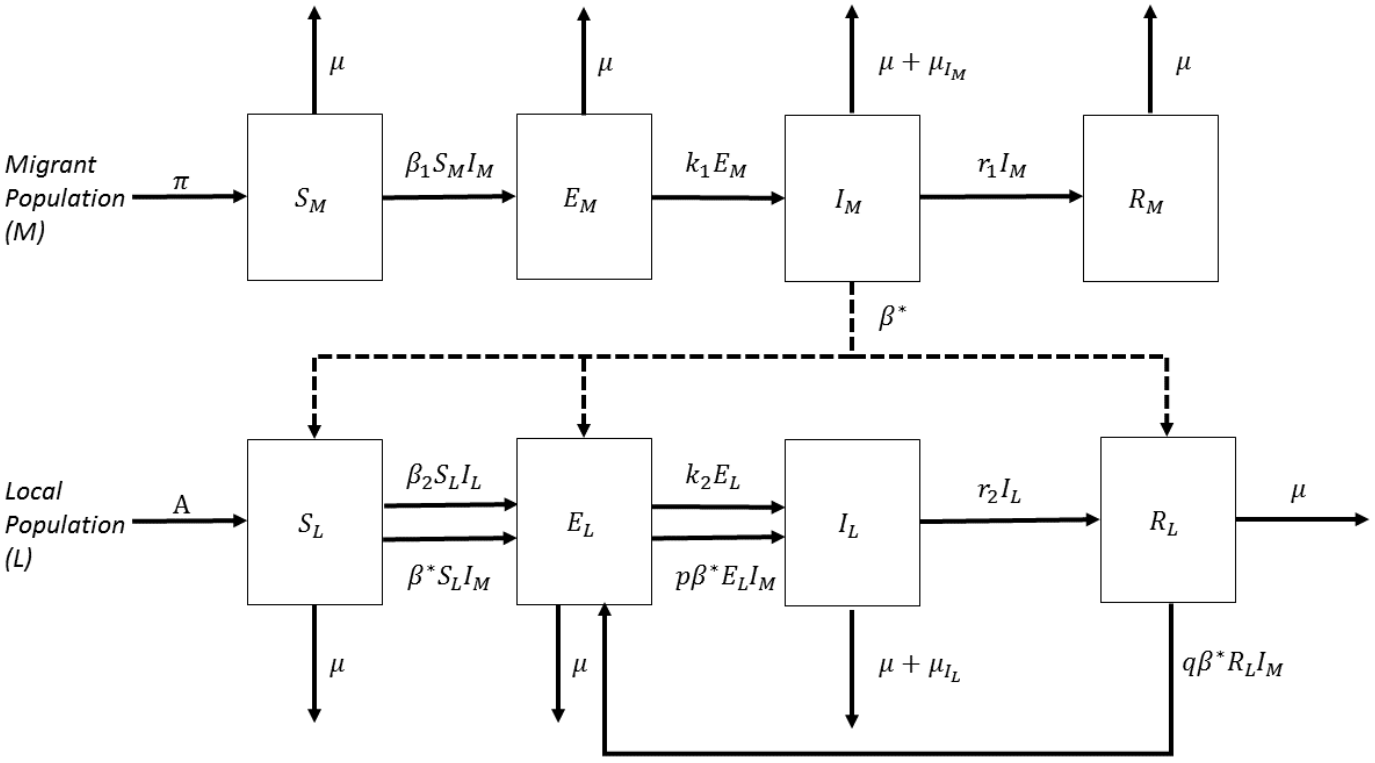

Figure 3.7: SEIR compartment model with foreign-born sub-population partition

The study refers to the above as a "*Basic Model*". The paper goes on to include an "*Extended Model*" within which the recruitment parameter π is divided among the E_M and I_M compartments. This is achieved by way of including additional proportional parameters $0 \leq v \leq 1$ and $0 \leq w \leq 1$, with $0 \leq v + w \leq 1$. The recruitment rate for S_M becomes $\pi(1 - v - w)$, and additional recruitment rates, $v\pi$ and $w\pi$, assigned to E_M and I_M respectively. The author notes that within the extended model, the disease cannot die out, and will remain endemic in the population due to the migration parameters $v > 0$ and $w > 0$.

As noted when simulation was carried out: "*Good agreement was found between the model and data when β^* was sufficiently small*". This correlates with an EU systematic review by Sandgren and colleagues which reported that "*TB in a foreign-born population does not have a significant influence on TB in the native population in EU/EEA*" [94] .

Further work and limitations of the model were reviewed in the discussion section and include the model neglecting to consider a directional effect of the local infectious population on the migrant population.

Discussion of Texts That Did Not Pass Systematic Review

There were twenty five texts that were given a full text evaluation, they will now be briefly discussed.

Seasonal Models

- Bowen et al. [171] - Two TB models are established, one seasonal and one not. The seasonal model was fit to incidence data from Cameroon. The model was based on the model of Liu et al. [89], which is the selected model of this study. The model differs from the original by way of introduction of a reactivation rate within the recovered compartment. Individuals who have recovered will reactivate at some specified rate. This slightly different model was deemed unfit for modelling as certain assumptions of the recovered population are made later on.

- Hu [172] - In this text Hu constructs a mathematical model for TB and conducts a qualitative analysis calculating the basic reproductive number. The model considers a "*Loss Of Sight*" compartment. This compartment represents the population who were once infected and treated but dropped off surveillance. Due to the inclusion of this compartment this text was excluded.

- Leung et al. [173] - In this text, Leung and colleagues complete a statistical analysis of TB incidence. The study highlights the seasonality in Hong Kong TB incidence. A general linear statistical model is considered in this text, hence, it was excluded.

- Parrinello et al. [174] - This analysis performs a comprehensive analysis, examin-

ing the seasonal effects of TB incidence on multiple demographic. The results of this text are discussed later, however, the text considers a statistic model and was excluded.

- Rebelo et al. [157] - This text examines a very general seasonal epidemic model (without necessarily have application to TB). Three seasonal models are then discussed, one of which is a TB model. The TB model discussed is the model selected by this study (Liu et al. [89]).

- Soetens et al. [175] - In this text TB incidence between 1993 to 2014 is considered in the Netherlands. This study performs a statistical analysis within which significant seasonality was detected. Due to use of a statistical model, the text was excluded.

- Sun et al. [176] - A general seasonal model is considered in this text. The model constructed has susceptible, exposed, and infectious compartments and lacks a recovered compartment. Various qualitative analyses are conducted on this model but due to a lack of a recovered compartment it was excluded.

Migrant Models

- Bowong et al. [177] - Bowong and colleuges construct and analyse an SEI model for TB (the SEIR model without a recovered compartment, and a cyclic relationship between infectious and exposed compartments). This analysis did not consider a migrant population and was exlcuded.

- Cohen et al. [178] - Significant demographic differences are detected between individuals of different nationalities with respect to TB incidece. The text does not consider an epidemic model, but rather completes a statistical analysis.

- Colijn et al. [179] - This text reviews an array of models for TB, including age-structure models and spatial temporal models. The text summarises each model and discusses their strengths and weaknesses. A migrant TB model is not considered. This text was excluded on these grounds.

- Denholm et al. [180] - In this text, a mathematical model is constructed and simulated on Australian data. The model considers both immigrant and native populations. The model considered is a discrete stochastic model. The equations are not outline in the text or supplemental material. Due to both these methodologies (discrete, stochastic) the text was excluded.

- Guo et al. [181] - A continuous, deterministic TB model is constructed and analysed. The model considers a migrant population entering into the system (the susceptible, exposed, and infections compartments) at some specified rate. The model does not consider a set of compartments that model migrant populations seperately and because of this it was excluded.

- Guo et al. [182] - A TB model considering immigration is constructed and analysed. The model proposed considers early and late stage latent compartments. Migrants are categorized into high-low risk groups and modelled. Due to not considering a seperate compartment for migrants the model was exculded.

- Hill et al. [183] - Hill and colleagues consider an SEI model for TB. The model considers a migrant population entering into the exposed and susceptible compartments at some specified rate. The model does not consider seperate compartments for migrant populations. Due to not considering this or a recovered compartment, the text was excluded.

- Klotz et al. [184] - In this text an SELIR model is considered (Susceptible, Exposed, Latent, Infectious, Recovered). The exposed compartment in this instance is the set of individuals that have come into contact with a TB case and the latent population is the set of individuals with latent TB. The model does not consider individuals transitioning from latent to infectious. The migrant population enter into the system (through the exposed and susceptible compartments) at a specified rate. The model was excluded as it did not consider seperate compartments for the migrant population.

- Li et al. [185] - In this text a basic susceptible and infectious compartmental model

is considered. The model is a general one and does not have strict application to TB, although the author mentions the possibility of modelling TB with it. A migrant population is considered, however due to the simplicity of the model it was excluded.

- Ma et al. [186] - In this novel, the fundementals of epidemology are defined and discussed. A vast series of models are also listed and discussed throughout the course of the novel, including a model considering TB and the effects of migration. This specific model is a discrete model, and was excluded for this reason.

- Okuonghae et al. [187] - Okuonghae and colleagues construct a TB model derived from a standard SEIR model. The model divides the infectious compartment into two compartments. One is a typical infecious compartment, the other is an infecious compartment with individuals who are isolated from the population. The migrant population enters into the susceptible and exposed and infectious compartments at some specified rate. The model does not consider seperate compartments for migrant populations and was excluded.

- Varughese et al. [188] - In this text, the migrantn population in Canada between 1986 and 2002 was stratified by three incidence groups, two age groups, and three groups considering the time of arrival of the migrant individual. The model considered is a deterministic mathematical model, however, due the the unconvential compartments choosen the model was excluded.

- Wolleswinkel et al. [189] - Wolleswinkel and colleuges contruct an intricate model for TB with application to The Netherlands. The model examines the population entering through the treatment process and attempts the model the rate at which various treatments are successful. The model considers over 12 compartments. The paper models various immigration scenarios and their effects on the model predictions. Due to not directly considering migrant populations in the model this text was excluded.

- Yang et al. [190] - In this text, two epidemic models are considered and analysed. A basic SIR and SI (susceptible, infectious) models are considered. The models

do not necassarily have application to TB and do not consider a recovered/treated compartment. For this reason they were excluded.

- Zhou et al. [191] - A tuberculosis model is constructed to model the incidence of TB on the migrant and native populations in Canada. The model considers seperate compartments and is very similar to the one selected by this study [93]. However, the model considers discrete time, and for this reason was excluded.

3.6 Conclusion

This chapter presented classical mathematical models and various modelling types, along with a justification and review of certain models. The modelling process was introduced and a historical review of epidemic modelling was conducted, the review of such material providing context and foundation for further work to be carried out. A review of SIR and SEIR compartmental models was conducted illustrating their respective flowcharts and detailing the underlying basic reproductive number of the system of ODE's.

The chapter went on to produce a systematic search of the literature. The objective of the search was to acquire suitable models to analyse and simulate. The search was informed by an exploratory data analysis detailed of which are set out in the following chapter. After exclusion and inclusion criteria were applied, the resulting two models were then presented and discussed. The model considering seasonality achieves seasonal variation in the system of equations by changing the transmission and reactivation parameters from constants to periodic functions of time. The model stratifying the total population into a native and foreign-born sub-population mathematically considers this by dividing the underlying SEIR model into two groups, migrant and local. A one way interaction is present in the model between the migrant infectious class and the entire population. The parameter describing the interaction was discovered to be small. Models resulting from the systematic search will be built upon in §5 and §6.

Chapter 4

Exploratory Data Analysis

4.1 Introduction

This chapter presents the exploratory analysis of the study data as a preliminary step for deterministically modelling tuberculosis incidence in Ireland. Epidemic models are, in general, constructed from a realisation of random processes, hence it is necessary to establish the distributional properties of those processes in order to accurately model them. This will ultimately help to make informed decisions concerning which features the models incorporate and to ultimately assist minimising the residuals between the model and data.

In the following sections of this chapter an overview of the dataset is given, assessing the type of variables within it as well as the completeness of those variables. A descriptive analysis is then completed on variables individually and each variable is commented on. A descriptive time series analysis is then conducted to investigate seasonality within notifications. Standard time series methods are introduced and implemented from which results are derived. The chapter concludes with a descriptive multivariate analysis, and insights from the data are translated into attributes of the underlying epidemic model.

Study Area

Ireland is an island on the western fringe of Europe. It one of 28 countries in the European Union. There is a strong and continuing migration from rural areas to towns and cities; 52% of the population now live in urban areas of 1,500 inhabitants or more. The influence of Dublin (the capital of Ireland) and other cities and urban areas have a very clear impact on the population throughout the country.

Ireland has a temperate maritime climate, ranging from 4 degrees Celsius to 7 degrees Celsius in the coldest months during January to February and the warmest months in July and August ranging from 14 degrees Celsius to 16 degrees Celsius.

Economically Ireland underwent a period of increased economic growth from the mid 1990's to the mid 2000's. This period is referred to as the "Celtic Tiger". During this period and up to 2006 the population increase by approximately 500,000 individuals, the majority of which was made up by non-Irish citizens. This led, therefore, to a large increase in the overall population. In 2008, Ireland went into economic recession resulting in a net outward migration from 2007 until present (2015).

4.2 Methodology

4.2.1 Data Quality And Acquirement

All denominator data (data referring to population) were obtained from Irish national census data [96]. The data obtained spanned five years. Intermediate years were calculated by linear interpolation, which follows the World Health Organisations guidelines for estimating denominator data [192]. This study ackowledges possible over or under reporting effects on these data. When foreign-born populations are considered, these data are possibly under recorded. The effects of under reporting on the foreign-born population will be an inflated incidence rate of TB for the foreign-born population. This is due to the fact that TB notification is system is mandatory (referred to as notifiable) in Ireland, meaning each

case, once recognised in a public or private healthcare establishment, must be recorded. There may exist foreign-born living in Ireland who have not been documented, and this will deflate denominator data for this population.

A descriptive analysis is conducted on foreign-born TB incidence and comparisons made between the incidence of TB in the foreign-born individuals birth country and the incidence foreign-born from the same country living in Ireland exhibit. Incidence data for each country listed was acquired from the World Health Organisation [193]. The WHO recognise their database is updated regularly as countries notify WHO of corrections to previously submitted data. These data were extracted on January 2016. The data quality is subject to WHO data collection methods which further information can be found here [194].

All Irish data pertaining to tuberculosis notifications along with related demographic and risk factor data were obtained from the Health Protection Surveillance Centre (HPSC)[2]. The HPSC is an agency of the Health Service Executive (HSE) and is dedicated to the surveillance of communicable diseases in Ireland. The HPSC is a national repository of TB incidence data and does not directly collect data. Tuberculosis is a notifiable disease within Ireland. All practitioners are required to compile information from hospital charts and interviews through a national notification form (Appendix B). The TB notification form is completed by public health doctors for each case of TB notified. These forms summarise all available clinical, microbiological, histological and epidemiological data. Forms were then collated in the regional departments of public health, where data were entered onto the Computerised Infectious Disease Reporting (CIDR) system. Finalised data (with outcome information) were extracted from CIDR for this study. Laboratories report all positive specimens to the same system. As laboratory analysis takes time each specimen gets linked to the relative identification number at a later date. The notification form summarises all clinical, microbiological, historical, and epidemiological data. Further information on HPSC data collection methods, TB screening methods, and various HPSC definitions can be found on the HPSC website [147]. The quality of these data

are subject to the healthcare practitioners filling out the surveillance form. This study acknowledges that this is not a full-proof system and that, although notifiable under national guidelines, there may exist a bias and missing values within these data. This study presents the number of missing values and completeness of the data it received in §4.2.2.

Datasets were acquired from the HPSC on two occasions. The initial dataset was obtained in January 2014 after ethical approval was granted by the research ethics committee of the Adelaide and Meath Hospital. This is an ethical approval recognised by Trinity College's Faculty of Health Sciences Ethics Committee. This dataset contained national notification, demographic, and risk factor data. Records within the dataset began on the 1^{st} of January 2002 and continued until the 31^{st} of December 2012. Figure 4.1 displays the progression of events over time.

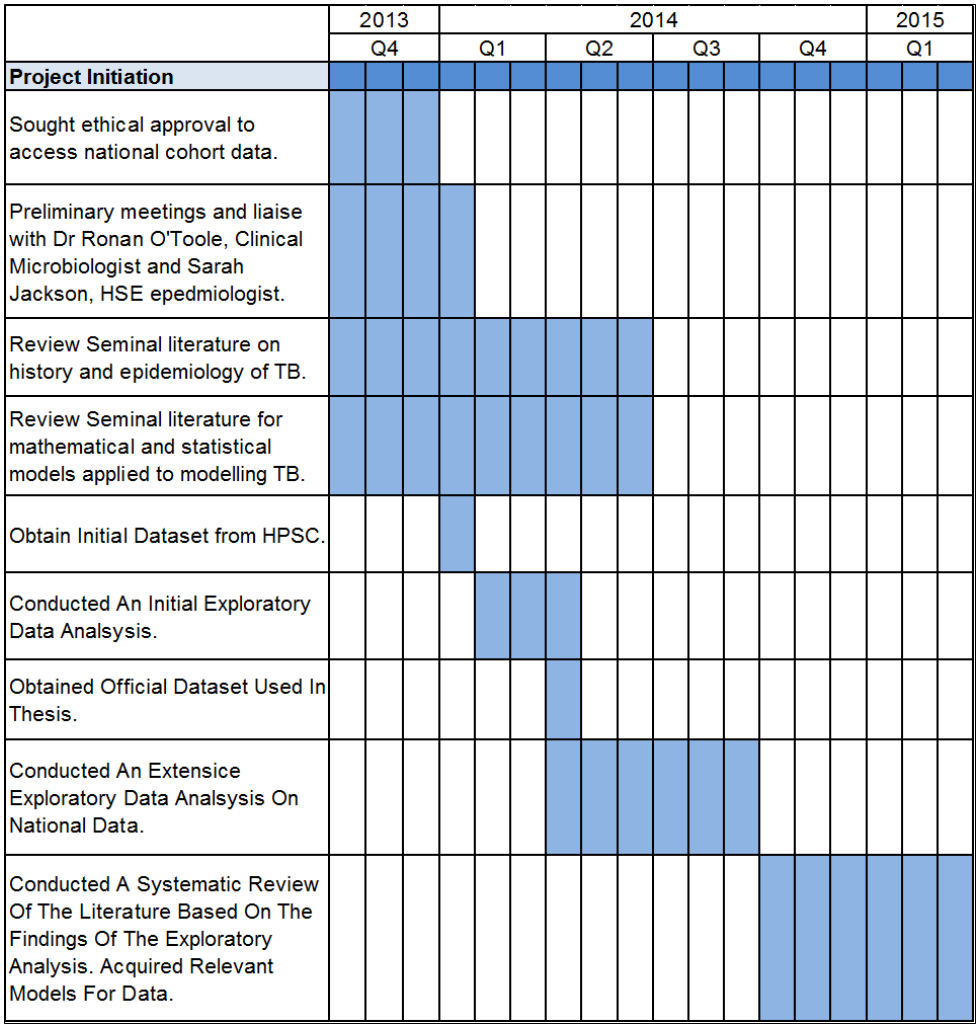

Figure 4.1: Gantt Chart Displaying Progression Of Research Over Time

A descriptive analysis was conducted on these data and the results presented to HPSC stakeholders in April 2014. After reviewing and discussing results, a memorandum of understanding was formed and signed between the research team and the HPSC (found in Appendix C). The purpose of the memorandum was to outline the nature of the data being made available to the research team and also to satisfy all parties involved that data protection legislation would not be breached. Once the memorandum was signed and all parties were in agreement, the relevant authority within the HPSC then sanctioned access to more

up-to-date data including national notification, risk factor, and demographic records collected from the 1^{st} of January 2002 to the 31^{st} of December 2013. All subsequent analysis was conducted using this, later, dataset and any time periods referred to reference this period. Further annual national notification data, from 1991 to 2002 were also acquired through the HPSC's annual surveillance reports [97].

4.2.2 Data description

The dataset contained variables that were partitioned into three subgroups: TB data, demographic data, and risk factor data. The completeness of the data was subject to healthcare practitioners completing the surveillance form correctly. The notification form has "Other" and "Unknown" options for a proportion of the questions. For the following tables (4.1-4.3) "Other" and "Unknown" are considered a valid response and do not contribute to the incompleteness of the data. Completeness refers to the number of responses not left blank or empty within the dataset. Data cleaning was not required for the majority of variables. Tables 4.1 to 4.3 below detail the variables within each of the three subgroups. Comments have been included for each table.

TB Data			
Variables	**Completeness N (%)**	**Type**	**Comment**
ID	5160 (100%)	Integer	Numeric identification
Year	5160 (100%)	Integer	Year of notification
Month	5160 (100%)	Integer	Month of notification
Date of Onset	3623 (70%)	Date	Estimated TB onset date
Date of Diagnosis	4748 (92%)	Date	Date of diagnosis
Date of Notification	5151 (100%)	Date	Date of Notification
Diagnosis Type	5155 (100%)	Nominal	Type of TB: Pulmonary, Extrapulmonary, Both
MDR/XDR	34 (1%)	Nominal	Multiple drug resistant TB strain/ Extensively drug resistant TB strain
TB Cause of Death	3013 (58%)	Boolean	Was the TB a cause of death
Date of Death	267 (5%)	Date	Date of Death

Table 4.1: Variable descriptor for dataset subgroup: TB data

Demographic Data			
	Completeness		
Variables	N (%)	Type	Comment
ID	5160 (100%)	Integer	Numeric identification
Year	5160 (100%)	Integer	Year of notification
Month	5160 (100%)	Integer	Month of notification
Sex	5160 (100%)	Nominal	Sex/Gender
Age	5152 (100%)	Integer	Age at time of notification
Employment Status	4685 (91%)	Nominal	Working status at time of notification
Current Living	4842 (94%)	Nominal	The type of structure the individual lived at the time of notification
Birth Country	5150 (100%)	Nominal	Country of birth
Race/Ethnicity	5057 (98%)	Nominal	Ethnicity classified as a combination of racial category and social background
Refugee	4938 (96%)	Boolean	Refugee Status
Year of Entry	492 (25%)	Integer	If foreign-born, what year did the individual enter Ireland

Table 4.2: Variable descriptor for dataset subgroup: Demographic data

Risk Factor Data			
Variables	Completeness N (%)	Type	Comment
ID	5160 (100%)	Integer	Numeric identification
Year	5160 (100%)	Integer	Year of notification
Month	5160 (100%)	Integer	Month of notification
Risk Factors	5059 (98%)	Boolean	Does the patient have one of the following risk factors
High Endemicity Origin	1885 (37%)	Boolean	Does the individual originate from a country of high endemicity
High Endemicity Residence	552 (11%)	Boolean	Does the individual have usual residence in a country of high endemicity
Anti-TNF	458 (9%)	Boolean	Is the individual undergoing treatment with tumor necrosis factor inhibitor
Contact	769 (15%)	Boolean	Was the individual in close contact with an active TB case
Diabetes	498 (10%)	Boolean	Is the individual diagnosed with Diabetes
Immunosuppressive Illness	656 (13%)	Boolean	Is the individual diagnosed with some form of Immunosuppressive illness
Immunosuppressive Medicine	505 (10%)	Boolean	Is the individual taking some form of Immunosuppressive medication
Alcohol Misuse	860 (17%)	Boolean	Does the individual have a history of alcohol misuse
Drug Misuse	507 (10%)	Boolean	Does the individual have a history of drug misuse
Tabacco	27 (1%)	Boolean	Does the individual smoke tobacco
Other	350 (7%)	Boolean	Does the individual have an unspecified risk factor
HIV Stautus	5001 (97%)	Boolean	Current HIV status

Table 4.3: Variable descriptor for dataset subgroup: Risk Factor data

TB Data Comment: The validity of some of the variables of type DATE within TB data are questionable. The possible orders of disease progression are either: Onset $\rightarrow$ Diagnosis $\rightarrow$ Notification $\rightarrow$ Death, Onset $\rightarrow$ Death $\rightarrow$ Diagnosis $\rightarrow$ Notification, or Onset $\rightarrow$ Diagnosis $\rightarrow$ Death $\rightarrow$ Notification. However a proportion of the notification and diagnosis dates pre-dates onset date, this implies onset of the disease occurred after a valid diagnosis was given. Date of death also occasionally pre-dates onset dates, which cannot be true. For this reason the data that had Onset date after Diagnosis, Onset date after Notification, and Onset date after Death, were labelled as invalid dates in the dataset.

Demographic Data Comment: No data cleaning was required for demographic data.

Risk Factor Data Comment: No data cleaning was required for risk factor data. Risk factors were not mutually exclusive (e.g. an individual could have Other and Tobacco). The variable "Year Of Entry" has been excluded from analysis as it contains null valued entries 75% of the time.

A small set of new variables were derived from the initial set. These included: "*Duration Of Illess*" which was created by subtracting the variable *Date of Diagnosis* from the variable *Date of Onset*, "*Substance Misuse*" (*Alcohol Misuse* OR *Drug Misuse*), and "*High Endemicity Affiliation*" (*High Endemicity Origin* OR *High Endemicity Resident*). Risk factor groups such as: organ transplant patients, resident/workers of high-risk congregate settings, mycobacteriology laboratory personnel, and other persons with clinical conditions placing them at high risk were not recorded within the national dataset and, hence, could not be examined.
Given the data had been cleaned and formatted, analysis then took place.

4.3 Descriptive Analysis

Modelling is typically time dependent, hence variables will be explored over time. The main aim of this and subsequent sections, is to lend insight to the modelling process.

The method of exploration was to examine the variables individually over time and then based on findings explore certain variables further. The majority of data and descriptive statistics can be found in Appendix A. Each variable is discussed on the merit of findings. Annual incidence rates were calculated in subsequent sections. Incidence is defined as the yearly count of reported cases/notifications per 100,000 resident population unless specified elsewhere.

4.3.1 National Notification and Incidence Data

National annual incidence can seen in figure 4.2. As the dataset contained data from 2002 to 2013, this section also incorporates additional notification data before 2002. These data were acquired from the HPSC's epidemiological reports [97].

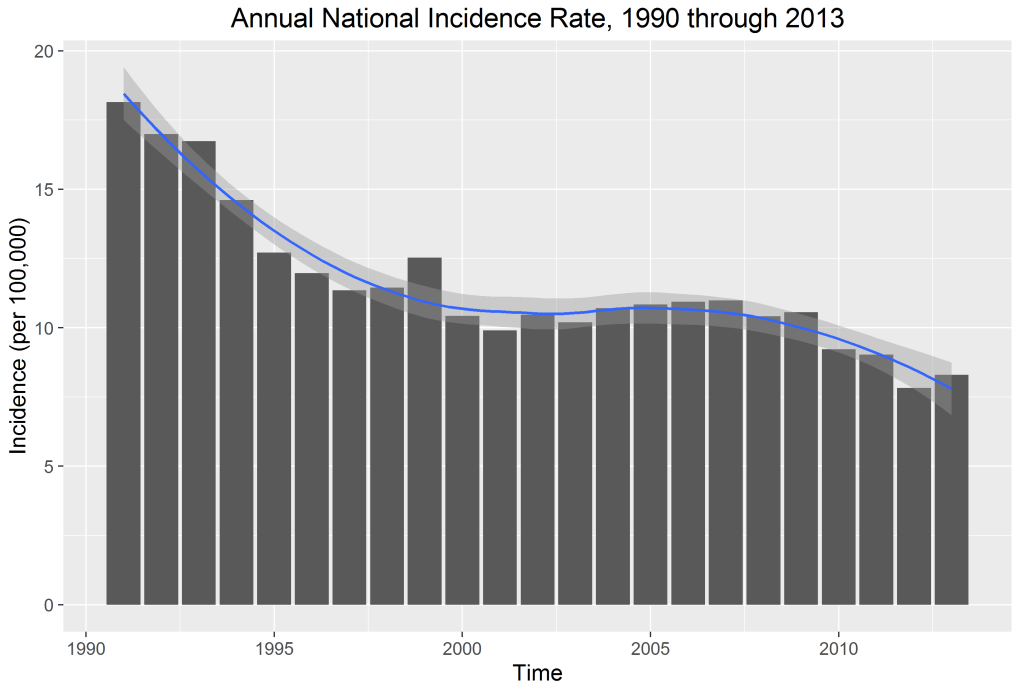

Figure 4.2: National incidence rates from 1991 to 2013

In 1991, the total number of reported cases was 640. In 2013, annual notifications had

dropped to 381, which accounts for a 60% decline in notifications since 1991. From the year 1991 to the year 2000 there was a decline of 27 cases on average, each year. Annual notifications began levelling out thereafter and even increasing; from 2000 to 2009 underlying cases increased on average nine per year. The increase in notifications thereafter appeared to stop, from 2009 to 2013 a decline was seen again, decreasing 25 cases on average, each year.

The WHO considers a country to have low incidence if the national annual incidence rate is 10 per 100,000 or less [98]. Ireland entered into this category in 2009 and has remained there for the last number of years.

4.3.2 TB, Demographical, And Risk Factor Data

For the following sections, unless specified, all incidence rates are per 100,000 of the respective population. For figures 4.2, 4.5 and 4.6 a local loess regression (Local Polynomial Regression) smoother [99] is applied to the data with a 95% confidence interval. The smoother is added to visually aid identifying the underlying trend in the data.

Gender

The ratio of male-to-female cases remained relatively constant over time. The percentage of notifications male ranged from 57.62% to 63.57% (Mean=61.12%, Std.Dev=1.81%) from 2002 to 2013, with an annual average change of -0.02%. The discrepancy between sexes has been noted in literature [100] as occurring in multiple other countries as possibly being a discrepancy due to inherent biological mechanisms between genders. Appendix table A.5/A.6 contains additional data and descriptives for gender.

Age

Figure 4.3 shows a density distribution of age in 2002 and 2013. The density goes from a bimodal density in 2002 to a uni-modal density in 2013. Intermediate years gradually show this transformation. The annual incidence rate of older populations has shown a decline

over time. In 2002 incidence of individuals aged 85 or older was 21.58 per 100,000, the population has seen an annual reduction in incidence of 1.38 units per year, in 2013 incidence was 6.44 per 100,000. A similar decline has been seen in ages over 60 years old. Despite a decline in incidence, these populations have a relatively high average incidence over the time period, ranging from 10.6 in the 60 to 64 year old population up to 21.14 within the population 85 or older. With regard children aged less than 15, a decline has been seen in incidence. Average incidence ranges between 1.43 in the five to nine year old population up to 3.09 in the less than five year old population. A notable incidence rate was observed in the population aged less than five in the year 2007. Incidence peaked at 9.02 increasing 125% from the previous year. The HPSC has attributed this peak to an outbreak occurring in a large crèche of which 21 children and three adults were infected [101].

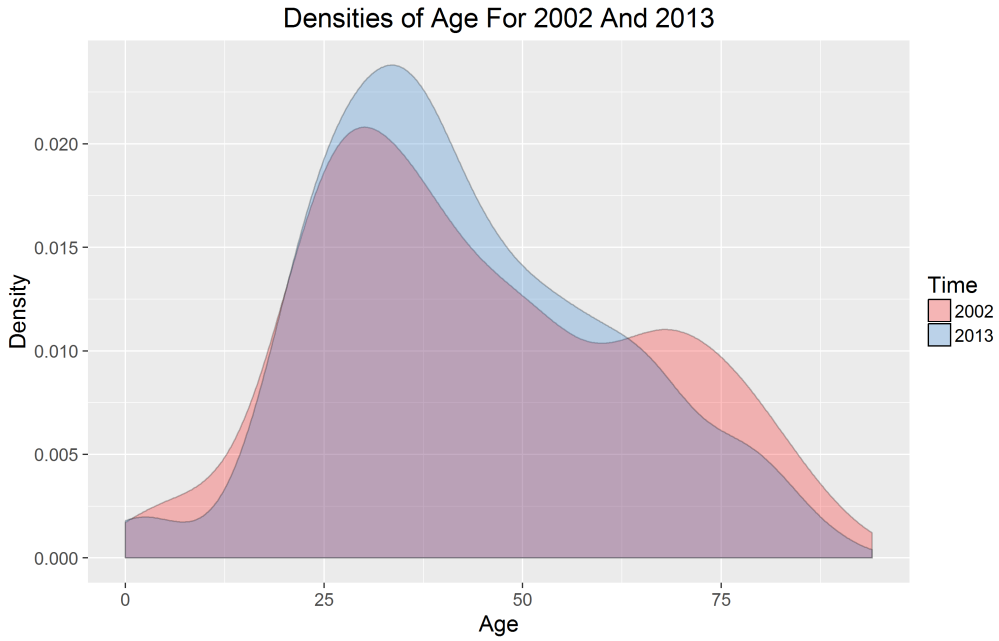

Figure 4.3: Comparing Density Of Age, 2000 and 2013

Other than older populations, incidence is relatively high within the 20 to 40 year old population, average incidence ranging between 12.1 per 100,000 individuals in the 20 to 24

year old category to 16.82 per 100,000 in the 25 to 29 year old population. The age group experiencing the largest increase in incidence is the population aged 20 to 24, followed by the population aged between 50 to 55. Multivariate analysis later gives an explanation of the change in the density of notifications. Appendix table A.7-A.9 contains data and descriptive statistics of the variable age categorized in five year intervals.

Occupation

Employment status contained nine categories into which the population was categorized: (1) Employed, (2) Unemployed, (3/4) Housewife/House-husband, (5) Retired, (6) Student, (7) Other, (8) Child, (9) Unknown. The proportion of cases occurring in each category was relatively consistent over time. The categories Unemployed and Unknown saw an average annual increase of 0.7% and 0.4% respectively, while the categories Retired and Employed saw the largest decrease, 0.5% and 0.4% respectively. Notable incidence rates were observed within the unemployed population, averaging 26.81 each year, peaking in 2006 reaching 45.63 per 100,000 of the population. The standard deviation of incidence for the Unemployed population is 14.24 over the time period. The high incidence rate was seen prior to 2009 when the population was within the range of 182,141 to 283,907 individuals. The unemployed population grew substantially from 2009 onward reaching a peak of 829,381 in 2011 due to the economic downturn. The incidence rate declined with this growth occurring in the population. In 2013 incidence was 13.65. Appendix table A.10/A.11 contains data and descriptives of the variable occupation and figure 4.4 illustrates the annual incidence rates between employed and unemployed individuals.

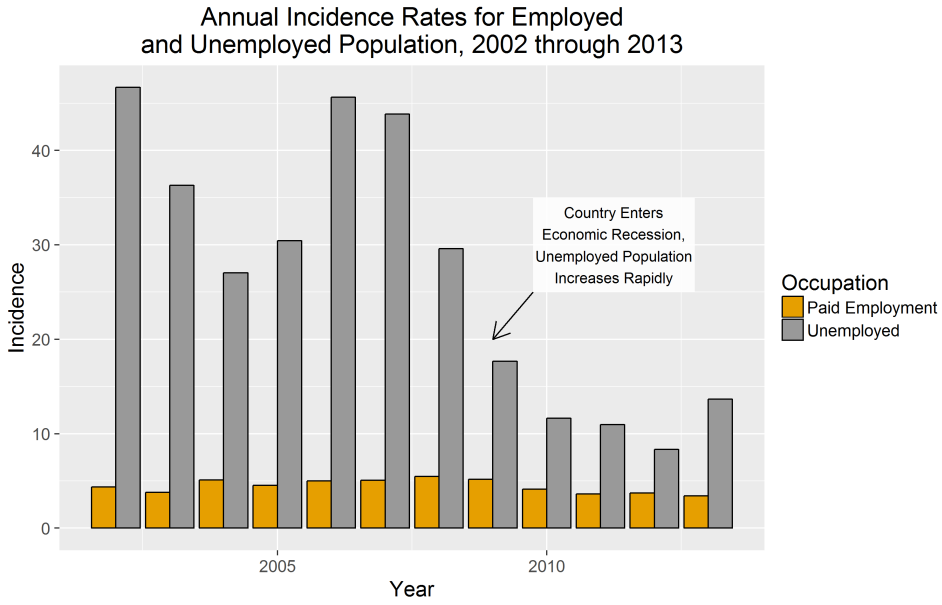

Figure 4.4: Incidence Rates for the Employed and Unemployed Populations, 2000 through to 2013

Current Living

The Current living variable had eight categories: (1) Home, (2) Hostel, (3) B&B/Hostel, (4) Homeless, (5) Prison, (6) Institution, (7) Other, and (8) Unknown. The majority of notifications occurred within the population living in a home. On average 85% of cases were in this category, decreasing 0.01% annually. The prison population witnessed a peak in notifications during 2011 with an incidence rate of 86.01 per 100,000 that year. The HPSC have attributed this peak in notifications to an outbreak that occurred in 2011 [112]. A total of 12 individuals residing in a prison were infected that year. The proportion of those cases attributed to the outbreak is unknown. Aside from 2011, the incidence for the prison population has fluctuated between 0 and 34.54; underlying cases ranging between 0 and 3 each year. For the homeless population, denominator data was only available for one year, 2011. The incidence for that year was 26.26 per 100,000. The underlying cases within the population have seen notifications in the range of zero to seven cases a year. Appendix A.12/A.13 table contains data and descriptives of the variable Current Living.

Birth Country

The values for the variable Birth Country were initially categorised as: (1) Ireland, (2) Other; Other meaning a foreign birthplace, and (3) Unknown. Analysis for each specific birth country can be found in §4.3.4, this section will simply examine native and foreign-born cases. Figure 4.5 illustrates foreign and native-born cases over the time period, the national notifications are included to give additional context.

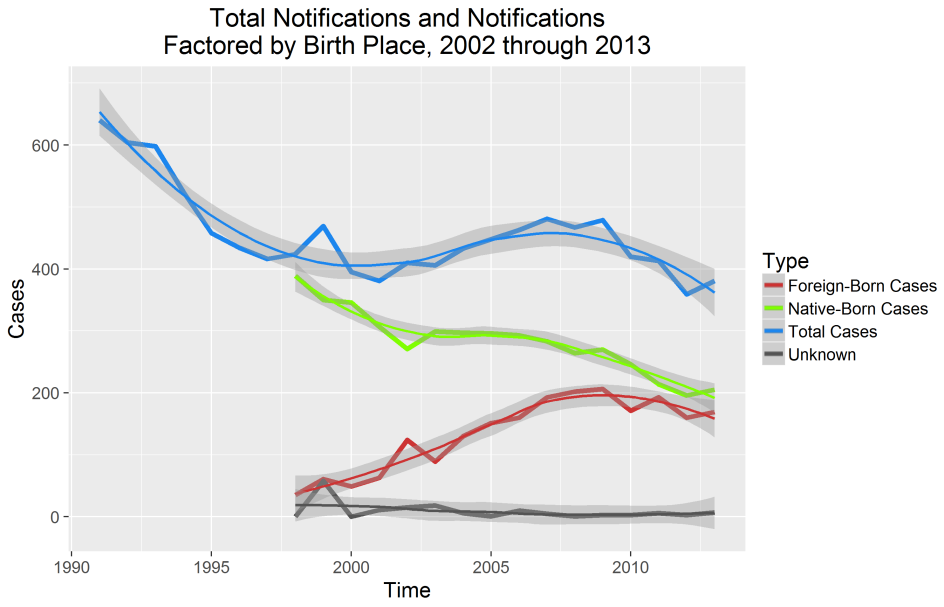

Figure 4.5: Foreign and Native Born Cases, 1998 through to 2013

The data indicate that the burden of TB falls more heavily on the foreign-born population. There were 271 notifications in 2002 attributed to the native-born population. This figure decreased 32%, reaching 205 cases in 2013. There was a close to linear decrease in native-born incidence over the time period (see figure 4.6). Underlying notifications on average decreased 1.1% each year. For the foreign-born population, notifications have, on average, increased 1.3% each year, with an average incidence rate of 26.27 per 100,000 and standard deviation of 4.62. Figure 4.6 and tables 4.4 and 4.5 detail the underlying incidence and data over time, respectively.

Birth-Place (N)	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
Ireland	271	299	297	296	293	283	264	270	246	214	196	205
Other	124	89	130	151	160	193	202	206	171	193	160	169
Unk.	15	18	6	1	10	5	1	3	3	6	3	7

Birth-Place (%)	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
Ireland	66.1%	73.6%	68.6%	66.1%	63.3%	58.8%	56.5%	56.4%	58.6%	51.8%	54.6%	53.8%
Other	30.2%	21.9%	30.0%	33.7%	34.6%	40.1%	43.3%	43.0%	40.7%	46.7%	44.6%	44.4%
Unk.	3.7%	4.4%	1.4%	0.2%	2.2%	1.0%	0.2%	0.6%	0.7%	1.5%	0.8%	1.8%

Birth-Place Incidence	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
Ireland	7.68	8.47	8.35	8.21	8.07	7.51	6.88	7.02	6.43	5.60	5.19	5.47
Other	31.79	19.84	26.62	28.58	26.59	31.74	31.18	29.95	23.50	25.65	19.82	19.95

Table 4.4: Notifications (Count, Percentage, Incidence) Categorized By Birthplace

Statistics (N)	Min	Q1	Median	Q3	Max	Mean	S.Dev	Mean Change
Ireland	196	238	270.5	293.75	299	261.17	37.49	-6.00
Other	89	145.75	164.5	193	206	162.33	35.05	4.09
Unknown	1	3	5.5	7.75	18	6.50	5.37	-0.73
Statistics (%)	Min	Q1	Median	Q3	Max	Mean	S.Dev	Mean Change
Ireland	51.8%	55.9%	58.7%	66.1%	73.6%	60.7%	6.8%	-1.1%
Other	21.9%	32.8%	40.4%	43.5%	46.7%	37.8%	7.6%	1.3%
Unknown	0.2%	0.7%	1.2%	1.9%	4.4%	1.5%	1.3%	-0.2%
Statistics Incidence	Min	Q1	Median	Q3	Max	Mean	S.Dev	Mean Change
Ireland	5.19	6.22	7.27	8.10	8.47	7.07	1.17	-0.20
Other	19.82	22.61	26.60	30.26	31.79	26.27	4.62	-1.08

Table 4.5: Statistics (Count, Percentage, Incidence) Categorized By Birthplace

The foreign-born population increased rapidly in Ireland during the review period. In 2002, the population made up approximately 9.96% of the total population; in 2013 it was estimated to be approximately 18.44% of the total population.

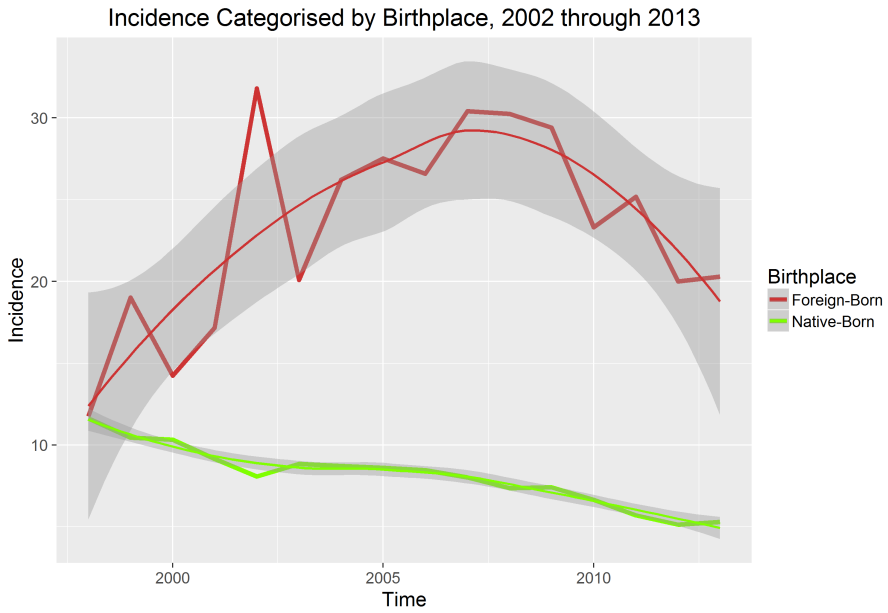

Figure 4.6: National Foreign and Native Born Incidence, 1998 through to 2013

The broad and ambiguous definition of "foreign-born" has led to a relatively larger variance in incidence rates for the population over time, the average standard deviation of incidence rates being four times that of native-born. Section 4.3.4 gives a further examination of foreign-born cases within Ireland.

Race/Ethnicity

The Race/Ethnicity variable has seven categories: (1) White, (2) South Asian, (3) Black, (4) Irish Traveller, (5) East/South East Asia, (6) Other, (7) Unknown. The proportion of cases occurring in each category in 2002 were 71% White, 8% South Asian, 11.5% Black, 0.2% Irish Traveller, 1.7% East/South East Asian, 1% Other, and 6.6% Unknown. The proportions in 2013 were 52% White, 12.6% Asian, 8.4% Black, 3.1% Irish Traveller,

7.1% East/South East Asian, 0.3% Other, and 16.5% Unknown. Average Incidence per 100,000 of the population over time within each category was 6.98 for White, 98.05 for Black, 116.2 for Asian (South Asian and East/South East Asian combined), 8.13 for Irish Traveller, 16.14 for Other, and 39.68 for Unknown. The average annual change in incidence over time was -0.30 for White, -11.46 for Black, -6.75 for Asian, +2.83 for Irish Traveller, -0.48 for Other, and +4.91 for Unknown. A large burden of incidence consistently occurs with Black and Asian populations. Underlying notifications have on average changed each year -8.45 for White, +1.36 for South Asian, -1.36 for Black, +1 for Irish Traveller, +1.82 for East/South East Asian, -0.27 for Other, and +3.27 for Unknown. Further data and descriptive statistics can be found in Appendix table A.14/A.15.

Additional Remark: A publication conducting descriptive statistics on the Irish traveller community can be found in Appendix E.

Refugee/Asylum Seeker

The variable Refugee/Asylum Seeker had responses (1) Yes, (2) No, and (3) an Unknown category. The annual incidence in the refugee population was one of the largest within the dataset, reaching a maximum of 855 per 100,000 refugees in 2002. However the incidence rate declined almost linearly since then reaching 150 in 2013; on average declining 64.09 units each year. This is one of the largest reductions occurring for incidence rates in the dataset. The underlying population was 5,380 individuals in 2002, climbing linearly to reach 9,730 in 2008, then began declining reaching 6,001 in 2013. The birth country composition of refugees has changed over time. In 2002, the global composition primarily consisted of Afghani and African refugees [106], of which high incidence rates were observed. More recently, in 2014/2015, the global refugee population comprised mostly of Syrian refugees [107]. Syria is categorised as low to medium incidence experiencing an incidence rate between 10-20 per 100,000 resident population annually. The change in the composition of the global refugee population ultimately changes its underlying incidence rate of TB . The cases occurring in Ireland range between nine and 46 annually, the

minimum number of cases occurring in 2013. Further data and descriptive statistics can be found in the Appendix table A.16.

Diagnosis Type

The variable Diagnosis Type or Disease Type had four categories: (1) Pulmonary, (2) Extrapulmonary, (3) Pulmonary&Extrapulmonary, and (4) Unknown. In 2002, the proportion of cases occurring in each category was as follows: 65.6% Pulmonary; 23.4% Extrapulmonary; 9% Pulmonary&Extrapulmonary; and 2% Unknown. The proportions changed slightly over the review period. In 2013 they were 59.8% Pulmonary, 33.6% Extrapulmonary, 6% Pulmonary&Extrapulmonary, and 0% Unknown. Within Europe, the WHO identified approximately 17% of all cases having Extrapulmonary TB in 2011. Ireland has experienced an average of 29.2% over the time period of the study indicating a slightly higher rate than other countries. Further data and descriptive statistics can be found in the Appendix table A.1/A.2. Figure 4.7 illustrates the proportion of cases for each disease type.

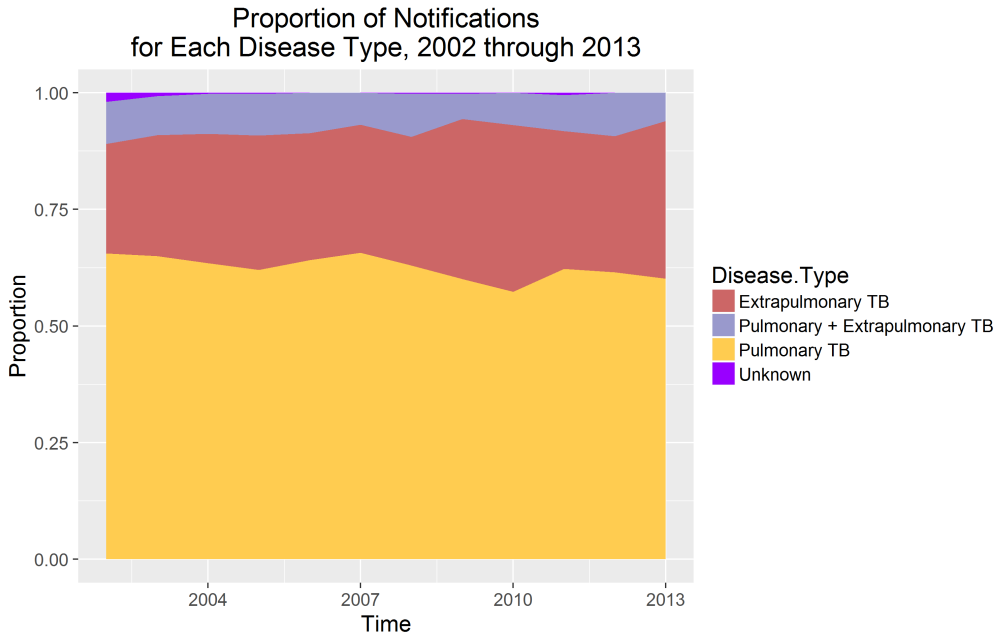

Figure 4.7: Proportion of Notifications with each Diagnosis Type, 2000 through to 2013

MDR/XDR

The variable MDR/XDR refers to the multiple drug-resistant strain and extensively drug-resistant strains of TB. The vast majority of cases in Ireland have been the standard drug susceptible strain, averaging 99.4% of all cases over the review period. Being drug-resistant, MDR and XDR TB strains are difficult to treat and although the number of cases is small, examining occurrences is important. The number of MDR-TB cases peaked in 2007, reaching seven cases. The next highest was in 2012 when there were five cases. Other years saw a range between zero to four cases. There was one recorded case of XDR-TB in Ireland that occurred in 2005. No deaths have been recorded due to MDR/XDR TB over the time period. The average age of individuals with MDR-TB was 35.7. The age of the individual with XDR TB was 26. With regards to employment, nine cases (27.27%) were unemployed, six were students (18.18%), five were in employment (15.15%), four were unknown (12.12%), and the remaining categories had three each; Other (9.1%), Retired (9.1%), and Husband/Housewife (9.1%). Further data and descriptive statistics can

be found in the Appendix table A.3.

Death Due To TB

The variable Death Due To TB had responses: (1) Yes, (2) No, and (3) Unknown. The number of cases classified as Unknowns were consistently large each year creating a high degree of uncertainty. To add to this uncertainty, a value of "Yes" was given only when individuals died solely due to TB. Consequently, TB being a contributing factor towards death was not sufficient to constitute a "Yes" response, TB needed to be the primary factor. The true count of deaths where TB was a factor is unknown in Ireland. The number of people dying from TB each year ranges between 0 and 11 individuals over the time period of the study, this corresponds to a death rate between 0 and 2.5% for individuals with TB. Table 4.6 shows the frequency of death with respect to age category, further data and descriptive statistics can be found in the Appendix table A.4.

20 Or Less	20 to 30	30 to 40	40 to 50	50 to 60	60 to 70	70 to 80	80 to 90	Over 90
0	4	9	6	14	15	20	13	4

Table 4.6: Death Due To TB Age Distribution

The mean age of death from TB was 62.7 years (median 61 years, range 26-95 years). The greatest number of deaths (n=20) occurred in the 70 to 80 age bracket.

Risk Factor

In order to asses if there was an association of TB with specific, known risks, patients were asked if they had been exposed to a risk factor and then, if yes, what the type of risk factor was. The risk factors available were: (1) Anti-TNF (Anti-tumor necrosis factor: a class of drugs that increases susceptibility to TB), (2) Contact, (3) Diabetes, (4) High Endemic Affiliation[1](A combination of high endemicity residence and high endemicity origin), (5) Immuno Suppresive Medication, (6) Immuno Suppresive Illness, (7) Other, (8) Substance

[1]Countries with annual TB notification rate greater than 40 cases per 100,000 population are considered areas of high endemicity

Misuse, (9) Tobacco, and (10) Unspecified Risk Factor. The risk factor with the greatest increase in frequency over the time period was High Endemicity Affiliation, on average increasing 1.6% annually. This was followed by individuals who came into contact with a case. This category grew on average 0.5% annually. Of the notifications associated with a risk factor, 49.1% had a High Endemicity Affiliation. The next largest frequency of risk factor was Substance Misuse accounting for 15% of all risk factors. Table 4.7 details the percentage of cases with a risk factor over time and figure 4.8 illustrates the frequency and number of individuals with more than one risk factor. Further data and descriptive statistics can be found in the Appendix table A.17/A.18. Table 4.7 and figure 4.8 detail risk factor data over the review period.

Risk Factor (%)	2002	2003	2004	2005	2006	2007
Yes	49.5%	45.8%	49.0%	46.2%	52.9%	54.9%
No	29.0%	32.8%	29.3%	29.7%	26.1%	27.2%
Unknown	21.5%	21.4%	21.7%	24.1%	21.0%	17.9%
Risk Factor (%)	2008	2009	2010	2011	2012	2013
Yes	65.7%	62.8%	61.2%	71.7%	55.7%	52.8%
No	23.6%	23.8%	24.8%	15.3%	23.4%	19.9%
Unknown	10.7%	13.4%	14.0%	13.1%	20.9%	27.3%

Table 4.7: Percentage of Cases With A Risk Factor

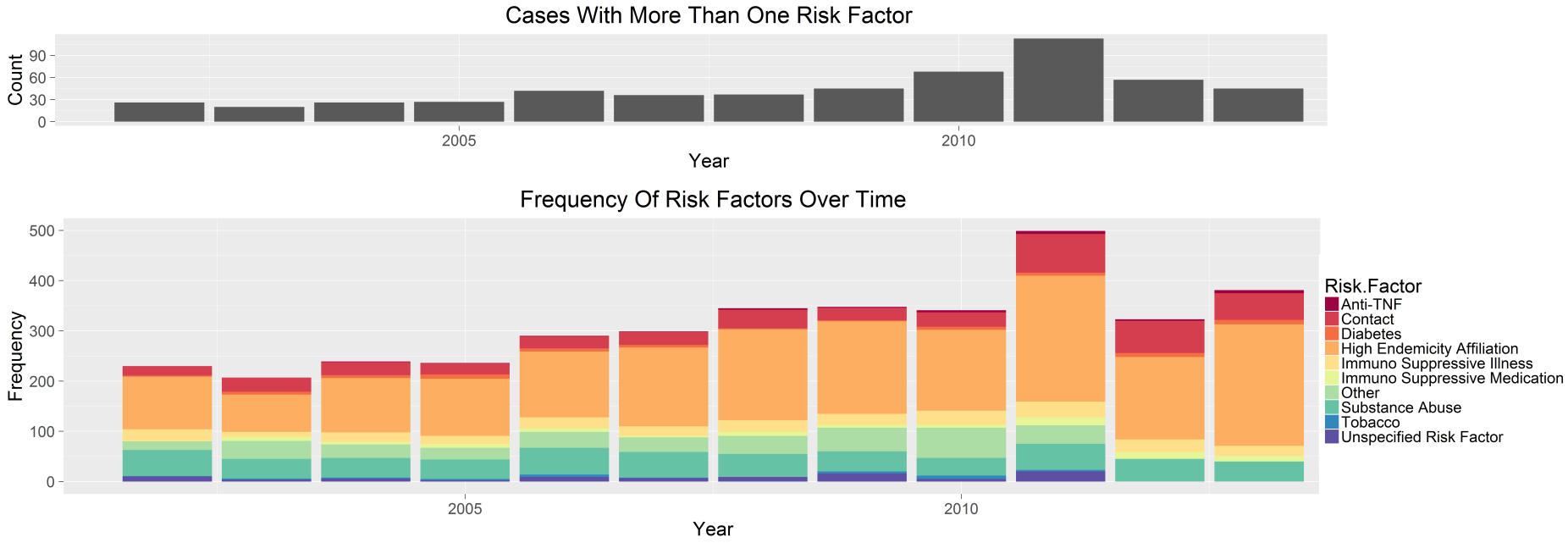

Figure 4.8: Top: Number Of Notifications With More Than One Risk Factor. Bottom: Frequency Of Risk Factors Over Time

4.3.3 Time Series Analysis

Time series analysis comprises methods for analysing data dependent on time in order to extract meaningful statistics from the data. Time series analysis was implemented to investigate seasonality within the TB data, as seasonality is an established attribute of some infectious diseases [108].

Quarterly, monthly, and weekly notifications were evaluated using simple moving average plots and boxplots, along with using an autocorrelation plot [109] to investigate the significance of the seasonality, should it exist. Appendix figure A.1 displays various moving averages being applied to monthly data.

Upon visual inspection of the run plot and moving averages (Appendix figure A.1) there appears to be evidence to suggest a seasonal trend within the data. Following seasonal detection methods [110] multiple box plots were used along with autocorrelation plots generated to check for seasonality. In addition to figures 4.9 and 4.10, weekly boxplots were also generated and can be seen in Appendix figure A.2.

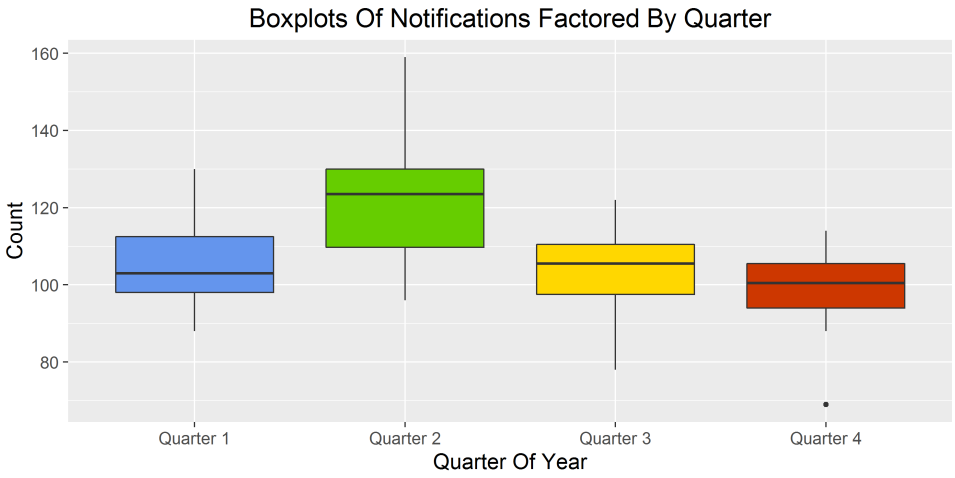

Figure 4.9: Boxplot Of Notifications Factored By Quarter

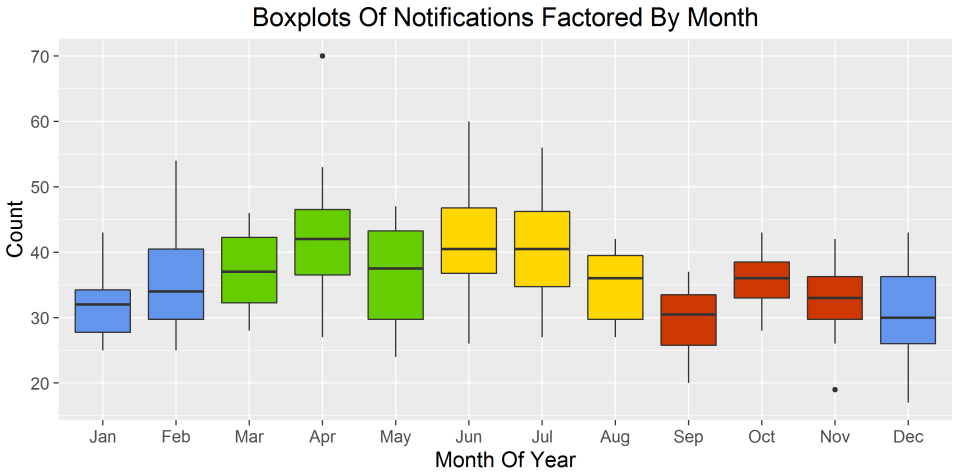

Figure 4.10: Boxplots Of Monthly Notifications Factored By Month Of Year

Autocorrelation, sometimes referred to as "lagged correlation", is a mathematical representation of the degree of similarity between a given time series and a lagged version of itself over successive time intervals. A 95% confidence band was calculated using the formula $\frac{Z_{1-\alpha/2}}{\sqrt{N}}$ for a measure of standard error. This establishes whether the data contains a significant dependency on past values (seasonality) or whether the data are a realisation of a white noise process. The margin or error calculated for quarterly data was found to be ± 0.28 and for monthly data the margin of error was ± 0.16. Figures 4.10 and 4.11 show the correlation of the series with itself, lagged by x time units.

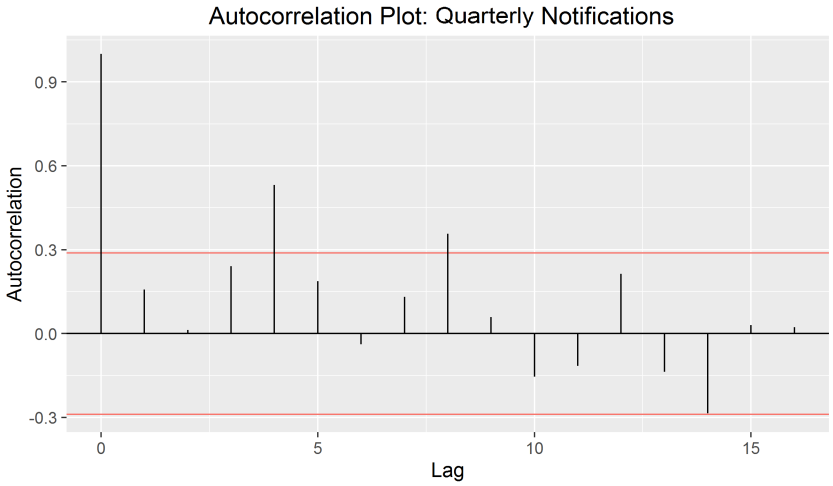

Figure 4.11: Autocorrelation Plot Of Quarterly Notifications with a 95% CI (±0.28)

.

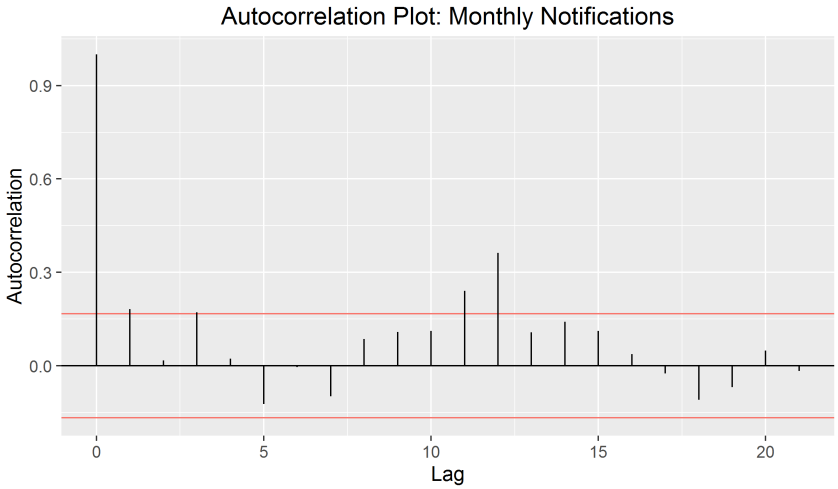

Figure 4.12: Autocorrelation Plot Of Monthly Notifications with a 95% CI (±0.16)

.

Within the quarterly autocorrelation plot there was a significant trend calculated for the fourth and eighth period, implying an annual seasonal cycle. Within the monthly autocorrelation plot there was a significant trend for the 11^{th} and 12^{th} period, again implying

an annual seasonal cycle. Furthermore, the monthly autocorrelation plot gives evidence to suggest the data are sinusoidal, which suggests continuity of the data going from a high seasonal period into a low seasonal period, and visa versa.

Further seasonal-trend decomposition for the time period can be found in the Appendix tables A.21/A.22, the decomposition procedure used followed from a loess smoother [111].

Ranking months in order of average notification we have: April (Mean=43.17, Std.Dev=10.99), June (Mean=41.92, Std.Dev=9.54), July (Mean=40.08, Std.Dev=8.59), March (Mean=37.08, Std.Dev=6.23), May (Mean=36.17, Std.Dev=8.23), February (Mean=36.08, Std.Dev=8.46), October (Mean=35.92, Std.Dev=4.10), August (Mean=34.83, Std.Dev=5.77), November (Mean=32.25, Std.Dev=6.17), January (Mean=32.08, Std.Dev=5.37), December (Mean=30.42, Std.Dev=7.17), and September (Mean=29.25, Std.Dev=5.75). Further descriptive statistics along with the count and percentage of notifications factored by month and quarter can be found in the Appendix table A.19/A.20. Due to a significant seasonality detected within the data, multivariate analysis will be carried out to examine whether seasonality affects the population uniformly.

Descriptive Multivariate Analysis On Seasonality

The most frequent six-month period for notifications typically run from February through to July, and the least frequent from August to January. The top six months on average had 20.4% more notifications than the lower six. This is potentially clinically relevant as it can inform planning for control and treatment strategies in the future. Hospitals and other clinical practices can now potentially predict when to maximise/minimise resources available. By calculating the percentage increase in notifications for certain demographics (e.g. do the unemployed experience this increase in notifications during seasonally high periods?) we can compare the increase in the categories for each demographic variable to the national increase to evaluate if there exists a significant difference. Table 4.8 shows the increase for some of the demographical data.

Variable	Category	Percentage Increase In A High Seasonal Period	Total Number Of Cases	Chi Square Between Categories	Chi Square Against National
	Nationally	20.4%	5160		
	Male	18.94%	3148		0.073 (p-value=0.787)
Gender	Female	23.10%	2004	0.358 (p-value=0.55)	0.174 (p-value=0.676)
	Employed	21.73%	1643		0.036 (p-value=0.849)
	Unemployed	16.63%	1003		0.212 (p-value=0.645)
	Housewife/ Husband	20.69%	448		0 (p-value>1)
	Retired	10.82%	799		1.188 (p-value =0.276)
	Unknown	35.48%	511		NA
	Student	16.74%	492		0.107 (p-value=0.744)
Employment	Other	37.38%	254	1.28 (p-value=.865)	NA
	Home	20.41%	4395		0 (p-value>1)
	Hostel	-1.43%	139		1.359 (p-value=0.244)
	B&B/Hotel	150.00%	21		2.382 (p-value=0.123)
	Homeless	-15.79%	35		1.115 (p-value=0.291)
	Prison	300.00%	30		7.75 (p-value=0.005)*
	Institution	45.24%	103		0.86 (p-value=0.354)
	Other	-3.70%	106		NA
Current Living	Unknown	20.55%	322	13.55 (p-value=0.019)*	NA
	Ireland	21.07%	3126		0.015 (p-value=0.909)
Birth Country	Foreign-Born	21.50%	1947	0.004 (p-value=0.987)	0.028 (p-value=0.868)
	White	20.36%	3334		0 (p-value>1)
	South Asian	25.38%	586		0.213 (p-value=0.645)
	Black	16.54%	550		0.132 (p-value=0.716)
	Unknown	11.25%	338		NA
	East/South East Asian	26.47%	231		0.131 (p-value=0.717)
	Other	23.68%	85		NA
Race/Ethnicity	Irish traveller	70.00%	27	1.246 (p-value=0.87)	0.753 (p-value=0.386)
Refugee	Refugees	26.45%	351		0.194 (p-value=0.659)

Table 4.8: Table Highlighting Percentage Increase In Cases During A Seasonally High Period For Each Demographic Variable

A similar density of age is observed when comparing densities of high seasonal periods to that of a low seasonal periods ($Mean_{[High]} = 42.14$, $Mean_{[Low]} = 43.07$,$SD_{[High]} = 20.55$, $SD_{[Low]]} = 20.52$, $Skewness_{[High]} = 0.433$, $Skewness_{[Low]} = 0.409$). A two sample independent mean t-test resulted in non-significant results when testing means between high and low period (*t-value*=-1.618, *p-value*=0.106). With regards diagnosis type, Pulmonary and Extrapulmonary TB saw a roughly similar increase in seasonally high periods, increasing 19.13% and 19.56% respectively. For Pulmonary & Extrapulmonary TB cases 172 were recorded in a seasonally low period, and 234 were recorded in a seasonally high period, accounting for a 36.05% increase. However, the difference between disease type groups was non-significant when tested ($\chi^2 = 1.59, p - value = 0.452$). For date of death there were 25 deaths recorded within a seasonally low period and 47 deaths recorded within a seasonally high period, which accounts for a 61.1% increase in death during seasonally high periods.

All demographic variables, excluding Current Living, did not show significant differences in the variable categories using a Chi-square test statistic. There was one category within Current living; Prison, that displayed a significant difference in notifications when compared to the national increase. The sample size for this population was 30. This increase in cases was investigated and it was discovered an outbreak occurred in 2011 of which 12 individuals were infected [112], which may have skewed the underlying results of the test statistic. In conclusion, seasonality was found to be a significant factor within the data. Following the multivariate analysis, there is a lack of significant differences between demographics used in the analysis. Consequently, the conclusion will be drawn that seasonality affects the total population uniformly.

4.3.4 Foreign-Born Incidence

In this section further analysis is conducted on the variable Birth Country. Denominator data was acquired for some populations living in Ireland, specifically the size of certain foreign-born populations in Ireland. Table 4.10 in a later section lists these countries. Calculation of incidence and descriptive multivariate analysis will now be presented.

Foreign-Born Descriptive Statistics For Notifications and Incidence

With respect to average annual notifications, the top 10 contributing countries were: India (25 cases annually), Pakistan (19 cases annually), Nigeria (12 annually), Philippines (11 annually), the UK (7 annually), Somalia (7 annually), Romania (6 annually), South Africa (6 annually), Poland (5 annually), Congo (4 annually), and Zimbabwe (4 annually). Out of the top 20 average annual contributors, 13 originate from a high burden country (HBC), described in §2.5.1. On average, the top 10 foreign-born contributors account for 62% of all foreign-born notifications, the top 5 on average contributors accounted for 45% of all foreign-born notifications. Out of the top 10 contributors the average incidence per 100,000 within Ireland over the review period was 140.7 for India, 224.41 for Pakistan, 33.08 for Nigeria, 121.83 for the Philippines, 2.02 for the UK, 50.47 for Romania, 108.02 for South Africa, 6.6 for Poland, 167.48 for the Congo, and 33.18 for Zimbabwe. Population data was unavailable for the Somalian population living in Ireland.

Distributional Properties Of Foreign-Born Incidence

Within a US study [113] it was found that persons who have migrated from areas of the world with high TB rates exhibit notification rates that approach those of their regions of origin for a number of years after arrival. This study now examines this claim with respect to Irish data. Denote $N_{[C,t]}$ the national incidence rate of country C at time t, and denote $I_{[C,t]}$ the incidence rate within Ireland for the population with birth country C at time t (e.g. $N_{[India,2002]}$ denotes India's TB incidence rate in 2002, and $I_{[India,2002]}$ denotes the incidence rate of TB for the Indian population living within Ireland). The expression $N_{[C,t]} - I_{[C,t]}$ is the discrepancy between what Irish populations experience in incidence and what the original birth country experiences in incidence. The underlying distribution of $N_{[C,t]} - I_{[C,t]}$ and the distributional statistics can be seen in figure 4.13 and 4.14 and tables 4.9 and 4.10. The aim of the following analysis is to highlight the difference, if any, between a foreign-born individuals birth country incidence and that of the individuals coming from that country living in Ireland. The expression $N_{[C,t]} - I_{[C,t]}$ will be close to zero if there is little to no difference between the incidence of the birth country and that of the individuals living in Ireland. This will therefore provide evidence of a possibile

relationship between the type of foreign-born populations living in Ireland and Ireland's overall national incidence.

For table 4.9, outliers were considered any data point that was more than 1.5 interquartile ranges below the first quartile or above the third quartile.

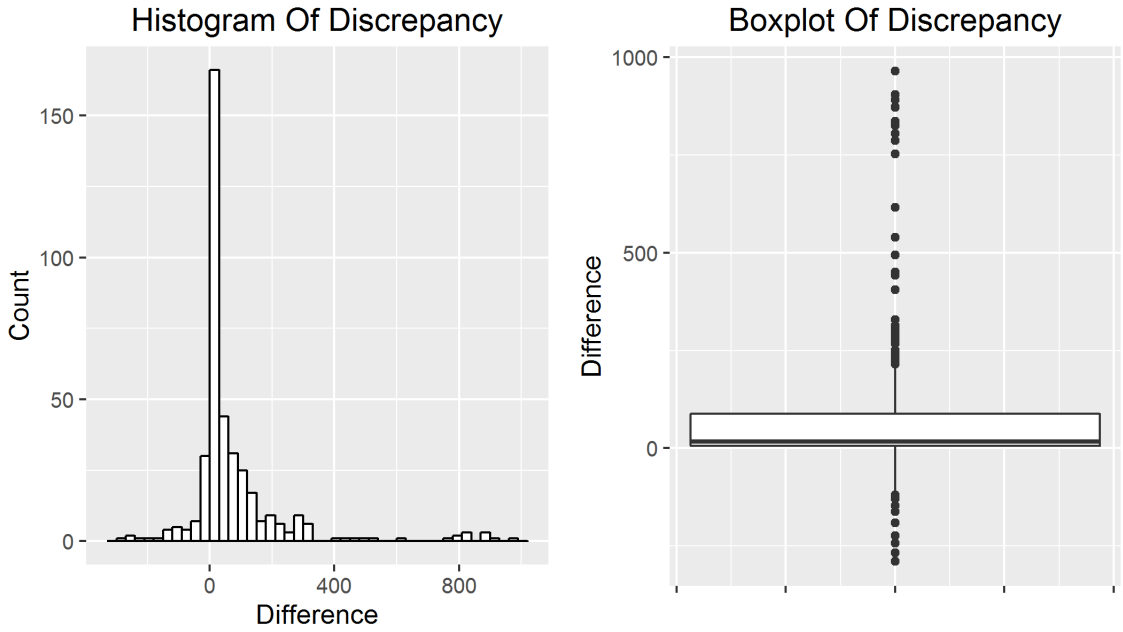

Figure 4.13: Histogram and Boxplot of $N_{[C,t]} - I_{[C,t]}$, the discrepancy for all countries over 2002 through to 2013.

Statistics	All Data	Outliers Removed
Min	-289.7	-114.8
Q1	6.1	6
Median	17.25	12
Q3	88.5	60.78
Max	964	207.6
Mean	73.39	35.19
Geometric Mean	29.26	24.97
S.Dev	166.8	55.44
Skewness	3.06	0.89
Kurtosis (Normal=2.96)	11.73	1.37

Table 4.9: Statistics Of $N_{[C,t]} - I_{[C,t]}$, The Discrepancy For All Countries Over 2002 Through To 2013.

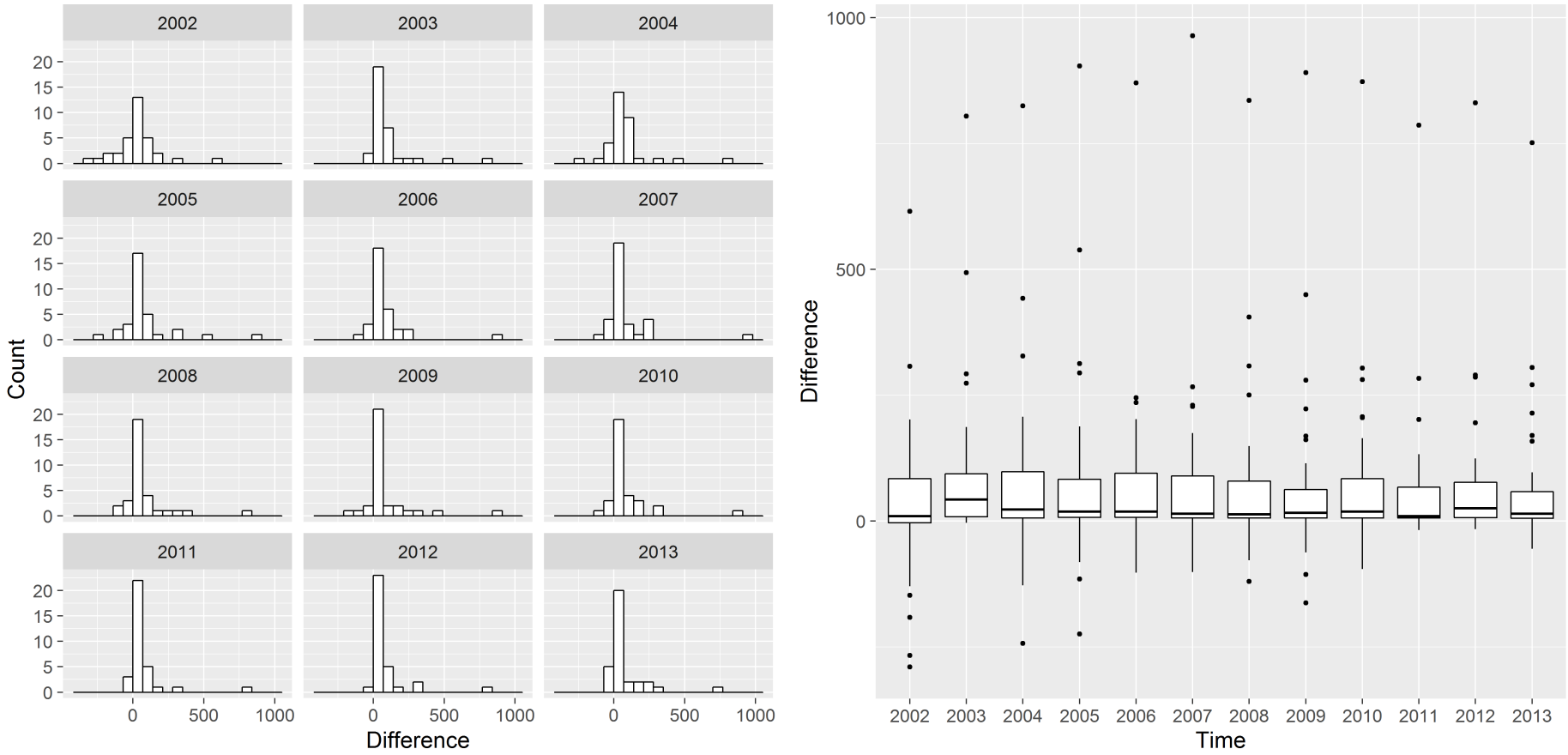

Figure 4.14: Histograms And Boxplots Of $N_{[C,t]} - I_{[C,t]}$ Factored By Year, Where $N_{[C,t]} - I_{[C,t]}$ Is The Discrepancy Of National Incidence And Irish Incidence.

Statistics	National Minimum Incidence ($\min_t(N_{[C,t]})$)	National Maximum Incidence ($\max_t(N_{[C,t]})$)	Ireland Minimum Incidence ($\min_t(I_{[C,t]})$)	Ireland Maximum Incidence ($\max_t(I_{[C,t]})$)	Mean Yearly Notifications
Australia	5.9	6.5	0	15.76	0.17
Belgium	9.1	13	0	71.99	0.17
Brazil	45	56	0	247.12	1.33
China	70	102	9.07	72.42	3.75
Congo	382	426	90.42	322.48	3.92
Denmark	6.7	9.8	0	136.05	0.08
Finland	5.8	9.5	0	112.49	0.17
France	8.9	12	0	25.32	0.50
Germany	5.8	10	0	18.53	0.50
Greece	5	7	0	296.74	0.25
Hong Kong	76	103	0	177.49	0.75
India	171	215	117.41	362.43	24.50
Italy	6.3	8.1	0	27.52	0.33
Latvia	50	104	0	44.56	1.92
Lithuania	65	97	0	21.34	3.67
Malaysia	75	97	0	116.67	1.08
Moldova	159	176	0	98.72	0.75
Netherlands	6	9.3	0	21.95	0.08
New Zealand	7.4	11	0	42.83	0.08
Nigeria	336	343	33.08	191.16	12.42
Pakistan	275	276	138.96	541.84	18.50
Philippines	292	355	14.98	169.80	11.00
Poland	21	30	0	46.79	4.58
Portugal	25	46	0	175.13	0.58
Romania	87	168	24.85	93.07	6.42
Russia	89	138	0	39.36	0.67
South Africa	746	977	13.01	141.17	5.92
Spain	13	21	0	29.30	0.67
Ukraine	96	127	0	68.26	0.50
UK	13	15	1.45	4.16	6.83
USA	3.4	6.1	0	9.53	0.75
Zimbabwe	304	617	0	703.73	3.83

Table 4.10: Minimum And Maximum Incidence Rates For Each Country ($N_{[C,t]}$) And For The Individuals Born Of That Country Who Live In Ireland ($I_{[C,t]}$).

With respect to table 4.9 and figure 4.13, the distributions are centered close to zero (median incidence 17.25, or 12 with outliers removed) which indicates a possible relationship between variables $N_{[C,t]}$ and $I_{[C,t]}$. The distribution has positive skewness which indicates a large discrepency between individuals coming from high-incidence countries and the incidence they experience in Ireland. A large standard deviation was calculated (166.8 or 55.4 with outliers removed), which indicates that there is considerable variability between birth country incidence and the incidence of individuals living in Ireland born of that country. This demonstrates that, while the distribution is cenetered close to zero, the incidence of an individuals birth country may not have a strong relationship with the incidence those individuals observe in Ireland. Figure 4.14 indicates the distribution of $N_{[C,t]} - I_{[C,t]}$ does not change over time, however, a more formal methodology would need to be carried out to support this claim (a linear regression, for instance).

To test the strength of the relationship between variables $N_{[C,t]}$ and $I_{[C,t]}$ a correlation coefficient was calculated. A significant correlation was calculated between the $N_{[C,t]}$ and $I_{[C,t]}$ [$R^2 = 0.224$, Adjusted R^2= 0.22, n = 167, $p \leq 0.001$].

To help illustrate the relationship between $N_{[C,t]}$ and $I_{[C,t]}$ both variables had their countries placed in categories based on their latest data. The categorise follow: Very High if $N_{[C,2013]} \geq 100$, High if $50 \leq N_{[C,2013]} \leq 100$, Medium if $10 \leq N_{[C,2013]} \leq 50$, and Low if $N_{[C,2013]} \leq 10$. Table 4.11 displays the average incidence for each category.

Countries Categorised By Incidence Rate	Very High Incidence Countries (N=8)	High Incidence Countries (N=7)	Medium Incidence Countries (N=5)	Low Incidence Countries (N=12)
Average Birth Country Incidence	347.63	82.86	23.40	6.61
Average Incidence Experienced in Ireland	151.03	30.84	20.10	11.79

Table 4.11: A Comparison Of $N_{[C,t]}$ And $I_{[C,t]}$ When Countries Are Categorized By Incidence

Table 4.11 suggest an increase in individuals originating for countries with large incidence will increase the incidence of the foreign-born population. When observed directly, the relationship is relatively weak. However, when ranges and categorizations are observed, there appears to be a trend between both variables.

Descriptive Multivariate Analysis On Birthplace

With respect to age, a two sample independent means t-test resulted in a significant difference when testing means between foreign-born and native-born groups (*t-value*=29.09, $p - value \leq 0.001$). The mean age of native-born populations was 48.49 and of foreign-born 32.59. The density of age for both populations for 2002 and 2013 can be seen in figure 4.15. With respect to sex, employment status, disease type, current living, and death, table 4.13 highlights the differences between native-born and foreign-born notifications. With regards Race/Ethnicity the proportion of native-born and foreign-born cases occurring in each category can be seen in table 4.12. The very small percentages observed for native-born Asian and Black populations, would suggest the high incidences observed within these groups may be a factor of birthplace, rather than concluding differences in biological mechanisms within the ethnic groups themselves. With regards to diagnosis type, an increased rate of Extrapulmonary TB is seen within the foreign-born population; native-born populations averaging 22.36% of cases Extrapulmonary, foreign-born averaging 40.2%. A similar figure is observed within the UK which, in 2014, had over 70% of notifications foreign-born and reported the proportion of cases with Extrapulmonary TB as 47.9% of total notifications [103]

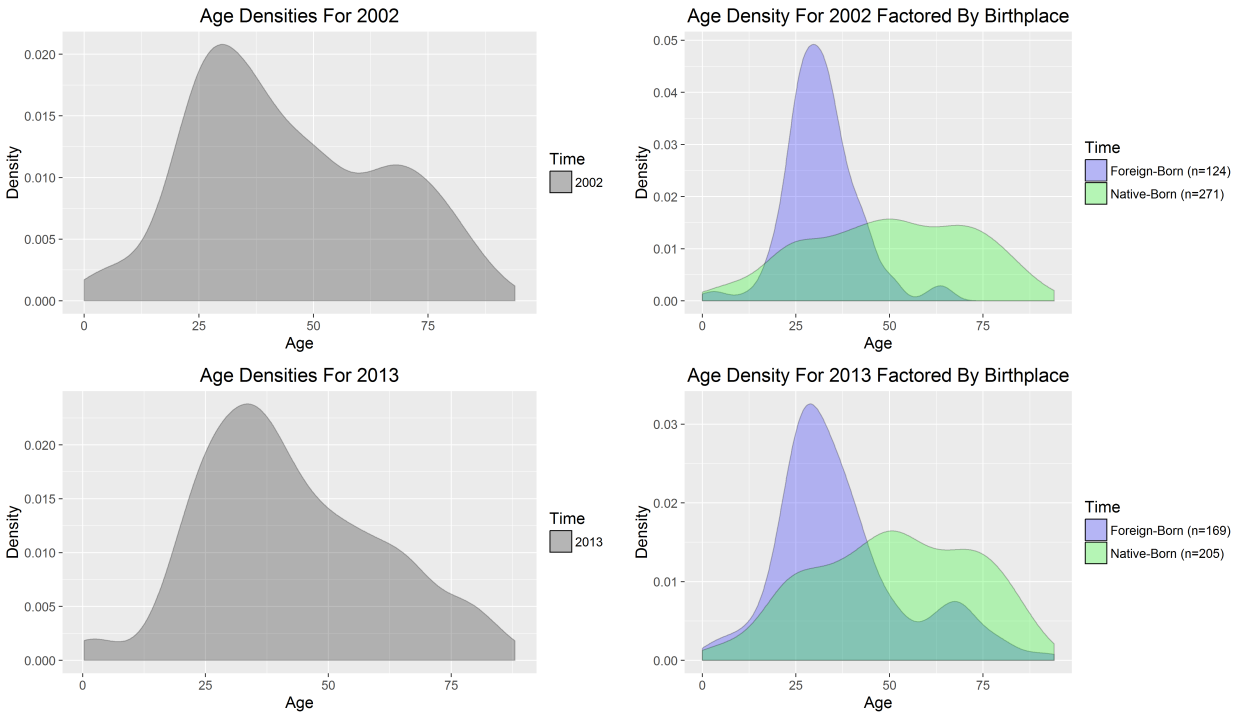

Figure 4.15: Density Plot Of Age Factored By Birthplace For The Year 2002 And 2013

	White	South Asian descent	Black	Irish traveller	East / South East Asian descent	Other	Unknown
Native-Born (n=3078)	95.22%	0.16%	0.65%	0.78%	0.10%	0.13%	2.96%
Foreign-Born (n=1908)	21.07%	30.40%	27.67%	0.16%	11.95%	4.25%	4.51%

Table 4.12: Distribution Of Birthplace Factored By Race/Ethnicity

Variable	Category	Irish-Born Notifications	Foreign-Born Notifications	Chi Square Between Groups
	Total	3134	1948	
Gender	Male	1966 (62.73%)	1130 (58.19%)	
	Female	1168 (37.26%)	812 (41.81%)	10.41 (p-value = 0.001)
Employment	Employed	855 (28.56%)	784 (40.27%)	
	Unemployed	579 (19.34%)	424 (21.78%)	
	Housewife/ Husband	256 (8.55%)	193 (9.91%)	
	Retired	765 (25.55%)	29 (1.49%)	
	Unknown	159 (5.31%)	156 (8.01%)	
	Student	215 (7.18%)	273 (14.02%)	
	Other	165 (5.51%)	88 (4.52%)	537.45 (p-value <0.001)
Current Living	Home	2751 (90.82%)	1635 (8402%)	
	Hostel	38 (1.25%)	100 (5.14%)	
	B&B/Hotel	9 (0.3%)	10 (0.51%)	
	Homeless	24 (0.79%)	12 (0.62%)	
	Prison	12 (0.43%)	16 (0.82%)	
	Institution	88 (2.91%)	13 (0.67%)	
	Other	56 (1.85%)	50 (2.57%)	
	Unknown	50 (1.65%)	110 (5.65%)	104.09 (p-value<0.001)
Diagnosis Type	Pulmonary	2220 (70.9%)	961 (49.33%)	
	Extrapulmonary	700 (22.36%)	783 (40.2%)	
	Pulmonary + Extrapulmonary	203 (6.48%)	202 (10.37%)	
	Unknown	8 (0.26%)	2 (0.1%)	242.74 (p-value<0.001)
Death Due To TB	Yes	70	15	

Table 4.13: Table Detailing Demographic Differences In Notifications Between Foreign-Born and Native-Born Groups

The Chi-square tests in table 4.13 become a more significant result under the assumption the underlying populations have similar proportions for each variable. Internationally, various socio-economic differences have been noted in some foreign-born populations[114]. Hence the significance of results in table 4.12 may be due to the composition of the underlying populations. An improved comparison would be to compare incidence, rather than count data, however, population data was not available to establish incidence rates. What can be concluded is that there is a difference in the underlying populations, or there is a significant difference between birthplace for each of the variables.

With respect to the type of strain, foreign-born MDR-TB has accounted for 27 of the 33 cases (81.82%). There was only one XDR-TB case that occurred in 2005, the birthplace of that individual was Lithuanian. The WHO have categorised Lithuania as a High MDR TB burden country [115] reporting 310 (13.48% of total cases) MDR-TB cases in 2010. With respect to MDR-TB, Lithuanian born individuals accounted for 7 of the 33 cases (21.21%), Irish born cases accounted for 6 of the 33 (18.18%), and Indian born 3 of the 33 (9.09%). Mongolia, Zimbabwe, South Africa, Nigeria, and Latvia each separately accounted for 2 of the 33 cases (6.06%), and Georgia, Azerbaijan, China, Romania, Ukraine, Somalia, and Russia all separately accounted for 1 of the 33 cases (3.03%).

In conclusion very different demographics and disease types were observed within the foreign-born population.

4.4 Conclusion

The exploratory analysis of the tuberculosis cases conducted in this chapter aimed at estimating the distributional properties of the incidence cases. This was completed to lend insight of a suitable deterministic epidemiological model. First, the methodology was presented including data acquisition methods from the HPSC. The completeness of that data was also reviewed. The analysis began by observing variables individually over time through tables, graphs, and descriptive statistics. The variables: Ethnic Group, Birthplace, Employment status, and Refugee Status all appeared to contain categories within which

notably large incidence rates were identified. This indicates a larger burden on the respective populations. The variables Ethnic Group and Refugee Status are intuitively dependent on Birthplace; hence, further analysis on birthplace was carried out in §4.3.4.

A descriptive time series analysis was then carried out revealing a seasonal trend within notifications, deemed to be statistically significant. The seasonality was observed to peak within a six month time frame, and appeared to be consistent with a sinusoid model. For the majority of cases, seasonality affected the categories within demographic variables uniformly. The exception being individuals currently living within a prison. Due to the small sample size of this population, it will be assumed that seasonality affects the national population uniformly. Further analysis of foreign-born cases indicated a relationship between national incidence of birthplace and the incidence levels experienced in Ireland. Significantly different demographic characteristics were also noted between native and foreign-born populations.

With respect to epidemic modelling these results will be translated into the model. As seasonality is being deemed to effect the population uniformly, a seasonal model will be considered which models the population as a whole. In addition, due to the impact foreign-born notifications are having a foreign-born/native-born model will be considered. As significant differences were seen between native and foreign-born populations the model will consider different parameter sets and compartments when modelling each population. This ensures the population as a whole is modelled accurately.

The results of this chapter informed the literature review. Chapters 5 and 6 will now model the population given these results.

Chapter 5

A Mathematical Model For Seasonal TB

5.1 Introduction

This chapter aims to contribute to the on-going research into seasonality in tuberculosis notifications by conducting a qualitative analysis and estimating the epidemiological parameters of a seasonal model for Irish data.

It has been established that the incidence of many respiratory infections shows seasonal variation, and it is not as well documented for TB [116]. The exact mechanism underlying the fluctuation of tuberculosis rates at any given time of the year is not clear. Researchers have suggested the environmental and social factors such as temperature, humidity, sunlight, as well as crowding and interpersonal contact are a source of TB seasonality, particularly in winter time [117].

The following sections present and refine the model acquired from the literature review (§3.5). Once established, a qualitative analysis of the model is conducted followed by an assessment on the model's epidemiological parameters and initial conditions. Statistical inference techniques that incorporate uncertainty are implemented to estimate a subsection of the model parameters. Finally the model is simulated and extrapolated and the basic reproductive number for Ireland is calculated.

5.2 A Seasonal TB Model

The model obtained from the literature review is represented by the ordinary differential equation (ODE) system displayed below (equations 5.1-5.5). The underlying compartments follow an SEIR model (Susceptible, Exposed/Latent, Infectious, and Recovered) discussed in §3.4.2.

$$\frac{dS}{dt} = \Lambda - \beta(t)\frac{SI}{N} - \mu S \tag{5.1}$$

$$\frac{dE}{dt} = (1-q)\beta(t)\frac{SI}{N} - (\mu + k(t))E \tag{5.2}$$

$$\frac{dI}{dt} = q\beta(t)\frac{SI}{N} + k(t)E - (\mu + d + r)I \tag{5.3}$$

$$\frac{dR}{dt} = rI - \mu R \tag{5.4}$$

$$N = S + E + I + R \tag{5.5}$$

Where

- $S(t)$ is the number of people in the population at time t susceptible to TB.
- $E(t)$ is the number of people in the population at time t currently infected with the latent form of TB.
- $I(t)$ is the number of people in the population at time t actively infected and spreading the disease within the population.
- $R(t)$ is the number of people in the population at time t who have recovered, are treated, or have immunity to TB.
- Λ is the recruitment rate (called a rate in literature, however is a number), the number of individuals entering into the relative compartment at each time step.
- $\beta(t)$ is the transmission rate, the rate at which susceptibles are initiated into either of the infectious compartments. This is a time dependent variable meaning the transmission rate will not be constant as time progresses.

- μ is the rate of population change, the combined effect of births and deaths.
- q is the proportion of new infections that develop as a result of "fast" progression; cases that proceed directly to the infectious compartment from the susceptible compartment.
- *k(t)* is the progression rate, the rate at which individuals go from being latent to infectious. This is a time dependent variable meaning the progression rate will not be constant as time progresses.
- d is the disease-induced death rate. The rate at which individuals die because of the disease.
- r is the removal or recovery rate. The rate at which individuals are recovered or treated for TB.

The model makes the following assumptions

- The entire population recruits new individuals at a rate Λ into the susceptible population. Individuals entering into the population become susceptible automatically.
- The rate of infection of susceptibles into the infective or exposed compartment, β, is proportional to the number of current infectives, hence the infection of susceptibles is proportional to the number of individuals currently infected and the number of individuals who are susceptible. This is referred to as the mass action principle discussed in §3.
- After users recover, or have entered into the recovered compartment, they are no longer susceptible to the disease.
- The system presumes fast and slow progression of TB. The initially exposed individuals have a higher risk of developing active TB. With time passing, those individuals still face the possibility of progressing to infectious TB, but the rate of progression slows down. In other words, the likelihood of becoming an active infectious case decreases with the age of the infection. A proportion $q\frac{\beta SI}{N}$ gives rise to immediate

active cases (fast progression), while the rest $(1-q)\frac{\beta SI}{N}$ gives rise to latent-TB cases with a low risk of progressing to active. TB (slow progression).

- The transmission rate, $\beta(t)$, and progression rate, $k(t)$, are both time dependent functions. The rate at which individuals progress through compartments will change depending on the month of year.
- $N(t) = S(t) + E(t) + I(t) + R(t)$ is the total population size at time t. The rate of change can be expressed as $\frac{dN}{dt} = \Lambda - \mu N(t) - dI(t)$.

The original model makes the assumption that recruitment occurs directly into the susceptible compartment. That is, all individuals born or who are immigrating to the country will automatically be considered susceptible to TB. In Ireland, universal vaccination occurs from birth [124]. To assume vaccinated individuals are susceptible is not a rational assumption for the model. Hence an alternative model can be constructed with a change in placement on the recruitment parameter. This will account of new individuals entering the system who are also vaccinated. The alternative model follows:

$$\frac{dS}{dt} = w\Lambda - \beta(t)\frac{SI}{N} - \mu S \tag{5.6}$$

$$\frac{dE}{dt} = (1-q)\beta(t)\frac{SI}{N} - (\mu + k(t))E \tag{5.7}$$

$$\frac{dI}{dt} = q\beta(t)\frac{SI}{N} + k(t)E - (\mu + d + r)I \tag{5.8}$$

$$\frac{dR}{dt} = (1-w)\Lambda + rI - \mu R \tag{5.9}$$

$$N = S + E + I + R \tag{5.10}$$

Where w is the proportion of births or immigrants that enter into the susceptible compartment immediately, and where $(1-w)$ is the proportion of births or immigrants entering into the recovered compartment immediately. This model can now consider a vaccinated population. If individuals are vaccinated, a proportion of them will have immunity. They can then immediately enter into the recovered compartment. A schematic can be seen in figure 5.1.

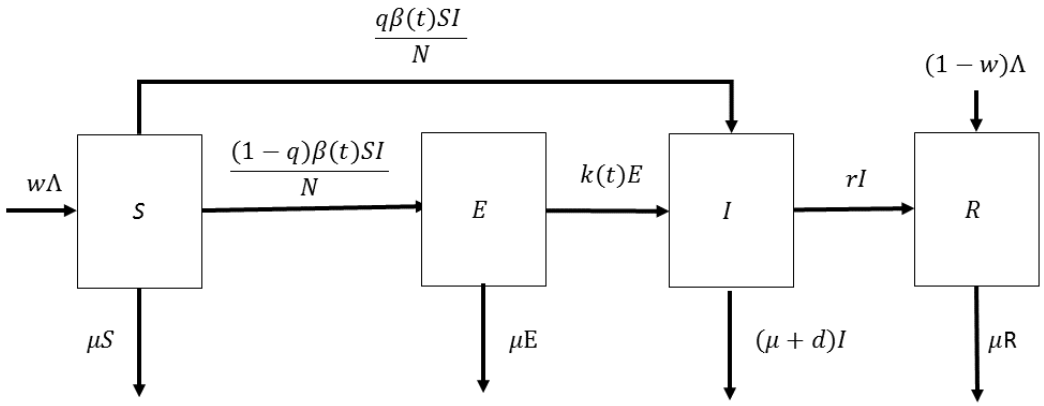

Figure 5.1: Schematic Of altered model

The following section conducts qualitative analysis on the altered model.

5.3 Qualitative Analysis

5.3.1 Equilibrium States

An equilibrium point is a constant solution to the ODE system. With respect to the mathematical modelling of epidemics there exist two main equilibrium points: the *disease-free equilibrium* and the *endemic equilibrium*.

Definition 5.1. *A **disease-free equilibrium** is defined as the point at which no disease is present in the population (infectious or exposed) and the change in the number of infec-*

*tions over time is zero. The **endemic equilibrium** is defined as the point which disease is present in the population and the change in the number of infections is zero.*

The equilibrium points of equations 5.6-5.9 are calculated through setting each equation to zero (no change in compartments), and solving the system of equations that result. The disease-free equilibrium is obtained by calculating the above and setting $I(t) = 0$, and $E(t) = 0$. Denote $D_0 = (S_{DF_0}, E_{DF_0}, I_{DF_0}, R_{DF_0})$ the solution to the disease-free equilibrium, solving system 5.6-5.9 results in

$$D_0 = \left(S_{DF_0} = \frac{w\Lambda}{\mu}, E_{DF_0} = 0, I_{DF_0} = 0, R_{DF_0} = \frac{(1-w)\Lambda}{\mu}\right).$$

This solution can usually be acquired analytically by hand. The endemic equilibrium is the equilibrium for which the change in each compartment is zero, however, the number of infectious/exposed is greater than zero. Depending on the system, the endemic equilibrium point can be a very long algebraic expression with hundreds of terms and software can often be required to calculate it. For equations 5.6-5.9 the equilibrium existed and was calculated, however, was hundreds of terms in length. Due to it not having utility with respect to this study, it will be omitted. The following section presents methods to calculate the basic reproductive number.

5.3.2 Basic Reproductive Number

The basic reproductive number was previously introduced in §3. It is a defining value of a deterministic model. If greater than one, the model will experience an epidemic. If less than one, no epidemic will occur and the disease will die off. What follows is the calculation of the basic reproductive number, R_0, using the next generation method [133]. With respect to the disease free equilibrium, the equations 5.7 and 5.8 of the linearised system become

$$\frac{dE}{dt} = (1-q)\beta(t)\frac{\frac{w\Lambda}{\mu}I}{N} - (\mu + k(t))E,$$
$$\frac{dI}{dt} = q\beta(t)\frac{\frac{w\Lambda}{\mu}I}{N} + k(t)E - (\mu + d + r)I.$$

At the disease-free equilibrium the population size, N, is the sum of the disease-free equation for each compartment. Hence we can replace $N = \frac{w\Lambda}{\mu} + \frac{(1-w)\Lambda}{\mu} = \frac{\Lambda}{\mu}$ into the above system to get

$$\frac{dE}{dt} = (1-q)\beta(t)wI - (\mu + k(t))E, \tag{5.11}$$
$$\frac{dI}{dt} = q\beta(t)wI + k(t)E - (\mu + d + r)I. \tag{5.12}$$

The next generation method obtains R_0 through calculating $\rho(FV^{-1})$, where ρ is the spectral radius of the matrix FV^{-1}, defined as the eigenvalue of the matrix with largest magnitude. The above equations (5.11 and 5.12) are used to construct the matrices F and V. The entries of F and V are determined by the effect of an infection event. If an infection event causes a gain to the compartment its derivative makes up F. If an infection event causes a loss to a compartment its derivative makes up V. The matrices F and V follow for system of equations 5.6 - 5.9.

$$F(t) = \begin{pmatrix} 0 & (1-q)\beta(t)w \\ 0 & q\beta(t)w \end{pmatrix},$$
$$V(t) = \begin{pmatrix} \mu + k(t) & 0 \\ -k(t) & \mu + d + r \end{pmatrix}.$$

It is not theoretically possible to compute the basic reproductive number when there are periodic functions of time as parameters [89], hence to calculate R_0 the assumption will be made that the parameters are constant. In the case of $\beta(t) = \beta$ and $k(t) = k$, the matrices

$F(t)$ and $V(t)$ become F and V, respectively. The matrix FV^{-1} was calculated to be

$$\begin{aligned} FV^{-1} &= \begin{pmatrix} 0 & (1-q)\beta w \\ 0 & q\beta w \end{pmatrix} \begin{pmatrix} \mu+k & 0 \\ -k & \mu+d+r \end{pmatrix}^{-1} \\ &= \begin{pmatrix} 0 & (1-q)\beta w \\ 0 & q\beta w \end{pmatrix} \begin{pmatrix} \frac{1}{\mu+k} & 0 \\ \frac{k}{(\mu+d+r)(k+\mu)} & \frac{1}{\mu+d+r} \end{pmatrix} \\ &= \begin{pmatrix} \frac{\beta kw(1-q)}{(k+\mu)(d+r+\mu)} & \frac{\beta w(1-q)}{d+r+\mu} \\ \frac{\beta kqw}{(k+\mu)(d+r+u)} & \frac{\beta qw}{d+r+\mu} \end{pmatrix}. \end{aligned}$$

The above matrix has eigenvalues

$$\lambda_1 = \frac{\beta w(k+q\mu)}{(\mu+d+r)(\mu+k)}, \lambda_2 = 0$$

Due to the spectral radius, or largest eigenvalue, of FV^{-1} being defined as R_0 through the next generation method and due to all parameters being positive constants, we define $R_0 = \lambda_1$ or

$$R_0 = \frac{\beta w(k+q\mu)}{(\mu+d+r)(\mu+k)}. \tag{5.13}$$

To overcome the omission of accepting constants for parameters $\beta(t)$ and $k(t)$ an average will be taken of both functions, and the average values will help acquire an estimate of R_0.

The elements of the matrix V^{-1} have epidemiological interpretations in themselves. In, for example, Diekmann el al.[195] it is shown that the element $V^{-1}_{i,j}$ is the expected time that an individual who is presently within state j will spend in state i. For the above V^{-1} was calculated as

$$\begin{pmatrix} \frac{1}{\mu+k} & 0 \\ \frac{k}{(\mu+d+r)(k+\mu)} & \frac{1}{\mu+d+r} \end{pmatrix}$$

Individuals who are exposed will spend on average $\frac{1}{\mu+k}$ units of time being exposed. The same individuals will spend on average $\frac{k}{(\mu+d+r)(k+\mu)}$ units of time being infected. The

individuals who are infected will stay infected for $\frac{1}{\mu+d+r}$ units of time.

5.4 Parameter Estimation

A total of eight parameters needed to be estimated in the model, along with the total population size and the relative proportions each compartment takes initially. The parameters Λ and μ, the recruitment rate and death rate, respectively, and the population size N were calculated from national surveillance data published by the World Bank [196]. The parameters d; the death rate due to TB, and r; the recovery or removal rate were calculated using the dataset acquired and used in §3. The parameters $\beta(t)$; the transmission rate, and $k(t)$; the progression rate, were estimated using statistical inference methods. The model's initial conditions were estimated using a combination of data and assumptions. The total population and initial infected population were estimated using World Bank data and the national TB dataset, respectively. Estimates of the initially exposed and recovered populations were guided by the literature by method of assumption.

5.4.1 Recruitment And Death Rate Parameters

The recruitment rate parameter is dependent on both births and net migration. Hence the recruitment rate will be defined as the average monthly birth rate plus the average monthly net migration rate. The vital statistics are displayed in the table 5.1.

Year	Population	Births	Migration	Monthly Migration +Births (Recruitment Rate)	Deaths	Monthly Death Rate
2002	3900000	60800	32100	7742	28900	0.00062
2003	3964000	62000	31600	7800	28600	0.00060
2004	4029000	61400	49500	9242	27900	0.00058
2005	4112000	61200	61800	10250	27000	0.00055
2006	4208000	65400	94600	13333	28000	0.00055
2007	4340000	71300	74700	12167	28000	0.00054
2008	4458000	75100	15900	7583	28000	0.00052
2009	4521000	75600	-19600	4667	28000	0.00052
2010	4549000	75100	-25100	4167	28000	0.00051
2011	4571000	74700	-33700	3417	29000	0.00053
2012	4583000	72200	-35200	3083	29000	0.00053
2013	4591000	68900	-23900	3750	30000	0.00054
Average	4318833	68642	18558	7267	28367	0.00055

Table 5.1: Annual Population, Birth, Migration, And Mortality Data for Ireland, 2002 through 2013.

Using the above data, the initial population size will be $N(0) = 3,900,000$. The recruitment rate will be the average monthly recruitment rate, $\Lambda = 7,267$, and the death rate will be the the average monthly death rate, $\mu = 0.00055$.

5.4.2 The Proportion Of Current/New Individuals With Immunity

The BCG vaccine has shown to have 69% effectiveness against TB infection [40]. As exact proportions are not known, this study makes the assumption for modelling purposes that 69% of individuals who have been vaccinated will have immunity. Due to the uncertainty surrounding this parameter it will be given further evaluation in §7, where a sensitivity analysis is conducted and a scenario analysis that evaluates a range of values. The coverage of the BCG vaccine in Ireland is approximately 94% [124]. Hence, for

modelling purposes, the assumption is made that the proportion of the population that can be considered immune is $69\% \times 94\% = 65\%$. The model will assume 65% of the population will be immune or recovered, hence the initial recovered population will be $R(0) = 0.65 \times N(0) = 2,535,000$. The model also makes the assumption that the recruitment rate parameter, Λ, can be divided in this way. As it is assumed that 65% of the current population will be immune, it will also be assumed that 65% of recruited individuals entering into the population (birth plus migration) will have immunity and 35% of the population will not, hence $w = 0.35$. The effects on the model when this rate is altered will be examined in §7.

5.4.3 Death and Recovery Rate

The death rate due to TB, and the recovery rate were calculated using the national TB dataset. The annual count of deaths due to TB can be seen in table 5.2. Death due to TB was defined as individuals who died while infected with TB, and whose death could not be attributed to any other factors other than TB. This excludes all individuals who died with HIV and TB, or any other illness and TB.

TB Cause of Death	Unknown	Yes	No	Total	Yes Proportion
2002	409	0	1	410	0.00000
2003	378	6	22	406	0.01478
2004	413	5	15	433	0.01155
2005	414	11	23	448	0.02455
2006	431	10	22	463	0.02160
2007	444	7	30	481	0.01455
2008	434	9	24	467	0.01927
2009	447	10	22	479	0.02088
2010	398	8	14	420	0.01905
2011	391	10	12	413	0.02421
2012	350	3	6	359	0.00836
2013	370	7	4	381	0.01837
Average	407	7	16	430	0.01643

Table 5.2: Number of Deaths Attributed to TB, 2002 through 2013.

The average annual death rate will be the death rate of the model, $d = 0.01643$. The recovery rate can be defined as the inverse of the average duration of time the population is ill [126]. The true duration of illness is unknown. The dataset contains the following variables: date of onset, and date of diagnosis. The date of onset is an estimate that doctors make based on an consultation with the patient. Table 5.3 below shows descriptive statistics on the dataset for the duration of illness.

	Duration Of Illness (Days)
Mean	107
Median	62.5
Mode	31
Min	1
Max	2793
Skewness	7.13

Table 5.3: Duration of Illness Statistics

Due to the skewness of the data, the median will be used to estimate the expected duration of illness. Hence $r = \frac{1}{62.5} = 0.016$.

5.4.4 Fast Progression, Initial Infected/Exposed/Susceptible Population

Approximately 5% of individuals will become actively infectious in the months directly after coming into contact with another infectious case [92]. This study assumes the same proportion for Ireland, hence the fast progression rate will take the value $q = 0.05$.

The initial infected population was calculated from the dataset and will be the number of notified cases for the month of January 2002, as this is the beginning of the study. The initial value $I(0) = 34$ will be used.

The number of initial exposed or latent individuals is assumed to be proportional to the number of initially infected individuals. The WHO [104] cite that one in ten individuals with latent TB will progress to be an active case. There is an absence of literature on latent TB prevalence and incidence in Ireland and for most countries. Due to the latent phase being asymptomatic, estimating prevalence can be a problem. The only Irish data available is outbreak data published by the HPSC [127] From 2004 to 2013 the total number of outbreak cases that led to either an infectious case or a latent TB case are shown in the table 5.4 below.

Year	Active TB	Latent TB	Ratio
2004	3	0	0.00
2005	8	0	0.00
2006	20	54	2.70
2007	73	160	2.19
2008	45	20	0.44
2009	28	53	1.89
2010	41	60	1.46
2011	42	15	0.36
2012	24	4	0.17
2013	42	174	4.14

Table 5.4: Annual Outbreak and Infection Type Data, 2004 through 2013

The above data is for outbreaks only. In reality the number of latent cases will likely have a much larger ratio. The annual risk of infection identified by Yeh and colleagues [128] was 0.4%. From the proportion of individuals who are not immune (1-0.846) we can assume 0.4% of that population is latent, or $0.004 \times 0.154 = 0.0000616$ which results in an estimated exposed population of $0.0000616 \times 3,900,000 = 240$. Hence $E(0) = 240$. This results in an initial ratio of latent to active cases of 7.07.

We have established the initial conditions of four of the five required subgroups for the system (equations 5.6-5.10) populations. By direct computation we can compute $S(0)$ by using the formula $N(0) = S(0) + E(0) + I(0) + R(0)$. This results in $S(0) = 584,726$.

5.5 Statistical Inference for Transmission Parameters

The transmission parameters, $\beta(t)$ and $k(t)$, are assumed to be periodic functions of time. Within the exploratory data analysis chapter, the seasonality in TB data appeared to be sinusoidal. Hence the transmission parameters will take the form

$$\beta(t) = \beta_0(1 + sin(\frac{2\pi t}{12})), \tag{5.14}$$

and

$$k(t) = k_0(1 + sin(\frac{2\pi t}{12})). \tag{5.15}$$

To estimate parameters β_0 and k_0, two statistical inference methods will be implemented: an approximate Bayesian computation and the Metropolis-Hasting algorithm. The methodology of each approach is detailed followed by application. Both methods are established within epidemiological inference methods [129,130]. The approximately Bayesian Computation does not assume a likelihood function, whereas the Metropolis-Hastings algorithm does.

Approximate Bayesian Computation (ABC Method)

The ABC method was implemented to acquire suitable parameter estimates for β_0 and k_0. The algorithm follows:

Denote infectious notification data $\mathfrak{D}$ and modelled data $\mathfrak{M}$ dependent on parameter set $\theta \in \mathbb{R}^N$.

1. Generate a parameter set θ from prior probability distributions $p_1(.), p_2(.), ..., p_N(.)$. i.e. generate a sample for each parameter from some specified distribution.

2. Calculate $\mathfrak{M}$ given parameter set θ. i.e. Simulate the model with the generated parameter set.

3. Calculate discrepancy $\rho(\mathfrak{M}, \mathfrak{D})$.

This is repeated numerous times. The algorithm relies on a suitable function of ρ. The function used for this study is the sum of squares difference

$$\rho(\mathfrak{M}, \mathfrak{D}) = \sum_{i=0}^{T} (\mathfrak{M}(i) - \mathfrak{D}(i))^2. \tag{5.16}$$

This algorithm is sometimes referred to as a Monte-Carlo parameter sweep. The parameter set resulting in the smallest value of ρ will be the selected parameter set to simulate Irish TB. Given the criteria ascribed, this method can also be referred as the ordinary least squares method of parameter estimation.

The work of Chavez and colleagues [75] highlights the equations of a slow and fast progression TB model. When discussion of the parameter k (k being a constant in the original model) is brought up, Chavez states an acceptable assignment lies within the range 0.00256 to 0.00527. Using this information as an approximation, the probability distribution for k_0 will take the form $U(0.002, 0.006)$. The probability distribution for β_0 will take the form $U(0,1)$. The above algorithm was run 10,000 times. The resulting distribution can be seen in the figures 5.2 and 5.3.

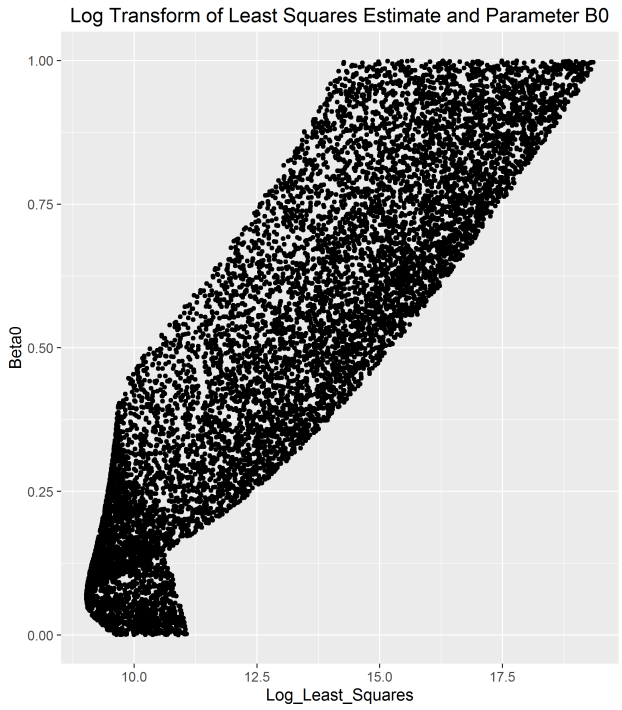

Figure 5.2: Log Transform Of The Sums of Squares Estimator For β_0 (Beta0)

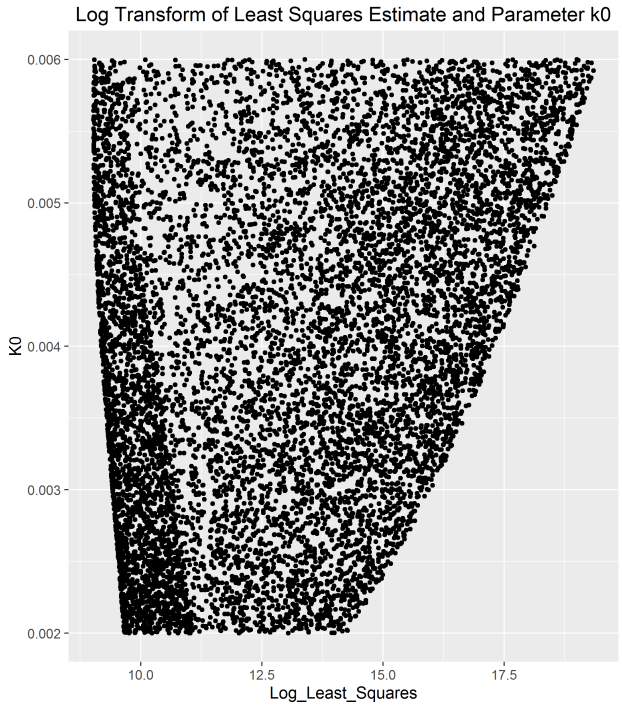

Figure 5.3: Log Transform of The Sums of Squares Estimator for k_0 (K0)

The values for β_0 and k_0 yielding the minimum sum of squares estimator were $\beta_0 = 0.0682$, and $k_0 = 0.0055$. Hence, by the ABC method, the rate at which infectious transmit the disease to the susceptible is 6.82% and the progression rate of exposed to infectious is 0.55%.

5.5.1 Metropolis-Hasting Algorithm with Sample Based Error Variance

The Metropolis-Hastings algorithm is a Bayesian technique that implements a Monte-Carlo Markov Chain (MCMC) method. The parameters in this case are considered to be random variables, with associated densities that incorporate known information. An MCMC process is a one that satisfies the property that random variable X_t depends only on X_{t-1} or $P(X_{t+1} = y|X_t = x_t, ..., X_0 = x_0) = P(X_{t+1} = y|X_t = x_t)$. The goal of an MCMC

process is to draw samples from a probability distribution, without being able to calculate its exact density. In this instance samples will be drawn for the parameters β_0 and k_0. The implemented algorithm follows the work of Smith [131].

General Algorithm Outline and Intuition

Given parameter set $\theta \in \mathbb{R}^N$, the Metropolis-Hastings algorithm takes current parameter set, θ^{k-1}, and proposes a new set[1] $\theta^* = \theta^{k-1} + U(.)$. The algorithm then takes

$$\theta^k = \begin{cases} \theta^* & \text{with probability } \alpha(\theta^*|\theta^{k-1}) \\ \theta^{k-1} & \text{with probability } (1-\alpha(\theta^*|\theta^{k-1})). \end{cases}$$

A note to be made is that $\alpha(\theta^*|\theta^{k-1})$ is not a conditional probability, but the probability of accepting θ^* given that it has been generated from the value θ^{k-1}.

The calculation of $\alpha(\theta^*|\theta^{k-1})$ is established, in part, by calculating the ratio of the posterior distribution evaluated at θ^* with the posterior distribution evaluated at θ^{k-1}, or through calculating

$$r(\theta^*|\theta^{k-1}) = \frac{\pi(\theta^*|\mathfrak{D})}{\pi(\theta^{k-1}|\mathfrak{D})},$$

where $\mathfrak{D}$ denotes the observed data. The calculation of a posterior distribution, $\pi(\theta|\mathfrak{D})$, relies on establishing a likelihood function. To achieve this, the assumption is made that the error measurements between the model and data are independent and identically distributed, and take value $\varepsilon_i \sim N(0,\sigma^2)$. From this, a likelihood function is calculated. Denote $SS_\theta = \sum_{i=1}^{n}(\mathfrak{M}_i(\theta) - \mathfrak{D}_i)^2$, where $\mathfrak{M}_i$ and $\mathfrak{D}_i$ represent the model and data at time i, respectively. The likelihood of the model takes the form

$$\pi(\mathfrak{D}|\theta) = L(\theta,\sigma|\mathfrak{D}) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-SS_\theta/2\sigma^2}$$

[1]The proposed new value does not have to take the exact form of θ^* as this study takes, the formal assignment is $\theta^* \sim J(\theta^*|\theta^{k-1})$. An alternative assignment could possibly be $\theta^* \sim N(\theta^{k-1},V)$.

The ratio $r(\theta^*|\theta^{k-1})$ is calculated using bayes rule,

$$\frac{\pi(\theta^*|\mathfrak{D})}{\pi(\theta^{k-1}|\mathfrak{D})} = \frac{\pi(\mathfrak{D}|\theta^*)\pi_0(\theta^*)}{\pi(\mathfrak{D}|\theta^{k-1})\pi_0(\theta^{k-1})}.$$

Where $\pi_0(\theta)$ is the prior probability of θ. Assuming a uniform prior, $\pi_0(\theta)$ becomes a constant function. Hence, $\frac{\pi_0(\theta^*)}{\pi_0(\theta^{k-1})} = 1$, and the ratio simply reduces to being the ratio of the likelihood distributions. From this we have

$$r(\theta^*|\theta^{k-1}) = \frac{\pi(\mathfrak{D}|\theta^*)}{\pi(\mathfrak{D}|\theta^{k-1})} = \frac{e^{-SS_{\theta^*}/2\sigma^2}}{e^{-SS_{\theta^{k-1}}/2\sigma^2}} = e^{-\frac{(SS_{\theta^*}-SS_{\theta^{k-1}})}{2\sigma^2}}$$

The acceptance probability $\alpha(\theta^*|\theta^{k-1})$ is defined as $\min(1,r)$.

The underlying intuition is that the likelihood function will be maximised when the sum of squares estimator is minimised. If the proposed parameter set θ^* simulates a "better" model than that of θ^{k-1}, then the ratio $\frac{\pi(\mathfrak{D}|\theta^*)}{\pi(\mathfrak{D}|\theta^{k-1})}$ will be large (as the likelihood $\pi(\mathfrak{D}|\theta^*)$ is greater than $\pi(\mathfrak{D}|\theta^{k-1})$) and the algorithm will have a high probability of accepting θ^* for it's next iteration. Similarly, if θ^* simulates a "worse" model then there is a low probability of it being accepted by the algorithm.

Pseudo-Code for Algorithm

The algorithm is as follows

1. Set the number of algorithm iterations M and initialise parameters n_s and σ_s^2, and the model parameter set θ^0.
2. Set $SS_{\theta^0} = \sum_{i=1}^{n}(\mathfrak{M}_i(\theta^0) - \mathfrak{D}_i)^2$.
3. Compute initial variance estimate (Chi-square estimate): $s_0^2 = \frac{SS_{\theta 0}}{n-N}$, where N is the number of parameters.
4. For $j = 1,...,M$ do
   (a) Construct $1 \times N$ vector u where $u_w \sim U(.)$, for $w = 1,2,..,N$.

(b) Construct candidate parameter set $\theta^* = \theta^{j-1} + u$.

(c) Compute $SS_{\theta^*} = \sum_{i=1}^{n} (\mathfrak{M}_i(\theta^*) - \mathfrak{D}_i)^2$.

(d) Compute $\alpha(\theta^* | \theta^{j-1}) = \min(1, e^{\frac{SS_{\theta^*} - SS_{\theta^{j-1}}}{2s_{k-1}^2}})$.

(e) If $U(0,1) < \alpha$
Set $\theta^j = \theta^*$, $SS_{\theta^j} = SS_{\theta^*}$.
else
Set $\theta^j = \theta^{j-1}$, $SS_{\theta^j} = SS_{\theta^{j-1}}$.

(f) Update $s_k \sim$ Inv-gamma(a_{val}, b_{val}), where

$$a_{val} = 0.5(n_s + n), b_{val} = 0.5(n_s \sigma_s^2 + SS_{\theta^j}).$$

The above algorithm considers σ^2 as a random parameter by calculating and updating s_k. The likelihood function is given by

$$\pi(\mathfrak{D}, \theta | \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{\frac{-SS_\theta}{2\sigma^2}}$$

With prior

$$\pi_0(\sigma^2) \propto (\sigma^2)^{-(c+1)} e^{-(d/\sigma^2)}$$

Which results in the posterior density

$$\sigma^2 | (\mathfrak{D}, \theta) \sim Inv - gamma\Big(\frac{n_s + n}{2}, \frac{n_s \sigma_s^2 + SS_\theta}{2}\Big)$$

Where $n_s = 2c$ and $\sigma_s^2 = \frac{c}{d}$. The value n_s typically takes values less than one and can be interpreted as the number of observations that provided information encoded in the prior.

As this method is an MCMC process, post-hoc diagnostics will be carried out in order to test convergence of the underlying process. The methodology used will follow that constructed by Gelman and Rubin [197]. This methodology implements $m > 1$ chains, and

convergence is diagnosed when sufficient burn-in time has elapsed and the output from all chains is indistinguishable. The underlying theory bases itself in the assumption that if two chains have converged, the mean of the variances of the individual chain should be indistinguishable from the global variance; that is, the variance of the all chains combined. The convergence diagnostic itself is given by the follow:

$$R = \sqrt{\frac{(\frac{2\hat{V}^2}{Var(\hat{V})} + 3)\hat{V}}{(\frac{2\hat{V}^2}{Var(\hat{V})} + 1)W}}$$

,

where $\hat{V} = \hat{\sigma}^2 + \frac{B}{mn}$, $\hat{\sigma}^2 = \frac{(n-1)W}{n} + \frac{B}{n}$, and W is the mean of the variance within each chain, n is the number of iterations, and $\frac{B}{n}$ is the between-chain variance.

While no strict criteria are given for R, the study states if the upper limit of the 95^{th} confidence interval on this estimate differs substantially from one, this indicates a lack of convergence.

Results

The Metropolis-Hastings algorithm was run $M = 10,000$ times to calculate β_0 and k_0 with initial parameter values $\beta_0 = 0.5$ and $k_0 = 0.005$. The parameters were set to these relatively arbitrary values in this instance to illustrate the algorithm, however, the ABC method is used in §6 to establish initial parameter values. The jump distribution for each parameter, denoted u_j in the above algorithm, was presumed to be $U(-0.05, 0.05)$ for β_0 and $U(-0.001, 0.001)$ for k_0. Smith notes, the values of n_s and σ_s are subjectively specified. The assumption of $n_s = 0.01$ and $\sigma_s = 0.01$ was made.

Each iteration of the algorithm can be seen in figures 5.4 and 5.5, allowing a burn-in time of 4,000 iterations we can see the posterior distribution of both parameters.

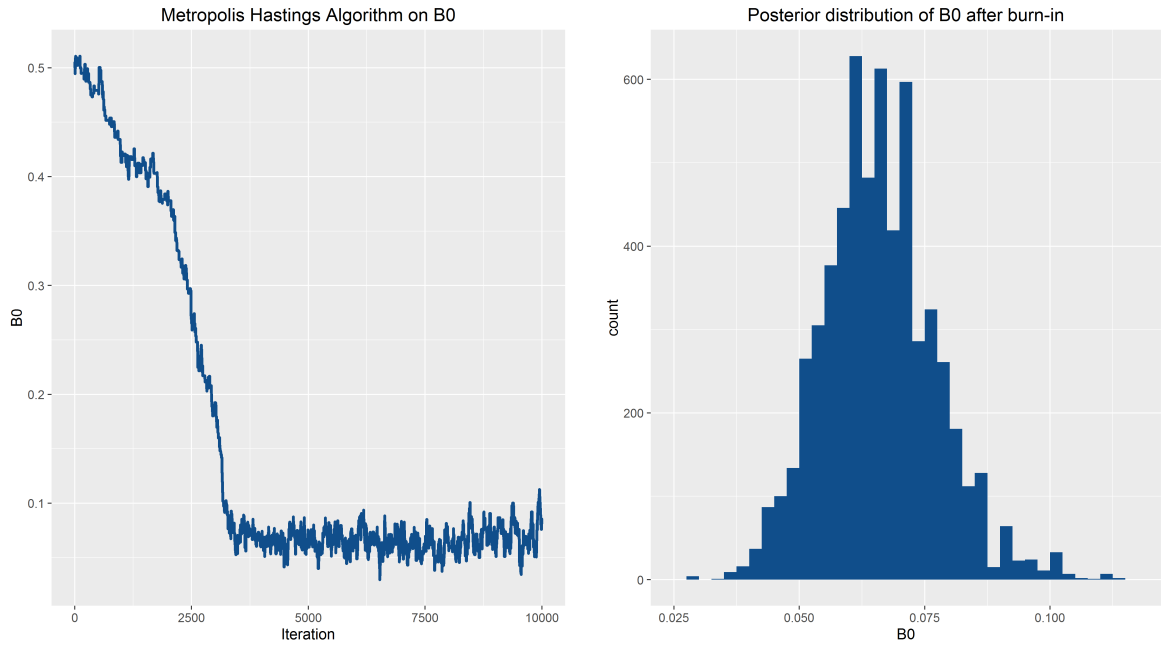

Figure 5.4: Metropolis Hastings Algorithm and Posterior Distribution for β_0 (B0) Given 10,000 Iterations and a Burn-in Time of 4,000 Iterations.

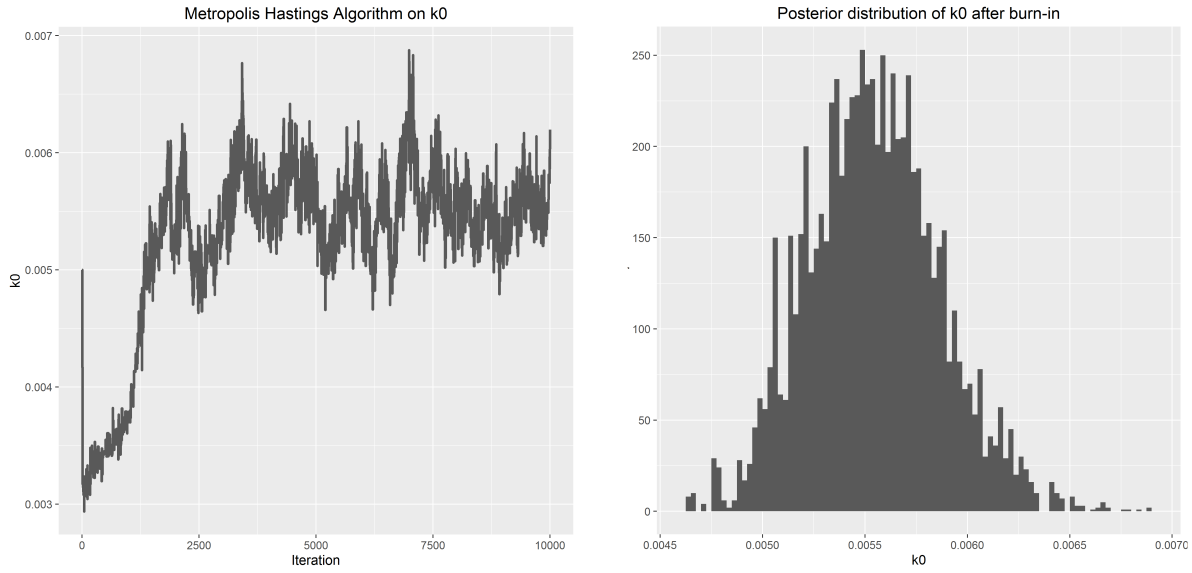

Figure 5.5: Metropolis Hastings Algorithm and Posterior Distribution for k_0 (K0) Given 10,000 Iterations and a Burn-in Time of 4,000 Iterations.

The statistics for each parameter are contained in table 5.5. The quantiles can be viewed as credible intervals for the parameters.

Statistic	β_0	k_0
Mean	0.06595569	0.005569941
Median	0.06552259	0.00557109
Standard Deviation	0.01130416	0.0003185046
Range	0.0829363	0.002421147
Skewness	0.3775253	-0.05297866
2.5% Quantile	0.04483685	0.004934423
97.5% Quantile	0.09067681	0.006204156

Table 5.5: Statistics for the Posterior Distributions of β_0 and k_0.

Post-Hoc Diagnostics

The Gelman and Rubin's convergence diagnostic was calculated for $m = 3$ chains. The results for the three chains are displayed in figure 5.6 and table 5.6

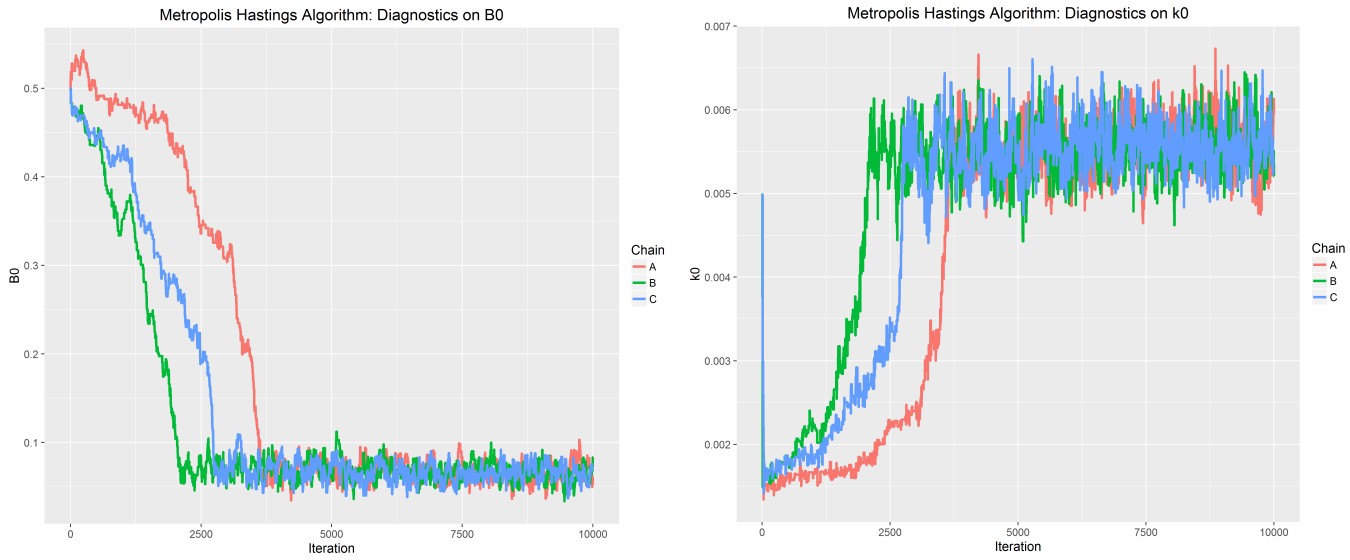

Figure 5.6: Convergence Of Three Chains For β_0 and k_0.

Parameter	Point Estimate	Upper 95% CI
β_0	1.01	1.01
k_0	1.00	1.01

Table 5.6: Potential Scale Reduction Factors For Parameters β_0 and k_0.

The upper confidence interval for both estimates are close to one, indicating convergence.

Both the ABC method and the Metropolis Hastings method gave output with very similar parameter estimates. The percentage difference between estimates for β_0 is approximately 3.3% and the percentage difference for k_0 is approximately -1.1%. Looking at both methods it appears that, given the range imposed on the parameters, there appears to be only one set of values for the parameters β_0 and k_0 within which the model can simulate the data. The mean of the posterior distributions will be used to calculate the basic

reproductive number and to simulate the model.

5.6 Calculation Of the Basic Reproductive Number and Simulation

5.6.1 The Basic Reproductive Number

An average of the seasonal parameters was taken to calculate the basic reproductive number. The season parameters are $\beta(t) = \beta_0(1 + sin(\frac{2\pi t}{12}))$ and $k(t) = k_0(1 + sin(\frac{2\pi t}{12}))$. Each of the functions have period 12 (due to §4 establishing seasonality of notifications as an annual seasonality), and will have one local maximum and one local minimum for each period. The selection of such functions is deliberate to match the sinusoidal data observed in §4.3.3. For $\beta(t)$, the functions has average value

$$\bar{\beta} = \frac{1}{T}\int_0^T \beta_0(1 + sin(\frac{2\pi t}{12}))$$

$$= \frac{1}{T}\frac{\beta_0(\pi T - 6cos(\frac{\pi T}{6}) + 6)}{\pi}$$

Since $\beta(t)$ is periodic, evaluating the integral over one period ($T = 12$) will obtain the average value.

$$\bar{\beta} = \frac{1}{12}\frac{\beta_0(\pi 12 - 6cos(\frac{\pi 12}{6}) + 6)}{\pi} \tag{5.17}$$

$$= \beta_0 \tag{5.18}$$

Similarly denote $\bar{k}$ the average value of $k(t)$, it too will have average value k_0 as it takes a similar function to $\beta(t)$.

This implies the estimate for the basic reproductive number takes the form

$$R_0 = \frac{\bar{\beta} w(\bar{k} + q\mu)}{(\mu + d + r)(\mu + \bar{k})} = \frac{\beta_0 w(k_0 + q\mu)}{(\mu + d + r)(\mu + k_0)} \tag{5.19}$$

With all parameters estimated, the basic reproductive number can be calculated.

$$R_0 = \frac{(0.065955)(0.35)(0.0055699 + (0.05)(0.00055))}{(0.00055 + 0.01667 + 0.016)(0.00055 + 0.005542)} = 0.6385 \tag{5.20}$$

Given the uncertainty on the transmission parameters generated from the Metropolis-Hastings algorithm, the percentiles of R_0 are given in table 5.7 below.

Percentile	0%	2.5%	5%	10%	90%	95%	97.5%	100%
R_0 Value	0.29	0.436	0.469	0.505	0.768	0.819	0.863	0.96

Table 5.7: Uncertainty of R_0 Given the Uncertainty of Parameters β_0 and k_0.

Due to the basic reproductive number value being less than one, it can be implied that for the seasonal model an epidemic has not occurred over the time period. In addition, due to the uncertainty surrounding the transmission parameters, the distribution of R_0 for the samples generated has maximum value 0.96. This implies with a degree of certainty, for the entire population, the disease cannot sustain itself and will die off given the current rates established.

Using the elements of the matrix (V^{-1}) used to calculate R_0, the calculation was made using the Metropolis-Hastings parameters that individuals who are exposed will spend on average $\frac{1}{0.00055+0.00557} = 163$ months being exposed. The same individuals will spend on average $\frac{0.00556}{(0.00055+0.01667+0.016)(0.00556+0.00055)} = 27$ months being infected. The individuals who are infected tend to stay infected for approximately $\frac{1}{0.00055+0.01667+0.016} = 30$ months.

5.6.2 Simulation

A simulation of the model with the given parameter set can be seen in figures 5.7 and 5.9, and the residual distribution and statistics can be seen in figure 5.8 and table 5.8. Simulation is conducted using the R package "deSolve" [198]. This package was specified to use a fourth order Runge-Kutta method [199] to simulate the underlying ODE system. Given the uncertainty in the transmission parameters, and given the distribution generated for each parameter using the Metropolis-Hastings algorithm, an upper and lower credibility interval was established on the point estimate for infections. This was achieved through simulating the set of generated parameters (after burn-in) and then extracting the 2.5^{th} and 97.5^{th} percentile of the resulting modelled infectious.

The resulting data from figures 5.7 and 5.9 indicate a downward trend in infectious and exposed individuals, and an upward trend in susceptible and recovered individuals.

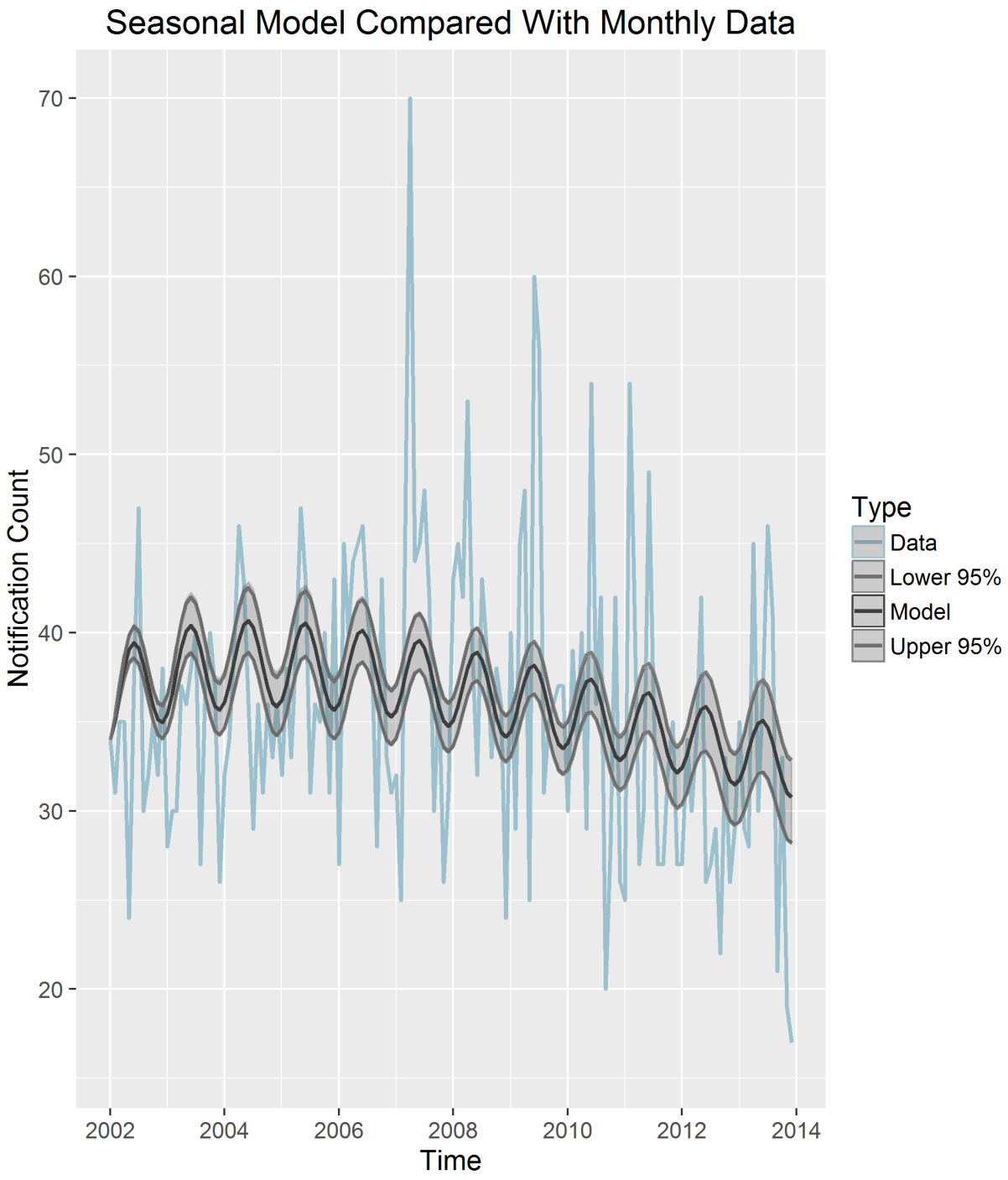

Figure 5.7: Notification Data and the Seasonal Model Simulation. The Shaded Interval is a 95% Credibility Region Given the Uncertainty of the Transmission Parameters.

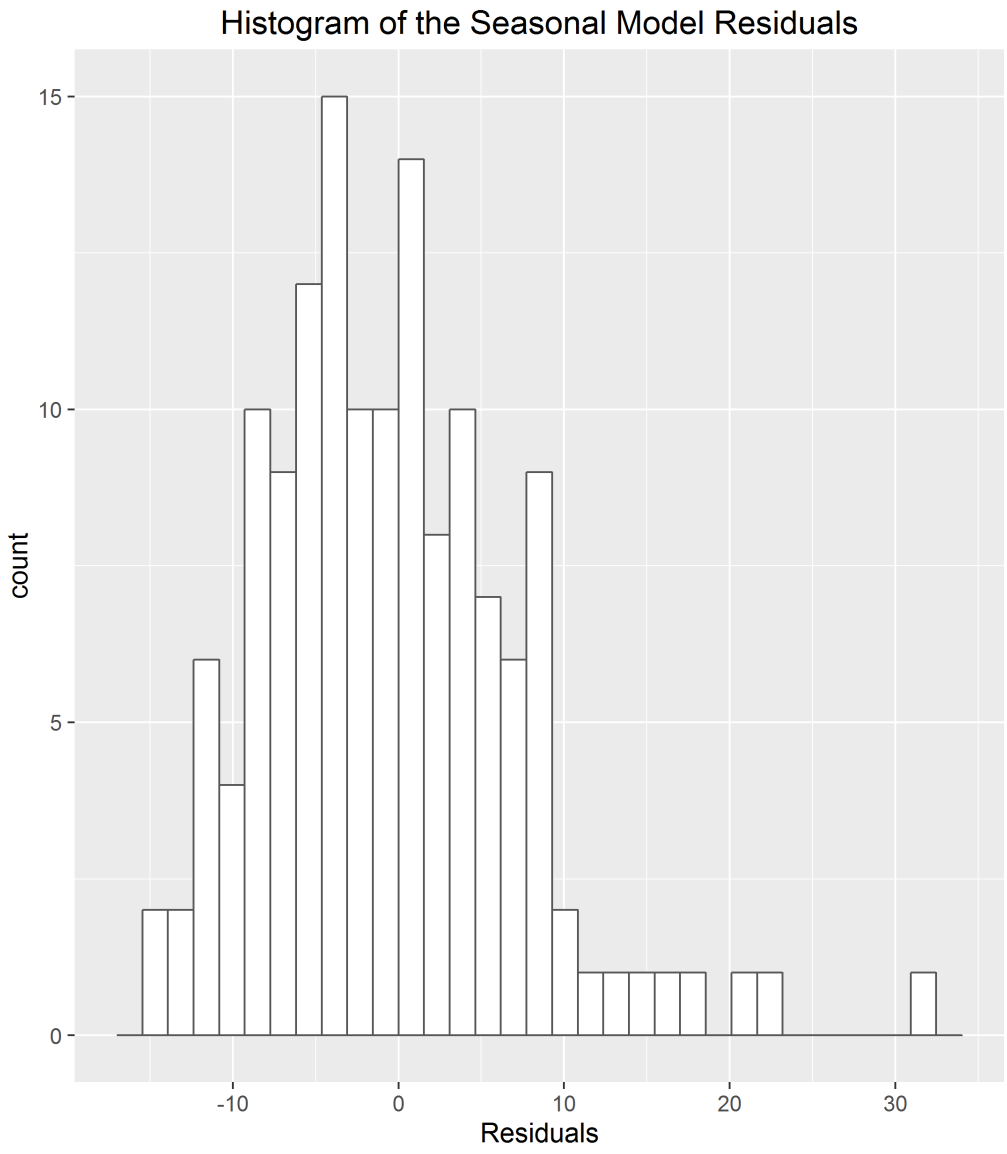

Figure 5.8: Histogram of Seasonal Model Residuals

Statistic	Mean	Standard Dev.	0% Quantile	25% Quantile	50% Quantile	75% Quantile	100% Quantile	Skew-ness	Kurt-osis
Seasonal Model Residuals	-0.5215	7.61	-15.05	-5.63	-1	3.57	31.35	0.855	1.619

Table 5.8: Statistics and Quantiles of Seasonal Model Residuals

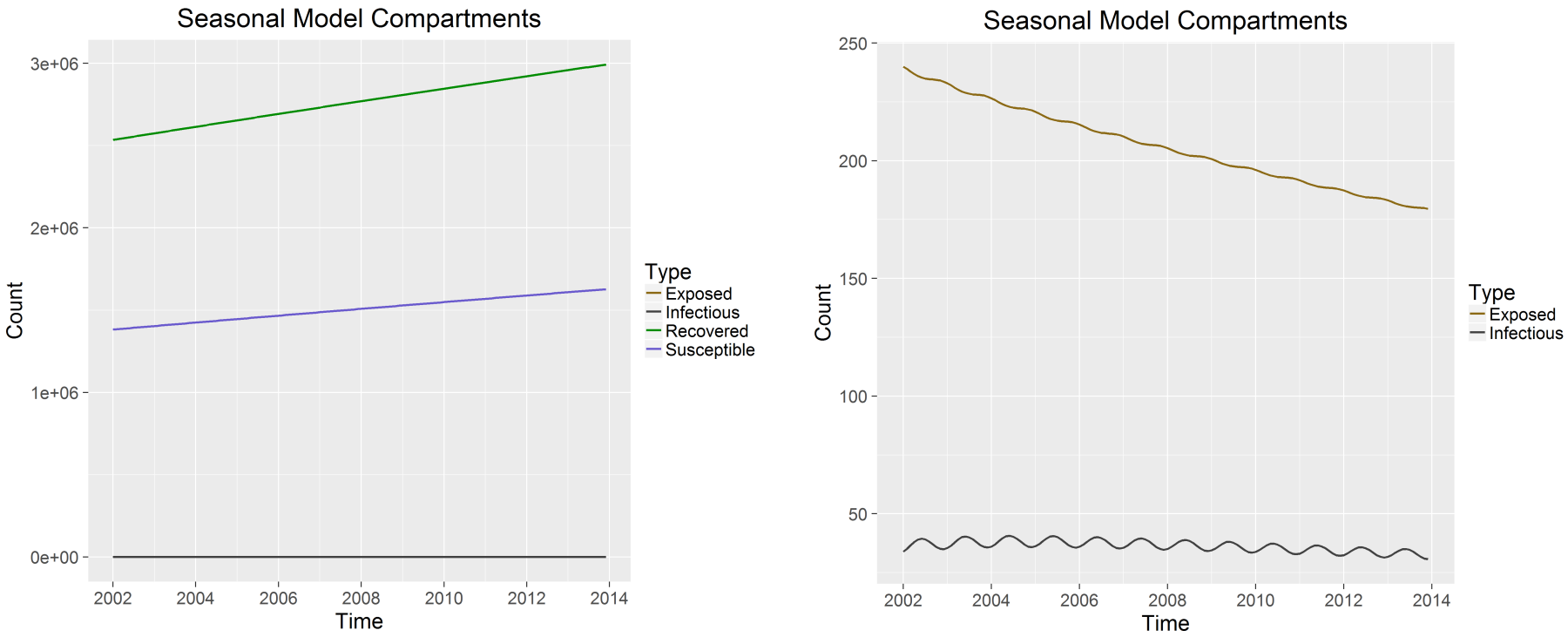

Figure 5.9: Seasonal Model Simulation - All Compartments (Left) And The Exposed And Infectious Compartments (Right), 2002 through to 2013

5.6.3 Model Extrapolation

To extrapolate the seasonal model is to presume no external factors or interventions will occur within the population. Figure 5.10 illustrates an extrapolation of the model 10 years into the future. The same methodology applies here as in the previous simulation section.

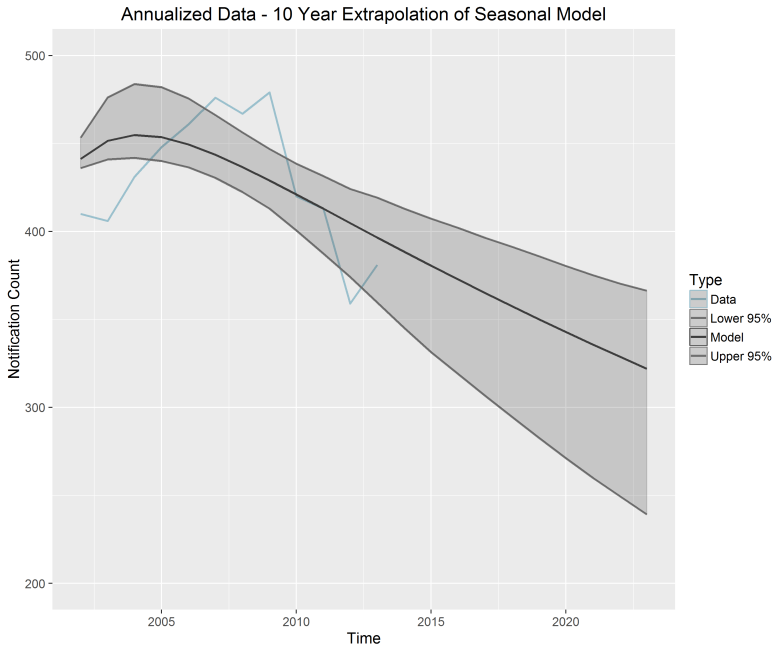

Figure 5.10: Annualized Seasonal Model Extrapolation 10 Years into the Future. The Shaded Interval is a 95% Credibility Region Given the Uncertainty of the Transmission Parameters.

Table 5.9 displays a comparison of annualized model values and data and table 5.10 displays the annualized extrapolated data along with uncertainty intervals.

Year	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
Model	441	452	455	454	450	444	437	429	421	413	405	397
Data	410	406	431	448	461	476	467	479	420	413	359	381
Error	7.6%	11.2%	5.5%	1.3%	-2.5%	-6.8%	-6.5%	-10.4%	0.3%	0.0%	12.7%	4.1%

Table 5.9: Annualized Seasonal Model Values Compared with Data.

Year	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
Extrapolated Model	389	381	373	365	358	350	343	336	329	322
Lower 95% CI	345	331	319	307	295	283	271	260	249	239
Upper 95% CI	413	407	402	396	391	386	380	375	370	366

Table 5.10: Annualized Seasonal Model Extrapolated 10 Years into the Future. The Upper and Lower Credibility Intervals were Calculated given the Uncertainty of the Transmission Parameters

Given a seasonal model, a decline is forecast to occur in the number of infections over time. Notifications are expected to decline approximately 8 (95% CI -11 to -4) cases annually for the next 10 years, accounting for an approximate -2.3% (95% CI -3.48% to -0.97%) change annually.

5.7 Conclusion

This chapter examined and simulated a homogeneous seasonal tuberculosis ODE model. The model was adapted to simulate Irish data, which was achieved by dividing the recruitment parameter between the susceptible and recovered compartments. The limitations and assumptions of the model were detailed, including the assumption of slow and fast progression to the infectious compartment. A theoretical qualitative analysis was conducted on the model detailing the disease-free and endemic equilibrium states. The basic reproductive number was calculated. Since it is not theoretically possible to calculate the basic reproductive number with periodic parameters being used in the model, an average was calculated of the periodic functions and that value was used to estimate the basic reproductive number.

Parameters were estimated using a combination of assumptions, literature, data, and statistical inference methods. A total of eight parameter estimates were required and the study systematically proposed rational estimates. The recruitment rate, death rate, disease induced death rate, and recovery rate were calculated using national vital statistics and the national TB data set. The progression rate, the fast progression parameter were acquired

from literature. The proportion of the recruitment rate entering into the susceptible compartment was assumed to be in line with national vaccination data. The remaining transmission parameters were estimated using both an Approximate Bayesian Computation method and the Metropolis Hastings algorithm. Both methods estimated similar parameter values indicating one viable parameter set. The Recovered and Exposed compartments had initial conditions that were derived under the assumption of national vaccination coverage. The initially infected compartment was acquire from the national data set, and by deduction, the Susceptible population was calculated.

The basic reproductive number was calculated using an averaging method and it was found to be less than one for the given parameter set, implying an epidemic is not a current threat. Simulation of the model was then carried out to observe the underlying dynamics over time, and the model was extrapolated 10 years into the future and ultimately a 2.3% annual decline was seen in notifications.

Strengths of this chapter include the construction of a viable epidemic model for Irish data that considers seasonal variation. The advanced inference methods used gave converged estimates on the transmission parameters. All parameters calculated set a bade for future modelling within Ireland or countries with similar attributes. the weaknesses of this chapter include the inability to theoretically derive an exact R_0, although an approximate value was obtained. The parameters estimated are done so with the best information possible, however, they are susceptible to error. This could potentially yield in inaccurate model. The fitted values are also subject to the possible bias of the underlying data they are fit.

The following chapter examines and simulates a model that divides the population into two groups: the local and migrant population. Alterations will be made to the model and a qualitative analysis conducted along with simulation.

Chapter 6

A TB Model Considering Migration

6.1 Introduction

In this chapter an epidemiological model considering migrant and local populations is presented. Increases in foreign-born TB have been seen in multiple other European countries (Appendix D) and deterministic models have previously been constructed, analysed, and applied to such locations as Quebec [148], Nigeria [149], and Canada [150,151] to consider the impact that an external population has on infection rates.

The migrant model discussed in §3 was found using a systematic search strategy informed by an exploratory analysis conducted in §4. The model identified from the literature review will now be refined to accommodate an Irish population accurately. Two separate models are derived: a model considering no interaction between the migrant and local populations, and a model considering an interaction. Once both models are established, a qualitative analysis is conducted. The basic reproductive number could not be calculated due to the non-existence of a disease-free equilibrium. However, it is shown that, given a minor alteration to each model, a basic reproductive number does exists.

The model parameters were estimated using data, literature, and assumptions. The transmission parameters of each model are estimated using Approximate Bayesian Compu-

tation and Metropolis-Hastings methods. These methods were developed in §5.5. Once parameter sets are estimated for each model, a simulation was carried out and the residuals between the model and data examined. Finally, the models were extrapolated 10 years into the future and the resulting data from the forecasts detailed.

6.2 Model Formulation

The migrant model obtained from the literature review (§3.5.3) will now be refined. The underlying dynamics follow SEIR compartments (Susceptible, Exposed/Latent, Infectious, and Recovered). The extended version of the model is presented within which recruitment (or immigration) for the migrant population does not just occur within the susceptible compartment, but also the exposed and infectious compartments. The system is represented by equations 6.1-6.10.

$$\frac{dS_M}{dt} = (1-v-w)\pi - \frac{\beta_1 S_M I_M}{N_N} - \mu S_M \tag{6.1}$$

$$\frac{dE_M}{dt} = v\pi + \frac{\beta_1 S_M I_M}{N_M} - (k_1+\mu)E_M \tag{6.2}$$

$$\frac{dI_M}{dt} = w\pi + k_1 E_M - (r_1+\mu+\mu_{I_M})I_M \tag{6.3}$$

$$\frac{dR_M}{dt} = r_1 I_M - \mu R_M \tag{6.4}$$

$$N_M = S_M + E_M + I_M + R_M \tag{6.5}$$

and

$$\frac{dS_L}{dt} = \Lambda - \frac{\beta_2 S_L I_L}{N_L} - \frac{\beta^* S_L I_M}{N_L} - \mu S_L \tag{6.6}$$

$$\frac{dE_L}{dt} = \frac{\beta_2 S_L I_L}{N_L} + \frac{\beta^* S_L I_M}{N_L} - p\frac{\beta^* I_M E_L}{N_L} + q\frac{\beta^* I_M R_L}{N_L} - (k_2 + \mu)E_L \tag{6.7}$$

$$\frac{dI_L}{dt} = p\frac{\beta^* I_M E_L}{N_L} + k_2 E_L - (r_2 + \mu + \mu_{I_L})I_L \tag{6.8}$$

$$\frac{dR_L}{dt} = r_2 I_L - q\frac{\beta^* I_M R_L}{N_L} - \mu R_L \tag{6.9}$$

$$N_L = S_L + E_L + I_L + R_L \tag{6.10}$$

Where the compartment subscripts M and L denote the migrant and local population, respectively, and where:

- $S_M(t)$ and $S_L(t)$ are the number of people at time t in the migrant and local population susceptible to Tuberculosis, respectively.
- $E_M(t)$ and $E_L(t)$ are the number of people at time t in the migrant and local population currently infected with latent Tuberculosis.
- $I_M(t)$ and $I_L(t)$ are the number of people at time t in the migrant and local population actively infected and spreading the disease within the population
- $R_M(t)$ and $R_L(t)$ are the number of people at time t in the migrant and local population who have recovered, are treated, or have immunity to the disease.
- π is the recruitment rate within the migrant population.
- v and w are the proportions of the recruitment rate being partitioned between the exposed and infectious compartments, respectively.
- β_1 and β_2 are the transmission rates, the rate at which the susceptible population transition into the the latent compartment, of the migrant and local populations within their own respective compartments.

- μ is the universal death rate
- μ_{I_M} and μ_{I_L} are the death rates due to infection of the migrant and local population, respectively.
- k_1 and k_2 are the progression rates, the rate at which individuals progress from latent to active TB, for the migrant and local populations within their respective systems.
- r_1 and r_2 are the recovery rates, the rate at which individuals recover or are treated for tuberculosis, for the migrant and local population within their respective compartments.
- β^* is the transmission rate of the migrant population to the local population
- q is the dampening factor on the transmission rate affecting the local recovered population that becomes reinfected with latent tuberculosis due to the transmission from the migrant infectious compartment.
- p is the dampening factor on the transmission rate effecting the local exposed population that transition to active tuberculosis due to the transmission of the migrant population.

The model makes the following assumptions

- The recruitment of new individuals into the population occurs in the susceptible, exposed, and infectious compartments for migrant populations. For the local population all individuals get recruited into the susceptible compartment.
- The death rate is constant and independent of each compartment. The assumption is made that the death rates between the local and migrant populations are equal.
- When the transmission of the disease occurs, it is proportional to the number of infective individuals of the respective compartment. This is referred to as the mass action principle [63].

- There exists an interaction between the migrant infectious compartment and the susceptible, exposed, and recovered compartments of the local population. There is a reactivation rate of the local recovered population due to the infectious migrant compartment and there is an additional progression term from exposed to infectious due to the infectious migrant population.

- $N(t) = N(t) + N_M(t) = S_L + E_L + I_L + R_L + S_M + E_M + I_M + R_M$ is the total population at time t. Its change can be expressed as $\frac{dN}{dt} = \pi + \Lambda - \mu N(t) - \mu_{I_M} I_M - \mu_{I_L} I_L$.

6.2.1 Alternative Model Construction

In this section, two alternative models will be constructed. Similar to the previous model, recruitment rate parameters will be altered. In the original model, recruitment (birth/migration) occurs in the susceptible compartment for the local population. This will be altered to allow recruitment to occur within the recovered compartment. Within the foreign-born or migrant population, it will be assumed recruitment can occur in either the susceptible, exposed, or recovered compartments (i.e. a foreign-born individual can migrate being susceptible, exposed, or recovered). It will not be assumed a foreign-born individual who is diagnosed with active TB can immigrant into the population.

No Interaction Between Local And Migrant Populations

A model will be considered in which there is no interaction between local and migrant infectious compartments. This is achieved by setting $\beta^* = 0$ in equations 6.6 to 6.9. Due to no interaction occurring, this essentially reduces systems 6.1-6.4 and 6.6-6.9 into two independent SEIR models with both simulating for the local and migrant populations individually. The system for the migrant population can be seen in equations 6.11-6.15 and

the for local population equations 6.16-6.20.

$$\frac{dS_M}{dt} = (1 - v_1 - v_2)\pi - \frac{\beta_1 S_M I_M}{N_N} - \mu S_M \tag{6.11}$$

$$\frac{dE_M}{dt} = v_1\pi + \frac{\beta_1 S_M I_M}{N_M} - (k_1 + \mu)E_M \tag{6.12}$$

$$\frac{dI_M}{dt} = k_1 E_M - (r_1 + \mu + \mu_{I_M})I_M \tag{6.13}$$

$$\frac{dR_M}{dt} = v_2\pi + r_1 I_M - \mu R_M \tag{6.14}$$

$$N_M = S_M + E_M + I_M + R_M \tag{6.15}$$

where v_1, and v_2 represent the partitioning variables of the recruitment rate parameter (replacing parameters v and w in equations 6.1-6.3), subject to $v_1 + v_2 \leq 1$ and $v_1 \geq 0$, $v_2 \geq 0$. The change in the local population takes the form

$$\frac{dS_L}{dt} = w_1\Lambda - \frac{\beta_2 S_L I_L}{N_L} - \mu S_L \tag{6.16}$$

$$\frac{dE_L}{dt} = \frac{\beta_2 S_L I_L}{N_L} - (k_2 + \mu)E_L \tag{6.17}$$

$$\frac{dI_L}{dt} = k_2 E_L - (r_2 + \mu + \mu_{I_L})I_L \tag{6.18}$$

$$\frac{dR_L}{dt} = (1 - w_1)\Lambda + r_2 I_L - \mu R_L \tag{6.19}$$

$$N_L = S_L + E_L + I_L + R_L \tag{6.20}$$

The system of equations 6.11-6.20 will be referred to as the "migrant model with no interaction". The reason for excluding interaction arises from a systematic review by Sandgren and colleagues which concluded "TB in a foreign-born population does not have a significant influence on TB in the native population in EU/EEA" [94]

Interaction Between Local And Migrant Populations

The original model (equations 6.1-6.10) only investigates a one-way interaction between the migrant infectious compartment and the local susceptible, exposed, and recovered

compartments. The authors of the original model go on to discuss the possibility of a two-way interaction model. As such, a model to examine a two-way interaction between the migrant and local populations was developed. The original model was refined to consider this. In the refined model, interaction is assumed to occur between the local infectious compartment and the migrant susceptible and exposed compartments and an interaction is also assumed to occur between the migrant infectious compartment and local susceptible and exposed compartments. The interaction between infectious and recovered will be omitted, as when estimating the model parameters, the assumption is made that recovered individuals have achieved immunity and thus cannot be affected by infectious individuals. The refined system can be seen in equations 6.21 - 6.30.

$$\frac{dS_M}{dt} = (1 - v_1 - v_2)\pi - \frac{\beta_1 S_M I_M}{N_N} - \frac{\beta_2^* S_M I_L}{N_M} - \mu S_M \tag{6.21}$$

$$\frac{dE_M}{dt} = v_1\pi + \frac{\beta_1 S_M I_M}{N_M} + \frac{\beta_2^* S_M I_L}{N_M} - p_2\frac{\beta_2^* I_L E_M}{N_M} - (k_1 + \mu)E_M \tag{6.22}$$

$$\frac{dI_M}{dt} = p_2\frac{\beta_2^* I_L E_M}{N_M} + k_1 E_M - (r_1 + \mu + \mu_{I_M})I_M \tag{6.23}$$

$$\frac{dR_M}{dt} = v_2\pi + r_1 I_M - \mu R_M \tag{6.24}$$

$$N_M = S_M + E_M + I_M + R_M \tag{6.25}$$

and

$$\frac{dS_L}{dt} = w_1\Lambda - \frac{\beta_2 S_L I_L}{N_L} - \frac{\beta_1^* S_L I_M}{N_L} - \mu S_L \tag{6.26}$$

$$\frac{dE_L}{dt} = \frac{\beta_2 S_L I_L}{N_L} + \frac{\beta_1^* S_L I_M}{N_L} - p_1\frac{\beta_1^* I_M E_L}{N_L} + -(k_2 + \mu)E_L \tag{6.27}$$

$$\frac{dI_L}{dt} = p_1\frac{\beta_1^* I_M E_L}{N_L} + k_2 E_L - (r_2 + \mu + \mu_{I_L})I_L \tag{6.28}$$

$$\frac{dR_L}{dt} = (1 - w_1)\Lambda + r_2 I_L - \mu R_L \tag{6.29}$$

$$N_L = S_L + E_L + I_L + R_L \tag{6.30}$$

Where the subscripts M and L denote the migrant and local populations, respectively. A schematic of the new system can be visualised in figure 6.1 below.

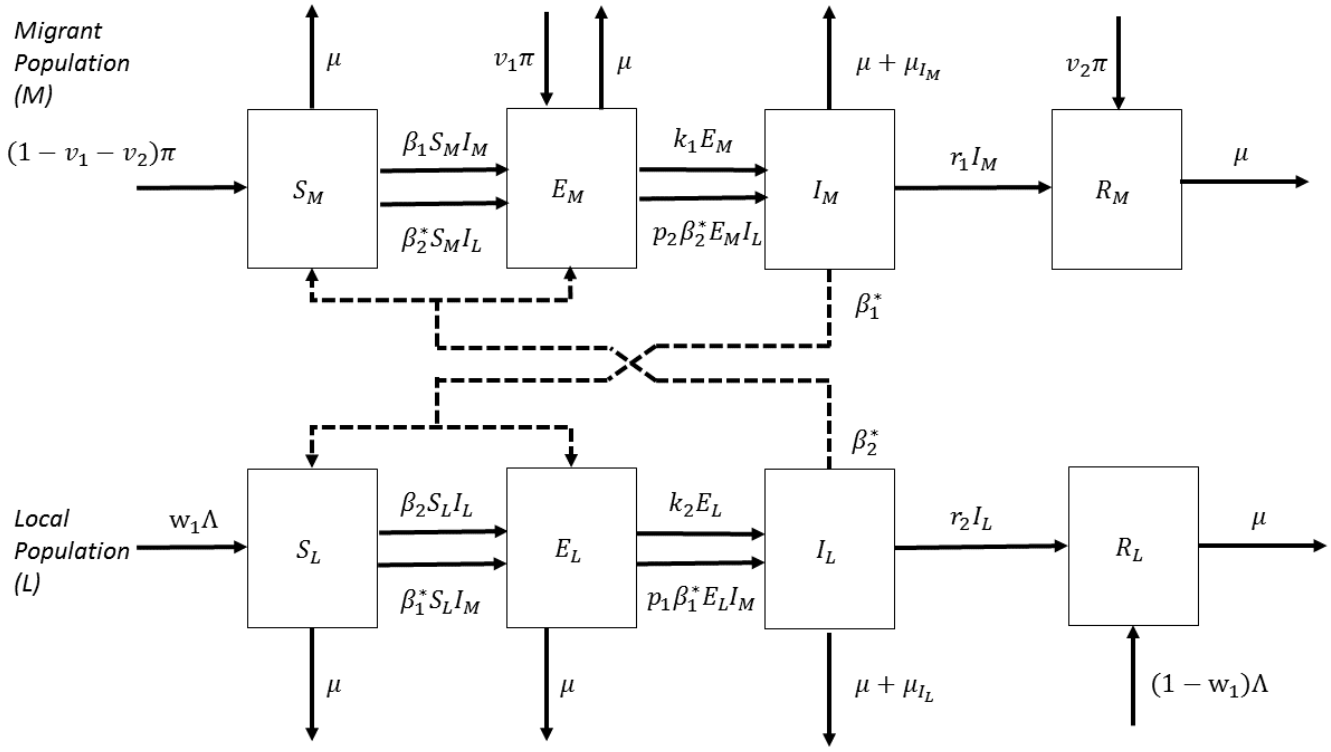

Figure 6.1: A Schematic of the Migrant Model With Interaction Occurring Between both Migrant and Local Populations

The above model (equations 6.21-6.30) will henceforth be referred to as the "migrant model with interaction". For both models (migrant models with and without interaction), the definition of the migrant population is as follows: the top 20 foreign-born populations that contributed to TB in Ireland from 2002 to 2013. These countries will be referred to as the migrant population henceforth. The top 20 countries account for approximately 80.3% of all foreign-born notifications and have accounted for 32.3% of national notifications. The local population for both models is defined as the complement of the migrant population.

6.3 Qualitative Analysis

For both models (equations 6.11-6.20, 6.21-6.30) a disease-free equilibrium cannot be calculated. A corollary of this is that the basic reproductive number also cannot be calculated as it is dependent on a disease-free equilibrium [77]. The disease-free equilibrium cannot be calculated in either system because of the recruitment parameter within the exposed compartment for the migrant population is present in both models. For the model with no interaction, equation 6.12 has a positive recruitment rate $v_1\pi$. Similarly, within the migrant model with interaction equation 6.22 has an identical positive recruitment rate parameter. This is an example of when an attempt to make a model empirically accurate contributes to the complexity of analysing the model.

A disease-free equilibrium does not exist because if the number of exposed and infectious individuals is set to zero, equations 6.12 and 6.22 reduce to

$$\frac{dE_M}{dt} = v_1\pi$$

As parameters $v_1 > 0$ and $\pi > 0$, the above equation implies the change in the exposed compartment is always positive, and hence, can never transition into disease-free state.

One of two assumptions must be made to calculate the disease-free equilibrium in both models: either $v_1 = 0$ in equations 6.12 and 6.22, or $\pi = 0$ for both systems. In an attempt at calculating a reproductive number for the systems, the study will assume $v_1 = 0$ as this essentially assumes a model with no recruitment of exposed individuals into the migrant population. Assuming $\pi = 0$ results in no migration occurring within the migrant population which is a more empirically inaccurate assumption.

6.3.1 Basic Reproductive Number For Migrant Model With No Interaction.

With the assumption of $v_1 = 0$, the disease-free equilibrium can be acquired by setting $E_M(t) = 0$, $E_L(t) = 0$, $I_M(t) = 0$, and $I_L(t) = 0$. Denote

$$D_0 = (S_{MDF_0}, E_{MDF_0}, I_{MDF_0}, R_{MDF_0}, S_{LDF_0}, E_{LDF_0}, I_{LDF_0}, R_{LDF_0})$$

the solution set to the disease-free equilibrium, solving system 6.11-6.15 in the above manner results in

$$D_0 = (\frac{(1-v_2)\pi}{\mu}, 0, 0, \frac{v_2\pi}{\mu}, \frac{w_1\Lambda}{\mu}, 0, 0, \frac{\Lambda(1-w_1)}{\mu}).$$

Given the disease-free equilibrium, the basic reproductive number can now be calculated through the next generation method [133], as described in §5.3.2.

The matrices F and V are calculated to be:

$$F = \begin{pmatrix} 0 & \beta_1(1-v_2) & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \beta_2 w_1 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

and

$$V = \begin{pmatrix} k_1+\mu & 0 & 0 & 0 \\ -k_1 & r_1+\mu+\mu_{I_M} & 0 & 0 \\ 0 & 0 & k_2+\mu & 0 \\ 0 & 0 & -k_2 & r_2+\mu+\mu_{I_L} \end{pmatrix}.$$

The basic reproductive number is achieved through the calculation $\rho(FV^{-1})$, where the function ρ is the spectral radius (largest eigenvalue) of the matrix. Letting λ_i be the roots of the characteristic equation, the eigenvalues of FV^{-1} for the migrant model without

interaction follow.

$$\lambda_1 = 0, \lambda_2 = 0, \lambda_3 = \frac{\beta_2 k_2 w_1}{(k_2+\mu)(r_2+\mu+\mu_{I_L})}, \lambda_4 = \frac{\beta_1 k_1 (1-v_2)}{(k_1+\mu)(r_1+\mu+\mu_{I_M})}. \tag{6.31}$$

As the model does not have any interaction, λ_3 can be seen as the basic reproductive number for the local population as it is independent of parameters used to model the migrant population. Hence denote $R_{(0)L} = \lambda_3$. Similarly λ_4 can be seen as the basic reproductive number for the migrant population. Hence denote $R_{(0)M} = \lambda_4$. The basic reproductive number for the model as a whole is

$$R_0 = \max(R_{(0)M}, R_{(0)L}). \tag{6.32}$$

This is the largest eigenvalue of the entire model. Similar to §5, the matrix V^{-1} has it's own interpretation, the elements being the expected time an individual spends in each state. The matrix V^{-1} is given as:

$$V^{-1} = \begin{pmatrix} \frac{1}{k_1+\mu} & 0 & 0 & 0 \\ \frac{k_1}{(k_1+\mu)(r_1+\mu+\mu_{I_M})} & \frac{1}{r_1+\mu+\mu_{I_M}} & 0 & 0 \\ 0 & 0 & \frac{1}{k_2+\mu} & 0 \\ 0 & 0 & \frac{k_2}{(k_2+\mu)(r_2+\mu+\mu_{I_L})} & \frac{1}{r_2+\mu+\mu_{I_L}} \end{pmatrix}.$$

Hence, the the migrant population are expected to remain exposed and infectious for $\frac{1}{k_1+\mu}$ and $\frac{1}{r_1+\mu+\mu_{I_M}}$ units off time. Individuals who are exposed are expected to be infectious for $\frac{k_1}{(k_1+\mu)(r_1+\mu+\mu_{I_M})}$ units of time. The same applies to the local population, they are expected to remain exposed and infectious for $\frac{1}{k_2+\mu}$ and $\frac{1}{r_2+\mu+\mu_{I_L}}$ units off time. Individuals who are exposed are expected to be infectious for $\frac{k_2}{(k_2+\mu)(r_2+\mu+\mu_{I_L})}$ units of time.

The basic reproductive number for the model with interaction will now be calculated.

6.3.2 Basic Reproductive Number For Migrant Model With Interaction.

The disease-free equilibrium for the model with interaction was calculated to be the same as the model without interaction.

$$D_0 = (\frac{(1-v_2)\pi}{\mu}, 0, 0, \frac{v_2\pi}{\mu}, \frac{w_1\Lambda}{\mu}, 0, 0, \frac{\Lambda(1-w_1)}{\mu}).$$

The underlying calculation of R_0 differs as the calculation of matrix F changes slightly due to the inclusion of the interaction variables within the model. The matrices F and V were calculated to be

$$F = \begin{pmatrix} 0 & \beta_1(1-v_2) & 0 & \beta_2^*(1-v_2) \\ 0 & 0 & 0 & 0 \\ 0 & \beta_1^* w_1 & 0 & \beta_2 w_1 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

and

$$V = \begin{pmatrix} k_1+\mu & 0 & 0 & 0 \\ -k_1 & r_1+\mu+\mu_{I_M} & 0 & 0 \\ 0 & 0 & k_2+\mu & 0 \\ 0 & 0 & -k_2 & r_2+\mu+\mu_{I_L} \end{pmatrix}.$$

The calculation of the basic reproductive number results in a long algebraic expression in this instance. In order to overcome this, denotations will be made. Keeping similar notation for the model without interaction (equations 6.31) let $\lambda_3 = \frac{\beta_2 k_2 w_1}{(k_2+\mu)(r_2+\mu+\mu_{I_L})}$ (previously considered the basic reproductive number for the local population in the model without interaction) and let $\lambda_4 = \frac{\beta_1 k_1 (1-v_2)}{(k_1+\mu)(r_1+\mu+\mu_{I_M})}$ (previously considered the basic reproductive number for the migrant population in the model without interaction). In addition, denote $\lambda_3' = \frac{\beta_2^* k_2 (1-v_2)}{(k_2+\mu)(r_2+\mu+\mu_{I_L})}$ and $\lambda_4' = \frac{\beta_1^* k_1 w_1}{(k_1+\mu)(r_1+\mu+\mu_{I_M})}$. Letting $\hat{\lambda}_i$ be the roots of the characteristic equation, the eigenvalues of FV^{-1} for the migrant model with interaction follow.

$$
\begin{aligned}
\hat{\lambda_1} =& 0, \\
\hat{\lambda_2} =& 0, \\
\hat{\lambda_3} =& \frac{\lambda_3 + \lambda_4 + \sqrt{(\lambda_3 - \lambda_4)^2 + 4\lambda_3'\lambda_4'}}{2}, \\
\hat{\lambda_4} =& \frac{\lambda_3 + \lambda_4 - \sqrt{(\lambda_3 - \lambda_4)^2 + 4\lambda_3'\lambda_4'}}{2}.
\end{aligned}
$$

Hence the basic reproductive number of the system is given by

$$R_0 = \max(\hat{\lambda_3}, \hat{\lambda_4})$$

As the matrix V is identical to the model with no interaction, the matrix V^{-1} will have same elements and, hence, can be interpreted the same.

Remark

For the above eigenvalues $(\hat{\lambda_3}, \hat{\lambda_4})$, if it is assumed that there was no interaction between local and migrant populations, in the model this would translate to $\beta_2^* = \beta_1^* = 0$ within the equations 6.21-6.30. This results in both $\lambda_3' = 0$, and $\lambda_4' = 0$. In this case,

$$
\begin{aligned}
\hat{\lambda_3} &= \frac{\lambda_3 + \lambda_4 + \sqrt{(\lambda_3 - \lambda_4)^2}}{2} \\
\hat{\lambda_3} &= \lambda_3
\end{aligned}
$$

Likewise,

$$
\begin{aligned}
\hat{\lambda_4} &= \frac{\lambda_3 + \lambda_4 - \sqrt{(\lambda_3 - \lambda_4)^2}}{2} \\
\hat{\lambda_4} &= \lambda_4.
\end{aligned}
$$

Hence, when the interaction terms are set to zero, the systems basic reproductive number reduces to the basic reproductive number of the model when no interaction is considered.

In addition, if similar rates are observed between the migrant and local populations, namely if $\beta_1 = \beta_1^*$ and $\beta_2 = \beta_2^*$ holds for the transmission rates and similar recruitment rates are observed $w_1 = 1 - v_2$, then this results in $\lambda_3 = \lambda_3'$ and $\lambda_4 = \lambda_4'$. Which brings about

$$\begin{aligned}
\hat{\lambda_3} =& \frac{\lambda_3 + \lambda_4 + \sqrt{(\lambda_3 - \lambda_4)^2 + 4\lambda_3\lambda_4}}{2} \\
=& \frac{\lambda_3 + \lambda_4 + \sqrt{(\lambda_3 + \lambda_4)^2}}{2} \\
=& \frac{\lambda_3 + \lambda_4 + (\lambda_3 + \lambda_4)}{2} \\
=& \lambda_3 + \lambda_4,
\end{aligned}$$

and by a similar calculation, $\hat{\lambda_4} = 0$. The sum of the basic reproductive number for each sub-population can be interpreted as the basic reproductive for the entire population. This result is dependent on (i) the assignments of the parameter and (ii) an interaction is occurring between the two sub-populations.

6.4 Parameter Estimation

A total of 14 parameters required estimation for the model without interaction, and a total of 18 parameters required estimation for the model with interaction. Both models will share common parameter estimates. The parameter estimates that vary between each model are $\beta_1, \beta_2, \beta_1^*, \beta_2^*, k_1, k_2$; these parameters will be referred to as the transmission parameters. The remaining parameters: $v_1, v_2, \pi, \Lambda, w_1, \mu, \mu_{I_M}, \mu_{I_L}, r_1, r_2, p_1, p_2$ will remain constant regardless of the type of migrant model. Similarly the initial conditions: $S_M(0), E_M(0), I_M(0)$,$R_M(0), S_L(0), E_L(0), I_L(0)$, and $R_L(0)$, shall not change between models.

The following parameters were calculated from data acquired from the Central Statistics Office [96]: π, the recruitment rate for the migrant population; Λ, the recruitment rate for the local population; and μ, the population death rate. The parameters: μ_{I_M}; the death rate due to TB of the migrant population, μ_{I_L}; the death rate due to TB of the local population, r_1; the recovery or removal rate for the migrant population, r_2; the recovery or removal rate for the local population, were calculated using the dataset acquired and used in §3. The model's initial conditions were estimated using the initial conditions established in §5.4.4 along with additional assumptions. The proportional parameters: v_1, v_2, w_1, p_1 and p_2 were estimated using a combination of literature and assumptions. The remaining transmission parameters : $\beta_1, \beta_2, \beta_1^*, \beta_2^*, k_1$, and k_2, shall be estimated by way of statistical inference.

6.4.1 Recruitment, Death, and Recovery Rate Parameters

Recruitment Rates

The recruitment rate for the migrant population will be the net migration rate for the foreign-born population. The recruitment rate for the local population will be the recruitment rate established in §5 minus the recruitment rate for the migrant population. Census data for the stock total of the migrant population are illustrated in figure 6.2.

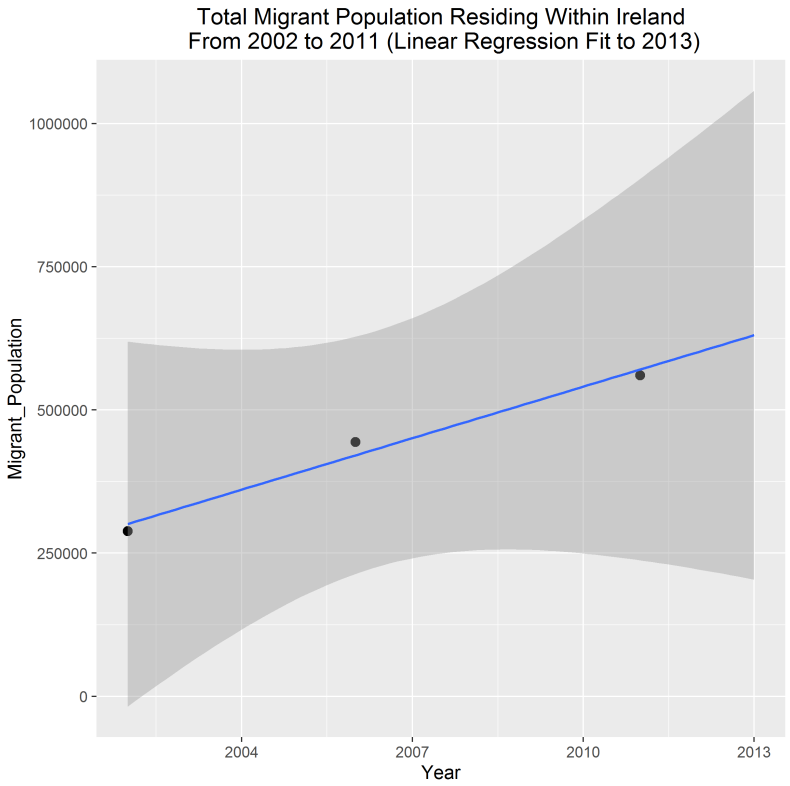

Figure 6.2: Total Migrant (Top 20 Contributors to TB) Population in Ireland. Data available for 2002, 2006, and 2011 with a linear regression fit for 2002 to 2013.

As Ireland does not record census data annually, interpolation and extrapolation of the population data was done using a linear trend. Data were only available for the years 2002, 2006, and 2011, all other years from 2002 to 2013 have been calculated with the linear trend. Within the resulting regression, the average annual change in migrants was 29,966 individuals per year. This converts to an average monthly change of $\pi = \frac{29{,}966}{12} = 2{,}497$ individuals. By deduction the recruitment rate parameter for the local population is $\Lambda = 7{,}267 - 2{,}497 = 4{,}770$.

Death Rate Parameter

The universal death rate will be the death rate established in §5, $\mu = 0.00055$. The death rates caused through TB are calculated from table 6.1 constructed using the national dataset.

TB Cause Of Death Migrant (M) Local (L)	Yes (M)	Total (M)	Yes Proportion (M)	Yes (L)	Total (L)	Yes Proportion (L)
2002	0	99	0.0000	0	271	0.0000
2003	0	57	0.0000	5	299	0.0167
2004	0	90	0.0000	5	297	0.0168
2005	0	105	0.0000	10	296	0.0338
2006	4	117	0.0342	6	293	0.0205
2007	0	142	0.0000	7	283	0.0247
2008	2	164	0.0122	7	264	0.0265
2009	1	157	0.0064	9	270	0.0333
2010	0	138	0.0000	8	246	0.0325
2011	0	162	0.0000	10	214	0.0467
2012	2	124	0.0161	1	196	0.0051
2013	5	140	0.0357	2	205	0.0098
Average	1	125	0.0087	6	261	0.0222

Table 6.1: Count and Proportion of Individuals who had Yes/No Filled Out On Their Notification Form When Assessed Whether They Had Died Due To TB.

The death rate parameter for the migrant and local population will be the average rate from 2002 to 2013, hence $u_{I_M} = 0.00872$ and $u_{I_L} = 0.02221$.

Recovery-Rate Parameters

The recovery-rate parameters were calculated from table 6.2 constructed from the national dataset. The recovery rate can be estimated as the inverse of the average recovery time.

Statistic	Local	Migrant
Mean	114	102
Median	66	59
Mode	31	31
Main	1	1
Max	2650	2422
Skewness	18.63	6.76

Table 6.2: Statistics On Recovery Time For The Local And Migrant Populations

Due to the large skewness in these data, the median will be used as an estimate for recovery time. Hence, $r_1 = \frac{1}{59} = 0.01695$ and $r_2 = \frac{1}{66} = 0.01515$.

6.4.2 Initial Conditions

Migrant Population

The initial conditions established in the previous chapter were as follows

$$S(0) = 1,382,374, E(0) = 240, I(0) = 34, R(0) = 2,535,000$$

The proportion of the total population made up by the migrant population was approximately 7.4% in 2002. A naive initial assumption is to assume the 7.4% distributes uniformly among the above compartments. This results in the following initial conditions

$$S_M(0) = 102,296, E_M(0) = 18, I_M(0) = 3, R_M(0) = 187,590$$

Given this calculation, the condition $S_M(0)+E_M(0)+I_M(0)+R_M(0) = N_M(0)$ is held. The initial infected population was acquired from the dataset and calculated to be $I_M(0) = 7$. For the year 2002, active TB cases within the total population were 3.14 times that of the migrant population. Using this multiple as an estimate for the exposed compartment we have $18 \times 3.14 = 57$. To support this estimate, the actual number of infectives in the year 2002 was 99 for the migrant population. The ratio of infectives to exposed found in chapter 5 was found to be 7.07. Applying that ratio to the number of infective migrants

results in an annual total exposed population of $7.07 * 99 = 700$, converting this value into a monthly value results in $700/12 = 58$. Because of these two estimates being very similar the initial Exposed will be $E(0) = 58$.

Very little data are available of the vaccination rate on foreign-born individuals. In an American study conducted in 2003 [152] approximately 25% of foreign-born individuals reported having received the BCG vaccination. Within the previous chapter the BCG vaccine was stated as being approximately 69% effective against TB. Hence the study assumes $0.25 \times 0.69 = 0.173$ or 17.3% of foreign-born are immune to TB. This proportion will be used to calculate the initial population estimate $R_M(0)$ and as a result $S_M(0)$. Out of the foreign-born population it will be assumed 25% are vaccinated, which will result in $R_M(0) = 0.173 \times N_M(0) = 50,136$. Using the condition $S_M(0) + E_M(0) + I_M(0) + R_M(0) = N_M(0)$, the initial susceptible population was calculated to be $S_M(0) = 239,603$.

Local Population

The local-population made up approximately 92.6% of the total population in 2002. Using the initial conditions from §5.4.4, it would appear reasonable to assume the proportion of individuals within each compartment for the seasonal model would be relatively representative for the proportion of individuals in each compartment for the local population. The proportion of individuals in each compartment within the seasonal model are given below.

$$S(0) = 35.286\%, E(0) = 0.0061\%, I(0) = 0.0009\%, R(0) = 64.707\%$$

Given the local population was $N_L(0) = 92.6\% \times 3,917,648 = 3,627,742$, multiplying the proportions by the initial total local population will result in an estimate for each compartment.

$$S_L(0) = 1,280,078, E_L(0) = 222, I_L(0) = 27, R_L(0) = 2,347,409$$

6.4.3 The Proportion Parameters

The proportional parameters are parameters that divide other parameters into proportions. Within the model the proportional parameters are: v_1; the proportion of migrant individuals entering into the exposed compartment, v_2; the proportion of migrant individuals entering into the recovered compartment, w_1; the proportion of local individuals entering into the susceptible compartment, p_1; the dampening factor on the transmission rate affecting the local population infected by the migrant population, and p_2; the dampening factor on the transmission rate affecting the migrant population who were infected by the local population.

Data examining the proportions v_1 and v_2 within the literature could not be found. The proportion of individuals entering into the exposed compartment will be derived from the initial condition proportions. The study assumes approximately 0.002% of the migrant population are exposed initially and it will be assumed this rate can be applied to the incoming migrant population. Hence, $v_1 = 0.00002$. Likewise, the study assumes approximately 17.3% of the migrant population are recovered initially, and will assume this rate can be applied to the incoming migrant population. Hence, $v_2 = 0.173$. For the proportion of local individuals entering into the susceptible compartment, as the local population made up a 92.7% of the total population in 2002, it will be assumed the proportion established in chapter 4 will act as a reliable estimate for the local population. The rate established in chapter 4 was $w_1 = 0.35$. For the dampening factors, the rates used for numerical simulation in the original model [93] will be used, $p_1 = 0.80$ and $p_2 = 0.80$.

6.4.4 Transmission Parameters

As mentioned, the transmission parameters will vary between the two models being considered. Hence, two sets of estimations must be made. Statistical inference will be used to estimate these parameters. The methods used will be the methods implemented in Chapter 5: The Approximate Bayesian Computation (ABC) method and the Metropolis-Hasting algorithm.

Model With No Interaction Considered

The parameters requiring estimation for this model are the transmission rates: $\beta_1, \beta_2, k_1, k_2$. The ABC method will be implemented along with the metropolis-hasting algorithm.

ABC Method

The following assumptions will be made for the probability distribution of β_1 and β_2, the transmission rate parameters from susceptible to exposed.

$$\beta_1 \sim U(0,1), \beta_2 \sim U(0,1)$$

In addition, following the work of Chavez and colleagues [75] the recommended assignment of the progression transmission rates, k_1 and k_2, will lay within the range 0.00256 to 0.00527. Using this information as an approximation the probability distribution for k_1 and k_2 will take the form

$$k_1 \sim U(0.002, 0.006), k_2 \sim U(0.002, 0.006)$$

Samples using the above distributions were generated 10,000 times and the parameter set that minimised the least squares estimator was selected as the parameter set for simulation. The distribution of the transmission parameters with the least squares estimate can be seen in figure 6.3.

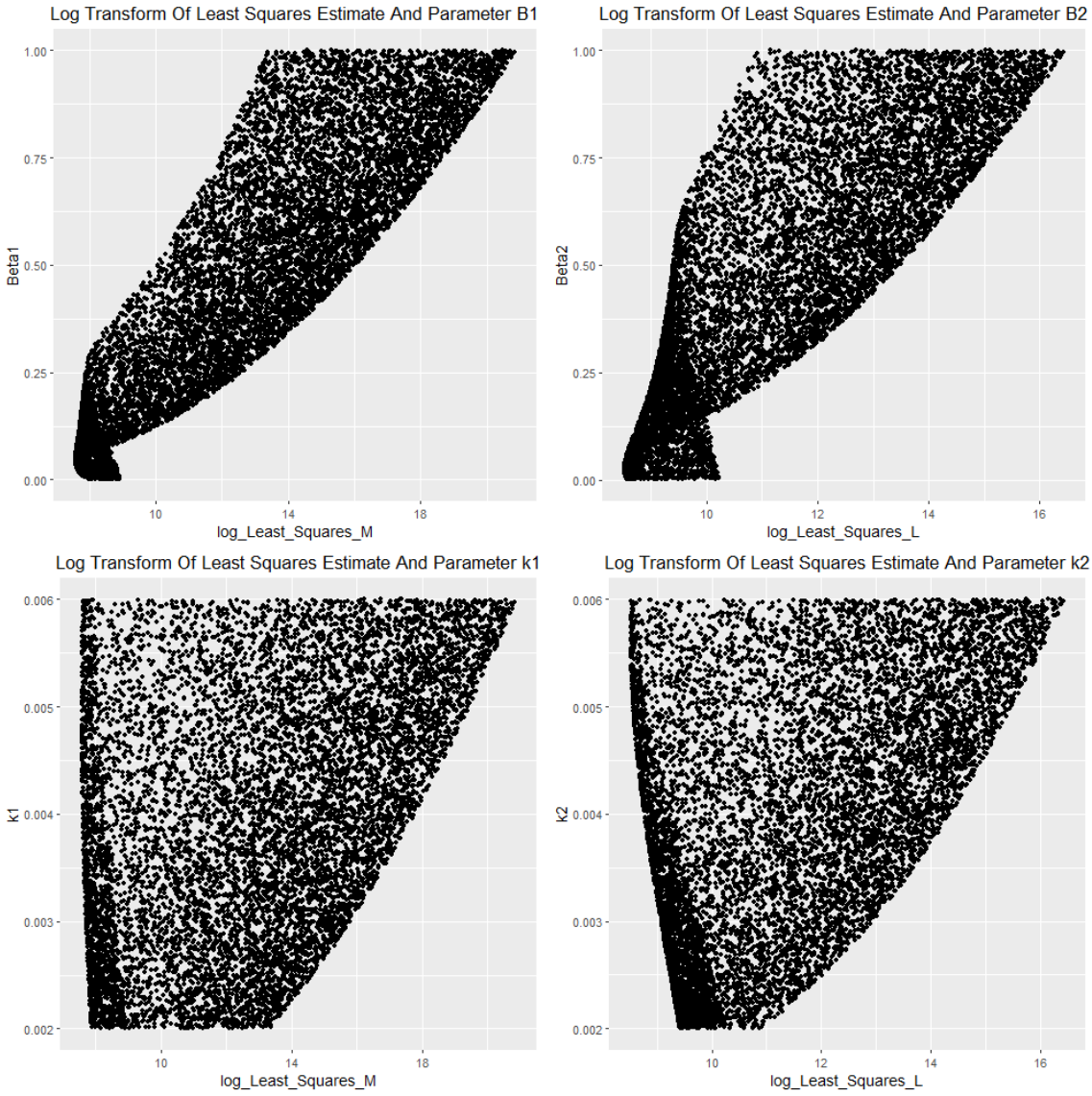

Figure 6.3: Leasts Squares Estimator for the Transmission Parameters of the Migrant Model without Interaction.

The parameter estimates that minimised the least squares estimate were $\beta_1 = 0.0514155, \beta_2 = 0.030651, k_1 = 0.0048265$, and $k_2 = 0.005630$. These results show a rough agreement occurring between the transmission parameters, although the transmission values for the

local population appear to be slightly larger than the migrant population. This could possibly be due to the number of susceptible within each population. A sensitivity analysis is completed on these parameters in §7. The Metropolis-Hastings algorithm will now be implemented on the same transmission parameters.

The Metropolis-Hastings Algorithm

As the number of dimensions increases within the parameter set, the error (least squares estimator) becomes increasingly sensitive to changes in the parameter set. As such, as all parameters are changing at once within the Metropolis-Hastings algorithm, in order for the model to accept new parameter values one must decrease the standard deviation or range of the jump distributions. This result in smaller changes being made to the parameter set at each iteration. Due to smaller changes being made, more iterations of the algorithm may be required in order for the parameters to converge. The initial parameter estimate is also important as the less the parameters have to "travel" to find convergence the quicker convergence will occur. Because of this, the estimates acquired from the ABC method in §6.4.4 will be used as initial estimates.

The Metropolis-Hastings algorithm was run with $M = 10,000$ iterations to calculate the posterior distributions of β_1, β_2, k_1, and k_2. The algorithm used initial parameter values $\beta_1 = 0.0514155, \beta_2 = 0.030651, k_1 = 0.0048265$, and $k_2 = 0.005630$. The distribution of the jump that each parameter experienced came from a uniform distribution. The range of each jump distribution varied from parameter to parameter. Each range was selected based on whether or not the proposed parameter set was being accepted an adequate number of times, having a large range for any given parameters jump distribution results in very few proposed parameters being accepted by the algorithm. The jump distribution was set as a $U(-0.05, 0.05)$ for β_1, $U(-0.005, 0.005)$ for k_2, $U(-0.01, 0.01)$ for β_2, and $U(-0.001, 0.001)$ for k_2. Figures 6.4 and 6.5 illustrate each iteration of the algorithm and the posterior distribution of the parameters after a burn-in of 2,500 iterations. Descriptive statistics are given for the posterior distributions in table 6.3.

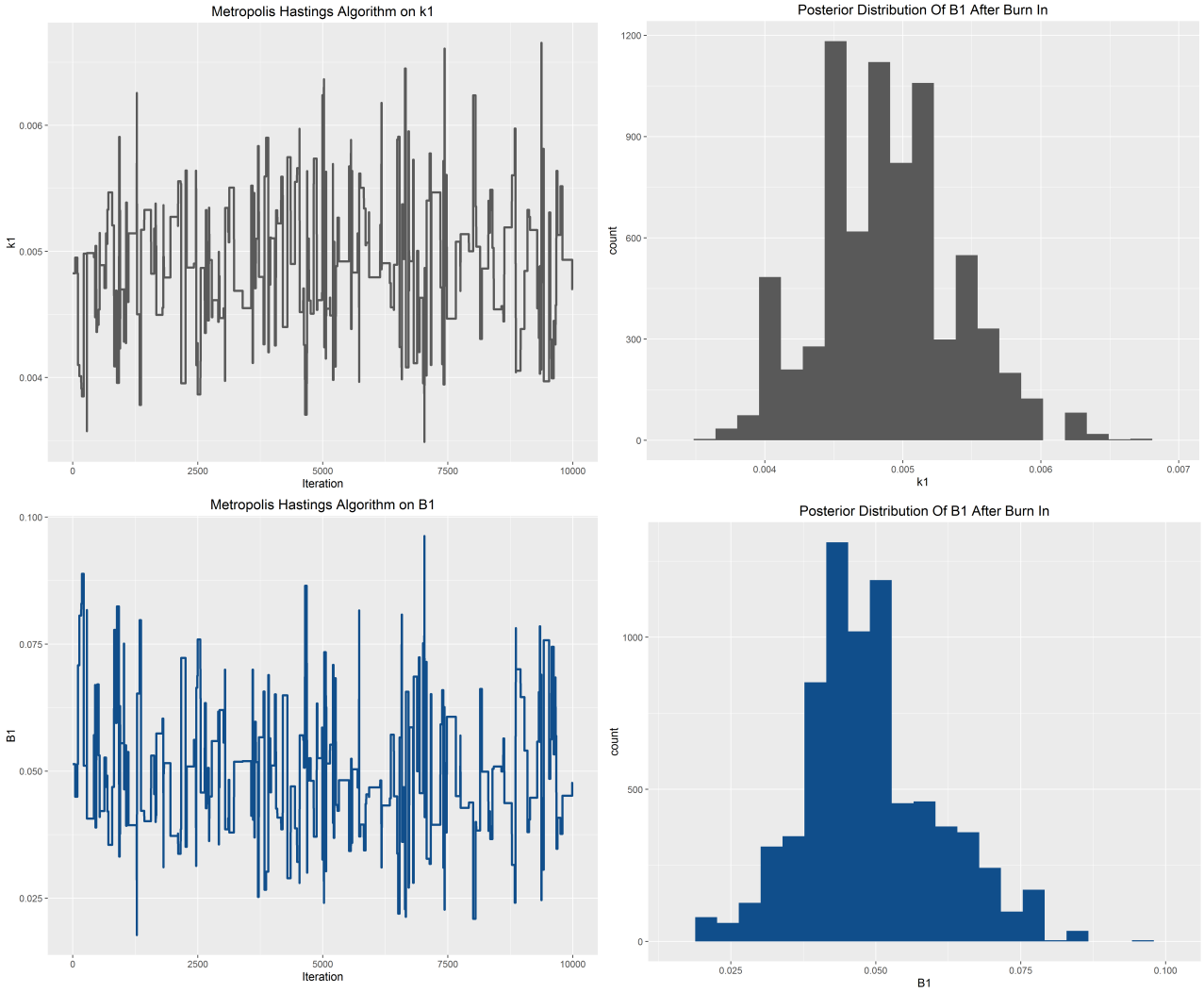

Figure 6.4: Metropolis-Hastings Algorithm Applied to the Transmission Parameters of the Migrant Population. *Left: Each Iteration of the Algorithm, Right: The Posterior Distribution after 2,500 Iterations*

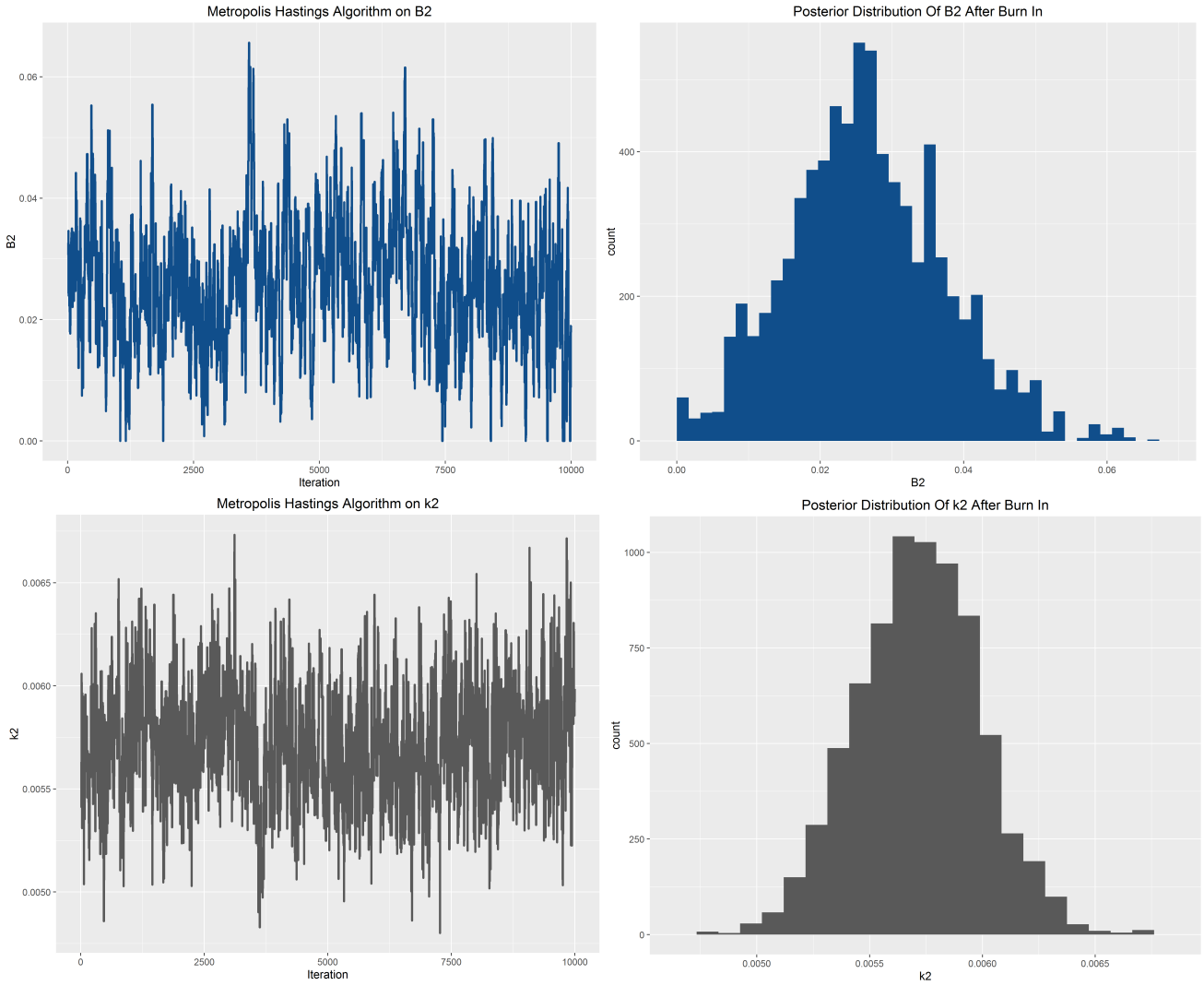

Figure 6.5: Metropolis-Hastings Algorithm Applied to the Transmission Parameters of the Local Population. *Left: Each Iteration of the Algorithm, Right: The Posterior Distribution after 2,500 Iterations*

Statistic	β_1	β_2	k_1	k_2
Mean	0.04913334	0.02644336	0.004893855	0.005715805
Median	0.04813421	0.02578991	0.004894933	0.005718408
Standard Deviation	0.01146589	0.01094568	0.0004969504	0.0002737901
Range	0.07536945	0.06562568	0.003166518	0.001931821
Skewness	0.4068527	0.2615066	0.2540014	0.04646015
2.5% Quantile	0.02802169	0.006613588	0.003970225	0.005203268
97.5% Quantile	0.07579892	0.04939109	0.005903113	0.006263031

Table 6.3: Descriptive Statistics On Transmission Parameter Distribution For the Metropolis-Hasting Algorithm on the Model with No Interaction.

The results of the Metropolis-Hastings methods are approximately the same as the ABC method, this is due to the choice of initial parameter values set for the algorithm. The transmission parameters within the migrant population have a similar posterior mean to that of the local population. The standard deviation and range of the transmission parameters for the migrant population were smaller than that of the local population, indicating a greater uncertainty for the transmission parameters on the local population than that of the migrant population.

Post-Hoc Diagnostics

The Gelman and Rubin's convergence diagnostic was calculated for $m = 3$ chains. The results for the three chains are displayed in figure 6.6 and table 6.4.

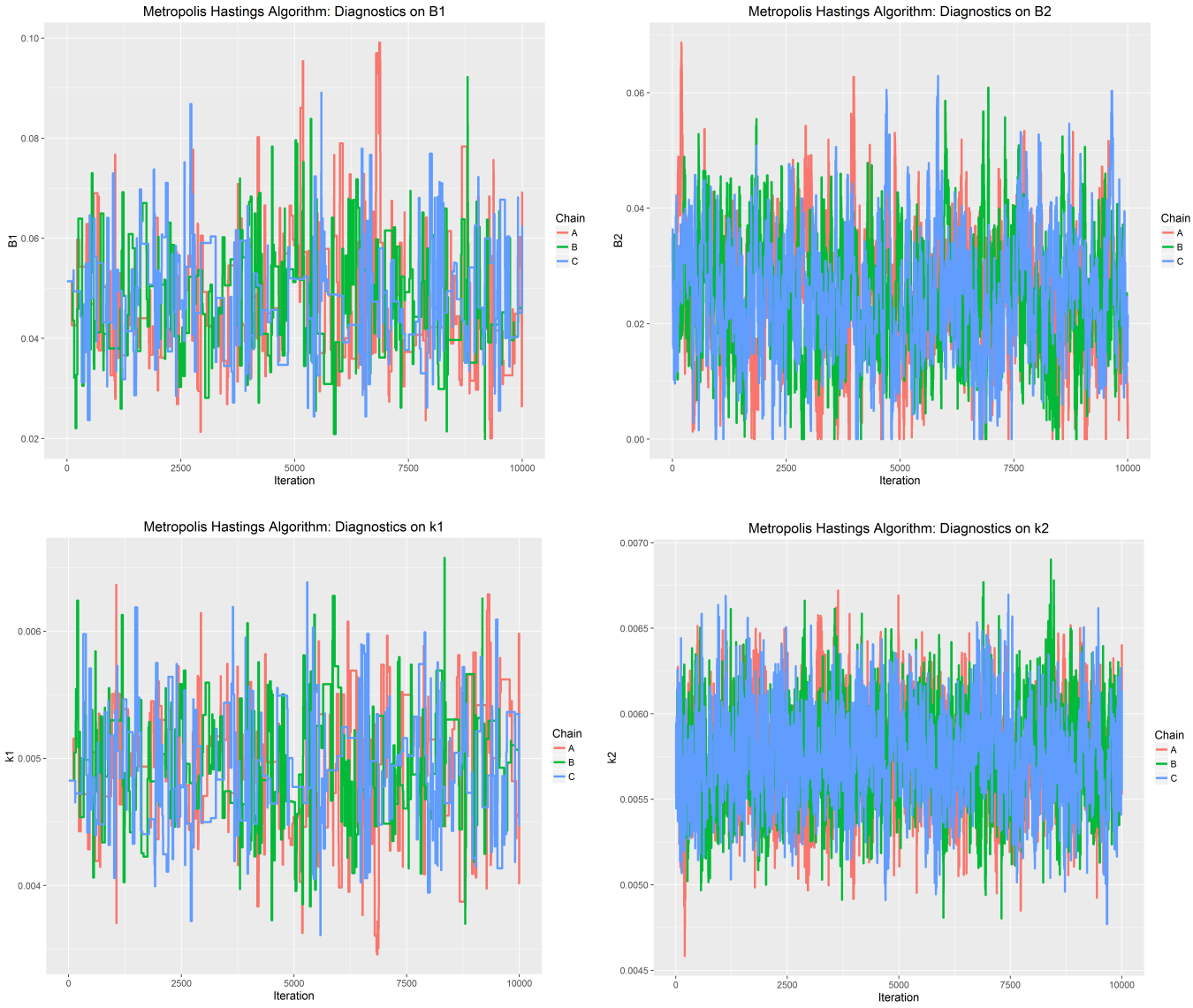

Figure 6.6: Convergence Of Three Chains For β_1, β_2, k_1, and k_2.

Parameter	Point Estimate	Upper 95% CI
β_1	1.02	1.02
k_1	1.01	1.02
β_2	1.01	1.03
k_2	1.01	1.02

Table 6.4: Potential Scale Reduction Factors For Parameters β_1, β_2, k_1, and k_2.

The upper confidence interval for the four parameter estimates are close to one, indicating convergence.

What follows is the approximation of the parameters for the model considering interaction between the migrant and local populations.

Model With Interaction Considered

The model with interaction has additional transmission parameters β_1^* (the effect of the migrant infectious on the local susceptible and exposed) and β_2^* (the effect of the local infectious on the migrant susceptible and exposed). These additional transmission rates are expected to influence the estimation of the other transmission parameters.

What follows is the estimation of these parameters using the ABC method and Metropolis-Hastings Algorithm. Figure 6.7 illustrates the least squares estimator for varying transmission parameter values.

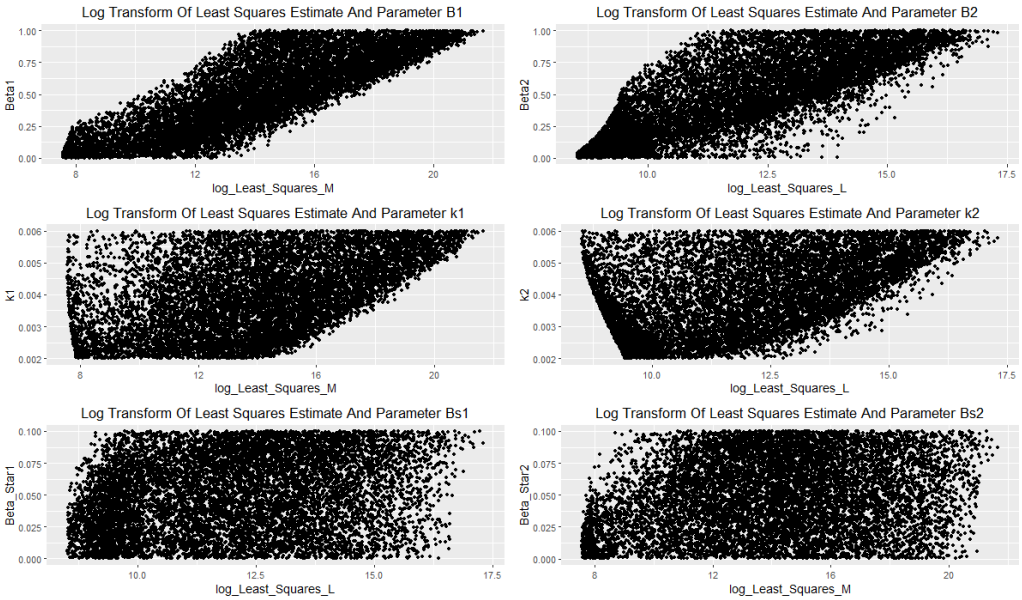

Figure 6.7: ABC Implemented on the Transmission Parameters for a Model with Interaction

The values that minimised the least square estimators were $\beta_1 = 0.02814503$, $\beta_2 = 0.02520684$, $k_1 = 0.004931357$ $k_2 = 0.005589813$, $\beta_1^* = 0.006560251$, and $\beta_2^* = 0.009016213$. An agreement was seen between the transmission parameter values for each population. However, a difference could be seen between interaction parameters. The effects of the migrant infectious compartment on the local compartments appeared to be greater than the local infectious on the migrant compartments. The Metropolis-Hastings algorithm will now be implemented to find the posterior distribution of the parameters. Figures 6.8 and 6.9 illustrate each iteration of the algorithm and the posterior distribution of the parameters after a burn-in of 5,000 iterations. Descriptive statistics are given for the posterior distributions in table 6.5.

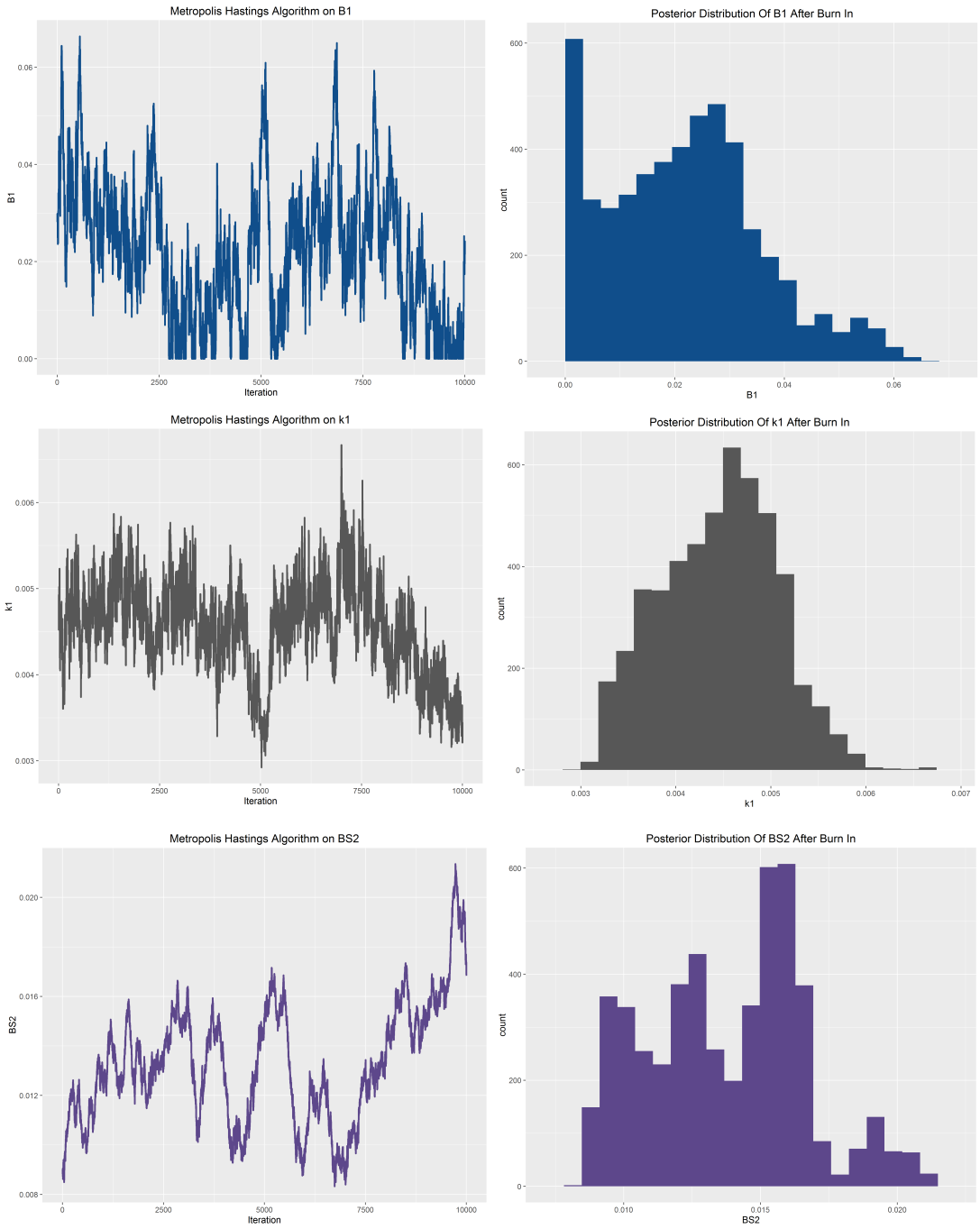

Figure 6.8: Metropolis-Hastings Algorithm and Posterior Distribution for β_1 (B1), k_1 (k1), and β_2^* (Beta_star2) given 10,000 Iterations and a Burn-in Time of 5,000 Iterations.

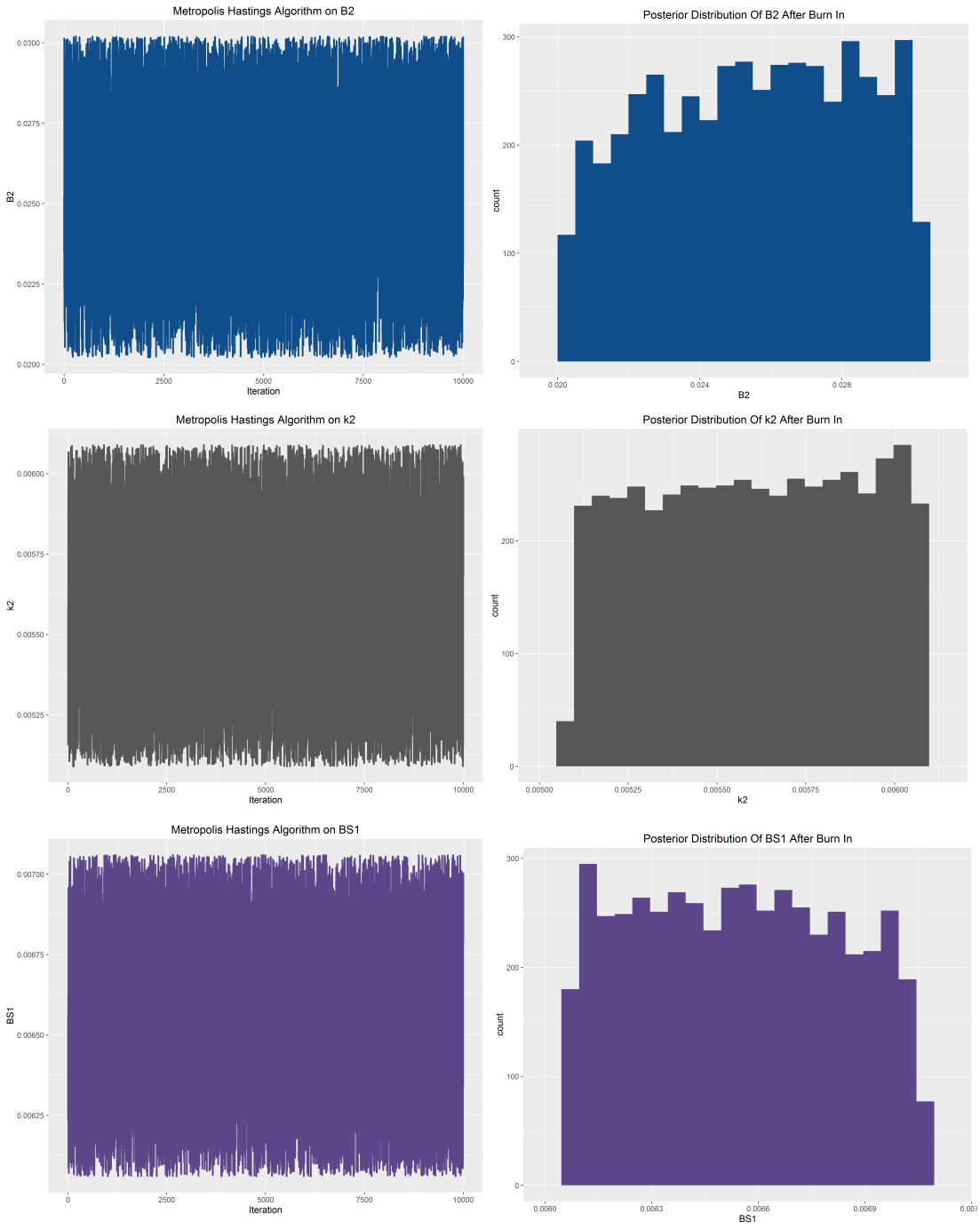

Figure 6.9: Metropolis-Hastings Algorithm and Posterior Distribution for β_2 (B2), k_2 (k2), and β_1^* (Beta_star1) given 10,000 Iterations and a Burn-in Time of 5,000 Iterations.

Statistic	β_1	β_2	k_1	k_2	β_2^*	β_1^*
Mean	0.0224	0.00464	0.02554	0.00559	0.00657	0.01279
Median	0.02248	0.00465	0.02565	0.00558	0.00658	0.01285
Standard Deviation	0.01356	0.00044	0.00278	0.00029	0.00029	0.0017
Range	0.06636	0.00259	0.00999	0.001	0.001	0.00814
Skewness	0.309	-0.18684	-0.11493	0.00704	-0.02297	-0.18363
2.5% Quantile	0	0.00368	0.02058	0.00512	0.00609	0.00963
97.5% Quantile	0.05027	0.00549	0.02998	0.00606	0.00704	0.01563

Table 6.5: Statistics for the Posterior Distribution of Transmission Variables.

The results in table 6.5 show contrasting dynamics occurring between the local and migrant population when compared to that of the parameters generated from the ABC method. The transmission rate for the migrant population is approximately the transmission rate β_2^*. This implies the local infectious compartment roughly contributes to infections within the migrant population the same as the migrant infectious class does. If the same was true for the local population, this would possibly indicate a homogeneity between the infectious compartments. However, the local population had a transmission rate approximately 10 times that of β_1^*. This is a contrasting result, suggesting the migrant infectious compartment contributes little, if at all, to local infections. This result also conflicts with the results of the ABC parameters, as little contribution was seen from the local infectious to the migrant population.

Relative to the mean, the standard deviation statistic (coefficient of variation) was largest within the transmission parameters for the local population. This indicates a larger uncertainty of the transmission parameter values within the local population.

Post-Hoc Diagnostics

The Gelman and Rubin's convergence diagnostic was calculated for $m = 3$ chains. The results for the three chains are displayed in figure 6.10 and table 6.6.

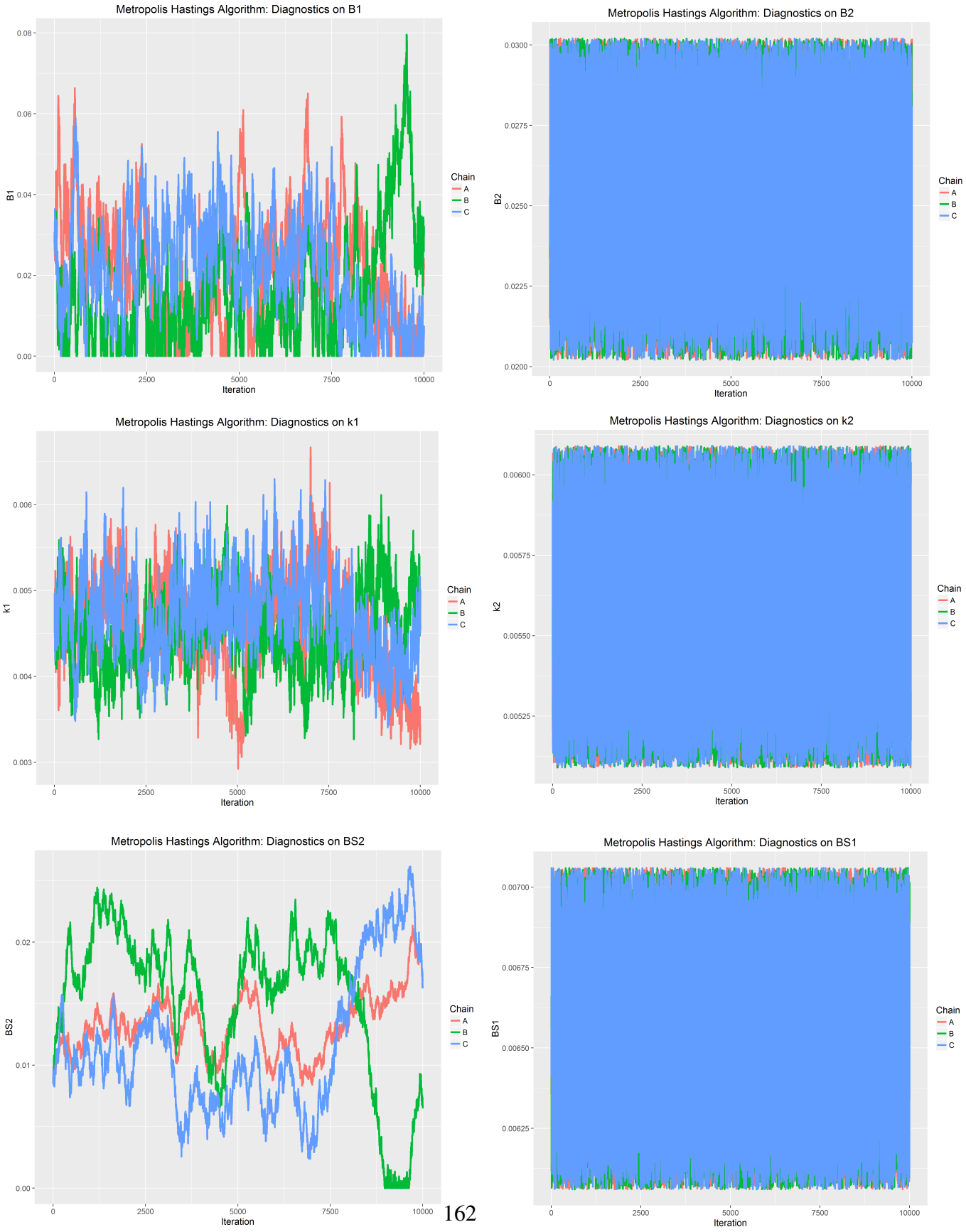

Figure 6.10: Convergence Of Three Chains For β_1, β_2, k_1, k_2, β_1^*, and β_2^*.

Parameter	Point Estimate	Upper 95% CI
β_1	1.03	1.08
k_1	1.04	1.13
β_2	1	1
k_2	1	1
β_1^*	1	1
β_2^*	1.01	1.02

Table 6.6: Potential Scale Reduction Factors For Parameters β_1, β_2, k_1, k_2, β_1^*, and β_2^*..

The upper confidence interval for the four parameter estimates are close to one, indicating convergence.

The following section simulates the models and calculate the basic reproductive number for each model, along with calculating residual statistics for each model.

6.5 Simulation And Calculation Of The Basic Reproductive Numbers

As two distinct parameter sets were generated for the models with and without interaction, the basic reproductive number and simulations will differ depending on what parameter set is used. Both parameter sets will be used which will result in two basic reproductive numbers and two simulations being calculated for both models.

6.5.1 Basic Reproductive Number

The condition for which the basic reproductive numbers were calculated for the migrant model with and without interaction such that the migrant population did not recruit individuals into the exposed class. This assumption translates to $\nu_1 = 0$ within both models.

As it was calculated that $v_1 = 0.00002$, this can be seen as a value quite close to zero. It was also calculated $\pi = 2,497$, as $v_1 \times \pi = 0.05$ it appears that the recruitment rate as a whole within the exposed compartment is small for each time step the system is iterated. As such the basic reproductive number calculated in section 6.3 should serve as reasonable estimates for the system.

The basic reproductive numbers derived in section 6.3.1 follow

Model With No Interaction

The basic reproductive number for the entire system and for the local and migrant populations is given below

$$R_0 = \max(R_{(0)M}, R_{(0)L}),$$

where

$$R_{(0)L} = \frac{\beta_2 k_2 w_1}{(k_2+\mu)(r_2+\mu+\mu_{I_L})}, \qquad R_{(0)M} = \frac{\beta_1 k_1 (1-v_1)}{(k_1+\mu)(r_1+\mu+\mu_{I_M})}.$$

The calculation of R_0, $R_{(0)L}$, and $R_{(0)M}$ for both parameter estimation methods are in table 6.7.

Estimate	ABC	Metropolis-Hastings (Distribution Mean)	Estimate	ABC	Metropolis-Hastings (Distribution Mean)
β_1	0.05141	0.04913	R_0	1.7636	1.6877
β_2	0.03065	0.0264	$R_{(0)L}$	0.2578	0.2227
k_1	0.00482	0.0049	$R_{(0)M}$	1.7636	1.6877
k_2	0.00563	0.00571			

Table 6.7: Transmission Parameter Estimates and Basic Reproductive Numbers for the Non-Interactive Model

For the model without interaction R_0 was estimated to be greater than one for both parameter estimation methods. It was observed $R_{(0)L} < 1$ implying if there is no inter-

action between migrant and local populations, then the local population is not at risk of an outbreak. It was also observed $R_{(0)M} > 1$, which implies an outbreak within the migrant population. The R_0 for population as a whole (migrant and local populations) was calculated as being the maximum between the basic reproductive number for the local population and the basic reproductive number for the migrant population. As the basic reproductive number was greater than one for the migrant population, the entire population has $R_0 > 1$. Due to the uncertainty of the transmission parameters, the percentiles of $R_{(0)L}$ and $R_{(0)M}$ follow in table 6.8.

Percentile	0%	2.5%	5%	10%	90%	95%	97.5%	100%
$R_{(0)L}$ Value	0.0000	0.035302	0.0566	0.0880	0.318	0.3534	0.391	0.521
$R_{(0)M}$ Value	0.704	1.0396	1.0707	1.1358	2.1697	2.477	2.652	3.274

Table 6.8: Uncertainty of $R_{(0)L}$ and $R_{(0)M}$ given the Uncertainty of the Transmission Parameters.

The percentile values given in table 6.8 suggest that if no interaction is occurring between migrant and local populations, the local populations is almost surely not at risk of an outbreak. By comparison the migrant population is likely undergoing an epidemic, since 98.38% of all parameter sets generated resulted in the basic reproductive number being greater than one.

Using the elements of the matrix (V^{-1}) used to calculate R_0, the calculation was made using the Metropolis-Hastings parameters that the migrant population are expected to remain exposed and infectious for 183 and 38 months, respectfully. Individuals who are exposed are expected to be infectious for 34 months.

The same applies to the local population, they are expected to remain exposed and infectious for 161 and 26 months, respectfully. Individuals who are exposed are expected to be infectious for 24 months.

The following section calculates the basic reproductive number for the model that considers an interaction.

Model With Interaction

The basic reproductive number for the system as a whole follows along with its various components.

$$R_0 = \max(\hat{\lambda_3}, \hat{\lambda_4}),$$

where

$$\hat{\lambda_3} = \frac{\lambda_3 + \lambda_4 + \sqrt{(\lambda_3 - \lambda_4)^2 + 4\lambda_3'\lambda_4'}}{2}, \qquad \hat{\lambda_4} = \frac{\lambda_3 + \lambda_4 - \sqrt{(\lambda_3 - \lambda_4)^2 + 4\lambda_3'\lambda_4'}}{2}.$$

$$\lambda_3 = \frac{\beta_2 k_2 w_1}{(k_2+\mu)(r_2+\mu+\mu_{I_L})} \qquad \lambda_4 = \frac{\beta_1 k_1 (1-v_1)}{(k_1+\mu)(r_1+\mu+\mu_{I_M})}$$

$$\lambda_3' = \frac{\beta_2^* k_2 (1-v_2)}{(k_2+\mu)(r_2+\mu+\mu_{I_L})} \qquad \lambda_4' = \frac{\beta_1^* k_1 w_1}{(k_1+\mu)(r_1+\mu+\mu_{I_M})}$$

The estimates for $R_0, \hat{\lambda_3}, \hat{\lambda_4}, \lambda_3, \lambda_4, \lambda_3'$, and λ_4' follow in table 6.9 for both parameter estimation methods.

Estimate	ABC Method	Metropolis-Hastings (Distribution Mean)	Estimate	ABC Method	Metropolis-Hastings (Distribution Mean)
β_1	0.02814	0.01886	R_0	0.697	0.5132
β_2	0.0252	0.02548	$\hat{\lambda_3}$	0.697	0.5132
k_1	0.00493	0.00447	$\hat{\lambda_4}$	0.1827	0.14411
k_2	0.0055	0.0056	λ_3	0.2118	0.2142
β_1^*	0.00656	0.0065532	λ_4	0.6679	0.4431
β_2^*	0.00901	0.01354	λ_3'	0.179068	0.26895
			λ_4'	0.07878	0.07791

Table 6.9: The Values of the Transmission Parameters, and Calculation of R_0 for the System

The basic reproductive number was calculate to be less than one for both parameter estimation methods, implying no epidemic will occur within the population. Due to the

uncertainty of the transmission parameters, the percentiles of R_0 are given in table 6.10.

Percentile	0%	2.5%	5%	10%	90%	95%	97.5%	100%
R_0 Value	0.2406	0.280	0.2951	0.315	0.955	1.096	1.2496	1.5448

Table 6.10: Percentiles of R_0 given the Transmission Parameter Values

Table 6.8 calculates R_0 given the posterior distribution of the transmission parameters obtained from the Metropolis-Hastings algorithm. For approximately 8.81% of transmission parameter sets generated after burn-in time, the basic reproductive number was greater than one. This implies if the model with interaction is a feasible model, and the parameter sets and initial conditions are also feasible, then there exists a possibility there is an outbreak occurring within the population as a whole.

Using the elements of the matrix (V^{-1}) used to calculate R_0, the calculation was made using the Metropolis-Hastings parameters that the migrant population are expected to remain exposed and infectious for 198 and 38 months, respectfully. Individuals who are exposed are expected to be infectious for 33 months.

The same applies to the local population, they are expected to remain exposed and infectious for 162 and 26 months, respectfully. Individuals who are exposed are expected to be infectious for 9 months.

6.5.2 Simulation

Simulation will now be carried on both models considering and not considering an interaction. The different transmission parameter sets are also simulated. Figure 6.11 simulates each transmission parameter set for the model without interaction, and Figure 6.12 simulates each transmission parameter set for the model with interaction.

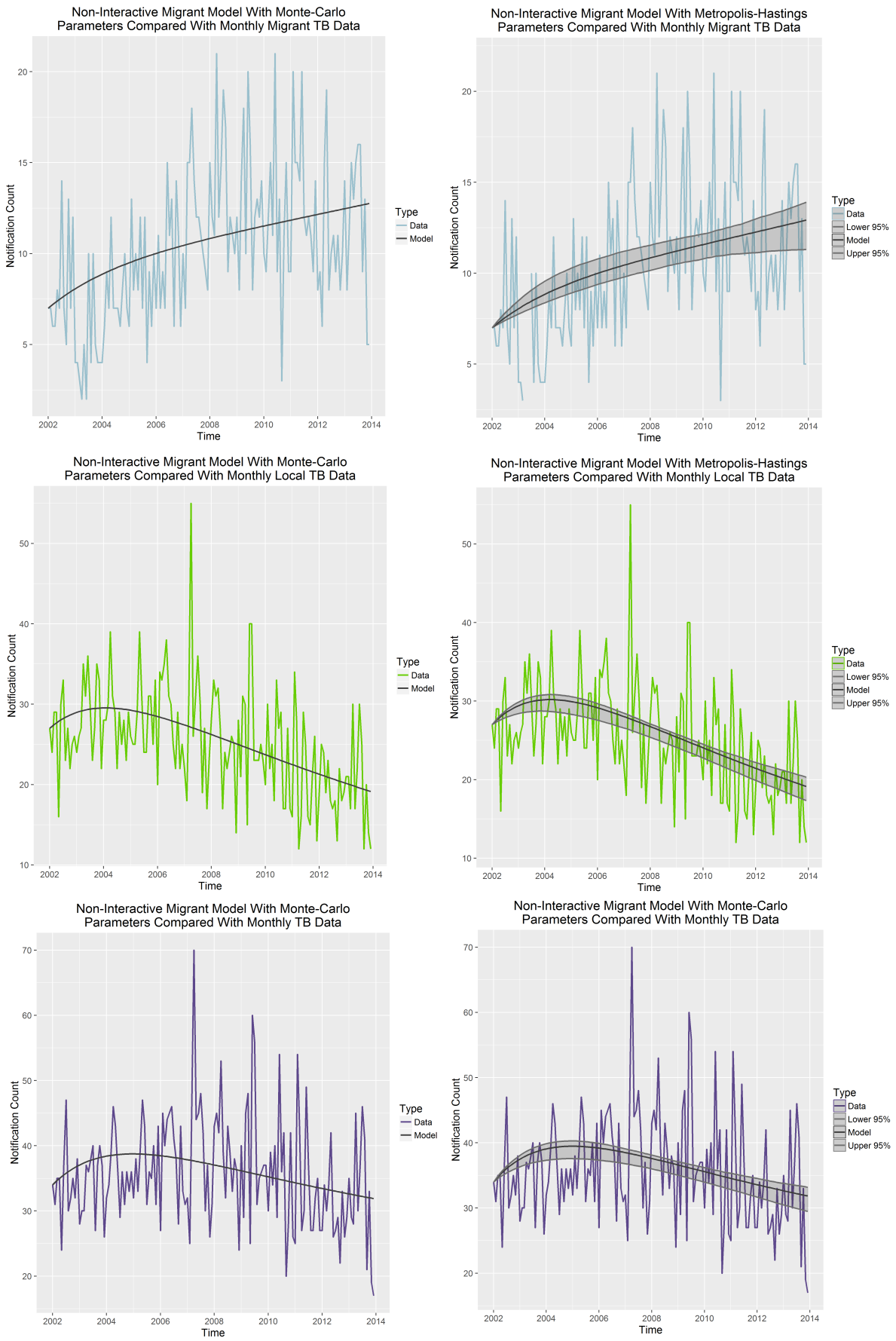

Figure 6.11: The Non-Interactive Migrant Model Simulation with the ABC Method Parameters and Metropolis-Hastings Parameters

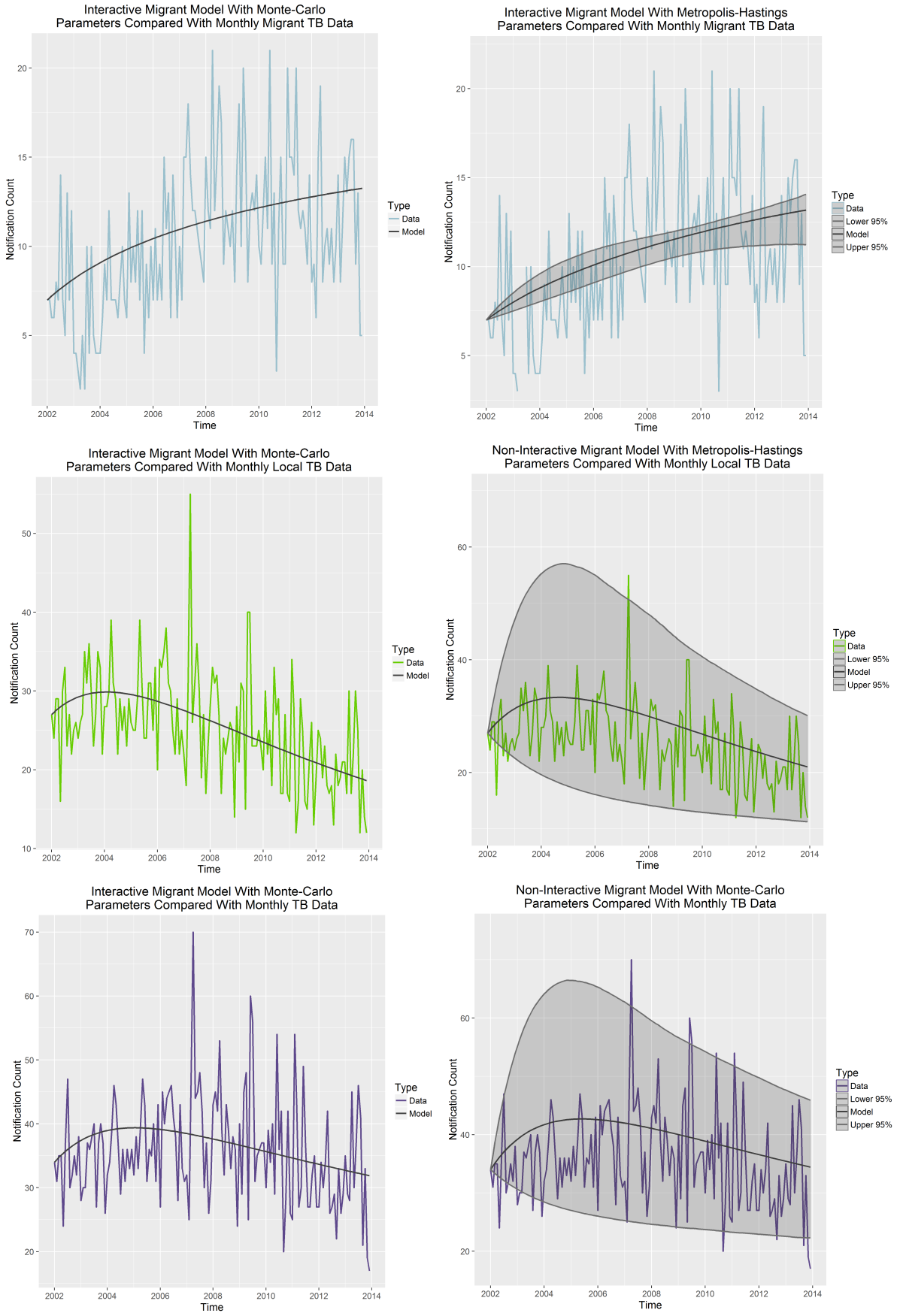

Figure 6.12: The Interactive Migrant Model Simulation with the ABC Method Parameters and Metropolis-Hastings Parameters

6.5.3 Various Model Residuals

Figure 6.13 and Table 6.11 display information on the residuals of both migrant models for the ABC and Metropolis-Hastings methods.

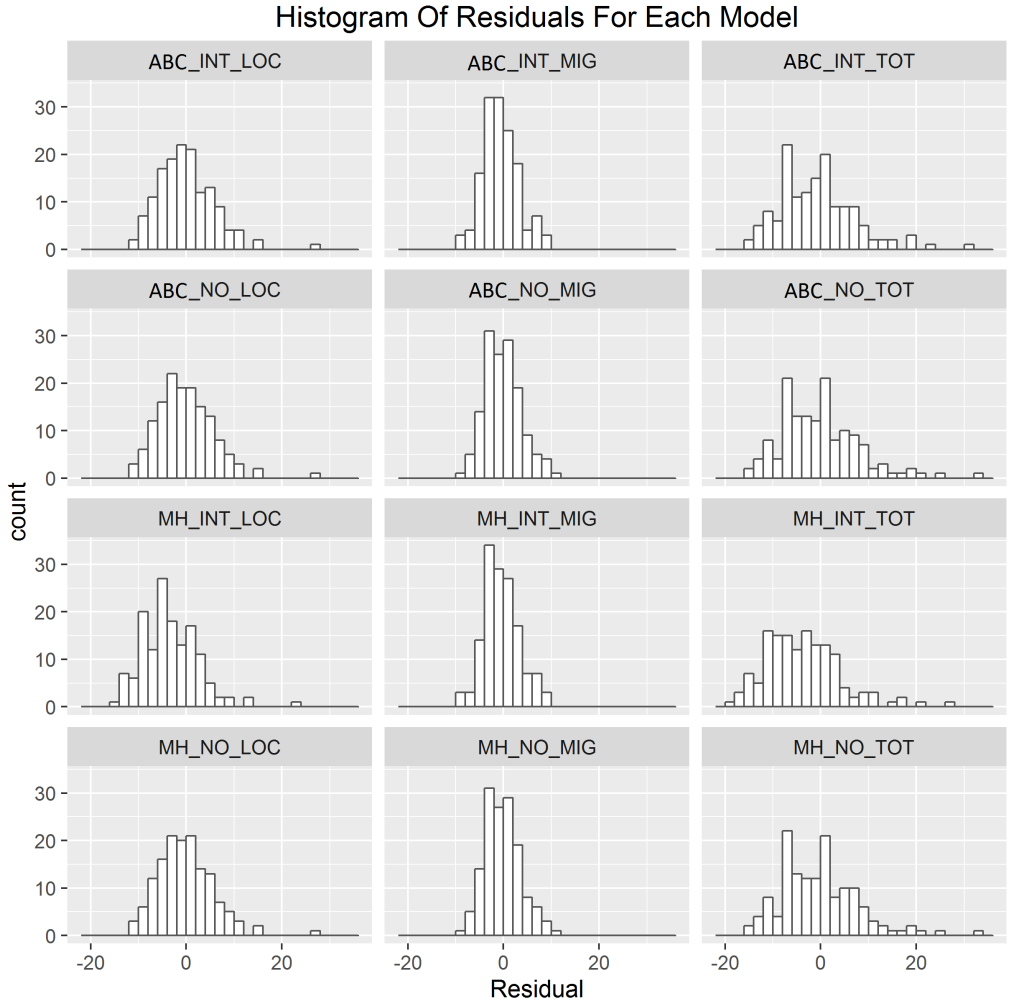

Figure 6.13: Distribution Of Residuals For Each Model. *Model Type Abbreviations: ABC=Approximate Bayesian Computation, MH = Metropolis-Hastings, NO = No Interaction, INT = Interaction, MIG = Migrant, LOC = Local, TOT = Total Population*

Model Type (Abreviated)	Mean	Stand. Dev.	0% Quantile	25% Quantile	50% Quantile	75% Quantile	100% Quantile	Skew.	Kurt.
ABC_NO_MIG	-0.173	3.689	-8.731	-2.881	-0.623	2.342	10.083	0.454	0.100
ABC_NO_LOC	-0.120	5.916	-11.864	-4.173	-0.588	3.630	27.871	0.962	2.728
ABC_NO_TOT	-0.293	7.997	-14.895	-6.263	-0.545	4.541	32.330	0.939	1.689
MH_NO_MIG	-0.209	3.688	-8.805	-2.910	-0.646	2.277	10.062	0.447	0.110
MH_NO_LOC	-0.175	5.915	-11.906	-4.320	-0.642	3.508	27.767	0.962	2.711
MH_NO_TOT	-0.384	7.997	-14.977	-6.371	-0.626	4.429	32.223	0.940	1.688
ABC_INT_MIG	-0.635	3.650	-9.382	-3.447	-0.954	1.861	9.489	0.441	0.107
ABC_INT_LOC	-0.082	5.926	-11.975	-4.472	-0.772	3.573	27.783	0.963	2.669
ABC_INT_TOT	-0.716	7.981	-14.994	-6.693	-1.002	3.912	31.710	0.942	1.675
MH_INT_MIG	-0.453	3.651	-9.376	-3.140	-0.745	1.832	9.740	0.408	0.140
MH_INT_LOC	-3.342	5.906	-14.716	-7.485	-4.369	0.373	23.715	0.876	2.303
MH_INT_TOT	-3.795	7.923	-18.275	-9.796	-4.442	0.471	27.993	0.896	1.577

Table 6.11: Residual Statistics Of Each Model Fit. *Model Type Abbreviations: ABC = Approximate Bayesian Computation, MH = Metropolis-Hastings, NO = No Interaction, INT = Interaction, MIG = Migrant, LOC = Local, TOT = Total Population*

Although the model residuals were roughly centred around zero (evident from the mean and median values in table 6.11), each simulation appeared to slightly over estimate notifications. This is evident by the consistent negative mean (and median) calculated for the residuals. The worst performing simulation was the Metropolis-Hastings simulation for the local population on the model considering an interaction (mean residual of -3.24). For this simulation the value of $\beta_1^* = 0.0051$, the interaction of the migrant population on the local population, was approximately one fifth of the value estimated through the ABC method. The value of $\beta_2 = 0.06$, the transmission rate for the local population, was approximately double the value estimated through the ABC method. Due to the small size of β_1^* this is equivalent to considering no effect occurring between the migrant infectious on the local population. However, there was a greater contribution of the local infectious population to the migrant population, which is a claim yet to be supported by literature. This result taking with the inferior residual statistic leads to the conclusion that the Metropolis-Hastings algorithm did not perform as well as the ABC method for the model considering

an interaction.

The standard deviation of the residuals for the local population appeared to be consistently larger than the migrant residuals. This indicates the data was more volatile relative to the model for the local infectious than that of the migrant infectious. This may be related to the results of chapter 4 section 4.3, within which the average number of monthly notifications were shown to significantly positively correlate to the standard deviation of cases (e.g. if the last 10 years, the month April had a large average notifications, then April would also likely be a month of large fluctuations in notifications). As the local population has a larger count of infections, this may result in a large standard deviation of notifications, and hence this would contribute to the standard deviation of the residual statistic.

The models are now extrapolated forward, figures and data tables will present the extrapolations.

6.5.4 Model Extrapolation

To extrapolate the migrant model is to presume no external factors or interventions will occur within the population. The following (figures 6.14-6.15, tables 6.12-6.15) are annual data extrapolated forward 10 years using both parameter sets generated.

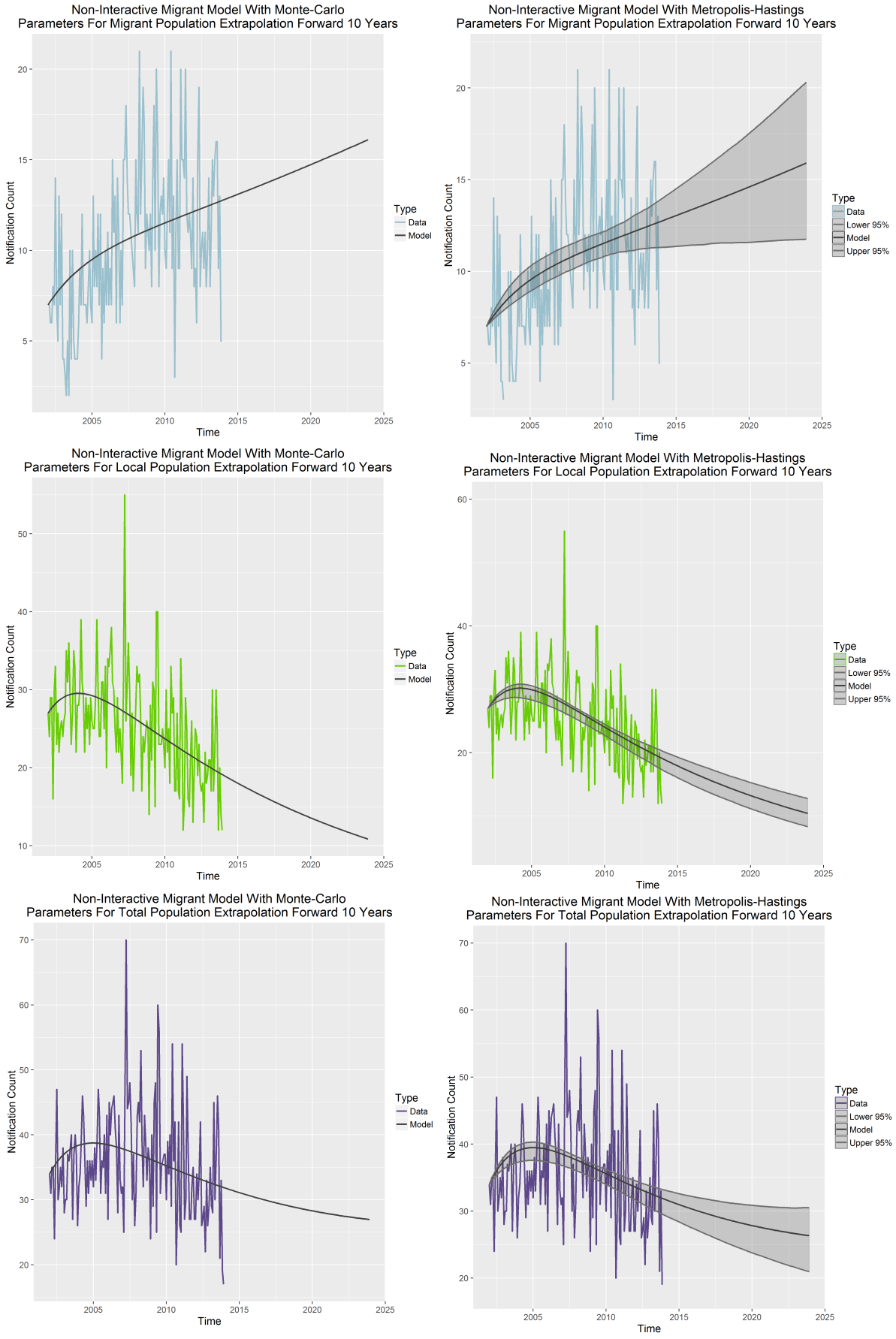

Figure 6.14: Migrant Model Extrapolation Not Considering Interaction for the ABC and Metropolis-Hastings Parameters

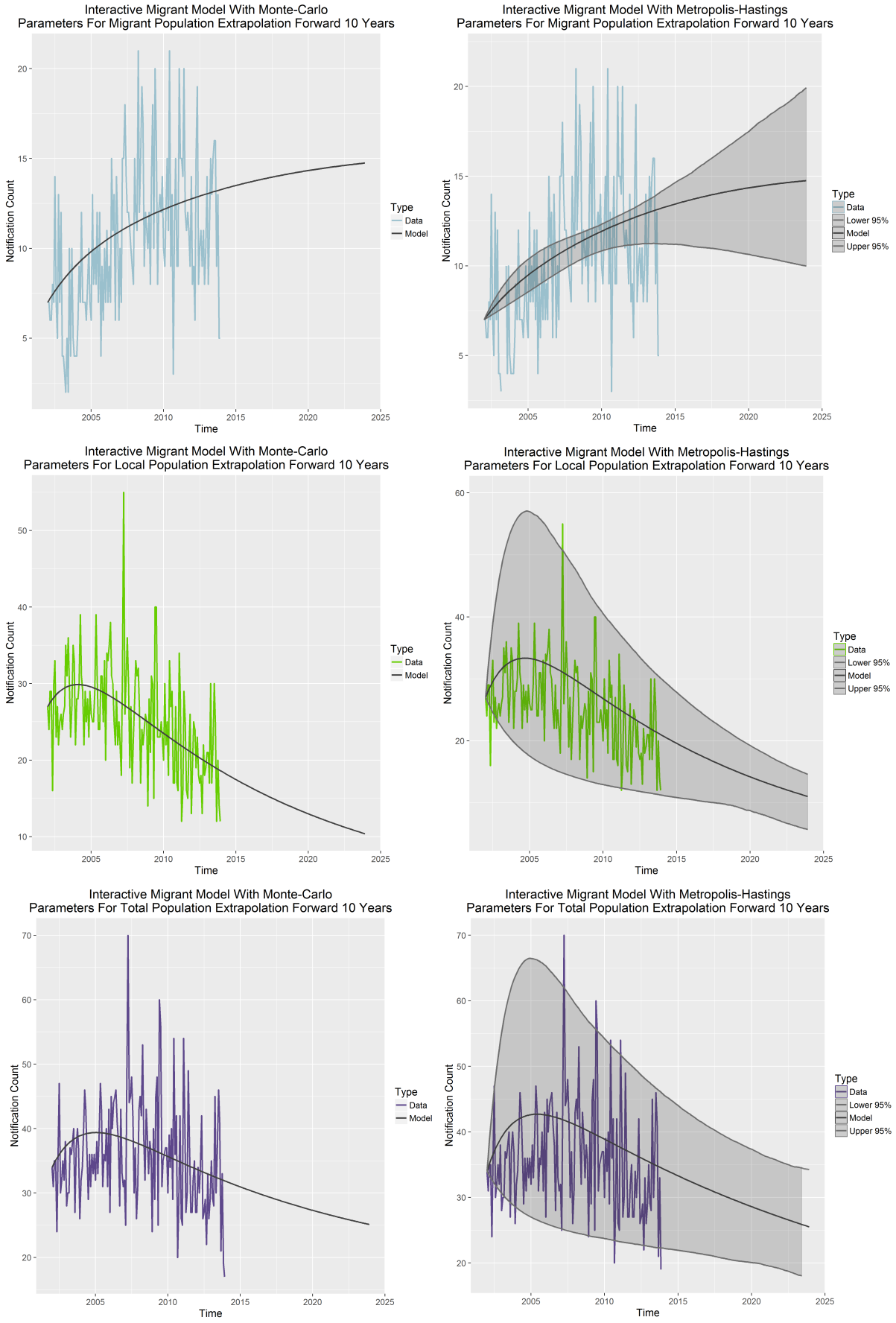

Figure 6.15: Migrant Model Extrapolation Considering Interaction for the ABC and Metropolis-Hastings Parameters

The extrapolated data for each compartment from the models follow in tables 6.12-6.15.

Time	No Interaction Migrant Model Population Estimates: ABC Parameters							
	Susceptible Migrant	Exposed Migrant	Infectious Migrant	Recovered Migrant	Susceptible Local	Exposed Local	Infectious Local	Recovered Local
2002	244400	697	90	70976	548978	2590	337	3108944
2003	265933	703	101	77230	553922	2438	352	3136992
2004	287324	712	110	83444	558832	2297	353	3164855
2005	308575	725	117	89616	563711	2167	347	3192535
2006	329685	739	123	95749	568557	2045	336	3220032
2007	350657	755	127	101841	573372	1930	323	3247349
2008	371490	772	132	107893	578155	1822	308	3274485
2009	392186	790	136	113905	582907	1721	293	3301443
2010	412746	809	140	119878	587628	1626	278	3328223
2011	433170	828	144	125811	592318	1535	263	3354827
2012	453460	847	148	131706	596977	1450	249	3381256
2013	473616	867	151	137562	601605	1370	236	3407510
2014	493640	888	155	143379	606204	1294	223	3433592
2015	513531	909	159	149158	610772	1223	211	3459502
2016	533292	930	163	154899	615310	1155	199	3485241
2017	552922	951	167	160603	619818	1091	188	3510811
2018	572424	973	171	166269	624297	1031	178	3536212
2019	591796	996	175	171898	628747	974	168	3561447
2020	611042	1019	179	177490	633167	920	159	3586515
2021	630160	1042	183	183045	637558	869	150	3611418
2022	649153	1065	187	188564	641921	821	142	3636157
2023	668020	1089	191	194046	646254	775	134	3660733

Table 6.12: Non-Interactive Migrant Model Population Estimates With ABC Parameters. **Note:** Susceptible and Recovered Populations Are Year End Estimates, Exposed and Infectious Populations are Year Total Estimates.

Time	No Interaction Migrant Model Population Estimates: Metropolis-Hastings Parameters							
	Suscept. Migrant	Exposed Migrant	Infectious Migrant	Recov. Migrant	Suscept. Local	Exposed Local	Infectious Local	Recov. Local
2002	244399	698	90 (88,93)	70976	548978	2589	337 (333,342)	3108944
2003	265933	705	101 (97,108)	77230	553922	2433	353 (343,365)	3136992
2004	287324	717	110 (104,119)	83444	558833	2290	355 (343,369)	3164855
2005	308574	731	117 (109,126)	89616	563711	2156	349 (336,363)	3192535
2006	329685	747	122 (115,132)	95749	568558	2032	338 (325,350)	3220032
2007	350656	765	128 (120,136)	101841	573373	1915	324 (313,334)	3247349
2008	371489	784	132 (124,140)	107893	578156	1806	309 (298,318)	3274485
2009	392185	804	137 (128,144)	113905	582908	1703	293 (282,302)	3301443
2010	412744	824	141 (131,148)	119878	587629	1606	278 (265,287)	3328223
2011	433168	845	145 (133,153)	125811	592319	1515	263 (248,274)	3354827
2012	453458	867	149 (134,158)	131706	596978	1429	249 (231,261)	3381256
2013	473614	890	153 (135,164)	137562	601607	1348	235 (215,249)	3407510
2014	493637	912	157 (136,170)	143379	606205	1272	222 (200,238)	3433592
2015	513529	936	161 (137,177)	149158	610774	1200	209 (186,227)	3459502
2016	533289	959	166 (137,184)	154900	615312	1132	197 (173,217)	3485241
2017	552919	984	170 (138,191)	160603	619820	1067	186 (161,207)	3510811
2018	572420	1008	174 (139,199)	166269	624299	1007	176 (150,197)	3536212
2019	591792	1034	179 (139,206)	171898	628749	950	166 (140,188)	3561447
2020	611037	1059	183 (139,214)	177490	633169	896	156 (130,180)	3586515
2021	630155	1086	188 (140,222)	183046	637560	845	148 (120,172)	3611418
2022	649147	1112	193 (140,231)	188564	641923	797	139 (112,164)	3636157
2023	668014	1140	198 (141,240)	194047	646257	752	131 (104,157)	3660733

Table 6.13: Non-Interactive Migrant Model Population Estimates With Metropolis-Hastings Parameters. **Note:** Susceptible and Recovered Populations Are Year End Estimates, Exposed and Infectious Populations are Year Total Estimates.

Time	Migrant Interactive Model Population Estimates: ABC Parameters							
	Susceptible Migrant	Exposed Migrant	Infectious Migrant	Recovered Migrant	Susceptible Local	Exposed Local	Infectious Local	Recovered Local
2002	244399	701	91	70976	548979	2583	338	3108944
2003	265932	713	103	77230	553923	2417	355	3136992
2004	287322	729	113	83444	558835	2264	357	3164855
2005	308572	746	121	89616	563714	2122	350	3192535
2006	329682	763	128	95749	568561	1991	338	3220033
2007	350654	780	134	101841	573377	1869	323	3247349
2008	371487	796	139	107893	578161	1755	307	3274486
2009	392183	811	144	113905	582913	1649	291	3301443
2010	412743	825	148	119878	587634	1550	275	3328223
2011	433168	838	151	125812	592325	1458	259	3354827
2012	453458	850	155	131707	596984	1371	244	3381256
2013	473615	861	158	137563	601613	1291	230	3407510
2014	493639	871	161	143380	606211	1215	216	3433592
2015	513532	880	163	149159	610779	1145	204	3459502
2016	533294	888	165	154901	615318	1079	192	3485241
2017	552926	896	167	160604	619826	1018	181	3510810
2018	572428	902	169	166270	624305	960	171	3536212
2019	591803	909	171	171899	628754	906	161	3561446
2020	611050	914	173	177491	633175	856	152	3586514
2021	630170	919	174	183046	637566	809	143	3611417
2022	649165	923	175	188564	641928	765	135	3636156
2023	668035	927	176	194047	646262	724	128	3660732

Table 6.14: Migrant Interactive Model Population Estimates With ABC Parameters. **Note:** Susceptible and Recovered Populations Are Year End Estimates, Exposed and Infectious Populations are Year Total Estimates.

Time	Migrant Interactive Model Population Estimates: Metropolis-Hastings Parameters							
	Suscept. Migrant	Exposed Migrant	Infectious Migrant	Recov. Migrant	Suscept. Local	Exposed Local	Infectious Local	Recov. Local
2002	244398	709	89 (87,93)	70975	548978	2574	352 (298,450)	3108944
2003	265929	740	99 (93,109)	77230	553921	2391	389 (253,618)	3136993
2004	287319	774	108 (99,120)	83443	558831	2227	402 (223,678)	3164857
2005	308567	809	116 (106,128)	89616	563709	2077	399 (202,675)	3192537
2006	329676	841	123 (112,134)	95748	568555	1940	388 (187,641)	3220035
2007	350646	871	130 (118,138)	101840	573369	1813	372 (176,600)	3247353
2008	371479	898	136 (124,142)	107893	578152	1695	353 (166,555)	3274490
2009	392174	922	142 (128, 146)	113905	582903	1585	334 (159,508)	3301448
2010	412734	942	147 (131,150)	119878	587623	1484	315 (153,470)	3328229
2011	433158	960	152 (133,155)	125811	592313	1388	296 (147,436)	3354833
2012	453448	974	156 (134,160)	131706	596971	1300	278 (142,403)	3381262
2013	473605	986	160 (135,166)	137562	601600	1217	261 (137,374)	3407517
2014	493629	996	163 (135,172)	143380	606198	1139	244 (133,348)	3433599
2015	513522	1003	166 (134,178)	149159	610765	1067	229 (128,323)	3459509
2016	533284	1009	168 (133,185)	154900	615303	999	214 (124,300)	3485249
2017	552916	1012	170 (132,192)	160604	619811	936	201 (120,278)	3510819
2018	572420	1014	172 (130,199)	166270	624290	877	188 (116,259)	3536220
2019	591795	1014	173 (129,206)	171899	628739	822	176 (109,240)	3561454
2020	611042	1013	174 (127,214)	177491	633159	770	165 (100,224)	3586522
2021	630163	1010	175 (125,221)	183046	637551	722	155 (91,208)	3611425
2022	649159	1007	176 (123,228)	188564	641913	677	145 (81,194)	3636165
2023	668029	1002	177 (121,235)	194047	646246	635	136 (72,181)	3660741

Table 6.15: Migrant Interactive Model Population Estimates With Metropolis-Hastings Parameters. **Note:** Susceptible and Recovered Populations Are Year End Estimates, Exposed and Infectious Populations are Year Total Estimates.

For the model not considering an interaction (table 6.14, 6.15), the resulting data indicates an annual increase in the infectious migrant population, on average increasing four cases per year after 2013. The local population is estimated to decline on average 10 cases per year after 2013. The number of exposed are expected to increase approximately

22 cases a year after the year 2013 for the migrant population and decrease 58 per year for the local population. These averages do not differ for the ABC parameter set and Metropolis-Hasting parameter set. These data are consistent with the basic reproductive numbers calculated for the model without interaction.

For the model considering an interaction, the ABC parameters resulted in data indicate an increase of two cases per year within the migrant population after the year 2013, and decrease of 11 cases per year for the local population. The number of exposed is expected to increase seven per year for the migrant population and decrease 58 cases per year for the local population. The Metropolis-Hasting parameters resulted in similar data to the ABC parameter data, after 2013 the migrant infectious population is expected to increase 2 per year, while the local infectious are expected to decrease 13 per year. The migrant exposed population are expect to increase two per year and the local exposed are expected to decrease 60 per year. As mentioned in section 6.5.3, the Metropolis-Hasting parameters for the model without interaction are suspected to be empirically inaccurate, so the later projections may possibly be inaccurate.

6.6 Conclusion

This chapter examined and simulated two homogeneous migrant tuberculosis ODE models. The models ware adapted to simulate Irish data. This was achieved by dividing the recruitment parameter between the susceptible and recovered compartments. The limitations and assumptions of the model were detailed, including the assumption of slow and fast progression to the infectious compartment. A theoretical qualitative analysis was conducted on the model detailing the disease-free and endemic equilibrium states. The basic reproductive number was calculated. Due to it not being theoretically possible to calculate the basic reproductive number without the model harbouring a disease-free equilibrium, assumptions were made to allow for the equilibrium to exist. The basic reproductive number of each model was then derived. Parameters were estimated using a combination of assumptions, literature, data, and statistical inference methods. A total of 16 parameters

required estimating and the study systematically proposed rational estimates. The recruitment rate, death rate, disease induced death rate, and recovery rate were calculated using national vital statistics and the national TB data set. The progression rate, the fast progression parameter were acquired from literature. The proportion of the recruitment rate entering into the susceptible compartment was assumed to be in line with national vaccination. The remaining transmission parameters were estimated using both an Approximate Bayesian Computation method and the Metropolis-Hastings algorithm, both methods estimated similar parameter values indicating one viable parameter set. The Recovered and Exposed compartments had initial conditions that were derived under the assumption of national vaccination coverage. The initially infected compartment was acquire from the national data set, and by deduction, the Susceptible population calculated. The basic reproductive numbers calculated were found to be less than one for the model considering an interaction between migrant and local populations. For the model not considering an interaction, the basic reproductive number was found to be less than one for the local population and greater than one for the migrant population. Simulation of the model was then carried out to observe the underlying dynamics over time, and the model was extrapolated 10 years into the future. Given the extrapolation, the number of infectious is expected to decline nationally each year. The decline in cases was seen in both models.

The following chapter conducts a sensitivity and scenario analysis of the models presented thus far.

Chapter 7

Sensitivity And Scenario Analysis

7.1 Introduction

Sensitivity analysis is the study of how the uncertainty in the output of a mathematical model can be apportioned to different sources of uncertainty in its inputs. The simultaneous estimation of several parameters raises questions of parameter identifiability [134], even if the model being fitted is simple. Often times, parameter estimates are correlated: the values of two or more parameters cannot be estimated independently. In this chapter the study aims to implement a sensitivity analysis to quantify the uncertainties associated with parameter estimates.

The models and parameter sets that have been developed in §5 and §6 will be examined. A specific methodology is carried out on the parameters and various correlation statistics calculated. The chapter then goes on to conduct two scenario analyses. Additional analysis is conducted and the basic reproductive numbers are updated to accommodate the scenarios. Recommendations are given and results of the scenarios discussed.

7.2 Sensitivity Analysis

The methodology used to conduct the sensitivity analysis will be the methods described in the work of Marino and colleagues [135]. The methodology follows the following steps.

1. Construct mathematical model $\mathfrak{M}_t(\theta)$ dependent on parameter set $\theta \in \mathbb{R}^N$.

2. Specify a probability distribution for each parameter then generate M Latin Hypercube samples for each of them. This results in an $M \times N$ matrix of parameters which we will denote X. Each row of X is a viable parameter set that can be used within the model, and each of the columns are the samples generated for each parameter.

3. For each of the M parameter sets generated, simulate the model $\mathfrak{M}_t(X_i)$, for $i = 1,..,M$.

4. Using the M models generated calculate some single dimensioned output vector y. The definition of y_i for $i = 1,2,...,M$ is usually something that is of importance within the model, however must be a single value for each i (e.g. y_i= number of infectious within the population at $t = 100$, or y_i = the sum of all exposed individuals over the time period of the study).

5. **Result 1:** Calculate Pearson's correlation coefficient (CC_P) for each of the columns of X and y.

6. Numerically rank the columns of X and denote the new matrix X_R and also numerically rank y and denote this new vector y_R.

7. **Result 2:** Calculate Spearman's correlation coefficient (CC_S) and the partial rank correlation coefficient ($PRCC$) between each of the columns of X_R and y_R.

The reasoning behind this methodology is the following: Latin hypercubes are sampled from as they ensure stratified sampling, which ensures a representative sample is taken from each distribution. Other methodology could be used here, such as generating samples from low discrepancy sequences (e.g. The Sobol sequence [153]). Various correlation

coefficients are calculated as they measure different relationships between the variables, for each correlation (Pearson's correlation, ranked correlation, partial rank correlation), the linear relationship between parameter and output variations can become more or less apparent and the various correlations highlight this. Pearson's correlation coefficient is calculated as it works well for measuring linear relationship measures between two variables. It is, however, not robust to outliers, and this is a recognised limitation of this calculation. For non-linear relationships but monotonic ranked transforms work well, such as Spearmans's correlation coefficient and the Partial ranked correlation coefficient. The Partial ranked correlation coefficient differs from Spearman's coefficient as the partial ranked correlation considers the relationship between two variables, an input and an output, after discounting or controlling for remaining variables. For example, in this instance the partial ranked correlation measures the monotonic relationship between say, the transmission rate and the total number of infectious, it considers this relationship after the relationship between all other parameters effecting the total number of infectious have been controlled for. The reason for calculating all three correlation coefficients is due to the fact that the underlying relationships between the outcome variable and the parameters is unknown. It may be linear, it may not. The calculation of all three is done to provide insight, however the partial rank correlation coefficient will undoubtedly be the most reliable as it's controlling for multiple effects.

The above procedure will be run with varying definitions of the vector y. The seasonal model will have the procedure run four times, and run eight times for both the migrant model without interaction and with interaction. The definitions of various y's follow with under-scripts to help distinguish them. As the model has been simulated for 12 years (144 months) in previous chapters, the model will continue to be simulated for this duration within this chapter and the initial conditions previously established for both the seasonal and migrant models will also be used.

Sensitivity values for the seasonal model established in §5 can be seen in table 7.1, and for the migrant model established in §6 can be seen in table 7.2.

Sensitivity Value	Formula	Interpretation
$y_{S(144)}$	$S(144)$	Number of Susceptible individuals after 144 months.
$y_{\Sigma I}$	$\sum_{j=0}^{144} I(j)$	The total sum of monthly infectious individuals.
$y_{\Sigma E}$	$\sum_{j=0}^{144} E(j)$	The total sum of monthly exposed individuals.
$y_{R(144)}$	$R(144)$	Number of Recovered individuals after 144 months.
y_{R_0}	R_0	The basic reproductive number.

Table 7.1: Sensitivity Values for Seasonal Model

Sensitivity Value	Formula	Interpretation
$y_{SM(144)}$	$SM(144)$	Number of Susceptible Migrant individuals after 144 months.
$y_{SL(144)}$	$SL(144)$	Number of Susceptible Local individuals after 144 months.
$y_{\Sigma EM}$	$\sum_{j=0}^{144} E_M(j)$	The sum of monthly migrant exposed individuals.
$y_{\Sigma EL}$	$\sum_{j=0}^{144} E_L(j)$	The sum of monthly local exposed individuals.
$y_{\Sigma IM}$	$\sum_{j=0}^{144} I_M(j)$	The sum of monthly migrant infectious individuals.
$y_{\Sigma IL}$	$\sum_{j=0}^{144} I_L(j)$	The sum of monthly local infectious individuals.
$y_{RM(144)}$	$RM(144)$	Number of Recovered Migrant individuals after 144 months.
$y_{RL(144)}$	$RL(144)$	Number of Recovered Local individuals after 144 months.
$y_{R_{(0)M}}$	$R_{(0)M}$	The basic reproductive number for the migrant population.
$y_{R_{(0)L}}$	$R_{(0)L}$	The basic reproductive number for the local population.
y_{R_0}	R_0	The basic reproductive number for the entire population.

Table 7.2: Sensitivity Values for Migrant Models

The probability distributions assigned to each parameter will be detailed at the beginning of each section followed by the correlation results and discussion. If a variable has its values sampled from the normal distribution with mean $\hat{\mu}$, and variance $\hat{\sigma^2}$, then it will be denoted $N(\hat{\mu}, \hat{\sigma}^2)$. If a variable is generated from a continuous uniform distribution with minimum $\hat{a}$ and maximum $\hat{b}$, it will be denoted $U(min = \hat{a}, max = \hat{b})$. A broad range of probability distributions could be used however this study will primarily draw from the normal distribution. The mean of each distribution for each parameter will the values estimated in §5 and §6. The standard deviations assigned to each distribution will

be based on data when possible, however when data are not available the standard deviations are assumed. The transmission parameters will have standard deviations of those calculated from the posterior distributions established in §5 and §6. The number of latin hypercube samples generated for each parameter is $M = 10,000$ for all models. For sections §7.2.1-7.2.2 When discussion is taking place about correlation, the study is referring to the partial ranked correlation coefficient (PRCC). In addition to this categorise of correlation are established following the categorise established in literature [136]. A very strong correlation is that with an r-square magnitude of 0.90 or greater, a strong correlation is that of magnitude between 0.70 and 0.90, a moderate correlation is that of magnitude between 0.50 and 0.70, a weak correlation has magnitude between 0.30 and 0.50, and a very weak correlation has magnitude less than 0.30. The results of the sensitivity analysis follow.

7.2.1 Seasonal Model Parameter Sensitivity Analysis

The standard deviations of w and q were assumed, all remaining parameter standard deviations were calculated from data. The distributions each parameter was generated follow in table 7.3.

Parameter	Distribution	Parameter	Distribution
Λ	$N(7267, 3358^2)$	μ	$N(0.00055, 0.0000325^2)$
w	$N(0.15, 0.05^2)$	d	$N(0.01643, 0.0068^2)$
r	$N(0.016, 0.000358^2)$	q	$N(0.05, 0.01^2)$
β_0	$N(0.0659, 0.0113^2)$	k_0	$N(0.00556, 0.000318^2)$

Table 7.3: Parameter Distributions for Sensitivity Analysis of Seasonal Model

Where Λ is the recruitment rate into the population, μ is the population death rate, w is the proportion of new recruits entering into the susceptible compartment, d is the death rate due to the disease, r is the recovery rate, q is the proportion of cases that go directly to infectious, β_0 is the transmission parameter, and k_0 is the progression rate from exposed to infectious.

The results of the varying correlations can be visualized in table 7.4 and figures 7.1 and detail the correlation results between the parameters and the total number of infections.

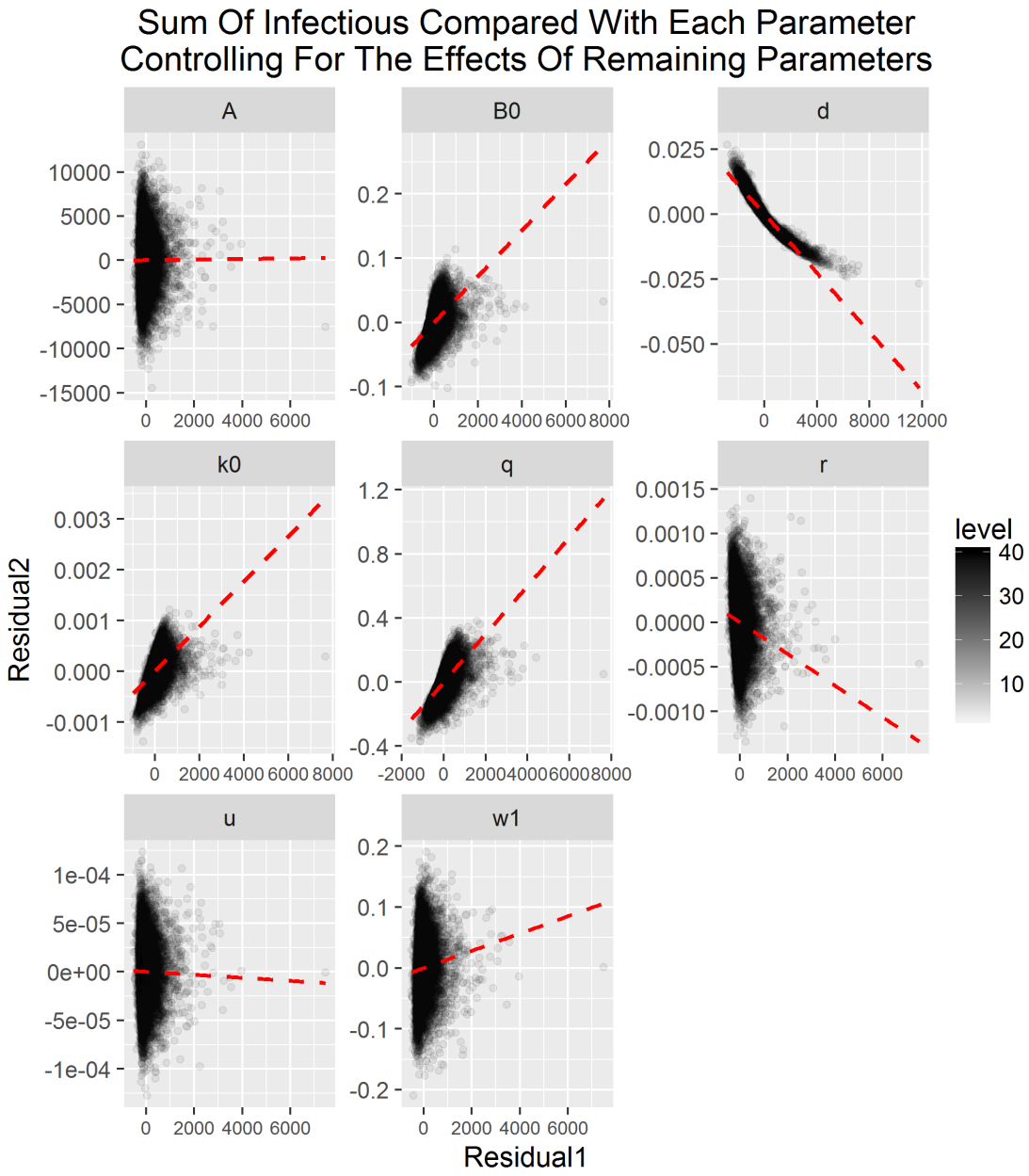

Figure 7.1: Parameter Values Compared with the Total Sum of Infectious, Controlling for the Effects of all Remaining Parameters

Parameter	Correlation Type	$y_{\Sigma I}$	$y_{\Sigma E}$	y_{R_0}	$y_{S(144)}$	$y_{R(144)}$
β_0	CC_P	0.147***	0.713***	0.357***	-0.015	-0.009
	CC_S	0.145***	0.715***	0.354***	-0.02*	-0.012
	PRCC	0.711***	0.943***	0.808***	-0.008	-0.004
k_0	CC_P	0.168***	-0.175***	0.01	0.007	0.013
	CC_S	0.172***	-0.183***	0.003	0.008	0.011
	PRCC	0.757***	-0.567***	0.024*	-0.006	0.021*
Λ	CC_P	-0.001	-0.015	-0.004	0.779***	0.989***
	CC_S	-0.005	-0.021*	-0.007	0.777***	0.99***
	PRCC	0.015	-0.015	-0.002	0.939***	0.998***
μ	CC_P	-0.02*	-0.05***	-0.024*	-0.034***	-0.039***
	CC_S	-0.031**	-0.055***	-0.033**	-0.04***	-0.04***
	PRCC	-0.051***	-0.112***	-0.042***	-0.117***	-0.505***
q	CC_P	0.234***	-0.226***	0.018	0.013	0.005
	CC_S	0.225***	-0.228***	0.02*	0.008	0.005
	PRCC	0.83***	-0.649***	0.076***	-0.007	-0.003
r	CC_P	-0.054***	-0.035***	-0.012	0.022*	0.008
	CC_S	-0.056***	-0.034***	-0.013	0.022*	0.008
	PRCC	-0.275***	-0.108***	-0.082***	0.009	-0.013
d	CC_P	-0.9***	-0.595***	-0.502***	-0.005	0.005
	CC_S	-0.929***	-0.572***	-0.464***	-0.003	0.01
	PRCC	-0.986***	-0.914***	-0.862***	-0.004	0.015
w_1	CC_P	0.037***	0.112***	0.72***	0.571***	-0.123***
	CC_S	0.034***	0.1***	0.764***	0.563***	-0.116***
	PRCC	0.143***	0.384***	0.943***	0.891***	-0.871***

Table 7.4: Various Correlation Results for each Parameter on the Model Output Values $y_{\Sigma I}$, $y_{\Sigma E}$, y_{R_0}, $y_{S(144)}$,and $y_{R(144)}$. *Significance codes:* * $p-value \leq 0.05$, ** $p-value \leq 0.01$, *** $p-value \leq 0.001$.

The sensitivity analysis results show that all parameters, excluding Λ the recruitment rate, significantly effect the total number of infectious and total number of exposed for the model over a 12 year period. Each parameter will now be discussed

- β_0: The transmission rate - The parameter was found to significantly positively correlate with the total number of infectious, exposed, and the basic reproductive number. The parameter showed a very strong relationship between the total number of exposed after discounting the effects of other parameters. Control measures for the transmission rate include a combination of isolating infectious and the self management methods discussed in §2.4.3.

- k_0: The progression rate - This rate was found to significantly positively correlate with the total number of infectious and exposed. Although a significant correlation was found for the basic reproductive number, it was calculated to be a very weak positive relationship. Control measures for this parameter involve detecting latent individuals. The Tuberculin skin test discussed in §2.4.1 is capable of detecting latent infectious and as such, reduction of the progression rate can be achieved through early detection before individuals begin experiencing symptoms.

- Λ: The recruitment rate - This parameter significantly contributed to the ending number of Susceptible and Recovered population in the model. Given the low notification rate relative to the recruitment of new individuals and given design of the model this is to be expected.

- μ: The population death rate - The population death rate significantly contributed to the decline in all compartments. An increase in the death rate parameter ultimately showed a decline in the total population, which is to be expect from the model construction. A note to be made is an increase in the death rate happened to decrease the basic reproductive number, however, although highly significant, the correlation measured a very weak relationship.

- q: The quick progression parameter. An increase in the number of individuals going from susceptible to infectious and missing the latent period resulted in an increase in the number of infectious, which given the design of the model is to be expected. A very weak positive correlation was calculated for the basic reproductive number

- r: The recovery rate. An increased recovery rate decreased the total number of

infectious and exposed over the period of the model, although the correlation was very weak. There were also significant very weak effects on the basic reproductive number.

- d - Death rate of infectious - This parameter was calculated have a strong relationship for the the basic reproductive number and to have a very strong relationship for both the total number of exposed and total number of infectious. An increased death rate implies less infectious which implies lower transmission of the disease. While this parameter cannot be intervened in, one cannot force the death rate to increase, it is telling about the underlying disease dynamics. Dying is equivalent to the removal of individuals, which can be equivalent to isolating individuals. This relates the the significance of the recovery rate as a large recovery rate implies a quick recovery which impies less time spent being infectious. Hence, the model suggests whether or not infectious are isolated effectively plays a crucial role to the number of infectious.

- w_1 : The proportion of individuals entering into the susceptible class. This parameter significantly impacted all compartments and the basic reproductive number. An increase in this parameter (or a decrease in the number of recruited immune individuals) showed a very weak positive correctional with the total number of infectious and a weak correlation for the total number of exposed. An increase also resulted in an increase in the ending number of susceptible and a decrease in the ending number of recovereds, which was result by design of the model.

In conclusion the parameters β_0, k_0 q, and d showed strong correlation with the total number of infectious while the parameters μ, r, and w_1 were calculated as having weak effects on the total number of infectious.

7.2.2 Migrant Model Parameter Sensitivity Analysis

Migrant Model Without Interaction

The parameter standard deviations of parameters v_1, v_2, π, w_1 and Λ were assumed, all remaining parameter standard deviations were calculated from data. The distribution each parameter was generated from the follow in table 7.5.

Parameter	Distribution	Parameter	Distribution
β_1	$N(0.0491, 0.01146^2)$	β_2	$N(0.02644, 0.0109^2)$
k_1	$N(0.00489, 0.0005^2)$	k_2	$N(0.005715, 0.00027^2)$
v_1	$N(0.00002, 0.00001^2)$	w_1	$N(0.15, 0.05^2)$
r_1	$N(0.0169, 0.00516^2)$	r_2	$N(0.01515, 0.00148^2)$
π	$N(2497, 1000^2)$	Λ	$N(4,770, 1500^2)$
μ_{I_M}	$U(min = 0, max = 0.0357)$[1]	μ_{I_L}	$N(0.0222, 0.01^2)$
v_2	$N(0.225, 0.1^2)$	μ	$N(0.00055, 0.0127^2)$

Table 7.5: Parameter Distributions for Sensitivity Analysis of Migrant Model Without Interaction

Where β_1 and β_2 are the transmission rates for the migrant and local populations, respectively, k_1 and k_2 are the progression rates from exposed to infectious, respectively, π and Λ are the recruitment rates within the migrant and local populations, respectively, v_1 and v_2 are the proportions of the recruitment rate being partitioned between the exposed and infectious migrant compartments, respectively, μ is the universal death rate, μ_{I_M} and μ_{I_L} are the death rates due to infection of the migrant and local population, respectively, r_1 and r_2 are the rates at which individuals recover or are treated for tuberculosis for the migrant and local populations, respectively, and w_1 is the proportion of local recruits entering into the susceptible compartment.

The results of the varying correlation statistics can be seen in table 7.6 and 7.7. Figures 7.2 and 7.3 visualise the correlation results between the parameters and the total number of infections over the review period.

[1]The standard deviation of annual death rate variable was calculated to be large. This caused a normally distributed variable to take negative values which cannot be modelled. Hence, a uniformly distributed

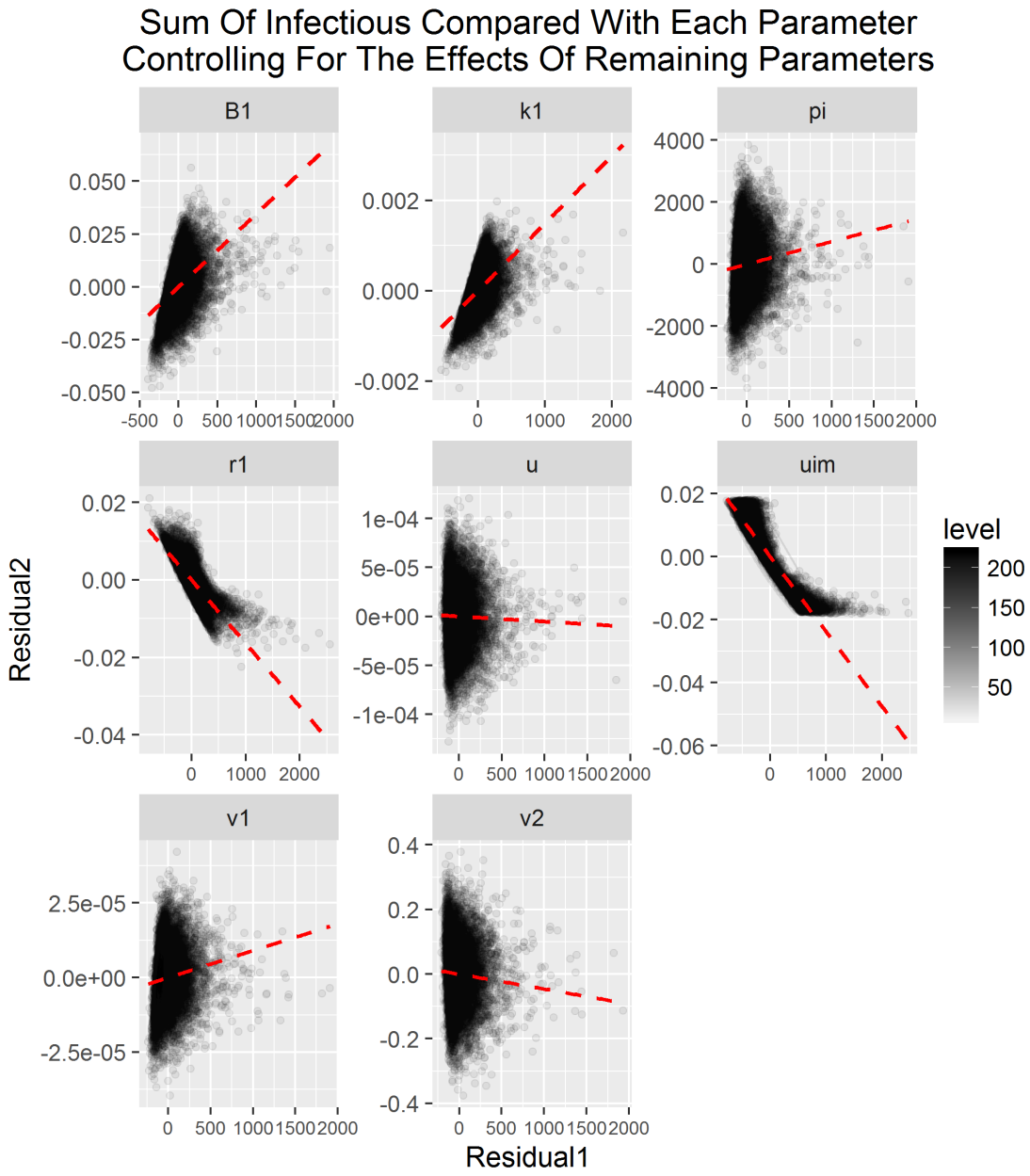

Figure 7.2: Parameter Values Compared with the Total Sum of Infectious, Controlling for the Effects of all Remaining Parameters

variable was simulated with minimum and maximum values of the data.

Parameter	Correlation Type	$y_{\Sigma IM}$	$y_{\Sigma EM}$	y_{R_0}	$y_{SM(144)}$	$y_{RM(144)}$
β_1	CC_P	0.167***	0.657***	0.456***	-0.001	0.01
	CC_S	0.157***	0.672***	0.561***	-0.007	0.01
	PRCC	0.697***	0.93***	0.917***	0.006	-0.002
k_1	CC_P	0.221***	-0.03**	0.02*	-0.006	-0.018
	CC_S	0.231***	-0.051***	0.031**	-0.006	-0.017
	PRCC	0.8***	-0.219***	0.103***	-0.001	0
π	CC_P	0.044***	0.175***	0.001	0.945***	0.633***
	CC_S	0.052***	0.179***	0.006	0.945***	0.629***
	PRCC	0.25***	0.538***	-0.003	0.988***	0.9***
v_1	CC_P	0.035***	0.184***	-0.024**	0.01	0.003
	CC_S	0.045***	0.203***	-0.0101	0.007	-0.005
	PRCC	0.299***	0.625***	0.005	0.006	-0.022*
v_2	CC_P	-0.021*	-0.09***	-0.228***	-0.32***	0.709***
	CC_S	-0.018	-0.077***	-0.262***	-0.303***	0.706***
	PRCC	-0.123***	-0.318***	-0.721***	-0.892***	0.918***
μ	CC_P	-0.009	-0.016	-0.011	-0.017	-0.014
	CC_S	-0.007	-0.016	-0.012	-0.013	-0.016
	PRCC	-0.037***	-0.076***	-0.071***	-0.078***	-0.033**
r_1	CC_P	-0.417***	-0.28***	-0.364***	-0.011	-0.006
	CC_S	-0.384***	-0.253***	-0.321***	-0.013	-0.008
	PRCC	-0.918***	-0.708***	-0.805***	0.008	-0.002
μ_{I_M}	CC_P	-0.793***	-0.559***	-0.633***	-0.004	-0.004
	CC_S	-0.856***	-0.56***	-0.713***	-0.002	-0.005
	PRCC	-0.981***	-0.904***	-0.946***	0.005	0.006

Table 7.6: Various Correlation Results for each Parameter on the Model Output Values $y_{\Sigma IM}$, $y_{\Sigma EM}$, y_{R_0}, $y_{SM(144)}$,and $y_{RM(144)}$. *Significance codes:* * $p-value \leq 0.05$, ** $p-value \leq 0.01$, *** $p-value \leq 0.001$.

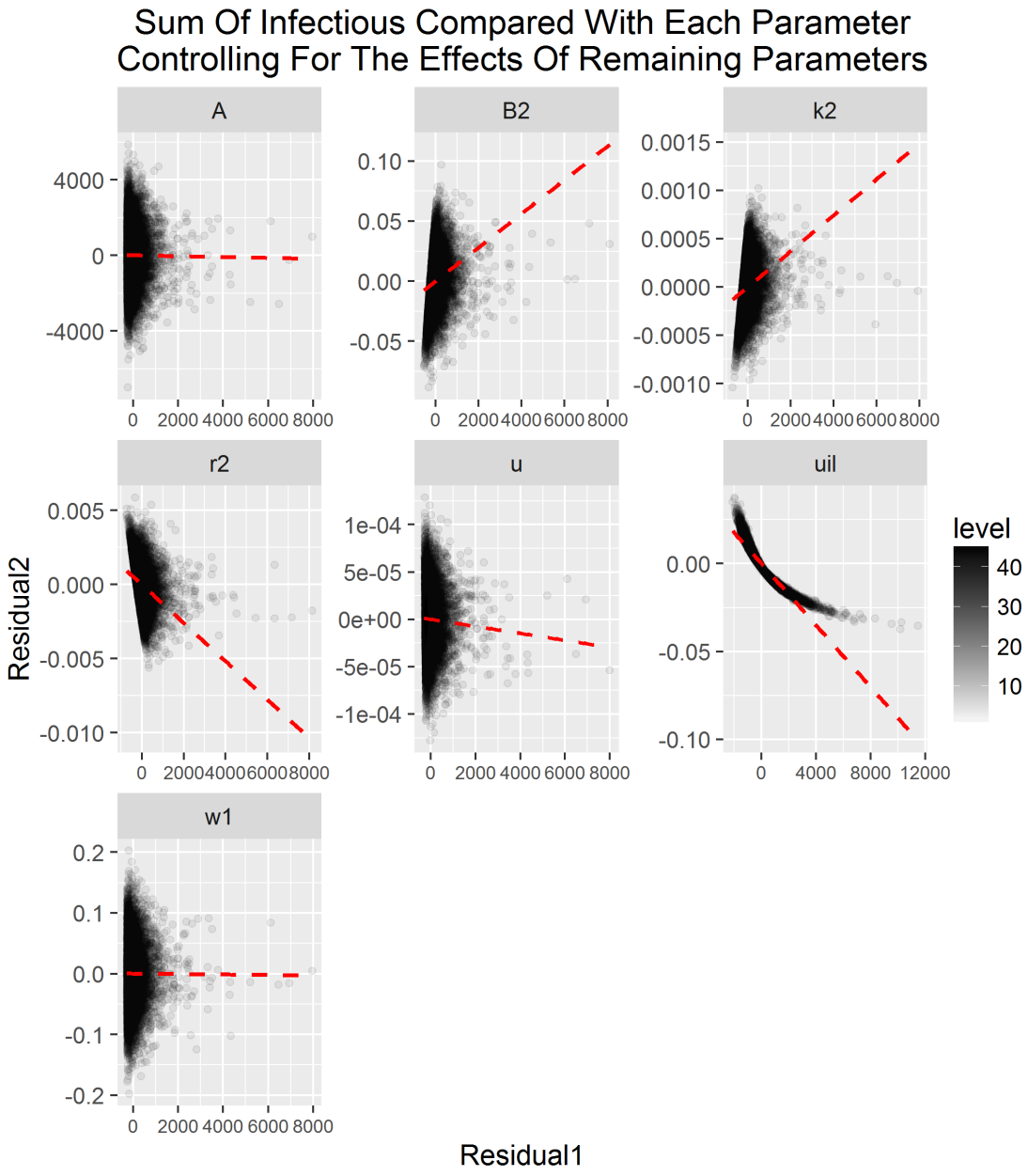

Figure 7.3: The Impact of Model Parameters on the Total Number of Local Infectious, Controlling the Effects of all Other Parameters.

Parameter	Correlation Type	$y_{\Sigma IL}$	$y_{\Sigma EL}$	y_{R_0}	$y_{SL(144)}$	$y_{RL(144)}$
β_2	CC_P	0.09***	0.81***	0.154***	-0.012	-0.013
	CC_S	0.091***	0.821***	0.662***	-0.015	-0.017
	$PRCC$	0.709***	0.96***	0.896***	0.005	-0.005
k_2	CC_P	0.094***	-0.313***	-0.001	0.005	0.005
	CC_S	0.094***	-0.325***	0.009	0.008	0.006
	$PRCC$	0.77***	-0.804***	0.044***	0.009	-0.007
w_1	CC_P	-0.019	-0.011	0.112***	-0.025*	-0.026**
	CC_S	-0.016	-0.004	0.524***	-0.023*	-0.025*
	$PRCC$	0.008	0.007	0.851***	0.007	-0.001
r_2	CC_P	-0.141***	-0.057***	-0.029**	-0.001	0.002
	CC_S	-0.141***	-0.048***	-0.052***	-0.003	-0.002
	$PRCC$	-0.839***	-0.231***	-0.191***	-0.002	0.007
Λ	CC_P	0.005	-0.01	0.003	0.946***	0.995***
	CC_S	0.008	-0.008	-0.017	0.974***	0.996***
	$PRCC$	-0.006	0.016	0.011	0.977***	0.999***
μ	CC_P	-0.03**	-0.06***	-0.019	-0.082***	-0.084***
	CC_S	-0.024*	-0.056***	-0.012	-0.082***	-0.081***
	$PRCC$	-0.108***	-0.207***	-0.035***	-0.325***	-0.831***
μ_{I_L}	CC_P	-0.917***	-0.426***	-0.151***	-0.012	-0.009
	CC_S	-0.976***	-0.393***	-0.408***	-0.012	-0.01
	$PRCC$	-0.996***	-0.848***	-0.784***	-0.01	0.01

Table 7.7: Various Correlation Results for each Parameter on the Model Output Values $y_{\Sigma IL}$, $y_{\Sigma EL}$, y_{R_0}, $y_{SL(144)}$,and $y_{RL(144)}$. *Significance codes:* * $p-value \leq 0.05$, ** $p-value \leq 0.01$, *** $p-value \leq 0.001$.

The sensitivity analysis concluded, after controlling for the effects of other parameters, all parameters modelling the migrant population were highly significant in effecting the total number of infectious an the total number of exposed over a 12 year period. Each parameter will now be discussed

- β_1, β_2: The transmission rate among the migrant and local population, respectively

- Both parameters showed strongly positive correlation with each populations basic reproductive number and the total number of exposed. Each parameter was calculated to have approximately the same impact on the total number of infectious after controlling for other parameter effects ($\beta_1 - PRCC = 0.7$, $\beta_2 - PRCC = 0.71$). Control strategies for the transmission rate include a reduction in contact with susceptible individuals and self management.

- k_1, k_2: The progression rate to infectious among the migrant and local populations, respectively - Both rates significantly impacted infectious and exposed totals along with having very weak positive correlations with the basic reproductive number. Both rates showed a negative correlation with the number of exposed after the effects of other parameters were considered, however the correlation was much stronger among the local population when compared to that of the migrant population. This may relate to the initial conditions assigned to the local population; the initial exposed population within the local population being almost four times that of the migrant, which would imply greater sensitivity to the progression rate parameter.

- r_1, r_2: The recovery rate among the migrant and local populations, respectively - The recovery rates for both populations were calculated as having highly significant negative correlations with the total number of infectious and exposed, and the basic reproductive number. A stronger to very strong negative correlation was seen for both parameters on the total number of infectious, however differences were seen in the effects on the total exposed population and the basic reproductive number. Strong negative correlations were seen between the recovery rate with the total exposed and the basic reproductive number for the migrant population, but very weak negative correlations were calculated between the recovery rate with the total exposed and the basic reproductive number for the local population. This difference could ne due to the count of infectious individuals being different for each population.

- μ_{I_M}, μ_{I_L}: The infectious death rate among the migrant and local populations, respectively - Both rates were calculated as having highly significant strong to very strong

negative correlations with the total infectious, total exposed, and basic reproductive numbers after the effects of other parameters were considered. Both parameters had the largest correlations for the total infectious and total exposed when compared with all other parameters. This result was seen in the sensitivity analysis in §7.2.1 for the seasonal model.

- μ: The population death rate - The population death rate significantly contributed to the decline in all compartments for both models. An increase in the death rate parameter ultimately showed a decline in both migrant and local populations. An increase in the death rate happened to decrease the basic reproductive number for both models. A remark to be made is there was a strong negative correlation seen between the death rate and the number of recovered individuals towards the end of the model, after the effects of all other parameters were considered. A possible cause of the strong correlation could have to do with the large initially recovered population, changes to a rate applied to this population could result in drastic changes to the population itself.

- π, Λ: The recruitment rates for the migrant and local populations, respectively. Both were calculated as having highly significant strong positive correlations with the susceptible and recovered compartments for each model. After the effects of other parameters were considered a close to perfect correlation (0.999) was observed between Λ and the ending number of local recovered individuals. For π a very weak positive correlation was observed for the total number of infectious, and a moderate positive correlation was observed for the for the total number of exposed. This result may be due to the design of the model, incorporating proportional parameters v_1 and v_2 on the migrant recruitment rate.

- v_1, v_2: The proportion of new recruits within the migrant population entering into the exposed and recovered compartments, respectively. Discounting the effects of other parameters, an increase in the proportion of individuals entering into the exposed class correlated positively with the total number of infectious (very weak positive correlation) and exposed (moderate positive correlation).The parameter v_1 was also

showed a very weak negative correlation with the ending number of recovereds of the model. The proportion of individuals entering into the recovered class significantly correlated with all compartments and the basic reproductive number. Negative correlations were observed between v_2 and the total number of infectious (very weak correlation), total number of exposed (weak correlation), the basic reproductive number (strong correlation) and the ending number of susceptible individuals (strong correlation). A very strong positive correlation was seen between v_2 and the ending number of recovered individuals, which due to model design is to be expected.

- w_1: The proportion of individuals entering into the susceptible class in the local population. This variable was strongly positively correlated with the basic reproductive number. No other significant correlations were observed.

The results suggest the parameters: β_1, k_1, r_1 and μ_{I_M}, showed moderate to very strong correlation with the total number of infectious migrants, while the parameters π, v_1, v_2, and μ showed very weak correlation with the total number of infectious. The parameters showing strong to very strong correlation with the basic reproductive number include β_1, v_2, r_1, and μ_{I_M}, were as the parameter k_1 showed very weak correlation with the basic reproductive number. The parameter β_1 can be reduced by way of supporting strategies that isolate infectious individuals, k_1 can be reduced through a reduction of exposed individuals within the population which can be done through first expanding detection methods, then treating the exposed individuals. The recovery rate, r_1, can be influenced by way of ensuring infectious complete treatment and to reduce individuals lost to follow up. The number of individuals entering into the recovered migrant class, v_2, can be increased by way of pre-screening migrant individuals, specifically high-risk migrant individuals and ensure vaccination has occurred. Such strategies have been cost effectively implemented in countries such as Australia, Austria, Canada, France, Israel, New Zealand, and USA [137].

For the local population the parameters: β_2, k_2, r_2 and μ_{I_L} showed a strong to very strong correlation with the total number of infectious locals, while the parameter μ showed a very weak correlation with the total number of infectious. The parameters strongly correlating

with the basic reproductive number for the local population were β_2, w_1, and μ_{I_L}. The sensitivity analysis for the migrant model with interaction follows.

Migrant Model With Interaction

The parameter standard deviations of parameters v_1, v_2, π, w_1, and Λ were assumed. The transmission parameter $(\beta_1, k_1, \beta_2, k_2, \beta_1^*, \beta_2^*)$ standard deviations were obtained from posterior distributions acquired from the Metropolis-Hastings algorithm in §6.4.4. All remaining parameter standard deviations were calculated from TB and Census data used within §4. The distribution each parameter was generated from follow in table 6.8. The results of the varying correlations for both local and migrant populations can be visualized in figures 7.4 and 7.5 and tables 7.9 - 7.12 detail the various correlation results.

Parameter	Distribution	Parameter	Distribution
β_1	$N(0.0224, 0.0135^2)$	β_2	$N(0.00464, 0.00044^2)$
k_1	$N(0.02554, 0.00278^2)$	k_2	$N(0.00559, 0.00029^2)$
v_1	$N(0.00002, 0.00001^2)$	w_1	$N(0.15, 0.05^2)$
r_1	$N(0.0169, 0.00516^2)$	r_2	$N(0.01515, 0.00148^2)$
π	$N(2497, 1000^2)$	Λ	$N(4770, 1500^2)$
μ_{I_M}	$U(min = 0, max = 0.0357)^2$	μ_{I_L}	$N(0.0222, 0.01^2)$
v_2	$N(0.225, 0.1^2)$	μ	$N(0.00055, 0.0127^2)$
β_2^*	$N(0.00657, 0.00029^2)$	β_1^*	$N(0.01279, 0.0017^2)$

Table 7.8: Parameter Distributions for Sensitivity Analysis of Migrant Model With Interaction

Where, β_1 and β_2 are the transmission rates for the migrant and local populations, respectively, k_1 and k_2 are the progression rates from exposed to infectious, respectively, π and Λ are the recruitment rates within the migrant and local populations, respectively, v_1 and v_2 are the proportions of the recruitment rate being partitioned between the exposed and infectious migrant compartments, respectively, μ is the universal death rate, μ_{I_M} and μ_{I_L} are the death rates due to infection of the migrant and local population, respectively, r_1 and r_2 are the rates at which individuals recover or are treated for tuberculosis for the migrant and local populations, respectively, w_1 is the proportion of local recruits entering into

the susceptible compartment, and where β_1^* and β_2^* are the cross-population transmission rates for the migrant and local populations, respectively.

[2]The standard deviation of annual death rate variable was calculated to be large. This caused a normally distributed variable to take negative values which cannot be modeled. Hence, a uniformly distributed variable was simulated with minimum and maximum values of the data.

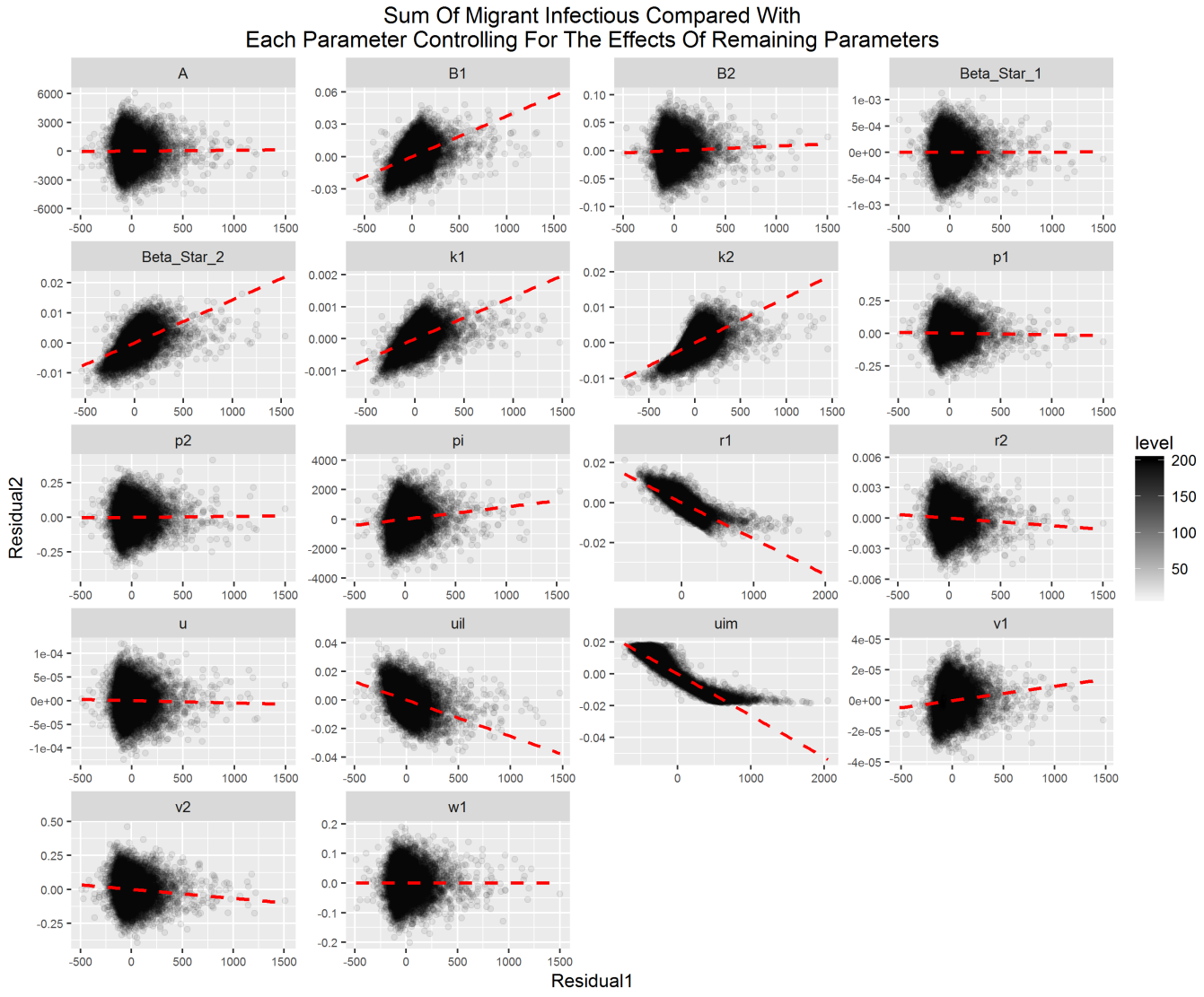

Figure 7.4: The Impact of Model Parameters on the Total Number of Migrant Infectious, Controlling the Effects of all Other Parameters.

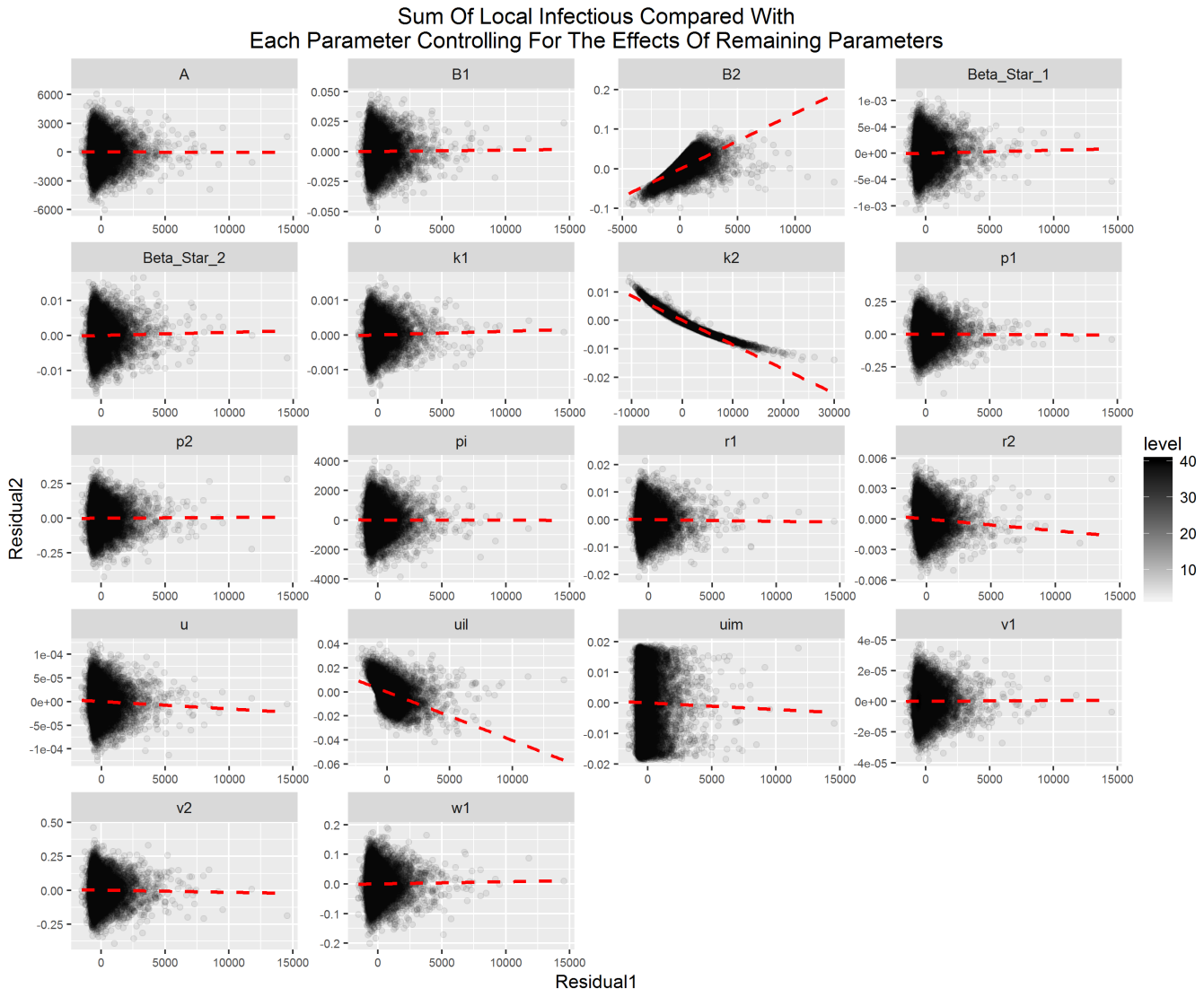

Figure 7.5: The Impact of Model Parameters on the Total Number of Local Infectious, Controlling the Effects of all Other Parameters.

Parameter	Correlation Type	$y_{\Sigma IM}$	$y_{\Sigma EM}$	y_{R_0}	$y_{\Sigma IL}$	$y_{\Sigma EL}$
β_1	CC_P	0.141***	0.385***	0.143***	0.022*	-0.023*
	CC_S	0.132***	0.39***	0.783***	0.017	-0.02
	PRCC	0.544***	0.754***	0.895***	0.012	-0.012
k_1	CC_P	0.19***	-0.084***	0.003	0.008	0.004
	CC_S	0.185***	-0.093***	0.005	0.008	-0.003
	PRCC	0.64***	-0.246***	0.032**	0.001	-0.009
r_1	CC_P	-0.401***	-0.066***	-0.043***	-0.012	0.007
	CC_S	-0.375***	-0.069***	-0.185***	-0.005	-0.005
	PRCC	-0.852***	-0.172***	-0.437***	-0.008	-0.001
β_2^*	CC_P	0.207***	0.495***	-0.009	0.002	0.007
	CC_S	0.209***	0.491***	0.024*	0.001	0.013
	PRCC	0.648***	0.809***	0.048***	0.014	0.008
p_2	CC_P	0	0.001	0	-0.009	0.006
	CC_S	-0.007	0	-0.001	-0.013	0.007
	PRCC	-0.008	0.002	0.008	0.006	0.002
μ_{I_M}	CC_P	-0.753***	-0.115***	-0.101***	-0.021*	0.01
	CC_S	-0.798***	-0.113***	-0.374***	-0.02*	0.008
	PRCC	-0.962***	-0.316***	-0.701***	-0.02*	0.002
π	CC_P	0.044***	0.098***	0	0.02*	-0.004
	CC_S	0.043***	0.101***	0.015	0.024*	-0.001
	PRCC	0.166***	0.258***	-0.009	-0.002	0.013
v_1	CC_P	0.047***	0.102***	-0.001	0.016	0
	CC_S	0.051***	0.104***	-0.008	0.014	-0.001
	PRCC	0.21***	0.309***	0.003	0.011	-0.005
v_2	CC_P	-0.04***	-0.09***	-0.002	0.019	-0.003
	CC_S	-0.04***	-0.088***	-0.014	0.024*	-0.004
	PRCC	-0.098***	-0.193***	0.009	0.005	0.007

Table 7.9: Various Correlation Results for each Parameter on the Model Output Values $y_{\Sigma IM}$, $y_{\Sigma EM}$, y_{R_0}, $y_{\Sigma IL}$,and $y_{\Sigma EL}$. *Significance codes:* * $p-value \leq 0.05$, ** $p-value \leq 0.01$, *** $p-value \leq 0.001$.

Parameter	Correlation Type	$y_{\Sigma IM}$	$y_{\Sigma EM}$	y_{R_0}	$y_{\Sigma IL}$	$y_{\Sigma EL}$
μ	CC_P	-0.004	0.011	0.012	-0.019	0.013
	CC_S	0.001	0.012	-0.006	-0.018	0.016
	PRCC	-0.009	-0.027**	-0.007	-0.064***	-0.012
β_2	CC_P	0.009	0.034***	0.032**	0.227***	0.062***
	CC_S	0.005	0.03**	0.13***	0.248***	0.054***
	PRCC	0.042***	0.097***	0.303***	0.831***	0.238***
k_2	CC_P	0.204***	0.562***	-0.036***	-0.95***	0.748***
	CC_S	0.208***	0.561***	-0.014	-0.957***	0.779***
	PRCC	0.703***	0.85***	0.019	-0.988***	0.937***
w_1	CC_P	-0.034***	-0.024*	0.019	0.009	-0.017
	CC_S	-0.033**	-0.017	0.115***	0.001	-0.015
	PRCC	0.005	-0.004	0.315***	0.002	-0.006
r_2	CC_P	-0.035***	-0.056***	-0.001	-0.013	-0.085***
	CC_S	-0.033***	-0.048***	-0.022*	-0.021*	-0.073***
	PRCC	-0.098***	-0.144***	-0.051***	-0.12***	-0.263***
Λ	CC_P	0.012	0.004	-0.007	-0.016	0.016
	CC_S	0.005	0.001	-0.015	-0.017	0.015
	PRCC	-0.011	-0.006	0.001	0.003	0.001
μ_{I_L}	CC_P	-0.141***	-0.362***	-0.028**	-0.092***	-0.548***
	CC_S	-0.145***	-0.331***	-0.131***	-0.104***	-0.548***
	PRCC	-0.489***	-0.679***	-0.323***	-0.577***	-0.883***
β_1^*	CC_P	0.006	-0.008	0.006	0.024*	-0.011
	CC_S	0.006	-0.005	0.019	0.026**	-0.009
	PRCC	0.003	0.002	0.014	0.019	0.006
p_1	CC_P	-0.011	0.01	0.017	-0.002	-0.003
	CC_S	-0.006	0.006	0.003	-0.002	-0.003
	PRCC	-0.007	-0.009	0.002	-0.008	-0.013

Table 7.10: Various Correlation Results for each Parameter on the Model Output Values $y_{\Sigma IM}$, $y_{\Sigma EM}$, y_{R_0}, $y_{\Sigma IL}$,and $y_{\Sigma EL}$. *Significance codes:* * $p-value \leq 0.05$, ** $p-value \leq 0.01$, *** $p-value \leq 0.001$.

Parameter	Correlation Type	$y_{SM(144)}$	$y_{RM(144)}$	$y_{SL(144)}$	$y_{RL(144)}$
μ	CC_P	0.01	0.001	-0.06***	-0.064***
	CC_S	0.007	0.005	-0.067***	-0.066***
	PRCC	-0.079***	-0.039***	-0.322***	-0.829***
β_2	CC_P	0.02*	0.023*	-0.023*	-0.021*
	CC_S	0.019	0.022*	-0.016	-0.018
	PRCC	0.009	0.006	0.009	0.001
k_2	CC_P	-0.008	-0.025*	0.014	0.012
	CC_S	-0.01	-0.024*	0.015	0.015
	PRCC	0.014	0.009	0.005	0.002
w_1	CC_P	0.005	0	0.002	-0.009
	CC_S	0.007	0.001	-0.001	-0.008
	PRCC	-0.002	-0.002	0.025*	-0.019
r_2	CC_P	0.02	-0.008	0.002	0.001
	CC_S	0.017	-0.006	0.009	0.005
	PRCC	0.016	0	0.014	-0.007
Λ	CC_P	-0.014	-0.002	0.948***	0.995***
	CC_S	-0.017	-0.001	0.974***	0.996***
	PRCC	0.003	0.022*	0.977***	0.999***
μ_{I_L}	CC_P	-0.008	-0.026**	-0.012	-0.011
	CC_S	-0.011	-0.027**	-0.008	-0.006
	PRCC	0.009	-0.018	-0.009	0.005
β_1^*	CC_P	-0.014	0.006	0.01	0.01
	CC_S	-0.012	0.007	0.018	0.015
	PRCC	-0.018	0.007	0.016	0.003
p_1	CC_P	0	-0.006	0.007	0.001
	CC_S	-0.006	-0.006	0.005	0.004
	PRCC	-0.007	0.017	0.005	-0.006

Table 7.11: Various Correlation Results for each Parameter on the Model Output Values $y_{SM(144)}$, $y_{RM(144)}$, $y_{SL(144)}$, and $y_{RL(144)}$. *Significance codes:* * $p-value \leq 0.05$, ** $p-value \leq 0.01$, *** $p-value \leq 0.001$.

Parameter	Correlation Type	$y_{SM(144)}$	$y_{RM(144)}$	$y_{SL(144)}$	$y_{RL(144)}$
β_1	CC_P	0.019	-0.007	-0.014	-0.02
	CC_S	0.023*	-0.004	-0.019	-0.02*
	PRCC	0.012	-0.003	0	-0.005
k_1	CC_P	0.006	0.011	0.007	0.009
	CC_S	0.008	0.007	0.008	0.008
	PRCC	0.009	0.005	0.003	0.012
r_1	CC_P	-0.007	-0.017	-0.002	-0.002
	CC_S	-0.003	-0.012	-0.002	-0.002
	PRCC	0.011	-0.004	-0.001	0.001
β_2^*	CC_P	0.009	0.013	0.003	0.003
	CC_S	0.01	0.006	0.005	0.004
	PRCC	-0.002	0.005	0.006	0.012
p_2	CC_P	-0.01	0.002	0.003	0.005
	CC_S	-0.012	0.006	0.008	0.01
	PRCC	-0.018	0.016	-0.005	0.001
μ_{I_M}	CC_P	-0.004	0.019	-0.001	-0.009
	CC_S	-0.004	0.016	-0.004	-0.007
	PRCC	-0.011	0.005	0.007	-0.017
π	CC_P	0.945***	0.643***	-0.016	-0.013
	CC_S	0.945***	0.634***	-0.019	-0.018
	PRCC	0.987***	0.9***	-0.002	0.01
v_1	CC_P	0.008	-0.003	0.006	0.004
	CC_S	0.005	0	0.002	0.002
	PRCC	0.002	0.018	0.002	-0.008
v_2	CC_P	-0.313***	0.701***	0.005	0.003
	CC_S	-0.304***	0.697***	0.005	0.005
	PRCC	-0.888***	0.914***	0	-0.001

Table 7.12: Various Correlation Results for each Parameter on the Model Output Values $y_{SM(144)}$, $y_{RM(144)}$, $y_{SL(144)}$, and $y_{RL(144)}$. *Significance codes:* * $p-value \leq 0.05$, ** $p-value \leq 0.01$, *** $p-value \leq 0.001$.

The sensitivity analysis concluded, after controlling for the effects of other parameters, all parameters used for the migrant population did not significantly effect the total count of local infectious or exposed individuals. No significant effect was observed for the ending number of susceptible or recovered individuals either. Out of the parameters used to model the local population, four had significant effect on the the migrant population - β_2, w_1, r_2, μ_{I_L}. This infers that given the parameter set calculated, a one way interaction is occurring, the local population effecting the migrant population but not the other way around. Each parameter will now be discussed

- β_1, β_2^*: The transmission rate for the migrant population, and the transmission rate of the local population onto the migrant population, respectively - Both parameters showed moderate positive correlation with the total number of migrant infectious, and showed strong positive correlation with the total number of exposed. The transmission among the migrant population showed strong positive correlation with the overall model's basic reproductive number, and the transmission of local onto the migrant population showed a very weak positive correlation with the basic reproductive number. This indicates transmission from the local population onto the migrant population plays a very small role with regards the basic reproductive number being greater than one. No significant interaction was seen between these parameters and the infectious/exposed local compartments.

- β_2, β_1^*: The transmission rate for the local population, and the transmission rate of the migrant population onto the local population, respectively - The local transmission rate significantly correlated with the infectious and exposed local compartments and the migrant infectious and exposed compartments. In addition the parameter was calculated as having weak positive correlation with the overall models basic reproductive number. The correlation of this parameter on the migrant population were classes as very weak, albeit significant. The transmission from the migrant population onto the local showed no significance with regards any compartments or the models basic reproductive number. This indicates the impact this parameter has on the model is not very significant.

- k_1, k_2: The progression rate from exposed to infectious among the migrant and local populations, respectively - A strong positive correlation was seen from the local progression rate onto the migrant infectious and exposed compartments, this correlation was larger than the impact the progression rate within the migrant population had, showing a moderate to weak correlation on the number of infectious/exposed. This indicates, given a significant interaction occurring from local to migrant compartments, the parameters estimates of the local population will effect the migrant population and not just the parameters designed specifically to interact (β_1^* and β_2^*). Hence, if this model is to be accepted, control and prevention measures applied to the local population (e.g. slowing progression rate from latent to active) will , as a corollary, impact the migrant population.

- r_1, r_2: The recovery rates among the migrant and local populations, respectively - Similar effects were observed for the recovery rates as for the progression rates. A strong to weak correlation was calculated between the migrant recovery rate and the migrant total infectious and exposed. A weak significant negatove correlation was seen with the overall model's basic productive number. The recovery rate for the local population significantly affected all infectious and exposed compartments and the model's basic reproductive number, although all correlations could be classed as very weak. This again suggests, if this interaction is accepted, interventions on the local population may impact the migrant population.

- π, Λ, v_1, v_2, w_1: The recruitment rates for the migrant and local population, the proportion recruited migrants exposed, and the proportions of recruited migrants and locals recovered, respectively. The recruitment rate among the migrant population showed a very weak significant correlation with the number of infectious and exposed for the migrant population, the recruitment rate did not correlate with the local population in anyway. This can more than likely be attributed to the proportional parameters associates with this recruitment rate. A change in the proportional parameters showed very weak correlation with the total number of exposed and infectious. For the local population no significant correlations were calculated between the recruitment rate and the total number of infectious/exposed. A change in

the proportional parameter associated with this recruitment rate also did not significantly correlate with the total number of infectious/exposed.

- μ_{I_M}, μ_{I_L}: The infectious death rate among the migrant and local populations, respectively - The death rate for the migrant infectious population showed very strong negative correlation with the total number of infectious. A weak correlation was calculated for the effects on the total number of exposed migrants. A strong correlation was seen impacting the basic reproductive number overall, and a very weak correlation was found significant on this death rate and the total number infectious within the local population. The death rate of the local infectious significantly negatively correlation with all infectious and exposed compartments, and negatively correlated with the model's basic reproductive number. In addition to the progression and recovery rate, this suggests an increased death rate of the local population impacts the migrant population. This can be adapted as a control strategy, suggesting isolation of local infectious population impacts the migrant infectious and exposed populations.

- μ: The population death rate - The population death rate did significantly negatively correlated with the total number of migrant exposed and the total number of local exposed individuals. However, both correlations were classed as very weak.

- p_1, p_2: the dampening factor on the transmission rate effecting the local population from the migrant population, and the dampening factor on the transmission rate effecting the migrant population from the local population, respectively. Neither parameter significantly correlated with any exposed or latent compartments, or the basic reproductive number. Given the standard deviations of each distribution, this suggests the dampening of the interactions is not important for intervention purposes.

The results suggest the parameters: β_1, k_1, r_1, β_2^* and μ_{I_M}, showed moderate to very strong correlation with the total number of infectious migrants, while the parameters π, v_1, and v_2, and μ showed very weak correlation with the total number of migrant infectious. Excluding the migrant death rare, there was no significant interaction between the remaining migrant parameters and the local infectious. The parameters β_2, k_2, and μ_{I_L} were observed

as having a moderate to very strong relationship with the local infectious. Interaction between these parameters and the migrant infectious proved to be significant, however, a weak to very weak correlation was observed. The parameters showing a moderate to very strong correlation with the basic reproductive number include β_1 and μ_{I_M}, were as the parameter k_1, r_1, β_2^*, β_2, w_1, r_2, and μ_{I_L} showed weak to very weak correlation with the basic reproductive number. The model suggests, given this interaction occurring within the population, that staging an intervention on parameters used to model the local population, namely β_2, k_2, r_2, and μ_{I_L}, may have an impact on the migrant infectious population. The following section is a scenario analysis on two of the models constructed.

7.3 Scenario Analysis

A scenario analysis is a process used to examine possible events that can take place in the future. Two scenarios will be investigated in this section to evaluate the effects on the model. The scenarios being considered are: a change in the vaccination rate for the local population, and a change in parameters at various times of the year given the seasonal dynamics of the disease. Both §7.3.1 and §7.3.2 evaluate these scenarios.

7.3.1 Universal Vaccination Outcome

A recent report produced by the HIQA [146] recommended a change to vaccination procedures from the one based on universal vaccination to selective vaccination of vulnerable populations. Other countries, such as Sweden [156], made a similar transition in 1975 and have continued to have one of the lowest incidences of TB in the world. This section will now attempt to simulate the migrant model considering similar relevant vaccination scenarios. the analysis is aimed at modelling a similar scenario.

Methods

This scenario will be simulated within the migrant model with no interaction occurring between the local and migrant populations. This model was presented in §6.2.1 and now be used to simulate changes within the local population in order to evaluate the sensitivity of the basic reproductive number to a changing vaccination strategy. When a simulation took place for this model in §6.5.2, it was assumed 65% of individuals recruited into the local population were immune, and the remaining 35% were susceptible to TB. As the model assumed homogeneous mixing (any given individual is equally likely to be infected), the implications on the proportion of individuals entering into the susceptible class is unknown when selective vaccination is implemented. This is because selective vaccination targets the high-risk groups identified by the HIQA report and the model does not consider these high-risk groups. This is a recognised limitation of the model simulating this scenario, and because of this no clinical conclusions will be made for this analysis.

An additional limitation to drawing clinical conclusions is that the local population is considered to be individuals not born to the top 20 countries contributing to TB notifications in Ireland. The local population is considered to comprise a mixture of "low contributing" foreign-born and native-born individuals. Although the local population dominantly contains native-born individuals, it does not completely represent the native-born population. Given these limitations, the model will be assessed for varying increases in the proportion entering into the susceptible compartment (w_1) after the year 2013. Varying increases will be assessed, as the impact on model parameters that selective vaccination will have is unknown. The parameter set used to simulate this scenario will be parameters resulting from ABC parameter estimation method calculated in §6.4. The ABC algorithm parameters were selected as the mean of the residuals were marginally better than that of the residuals for the Metropolis-Hastings algorithm.

With respect to the model, to simulate a change in vaccination the proportion of newly recruited individuals entering into the susceptible class, w_1, will increase to $\dot{w_1} = w_1 + \Delta$ at time point $t = T$. The parameter set and initial conditions being used will be the ones

already established for the migrant model in §6, and are as follows:

$$\theta = (\Delta, w_1 = 0.35, \Lambda = 4770, \beta_2 = 0.0264, k_2 = 0.0057, \\ r_2 = 0.01515, \mu_{I_L} = 0.0222, \mu = 0.00055),$$

and initial conditions:

$$S_L(0) = 1,280,078, E_L(0) = 222, I_L(0) = 27, R_L(0) = 2,347,409.$$

The value $T = 144$ was used, indicating a change in the model from w_1 to $\ddot{w}_1$ at a time point towards the end of 2013. The model was extrapolated forward 10 years and the sensitivity values $y_{\Sigma I}$ and y_{Σ_E} were measured. The results were compared to a model for which no increase in the number of susceptibles took place.

Results

Marginal effects were observed on the sensitivity values $y_{\Sigma I}$ and y_{Σ_E} for increasing values of Δ. A one percentage unit increased in Δ resulted in the total number of infectives being 0.003922% larger than that of a model were no increase in susceptibles took place. Similarly, a unit increase in Δ resulted in the total number of exposed seeing a 0.007113% increase when compared to a model where no increase in susceptible recruiting occurred. As Δ increased, a relatively large increase was seen in the number of susceptibles in the population at the end of the model. A one percentage unit increase in Δ resulted in a 0.8502% increase in the number of susceptibles after 10 years compared to a model with no increase occurring, whereas the number of recovered individuals after 10 years decreased approximately -0.1501% for every percentage unit increase in Δ. The effects on the sensitivity values can be seen in figure 7.6.

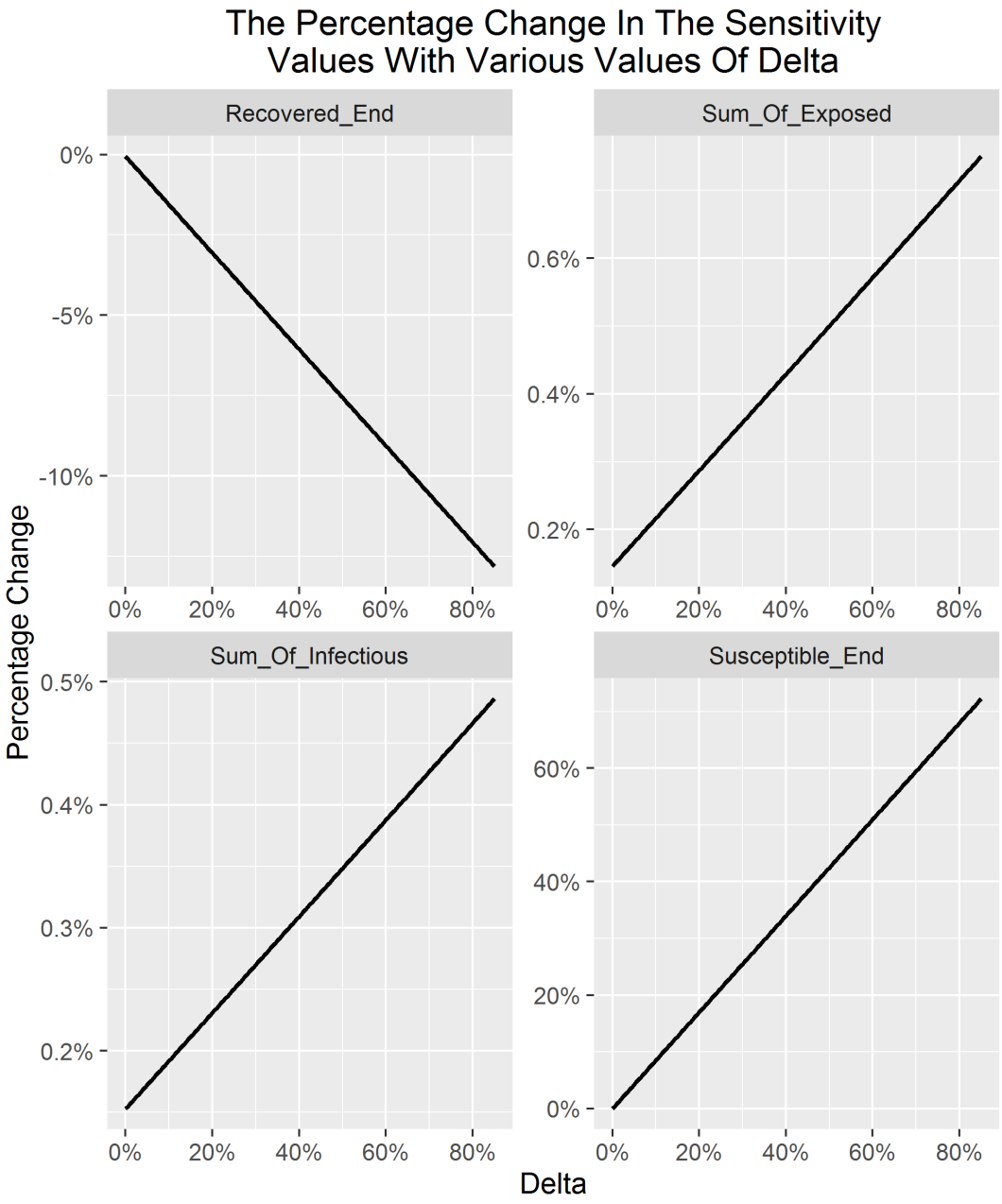

Figure 7.6: The Difference Between The Sensitivity Values For A Model Using w_1 Continuously And A Model The Suddenly Uses $\ddot{w_1}$ At $T = 144$. e.g. If w_1 increased from 15% to 55% (a 40% increase) at $T = 144$, then over the next ten years the total number of infectious would increase approximately 0.3% when compared to a model where w_1 did not increase.

For varying values of Δ, the model estimates that the number of additional cases per year would approximately be between 1 and 3 cases for the local population.

Additional Analysis on The Basic Reproductive Number

The basic reproductive number previously established for the local population within the model with no interaction is given below.

$$R_{(0)L} = \frac{\beta_2 k_2 w_1}{(k_2+\mu)(r_2+\mu+\mu_{I_L})}$$

This is dependent on w_1. Within this scenario w_1 has been replaced with the original w_1 from $t = 1, ..., T$ and replaced with $\ddot{w}_1 = w_1 + \Delta$ from time T onward. This is a function of time and can be represented as the following function

$$\hat{w_1}(t) = \Delta H(t-T) + w_1$$

Where $H(x)$ is the Heaviside step function[138], and take value one when $x > 0$ and value zero when $x < 0$. Hence when $t - T > 0$ or $t > T$ the variable $\Delta H(t-T) = \Delta$, similarly when $t - T < 0$ or $t < T$ the variable $\Delta H(t-T) = 0$.

The model has transitioned from constant parameters to having one parameter a function of time. Within the §5, for the seasonal model, the transmission parameters were functions of time. The basic reproductive number was derived using averages of parameters. This method can be implemented now to achieve an estimate for the basic reproductive number

within this scenario. Assuming the model is run until $t = T_{end} > T$, the average of $\hat{w_1}(t)$ is

$$\frac{1}{T_{end}} \int_0^{T_{end}} (\Delta H(t-T) + w_1) dt$$

$$= \frac{\Delta}{T_{end}} \int_0^{T_{end}} H(t-T) dt + \frac{1}{T_{end}} \int_0^{T_{end}} w_1 dt$$

$$= \frac{\Delta}{T_{end}} \int_0^{T} 0 dt + \frac{\Delta}{T_{end}} \int_T^{T_{end}} 1 dt + \frac{1}{T_{end}} \int_0^{T_{end}} w_1 dt$$

$$= \frac{\Delta}{T_{end}} \int_T^{T_{end}} 1 dt + \frac{1}{T_{end}} \int_0^{T_{end}} w_1 dt$$

$$= \frac{\Delta}{T_{end}} (T_{end} - T) + \frac{1}{T_{end}} (w_1 T_{end})$$

$$= \Delta - \frac{\Delta T}{T_{end}} + w_1$$

Hence the altered basic reproductive number calculated using averages is

$$R_{(0)L} = \frac{\beta_2 k_2 (\Delta - \frac{\Delta T}{T_{end}} + w_1)}{(k_2 + \mu)(r_2 + \mu + \mu_{I_L})}$$

Within this scenario $T = 144$ and the parameter values have already been established, hence

$$R_{(0)L} = 0.73(\Delta - \frac{\Delta 144}{T_{end}} + 0.35)$$

If we allow $T_{end} \to \infty$ we have

$$\lim_{T_{end} \to \infty} (R_{(0)L}) = \lim_{T_{end} \to \infty} (0.73(\Delta - \frac{\Delta 144}{T_{end}} + 0.35)) = 0.73(\Delta + 0.35)$$

The value $0.73(\Delta + 0.35)$ is greater than one when $\Delta > 0.99$. Seeing as w_1 is restricted to being a proportion between zero and one, $w_1 + \Delta$ must also be between zero and one. Hence, as $w_1 = 0.35$, Δ has upper bound 0.65. This implies given the parameter set and initial conditions, the basic reproductive number cannot be greater than one for all feasible values of Δ.

In additional, if $\frac{T}{T_{end}} \to 0$ as $T_{end} \to \infty$, the basic reproductive number is independent on of the value T. As long as this condition holds, the time point the new parameter $\ddot{w}_1 = w_1 + \Delta$ is introduced does not impact whether R_0 will be greater than one.

7.3.2 Optimal Intervention For A Seasonal TB Model

The seasonal model established in §5.2 will be simulated with reduced transmission and increased recovery rates. This is done to set a foundation for further analysis to be done. Initially, an optimal change in parameters is established through numerical simulation. A six month change in parameters is staged at various time intervals throughout the year and the results tabulated. Then, given the optimal time interval, varying degrees of changes to parameters are assessed ranging from a very small impact in parameter values up to a very large.

Optimal Intervention Period Given Fixed Intervention Effects

Methods

This section numerically assesses the optimal time of year to change certain parameters for a seasonal epidemic model in order to reduce the total infectious. This scenario was evaluated using the initial conditions, parameter set, and time intervals established in §5.4. The transmission rates and recovery rate determined were $\beta_0 = 0.1575, k_0 = 0.00554$ and $r = 0.016$. In this section, it will be assumed a 10% reduction in the transmission parameters will occur, and a 10% increase in the recovery rate parameter for a six month period. This results in the alternative parameters $\ddot{\beta}_0 = 0.14178, \ddot{k_0} = 0.005$ and $\ddot{r} = 0.0176$. These parameters will be used in the model for six consecutive months, then the original parameters will be used for the following six months. The parameters will alternate continuously in this way throughout the duration of time the model is run. The sensitivity values $y_{\Sigma I}$, $y_{\Sigma E}$, $y_{S(144)}$, and $y_{R(144)}$ established in §7.2 for the seasonal model will be used to assess the impact. The optimal six month period within the year will be found by simulating the model with the altered parameters at varying time intervals (January through June, February through July, April through August etc.) and finding the six month period that minimises the sensitivity parameter $y_{\Sigma I}$.

Results

The results over time of varying reduction to parameters are illustrated in figure 7.7. Table 7.13 details percentage change on the sensitivity values for various times.

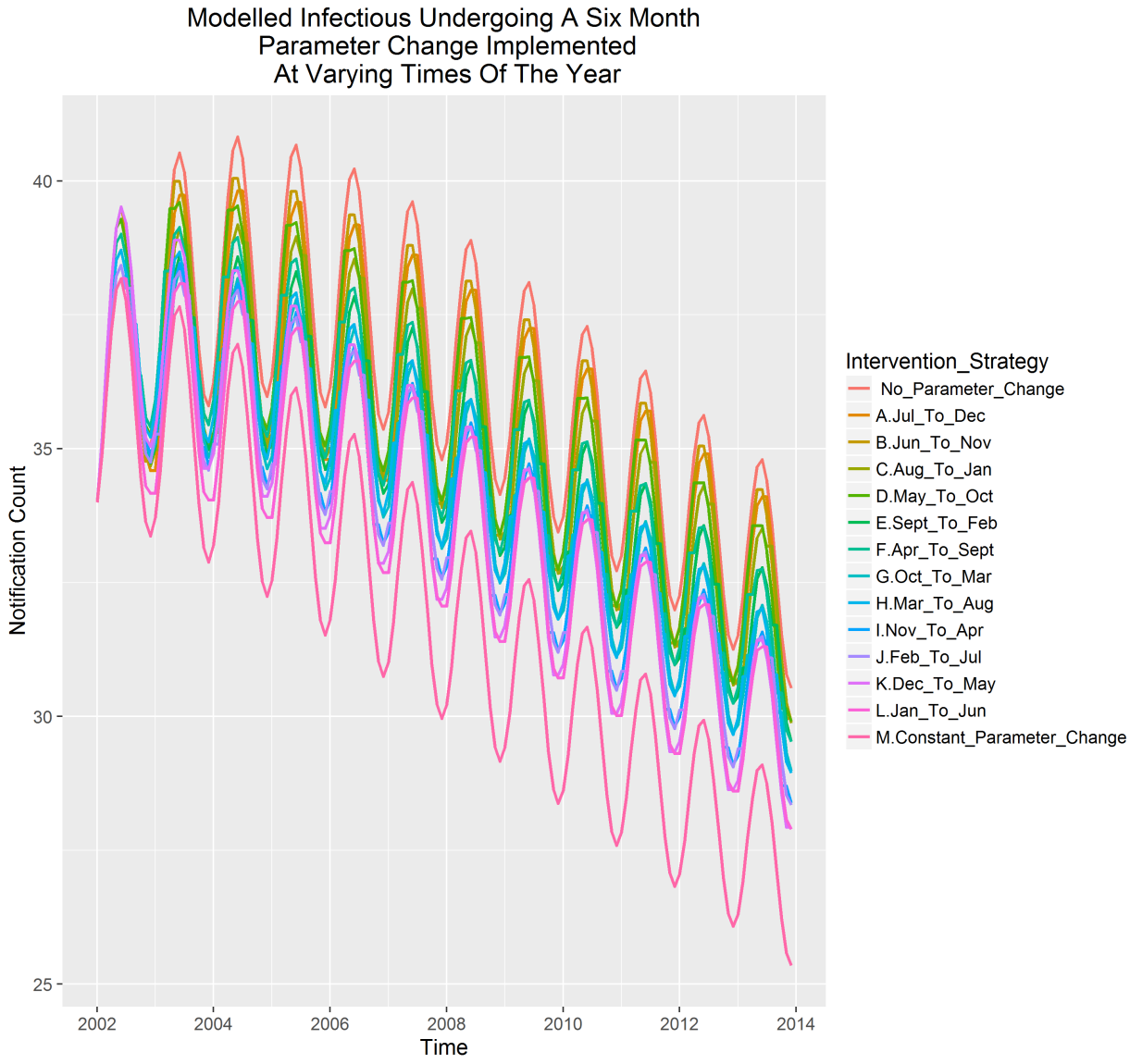

Figure 7.7: The Seasonal Infectious Compartment Over Time With Various Interventions Staged

Time Parameters Were Changed	Reduction In Total Infectious	Time Parameters Were Changed	Reduction In Total Infectious
Constant	-12%	Jun_To_Nov	-2%
Jan_To_Jun	-8%	Jul_To_Dec	-2%
Feb_To_Jul	-7%	Aug_To_Jan	-3%
Mar_To_Aug	-5%	Sept_To_Feb	-4%
Apr_To_Sept	-4%	Oct_To_Mar	-6%
May_To_Oct	-2%	Nov_To_Apr	-7%
		Dec_To_May	-7%

Table 7.13: The Percentage Change for Varying Six Month Intervention Intervals on the Sensitivity Value $y_{\Sigma I}$, when Compared to a Model without Intervention.

January to June was numerically calculated to be the optimal time to reduce parameter values. During this time interval is when the model increases monotonically each year. It can be concluded that the model suggests that the optimal effects of reduced transmission rates and increased recovery rates are seen when there is an increase in notifications. When reduced transmission rates and increased recovery rates were implemented constantly all year, there was an 12% reduction in notifications when compared to a model with no change in parameters. When the rates were altered from January to June, a 8% reduction in notifications was seen. The six month period having the least impact on notifications was July through to December, implementing a change in rates over this period resulted in a 2% reduction in total notifications.

Varying Intervention Effects

Methods

This section numerically examines varying impacts on model parameters. Denote the change on the parameters $0 \leq \delta \leq 1$. Given this change, the transmission parameters without change (β_0 and k_0) become $\ddot{\beta}_0 = \beta_0(1-\delta)$ and $\ddot{k_0} = k_0(1-\delta)$ and the recovery parameter becomes $r(1+\delta)$. Evaluating a range of values for δ will now occur, this in

contrast of the previous section, in which $\delta = 0.10 = 10\%$ was evaluated. Two scenarios will now be assessed: a six month change in parameters staged from January to June, and a constant change staged continuously.

Results

The results suggest implementation of a six month change in parameters can be highly effective when compared to that of a constant change in parameters. Figure 7.8 and table 7.9 detail the percentage decline in infections for varying values of δ.

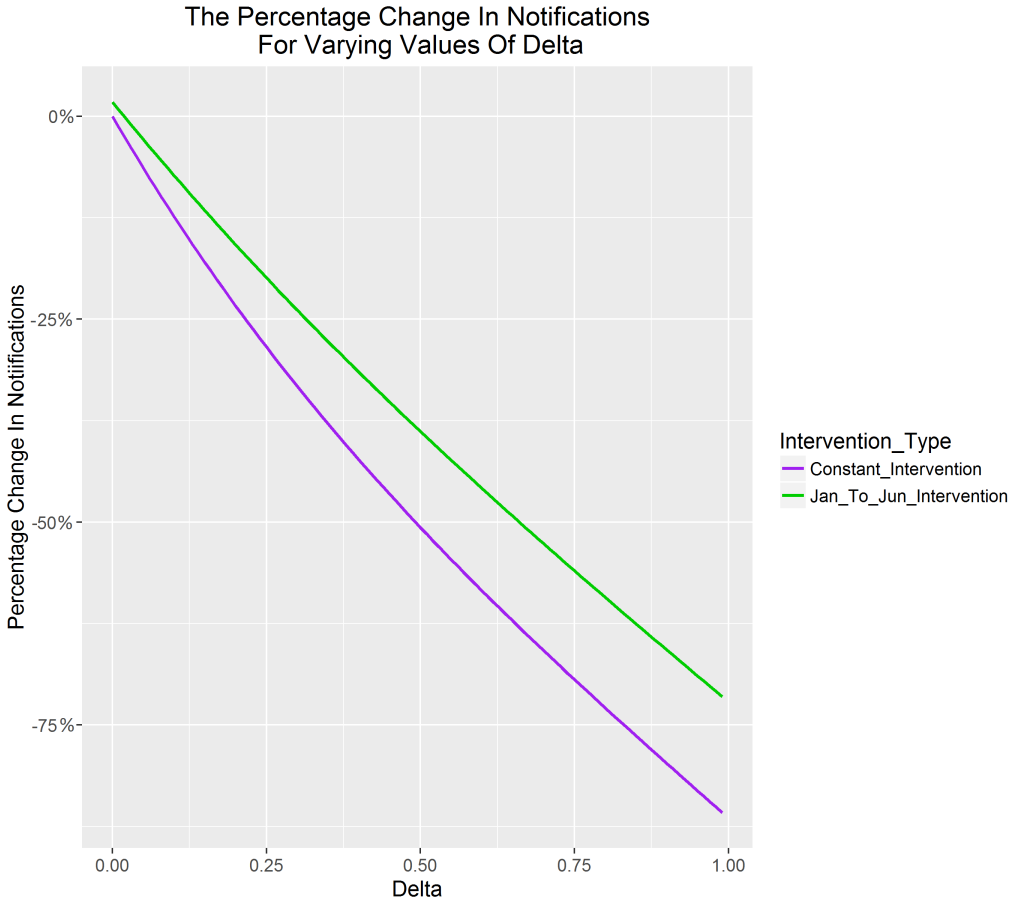

Figure 7.8: Percentage Reduction in Notifications (Between Constant Parameter Change and Seasonal) for both Intervention Strategies and Varying Values of δ (Delta)

Table 7.14 describes the sensitivity of $y_{\Sigma I}$, $y_{\Sigma E}$, $y_{S(144)}$, and $y_{R(144)}$ with respect to δ.

Model Sensitivity	Change Of Parameters From January To June				Change Of Parameters All Year			
δ Value	$y_{\Sigma I}$	$y_{\Sigma E}$	$y_{S(144)}$	$y_{R(144)}$	$y_{\Sigma I}$	$y_{\Sigma E}$	$y_{S(144)}$	$y_{R(144)}$
0% (Count)	5224	29558	693250	3927210	5224	29558	693250	3927210
10%	-8%	2%	-2%	-2%	-11%	-1%	0%	0%
20%	-15%	1%	-2%	-2%	-22%	-2%	0%	0%
30%	-23%	1%	-2%	-2%	-32%	-2%	0%	0%
40%	-31%	2%	-2%	-2%	-41%	-1%	0%	0%
50%	-38%	2%	-2%	-2%	-49%	0%	0%	0%
60%	-45%	3%	-2%	-2%	-58%	1%	0%	0%
70%	-52%	4%	-2%	-2%	-65%	3%	0%	0%
80%	-59%	5%	-2%	-2%	-72%	6%	0%	0%
90%	-65%	7%	-2%	-2%	-79%	9%	0%	0%
100%	-72%	9%	-2%	-2%	-86%	12%	0%	0%

Figure 7.9: Percentage Change in Sensitivity Values $y_{\Sigma I}$, $y_{\Sigma E}$, $y_{S(144)}$, and $y_{R(144)}$, Between the Varying Values of δ (Delta)

The model suggests that the difference between the reduction in the number of cases for a six month change in parameters and that of a constant change of parameters does not remain constant for increasing values of δ (e.g. for $\delta = 10\%$ the six month change in parameters reduces cases by 8%, whereas the constant change of parameters reduces cases 11% - this is a 68% ratio between to two percentages. For $\delta = 90\%$ this ratio grows to 80%, which indicates for larger values of δ a six month reduction becomes more effective).

7.4 Conclusion

In this chapter a sensitivity and scenario analysis was conducted of the models developed in §5 and §6. The model parameters and initial conditions for those parameters were used

in this chapter. A series of sensitivity values were constructed and the effects for varying parameter values on those values evaluated. Key parameters were identified and various forms of changes on those parameters were discussed. A scenario analysis was then conducted within which two primary scenarios were examined. Given some limitations, the model was simulated to investigate an increase in the number of susceptible individuals in the population at a specific time point. It was found that, given an indefinite duration for the model, the time point an alternative vaccination rate is implemented is considered to be arbitrary. In addition, it was also found that, given seasonal dynamic of TB, a change in parameters was found to be effective at times that coincided with an increase in notifications. It was also found a seasonal change in parameters could be as effective as a year round change in parameters.

Chapter 8

Discussion

8.1 Introduction

In this chapter we review the findings of thesis and examine the relationship to other research found in literature. This thesis has examined and modelled notifications of TB within an Irish national cohort. Ireland, like most western European countries, has shown a progressive reduction in the frequency of TB over the last century. Since 2013, Ireland has been categorized as a low incidence country by the World Health Organisation. Despite this, Ireland saw a resurgence in notifications, specifically during the Celtic Tiger economic boom [139]. This resurgence was believed to have been caused by an increase in foreign-born notifications. This study independently gave an evaluation of the epidemiological situation from 2002 to 2013.

The study applied various statistical and mathematical analyses. One of the primary objectives was to establish an empirically accurate model. In an attempt to formulate such a model, the study acquired national TB surveillance data. Following ethical approval, analysis of this data was conducted. The data was acquired from the HPSC which collects various information on all notified cases of TB in the Republic of Ireland. It was assumed that the data was reliable in terms of describing incidence of TB in Ireland. After the dataset was evaluated, two main factors of the profile of the disease were found: (i) sea-

sonality of notifications and (ii) an increase in the proportion of foreign-born notifications over time. A systematic approach to data exploration was then followed by a similarly systematic approach in establishing an empirically accurate model. A systematic search strategy was implemented identifying an appropriate model. Given the two key findings of the exploratory analysis, a systematic search was then implemented that resulted in two viable models being identified. These models were selected and refined to accommodate an Irish setting.

The results of the exploratory analysis suggested numerous pathways the study could have taken. In addition to the analysis of the Irish traveller community (see Appendix E), notably high notification rates were calculated for other vulnerable populations such as: refugees, prisoners, and the unemployed. Combining the results from the data exploration with literature, the study elected to construct and analyse a model which focused on the impact of migration on TB notifications. In addition, and due to the lack of research completed on seasonality of TB in an Irish and global setting, the study also elected to construct and analyse a model considering seasonality.

8.2 Data Quality

8.2.1 TB Data

The data from which a lot of the work in this thesis is conducted was obtained from the Irish national surveillance organisation, the HPSC. The HPSC uses a conservative EU-based clinical definition of TB. The disease is notifiable in Ireland under the Infectious Disease Regulations 1981 (and subsequent amendments). Only symptomatic cases are reported [200]. Despite this, this study must recognise the potential for incomplete or incorrect data from this data source. Possible areas of under-reporting include lost to sight cases, or the possibility of false-negatives positives occurring within the diagnosis procedure, or indeed over reporting with false-positives occurring.

The study further concedes that the sample of data acquired is only a splice of all data. The years 2002 through to 2013 are considered. The years prior this are not reported in

this study. Potential trends in the data could be missed by not having data prior 2002.

The effects of under/over reporting could drastically alter model projections and the descriptive analyses complete, which is a recognised limitation.

8.2.2 Denominator Data

When denominator data are used, they were acquired from the Central Statistics Office (CSO), Ireland [201] or the World Bank [196]. These data are used in order to calculate vital statistics for modelling and crude incidence rates. The CSO only has historical data for 2002, 2006, and 2011. This is a limitation of this study, as interpolation/extrapolation was required on the acquired data in order to get missing years. The CSO data and the World Bank both are subject to reporting error, which this study acknowledges as a limitation. This could also drastically alter model projections and the descriptive analyses complete.

8.3 Foreign-Born Tuberculosis

8.3.1 Discussion On Statistical Results

Foreign-born TB notifications have seen a rise in multiple other countries in recent years. Appendix D identifies nine European countries where low to medium national incidence rates were evaluated. All countries had an increase number of foreign-born TB notifications from 2000 to 2013 and three countries saw an increase in national incidence. This increase in notifications was not unique to Europe. Other countries such as the USA [141] Canada[142], and New Zealand [143], all have witnessed various forms of increases in foreign-born TB rates. Prior descriptive statistical work had been done on foreign-born notifications in Ireland, evaluating data from 1998 to 2005 [1]. The increase in foreign-born notification was noted then. While the study did not conduct a thorough statistical analysis of national surveillance data, it did highlight the age distribution of foreign and native-born cases. This result was replicated from 2002 to 2013 data in §4. It is apparent from figure 8.1 the two distributions have changed very little over the years.

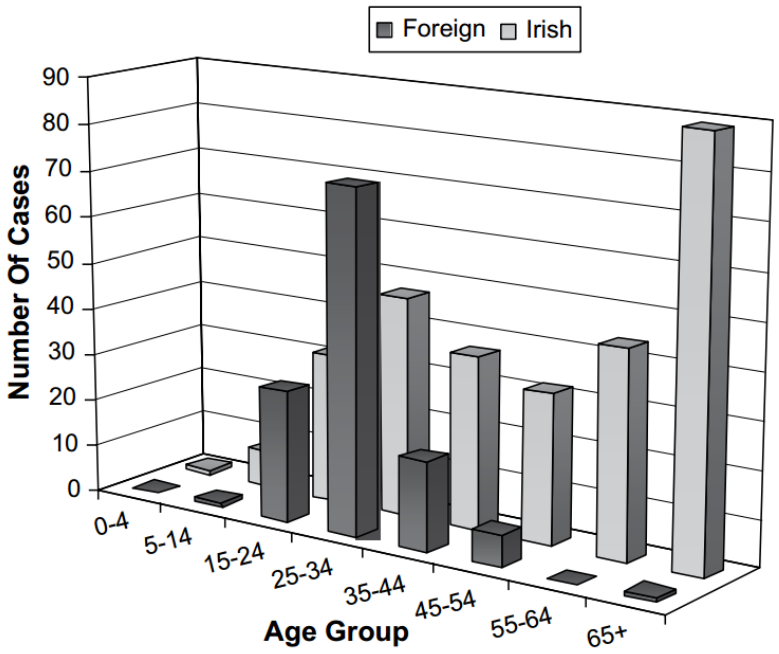

Figure 8.1: Source: [1] Age Distribution Of Foreign and Native-Born TB Notifications From 1998 to 2005.

For individuals with TB, this study found a significant difference between the foreign-born composition and native-born composition when the following variables are taken into account: gender, employment status, current living arrangement, and diagnosis type. This suggests either the composition of the underlying populations differ significantly or, that the populations are both similar and that foreign-born individuals experience TB differently to native-born. For either event, the model had to accommodate this apparent difference between the populations. In order to achieve this, it was decided that an approach using different compartments and parameter sets would help model the individual populations. Similar results were found in other developed countries. In 1998, the work of Chin and colleges in the USA [140] found homelessness, drug misuse, and positive HIV/AIDS status to be infrequent in the foreign-born population with TB. A UK study [144] using data from 2000 to 2011 found significant differences in variables such as gen-

der, age, race/ethnicity, site of disease, and HIV status between foreign and native-born populations. That study also notes that five countries (India, Pakistan, Bangladesh, Somalia and Zimbabwe) contributed to 61% of all foreign-born notifications. This study finds similar results, four of those countries (India, Pakistan, Somalia, and Zimbabwe) were among the top ten contributors of foreign-born TB in Ireland, this distribution of notifications remains constant over time (Pakistan being the top contributor in 2002, and India being top contributor in 2013).

An increase in notifications was documented in the USA from 1985 to 1992. Figure 8.2 details the research of Cantwell and colleagues [87], in which analysis of USA TB surveillance data was conducted.

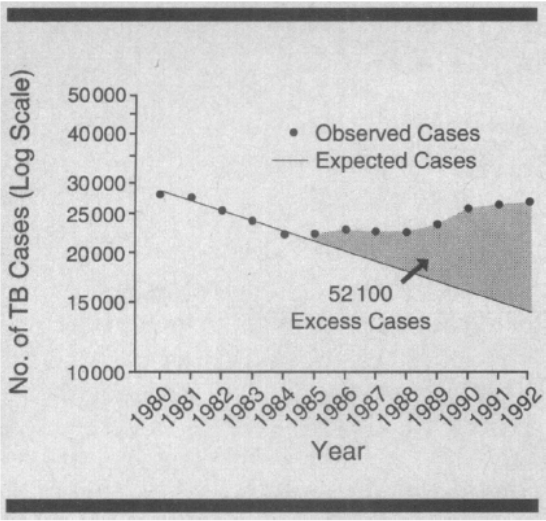

Fig 1.—Excess tuberculosis (TB) cases, United States, 1985 through 1992.

Characteristic	Total Excess Cases, %
Sex	
Male	70
Female	30
Race/ethnicity	
White	18
Black	29
Hispanic	33
Native American	<1
Asian	20
Age, y	
0-4	4
5-14	4
15-24	9
25-44	48
45-64	22
≥65	13

Table 2.—Distribution of 52 100 Excess Cases of Tuberculosis in the United States, 1985 Through 1992

Figure 8.2: Source: [87] Figure (Right): A Rise in US TB Notifications, 1985 through 1992. Table (Left): Demographics of Excess Cases.

Cantwell noted: *"From 1986 through 1992, the number of foreign-born cases increased by 2345 (48%) and accounted for 60% of the total increase in the number of*

US cases". The increase observed in Ireland appears to show a similar trend to that experienced of the USA from 1985 to 1992. Figure 4.5 illustrates very similar trends to Figure 8.2 (above). The study defines a distribution of people as "excess" and examines the demographics of that distribution. The "excess" cases were primarily found to be foreign-born cases. Similar trends and demographics were observed for Irish notifications. The mean age of the foreign-born population in Ireland over the review period was calculated to be 33 years of age. This is similar to the USA study in which it was found that the mode age of "excess" cases was within the 25-45 category. With regard ethnicity, both studies used different categories to place individuals. However, diverse proportions of ethnicity were observed within excess cases for the US study and diverse proportions were also observed in Irish data for foreign-born notification (Table 4.11).

8.3.2 Mathematical Modelling

The mathematical model used here was adapted from Jia and colleagues [93] with some modifications to accommodate an Irish setting. Due to the complexity of the foreign-born population it was decided to limit what would be modelled. The migrant population was used and defined as a subset of the foreign-born population. The justification of this was that the primary purpose of the chapter was to model TB and not to simulate a foreign-born/native-born population model. Consequently, the top contributors were selected for study. The top contributors were identified as those countries that contributed to 80% of foreign-born notifications (20 birth countries in total). Two models were constructed building on the original model identified during the literature review.

Rationale For Modelling

The first model considered did not allow for interaction between migrant and local populations while the second incorporated a two way interaction occurring between these populations. The reason for using the first model was that a previous systematic review suggested foreign-born notifications do not effect native-born [94]. The second model was

developed primarily for two reasons. Firstly, to evaluate if the assertion above regarding the lack of importance of interaction between populations was accurate. Secondly, the study from which the model is derived mentioned the importance of expanding the model to include the impact of interaction, and suggested a two way interaction model which to be used for further analysis. As a consequence it was decided that two models would be presented here for simulation and qualitative analysis.

Discussion Of Results

The qualitative analysis showed that the basic reproductive number of the entire population was the maximum of the basic reproductive number for the local population and the basic reproductive number for the migrant population. The parameters indicate an outbreak occurring within the migrant population $(R_{(0)M} = 1.76 > 1)$ and not for the local population ($(R_{(0)L} = 0.257 < 1)$). This result coincides with the increase in foreign-born notification in European countries [145].

Furthermore, the fact that the basic reproductive number was greater than one in one group and not in the other supports the suggestion that the populations have not interacted, as if interaction was occurring, the likelihood of an outbreak occurring in both populations or neither would increase as there would exist some form of correlation between the two populations. This result ultimately supports the work of Sandgren et al. [94].

Similar results were found between the simulations in this study and the study identified during the literature review (§3.5.3). That model simulated Canadian data. Here, the the basic reproductive number was calculated as greater than one for the migrant population and less than one for the local population. Figure 8.3 shows a distinct increase in prevalence rates within the foreign-born population and a decrease in the native-born population.

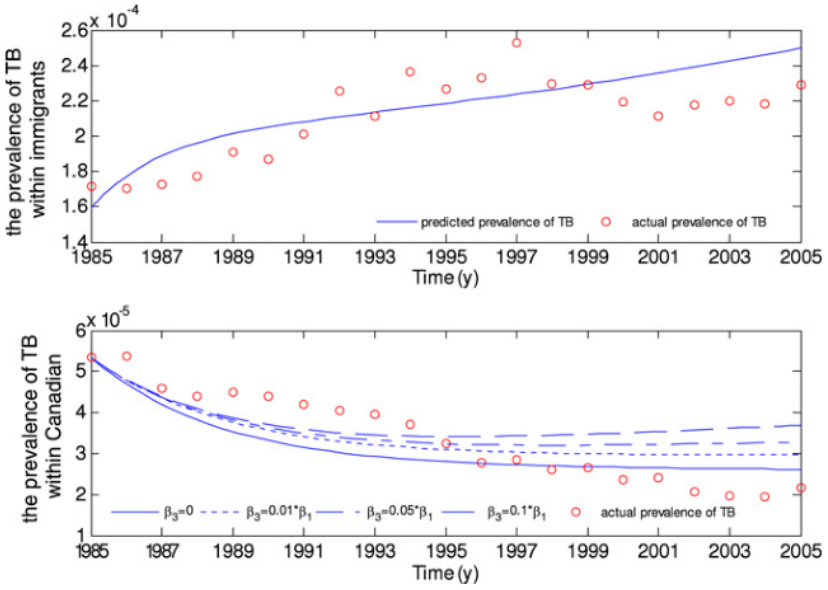

Fig. 5. A match between the basic model and the prevalence of TB in Canada from 1985 to 2005. The parameter values are $\pi = 200\,000$, $\Lambda = 80\,000$, $k_1 = k_2 = 0.00256$, $r_1 = 0.021$, $r_2 = 0.086$, $\mu = 0.0143$, $\mu_{I_M} = 0.461$, $\mu_{I_L} = 0.159$, $\beta_1 = 0.00000093$, $\beta_2 = 0.00000017$, $p = q = 0.8$. These parameter values give $R_{M0} = 3.9794$ and $R_{L0} = 0.6035$.

Figure 8.3: Source: [93] Numerical Simulation Results of Model From Literature Review. (§3.5.3) A Canadian Model with One-Way Interaction from Migrants to Locals.

The model was simulated to numerically evaluate the effects of increased recruitment into the susceptible population. It was found increased recruitment had marginal effects on the total number of infectious and exposed 10 years into the future. The number of susceptible individuals after 10 years was found to be notably larger when compared to a model without a change in vaccination procedure. A reduction in the number of recovered individuals was also noted. The analysis went on to examine the effects on the basic reproductive number for the local population. Given the parameters and conditions established, it was found to be impossible for the basic reproductive number to be larger than one, regardless of the increase in susceptibles. While this study does not put forward a recommendation, a report by HIQA recommended Ireland switch to a selective vaccination

strategy as opposed to universal vaccination [146]. The report estimated that an additional 3.6 cases would occur each year if selective vaccination was used instead of universal vaccination. If there is any legitimacy, in terms of representing Ireland, can be given to the scenario analysis conducted, then the results of this section corroborate the HIQA report.

Strengths Of Modelling

The model accurately partitions a very clear divide occurring within notifications, between foreign and native-born cases. Each of the parameter sets calculated converged at their local optimal values, which suggests they are reliable estimates. The models obtained from the systematic search were adapted to an Irish setting successfully and simulated with the relatively low error. These models provide a foundation for future modelling to occur and for methods to be expanded on. The results of the chapter corroborate literature, which indirectly indicates a type of legitimacy of the modelling approach. The calculation of the basic reproductive number, and the subsequent calculation and difference in values between the two populations, highlights a distinction between the foreign and native-born populations. This can provides insight into treatment methods to lower overall incidence.

Limitations Of Modelling

For both migrant models, the residuals consistently suggest under-prediction was occurring (albeit slight under-prediction). The model which incorporates interaction between local and migrant populations has limitations which should be acknowledged. The interaction parameters β_1^* and β_2^* were both estimated using inferential methods. Within the ABC method both parameters were assigned a uniform distribution with range between 0 and 0.1. This was selected to be one tenth the range of their counter parts β_1 and β_2. It was later established within the sensitivity analysis that β_2^* contributed to the total sum of infectious within the migrant population more so than β_1, even after the effects of the other parameters were controlled for. This implies the infectious local population contributes more to the number of infectious within the migrant population than the infectious migrant population does. If it is accepted that there is, in fact, an interaction between populations this is a plausible scenario, however, it does not coincide with literature [94].

In addition, in order to calculate the basic reproductive number for this model, the assumption of no recruitment into the exposed population was occurring, which in itself is a limitation. In general, the foreign-born population is a very complex categorization of a population, the models proposed are undoubtedly oversimplified representations of the true processes they aim to model.

8.4 Seasonality Of Tuberculosis

8.4.1 Discussion On Statistical Results

Some studies have shown variable periods of peak seasonality in TB incidence rates in late winter to early spring in South Africa [118], during summer in UK [119] and Hong Kong [120] , and during spring and summer in Japan [117].

While the cause of seasonal TB remains unknown, a possible link between vitamin D deficiency and impaired host defence to tuberculosis infection leading to primary TB has been hypothesised [122]. Furthermore, in winter and spring, viral infections such as influenza are more prevalent and cause immunological deficiency leading to reactivation of Tuberculosis [123].

This study examined claims of seasonal attributes within Irish TB and then proceeded to model and investigate these attributes. Seasonal TB models have previously been constructed and evaluated within countries such as China [125] and the United States [121], however, current research on the seasonal property of TB is very limited and the cause of the seasonality is for the most part unknown. The occurrence of a seasonal high and low periods also appears to differ from country to country.

This study identified seasonality within Irish notifications through data independent of external research. Various analyses were completed following the methods of Box and colleagues [110] to detect seasonality including simple moving averages, box plots, and autocorrelation plots. Seasonality was concluded to be a significant characteristic in

notifications over time. This study found that the high seasonal period for TB notifications in Ireland was during the spring and summer months. As discussed in §5.1, varying high seasonal periods occur from country to country. The high seasonal period in Ireland is similar to that of the UK, Hong Kong, the USA, and Japan. One study [148] using USA surveillance data from 1993 to 2008 reported results very similar to this study's findings. In that study there was a 21.4% increase in cases during the high season compared to that of the low season. This study found a 20.4% increase in notifications during the high season. A study conducted in New York, USA [132], conducted a similar autocorrelation analysis to this study and found significant positive correlations not only every 12th month, but significant negative correlations every sixth period (figure 8.4). This indicates that the notifications sixth month prior, regardless if it's a high or low period, negatively correlate with the current months notifications.

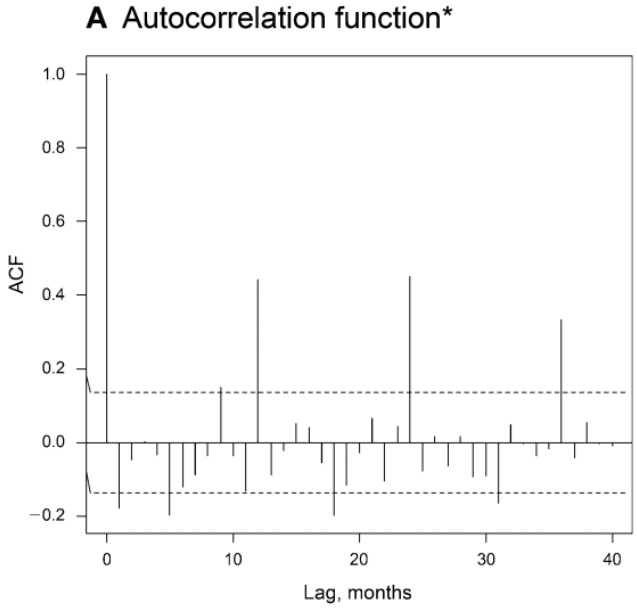

Figure 1 Autocorrelation function of all verified tuberculosis cases, New York City, January 1990–December 2007. *The dotted lines represent two standard errors. ACF = autocorrelation function.

Figure 8.4: Source: [132] Autocorrelation of Monthly Notifications In New York

It should also be noted that, given the available data associated with demographic

variables, this study did not find strong evidence to suggest that seasonality affects those demographics. No significant seasonal difference was detected for individuals with different gender, employment status, birth country, or race/ethnicity. With regards to "living arrangements" a significant difference was detected however for the prisoner population within the study. This maybe coincidental as the outbreak that occurred within the prisoner population occurred within the high seasonal period.

The seasonal attribute was unknown to both the HPSC researchers and the research team prior to discovery in this study. This discovery increases insight into the underlying dynamics of the disease and specifically relates to research in Ireland.

Typically, seasonality is modelled using time series methods. Autoregressive processes and exponentially weighted moving averages are the most common methods of forecasting and have been previously used to model epidemiological processes [154]. An epidemic model was selected for this research as it presented a deeper insight into the underlying dynamics of the disease.

Multiple reasons have been put forward, to explain the underlying causes of seasonality. However, the two leading hypotheses developed so far relate to:

- Vitiman D deficiency - Vitamin D has been shown to enhance cellular immunity against TB [57]. It has been hypothesized that the decline in vitamin D levels over the summer months contribute to the activation of latent TB. This translates to variable progression rates from latent to active TB throughout the year.

- Increased transmission rates due to overcrowding - Generally, the population spends more time indoors in winter months than in summer months. The resultant potential overcrowding, increased humidity, and low airflow provide a suitable environment for TB to transmit as research has shown.

Both of these factors were considered within the seasonal model put forward. The possible seasonal activation of latent TB is considered by allowing the progression rate parameter to become a function of time. Similarly, the transmission rate parameter is given a time dependent periodic functions represents overcrowding. As no prior deterministic modelling for TB had been developed for Ireland in the past, and because of the

apparent gap in the literature globally for work completed on seasonality in TB, this study established parameters and initial conditions for such a model.

8.4.2 Mathematical Modelling

Rationale For Modelling

The rationale behind using a seasonal deterministic model was to provide the basis for future modelling and to support further analysis of seasonal variation in TB. The exact reason why TB case notification rates vary by season is unknown, although several factors have been suggested. This is ultimately an open area of research and this study aims at contributing to it through the construction, analysis, and simulation of a seasonal epidemic model.

Discussion Of Results

The seasonal model was successfully constructed, analysed, and simulated and a basic reproductive number was calculated. For the system as a whole, the basic reproductive number was found to be less than one, indicating the infection will reach a disease-free equilibrium after a long enough period of time. The model predicted a decline in total notifications (8 per year expected), which is being observed in other European countries (Appendix D).

The seasonal model was selected to simulate a change in parameters to investigate the effects on notifications. With respect to the results of the scenario analysis, similar modelling completed by Griffin [46] on optimal malaria interventions recommended: *"The best time for indoor residual spraying or a vaccine which reduces infection rates is just before the high season"*. This indicates the effects of the interventions should take place during the high season period and, for that reason, lead to reduced transmission during the seasonal high period. The findings presented in §7.3.2 seem to corroborate this result, suggesting that a lower transmission and progression rate during the seasonally high pe-

riod reduces infection. In addition, the numerical analysis suggests a seasonal or pulsed intervention can be more cost effective than a constant intervention.

Strengths Of Modelling

Each of the parameter sets calculated converged at their local optimal values, which suggests they are reliable estimates. The model obtained from the systematic search were adapted to an Irish setting successfully and simulated and extrapolated with relatively low error. The model provides a foundation for future modelling to occur and for methods to be expanded on, this model is the first of it's kind to be applied to an Irish setting. The scenario analysis provides a basis for future, more clinically interpretable, scenarios to be evaluated. In addition interventions could be simulated using this model, or a similar model.

Limitations Of Modelling

Some of the limitations of the seasonal model include the inability to calculate the basic reproductive number using traditional approaches. This study used an averaging method to overcome this. However, this can be considered a limitation as it essentially represents a basic reproductive number for a model without seasonal transmission parameters. The seasonal functions assigned to the transmission rates are also subjectively selected. A possible alternative approach would be to allow an additional amplifier parameter on the periodic part of the function to allow the range of values to increase for each period. Also, it would be preferable to investigate if seasonality is increasing or decreasing in strength over time, as this study assumes seasonality is constant. In addition, seasonality may not necessarily effect the transmission parameters (little research is present as to what it does effect), and this study acknowledges it makes the assumption that it does.

8.5 Modelling Methods

8.5.1 Inference Methods

In an effort to calculate model parameters, established statistical inference methods were implemented. Bayesian methods are typically used in estimating parameters within compartmental models. Two common methods used in epidemiology are (a) the Approximate Bayesian Computation (ABC) rejection algorithm and (b) the Metropolis-Hastings algorithm [47]. The Metropolis-Hastings algorithm is used when it's difficult to sample from the target distribution (e.g., when the prior isn't conjugate to the likelihood). So you use a proposal distribution to generate samples and accept/reject them based on the acceptance probability. Within this study, the ABC algorithm served as a method for obtaining the initial parameter set for the Metropolis-Hastings algorithm. It was implemented despite the algorithm itself being a legitimate methodology for calculating parameters. The advantages of the Metropolis-Hasting algorithm is the posterior distribution it generates, and the uncertainty statistics that can be calculated for parameters. The results of both inference methods approximately coincided for both methods for each model excluding the migrant model with interaction (§6.4.4). The limitation of this method include the inability to choose the optimal variance for the jump distributions used when proposing the alternative parameter set.

When estimating parameters using these methods, as the number of parameters requiring estimation increases, the more viable parameter sets can occur. This is due to an increase in the number of dimensions which leads to more global minima occurring in the sums of squares value. The study has not concluded that the parameter sets generated are the sets that optimally fit the data, but rather happen to fit well when compared to thousands of other simulations. Interpretation of the parameters can also become difficult with inferential methods. For example, setting the usually positive death rate parameter to $\mu = -10^3$ may yield the smallest least squares value among all other simulations, however verifying this value in any empirical sense is impossible. As such multiple parameters required estimating and implementation of the above algorithms was limited to only a few parameters, namely the transmission parameters.

8.5.2 Sensitivity and Scenario Methods

For the bio-mathematical and bio-statistical community to continue to have an impact in important healthcare problems, it is clear that identifying uncertainty in models is of key importance. Through classifying this uncertainty, we can identify the parameters that are causing various model outputs. Some of the limitations to sensitivity analysis cited by Drummond [36] include:

1. Variation of uncertain parameters one at a time ignores possible interaction between parameters

2. The analyst has discretion as to which variables and what alternative values are included in sensitivity analysis

3. Interpretation is arbitrary as there are not guidelines/standards as to what degree of variation in results is acceptable evidence that the analysis is robust

The methods used in this study presented by Marino and colleagues [135] attempt to overcome the first point. The partial rank correlation calculates the correlation between an independent and dependent variable after the linear effects on the dependent variable from the remaining variables are considered. In an attempt to overcome the second point, data were used when possible to estimate the standard deviation and hence, the range the parameter values could take within the sensitivity analysis. Given the parameter sets calculated, this study provides a number of significant parameters from the partial rank correlation *p-values*. Thus the partial ranked correlations can be informative in terms of what parameters to target if we want to achieve specific goals. For example, within the seasonal model an increase in the recruitment rate parameter was shown to significantly increase the recovered population towards the end of the simulation, whereas an increase in the death rate parameter and the proportion of individuals being recruited into the susceptible compartment resulted in a decline in the recovered population towards the end

of the simulation. Hence, if a change in the number of recovered or immune individuals within the population was required, attempting an intervention to increase or decrease these parameters would yield the desired results.

8.6 Conclusion

In addition to the extending the overall discussion presented throughout this paper, in this chapter we also acknowledged some limitations of the study's methodology. The study's approach to modelling was discussed and it was posited that other approaches could also be taken. Foreign-born TB was discussed and compared to multiple other countries. Common trends were seen to be emerging. In relation to the seasonality of TB, possible causes of this attribute were considered and a comparison made to other research which considered the the timing of intervention strategies. The limitations of the methodologies used were then discussed and, in particular, the use of inferential methods for parameter estimation. The following chapter concludes the thesis.

Chapter 9

Conclusion and Further Work

9.1 Introduction

This thesis presented a range of recognised statistical and mathematical methodologies used in epidemiology. Two deterministic models were constructed from classical epidemic models and simulated to model TB infections within in the context of an Irish setting. The models considered a seasonal fluctuation in disease notifications and the impact of migrant individuals on total notifications. This chapter summarises the overall results in terms of the main objectives of the thesis and the individual aims of each chapter. The objectives of this study included:

- To describe and analyse existing cross-sectional TB data from a national source.
- To derive a deterministic model that accurately models underlying TB dynamics and incorporates attributes of the aforementioned analysis.
- For each model derive R_0, the basic reproductive number, for calculation and for sensitivity analysis.
- Given data and statistical inference methods, estimate epidemiological parameters and initial conditions for each model.

- For each model, simulate and extrapolate the underlying dynamics and numerically calculate the basic reproductive number.
- Perform a sensitivity analysis on the parameters and provide a scenario analysis for each model.

A brief summary of original findings is now presented to give context to the following conclusions section.

9.2 Summary Of Original Findings

Some of the original findings this study has produced includes:

- A thorough descriptive investigation on Irish national TB datasets.
- Proposal of seasonality in Irish TB notifications.
- Multivariate descriptives for TB.
- Construction three deterministic models with application to TB.
- Calculation of a basic reproductive number for each system proposed.
- Estimation of a range of epidemiological parameters, each with their own individual interpretation.
- Estimation of the basic reproductive number for each system accompanied by uncertainty intervals.
- Sensitivity analysis provided on the aforementioned parameters
- A scenario analysis of two models being considered.

9.3 Conclusions

9.3.1 Mathematical Conclusions

In §5 and §6 two deterministic differential equation models of Tuberculosis were constructed and simulated to model the Irish population. For each model attempts were made at deriving the basic reproductive number, R_0, a value that can be calculated from the model parameters. When R_0 has a value greater than one, implies an outbreak will occur in the population.

For the model considering seasonality, an averaging method was used to calculate the basic reproductive number. Given the parameters estimated, it was determined $R_0 = 0.65 < 1$ with a 95% credibility interval (0.43,0.87) which implies an outbreak would not be imminent in the total population. It was calculated, using a sensitivity analysis in §7, that the three parameters to target in order to reduce R_0 were (i) the proportion of individuals entering into the recovered class w_1, (ii) the transmission rate β_1, and (iii) the infectious death rate d. To reduce the number of overall infections it was identified that targeting the transmission rate β_0, the progression rate k_0, the quick progression rate q, and the infectious death rate d could result in a decrease in the total number of infections. Overall, the model forecast a downward trend in notifications up until the year 2023.

In §7 a scenario analysis was carried out. It was concluded that the number of infections could be effectively reduced if the relevant parameters were changed at the time when the model was witnessing an increase in the number of infections. The numerical results also suggested a seasonal or pulsed intervention could be more effective in reducing cases annually.

For the model considering migration, the basic reproductive number could not be derived. This was due to the fact that the model was designed to recruit latent infections at a constant rate. Because of this a disease-free equilibrium could not be calculated. To compensate for this limitation a basic reproductive number was calculated assuming no

recruitment was occurring in the latent infectious class. Two models were constructed: one that considered no interaction occurring between the migrant and local populations, another that considered an interaction.

For the model not considering an interaction, the basic reproductive number was calculated to be $R_{(0)L} = 0.201$ with a 95% credibility interval (0.05,0.37) which implies an outbreak was not imminent in the local population. For the migrant population the basic reproductive number was calculated to be $R_{(0)M} = 1.75$ with a 95% credibility interval (1.06,2.72). This implies an outbreak has occurred in the migrant population. It was also calculated, using a sensitivity analysis, that the four parameters to target in order to reduce the basic reproductive number in the migrant population $R_{(0)M}$ were (i) the transmission rate β_1, (ii) the infectious death rate μ_{I_M}, (iii) the recovery rate r_1, and (iv) the proportion of individuals being recruited into the recovered class v_2. For the the basic reproductive number $R_{(0)L}$ of the local population, it was identified the three key parameters to target included (i) the transmission rate β_2, (ii) the proportion of recruits entering into the susceptible compartment w_1, and (iii) the infectious death rate μ_{I_L}

For the model considering an interaction, the basic reproductive number was calculated to be $R_0 = 0.832 < 1$ with 95% credibility interval (0.26,1.75) which implies an there exists a chance an outbreak occurred in the population. It was also calculated, using sensitivity analysis, that the two parameters to target in order to reduce the basic reproductive number were (i) the transmission rate β_1, and (ii) infectious death rate μ_{I_M}, both in the migrant population. To target the total number of infections, the appropriate parameter to target in this instance was the progression rate of the local exposed class k_2. When simulation took place of this model the mean residuals were the worst relative to all other models indicating a poor fit.

For both migrant models a downward trend was identified when the models were extrapolated up until the year 2023. An increase is to be expected in the migrant infectious,

however, overall notifications are expected to decline.

Overall, the epidemiological parameters estimated contribute to on-going TB research. With no prior modelling completed in an Irish setting, the construction and simulation of viable models can form a basis for future research to be conducted.

9.3.2 Epidemiological Conclusions

Targeting key parameters in each model can serve as a method of disease prevention. Some of the key parameters identified included the transmission and death rate parameters which, as discussed in §7, indicate the importance of isolating infectious individuals to reduce overall notifications and prevent disease outbreak.

The numerical extrapolations of each model took place from 2013 to 2023. For the seasonal model (§5.6.3), the total number of infections projected are are expected to decrease approximately seven to eight cases per annum from 2013 to 2023. For the migrant model with no interaction (§6.5.4), it was projected that the number of infectious migrants would be expected to increase four to five cases per annum, and expected to increase one to three cases per year if interaction is included. For the local population, it was projected that the infectious population would be expected to decrease by between eight to 10 cases per year for the model not considering interaction and to decrease 10 to 12 cases a year if interaction is considered.

In §7 a scenario analysis was carried out and additional analysis was conducted on the basic reproductive number for the local population. After numerical simulation took place, the effects of altering the proportions of the population entering into the susceptible class was evaluated. If the model is allowed to continue indefinitely, it was found that the time point of vaccination alteration is arbitrary. Through numerical calculation it was also determined that the basic reproductive number would remain less than one regardless of how many newly recruited individuals enter into the susceptible compartment.

For the seasonal model, given a six month intervention, it was numerically calculated that January to June was the optimal time of year to implement an intervention. In this way, the total number of infectious individuals was minimised numerically. It was also concluded numerically that the larger impact an intervention has on the underlying parameters, the more cost effective a six month intervention can be when compared to that of a constant, ongoing intervention.

9.4 Further Work

In §4, an exploratory analysis is conducted in which seasonality was concluded as being a significant factor influencing notification rates. This study assumed seasonality effects all demographics similarly, however it only reviewed seasonality for categories of gender, employment status, currently living status, birth country, race/ethnicity, and refugee status. Whether there exists additional categories which would be significantly effected by seasonality is unknown. Possible future work could include spatial or temporal analysis to examine the effects of seasonality in TB, or to examine the effects of sunlight on vitamin D levels which may be influence seasonality.

The models used in this thesis were based on the work of Jai and colleagues [93] and Liu and colleagues [89]. Although the models incorporated various attributes of TB in Ireland, they may simplify the underlying dynamics and can be improved. For example, it was seen that TB effects vulnerable populations such as homeless, unemployed, and refugees. Constructing a model to consider these populations could be considered in future work. Construction of such a model would enable a more accurate evaluation of a vaccination strategy that targets such vulnerable groups.

Analytical work could be completed in the derivation of the basic reproductive number for the seasonal and migrant models, as problems arose for each derivation for the each one. For the seasonal model, the basic reproductive number should be a dimension-

less rate, and not incorporate seasonality. Further work to calculate the basic reproductive number using numerical methods instead of deriving an explicit formula would be a possible approach here to overcome the seasonality. For the migrant model, working towards a solution of obtaining an R_0 when there is a larger, global, system contributing to notifications. This recommendation is based on the finding of a basic reproductive number greater than one for this population, which warrants further investigation.

For the three models considered, additional work could be carried out on analysis of the endemic equilibrium states. The conditions of stability of each equilibrium could be considered. The predictability of each model could also be examined against other models, such as linear regression or time series models that incorporate seasonality. These model could be used as a baseline to measure the predictability of epidemic models.

The model with interaction (§6.4) could be improved to simulate Irish data. This is evident in the result of poor residual statistics relative to other models. Further work including application of restrictions to the interaction parameters β_1^* and β_2^* could be completed or more statistical analysis to fully quantify the level of interaction occurring, as the models proposed in this study can be considered over simplified. Statistical methods such as cross-correlation, for instance, could be implemented to achieve this.

To conclude, the research demonstrates how a range of models can be implemented for the benefit of surveillance, treatment, policy, and provision. While limitations are highlighted, the results presented can be used to refine existing strategies both in Ireland and globally.

Appendices

Appendix A

Data Tables

Disease Type (N)	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
Pulmonary	269	264	275	278	297	316	294	288	241	257	219	228
Extrapulmonary	96	105	120	129	126	132	129	164	150	122	104	128
Pulmonary & Extrapulmonary	37	34	37	40	40	33	43	26	29	32	33	23
Unknown	8	3	1	1	0	0	1	1	0	2	0	0
Disease Type (%)	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
Pulmonary	65.6%	65.0%	63.5%	62.1%	64.1%	65.7%	63.0%	60.1%	57.4%	62.2%	61.0%	59.8%
Extrapulmonary	23.4%	25.9%	27.7%	28.8%	27.2%	27.4%	27.6%	34.2%	35.7%	29.5%	29.0%	33.6%
Pulmonary & Extrapulmonary	9.0%	8.4%	8.5%	8.9%	8.6%	6.9%	9.2%	5.4%	6.9%	7.7%	9.2%	6.0%
Unknown	2.0%	0.7%	0.2%	0.2%	0.0%	0.0%	0.2%	0.2%	0.0%	0.5%	0.0%	0.0%
Disease Type Incidence	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
Pulmonary	6.87	6.63	6.80	6.73	7.02	7.22	6.56	6.35	5.29	5.62	4.78	4.96
Extrapulmonary	2.45	2.64	2.97	3.12	2.98	3.02	2.88	3.62	3.29	2.67	2.27	2.79
Pulmonary & Extrapulmonary	0.94	0.85	0.91	0.97	0.94	0.75	0.96	0.57	0.64	0.70	0.72	0.50

Table A.1: Notifications (Count, Percentage, Incidence) Categorized By Disease Type

Statistics (N)	Min	Q1	Median	Q3	Max	Mean	S.Dev	Mean Change
Pulmonary	219	253	272	289.5	316	268.83	28.97	-3.73
Extrapulmonary	96	116.25	127	129.75	164	125.42	18.95	2.91
Pulmonary & Extrapulmonary	23	31.25	33.5	37.75	43	33.92	5.93	-1.27
Unknown	0	0	1	1.25	8	1.42	2.27	-0.73
Statistics (%)	Min	Q1	Median	Q3	Max	Mean	S.Dev	Mean Change
Pulmonary	57.4%	60.8%	62.6%	64.4%	65.7%	62.5%	2.6%	-0.5%
Extrapulmonary	23.4%	27.4%	28.3%	30.6%	35.7%	29.2%	3.6%	0.9%
Pulmonary & Extrapulmonary	5.4%	6.9%	8.5%	9.0%	9.2%	7.9%	1.3%	-0.3%
Unknown	0.0%	0.0%	0.2%	0.3%	2.0%	0.3%	0.6%	-0.2%
Statistics Incidence	Min	Q1	Median	Q3	Max	Mean	S.Dev	Mean Change
Pulmonary	4.78	5.54	6.59	6.82	7.22	6.23	0.84	-0.17
Extrapulmonary	2.27	2.66	2.92	3.04	3.62	2.89	0.37	0.03
Pulmonary & Extrapulmonary	0.50	0.68	0.80	0.94	0.97	0.79	0.16	-0.04

Table A.2: Statistics(Count, Percentage, Incidence) Categorized By Disease Type

Strain (N)	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
Standard	410	405	431	445	459	474	465	478	418	410	354	377
MDR	0	1	2	2	4	7	2	1	2	3	5	0
XDR	0	0	0	1	0	0	0	0	0	0	0	0

Strain (%)	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
Standard	100%	99.8%	99.5%	99.3%	99.1%	98.5%	99.6%	99.8%	99.5%	99.3%	98.6%	100%
MDR	0.0%	0.2%	0.5%	0.4%	0.9%	1.5%	0.4%	0.2%	0.5%	0.7%	1.4%	0.0%
XDR	0.0%	0.0%	0.0%	0.2%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%

Strain Incidence	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
Standard	10.47	10.18	10.65	10.76	10.84	10.83	10.37	10.54	9.18	8.96	7.72	8.21
MDR	0.00	0.03	0.05	0.05	0.09	0.16	0.04	0.02	0.04	0.07	0.11	0.00
XDR	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Statistics (N)	Min	Q1	Median	Q3	Max	Mean	S.Dev	Mean Change
Standard	354	408.75	424.5	460.5	478	427.17	38.83	-3.00
MDR	0	1	2	3.25	7	2.42	2.07	0.00
XDR	0	0	0	0	1	0.08	0.29	0.00

Statistics (%)	Min	Q1	Median	Q3	Max	Mean	S.Dev	Mean Change
Standard	98.5%	99.2%	99.5%	99.8%	100%	99.4%	0.5%	0.0%
MDR	0.0%	0.2%	0.5%	0.8%	1.5%	0.6%	0.5%	0.0%
XDR	0.0%	0.0%	0.0%	0.0%	0.2%	0.0%	0.1%	0.0%

Statistics Incidence	Min	Q1	Median	Q3	Max	Mean	S.Dev	Mean Change
Standard	7.72	9.12	10.42	10.68	10.84	9.89	1.09	-0.21
MDR	0.00	0.02	0.05	0.07	0.16	0.06	0.05	0.00
XDR	0.00	0.00	0.00	0.00	0.02	0.00	0.01	0.00

Table A.3: Notifications (Count, Percentage, Incidence) Categorized By Strain Type, Statistics (Count, Percentage, Incidence) Categorized By Strain Type

TB Caused Death (N)	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
Yes	0	6	5	11	10	7	9	10	8	10	3	7
No	1	22	15	23	22	30	24	22	14	12	6	4
Unknown	409	378	413	414	431	444	434	447	398	391	350	370

TB Caused Death (%)	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
Yes	0.0%	1.5%	1.2%	2.5%	2.2%	1.5%	1.9%	2.1%	1.9%	2.4%	0.8%	1.8%
No	0.2%	5.4%	3.5%	5.1%	4.8%	6.2%	5.1%	4.6%	3.3%	2.9%	1.7%	1.0%
Unknown	99.8%	93.1%	95.4%	92.4%	93.1%	92.3%	92.9%	93.3%	94.8%	94.7%	97.5%	97.1%

TB Caused Death Incidence (per 10^6)	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
Yes	0.00	1.51	1.24	2.66	2.36	1.60	2.01	2.21	1.76	2.19	0.65	1.52

Statistics (N)	Min	Q1	Median	Q3	Max	Mean	S.Dev	Mean Change
Yes	0	5.75	7.5	10	11	7.17	3.27	0.64
No	1	10.5	18.5	22.25	30	16.25	9.08	0.27
Unknown	350	387.8	411	431.8	447	406.6	30.39	-3.55

Statistics (%)	Min	Q1	Median	Q3	Max	Mean	S.Dev	Mean Change
Yes	0.0%	1.4%	1.9%	2.1%	2.5%	1.6%	0.7%	0.2%
No	0.2%	2.6%	4.0%	5.1%	6.2%	3.7%	1.9%	0.1%
Unknown	92.3%	93.0%	94.0%	95.8%	99.8%	94.7%	2.4%	-0.2%

Statistics Incidence (per 10^6)	Min	Q1	Median	Q3	Max	Mean	S.Dev	Mean Change
Yes	0.00	1.44	1.68	2.19	2.66	1.64	0.75	0.14

Table A.4: Notifications (Count, Percentage, Incidence) Categorized By Death, Statistics (Count, Percentage, Incidence) Categorized By Death

N	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
Male	260	252	257	268	279	296	280	300	261	238	217	240
Female	149	153	176	180	184	185	187	176	159	175	141	139
Unknown	1	1	0	0	0	0	0	3	0	0	1	2
Total	410	406	433	448	463	481	467	479	420	413	359	381
%	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
Male	63.4%	62.1%	59.4%	59.8%	60.3%	61.5%	59.9%	62.6%	62.1%	57.6%	60.5%	63%
Female	36.3%	37.7%	40.7%	40.2%	39.7%	38.5%	40%	36.7%	37.9%	42.4%	39.3%	36.5%
Unknown	0.2%	0.3%	0.00%	0.00%	0.00%	0.00%	0.00%	0.63%	0.00%	0.00%	0.3%	0.5%
Incidence (per 10^5)	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
Male	13.36	12.75	12.77	13.00	13.18	13.51	12.51	13.29	11.54	10.48	9.56	10.56
Female	7.56	7.64	8.66	8.69	8.70	8.47	8.32	7.73	6.94	7.59	6.09	5.99

Table A.5: Notifications (Count, Percentage, Incidence) For The Variable Gender

(N) Average Change	N	%
Male	-1.82	-0.45%
Female	-0.91	-0.23%
Unknown	0.09	NA

(N) Overal Change	N	%
Male	-20	-7.69%
Female	-10	-6.71%
Unknown	1	100.00%

Statistics (%)	Male	Female	Unknown
Min	57.63%	36.34%	0.00%
Q1	59.92%	37.45%	0.00%
Median	60.99%	38.87%	0.00%
Q3	62.26%	40.08%	0.00%
Max	63.41%	42.37%	0.63%
Mean	61.02%	38.82%	0.16%
S.Dev	1.72%	1.87%	0.23%

(N) Largest Change	N	%
Male	-39	-13.00%
Female	-34	-19.43%
Unknown	-3	100%

Statistics For % Data	Male	Female
Min	9.56	5.99
Q1	11.29	7.40
Median	12.76	7.69
Q3	13.21	8.52
Max	13.51	8.70
Mean	12.21	7.70
S.Dev	1.34	0.95

Table A.6: Descriptive Statistics (Count, Percentage, Incidence) For The Variable Gender Notifications

Age Category (N)	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
0-4	7	14	2	14	12	28	8	8	7	6	2	8
5-9	7	6	2	6	5	5	3	3	4	5	2	2
10-14	5	6	7	8	6	8	4	7	10	9	6	0
15-19	10	19	20	21	22	17	19	14	24	13	21	11
20-24	34	51	41	51	41	42	41	37	46	30	26	37
25-29	52	48	74	56	70	80	85	64	66	47	42	42
30-34	42	37	47	38	53	49	52	63	39	62	37	47
35-39	41	30	28	38	52	48	46	39	32	47	42	48
40-44	30	22	32	31	28	40	23	43	37	27	25	32
45-49	27	22	29	25	22	26	31	36	24	39	39	31
50-54	27	26	13	29	23	20	31	25	26	25	18	24
55-59	18	18	23	17	20	18	20	27	14	19	23	26
60-64	17	15	22	25	26	21	20	27	12	16	20	19
65-69	30	22	23	19	12	14	23	23	22	24	16	21
70-74	21	22	22	23	24	18	18	21	20	13	13	8
75-79	18	23	17	16	15	18	18	21	15	13	12	13
80-84	15	12	16	19	21	17	12	15	7	10	8	8
>=85	9	13	15	12	11	12	13	6	15	8	6	4
Unknown	0	0	0	0	0	0	0	0	0	0	1	0

Age Category (%)	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
0-4	1.7%	3.4%	0.5%	3.1%	2.6%	5.8%	1.7%	1.7%	1.7%	1.5%	0.6%	2.1%
5-9	1.7%	1.5%	0.5%	1.3%	1.1%	1.0%	0.6%	0.6%	1.0%	1.2%	0.6%	0.5%
10-14	1.2%	1.5%	1.6%	1.8%	1.3%	1.7%	0.9%	1.5%	2.4%	2.2%	1.7%	0.0%
15-19	2.4%	4.7%	4.6%	4.7%	4.8%	3.5%	4.1%	2.9%	5.7%	3.1%	5.8%	2.9%
20-24	8.3%	12.6%	9.5%	11.4%	8.9%	8.7%	8.8%	7.7%	11.0%	7.3%	7.2%	9.7%
25-29	12.7%	11.8%	17.1%	12.5%	15.1%	16.6%	18.2%	13.4%	15.7%	11.4%	11.7%	11.0%
30-34	10.2%	9.1%	10.9%	8.5%	11.4%	10.2%	11.1%	13.2%	9.3%	15.0%	10.3%	12.3%
35-39	10.0%	7.4%	6.5%	8.5%	11.2%	10.0%	9.9%	8.1%	7.6%	11.4%	11.7%	12.6%
40-44	7.3%	5.4%	7.4%	6.9%	6.0%	8.3%	4.9%	9.0%	8.8%	6.5%	7.0%	8.4%
45-49	6.6%	5.4%	6.7%	5.6%	4.8%	5.4%	6.6%	7.5%	5.7%	9.4%	10.9%	8.1%
50-54	6.6%	6.4%	3.0%	6.5%	5.0%	4.2%	6.6%	5.2%	6.2%	6.1%	5.0%	6.3%
55-59	4.4%	4.4%	5.3%	3.8%	4.3%	3.7%	4.3%	5.6%	3.3%	4.6%	6.4%	6.8%
60-64	4.1%	3.7%	5.1%	5.6%	5.6%	4.4%	4.3%	5.6%	2.9%	3.9%	5.6%	5.0%
65-69	7.3%	5.4%	5.3%	4.2%	2.6%	2.9%	4.9%	4.8%	5.2%	5.8%	4.5%	5.5%
70-74	5.1%	5.4%	5.1%	5.1%	5.2%	3.7%	3.9%	4.4%	4.8%	3.1%	3.6%	2.1%
75-79	4.4%	5.7%	3.9%	3.6%	3.2%	3.7%	3.9%	4.4%	3.6%	3.1%	3.3%	3.4%
80-84	3.7%	3.0%	3.7%	4.2%	4.5%	3.5%	2.6%	3.1%	1.7%	2.4%	2.2%	2.1%
>=85	2.2%	3.2%	3.5%	2.7%	2.4%	2.5%	2.8%	1.3%	3.6%	1.9%	1.7%	1.0%
Unknown	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.3%	0.0%

Table A.7: Notifications (Count, Percentage) Categorized By Age

Incidence Age Category	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
0-4	2.52	4.92	0.69	4.72	3.97	9.02	2.47	2.39	2.02	1.69	0.55	2.19
5-9	2.65	2.23	0.73	2.14	1.73	1.68	0.98	0.97	1.27	1.56	0.62	0.60
10-14	1.75	2.13	2.52	2.90	2.19	2.89	1.41	2.41	3.38	2.99	1.96	0.00
15-19	3.19	6.24	6.70	7.17	7.56	5.83	6.43	4.76	8.29	4.63	7.62	4.01
20-24	10.36	15.27	12.20	15.02	11.80	11.30	10.97	10.44	14.10	10.05	9.36	14.30
25-29	16.63	15.12	22.74	16.22	18.66	20.01	20.82	15.98	17.30	12.95	12.31	13.13
30-34	13.78	11.71	14.46	11.32	15.23	13.56	14.07	16.82	10.23	15.76	9.43	12.13
35-39	14.09	10.12	9.23	12.18	16.23	14.32	13.20	10.95	8.90	12.94	11.53	13.19
40-44	11.03	7.88	11.19	10.54	9.34	12.98	7.33	13.58	11.50	8.20	7.45	9.33
45-49	10.82	8.67	11.19	9.42	8.04	9.25	10.75	12.18	7.98	12.82	12.66	9.93
50-54	11.70	11.06	5.46	11.93	9.33	7.90	12.02	9.53	9.72	9.13	6.47	8.48
55-59	9.12	8.71	10.75	7.68	8.93	7.90	8.60	11.46	5.84	7.81	9.27	10.35
60-64	11.02	9.36	13.25	14.53	14.46	11.01	10.01	13.00	5.61	7.37	9.09	8.50
65-69	22.47	16.30	16.73	13.49	8.50	9.72	15.35	14.69	13.46	13.95	8.82	11.09
70-74	18.73	19.33	19.01	19.59	20.43	15.06	14.79	16.85	15.64	9.99	9.80	5.79
75-79	20.04	25.70	18.89	17.58	16.41	19.46	19.05	21.69	15.14	12.82	11.58	12.29
80-84	25.47	19.45	25.36	29.69	32.61	25.91	18.21	22.42	10.23	14.33	11.27	11.02
>=85	21.58	31.03	34.56	26.26	23.01	24.29	25.05	11.11	26.74	13.75	9.92	6.44

Statistics (N) Age Catgory	Min	Q1	Median	Q3	Max	Mean	S.Dev	Mean Change
0-4	2	6.75	8.0	12.50	28	9.67	6.96	0.09
5-9	2	2.75	4.5	5.25	7	4.17	1.75	-0.45
10-14	0	5.75	6.5	8.00	10	6.33	2.61	-0.45
15-19	10	13.75	19.0	21.00	24	17.58	4.56	0.09
20-24	26	36.25	41.0	43.00	51	39.75	7.59	0.27
25-29	42	47.75	60.0	71.00	85	60.50	14.80	-0.91
30-34	37	38.75	47.0	52.25	63	47.17	9.12	0.45
35-39	28	36.50	41.5	47.25	52	40.92	7.77	0.64
40-44	22	26.50	30.5	33.25	43	30.83	6.53	0.18
45-49	22	24.75	28.0	32.25	39	29.25	6.09	0.36
50-54	13	22.25	25.0	26.25	31	23.92	4.93	-0.27
55-59	14	18.00	19.5	23.00	27	20.25	3.82	0.73
60-64	12	16.75	20.0	22.75	27	20.00	4.57	0.18
65-69	12	18.25	22.0	23.00	30	20.75	4.88	-0.82
70-74	8	16.75	20.5	22.00	24	18.58	4.87	-1.18
75-79	12	14.50	16.5	18.00	23	16.58	3.29	-0.45
80-84	7	9.50	13.5	16.25	21	13.33	4.56	-0.64
>= 85	4	7.50	11.5	13.00	15	10.33	3.68	-0.45
Unknown	0	0.00	0.0	0.00	1	0.08	0.29	0.00

Table A.8: Top: Notifications (Incidence) Categorized By Age. Bottom: Statistics Of Count Data (N) For Age Categorized

Statistics (%) Age Catgory	Min	Q1	Median	Q3	Max	Mean	S.Dev	Mean Change
0-4	0.5%	1.6%	1.7%	2.7%	5.8%	2.2%	1.4%	0.0%
5-9	0.5%	0.6%	1.0%	1.2%	1.7%	1.0%	0.4%	-0.1%
10-14	0.0%	1.3%	1.5%	1.7%	2.4%	1.5%	0.6%	-0.1%
15-19	2.4%	3.1%	4.3%	4.7%	5.8%	4.1%	1.1%	0.0%
20-24	7.2%	8.2%	8.8%	10.0%	12.6%	9.2%	1.7%	0.1%
25-29	11.0%	11.8%	13.0%	15.9%	18.2%	13.9%	2.5%	-0.2%
30-34	8.5%	10.0%	10.6%	11.7%	15.0%	11.0%	1.8%	0.2%
35-39	6.5%	8.0%	9.9%	11.3%	12.6%	9.6%	1.9%	0.2%
40-44	4.9%	6.4%	7.1%	8.3%	9.0%	7.2%	1.3%	0.1%
45-49	4.8%	5.5%	6.6%	7.7%	10.9%	6.9%	1.8%	0.1%
50-54	3.0%	5.0%	6.1%	6.4%	6.6%	5.6%	1.1%	0.0%
55-59	3.3%	4.2%	4.4%	5.4%	6.8%	4.8%	1.1%	0.2%
60-64	2.9%	4.1%	4.7%	5.6%	5.6%	4.6%	0.9%	0.1%
65-69	2.6%	4.4%	5.1%	5.4%	7.3%	4.9%	1.3%	-0.2%
70-74	2.1%	3.7%	4.6%	5.1%	5.4%	4.3%	1.0%	-0.3%
75-79	3.1%	3.4%	3.7%	4.0%	5.7%	3.9%	0.7%	-0.1%
80-84	1.7%	2.4%	3.0%	3.7%	4.5%	3.1%	0.9%	-0.1%
>= 85	1.0%	1.9%	2.4%	2.9%	3.6%	2.4%	0.8%	-0.1%
Unknown	0.0%	0.0%	0.0%	0.0%	0.3%	0.0%	0.1%	0.0%

Statistics (Incidence) Age Catgory	Min	Q1	Median	Q3	Max	Mean	S.Dev	Mean Change
0-4	0.55	1.93	2.43	4.16	9.02	3.09	2.32	-0.03
5-9	0.60	0.91	1.42	1.83	2.65	1.43	0.68	-0.19
10-14	0.00	1.91	2.30	2.89	3.38	2.21	0.90	-0.16
15-19	3.19	4.72	6.33	7.27	8.29	6.04	1.59	0.07
20-24	9.36	10.42	11.55	14.15	15.27	12.10	2.07	0.36
25-29	12.31	14.62	16.43	19.00	22.74	16.82	3.27	-0.32
30-34	9.43	11.61	13.67	14.65	16.82	13.21	2.27	-0.15
35-39	8.90	10.74	12.56	13.42	16.23	12.24	2.19	-0.08
40-44	7.33	8.12	9.94	11.27	13.58	10.03	2.11	-0.15
45-49	7.98	9.10	10.34	11.44	12.82	10.31	1.70	-0.08
50-54	5.46	8.33	9.43	11.22	12.02	9.39	2.10	-0.29
55-59	5.84	7.88	8.82	9.54	11.46	8.87	1.52	0.11
60-64	5.61	8.94	10.51	13.06	14.53	10.60	2.82	-0.23
65-69	8.50	10.74	13.72	15.59	22.47	13.71	3.94	-1.03
70-74	5.79	13.59	16.25	19.09	20.43	15.42	4.65	-1.18
75-79	11.58	14.56	18.24	19.61	25.70	17.55	4.16	-0.71
80-84	10.23	13.56	20.94	25.58	32.61	20.50	7.61	-1.31
>= 85	6.44	13.09	23.65	26.38	34.56	21.14	8.85	-1.38

Table A.9: Top: Statistics Of Percentage Data (%) Categorized By Age. Bottom: Statistics Of Incidence Data (Incidence) Categorized By Age

Emplyoment Status (N)	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
Employed	128	113	155	142	164	171	185	156	119	104	107	102
Unemployed	85	71	56	59	88	92	84	121	92	91	69	96
Housewife/ Husband	49	33	44	39	42	33	38	35	36	30	33	38
Retired	70	78	84	75	74	73	66	77	65	45	50	44
Student	21	40	47	54	39	44	37	41	64	36	43	26
Other	25	27	10	24	23	24	27	30	15	20	15	14
Child	0	1	0	0	0	0	0	0	0	0	0	0
Unknown	32	43	37	55	33	44	30	19	29	87	42	61

Emplyoment Status (%)	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
Employed	31.2%	27.8%	35.8%	31.7%	35.4%	35.6%	39.6%	32.6%	28.3%	25.2%	29.8%	26.8%
Unemployed	20.7%	17.5%	12.9%	13.2%	19.0%	19.1%	18.0%	25.3%	21.9%	22.0%	19.2%	25.2%
Housewife/ Husband	11.5%	8.1%	10.2%	8.7%	9.1%	6.9%	8.1%	7.3%	8.1%	7.3%	9.2%	10.0%
Retired	17.1%	19.2%	19.4%	16.7%	16.0%	15.2%	14.1%	16.1%	15.5%	10.9%	13.9%	11.5%
Student	5.1%	9.9%	10.9%	12.1%	8.4%	9.1%	7.9%	8.6%	15.2%	8.7%	12.0%	6.8%
Other	6.1%	6.7%	2.3%	5.4%	5.0%	5.0%	5.8%	6.3%	3.6%	4.8%	4.2%	3.7%
Child	0.0%	0.2%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Unknown	7.8%	10.6%	8.5%	12.3%	7.1%	9.1%	6.4%	4.0%	6.9%	21.1%	11.7%	16.0%

Emplyoment Status (Incidence)	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
Employed	4.36	3.80	5.10	4.51	5.01	5.07	5.48	5.18	4.11	3.63	3.73	3.41
Unemployed	46.67	36.30	27.05	30.42	45.63	43.85	29.59	17.67	11.62	10.97	8.32	13.65

Table A.10: Notifications (Count, Percentage, Incidence) Categorized By Employment Status

Statistics (N) Employment Status	Min	Q1	Median	Q3	Max	Mean	S.Dev	Mean Change
Employed	102	111.5	135	158	185	137.17	28.75	-2.36
Unemployed	56	70.5	86.5	92	121	83.67	17.86	1.00
Housewife/ Husband	30	33	36.5	39.75	47	37.17	5.13	-0.82
Retired	44	61.25	71.5	75.5	84	66.75	13.40	-2.36
Student	21	36.75	40.5	44.75	64	41.00	11.35	0.45
Other	10	15	23.5	25.5	30	21.17	6.28	-1.00
Housewife	0	0	0	0	2	0.33	0.78	-0.18
Child	0	0	0	0	1	0.08	0.29	0.00
Unknown	19	31.5	39.5	46.75	87	42.67	18.10	2.64

Statistics (%) Employment Status	Min	Q1	Median	Q3	Max	Mean	S.Dev	Mean Change
Employed	25.2%	28.2%	31.5%	35.5%	39.6%	31.6%	4.3%	-0.4%
Unemployed	12.9%	17.9%	19.2%	21.9%	25.3%	19.5%	3.9%	0.4%
Housewife/ Husband	6.9%	7.9%	8.4%	9.4%	11.5%	8.7%	1.4%	-0.1%
Retired	10.9%	14.1%	15.7%	16.8%	19.4%	15.5%	2.6%	-0.5%
Student	5.1%	8.3%	8.9%	11.1%	15.2%	9.6%	2.7%	0.2%
Other	2.3%	4.1%	5.0%	5.9%	6.7%	4.9%	1.3%	-0.2%
Child	0.0%	0.0%	0.0%	0.0%	0.2%	0.0%	0.1%	0.0%
Unknown	4.0%	7.1%	8.8%	11.8%	21.1%	10.1%	4.7%	0.7%

Statistics	Min	Q1	Median	Q3	Max	Mean	S.Dev	Mean Change
Employed	3.41	3.78	4.43	5.07	5.48	4.45	0.71	-0.09
Unemployed	8.32	13.14	28.32	38.19	46.67	26.81	14.24	-3.00

Table A.11: Descriptive Statistics (Count, Percentage, Incidence) For The Variable Employment Status

Current Living (N)	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
Home	342	359	364	380	396	431	410	429	382	302	294	313
Hostel	17	8	11	18	17	9	19	17	6	6	4	7
B&B/Hotel	6	0	2	1	3	1	2	1	1	2	2	0
Homeless	6	2	5	0	7	5	3	2	1	1	1	3
Prison	2	0	2	3	3	1	1	2	0	12	3	1
Institution	14	3	11	12	9	11	12	10	5	9	4	4
Other	6	8	13	12	14	11	8	6	10	8	8	2
Unknown	17	26	25	22	14	12	12	12	15	73	43	51

Current Living (%)	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
Home	83.4%	88.4%	84.1%	84.8%	85.5%	89.6%	87.8%	89.6%	91.0%	73.1%	81.9%	82.2%
Hostel	4.1%	2.0%	2.5%	4.0%	3.7%	1.9%	4.1%	3.5%	1.4%	1.5%	1.1%	1.8%
B&B/Hotel	1.5%	0.0%	0.5%	0.2%	0.6%	0.2%	0.4%	0.2%	0.2%	0.5%	0.6%	0.0%
Homeless	1.5%	0.5%	1.2%	0.0%	1.5%	1.0%	0.6%	0.4%	0.2%	0.2%	0.3%	0.8%
Prison	0.5%	0.0%	0.5%	0.7%	0.6%	0.2%	0.2%	0.4%	0.0%	2.9%	0.8%	0.3%
Institution	3.4%	0.7%	2.5%	2.7%	1.9%	2.3%	2.6%	2.1%	1.2%	2.2%	1.1%	1.0%
Other	1.5%	2.0%	3.0%	2.7%	3.0%	2.3%	1.7%	1.3%	2.4%	1.9%	2.2%	0.5%
Unknown	4.1%	6.4%	5.8%	4.9%	3.0%	2.5%	2.6%	2.5%	3.6%	17.7%	12.0%	13.4%

Current Living Incidence	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
Prison	20.58	0.00	22.68	34.54	30.93	10.30	9.15	16.21	0.00	86.01	21.65	7.66
Homeless	Na	Na	Na	Na	Na	Na	Na	Na	Na	26.26	Na	Na

Table A.12: Notifications (Count, Percentage, Incidence) Categorized By Current Living

Statistics (N)	Min	Q1	Median	Q3	Max	Mean	S.Dev	Mean Change
Home	294	334.75	372	399.5	431	366.83	46.90	-2.64
Hostel	4	6.75	10	17	19	11.58	5.60	-0.91
B&B/Hotel	0	1	1.5	2	6	1.75	1.60	-0.55
Homeless	0	1	2.5	5	7	3.00	2.26	-0.27
Prison	0	1	2	3	12	2.50	3.18	-0.09
Institution	3	4.75	9.5	11.25	14	8.67	3.73	-0.91
Other	2	7.5	8	11.25	14	8.83	3.38	-0.36
Unknown	12	13.5	19.5	30.25	73	26.83	19.23	3.09

Statistics (%)	Min	Q1	Median	Q3	Max	Mean	S.Dev	Mean Change
Home	73.1%	83.1%	85.2%	88.7%	91.0%	85.1%	4.9%	-0.1%
Hostel	1.1%	1.7%	2.3%	3.8%	4.1%	2.6%	1.2%	-0.2%
B&B/Hotel	0.0%	0.2%	0.3%	0.5%	1.5%	0.4%	0.4%	-0.1%
Homeless	0.0%	0.3%	0.6%	1.1%	1.5%	0.7%	0.5%	-0.1%
Prison	0.0%	0.2%	0.4%	0.7%	2.9%	0.6%	0.8%	0.0%
Institution	0.7%	1.2%	2.1%	2.5%	3.4%	2.0%	0.8%	-0.2%
Other	0.5%	1.7%	2.1%	2.5%	3.0%	2.0%	0.7%	-0.1%
Unknown	2.5%	2.9%	4.5%	7.8%	17.7%	6.5%	5.0%	0.8%

Statistics Incidence	Min	Q1	Median	Q3	Max	Mean	S.Dev	Mean Change
Homeless	0	8.78	18.40	24.74	86.01	21.64	23.05	-1.17
Prison	26.26	26.26	26.26	26.26	26.26	26.26	Na	Na

Table A.13: Statistics (Count, Percentage, Incidence) Categorized By Current Living

Race/ Ethnic (N)	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
White	291	312	301	325	329	320	296	300	266	207	196	198
South Asian	33	25	37	43	38	65	74	69	69	45	41	48
Black	47	25	47	41	55	53	48	72	47	42	41	32
Irish Traveller	1	1	1	1	0	1	1	1	2	4	2	12
East/South East Asian	7	10	11	22	13	23	30	18	19	30	21	27
Other	4	4	7	5	11	8	7	9	5	6	18	1
Unknown	27	29	29	11	17	11	11	10	12	79	40	63

Race/ Ethnicity (%)	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
White	71.0%	76.8%	69.5%	72.5%	71.1%	66.5%	63.4%	62.6%	63.3%	50.1%	54.6%	52.0%
South Asian	8.0%	6.2%	8.5%	9.6%	8.2%	13.5%	15.8%	14.4%	16.4%	10.9%	11.4%	12.6%
Black	11.5%	6.2%	10.9%	9.2%	11.9%	11.0%	10.3%	15.0%	11.2%	10.2%	11.4%	8.4%
Irish Traveller	0.2%	0.2%	0.2%	0.2%	0.0%	0.2%	0.2%	0.2%	0.5%	1.0%	0.6%	3.1%
East/South East Asian	1.7%	2.5%	2.5%	4.9%	2.8%	4.8%	6.4%	3.8%	4.5%	7.3%	5.8%	7.1%
Other	1.0%	1.0%	1.6%	1.1%	2.4%	1.7%	1.5%	1.9%	1.2%	1.5%	5.0%	0.3%
Unknown	6.6%	7.1%	6.7%	2.5%	3.7%	2.3%	2.4%	2.1%	2.9%	19.1%	11.1%	16.5%

Race/ Ethnicity Incidence	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
White	7.88	8.31	7.89	8.39	8.36	8.01	7.30	7.29	6.37	4.89	4.56	4.55
Black	169.6	78.46	130.5	102.1	124.1	109.3	91.22	126.8	77.14	64.54	59.22	43.61
Asian	151.1	106.3	121.8	141.7	97.43	149.6	159.3	121.3	112.5	88.56	68.01	76.82
Irish Traveller	6.00	5.53	5.12	4.77	0.00	4.20	3.97	3.75	7.13	13.56	6.47	37.10
Other	7.84	8.02	14.37	10.51	23.69	17.66	15.85	20.93	11.94	14.73	45.48	2.60
Unknown	36.54	39.46	39.67	15.13	23.51	15.30	15.38	14.06	16.97	112.3	57.20	90.61

Table A.14: Notifications (Count, Percentage, Incidence) For The Variable Race/Ethnicity

Statistics (N)	Min	Q1	Median	Q3	Max	Mean	S.Dev	Mean Change
White	196	251.3	298	314	329	278.4	50.04	-8.45
South Asian	25	37.75	44	66	74	48.92	16.22	1.36
Black	25	41	47	49.25	72	45.83	11.74	-1.36
Irish Traveller	0	1	1	2	12	2.25	3.22	1.00
East/South East Asian	7	12.5	20	24	30	19.25	7.75	1.82
Other	1	4.75	6.5	8.25	18	7.08	4.32	-0.27
Unknown	10	11	22	31.75	79	28.25	22.44	3.27

Statistics (%)	Min	Q1	Median	Q3	Max	Mean	S.Dev	Mean Change
White	50.1%	60.6%	65.0%	71.0%	76.8%	64.5%	8.5%	-1.7%
South Asian	6.2%	8.5%	11.2%	13.7%	16.4%	11.3%	3.3%	0.4%
Black	6.2%	9.9%	10.9%	11.4%	15.0%	10.6%	2.1%	-0.3%
Irish Traveller	0.0%	0.2%	0.2%	0.5%	3.1%	0.6%	0.9%	0.3%
East/South East Asian	1.7%	2.7%	4.7%	6.0%	7.3%	4.5%	1.9%	0.5%
Other	0.3%	1.1%	1.5%	1.7%	5.0%	1.7%	1.2%	-0.1%
Unknown	2.1%	2.4%	5.1%	8.1%	19.1%	6.9%	5.8%	0.9%

Statistics Incidence	Min	Q1	Median	Q3	Max	Mean	S.Dev	Mean Change
White	4.55	6.00	7.59	8.09	8.39	6.98	1.51	-0.30
Black	43.61	73.99	96.65	124.8	169.6	98.05	36.04	-11.46
Asian	68.01	95.21	116.9	143.7	159.3	116.2	30.16	-6.75
Irish Traveller	0.00	4.14	5.33	6.63	37.10	8.13	9.64	2.83
Other	2.60	9.89	14.55	18.48	45.48	16.14	10.94	-0.48
Unknown	14.06	15.36	30.03	44.06	112.3	39.68	32.21	4.91

Table A.15: Statistics (Count, Percentage, Incidence) For The Variable Race/Ethnicity

Refugee/ Asylum Seeker	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
Yes	46	28	32	36	37	29	30	41	19	30	14	9
No	346	344	373	394	404	445	426	431	372	288	283	271
Unknown	18	34	28	18	22	7	11	7	29	95	62	101

Refugee/ Asylum Seeker	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
Yes	11.2%	6.9%	7.4%	8.0%	8.0%	6.0%	6.4%	8.6%	4.5%	7.3%	3.9%	2.4%
No	84.4%	84.7%	86.1%	87.9%	87.3%	92.5%	91.2%	90.0%	88.6%	69.7%	78.8%	71.1%
Unknown	4.4%	8.4%	6.5%	4.0%	4.8%	1.5%	2.4%	1.5%	6.9%	23.0%	17.3%	26.5%

Refugee/ Asylum Seeker	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
Yes	855.0	468.9	489.1	506.1	467.3	310.7	308.3	428.4	208.6	363.7	221.3	150.0

Statistics (N)	Min	Q1	Median	Q3	Max	Mean	S.Dev	Mean Change
Yes	9	25.75	30	36.25	46	29.25	10.80	-3.36
No	271	330	372.5	409.5	445	364.8	59.73	-6.82
Unknown	7	16.25	25	41	101	36	32.54	7.55

Statistics (%)	Min	Q1	Median	Q3	Max	Mean	S.Dev	Mean Change
Yes	0.02	0.06	0.07	0.08	0.11	0.07	0.02	-0.01
No	0.70	0.83	0.87	0.9	0.9	0.84	0.07	-0.01
Unknown	0.01	0.04	0.1	0.1	0.3	0.1	0.09	0.02

Statistics Incidence	Min	Q1	Median	Q3	Max	Mean	S.Dev	Mean Change
Yes	150.0	286.6	396.1	474.0	855.0	398.1	187.1	-64.09

Table A.16: Notifications (Count, Percentage, Incidence) And Statistics For The Variable Refugee Status

Risk Factor	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
Tobacco	1	2	2	2	5	1	0	4	7	3	0	0
Substance Abuse	52	39	39	39	53	51	46	40	35	52	45	40
Immuno Suppressive Medication	1	8	5	8	7	3	8	6	6	16	14	11
Immuno Suppressive Illness	23	10	19	16	22	19	23	22	28	31	25	20
High Endemicity Affiliation	105	74	108	114	131	157	181	184	161	251	164	242
Diabetes	2	6	6	8	6	5	2	2	6	6	8	9
Contact	19	28	26	22	24	26	37	25	29	77	64	53
Anti-TNF	0	0	1	1	1	1	3	2	4	6	3	6
Other	17	36	27	23	32	29	36	47	60	37	0	0
Unspecified	10	4	6	3	9	7	9	16	5	20	0	0

Risk Factor	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
Tobacco	0.4%	1.0%	0.8%	0.8%	1.7%	0.3%	0.0%	1.1%	2.1%	0.6%	0.0%	0.0%
Substance Abuse	22.6%	18.8%	16.3%	16.5%	18.3%	17.1%	13.3%	11.5%	10.3%	10.4%	13.9%	10.5%
Immuno Suppressive Medication	0.4%	3.9%	2.1%	3.4%	2.4%	1.0%	2.3%	1.7%	1.8%	3.2%	4.3%	2.9%
Immuno Suppressive Illness	10.0%	4.8%	7.9%	6.8%	7.6%	6.4%	6.7%	6.3%	8.2%	6.2%	7.7%	5.2%
High Endemicity Affiliation	45.7%	35.7%	45.2%	48.3%	45.2%	52.5%	52.5%	52.9%	47.2%	50.3%	50.8%	63.5%
Diabetes	0.9%	2.9%	2.5%	3.4%	2.1%	1.7%	0.6%	0.6%	1.8%	1.2%	2.5%	2.4%
Contact	8.3%	13.5%	10.9%	9.3%	8.3%	8.7%	10.7%	7.2%	8.5%	15.4%	19.8%	13.9%
Anti-TNF	0.0%	0.0%	0.4%	0.4%	0.3%	0.3%	0.9%	0.6%	1.2%	1.2%	0.9%	1.6%
Other	7.4%	17.4%	11.3%	9.7%	11.0%	9.7%	10.4%	13.5%	17.6%	7.4%	0.0%	0.0%
Unspecified	4.3%	1.9%	2.5%	1.3%	3.1%	2.3%	2.6%	4.6%	1.5%	4.0%	0.0%	0.0%

Table A.17: Frequency Of Risk Factor Notifications (Count, Percentage)

Statistics (N)	Min	Q1	Median	Q3	Max	Mean	S.Dev	Mean Change
Tobacco	0	0.75	2	3.25	7	2.25	2.18	-0.09
Substance Abuse	35	39	42.5	51.25	53	44.25	6.40	-1.09
Immuno Suppressive Medication	1	5.75	7.5	8.75	16	7.75	4.27	0.91
Immuno Suppressive Illness	10	19	22	23.5	31	21.50	5.45	-0.27
High Endemicity Affiliation	74	112.5	159	181.8	251	156.00	53.88	12.45
Diabetes	2	4.25	6	6.5	9	5.50	2.39	0.64
Contact	19	24.75	27	41	77	35.83	18.63	3.09
Anti-TNF	0	1	1.5	3.25	6	2.33	2.10	0.55
Other	0	21.5	30.5	36.25	60	28.67	17.37	-1.55
Unspecified	0	3.75	6.5	9.25	20	7.42	5.98	-0.91

Statistics (%)	Min	Q1	Median	Q3	Max	Mean	S.Dev	Mean Change
Tobacco	0.0%	0.3%	0.7%	1.0%	2.1%	0.7%	0.7%	0.0%
Substance Abuse	10.3%	11.2%	15.1%	17.4%	22.6%	15.0%	3.9%	-1.1%
Immuno Suppressive Medication	0.4%	1.8%	2.4%	3.3%	4.3%	2.5%	1.1%	0.2%
Immuno Suppressive Illness	4.8%	6.3%	6.7%	7.8%	10.0%	7.0%	1.4%	-0.4%
High Endemicity Affiliation	35.7%	45.5%	49.3%	52.5%	63.5%	49.1%	6.6%	1.6%
Diabetes	0.6%	1.1%	1.9%	2.5%	3.4%	1.9%	0.9%	0.1%
Contact	7.2%	8.4%	10.0%	13.6%	19.8%	11.2%	3.8%	0.5%
Anti-TNF	0.0%	0.3%	0.5%	1.0%	1.6%	0.7%	0.5%	0.1%
Other	0.0%	7.4%	10.1%	11.8%	17.6%	9.6%	5.6%	-0.7%
Unspecified	0.0%	1.4%	2.4%	3.3%	4.6%	2.3%	1.5%	-0.4%

Table A.18: Statistics Of Risk Factor Notifications (Count, Percentage)

Count	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
Jan	34	28	32	32	27	32	43	40	30	25	27	35
Feb	31	30	34	38	45	25	45	29	39	54	34	29
Mar	35	30	39	33	40	46	42	45	34	43	30	28
Apr	35	37	46	39	44	70	53	48	40	27	34	45
May	24	36	43	47	45	44	39	25	29	30	42	30
Jun	37	38	36	43	46	45	32	60	54	49	26	37
Jul	47	40	29	31	41	48	43	56	36	37	27	46
Aug	30	27	36	36	38	42	39	31	42	27	29	41
Sept	32	37	31	35	28	30	33	35	20	27	22	21
Oct	35	40	36	40	43	37	38	36	28	32	33	33
Nov	32	37	33	31	33	26	36	37	42	35	26	19
Dec	38	26	36	43	31	31	24	37	26	27	29	17

Annual Statistics	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	Mean
Min	24	26	29	31	27	25	24	25	20	25	22	17	24.58
Q1	31.75	29.50	32.75	32.75	32.50	30.75	35.25	34.00	28.75	27.00	26.75	26.25	30.67
Median	34.50	36.50	36.00	37.00	40.50	39.50	39.00	37.00	35.00	31.00	29.00	31.50	35.54
Q3	35.50	37.25	36.75	40.75	44.25	45.25	43.00	45.75	40.50	38.50	33.25	38.00	39.90
Max	47	40	46	47	46	70	53	60	54	54	42	46	50.42
Range	23	14	17	16	19	45	29	35	34	29	20	29	25.83
IQ Range	3.75	7.75	4.00	8.00	11.75	14.50	7.75	11.75	11.75	11.50	6.50	11.75	9.23
Mean	34.17	33.83	35.92	37.33	38.42	39.67	38.92	39.92	35.00	34.42	29.92	31.75	35.77
Std.Dev	5.47	5.22	4.87	5.25	6.93	12.50	7.34	10.58	9.14	9.59	5.23	9.65	7.65
Skewness	0.68	-0.36	0.82	0.40	-0.62	1.19	-0.19	0.68	0.42	1.08	0.96	-0.05	0.42
Kurtosis	2.73	-1.69	0.49	-0.86	-1.23	2.12	1.09	-0.19	0.44	0.04	1.55	-0.99	0.29

Monthly Statistics	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sept	Oct	Nov	Dec	Mean
Min	25	25	28	27	24	26	27	27	20	28	19	17	24.42
Q1	27.75	29.75	32.25	36.5	29.75	36.75	34.75	29.75	25.75	33	29.75	26	30.98
Median	32	34	37	42	37.5	40.5	40.5	36	30.5	36	33	30	35.75
Q3	34.25	40.5	42.25	46.5	43.25	46.75	46.25	39.5	33.5	38.5	36.25	36.25	40.31
Max	43	54	46	70	47	60	56	42	37	43	42	43	48.58
Range	18	29	18	43	23	34	29	15	17	15	23	26	24.17
IQ Range	6.5	10.75	10	10	13.5	10	11.5	9.75	7.75	5.5	6.5	10.25	9.33
Mean	32.08	36.08	37.08	43.17	36.17	41.92	40.08	34.83	29.25	35.92	32.25	30.42	35.77
Std.Dev	5.37	8.46	6.23	10.99	8.23	9.54	8.59	5.77	5.75	4.10	6.17	7.17	7.20
Skewness	0.80	0.86	0.00	1.19	-0.21	0.33	0.11	-0.14	-0.47	-0.17	-0.74	0.00	0.13
Kurtosis	0.21	0.21	-1.48	2.55	-1.63	-0.08	-0.46	-1.67	-1.04	0.09	0.78	-0.08	-0.22

All Data	Min	Q1	Median	Q3	Max	Range	IQR	Mean	S.Dev	Skew	Kurt
Statistic	17	30	35	41	70	53	11	35.77	8.3	0.73	1.53

Table A.19: Descriptive Statistics Of Notifications For Each Month Of Year

Count	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
Quarter 1	100	88	105	103	112	103	130	114	103	122	91	92
Quarter 2	96	111	125	129	135	159	124	133	123	106	102	112
Quarter 3	109	104	96	102	107	120	115	122	98	91	78	108
Quarter 4	105	103	105	114	107	94	98	110	96	94	88	69
Statistics	Min	Q1	Q2	Q3	Max	Range	IQ Range	Mean	S.Dev	Skew	Kurt	
Quarter 1	88	98	103	112.5	130	42	14.5	105.25	12.56	0.57	-0.08	
Quarter 2	96	109.75	123.5	130	159	63	20.25	121.25	17.24	0.64	0.80	
Quarter 3	78	97.5	105.5	110.5	122	44	13	104.17	12.43	-0.56	0.48	
Quarter 4	69	94	100.5	105.5	114	45	11.5	98.58	11.97	-1.33	2.57	

Table A.20: Notifications (Count, Percentage, Incidence) Of Cases Categorized By Quarter Of Year

Trend	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
Quarter 1	106.17	106.42	106.93	110.86	113.47	112.59	112.47	114.60	110.01	99.97	94.29	97.71
Quarter 2	106.18	106.49	107.88	111.67	113.57	112.24	113.12	114.01	107.37	98.00	95.08	98.80
Quarter 3	106.19	106.56	108.83	112.48	113.68	111.89	113.77	113.41	104.72	96.04	95.87	99.90
Quarter 4	106.31	106.75	109.85	112.97	113.13	112.18	114.19	111.71	102.35	95.17	96.79	101.19
Seasonality	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
Quarter 1	-5.59	-5.59	-5.59	-5.59	-5.59	-5.59	-5.59	-5.59	-5.59	-5.59	-5.59	-5.59
Quarter 2	13.59	13.59	13.59	13.59	13.59	13.59	13.59	13.59	13.59	13.59	13.59	13.59
Quarter 3	-2.68	-2.68	-2.68	-2.68	-2.68	-2.68	-2.68	-2.68	-2.68	-2.68	-2.68	-2.68
Quarter 4	-5.31	-5.31	-5.31	-5.31	-5.31	-5.31	-5.31	-5.31	-5.31	-5.31	-5.31	-5.31
Irrefularity	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
Quarter 1	-0.58	-12.83	3.66	-2.27	4.12	-4.00	23.12	4.99	-1.42	27.62	2.30	-0.12
Quarter 2	-23.77	-9.08	3.53	3.74	7.84	33.18	-2.71	5.41	2.05	-5.59	-6.67	-0.39
Quarter 3	5.49	0.12	-10.15	-7.80	-3.99	10.79	3.91	11.27	-4.04	-2.36	-15.18	10.79
Quarter 4	4.01	1.57	0.47	6.34	-0.82	-12.87	-10.88	3.60	-1.03	4.15	-3.47	-26.87

Table A.21: The three additive components of quarterly notifications obtained from a robust STL decomposition with flexible trend and fixed seasonality.

Trend	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
Jan	31.95	33.98	36.26	36.15	39.08	39.87	39.15	38.05	37.43	35.23	31.64	31.78
Feb	32.22	33.85	36.04	36.33	39.24	40.25	39.02	38.41	36.93	35.01	31.15	32.39
Mar	32.50	33.72	35.82	36.50	39.39	40.62	38.88	38.76	36.43	34.80	30.66	33.01
Apr	32.74	33.82	35.63	36.58	39.28	40.67	38.93	39.00	36.05	34.67	30.31	32.69
May	32.99	33.93	35.44	36.66	39.17	40.71	38.97	39.24	35.67	34.53	29.95	32.37
Jun	33.16	34.14	35.54	36.88	38.88	40.77	38.71	39.43	35.44	34.18	29.91	31.85
Jul	33.34	34.36	35.64	37.09	38.58	40.83	38.44	39.62	35.21	33.83	29.86	31.32
Aug	33.50	34.77	35.66	37.41	38.37	40.84	37.97	39.54	35.42	33.40	29.99	30.80
Sept	33.66	35.18	35.67	37.72	38.16	40.86	37.49	39.46	35.63	32.97	30.12	30.28
Oct	33.88	35.65	35.72	38.09	38.49	40.46	37.36	38.98	35.64	32.71	30.38	29.71
Nov	34.11	36.13	35.77	38.47	38.81	40.05	37.23	38.51	35.66	32.45	30.64	29.13
Dec	34.05	36.19	35.96	38.77	39.34	39.60	37.64	37.97	35.44	32.05	31.21	28.48

Seasonality	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
Jan	-3.82	-3.82	-3.82	-3.82	-3.82	-3.82	-3.82	-3.82	-3.82	-3.82	-3.82	-3.82
Feb	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16
Mar	1.14	1.14	1.14	1.14	1.14	1.14	1.14	1.14	1.14	1.14	1.14	1.14
Apr	7.28	7.28	7.28	7.28	7.28	7.28	7.28	7.28	7.28	7.28	7.28	7.28
May	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33
Jun	6.14	6.14	6.14	6.14	6.14	6.14	6.14	6.14	6.14	6.14	6.14	6.14
Jul	4.36	4.36	4.36	4.36	4.36	4.36	4.36	4.36	4.36	4.36	4.36	4.36
Aug	-0.86	-0.86	-0.86	-0.86	-0.86	-0.86	-0.86	-0.86	-0.86	-0.86	-0.86	-0.86
Sept	-6.41	-6.41	-6.41	-6.41	-6.41	-6.41	-6.41	-6.41	-6.41	-6.41	-6.41	-6.41
Oct	0.27	0.27	0.27	0.27	0.27	0.27	0.27	0.27	0.27	0.27	0.27	0.27
Nov	-3.39	-3.39	-3.39	-3.39	-3.39	-3.39	-3.39	-3.39	-3.39	-3.39	-3.39	-3.39
Dec	-5.20	-5.20	-5.20	-5.20	-5.20	-5.20	-5.20	-5.20	-5.20	-5.20	-5.20	-5.20

Irregularity	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
Jan	5.88	-2.16	-0.44	-0.33	-8.26	-4.05	7.67	5.77	-3.60	-6.40	-0.82	7.05
Feb	-1.38	-4.01	-2.19	1.52	5.61	-15.40	5.83	-9.56	1.92	18.83	2.69	-3.55
Mar	1.37	-4.86	2.05	-4.64	-0.53	4.24	1.98	5.10	-3.57	7.06	-1.80	-6.15
Apr	-5.02	-4.10	3.10	-4.86	-2.56	22.06	6.80	1.72	-3.33	-14.94	-3.58	5.03
May	-9.33	1.74	7.23	10.00	5.50	2.95	-0.31	-14.57	-7.00	-4.87	11.71	-2.70
Jun	-2.30	-2.28	-5.68	-0.02	0.98	-1.91	-12.84	14.43	12.42	8.68	-10.05	-0.99
Jul	9.30	1.28	-11.00	-10.45	-1.94	2.81	0.20	12.02	-3.57	-1.19	-7.22	10.32
Aug	-2.64	-6.91	1.20	-0.55	0.48	2.01	1.89	-7.68	7.44	-5.54	-0.13	11.05
Sept	4.75	8.23	1.74	3.69	-3.75	-4.45	1.92	1.95	-9.22	0.44	-1.71	-2.87
Oct	0.85	4.08	0.01	1.64	4.25	-3.72	0.37	-3.25	-7.91	-0.98	2.35	3.02
Nov	1.28	4.26	0.62	-4.08	-2.42	-10.67	2.16	1.88	9.73	5.93	-1.25	-6.75
Dec	9.15	-5.00	5.23	9.42	-3.15	-3.41	-8.44	4.23	-4.25	0.15	2.99	-6.28

Table A.22: Monthly notification additive components obtained from a robust STL decomposition with flexible trend and fixed seasonality.

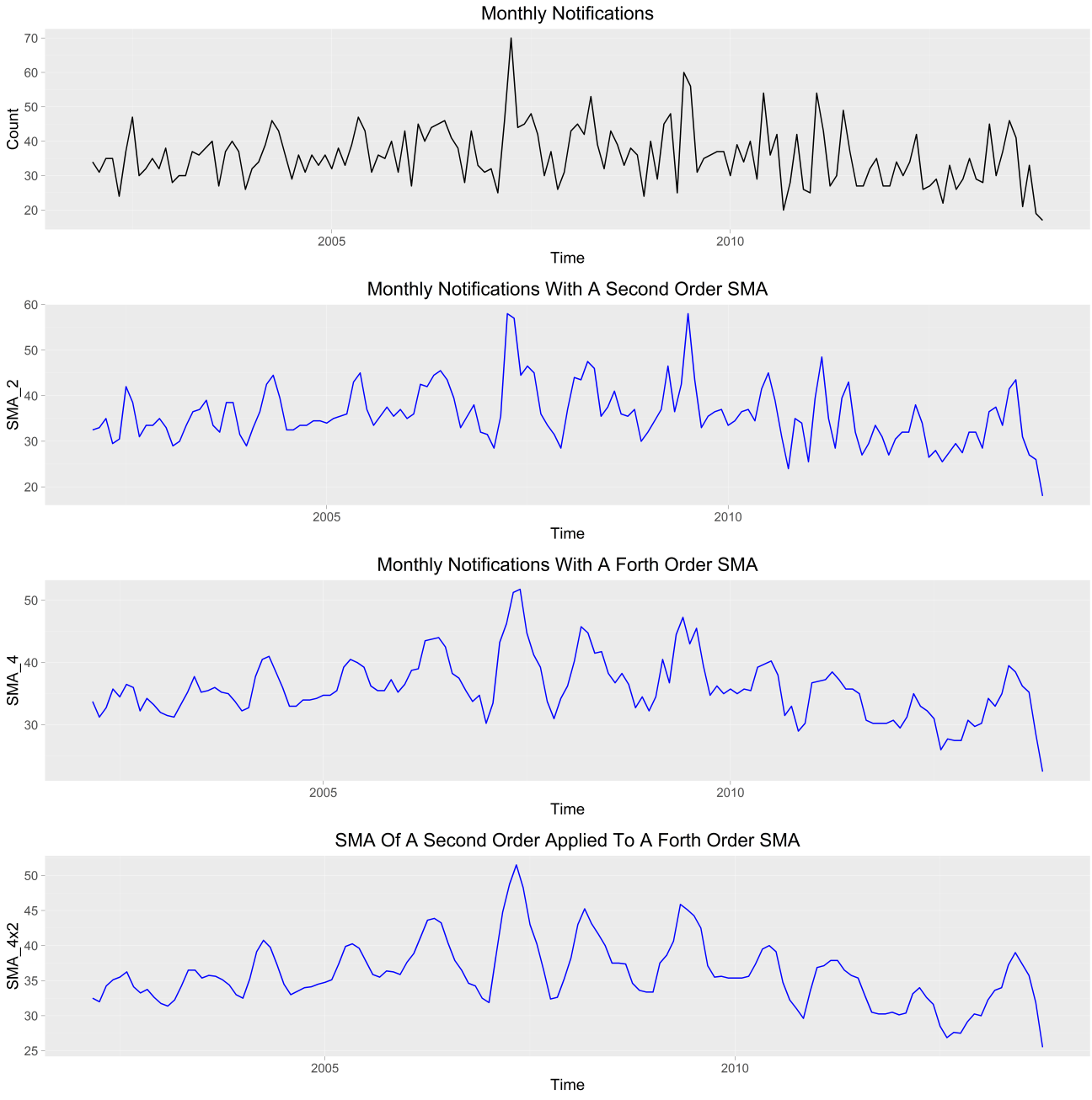

Figure A.1: Monthly Notification Data With A Second, Fourth and Second on Fourth Order Moving Average

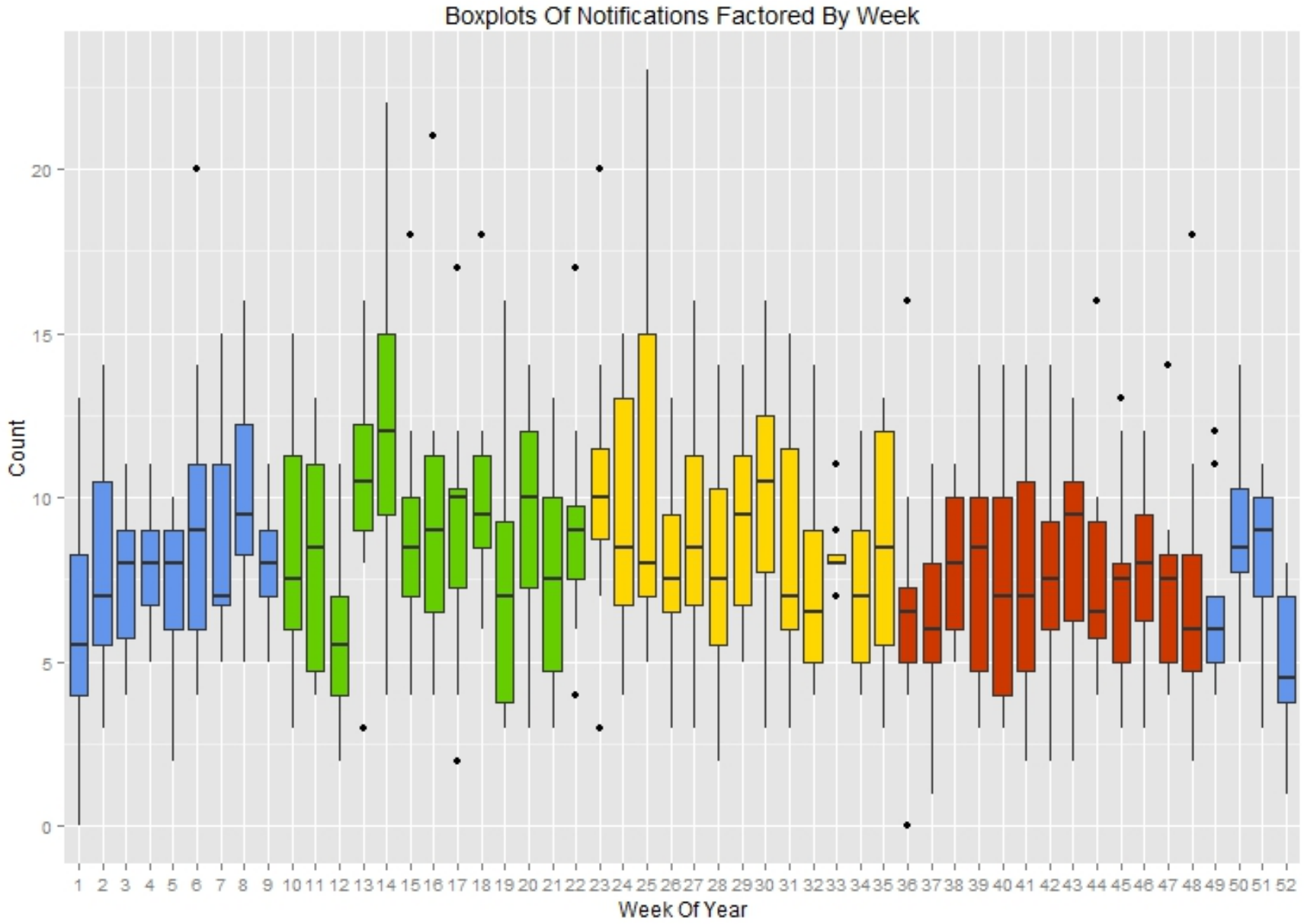

Figure A.2: Boxplots Of Weekly Notifications Factored By Week Of Year

Appendix B

National TB Notifications form

National Tuberculosis Notification Form v2.5

hpsc

A. PATIENT DETAILS

CIDR EVENT ID | **HSE ID**

HSE area | **County** | **CCA** | **DED name/code**

Patient forename | **Patient surname**

Patient address

Hospital name

Hospital number

Treating Physician

Phone

First notified by: ☐ Laboratory ☐ Occupational Health ☐ GP ☐ Public Health ☐ Hospital clinician ☐ Other

School/ college address

If other notification source, please specify:

Work address

B. SOCIODEMOGRAPHIC DETAILS

Sex: ☐ Male ☐ Female

Date of Birth

Age (years)

Current employment status

☐ Paid employment ☐ Retired
☐ Housewife/husband ☐ Student
☐ Unemployed ☐ Other

Other (please specify):

Current/most recent occupation (within last 2 years)

Current living status

☐ Home (private/rented) ☐ Hostel
☐ B&B/hotel ☐ Prison
☐ Homeless ☐ Institution
☐ Other (please specify):

Country of birth

☐ Ireland ☐ Other (please specify):

If born outside Ireland, year of entry into Ireland:

Race or ethnic group

☐ Black ☐ South Asian descent
☐ White ☐ East/south east Asian descent
☐ Irish Traveller ☐ Other (please specify):

Refugee / asylum seeker ☐ Yes ☐ No ☐ Unk

C. CLINICAL DETAILS

Symptomatic ☐ Yes ☐ No

Date of onset of symptoms

Date diagnosed

Date of notification

Date treatment commenced

Date contact tracing commenced

Diagnosis (tick one only)

☐ Pulmonary ☐ Extrapulmonary
☐ Pulmonary & Extrapulmonary (P+E)

If Extrapulmonary or P+E, please specify site(s):

EP site 1

EP site 2

Chest x-ray

☐ Active Cavitary TB ☐ Pleural ☐ Normal
☐ Active Non-cavitary TB ☐ Inactive/Old TB ☐ Not done
☐ Other

If other X-ray result, please specify:

CT thorax

☐ Abnormal with cavitation ☐ Normal ☐ Other CT result
☐ Abnormal without cavitation ☐ Not done ☐ Unknown

If other CT result, please specify:

Was this case hospitalised due to TB? ☐ Yes ☐ No ☐ Unk

This case was found by

☐ Presenting as case ☐ Post-mortem diagnosis
☐ Contact tracing ☐ Pre-employment screening
☐ Immigrant screening ☐ Other (please specify):

Did this case previously undergo TB screening in Ireland? ☐ Yes ☐ No ☐ Unk

If yes, please specify:

Previous history of TB (specify below)	☐ Yes	☐ No	☐ Unk
(a) Previous year of diagnosis			
(b) Previous treatment (>1 month)	☐ Yes	☐ No	☐ Unk
(c.) Previous treatment completed	☐ Yes	☐ No	☐ Unk
History of BCG vaccination	☐ Yes	☐ No	☐ Unk
If yes, year of BCG vaccination			
BCG scar present	☐ Yes	☐ No	☐ Unk
Risk factors present (specify below)	☐ Yes	☐ No	☐ Unk
Anti-TNF treatment	☐ Yes	☐ No	☐ Unk
Other immunosuppressive medication	☐ Yes	☐ No	☐ Unk
Immunosuppressive illness	☐ Yes	☐ No	☐ Unk
Diabetes	☐ Yes	☐ No	☐ Unk
Born in country of high endemicity	☐ Yes	☐ No	☐ Unk
Residence in country of high endemicity	☐ Yes	☐ No	☐ Unk
Contact of case	☐ Yes	☐ No	☐ Unk
Alcohol misuse	☐ Yes	☐ No	☐ Unk
Drug misuse	☐ Yes	☐ No	☐ Unk

If other/additional risk factors present (please specify)

Immune code ☐ Positive ☐ Negative ☐ Unk

Is this case currently on ARV* treatment? ☐ Yes ☐ No ☐ Unk

Is this case linked to an outbreak? ☐ Yes ☐ No ☐ Unk

If YES, please specify outbreak code:

D. DIAGNOSTIC DETAILS

Direct sputum microscopy (DSM)

(a) 1st DSM result

☐ Positive
☐ Negative
☐ Not done

1st DSM date:

(b) 2nd DSM result

☐ Positive
☐ Negative
☐ Not done

2nd DSM date:

Microscopy of other specimens (e.g. BAL, gastric washings etc)

(a) 1st microscopy result

☐ Positive
☐ Negative
☐ Not done

1st microscopy date:

1st microscopy specimen type

(b) 2nd microscopy result

☐ Positive
☐ Negative
☐ Not done

2nd microscopy date:

2nd microscopy specimen type

Histology ☐ Positive ☐ Negative ☐ Not done

Histology specimen site

Culture results

(a) 1st Culture result

Culture positive ☐
Culture negative ☐
Not done ☐

1st Culture specimen type

1st Culture specimen site

(b) 2nd Culture result

Culture positive ☐
Culture negative ☐
Not done ☐

2nd Culture specimen type

2nd Culture specimen site

Mycobacterium tuberculosis complex (MTC) isolated? Yes ☐ No ☐ Unk ☐

If YES, please tick species identified (1 species only)

☐ *M. tuberculosis* ☐ *M. bovis* ☐ *M. africanum* ☐ *M. canetti* ☐ *M. caprae* ☐ *M. microti*

Drug sensitivities (R= res, S = sens, ND = not done)
(Please fill for each drug used)

1st line drugs	**S**	**R**	**ND**
Isoniazid	☐	☐	☐
Rifampicin	☐	☐	☐
Ethambutol	☐	☐	☐
Pyrazinamide	☐	☐	☐
Streptomycin	☐	☐	☐

2nd line drugs	**S**	**R**	**ND**
Amikacin	☐	☐	☐
Capreomycin	☐	☐	☐
Ciprofloxacin	☐	☐	☐
Ethionamide	☐	☐	☐
Kanamycin	☐	☐	☐
Moxifloxacin	☐	☐	☐
Ofloxacin	☐	☐	☐
Para-amino salicylic acid (PAS)	☐	☐	☐
Prothionamide	☐	☐	☐
Rifabutin	☐	☐	☐

Sensitivity / resistance pattern (tick 1 only)

	Yes	No	Unk
A) Pansensitive	☐	☐	☐
B) MDR-TB	☐	☐	☐
C) XDR-TB	☐	☐	☐

Nucleic acid amplification test (e.g. PCR)

☐ Positive for MTC ☐ Negative for MTC ☐ PCR not done

If positive, were genetic resistance determinants to the following drugs detected:

Isoniazid ☐ Yes ☐ No ☐ Unk
Rifampicin ☐ Yes ☐ No ☐ Unk

Genotyping MIRU done? ☐ Yes ☐ No ☐ Unk

MTC lineage

MIRU-VNTR

E. OUTCOME DETAILS

Laboratory results : (Pulmonary cases ONLY):

	Direct Sputum microscopy				**Culture**			
	Pos	Neg	Not done	Sputum N/A	Pos	Neg	Not done	Sputum n/a
During treatment (at least 2 months)	☐	☐	☐	☐	☐	☐	☐	☐
Treatment end	☐	☐	☐	☐	☐	☐	☐	☐

Treatment Outcome (at 12 months)	Completed - cured ☐	Completed - status unknown ☐	Interrupted ☐	Transferred ☐
	Completed - failed ☐	Still on treatment ☐	Lost to follow up ☐	Died ☐
Treatment Outcome for MDR TB (at 24 months)	Completed - cured ☐	Completed - status unknown ☐	Interrupted ☐	Transferred ☐
	Completed - failed ☐	Still on treatment ☐	Lost to follow up ☐	Died ☐
Treatment Outcome for XDR TB (at 36 months)	Completed - cured ☐	Completed - status unknown ☐	Interrupted ☐	Transferred ☐
	Completed - failed ☐	Still on treatment ☐	Lost to follow up ☐	Died ☐

Did drug resistance develop during treatment? Yes ☐ No ☐ Unk ☐ **If YES:** MDR ☐ XDR ☐ Other resistance ☐

If other resistance, please specify:

DOTS recommended? Yes ☐ No ☐ Unk ☐ **DOTS commenced?** Yes ☐ No ☐ Unk ☐ **DOTS successful?** Yes ☐ No ☐ Unk ☐

If treatment completed, date of completion

If deceased, was TB the direct cause? Yes ☐ No ☐ Unk ☐

If deceased, date of death

Case denotified (i.e. was diagnosis changed?) Yes ☐ No ☐ Unk ☐

If YES, please specify new diagnosis

Case classification (tick 1 only): Possible ☐ Probable ☐ Confirmed ☐

F. CONTACT TRACING DETAILS

Is this case: Index case ☐ **OR** Contact of another case ☐ *(please tick one)*

If this case is a contact of another case, please complete the following questions:

Nature of contact:

Family ☐ Healthcare setting ☐ Work ☐ Other ☐
School/college ☐ Longstay care facility ☐ Prison ☐
If other, please specify:

Did this case comply with contact tracing? **Yes** ☐ **No** ☐

Name of index case
Date of notification of index case
CIDR Event ID of index case

COMPLETING DOCTOR SIGNATURE

Tick section(s) completed:

				A	B	C	D	E
Signature 1		**Date 1**		Section completed:					
Signature 2		**Date 2**		Section completed:					
Signature 3		**Date 3**		Section completed:					
Signature 4		**Date 4**		Section completed:					
Signature 5		**Date 5**		Section completed:					

COMMENTS

EU Case Definition for TB

Irish standardised case definitions for notification of a TB case:

under S.I. No. 452/2011 Infectious Diseases (Amendment) Regulations 2011

Tuberculosis (*Mycobacterium tuberculosis* complex including; *M. africanum*, *M. bovis*, *M. canetti*, *M. caprae*, *M. microti*, *M. pinnipedii* and *M. tuberculosis*)

Clinical Criteria - Any person with:

- Signs, symptoms and/or radiological findings consistent with active tuberculosis in any site
  AND
- A clinician's decision to treat the person with a full course of anti-tuberculosis therapy

OR

- A case discovered post-mortem with pathological findings consistent with active tuberculosis that would have indicated anti-tuberculosis antibiotic treatment had the patient been diagnosed before dying

Possible case - A person meeting the clinical criteria without laboratory confirmation

Probable case - A person meeting the clinical criteria with at least one of the following:

- Microscopy positive for acid-fast bacilli or equivalent fluorescent staining bacilli on light microscopy

OR

- Detection of *Mycobacterium tuberculosis* complex nucleic acid in a clinical specimen

OR

- Histological appearance of granulomata

Confirmed case - A person meeting the clinical criteria with:

- Detection of *M. tuberculosis* complex nucleic acid in a clinical specimen
  AND
- Positive microscopy for acid-fast bacilli or equivalent fluorescent staining bacilli on light microscopy

OR

- Isolation of *M. tuberculosis* complex (excluding *M. bovis*-BCG) from a clinical specimen

Abbreviations:

***ARV treatment:** Anti-retroviral treatment

Appendix C

Memorandum Of Understanding

Feidhmeannacht na Seirbhíse Sláinte
Health Service Executive

hpsc

Memorandum of Understanding between Trinity College Dublin (Dr. Ronan O'Toole, Professor Catherine Comiskey, Aidan Hanway, and Dr. Katy Tobin) and HSE- Health Protection Surveillance Centre (HSE-HPSC), regarding the provision of TB data for a study on (i) Risk Factors associated with TB in Ireland

April 2014

The HSE- Health Protection Surveillance Centre (HSE-HPSC) has been collecting surveillance data on TB since 1998. The collection of these data is a collaborative process with TB cases notified to regional departments of Public Health by clinicians and laboratories. Data collected include all available clinical, microbiological, histological and epidemiological data. Surveillance forms are then collated in the regional departments of public health, where data were entered onto an Epi2000 database (NTBSS) up to 2011. Data from 2011onwards are reported via the Computerised Infectious Disease Reporting (CIDR) system. Each HSE area provides finalised annual data (with outcome information) to HSE-HPSC. Data are validated with each area and national data are collated from which annual TB reports are produced.

A team from Trinity College Dublin which includes Dr. Ronan O'Toole (Department of Clinical Microbiology, School of Medicine) and Professor Catherine Comiskey, Aidan Hanway, and Dr. Katy Tobin (School of Nursing and Midwifery, Trinity College Dublin) are currently collaborating with HSE-HPSC on exploring the following issue: (i) risk factors associated with TB in Ireland. Mr. Hanway is using these data to form the basis of his PhD thesis. This work will be conducted in accordance with AMNCH/SJH Research Ethics Committee approval (Reference REC: 2013/3/12).

(i) The aim of this work is to use bio-mathematical and bio-statistical modelling techniques to model the impact of host-specific and environmental risk factors on the development of TB in Ireland. Univariate and multivariate statistical analyses will be used to analyse data on risk factors for TB. TB case factors that will be analysed in this study include the following:

- Birth or Residence in a country of high endemicity
- Diabetes
- Alcohol or drug misuse
- Immunosuppressive illness or medication
- Contact with a confirmed case
- Homelessness (where data are available)
- Imprisonment (where data are available)
- Other risk factors where data are available

Non linear ordinary differential equation compartment models will be generated. Using the refined models and the epidemiological parameters, a series of simulations and scenario analyses will be performed to predict the impact that increasing, or decreasing level of specific risk factors would have on the incidence of TB in Ireland. The knowledge generated will assist in identifying changes, or emerging trends, with respect to the effect of specific risk factors on TB incidence over time.

In order to facilitate the research outlined above, HSE-HPSC has made the national TB surveillance data available to the aforementioned team from Trinity College Dublin in order to further explore the possibilities for these studies. In addition, the aforementioned team will work with these data at TCD to explore their potential for use in the research.

In the above context, HSE-HPSC is requesting that they are advised of the progress of this research work e.g. used for PhD thesis or any other projects, publications etc. It is critical that this request is met as HPSC have legal and ethical obligations as trustees of this national data.

In the context of this study, if ethnicity data are used in the analyses, consultation will be made, and agreement sought with the Director of the Pavee Point Traveller and Roma Centre prior to publication of data or findings specifically referring to TB in the Irish Traveller population. This work will be conducted in accordance with AMNCH/SJH Research Ethics Committee approval (Reference REC: 2014/037 2014/02).

In addition, HSE-HPSC requests that it be included as joint authors in any publications relating to these data as per the criteria outlined by the International Committee of Medical Journal Editors as outlined below (ICMJE)

- Substantial contributions to the conception or design of the work; or the acquisition, analysis, or interpretation of data for the work; AND
- Drafting the work or revising it critically for important intellectual content; AND
- Final approval of the version to be published; AND
- Agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

The purpose of this memorandum is to outline the nature of the data being made available to the Trinity College team by HSE-HPSC and also to satisfy all parties involved that Data Protection legislation is not being breached.

To facilitate the afore-mentioned research, HSE-HPSC will provide following data to the Trinity College team:

- Anonymised data on the recording of host and environmental risk factors associated with the incidence of TB

The HPSC-HSE will not provide the Trinity College team with any data that could be used to identify an individual under any circumstance.

The Trinity College team will not seek to link-back the data received from HSE-HPSC to any individual patient under any circumstance.

The Trinity College team will also not share the data with any third party without prior agreement with HSE-HPSC. For the avoidance of doubt, TCD

employees and/students working on the dataset in conjunction with the Trinity College team are not considered third parties.

Signed by:

_____

Dr. Ronan O'Toole (TCD):

Ronan O'Toole

Date: 16 April 2014

_____

Professor Catherine Comiskey (TCD):

Catherine Comiskey

Date: 9 June 2014

_____

Aidan Hanway (TCD)

Aidan Hanway

Date: 9 June 2014

_____

Dr. Katy Tobin (TCD)

Katy Tobin

Date: 9 June 2014

_____

Dr. Joan O'Donnell:
(on behalf of HPSC) _____

Date:

Appendix D

Publication A

D.1 Declaration Of Work Contributed

All work in the following publication was primarily conducted by myself and overseen by the co-authors. The manuscript was accepted July 2016.

Tuberculosis 101 (2016) 67–74

Contents lists available at ScienceDirect

Tuberculosis

journal homepage: http://intl.elsevierhealth.com/journals/tube

EPIDEMIOLOGY

Relating annual migration from high tuberculosis burden country of origin to changes in foreign-born tuberculosis notification rates in low-medium incidence European countries

Aidan Hanway [a], Catherine M. Comiskey [a], Katy Tobin [a, b], Ronan F. O'Toole [c, *]

[a] School of Nursing and Midwifery, Trinity College Dublin, Ireland
[b] Academic Unit of Neurology, Trinity College Dublin, Ireland
[c] Breathe Well Centre, School of Medicine, University of Tasmania, Hobart, Australia

ARTICLE INFO

Article history:
Received 31 March 2016
Received in revised form
22 July 2016
Accepted 31 July 2016

Keywords:
Tuberculosis
Migration
High TB burden country

SUMMARY

The level of immigration from high tuberculosis (TB) burden countries (HBCs) which impacts on the foreign-born TB notification rate is largely unknown. In this work, we performed a cross-sectional analysis of epidemiological data from 2000 to 2013 from nine European countries: Austria, Denmark, Finland, Hungary, Netherlands, Norway, Spain, Sweden, and the United Kingdom. Crude notification rates were calculated for foreign- and native-born populations and a multiple-linear regression model predicting notification rates with HBC population data was generated. From 2000 to 2013, the population percentage with a foreign birthplace increased on average each year in all nine countries, ranging from +0.11%/year in the Netherlands to +0.66%/year in Spain. An annual increase in HBC migrants above +0.43% per year (95% Confidence Interval: 0.24%–0.63%) corresponded with higher TB notification rates in the foreign-born population of the countries analyzed. This indicates that migration from HBCs can exert a measurable effect on the foreign-born TB notification rate. However, an increase in the foreign-born TB notification rate coincided with an average annual rise in national TB notification rates only in countries, Norway (+3.85%/year) and Sweden (+2.64%/year), which have a high proportion (>80%) of TB cases that are foreign-born.

1. Introduction

TB is on the decline in most European countries. In the World Health Organization (WHO) European Region, the annual notification rate of TB has fallen from 47.2 cases per 100,000 in 2003 [1] to 39 cases per 100,000 in 2013 [2]. In the European Union (EU) and European Economic Area (EEA), the notification rate has continued to decrease from 19.8 per 100,000 in 2003 [3] to 12.7 per 100,000 in 2013 [2]. Within individual countries, there has also been a shift in the demographics of TB during this time with an increasing proportion of cases occurring in the foreign-born population. In the United Kingdom (UK), for example, the proportion of TB cases that occur in the foreign-born population, as a percentage of total cases, rose from 51% in 2000 to 70% in 2013. Foreign-born TB cases account for a high proportion of notifications in non-European high-income countries including Australia (88%) [4], New Zealand (79.5%) [5], Canada (71%) [6], and the USA (66%) [7].

It should be noted that a number of studies have found that immigrants do not appear to be a major source of TB for the native-born population in Europe. Using IS6110 DNA fingerprinting and spoligotyping, Barniol and co-workers in Germany determined that of 16 mixed TB clusters, defined as clusters containing both native- and foreign-born persons, 7 were first identified in German-born individuals [8]. They estimated the proportion of German TB cases that were caused by foreign-born cases to be 18.3% (95% CI: 8.3–28.3%) [8]. In Denmark, this proportion was between 2.7% (95% CI: 2.0–3.6) and 5.8% (95% CI: 4.8–7.0) in terms of cases in Danes caused by foreign-born cases. Conversely, 6.5% (95% CI: 5.6–7.5) to 7.9% (95% CI: 7.0–8.9) of migrants were estimated to be infected by Danes [9]. In a study from Spain, the prevalence of latent TB infection (LTBI) and of TB disease was higher in contacts of immigrant TB cases (51.5% and 1.8%,

* Corresponding author. Breathe Well Centre, School of Medicine, Faculty of Health, University of Tasmania, Medical Science 1 Building, 17 Liverpool Street, Hobart, TAS 7000, Australia.
E-mail address: ronan.otoole@utas.edu.au (R.F. O'Toole).

http://dx.doi.org/10.1016/j.tube.2016.07.019
1472-9792/

respectively) compared to contacts of indigenous cases (29.3% and 1.3%, respectively) [10]. However, when controlling for country of origin and BCG vaccination status, the investigators did not find evidence that immigrant index cases transmitted TB more than indigenous index cases [10]. This is in agreement with a study from Switzerland which found that immigrants were not more significantly linked to recent transmissions than Swiss-born TB patients (adjusted odds ratio of 1.58, 95% CI: 0.73–3.43, $p = 0.25$) [11]. In their systematic review, Sandgren and colleagues concluded that "TB in a foreign-born population does not have a significant influence on TB in the native population in EU/EEA" [12].

Nevertheless, a major focus of TB control in a number of low incidence countries has been on detecting TB in the foreign-born population through pre-immigration screening [19]. For example, of an estimated 378,939 visa applicants to Australia who underwent pre-immigration medical examinations during the 2009–2010 financial year, 519 people were diagnosed with active TB [13]. This corresponded to a TB prevalence of 137 per 100,000 examined population compared to 7.9 per 100,000 general population in Australia in 2009 [13]. Thus, the authors of this study concluded that premigration health screening of intending migrants is successful in identifying substantial numbers of people who would otherwise require treatment for TB after arriving in Australia [13]. But pre-entry screening of foreign-born individuals is not always possible in the case of asylum seekers or humanitarian refugees, or with regard to the provision of passport-free movement of people within the EU Schengen Area. Furthermore, it is believed that most cases of TB in the foreign-born population occur due to reactivation of latent LTBI rather than continuation of an existing case of active TB [14]. For example, among 65,529 new TB cases reported to the National Tuberculosis Surveillance System in the USA from 2005 to 2009, 83.4% of cases in foreign-born persons were attributed to reactivation of latent TB infection [15]. Therefore, even with active pre-immigration screening measures, a substantial proportion of TB that occurs in the foreign-born population in low TB incidence countries may not be detectable prior to entry.

The objective of this work was to explore the possibility that temporal monitoring of the demographics of a country's population may offer a measure that could be used to calculate whether TB incidence in a given section of the population is expected to increase. Statistical and mathematical modelling was performed on population census and TB surveillance data from 2000 to 2013. These analyses identified a predictor of changes in TB notification rates in the foreign-born population.

2. Materials and methods

2.1. Data sources

TB disease notifications along with foreign and native-born notifications from 2000 through 2013 were obtained from the Surveillance Atlas of Infectious Diseases, a tool publicly hosted by the European Centre for Disease Prevention and Control (ECDC) [16]. The ECDC defines each country within this study as having compulsory reporting of notifications, having comprehensive reporting, having case-based notification rates, and as having national coverage [2]. Foreign- and native-born notification data were complete (2000–2013) for all countries that passed the selection criteria except for Spain and the Netherlands, which had data available from 2007 to 2013 and 2005 to 2013, respectively. Publicly-available denominator data was obtained from the Organization for Economic Cooperation and Development (OECD) [17]. Denominator data of EU/EEA countries included total population estimates, foreign-born population estimates (aggregate total for each year), and each country's population estimates for individuals originating from high TB burden countries (HBC).

The 30 HBCs analysed in this study were as per the WHO's updated HBC definition *i.e.* the 20 countries with the highest estimated numbers of incident TB cases (Angola, Bangladesh, Brazil, China, DPR Korea, DR Congo, Ethiopia, India, Indonesia, Kenya, Mozambique, Myanmar, Nigeria, Pakistan, Philippines, Russian Federation, South Africa, Thailand, UR Tanzania, Viet Nam), plus the top 10 countries with the highest estimated TB incidence rate per 100,000 population, that are not in the top 20 by absolute number, with a minimum threshold of 10,000 estimated incident TB cases per year (Cambodia, Central African Republic, Congo, Lesotho, Liberia, Namibia, Papua New Guinea, Sierra Leone, Zambia, Zimbabwe).

The UK had population data for 18 of the 30 HBC populations living within their states, Spain had 19 of the 30 HBC populations, and the Netherlands had 29 of the 30 HBC populations living within its state. The remaining six countries had population data for all of the 30 HBC populations living within their jurisdiction.

2.2. Country selection criteria

Country selection criteria were applied for the purposes of establishing common underlying trends within countries with large foreign-born populations. The selection criteria were based on foreign-born population estimates and the quality of countries population surveillance data. A country was selected for analysis if:

1. It had documentation for at least half[1] of the 30 HBC populations living within its jurisdiction.
2. In the year 2013, it had an aggregate total of 300,000 foreign-born individuals or more living within its jurisdiction.

Applying these criteria resulted in 9 European Union countries being selected for analysis *i.e.* Austria, Denmark, Finland, Hungary, Netherlands, Norway, Spain, Sweden, and the UK. As each of these countries have large foreign-born populations, an aggregation of foreign-born data was performed to determine whether mutual trends exist within the population, specifically trends relating to country of origin.

2.3. Statistical analysis

All statistical analyses were completed in R and Excel. Crude notification rates were calculated for foreign- and native-born populations in each respective country for each year. Odds ratios with 95% confidence intervals were calculated for foreign-born populations of each country in each year. A multiple linear regression model was calculated to estimate foreign-born notification rates using the predictors: time; and HBC population density using all country data. The regression took the following form:

$$Y(t) = B_0 + B_1 * X(t) + B_2 * t + \varepsilon$$

Where B_0, B_1, and B_2 are the constant coefficients of the regression model that represent the underlying change in the dependent variable ($Y(t)$) which is a function of the independent variables ($X(t)$, t). The variable $Y(t)$ represents a country's foreign-born notification rate (per 100,000) at time t, the variable $X(t)$

[1] The study included countries that documented, at a minimum, half of the 30 high TB burden countries of origin. The nine countries which met this criterion had surveillance data for 18 of the HBCs.

represents the percentage of a country's foreign-born population of HBC origin at time t, and where t is time (in years, from 2000 to 2013).

The equation was derived by simply examining the relationship between countries' foreign-born notification rates ($Y(t)$) and countries' respective densities of HBC individuals ($X(t)$). This would be represented by the simple linear regression $Y(t) = B_0 + B_1 * X(t) + \varepsilon$, for which a time variable, t, has been added. This was due to the fact that there also exist an underlying trend within foreign-born notifications as time progresses. This trend represents other factors the model has not considered. The original equation can be summarized within the statement "Foreign-born notification rates depend on the density of HBC individuals and the time point a country is currently at".

The equation has the expectancy

$$E[Y(t)] = B_0 + B_1 * X(t) + B_2 * t.$$

Taking the derivative with respect to time gives

$$\frac{dE[Y(t)]}{dt} = B_1 * \frac{dX(t)}{dt} + B_2$$

which is an estimate of change over time of foreign-born TB incidence. The equation yields the following equality:

$$\frac{dE[Y(t)]}{dt} = 0 \quad \text{holds when} \quad \frac{dX(t)}{dt} = -\frac{B_2}{B_1}$$

This states that when HBC density has changed exactly $-\frac{B_2}{B_1}$ units, expected foreign born notification rate will not change. The inequalities are also true:

$$\frac{dE[Y(t)]}{dt} > 0 \quad \text{holds when} \quad \frac{dX(t)}{dt} > -\frac{B_2}{B_1},$$
$$\frac{dE[Y(t)]}{dt} < 0 \quad \text{holds when} \quad \frac{dX(t)}{dt} < -\frac{B_2}{B_1}$$

Put simply, this states that when HBC density has increased from one year to the next at a rate greater than $-\frac{B_2}{B_1}$ units, the expected foreign-born notification rate will increase. When HBC density has changed less than $-\frac{B_2}{B_1}$ units, expected foreign-born notification rate will decrease. Calculation of confidence interval values for the estimate $-\frac{B_2}{B_1}$ required advanced bootstrapping techniques. Bootstrapping methods were used due to the underlying theoretical distribution of $-\frac{B_2}{B_1}$ being a Cauchy distribution (with undefined variance) [18]. The R package 'boot' was utilized to acquire the interval [24]. A 95% adjusted bootstrap percentile interval [19] was calculated on the estimate $-\frac{B_2}{B_1}$. To demonstrate the above inequalities empirically, categories were established. A large increase in HBC density was classed to be any increase greater than the upper limit of the 95% confidence interval, and a moderate increase was classed as being the range between the point estimate and the upper 95% confidence limit. A small increase was classed as being the range between the point estimate and the lower 95% confidence limit, and lastly a very small increase/decrease was constructed to be less than the lower 95% confidence interval limit.

3. Results

3.1. Population proportion that is foreign born from 2000 to 2013

The foreign-born population, as a percentage of total population, increased in all nine countries between 2000 and 2013. The descriptive data can be seen in Table 1. The largest average annual percentage increase in foreign-born population was observed for Spain. The foreign-born population for 2013 in Spain was more than three times higher than in 2000. When the data from all nine countries are aggregated, the foreign-born populations have doubled in percentage terms from 2000 to 2013 (*i.e.* 1.97 times their level in 2000).

3.2. TB notifications between 2000 and 2013

Between 2000 and 2013, the national TB notification rate (per 100,000) increased in three of the nine countries: Norway, Sweden, and the UK. The remaining countries saw a decline notification rates. Data for each country can be seen within Table 2. The greatest decline was seen in Hungary, whereby, notification rates in 2013 were approximately 30% of their level in 2000.

3.3. Proportion of TB notifications among foreign- and native-born residents from 2000 to 2013

Between 2000 and 2013, there has been an increase in the proportion of TB cases that occur in the foreign-born population, compared to native born, in eight of the nine countries *i.e.* Austria, Finland, Hungary, the Netherlands, Norway, Spain, Sweden, and the UK (Figure 1). In 2013, foreign-born TB accounted for 51.2% of all TB notifications in Austria (+1.84% on average annually from 2000 to 2013), 60.4% of TB notifications in Denmark (−0.27% on average annually), 31.4% of TB notifications in Finland (+1.81% on average annually), 3.3% of TB notifications in Hungary (+0.13% on average annually), 73.9% of TB notifications in the Netherlands (+5.68% on average annually), 86% of TB notifications in Norway (+1.19% on average annually), 31.7% of TB notifications in Spain (+2.44% on average annually), 88.7% of TB notifications in Sweden (+1.7% on average annually), and 70.1% of TB notifications in the UK (+1.44% on average annually).

From 2000 to 2013, there has been a decline in the notification rate of TB in native-born populations for all nine countries. The average annual percentage changes in native-born TB incidence between 2000 and 2013 are as follows: Austria (−7.38%), Denmark (−2.37%), Finland (−5.97%), Hungary (−8.21%), Netherlands (−5.93%), Norway (−0.35%), Spain (−4.79%), Sweden (−2.68%), and the UK (−0.17%). The foreign-born TB notification rate has shown an average annual decline in six countries *i.e.* Austria (−2%), Denmark (−6.44%), Hungary (−2.44%), the Netherlands (−3.28%), Spain (−2.5%), and the UK (−1.04%). On the other hand, the foreign-born notification rate has exhibited an average annual increase in three of the nine countries *i.e.* Finland (+6.19%), Norway (+0.07%), and Sweden (+2.09%).

3.4. Odds ratios of TB in the foreign-born population

Odds ratios were constructed to minimize any mutual trends occurring between foreign-born and native-born TB notification rates. Table 3 details the annual odds ratios of TB for the foreign-born population compared to that of the native-born population. Norway has consistently had the largest odds ratio, averaging 44.1 each year (with a peak of 75.21 in 2009). Finland has shown the greatest average percentage growth in the odds ratio (+14.19% on average annually) while Denmark has shown the greatest decline (−3.91% on average annually). For the nine countries except the UK and Denmark, the odds ratio of foreign-born to native-born TB notification rate has increased between 2000 and 2013 (Table 3).

3.5. Predicting foreign-born TB notification rates

The proportion of the foreign-born population originating from high TB burden countries has on averaged changed

Table 1
Descriptive population data and rates of changes for foreign-born populations in each of the European countries.

Foreign population statistics	2000		2013		Average annual increase
	N	Percentage of total population	N	Percentage of total population	
Austria	843,000	10.52%	1,414,624	16.70%	+0.48%
Denmark	308,674	5.78%	476,059	8.50%	+0.21%
Finland	136,203	2.63%	304,268	5.59%	+0.23%
Hungary	294,573	2.88%	447,657	4.53%	+0.13%
Netherlands	1,615,377	10.14%	1,953,436	11.63%	+0.11%
Norway	305,035	6.79%	704,511	13.87%	+0.54%
Spain	1,969,270	4.86%	6,263,693	13.44%	+0.66%
Sweden	1,003,798	11.31%	1,533,493	16.10%	+0.37%
United Kingdom	4,184,429	7.11%	7,860,000	12.56%	+0.42%

Table 2
TB notification rates for the years 2000 and 2013 in each of the European countries, along with the average annual percentage change.

Country notification rates (per 100,000)	2000	2013	Average annual percentage change
Austria	15.29	7.66	−5.00%
Denmark	10.26	6.35	−3.28%
Finland	9.84	4.98	−4.13%
Hungary	35.24	10.57	−8.37%
Netherlands	8.82	5.05	−3.94%
Norway	5.28	7.89	3.85%
Spain	20.70	11.89	−4.05%
Sweden	5.16	6.80	2.64%
United Kingdom	11.42	12.61	0.87%

annually +0.08% for Austria, +0.10% for Denmark, +0.60% for Finland, +0.22% for Hungary, −0.09% for the Netherlands, +0.05% for Norway, +0.24% for Spain, +0.34% for Sweden, and −0.40% for the UK. Examining foreign-born notification rates as a function of the concentration of the HBC population yields the following model, Figure 2 displays a scatter plot indicating a potential relationship.

A multiple linear regression model was calculated using the data from all nine countries in order to construct a general model that predicts foreign-born notification rate based on HBC density within the foreign-born population and on a country's time point. A significant regression equation was found ($F(2,109) = 250.89$, $p < 0.001$), with an R^2 of 0.82 and adjusted R^2 of 0.818, which indicates a potential relationship between a country's high TB burden population proportion and its overall foreign-born TB notification rate. Tables 4 and 5 illustrate the regression model's coefficients and residual statistics for each individual country. The model predicts that when the percentage of HBC born individuals is set to zero within a country and time is set to one (the year 2000), the foreign-born notification rate is expected to be 16.43 and will decline 0.96 units each year thereafter. In contrast, if a country has 100% of its foreign-born population originating from a HBC and time is set to one, the foreign-born notification rate is expected to be 240.94 with an annual decline of 0.96 units each year.

The underlying regression equation has threshold $-\frac{B_2}{B_1}$, which is the value $-\frac{(-0.96)}{224.51}$, which is 0.43%. A 95% confidence interval was constructed on this value using an adjusted bootstrap percentile (BCa) interval [19] ($N = 10^6$, std. error = 0.099%) with lower limit 0.24% and upper limit 0.63%. The result of 0.43% is a threshold value which relates to foreign-born TB notification rate within the model. If the annual change in HBC density is greater than +0.43% for a specific year, the foreign-born TB notification rate is expected to rise. If the change is less than +0.43% for a specific year, the foreign-born TB notification rate is predicted to decline. Using the point estimate, along with the upper and lower limits of the confidence interval, distinct categories were established. A large increase in HBC density is considered to be a year-to-year change of greater than +0.63%. A moderate increase in HBC density is considered to be a year-to-year change between +0.43% and +0.63%. A small increase in HBC density is considered to be a year-to-year change between +0.24% and +0.43%. Lastly, a very small increase or a decline in HBC density is considered to be a year-to-year change less than +0.24%. The threshold can be empirically visualized in Table 6 using the data from all nine countries.

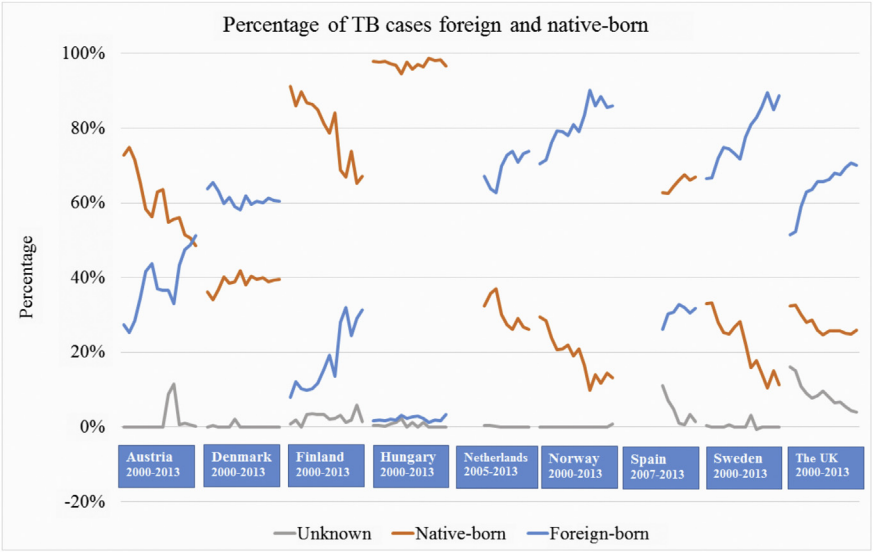

Figure 1. Trends in the proportion of TB cases among the native-born and foreign-born populations from 2000 to 2013. The proportion of TB cases that occurred in the foreign-born population, compared to native born, was examined between 2000 and 2013. An increase in the proportion of TB cases that occurred in the foreign-born population was observed in eight of the nine countries. Denmark was the only country to exhibit an average annual decrease in the proportion of TB cases that occurred in foreign born (−0.27% on average per year). Note, the observed simultaneous increase in both of the knowns *i.e.* native-born and foreign-born, as a percentage of total TB cases for Spain is due to a decrease in the "Unknown" variable with respect to birthplace between 2000 and 2013.

Table 3
Odds ratios of foreign-born to native-born TB notification rates in 2000, 2005, 2010 and 2013. A 95% confidence interval on the point estimate is provided (lower 95%, upper 95%).

Country	2000	2005	2010	2013
Austria	3.19 (2.82, 3.62)	4.57 (4.03, 5.18)	4.23 (3.64, 4.92)	5.27 (4.51, 6.14)
Denmark	28.88 (24.26, 34.38)	21.86 (17.98, 26.58)	17.99 (14.59, 22.17)	16.44 (13.29, 20.33)
Finland	3.21 (2.32, 4.42)	3.92 (2.83, 5.43)	9.83 (7.76, 12.46)	7.89 (6.1, 10.2)
Hungary	0.55 (0.42, 0.71)	1 (0.77, 1.28)	0.26 (0.17, 0.4)	0.72 (0.51, 1.01)
*Netherlands		17.45 (15.4, 19.76)	22.35 (19.5, 25.62)	21.53 (18.47, 25.09)
Norway	32.81 (24.82, 43.38)	39.8 (30.1, 52.63)	46.66 (34.28, 63.49)	40.48 (30.31, 54.07)
†Spain			2.97 (2.83, 3.12)	3.05 (2.88, 3.23)
Sweden	15.82 (13.02, 19.23)	19.28 (15.99, 23.25)	34.89 (28.07, 43.36)	40.91 (32.08, 52.17)
United Kingdom	20.76 (19.67, 21.9)	24.17 (23, 25.41)	21.21 (20.18, 22.29)	18.36 (17.46, 19.3)

* Foreign-born TB notification data for the Netherlands were available from 2005.
† Foreign-born TB notification data for Spain were available from 2007.

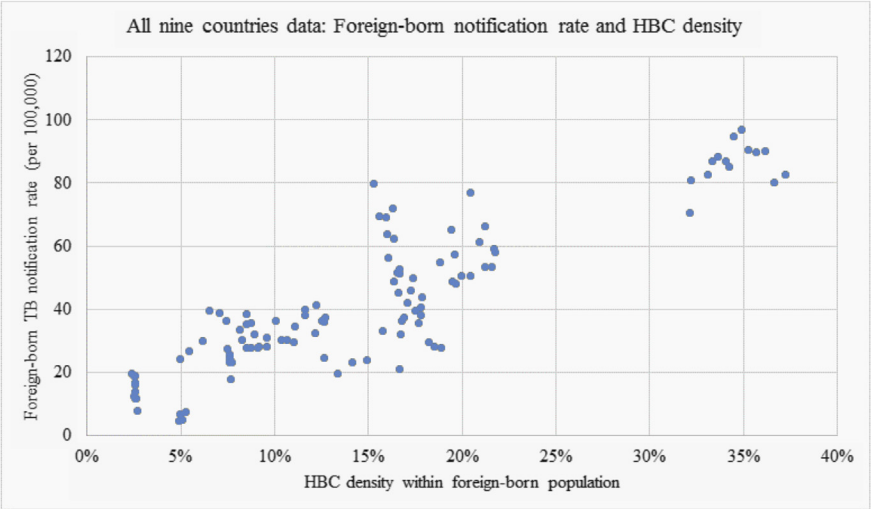

Figure 2. The TB notification rate in the foreign-born population versus the percentage of foreign born from a high TB burden country of origin. The proportion of the foreign-born population originating from high TB burden countries (HBC) was examined. A scatter plot illustrates an apparent linear relationship between the foreign-born TB notification rate and HBC population density as a proportion of the foreign-born population.

3.6. Model reliability

Data for Spain and the Netherlands were limited to the years 2007–2013 and 2005–2013, respectively. An alternative model was calculated excluding Spain and the Netherlands. An additional altered model was calculated excluding 2013 data to validate the predictability of the model. The results of both regression models are presented in Table 7.

Examining the direct effects of HBC density on foreign-born notifications, a linear model displayed one of the largest R-square values (R-square = 0.79) when compared with a(n): exponential model (R-square = 0.64), logarithmic model (R-square = 0.67), and a power model (R-square = 0.69). In addition, polynomial models were calculated up to order six. The R-square and Akaike Information Criterion (AIC) estimates for polynomial of order one were: R-square = 0.79, AIC = 527.66; of order two: R-square = 0.79, AIC = 529.67; of order three: R-square = 0.79, AIC = 531.63; of order four: R-square = 0.80, AIC = 532.64; of order five: R-square = 0.80, AIC = 533.99; and of order six: R-square = 0.80, AIC = 535.99. The R-square value was observed to be largest within a polynomial model of order six, however, the AIC estimate was minimised within a linear model.

Table 4
Coefficients, standard errors, and significance of the multiple linear regression.

Model	Coefficients		*p* value
	Beta	SE-B	
Intercept	16.43	2.55	<0.001
Percentage of foreign-born population from a high TB burden country	224.51	10.12	<0.001
Time	−0.96	0.24	<0.001

4. Discussion

In this work, we used population census and TB surveillance data to analyze the demographics of TB in nine low to medium incidence countries within the EU and EEA with the most complete TB data available. In all nine countries, there was an average annual increase in the foreign-born population with the highest increases of +0.54% and +0.66% per year occurring in Norway and Spain, respectively. However, the national TB notification rate (per 100,000) increased in just three of the nine countries *i.e.* Norway, Sweden, and the UK, but declined in Austria, Denmark, Finland, Hungary, Netherlands, and Spain. In all nine countries, there was an annual decline in the notification rates of TB within the native-born populations. An examination of odds ratios of foreign-born to native-born TB notification rates revealed that Norway has consistently had the largest odds ratio, averaging 44.1 each year between 2000 and 2013 (Table 3). An average annual increase in the TB notification rate in the foreign-born population was

Table 5
Regression model residual statistics for each individual country.

Country	Model residual statistics				2013 data	
	Max	Min	Range	Average	Model notification rate	Actual notification rate
Austria	11.20	−6.01	17.21	2.51	20.0	23.5
Denmark	31.96	3.47	28.49	15.59	40.3	45.2
Finland	10.63	−24.21	34.84	−13.14	45.3	28.0
Hungary	2.47	−17.55	20.02	−4.29	14.8	7.7
Netherlands	0.33	−12.68	13.01	−6.65	40.5	32.1
Norway	18.37	−8.42	26.79	2.33	46.7	49.0
Spain	7.48	−2.00	9.48	7.41	24.6	28.0
Sweden	11.76	−5.82	17.58	0.33	31.5	37.5
United Kingdom	8.02	−16.56	24.59	−0.54	75.1	70.4

observed for Finland (+6.19%), Norway (+0.07%), and Sweden (+2.09%) but not for the other six countries. This raised the question as to why the notification rate of TB in the foreign-born population was increasing in some countries but declining in others. We therefore examined whether different levels of annual immigration from HBCs corresponded with a decrease, or increase, in the foreign-born TB notification rate.

The average annual change in the proportion of the foreign-born population originating from high TB burden countries between 2000 and 2013 ranged from −0.4% for the UK to +0.6% for Finland. Regression analysis indicated that a significant relationship ($p < 0.001$, $R^2 = 0.82$) exists between a country's high TB burden population proportion, and its overall foreign-born TB notification rate. The effects of changes in the HBC density on the TB notification rate of the foreign-born population are illustrated in Table 6 and Figure 3. Very small or small increases in HBC density have a negligible effect on the foreign-born TB notification rate. Moderate or large increases in HBC density have a measurable effect on the foreign-born TB notification rate. A specific threshold was observed with respect to the level at which immigration from HBCs corresponded with a higher foreign-born TB notification rate in the countries analysed. Under the regression model, if the annual change in the HBC population, as a proportion of the foreign-born population, is greater than +0.43% (95% CI: 0.24%–0.63%) for a specific year, the foreign-born TB notification rate is predicted to rise. The threshold is an additive value rather than a percentage increase value or a multiplication factor. For example, if a country has 10% of its foreign-born population originating from a HBC in year 1, and 10.2% in year 2, the foreign-born TB notification rate would not be expected to increase going from year 1 to year 2. In contrast, if the foreign-born population originating from a HBC in year 2 was 10.8%, the foreign-born TB notification rate would be expected to increase going from years 1 to 2.

Two examples from the nine countries that illustrate the threshold are Finland and the Netherlands. From 2000 to 2009, Finland had 9 consecutive years in which its HBC density increased on average +0.71% each year, which is above the annual +0.43% threshold. HBC density started at 11.01% in the year 2000 and increased until 2009 where it reached 17.4%. The foreign-born TB notification rate also increased on average 1.123 units each year (per 100,000 population) in Finland. From 2005 to 2013 in the Netherlands, HBC density started at 17.85% and changed on average −0.14% each year (less than +0.43%). During this time, the foreign-born notification rate on average decreased 1.45 units each year (per 100,000 population) in the Netherlands. Examining the residuals for each individual country, the model appeared to fit best for Sweden and the UK (average residual 0.33 and −0.54, respectively) (Table 5). The models coefficients changed marginally when subsets of the data were used. When Spain and the Netherlands were excluded from the analysis, there was a minimal increase in the threshold from 0.43% to 0.44%. When 2013 data were excluded, the threshold decreased from 0.43% to 0.39%, within the confidence intervals of the threshold. It is worth noting that the calculation of the threshold can only be done within this specific regression model. While the model's coefficients and the underlying threshold exhibited minor variation within the alternative models detailed in Section 3.6, it is uncertain whether replication of this work can be accomplished within other temporal or spatial settings. The model appears appropriate when a linear relationship exists between two variables over a time interval.

Between 2000 and 2013, the proportion of TB cases that occurred in the foreign-born population, compared to native born, increased in eight of the nine countries. Figure 1 illustrates the “stage” each country was positioned with respect to this trend between 2000 and 2013. Sweden and Norway are within an apparent “late stage” as most of their TB is foreign born (>80%). An observed annual average increase in the foreign-born TB notification rate in Norway (+0.07% per year) and Sweden (+2.09% per year) coincided with an increase in the national TB notification rate in both countries *i.e.* +3.85% and 2.64% per year, respectively. Hence, in a so-called “late stage” country, an increase in the foreign-born TB notification rate can correspond with an increase in the national TB notification rate.

There are a number of limitations of this study that require acknowledgement. The term foreign-born is a broad term as it refers to an individual from any one of the other 197 countries in the world. Secondly, there are other potential factors besides HBC

Table 6
Effect of changes in HBC density on the foreign-born TB notification rate.

Annual change in a country's *HBC density	Average change in foreign-born TB notification rate	Average change in foreign-born TB cases (N)	Frequency of occurrence within the data (N)
Decrease/very small increase	−2.32	+7	63
Small increase	−2.12	+21	16
Moderate increase	+0.7	+19	13
Large increase	+3.48	+75	13

* HBC density is defined as the proportion of the foreign-born population from a high TB burden country of origin.

Table 7
Re-calculating the model with subsets of the data.

	Intercept coef. (B_0)	HBC density coef. (B_1)	Time coef. (B_3)	Threshold value	Model R−square
Original model (all data)	16.43^{***} (SE = 2.55)	224.51^{***} (SE = 10.12)	-0.96^{***} (SE = 0.24)	0.43%	0.82
Altered model 1 (Spain and the Netherlands excluded)	15.96^{***} (SE = 2.67)	229.8^{***} (SE = 10.62)	-0.994^{***} (SE = 0.26)	0.44%	0.83
Altered model 2 (exclusion of the year 2013)	15.78^{***} (SE = 2.7)	225.86^{***} (SE = 10.58)	-0.88^{***} (SE = 0.27)	0.39%	0.82

***Significant at the $p < 0.001$ level.

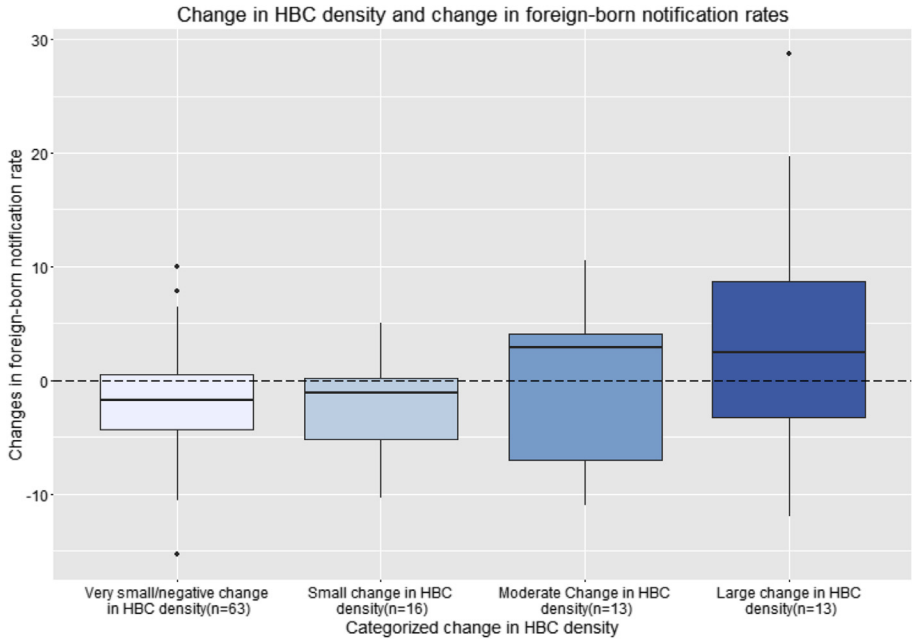

Figure 3. Effects of changes in the HBC density on the TB notification rate of the foreign-born population. The impact of the following changes: very small increase/negative change (less than +0.24%); small increase (between +0.24% and +0.43%); moderate increase (between +0.43% and +0.63%); large increase (greater than +0.63%), in annual migration from HBCs as a proportion of the foreign-born population, on the foreign-born TB notification rate was calculated. Very small or small increases in HBC density have negligible effect on the foreign-born TB notification rate. Moderate or large increases in HBC density have a measurable effect on the foreign-born TB notification rate. HBC density is defined as the proportion of the foreign-born population from a high TB burden country of origin.

proportion that could affect the TB notification rate in the foreign-born population. The TB notification rate of the foreign-born population of a given country may be influenced by the strength of TB risk factors that are present in its HBC cohort, in addition to the proportion of that population that originate from a HBC. Factors including homelessness, drug misuse, HIV/AIDS status, and diagnostic delay have been reported to potentially impact TB in foreign-born populations [20,21]. However, in one study, homelessness, drug misuse, and positive HIV/AIDS status were found to be infrequent in the foreign-born population [22]. Our study does not measure the contribution of individual risk factors but examines TB notification rates as a function of population density. Nevertheless, the validity of comparing notification rates is consistent with the finding from the USA that persons who have immigrated from areas of the world with high TB rates exhibit notification rates that approach those of their regions of origin for a number of years after arrival [23]. Despite limitations, our study indicates that the demographics of the immigration population can assist in predicting the probability of an increase in the foreign-born TB notification rate. It is not advocated or expected that a threshold alone be used in determining whether to introduce post-immigration TB screening. There are multiple other factors that will be considered by public health departments prior to the deployment of post-immigration TB screening. These factors, and the cost-effectiveness of a post-immigration TB screening strategy, may differ from one jurisdiction to another.

In summary, we have found that migration from HBCs can exert a measurable effect on the foreign-born TB notification rate. An annual increase in HBC migrants above +0.43% per year (95% CI: 0.24%–0.63%) in the nine countries analysed corresponded with higher TB notification rates in the foreign-born population. Only in countries Norway and Sweden, which have a high proportion of TB cases that are foreign-born (>80%), did an increase in the foreign-born TB notification rate, +3.85%/year and +2.64%/year, respectively, coincide with an average annual rise in national TB notification rates.

Funding: AH was supported by a Trinity College Postgraduate Research Studentship.

Competing interests: None declared.

Ethical approval: Not required.

References

[1] EuroTB and the national coordinators for tuberculosis surveillance in the WHO European Region. Surveillance of tuberculosis in Europe – report on tuberculosis cases notified in 2003. 2005.
[2] European Centre for Disease Prevention and Control/WHO Regional Office for Europe. Tuberculosis surveillance and monitoring in Europe 2015. 2015.
[3] Hollo V, Amato-Gauci A, Kodmon C, Manissero D. Tuberculosis in the EU and EEA/EFTA countries: what is the latest data telling us? EuroSurveillance 2009;14.
[4] Toms C, Stapledon R, Waring J, Douglas P. Tuberculosis notifications in Australia, 2012 and 2013. Commun Dis Intell 2015;39:E217–35.
[5] Institute of Environmental Science and Research Ltd. Tuberculosis in New Zealand. 2015. Annual Report 2013.
[6] Public Health Agency of Canada. Tuberculosis in Canada 2013-pre-release. 2015.
[7] Centers for Disease Control and Prevention. Reported tuberculosis in the United States, 2014. 2015.
[8] Barniol J, Niemann S, Louis VR, Brodhun B, Dreweck C, Richter E, Becher H, Haas W, Junghanss T. Transmission dynamics of pulmonary tuberculosis between autochthonous and immigrant sub-populations. BMC Infect Dis 2009;9:197.
[9] Kamper-Jorgensen Z, Andersen AB, Kok-Jensen A, Kamper-Jorgensen M, Bygbjerg IC, Andersen PH, Thomsen VO, Lillebaek T. Migrant tuberculosis: the extent of transmission in a low burden country. BMC Infect Dis 2012;12:60.
[10] Godoy P, Cayla JA, Carmona G, Camps N, Alvarez J, Rodes A, Altet N, Pina JM, Barrabeig I, Orcau A, Parron I, Alseda M, March J, Follia N, Minguell S, Dominguez A. Immigrants do not transmit tuberculosis more than indigenous patients in Catalonia (Spain). Tuberculosis (Edinburgh) 2013;93:456–60.
[11] Fenner L, Gagneux S, Helbling P, Battegay M, Rieder HL, Pfyffer GE, Zwahlen M, Furrer H, Siegrist HH, Fehr J, Dolina M, Calmy A, Stucki D, Jaton K, Janssens JP, Stalder JM, Bodmer T, Ninet B, Bottger EC, Egger M. Mycobacterium tuberculosis transmission in a country with low tuberculosis incidence: role of immigration and HIV infection. J Clin Microbiol 2012;50:388–95.
[12] Sandgren A, Schepisi MS, Sotgiu G, Huitric E, Migliori GB, Manissero D, van der Werf MJ, Girardi E. Tuberculosis transmission between foreign- and native-

born populations in the EU/EEA: a systematic review. Eur Respir J 2014;43:1159–71.
[13] King K, Douglas PJ, Beath K. Is premigration health screening for tuberculosis worthwhile? Med J Aust 2011;195:534–7.
[14] White RG, Houben RM. Towards elimination in industrialised countries: expanding diagnosis and treatment of LTBI among immigrants. Int J Tuberc Lung Dis 2014;18:380.
[15] Ricks PM, Cain KP, Oeltmann JE, Kammerer JS, Moonan PK. Estimating the burden of tuberculosis among foreign-born persons acquired prior to entering the U.S., 2005–2009. PLoS One 2011;6:e27405.
[16] Surveillance atlas of infectious diseases. Available online: http://ecdc.europa.eu/en/data-tools/atlas/Pages/atlas.aspx [accessed 28.01.15].
[17] International Migration Database. Available at: http://stats.oecd.org/Index.aspx?DataSetCode=MIG# [accessed 28.01.15].
[18] Weisstein EW. Normal ratio distribution. Available from: http://mathworld.wolfram.com/CauchyDistribution.html.
[19] Efron B. Better bootstrap confidence intervals. Stanford Univ. Dept. Statistics; 1984. Tech. Rep.
[20] Tomas BA, Pell C, Cavanillas AB, Solvas JG, Pool R, Roura M. Tuberculosis in migrant populations. A systematic review of the qualitative literature. PLoS One 2013;8.
[21] Ingrosso L, Vescio F, Giuliani M, Migliori GB, Fattorini L, Severoni S, Rezza G. Risk factors for tuberculosis in foreign-born people (FBP) in Italy: a systematic review and meta-analysis. PLoS One 2014;9:e94728.
[22] Chin DP, DeRiemer K, Small PM, de Leon AP, Steinhart R, Schecter GF, Daley CL, Moss AR, Paz EA, Jasmer RM, Agasino CB, Hopewell PC. Differences in contributing factors to tuberculosis incidence in U.S.-born and foreign-born persons. Am J Respir Crit Care Med 1998;158:1797–803.
[23] McKenna MT, McCray E, Onorato I. The epidemiology of tuberculosis among foreign-born persons in the United States, 1986 to 1993. N Engl J Med 1995;332:1071–6.
[24] Canty A, Ripley B. Package 'boot'. 2012. Available online at cran.r-project.org/web/packages/boot/boot.pdf; last accessed May 6, 2016.

Appendix E

Publication B

E.1 Declaration Of Work Contributed

Within the following paper statistical analysis was completed on the Irish traveller community. The work contributed by myself included: All statistical analysis that examined various demographics of the Irish traveller community, discussion of those statistical results, calculation of all confidence intervals presented within the study, and a written contribution to the to the reviewers responses of the study.

Epidemiol. Infect. (2015), **143**, 2849–2855. © Cambridge University Press 2015
doi:10.1017/S0950268815000138

Tuberculosis incidence in the Irish Traveller population in Ireland from 2002 to 2013

R. F. O'TOOLE[1,2]*, S. JACKSON[3], A. HANWAY[4], J. O'DONNELL[3], C. M. COMISKEY[4], T. R. ROGERS[2] AND D. O'FLANAGAN[3]

[1] *Breathe Well NHMRC Centre of Research Excellence, School of Medicine, University of Tasmania, Hobart, Australia*
[2] *Department of Clinical Microbiology, School of Medicine, Trinity College Dublin, Ireland*
[3] *Health Protection Surveillance Centre, Dublin, Ireland*
[4] *School of Nursing and Midwifery, Trinity College Dublin, Ireland*

Received 16 October 2014; Final revision 6 January 2015; Accepted 18 January 2015; first published online 12 February 2015

SUMMARY

The health status of the Irish Traveller ethnic minority is low compared to the general population in Ireland in terms of infant mortality rates and life expectancies. Respiratory disease is an area of health disparity manifested as excess mortalities in Traveller males and females. In this study, we examined the available data with regard to tuberculosis (TB) notifications in Ireland from 2002 to 2013. We found an increase in TB notifications in Irish Travellers from 2010 onwards. This resulted in a crude incidence rate for TB in Irish Travellers that was approximately threefold higher than that of the white Irish-born population in 2011 and 2012. An outbreak of TB in Irish Travellers in 2013 increased this differential further, but when outbreak-linked cases were excluded, a higher incidence rate was still observed in Irish Travellers relative to the general population and to white Irish-born. The mean age of a TB patient was 26 years in Irish Travellers compared to 43 years in the general population, and 49 years in white Irish-born. Based on available data, Irish Travellers exhibit a higher incidence rate and younger age distribution of TB compared to white Irish-born and the general population. These observations emphasize the importance of routine use of ethnicity identifiers in the management of TB and other notifiable communicable illnesses in Ireland. They also have implications for the orientation of preventive services to address health disparities in Irish Travellers and other ethnic minority groups.

Key words: Epidemiology, mycobacteria, tuberculosis (TB).

INTRODUCTION

A survey of the Irish Traveller population in Ireland and Northern Ireland published in 2010 found that infant mortality rates are about 3·5 times higher than in the general population [1]. Life expectancy at birth in Irish Traveller males and females is reportedly 15·1 and 11·5 years lower, respectively, compared to the national average [1]. In the 2011 Census of England and Wales, the median age of the Gypsy or Irish Travellers group was 26 years compared to the national median age of 39 years in the in the UK [2]. A local baseline census study indicated poor life

* Author for correspondence: Dr R. F. O'Toole, Breathe Well NHMRC Centre of Research Excellence, School of Medicine, University of Tasmania, Hobart, Australia.
(Email: ronan.otoole@utas.edu.au)

expectancy for Gypsies and Travellers in Leeds compared to the general Leeds population [3]. Currently, our understanding of the burden of communicable illnesses in the Traveller community, and the extent of its contribution to mortality and morbidity rates is relatively limited.

With respect to certain health conditions, general information is available in relation to mortalities in Irish Travellers. The All Ireland Traveller Health Study (AITHS) found that 13% of General Register Office-confirmed deaths in Irish Travellers in Ireland during a 1-year period (October 2007 to October 2008) were due to respiratory disease. The standardized mortality ratios due to respiratory disease were about 5·4- and 7·5-fold higher in Traveller females and males, respectively, than in the general Irish population [1]. Data outlining the proportion of Irish Traveller deaths in Ireland that were caused by respiratory infectious disease are not currently available. This limits the characterization of differential respiratory disease rates, and their associated determinants.

Tuberculosis (TB) is the leading cause of mortality due to respiratory infection worldwide, killing ~1·5 million people each year [4]. Using routine Irish enhanced TB surveillance data [5], we examined in this work the available case notification data with respect to TB in the Irish Traveller population.

METHODS

TB disease notifications for the period from 2002 to 2013 were extracted from the National TB Surveillance System and the Computerised Infectious Disease Reporting (CIDR) system (since 2011) by the Health Protection Surveillance Centre (HPSC). These data are collected by local departments of Public Health in Health Service Executive (HSE) areas in Ireland on each case of TB notified. Crude incidence rates (CIR) based on reported ethnicity were calculated using denominator data from the Census 2002, 2006, and 2011 data from the Central Statistics Office (CSO), Ireland [6], and also the AITHS data [1] on Traveller population size. Crude incidence rates (CIR)/100000 population were determined for each year from 2002 to 2013. Five-year cumulative CIR values were calculated for the period 2009–2013. Stratification of case based TB notifications by association with an outbreak is available from 2011 onwards. Average incidence rates for the period 2002–2013 were determined for each age group using Census data from the CSO [6]. Ethical approval for this study was obtained from the St James's Hospital/Adelaide & Meath National Children's Hospital Research Ethics Committee, Dublin, Ireland (reference: 2014/037/2014/02).

RESULTS

As illustrated in Figure 1, the CIR of TB in the Irish Traveller population, based on the number of cases notified, was at a level of <5 cases/100000 population per annum from 2002 to 2009. From 2010 onwards, an increase in notifications of TB cases in Irish Travellers was recorded. This resulted in a CIR >10 cases/100000 population per annum in 2011 to 2013 using the CSO Traveller population enumeration. In the general population during the same period, the CIR decreased from 10·5 new TB cases/100 000 population [95% confidence interval (CI) 9·5–11·5] in 2002 to 8·3/100000 population (95% CI 7·5–9·1) in 2013. The CIR for TB in Irish Travellers was about threefold higher than that of the white Irish-born population in the years 2011 and 2012. In 2013, an outbreak of TB in the Irish Traveller population resulted in a CIR of 40·6/100 000 (95% CI 21·0–71·1). When excluding outbreak cases, the CIR of TB in the Traveller population for 2013 was still higher (16·9/100000, 95% CI 5·5–39·6) than that of the general population (8·3/100000, 95% CI 7·5–9·1), and of the white Irish-born population (4·2/100000, 95% CI 3·5–4·8) (Fig. 1).

To allow for annual variations in TB cases in different populations, 5-year cumulative crude incidence rate values were calculated for the period 2009–2013. As illustrated in Table 1, the 5-year cumulative CIR of TB in Irish Travellers was 81·4/100000 (95% CI 29·6–197·7) compared to 45·5/100000 (95% CI 41·9–50·8) in the general population and 27·3/100000 (95% CI 22·9–30·3) in the white Irish-born population for 2009–2013.

The CIR is dependent upon the population data used for Irish Travellers. The total number of Irish Travellers in Ireland, enumerated from the April 2011 Census by the CSO, was 29573 (0·64%) of the total population [6]. The total number of Irish Travellers in Northern Ireland, enumerated from the March 2011 Census by the Northern Ireland Statistics and Research Agency, was 1301 (0·07%) of the total population [7]. This corresponds to an Irish Traveller population size in the region of 30874 for the island of Ireland in March/April 2011. By contrast, the AITHS conducted from 2007 to 2010

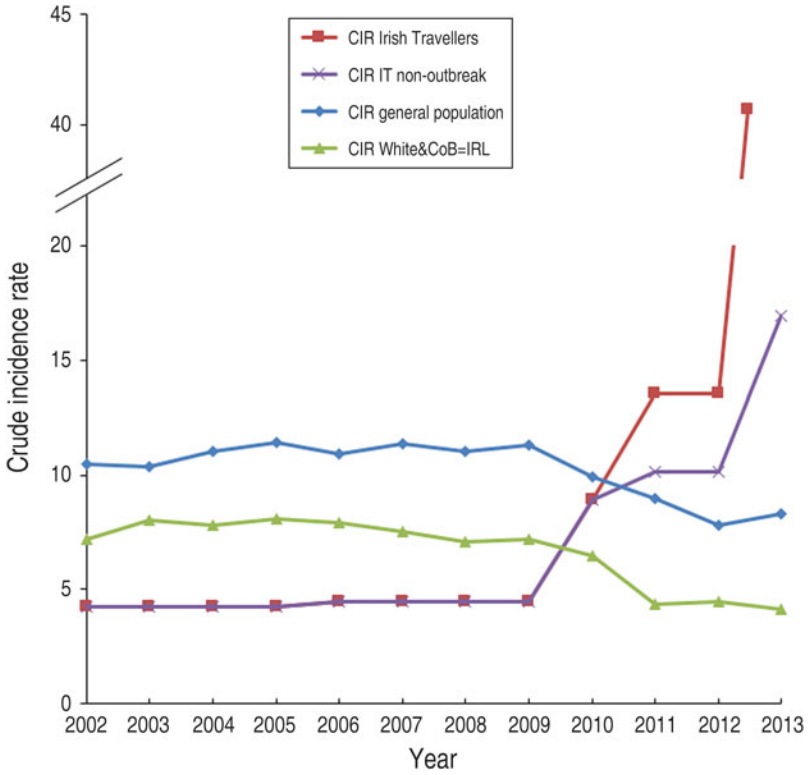

Fig. 1. Crude incidence rate (CIR)/100000 of TB in Ireland from 2002 to 2013. CIR of TB in the general population, white Irish-born in Ireland (white&CoB = IRL) and Irish Travellers are illustrated. The CIR in Irish Travellers excluding outbreak cases (IT non-outbreak) is also illustrated.

Table 1. *Cumulative notifications and crude incidence rates of TB from 2009 to 2013**

Ethnicity	Cumulative no. of TB notifications, 2009–2013	Cumulative TB CIR, 2009–2013 (Census 2011 population)	Cumulative TB CIR, 2009–2013 (AITHS population)
All ethnicities (including unknown)	2060	45·5	n.a.
White & CoB = IRL	983	27·3	n.a.
Irish Traveller CoB = IRL + UK	24	81·4	66·3

CIR, Crude incidence rate/100000; AITHS, All Ireland Traveller Health Study; CoB, country of birth; n.a., not available.
* Notification data for 2013 are provisional.

estimated the Irish Traveller population to be 36 224 in Ireland and 3905 in Northern Ireland, or 40 129 for the island of Ireland. As can be seen from Table 1, where the AITHS population data is used instead of the official Census data, the CIR of TB in the Irish Traveller population is lower. Where the higher enumeration of the Traveller population size from the AITHS is used, the 5-year cumulative CIR was still higher in Irish Travellers than in the general population and in the white Irish-born population (Table 1).

When the average incidence of TB for the period 2002–2013 is analysed by age, notable differences emerge between Irish Traveller cases and those in the general and white Irish-born populations. Figure 2 shows age-specific rates by Irish Traveller ethnicity from 2002 to 2013. The majority of cases in Irish Travellers were in the 0–34 years age groups with a mean age of 26 years and median age of 24 years. In the general population, the majority of cases occurred in the 25 to $\geqslant$65 years age groups

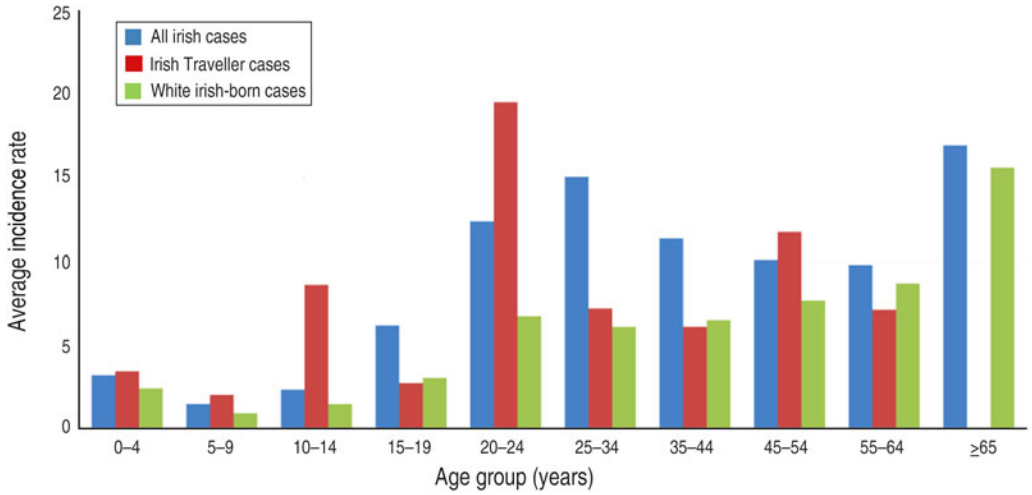

Fig. 2. Average TB incidence rate (per 100000) for the period 2002–2013 by age group (years) and age-specific rates by Irish Traveller ethnicity, and white Irish-born ethnicity.

with a mean age of 43 years and median age of 38 years. In the white Irish-born population, the majority of cases occurred in the 55 to ⩾65 years age groups with a mean age of 49 years and a median age of 49 years.

DISCUSSION

As recently noted by Tollefson *et al.*: 'the paucity of published information on TB burden among indigenous peoples highlights the need to implement and improve TB surveillance to better measure and understand global disparities in TB rates' [8]. For indigenous minority populations in the USA, the ethnicity data available have revealed that American Indians and Alaskan Natives have a 5·4-fold higher rate of TB compared to non-Hispanic Caucasians. The TB incidence rate in the Aboriginal population is 5·3- and 12·9-fold higher in Australia and Canada, respectively, compared to the domestic-born non-Aboriginal population [9, 10]. In New Zealand, Māori have an approximate sixfold higher rate of TB compared to people of European descent [11, 12]. An earlier study conducted on Roma in Barcelona found that they have a TB incidence rate which is 5·3-fold higher than the national incidence rate for Spain [8]. While one cannot deduce from the above findings alone that Irish Travellers as an indigenous minority experience a higher rate of TB with respect to the general population, they share with indigenous minority populations in other countries a number of determinants of poor health status such as lower levels of employment and education attainment, as well as social exclusion. While studies in the UK have identified higher numbers of Gypsy Travellers (including Irish, English, Scottish and Welsh Gypsies and Travellers) reporting illnesses that include bronchitis and asthma with respect to an age-matched comparator group [13, 14], data with respect to the incidence of TB in Irish Travellers in the UK have not been reported [15].

As a first step in examining the incidence of TB in the Irish Traveller population, we examined TB case notifications in Ireland for the period from 2002 to 2013. Based on the available data, notified cases of TB in Irish Travellers increased from 4·2 cases/100 000 in 2002 to more than 10/100000 from 2011 onwards. In comparison, notified cases of TB in the general population decreased from 10·5 cases/100 000 in 2002 to 8·3/100000 in 2013. The available notification data from 2011 onwards indicate an approximate threefold higher incidence rate of TB in Irish Travellers compared to their white Irish-born compatriots (Fig. 1). This differential is also seen in the respective 5-year CIR of TB for the period 2009–2013 for Irish Travellers (81·4/100000) and the white Irish-born population (27·3/100000) (Table 1). In terms of the age-specific distribution of TB cases, the mean age of cases in Irish Travellers (26 years) was notably lower than that of the general population (mean age of 43 years) and of the white Irish-born population (mean age of 49 years).

Studies in different jurisdictions may provide an insight into why the incidence of TB could be higher in Irish Travellers. Socioeconomic position is known to affect people's health. As noted by Fiscella &

Williams: 'Differences in socioeconomic status, whether measured by income, educational achievement, or occupation, are associated with large disparities in health status. This association persists across the life cycle and across measures of health, including health status, morbidity, and mortality. Although effects are largest for those living in poverty, gradients of disparity are seen across the socioeconomic spectrum' [16]. More recently, Boccia and colleagues established a specific association between household socioeconomic position and TB [17]. It is widely accepted that Irish Travellers experience a lower socioeconomic status as indicated by parameters such as unemployment and education. Unemployment was measured at 84·3% in the Irish Traveller community (percentage of people in the labour force who were either looking for their first job or unemployed) in the Irish Census 2011 [6]. This compared to 14·6% unemployment in the general population in 2011. In terms of education, only 3·1% of Irish Travellers continued their education beyond the age of 18 years in contrast to 41·2% of the general population [6].

A lower socioeconomic status in Irish Travellers may coincide with a higher prevalence of one or more known risk factors for the development of TB. Poor accommodation has previously been cited as a factor in the lower health status of Irish Travellers [1]. According to the Irish Census 2011, the average number of rooms in Irish Traveller households was 4·3 compared to 5·5 rooms for all private households in the state in 2011. Two and a half per cent of Irish Traveller households have $\geqslant 10$ persons compared to 0·04% of non-Irish Traveller households. Therefore, Irish Traveller households have a higher occupant density compared to non-Traveller households. It is worthy of note that in New Zealand, household crowding has been identified as a risk factor for increased TB incidence [18], and that exposure to extreme crowding was high in the indigenous Māori population compared to the European/Other group [19].

In terms of other known risk factors for the development of TB, a study published in 2009 reported that the prevalence of pre-diabetes and diabetes in the Irish Traveller population was about twofold higher than in the background population in Ireland [20]. In the UK, relatively high rates of smoking prevalence have been reported for Irish Traveller males and females [14, 21]. Further research is needed to establish and quantify each of the specific risk factors that underlie a higher incidence of TB in Irish Travellers. This information would be valuable in guiding intervention measures to reduce the risk of TB development in Irish Travellers.

Although this study illustrates differences in the incidence of TB in the Irish Traveller population compared to the general and white Irish populations, a number of limitations of the study should be noted. One limitation pertains to the completeness of the available ethnicity data. The proportion of TB cases in the Irish Traveller population that was correctly identified with respect to ethnicity is not known. A similar limitation was reported recently with respect to an assessment of meningococcal surveillance data in Ireland [22]. The definition of the Irish Traveller community in Irish Law (Equal Status Act, 2000, Part I) is given as follows: '"Traveller community" means the community of people who are commonly called Travellers and who are identified (both by themselves and others) as people with a shared history, culture and traditions including, historically, a nomadic way of life on the island of Ireland.' In the Irish Census 2011, people were asked the question: 'What is your ethnic or cultural background?' Hence, for census purposes, self-declaration of ethnicity is used. In terms of the surveillance of notifiable infectious diseases, it is possible for local departments of public health to record ethnicity for all notifications of notifiable diseases as per the CSO ethnicity categories. However, it is not clear if the practice of self-declaration of Irish Traveller ethnicity is in widespread use with respect to notifiable disease surveillance in Ireland. Incomplete recording of ethnicity could potentially result in an underestimation of the number of TB cases in the Traveller community. National guidance on the standardized collection of ethnicity identifiers has been noted as a possible future project within the HPSC of the HSE, Ireland in conjunction with relevant stakeholders.

Another limitation of the study relates to the relatively small population size of Irish Travellers who constitute 0·64% of the total population in Ireland [6]. An outbreak of TB in the Traveller population can significantly affect the CIR/100000. This is evident in wide confidence intervals that were obtained with respect to CIRs for Irish Traveller cases. In 2013, an outbreak of TB in the Irish Traveller community resulted in a CIR of 40·6/100000 per annum, using the CSO traveller population as a denominator. Due to the impact of this outbreak on the incidence rate, we examined the data excluding cases linked to the outbreak. The incidence rate of TB in Irish Travellers in 2013 was still higher (16·9/100000) than

in the general population (8·3/100000) and in the white Irish-born population (4·2/100000) when Irish Traveller outbreak TB cases were excluded (Fig. 1).

With an increase in notifications of a notifiable disease, there is often the consideration of whether it may be due to enhanced surveillance, or due to a true increase in incidence. An earlier study found that under-notification of TB in different studies in the UK ranged from 7% to 27% of cases [23]. For the whole of the UK, data completeness with respect to known ethnicity of TB cases was 98% in 2013 [15]. The completeness of the ethnicity data for TB cases in Ireland before an increase in TB notifications in Irish Travellers stood at 97·9% in 2009, compared to ethnicity data completeness of 83·5% in 2013 where notifications in Irish Travellers were highest. Therefore, a surge in ethnicity reporting for all TB cases in Ireland would not appear to be responsible for the increase in TB cases in Irish Travellers. This does not preclude the possibility of a differential enhancement of TB surveillance in a given population group such as Irish Travellers. At this stage, it is not possible to conclude whether enhanced surveillance or increased incidence, or both, are responsible for the increase in TB notifications in Irish Travellers. Nevertheless, the data point to a dissimilar incidence rate and age distribution of TB in Irish Travellers with respect to the general and white Irish populations and provide evidence that the monitoring of TB over a longer period in Irish Travellers is warranted.

In terms of ethnicity identifier use for Irish Travellers, the Irish Department of Health and Children's document 'Traveller Health: A National Strategy 2002–2005' made the following points:

> Making significant progress in tackling Traveller health status will be difficult unless an adequate system can be put in place to gather data on an ongoing basis on Traveller health. This data, effectively the baseline from which progress can be measured and by which services can be planned and monitored, is now an urgent necessity. However, for the purpose of effectively gathering information on the health status of the Traveller community, it may be necessary to modify existing health information systems in order to identify Travellers as an ethnic group [24].

The above recommendation has been employed by HPSC with regard to the use of enhanced surveillance forms for TB, invasive meningococcal disease/bacterial meningitis, salmonellosis, which specifically identify Irish Travellers under ethnicity. In addition, the CIDR system, which is used to record all cases of notifiable infectious disease in Ireland, contains a core variable which allows ethnicity to be recorded for every case of notifiable disease reported. These developments, where applied routinely with case investigation, would reduce passive reporting or third-party declaration of Irish Traveller ethnicity in disease surveillance, and consequently, increase the completeness and robustness of Irish Traveller health data.

In summary, it has been reported previously that the population age structure of the Irish Traveller population in Ireland resembles that of a developing country with high mortality rates at a younger age [1]. The extent to which infectious diseases may contribute to this observation deserves investigation. Acknowledging the limitations of the data available, this study constitutes one of the first reports to examine the incidence of the respiratory disease, TB, in the Irish Traveller population. We detected in our study a higher incidence rate of TB in Irish Travellers relative to the general population, and in particular, with respect to the white Irish-born population. With the general decline in TB in the white Irish-born population, TB cases may be concentrated in marginalized groups such as Irish Travellers. Routine recording of Irish Traveller ethnicity for TB, as well as other notifiable infectious diseases, could strengthen the detection of inequities in infection rates and the characterization of their respective determinants. This would in turn provide guidance on appropriate preventive measures that may reduce communicable disease morbidity and mortality in the Irish Traveller population.

ACKNOWLEDGEMENTS

The authors thank Ronnie Fay and Nurul Amin at the Pavee Point Traveller and Roma Centre, Dublin, for their valued discussions.

DECLARATION OF INTEREST

None.

REFERENCES

1. **All Ireland Traveller Health Study Team.** All Ireland Traveller Health Study, 2010 (http://www.ucd.ie/issda/data/allirelandtravellerhealthstudy/).
2. **Office for National Statistics.** 2011 Census for England and Wales.
3. **Baker M.** Gypsies and Travellers – Leeds Baseline Census 2004–2005. 2005.

4. **World Health Organisation.** Global tuberculosis report, 2014.
5. **Health Protection Surveillance Centre.** National TB surveillance, a report by the Health Protection Surveillance Centre, Quarter 1–4, 2013 TB Report, 2014.
6. **Central Statistics Office.** Census of Ireland 2011.
7. **Northern Ireland Statistics and Research Agency.** Census 2011.
8. **Tollefson D, *et al.*** Burden of tuberculosis in indigenous peoples globally: a systematic review. *International Journal of Tuberculosis and Lung Disease* 2013; **17**: 1139–1150.
9. **Barry C, *et al.*** Tuberculosis notifications in Australia, 2008 and 2009. *Communicable Diseases Intelligence Quarterly Report* 2012; **36**: 82–94.
10. **Public Health Agency of Canada.** Tuberculosis in Canada 2008. 2012.
11. **Bissielo A, Lim E, Heffernan H.** Tuberculosis in New Zealand: Annual Report 2011. Institute of Environmental Science and Research Ltd.
12. **Lim E, Heffernan H.** Tuberculosis in New Zealand: Annual Report 2012. Institute of Environmental Science and Research Ltd.
13. **Parry G, *et al.*** The health status of gypsies and travellers in England. University of Sheffield, 2004.
14. **Parry G, *et al.*** Health status of Gypsies and Travellers in England. *Journal of Epidemiology and Community Health* 2007; **61**: 198–204.
15. **Public Health England.** Tuberculosis in the UK – 2014 report. 2014.
16. **Fiscella K, Williams DR.** Health disparities based on socioeconomic inequities: implications for urban health care. *Academic Medicine* 2004; **79**: 1139–1147.
17. **Boccia D, *et al.*** The measurement of household socioeconomic position in tuberculosis prevalence surveys: a sensitivity analysis. *International Journal of Tuberculosis and Lung Disease* 2013; **17**: 39–45.
18. **Baker M, *et al.*** Tuberculosis associated with household crowding in a developed country. *Journal of Epidemiology and Community Health* 2008; **62**: 715–721.
19. **Baker M, *et al.*** Infectious diseases attributable to household crowding in New Zealand: a systematic review and burden of disease estimate. In *He Kainga Orangal Housing and Health Research Programme*, University of Otago, 2013.
20. **Tan S, *et al.*** Traveller health: prevalence of diabetes, pre diabetes and the metabolic syndrome. *Irish Medical Journal* 2009; **102**: 176–178.
21. **Aspinall PJ, Mitton L.** Smoking prevalence and the changing risk profiles in the UK ethnic and migrant minority populations: implications for stop smoking services. *Public Health* 2014; **128**: 297–306.
22. **Cotter S, *et al.*** Meningococcal disease in Ireland – can determinants of increased risk be identified? *Epi-Insight* 2014; **15**: 1–5.
23. **Pillaye J, Clarke A.** An evaluation of completeness of tuberculosis notification in the United Kingdom. *BMC Public Health* 2003; **3**: 31.
24. **Department of Health and Children.** Traveller health: a national strategy 2002–2005. 2002.

Glossary

- **Basic reproduction number** - The basic reproduction number (R_0) is defined as the average number of secondary cases caused by a single infectious individual in a totally susceptible population.
- **Exposed** - The term 'exposed' is used when an individual has encountered a disease causative pathogen. This is necessary for infection or transmission to take place. However, it is not necessarily the case that infection or transmission occurs.
- **Foreign-born** - Not native-born.
- **Immunity** - Immunity refers to an individual's resistance to infection or re-infection by a causative pathogen.
- **Incidence** - Incidence refers to the number of new cases of a disease over a period of time.
- **Infected** - The term 'infected' refers to an individual who has contracted a disease causative agent and infection (or transmission) has occurred.
- **Infectious** - Individuals who are infected and can transmit a pathogen (the cause of an infection) to other individuals.
- **Latent period** - The latent period is defined as the period of time between the occurrence of infection and the onset of infectiousness (when the infected individual becomes infectious).

- **Native-born** - An individual who was born to a country, and resides within that country.
- **Prevalence** - Prevalence is defined as the number of cases of a disease at a specific time point
- **Recovered** - Recovery refers to a transitional stage from the infectious state to another non-infectious state.
- **Susceptible** - Susceptible refers to a non-infected individual (or population) who may become infected through contact with individuals or environmental organisms that can transmit the disease
- **Vaccine efficacy** - Vaccine efficacy refers to the percentage reduction in the attack rate of unvaccinated and vaccinated cohorts as observed in a randomized control trial.

Code For Metropolis-Hastings Algorithm

```
library(deSolve)
MH_Seasonal = function(N=10000){
theta = array(dim=c(N+1,2))
theta2= array(dim=c(N,2))
error = array(dim=c(N))
error2 = array(dim=c(N))

init = c(S=1382374, E=240, I = 34, R = 2535000)

times = 1:144
nam1 = c(beta0 = .5,k0=.005,A=7267,u=0.000549905, q=.05,
r=.016 ,d=.01667, w1=0.35)

true = ode(y = init, times = times, func = seirs, parms = nam1)
tru=c(34,31,35,35,24,37,47,30,32,35,32,38,28,30,30,37,36,38,
40,27,37,40,37,26,32,34,39,46,43,36,29,36,31,36,33,36,
32,38,33,39,47,43,31,36,35,40,31,43,27,45,40,44,45,46,
41,38,28,43,33,31,32,25,46,70,44,45,48,42,30,37,26,31,
43,45,42,53,39,32,43,39,33,38,36,24,40,29,45,48,25,60,
56,31,35,36,37,37,30,39,34,40,29,54,36,42,20,28,42,26,
```

```
25,54,43,27,30,49,37,27,27,32,35,27,27,34,30,34,42,26,
27,29,22,33,26,29,35,29,28,45,30,37,46,41,21,33,19,17)

sigma02=.01
n0=0.001
aval=0.5*(n0*sigma02+length(times))
theta[1,] = nam1[1:2]
sigma=array(dim=c(N+1))

sigma[1]=sum((tru-ode(y = init, times = times, func = seirs,
parms = c(beta0=theta[1,1],k0=theta[1,2],A=7267,u=0.000549905,
q=.05, r=.016 ,d=.01667, w1=0.35))[,4])^2)/length(times)

modd=array(dim=c(N+1, 144))

for (i in 1:N){

modd[i,] = ode(y = init, times = times, func = seirs, parms =
c(beta0=theta[i,1],k0=theta[i,2],A=7267,u=0.000549905, q=.05,
r=.016 ,d=.01667,w1=0.35))[,4]

error[i]=sum((tru-modd[i,])^2)

theta2[i,1]=max(min(theta[i,1]+
runif(1,min=-0.01,max=0.01),1),0)

theta2[i,2]=max(min(theta[i,2]+
runif(1,min=-0.001,max=0.001),1),0)

error2[i]=sum((tru-ode(y = init, times = times, func = seirs,
parms = c(beta0=theta2[i,1],k0=theta2[i,2],A=7267,u=0.0005499,
```

```
q=.05,  r=.016  ,d=.01667, w1=0.35))[,4])^2)

theta[(i+1),] = theta[i,]
if(log(runif(1)) < -0.5*(error2[i] - error[i])/sigma[i]){
theta[(i+1),] = theta2[i,]
error[i]=error2[i]
}
bval=0.5*(n0*sigma02+error[i])
sigma[i+1]=1/rgamma(1,aval,bval)
print(100*round((i)/N,2))
}
return(theta)
}

MH_Foreign_Local = function(N=10000){
N=10000
theta = array(dim=c(N+1,2))
theta2= array(dim=c(N,2))
error = array(dim=c(N))
error2 = array(dim=c(N))

init = c(SL=1280078, EL=222, IL = 27, RL = 2347409)
times = 1:(144)
nam1 = c(w1=0.35,b2=0.03065148,k2=0.005630639,
r2=0.01515,uil=.02221,A=4770, u=0.00055)

inf2=c(7,7,6,6,8,7,14,7,5,13,7,12,4,4,3,2,5,2,10,4,
10,5,4,4,4,6,9,7,12,7,7,7,6,8,10,7,6,13,8,10,8,12,
7,12,4,9,6,10,7,11,7,9,7,15,11,13,6,14,11,6,10,7,
15,15,18,14,12,12,11,10,9,8,15,12,11,21,12,15,19,
17,9,12,11,10,12,8,14,18,10,20,16,8,12,13,12,14,
```

```
10,9,12,15,11,21,9,13,3,11,15,9,9,20,15,15,14,20,
12,11,12,11,9,14,8,9,6,15,19,8,10,11,9,11,8,10,14,
8,11,15,13,15,16,16,9,13,5,5)

inf=c(34,31,35,35,24,37,47,30,32,35,32,38,28,30,30,
37,36,38,40,27,37,40,37,26,32,34,39,46,43,36,29,36,
31,36,33,36,32,38,33,39,47,43,31,36,35,40,31,43,27,
45,40,44,45,46,41,38,28,43,33,31,32,25,46,70,44,45,
48,42,30,37,26,31,43,45,42,53,39,32,43,39,33,38,36,
24,40,29,45,48,25,60,56,31,35,36,37,37,30,39,34,40,
29,54,36,42,20,28,42,26,25,54,43,27,30,49,37,27,27,
32,35,27,27,34,30,34,42,26,27,29,22,33,26,29,35,29,
28,45,30,37,46,41,21,33,19,17)

tru=inf-inf2

sigma02=.01
n0=0.001
aval=0.5*(n0*sigma02+length(times))
theta[1,] = c(nam1[2],nam1[3])
sigma=array(dim=c(N+1))

sigma[1]=sum((tru-ode(y = init, times = times, func = seirf2,
parms = c(w1=0.35,b2=theta[1,1],k2=theta[1,2],r2=0.01515,
uil=.02221,A=4770,u=0.00055) )[,4])^2)/length(times)

for (i in 1:N){
error[i]=sum((tru-ode(y = init, times = times, func = seirf2,
parms = c(w1=0.35,b2=theta[i,1],k2=theta[i,2],r2=0.01515,
uil=.02221,A=4770,u=0.00055) )[,4])^2)
```

```
theta2[i,1]=max(min(theta[i,1]+
runif(1,min=-0.01,max=0.01),1),0.000001)

theta2[i,2]=max(min(theta[i,2]+
runif(1,min=-0.001,max=0.001),1),0.000001)

error2[i]=sum((tru-ode(y = init, times = times, func = seirf2,
parms = c(w1=0.35,b2=theta2[i,1],k2=theta2[i,2],r2=0.01515,
ui1=.02221,A=4770,u=0.00055) )[,4])^2)

theta[(i+1),] = theta[i,]
if(log(runif(1)) < -0.5*(error2[i] - error[i])/sigma[i]){
theta[(i+1),] = theta2[i,]
error[i]=error2[i]
}
bval=0.5*(n0*sigma02+error[i])
sigma[i+1]=1/rgamma(1,aval,bval)
print(100*round((i)/N,2))
}
return(theta)
}

MH_Foreign_Mig = function(N=10000){
theta = array(dim=c(N+1,2))
theta2= array(dim=c(N,2))
error = array(dim=c(N))
error2 = array(dim=c(N))

init = c(SM=239603, EM=58, IM = 7, RM = 50136)
times = 1:144
```

```
nam1 = c(v1=0.00002,v2=0.173,pp=2497,b1=0.0514155,
u=0.00055,k1=0.0048265,r1=.01695,uim=0.00872)

tru=c(7,7,6,6,8,7,14,7,5,13,7,12,4,4,3,2,5,2,10,4,
10,5,4,4,4,6,9,7,12,7,7,7,6,8,10,7,6,13,8,10,8,12,
7,12,4,9,6,10,7,11,7,9,7,15,11,13,6,14,11,6,10,7,
15,15,18,14,12,12,11,10,9,8,15,12,11,21,12,15,19,
17,9,12,11,10,12,8,14,18,10,20,16,8,12,13,12,14,10,
9,12,15,11,21,9,13,3,11,15,9,9,20,15,15,14,20,12,11,
12,11,9,14,8,9,6,15,19,8,10,11,9,11,8,10,14,8,11,15,
13,15,16,16,9,13,5,5)



sigma02=.01
n0=0.001
aval=0.5*(n0*sigma02+length(times))
theta[1,] = c(nam1[4],nam1[6])
sigma=array(dim=c(N+1))
sigma[1]=sum((tru-ode(y = init, times = times, func = seirf1,
parms = c(v1=0.00002,v2=0.173,b1=theta[1,1],k1=theta[1,2],
pp=2497,u=0.00055,r1=.01695,uim=0.00872))[,4])^2)/length(times)

for (i in 1:N){

error[i]=sum((tru-ode(y = init, times = times, func = seirf1,
parms = c(v1=0.00002,v2=0.173, b1=theta[i,1],k1=theta[i,2],
pp=2497,u=0.00055,r1=.01695,uim=0.00872))[,4])^2)

theta2[i,1]=max(min(theta[i,1]+
runif(1,min=-0.05,max=0.05),1),0.000001)
```

```
theta2[i,2]=max(min(theta[i,2]+
runif(1,min=-0.005,max=0.005),1),0.000001)

error2[i]=sum((tru-ode(y = init, times = times, func = seirf1,
parms = c(v1=0.00002,v2=0.173, b1=theta2[i,1],k1=theta2[i,2],
pp=2497,u=0.00055,r1=.01695,uim=0.00872))[,4])^2)

theta[(i+1),] = theta[i,]
if(log(runif(1)) < -0.5*(error2[i] - error[i])/sigma[i]){
theta[(i+1),] = theta2[i,]
error[i]=error2[i]
}
bval=0.5*(n0*sigma02+error[i])
sigma[i+1]=1/rgamma(1,aval,bval)
print(100*round((i)/N,2))
}
return(theta)
}


MH_Foreign_Int = function(N=10000){
theta = array(dim=c(N+1,6))
theta2= array(dim=c(N,6))
theta2a= array(dim=c(N,6))

error = array(dim=c(N))
error2 = array(dim=c(N))
errora = array(dim=c(N))
error2a = array(dim=c(N))

init = c(SM=239603, EM=58, IM = 7, RM = 50136,
```

```
SL=1280078, EL=222, IL = 27, RL = 2347409)
times = 1:144
nam1 = c(v1=0.00002,v2=0.173,pp=2497,b1=0.02814503,
u=0.00055,k1=0.004931357,r1=.01695,uim=0.00872,
w1=0.35,b2=0.02520684,k2= 0.005589813,r2=0.01515,
uil=.02221,A=4770,bs1=0.006560251,bs2=0.009016213)

trum=c(7,7,6,6,8,7,14,7,5,13,7,12,4,4,3,2,
5,2,10,4,10,5,4,4,4,6,9,7,12,7,7,7,
6,8,10,7,6,13,8,10,8,12,7,12,4,9,6,
10,7,11,7,9,7,15,11,13,6,14,11,6,10,
7,15,15,18,14,12,12,11,10,9,8,15,12,
11,21,12,15,19,17,9,12,11,10,12,8,14,
18,10,20,16,8,12,13,12,14,10,9,12,15,
11,21,9,13,3,11,15,9,9,20,15,15,14,20,
12,11,12,11,9,14,8,9,6,15,19,8,10,11,
9,11,8,10,14,8,11,15,13,15,16,16,9,13,5,5)

inf=c(34,31,35,35,24,37,47,30,32,35,32,38,28,
30,30,37,36,38,40,27,37,40,37,26,32,34,39,46,43,
36,29,36,31,36,33,36,32,38,33,39,47,43,31,36,35,
40,31,43,27,45,40,44,45,46,41,38,28,43,33,31,32,
25,46,70,44,45,48,42,30,37,26,31,43,45,42,53,39,
32,43,39,33,38,36,24,40,29,45,48,25,60,56,31,35,
36,37,37,30,39,34,40,29,54,36,42,20,28,42,26,25,
54,43,27,30,49,37,27,27,32,35,27,27,34,30,34,42,
26,27,29,22,33,26,29,35,29,28,45,30,37,46,41,21,
33,19,17)
trul=inf-trum

sigma02=.01
```

```
n0=0.001
aval=0.5*(n0*sigma02+length(times))


theta[1,] = c(nam1[4],nam1[6],nam1[10],nam1[11],
nam1[15],nam1[16])

sigma=array(dim=c(N+1))
sigmaa=array(dim=c(N+1))

sigma[1]=sum((trum-ode(y = init, times = times, func = seirfm,
parms =c(v1=0.00002,v2=0.173,pp=2497,b1=theta[1,1],u=0.00055,
k1=theta[1,2],r1=.01695,uim=0.00872,w1=0.35,b2=theta[1,3],
k2=theta[1,4],r2=0.01515,uil=.02221,A=4770,bs1=theta[1,5],
bs2=theta[1,6],p1=.8,p2=.8))[,4])^2)/length(times)

sigmaa[1]=sum((trum-ode(y = init, times = times, func = seirfm,
parms = c(v1=0.00002,v2=0.173,pp=2497,b1=theta[1,1],u=0.00055,
k1=theta[1,2], r1=.01695,uim=0.00872,w1=0.35,b2=theta[1,3],
k2=theta[1,4],r2=0.01515,uil=.02221,A=4770,bs1=theta[1,5],
bs2=theta[1,6],p1=.8,p2=.8))[,4])^2)/length(times)

for (i in 1:N){

error[i]=sum((trum-ode(y = init, times = times, func = seirfm,
parms =c(v1=0.00002,v2=0.173,pp=2497,b1=theta[i,1],u=0.00055,
k1=theta[i,2],r1=.01695,uim=0.00872,w1=0.35,b2=theta[i,3],
k2=theta[i,4],r2=0.01515,uil=.02221,A=4770,bs1=theta[i,5],
bs2=theta[i,6],p1=.8,p2=.8))[,4])^2)

theta2[i,1]=max(min(theta[i,1]+
```

```
runif(1,min=-0.005,max=0.005),1),0.000001)

theta2[i,2]=max(min(theta[i,2]
+runif(1,min=-0.0005,max=0.0005),1),0.000001)

theta2[i,6]=max(min(theta[i,6]+
runif(1,min=-0.0005,max=0.0005),1),0.000001)

theta2[i,3]=theta[i,3]
theta2[i,4]=theta[i,4]
theta2[i,5]=theta[i,5]

error2[i]=sum((trum-ode(y = init, times = times, func = seirfm,
parms =c(v1=0.00002,v2=0.173,pp=2497,b1=theta2[i,1],u=0.00055,
k1=theta2[i,2],r1=.01695,uim=0.00872,w1=0.35,b2=theta2[i,3],
k2=theta2[i,4],r2=0.01515,uil=.02221,A=4770,bs1=theta2[i,5],
bs2=theta2[i,6],p1=.8,p2=.8))[,4])^2)

theta[(i+1),] = theta[i,]
if(log(runif(1)) < -0.5*(error2[i] - error[i])/sigma[i]){
theta[(i+1),] = theta2[i,]
error[i]=error2[i]
}
bval=0.5*(n0*sigma02+error[i])
sigma[i+1]=1/rgamma(1,aval,bval)
print(100*round((i)/N,2))
}
for (i in 1:N){

errora[i]=sum((trul-ode(y = init, times = times, func = seirfm,
parms =c(v1=0.00002,v2=0.173,pp=2497,b1=theta[(i+1),1],
```

```
u=0.00055,k1=theta[(i+1),2],r1=.01695,
uim=0.00872,w1=0.35, b2=theta[(i+1),3],
k2=theta[(i+1),4],r2=0.01515,
uil=.02221,A=4770,bs1=theta[(i+1),5],bs2=theta[(i+1),6],
p1=.8,p2=.8))[,8])^2)

theta2a[(i),3]=max(min(theta[(i+1),3]+
runif(1,min=-0.005,max=0.005),1),0.000001)
theta2a[(i),4]=max(min(theta[(i+1),4]
+runif(1,min=-0.0005,max=0.0005),1),0.000001)
theta2a[(i),5]=max(min(theta[(i+1),5]
+runif(1,min=-0.0005,max=0.0005),1),0.000001)

theta2a[(i),1]=theta[(i+1),1]
theta2a[(i),2]=theta[(i+1),2]
theta2a[(i),6]=theta[(i+1),6]

error2a[i]=sum((trul-ode(y=init,times=times,func=seirfm,
parms=c(v1=0.00002,v2=0.173,pp=2497,b1=theta2a[i,1],u=0.00055,
k1=theta2a[i,2],r1=.01695,uim=0.00872,w1=0.35,b2=theta2a[i,3],
k2=theta2[i,4],r2=0.01515,uil=.02221,A=4770,bs1=theta2a[i,5],
bs2=theta2a[i,6],p1=.8,p2=.8))[,8])^2)

theta[(i+1),3] = theta[i,3]
theta[(i+1),4] = theta[i,4]
theta[(i+1),5] = theta[i,5]
if(log(runif(1)) < -0.5*(error2a[i] - errora[i])/sigmaa[i]){
theta[(i+1),3] = theta2a[i,3]
theta[(i+1),4] = theta2a[i,4]
theta[(i+1),5] = theta2a[i,5]
errora[i]=error2a[i]
```

```
}
bvala=0.5*(n0*sigma02+errora[i])
sigmaa[i+1]=1/rgamma(1,aval,bvala)
print(100*round((i)/N,2))

}
return(theta)
}


seirs <- function(time, state, parameters) {

with(as.list(c(state, parameters)), {

N= S+E+I+R
beta= beta0*(1+sin((2*pi*time)/(12)))
k= k0*(1+sin((2*pi*time)/(12)))

dS= w1*A-beta*S*I/N-u*S
dE= (1-q)*beta*S*I/N - k*E - u*E
dI= (q)*beta*S*I/N+k*E - r*I - (u+d)*I
dR= (1-w1)*A + r*I - u*R
return(list(c(dS,dE,dI,dR)))
})
}
seirfm <- function(time, state, parameters) {

with(as.list(c(state, parameters)), {
NM= SM+EM+IM+RM
NL= SL+EL+IL+RL
```

```
dSM= (1-v1-v2)*pp-b1*SM*IM/NM-bs2*SM*IL/NM-u*SM

dEM= v1*pp+b1*SM*IM/NM + bs2*SM*IL/NM
-p2*bs2*IL*EM/NM  - k1*EM - u*EM

dIM= k1*EM +p2*bs2*EM*IL/NM - (r1+u+uim)*IM
dRM= v2*pp + r1*IM - u*RM

dSL= w1*A-b2*SL*IL/NL - bs1*SL*IM/NL-u*SL
dEL= b2*SL*IL/NL +bs1*SL*IM/NL-p1*bs1*IM*EL/NL- k2*EL - u*EL
dIL= p1*bs1*IM*EL/NL+k2*EL - (r2+u+uil)*IL
dRL= (1-w1)*A + r2*IL - u*RL

return(list(c(dSM,dEM,dIM,dRM,dSL,dEL,dIL,dRL)))
})
}

seirf2 <- function(time, state, parameters) {

with(as.list(c(state, parameters)), {
NL= SL+EL+IL+RL

dSL= w1*A-b2*SL*IL/NL-u*SL
dEL= b2*SL*IL/NL - k2*EL - u*EL
dIL= k2*EL - (r2+u+uil)*IL
dRL= (1-w1)*A + r2*IL - u*RL


return(list(c(dSL,dEL,dIL,dRL)))
})
}
```

```
seirf1 <- function(time, state, parameters) {

with(as.list(c(state, parameters)), {
NM= SM+EM+IM+RM

dSM= (1-v1-v2)*pp-b1*SM*IM/NM-u*SM
dEM= v1*pp+b1*SM*IM/NM - k1*EM - u*EM
dIM= k1*EM - (r1+u+uim)*IM
dRM= v2*pp + r1*IM - u*RM


return(list(c(dSM,dEM,dIM,dRM)))
})
}
```

Bibliography

[1] Pringle, D. (2009). The resurgence of tuberculosis in the Republic of Ireland: Perceptions and reality. *Social Science & Medicine* **68** (4), 620-624.

[2] Health Service Executive. *Reports on the Epidemiology of TB in Ireland*. Retrieved August 24, 2016, from `http://www.hpsc.ie/A-Z/VaccinePreventable/TuberculosisTB/Epidemiology/AnnualReports/`.

[3] T. Paulson (2013), Epidemiology: A mortal foe, *Nature* **502**, S2–S3.

[4] Hayman J. (1984), Mycobacterium ulcerans: an infection from Jurassic time? *Lancet* **2** ,1015–6.

[5] Karlen, A. (1995). *Plagues progress, a social history of man and disease*. Londres, Grã-Bretanha.

[6] Galagan, J. (2014). Genomic insights into tuberculosis. *Nature Reviews*, **15**, 307-320.

[7] Hirsh, A. E., Tsolaki, A. G., DeRiemer, K., Feldman, M. W., & Small, P. M. (2004). Stable association between strains of Mycobacterium tuberculosis and their human host populations. *Proceedings of the National Academy of Sciences of the United States of America*, **101** (14), 4871-4876.

[8] Schoenlein, J. L. (1839). Zur pathogenie der impetigines. *Arch. Anat. Physiol. Wiss. Med*, **82**, 1839.

[9] Nobel Foundation. *The Nobel Prize in Physiology or Medicine 1905*. Accessed 01 April 2016.

[10] Waddington K (2004). To stamp out "So Terrible a Malady": bovine tuberculosis and tuberculin testing in Britain, 1890–1939. *Med Hist* **48** (1): 29–48.

[11] Bonah C (2005). The 'experimental stable' of the BCG vaccine: safety, efficacy, proof, and standards, 1921–1933. *Stud Hist Philos Biol Biomed Sci* **36** (4): 696–721.

[12] Comstock G (1994). The International Tuberculosis Campaign: a pioneering venture in mass vaccination and research. *Clin Infect Dis* **19** (3): 528–40.

[13] World Health Organization (2011). *Bugs, drugs and smoke: stories from public health*. Geneva: World Health Organization. p 99-117.

[14] Kilpatrick, D. (2013). *Saving Lives and Preventing Misery - The memoirs of Professor Sir John Wenman Crofton*. Peterborough UK, (ISBN 978-178035-541-2).

[15] Shields, T. W. (Ed.). (2005). General thoracic surgery (Vol. 1). Lippincott Williams & Wilkins.

[16] Lawn SD, Zumla AI (2011), Tuberculosis. *The Lancet* **378** (9785) 57 - 72 doi:10.1016/S0140-6736 (10)62173-3.

[17] Jindal, S. K., Shankar, P. S., Raoof, S., & Gupta, D. (2011). *Textbook of Pulmonary and Critical Care Medicine* **1**, **2**. JP Medical Ltd.

[18] Selwyn, P. A., Hartel, D., Lewis, V. A., Schoenbaum, E. E., Vermund, S. H., Klein, R. S., ... & Friedland, G. H. (1989). A prospective study of the risk of tuberculosis among intravenous drug users with human immunodeficiency virus infection. *New England journal of medicine*, **320** (9), 545-550. Chicago.

[19] Chandra, R. K. (1997). Nutrition and the immune system: an introduction. *The American journal of clinical nutrition*, **66** (2), 460S-463S.

[20] Chandra, R. K., & Kumari, S. (1994). Nutrition and immunity: an overview. *The Journal of nutrition*, **124** (8 Suppl), 1433S-1435S.

[21] Sinclair, D., Abba, K., Grobler, L., & Sudarsanam, T. D. (2011). *Nutritional supplements for people being treated for active tuberculosis*. The Cochrane Library.

[22] Brailey, M. (1940). A Study of Tuberculous Infection and Mortality in the Children of Tuberculous Households. *American Journal of Hygiene*, **31**, 1-43. Chicago.

[23] Alisjahbana B, van Crevel R, Sahiratmadja E, den Heijer M, Maya A, et al. (2006) Diabetes mellitus is strongly associated with tuberculosis in Indonesia. *Int J Tuberc Lung Dis* **10** 696–700.

[24] Blas, E., & Kurup, A. S. (2010). *Equity, social determinants and public health programmes*. World Health Organization.

[25] Lönnroth, K., Williams, B. G., Stadlin, S., Jaramillo, E., & Dye, C. (2008). Alcohol use as a risk factor for tuberculosis–a systematic review. *BMC public health*, **8** (1), 1.

[26] Bates, M. N., Khalakdina, A., Pai, M., Chang, L., Lessa, F., & Smith, K. R. (2007). Risk of tuberculosis from exposure to tobacco smoke: a systematic review and meta-analysis. *Archives of internal medicine*, **167** (4), 335-342.

[27] Pareek, M., Greenaway, C., Noori, T., Munoz, J., & Zenner, D. (2016). The impact of migration on tuberculosis epidemiology and control in high-income countries: a review. BMC medicine, 14(1), 48.

[28] Pareek M, Watson JP, Ormerod LP, et al. Screening of immigrants in the UK for imported latent tuberculosis: a multicentre cohort study and costeffectiveness analysis. Lancet Infect Dis. 2011;11:435–44.

[29] Pareek M, Bond M, Shorey J, et al. Community-based evaluation of immigrant tuberculosis screening using interferon γ release assays and tuberculin skin testing: observational study and economic analysis. Thorax. 2013, 68:230-9.

[30] Joshi, R., Reingold, A. L., Menzies, D., & Pai, M. (2006). Tuberculosis among health-care workers in low-and middle-income countries: a systematic review. *PLoS Med*, **3** (12), e494.

[31] World Health Organization. (2011). *Early detection of tuberculosis: an overview of approaches, guidelines and tools*.

[32] Bento, J., Silva, A. S., Rodrigues, F., & Duarte, R. (2011). Diagnostic tools in tuberculosis. *Acta medica portuguesa*, **24** (1), 145-154.

[33] National Collaborating Centre for Chronic Conditions (UK, & Centre for Clinical Practice at NICE (UK.)) (2011). *Tuberculosis: clinical diagnosis and management of tuberculosis, and measures for its prevention and control*.

[34] Centre for Disease Control and Prevention (2003). *Mantoux Tuberculin Skin Test*. [ONLINE] Available at: `http://www.cdc.gov/TB/education/Mantoux/images/mantoux.pdf`. Accessed: 21 September 2016.

[35] Nayak, S., & Acharjya, B. (2012). Mantoux test and its interpretation. *Indian dermatology online journal*, **3** (1), 2.

[36] Drummond, M. F., Sculpher, M. J., Claxton, K., Stoddart, G. L., & Torrance, G. W. (2015). *Methods for the economic evaluation of health care programmes*. Oxford university press.

[37] Colditz, G. A., Brewer, T. F., Berkey, C. S., Wilson, M. E., Burdick, E., Fineberg, H. V., & Mosteller, F. (1994). Efficacy of BCG vaccine in the prevention of tuberculosis: meta-analysis of the published literature. *Jama*, **271** (9), 698-702. Chicago.

[38] Colditz, Graham A.; Brewer, TF; Berkey, CS; Wilson, ME; Burdick, E; Fineberg, HV; Mosteller, F (1994). Efficacy of BCG Vaccine in the Prevention of Tuberculosis. *JAMA* **271** (9): 698–702.

[39] Fine PEM (1995). Variation in protection by BCG: implications of and for heterologous immunity. *Lancet* **346** (8986): 1339–45.

[40] Centers for Disease Control and Prevention. (1996). The role of BCG vaccine in the prevention and control of tuberculosis in the United States: a joint statement by the Advisory Council for the Elimination of Tuberculosis and the Advisory Committee on Immunization Practices. *MMWR Morb Mortal Wkly Rep*, **45** (1).

[41] Abubakar, I., Pimpin, L., Ariti, C., Beynon, R., Mangtani, P., Sterne, J. A. C., ... & Watson, J. M. (2013). *Systematic review and meta-analysis of the current evidence on the duration of protection by bacillus Calmette–Guerin vaccination against tuberculosis*.

[42] Aronson NE, Santosham M, Comstock GW (2004). Long-term efficacy of BCG vaccine in American Indians and Alaska Natives: A 60-year follow-up study. *JAMA*. **291** (17): 2086–91.

[43] World Health Organization. (2015). Global Tuberculosis Report 2015. Geneva, World Health Organization 2015.

[44] World Health Organization. (2016). *Tuberculosis*. Fact sheet 2015 N 104. Updated October 2015.

[45] Kumar, V., Abbas, A. K., & Aster, J. C. (2012). Robbins basic pathology. Elsevier Health Sciences.

[46] Griffin, J. T. (2015). The interaction between seasonality and pulsed interventions against malaria in their effects on the reproduction number. *PLoS Comput Biol* **11** (1), e1004057.

[47] Marjoram, P., Molitor, J., Plagnol, V., & Tavaré, S. (2003). Markov chain Monte Carlo without likelihoods. *Proceedings of the National Academy of Sciences* **100** (26), 15324-15328.

[48] Zumla, A., Mwaba, P., Huggett, J., Kapata, N., Chanda, D., & Grange, J. (2009). Reflections on the white plague. *The Lancet Infectious Diseases*, **9** (3), 197-202.

[49] Emergency, T. A. G. (1994). WHO Report on the TB Epidemic. *WH0 TB*/**94**, 177.

[50] World Health Organisation (2015), *Global strategy and targets for tuberculosis prevention, care and control beyond 2015*, Available online: http://www.who.int/entity/tb/post2015_TBstrategy.pdf?ua=1

[51] Surveillance of tuberculosis in Europe (2005). Report on tuberculosis cases notified in 2005. Available online: http://www.euroTB.org (Accessed 18 April 2016).

[52] Dara, M., Kremer, K., Huitric, E., Ködmön, C., & Zucs, P. (2012). *Tuberculosis surveillance and monitoring in Europe 2012. Tuberculosis surveillance and monitoring in Europe 2012*.

[53] Hayward, A. C., Darton, T., Van-Tam, J. N., Watson, J. M., Coker, R., & Schwoebel, V. (2003). Epidemiology and control of tuberculosis in Western European cities. *The international journal of tuberculosis and lung disease*, **7** (8), 751-757.

[54] Merrill, R. M. (2015). *Introduction to epidemiology*. Jones & Bartlett Publishers.

[55] Vynnycky, E., & White, R. (2010). *An introduction to infectious disease modelling*. Oxford University Press.

[56] Roberts, M., & Heesterbeek, H. (1993). Bluff your way in epidemic models. *Trends in microbiology*, **1** (9), 343-348.

[57] Coussens, A., Timms, P. M., Boucher, B. J., Venton, T. R., Ashcroft, A. T., Skolimowska, K. H., ... & Wilkinson, R. J. (2009). 1α, 25-dihydroxyvitamin D3 inhibits matrix metalloproteinases induced by Mycobacterium tuberculosis infection. *Immunology* **127** (4), 539-548.

[58] Barnes, B., & Fulford, G. R. (2014). *Mathematical Modelling with Case Studies: Using Maple and MATLAB* (Vol. 25). CRC Press.

[59] DServe archive persons - John Graunt, Available at: https://collections.royalsociety.org/DServe.exe?dsqIni=Dserve.ini\&dsqApp=Archive\&dsqCmd=Show.tcl\&dsqDb=Persons\&dsqPos=2\&dsqSearch=%28%28text%29%3D%27john%20graunt%27%29 (Accessed: 22 September 2016).

[60] Bailey, N. T. (1975). *The mathematical theory of infectious diseases and its applications*. Charles Griffin & Company Ltd, 5a Crendon Street, High Wycombe, Bucks HP13 6LE..

[61] Luque Fernandez, M. A., Schomaker, M., Mason, P. R., Fesselet, J. F., Baudot, Y., Boulle, A., & Maes, P. (2012). Elevation and cholera: an epidemiological spatial analysis of the cholera epidemic in Harare, Zimbabwe, 2008-2009. *BMC Public Health*, **12**, 442. http://doi.org/10.1186/1471-2458-12-442

[62] Ross SR. (1911) The prevention of malaria. New York, NY: Dutton

[63] Hamer W.H (1906) Epidemic disease in England: the evidence of variability and the persistence of type. *Lancet*. **167**, 733–739. doi:10.1016/S0140-6736(01)80340-8.

[64] Kermack, W. O., & McKendrick, A. G. (1927, August). A contribution to the mathematical theory of epidemics. *In Proceedings of the Royal Society of London A: mathematical, physical and engineering sciences* (Vol. **115**, No. 772, pp. 700-721). The Royal Society.

[65] Hethcote, H. W. (1976). Qualitative analyses of communicable disease models. *Mathematical Biosciences*, **28** (3), 335-356.

[66] Waltman, P. (2013). *Deterministic threshold models in the theory of epidemics* (Vol. **1**). Springer Science & Business Media.

[67] Anderson, R. M., May, R. M., & Anderson, B. (1992). *Infectious diseases of humans: dynamics and control* (Vol. **28**). Oxford: Oxford university press.

[68] Smith DL, Battle KE, Hay SI, Barker CM, Scott TW, McKenzie FE (2012) Ross, Macdonald, and a Theory for the Dynamics and Control of Mosquito-Transmitted Pathogens. *PLoS Pathog* **8** (4): e1002588. doi:10.1371/journal.ppat.1002588

[69] H. T. Waaler, A. Gese, & S. Anderson (1962) The use of mathematical models in the study of the epidemiology of tuberculosis, *Am. J. Publ. Health*, **52** 1002–1013.

[70] Flynn, J.L., and J. Chan. (2001). Tuberculosis: latency and reactivation. *Infect. Immun.* **69** 4195–4201.

[71] S. Brogger (1967). Systems analysis in tuberculosis control: A model, *Amer. Rev. Resp.Dis.*, **95** 419–434.

[72] C. S. ReVelle, W. R. Lynn, and F. Feldmann (1967). Mathematical models for the economic allocation of tuberculosis control activities in developing nations, *Am. Rev. Respir. Dis.*, **96** 893–909.

[73] ReVelle, C. S. (1967). *The economic allocation of tuberculosis control activities in developing nations*. Cornell University, June.

[74] Aparicio JP, Castillo-Chavez C (2009) Mathematical modelling of tuberculosis epidemics. Math Biosci Eng 6:209–237.

[75] Castillo-Chavez, C., & Song, B. (2004). Dynamical models of tuberculosis and their applications. *Mathematical biosciences and engineering*, **1**(2), 361-404.

[76] Feng, Z., Castillo-Chavez, C., & Capurro, A. F. (2000). A model for tuberculosis with exogenous reinfection. *Theoretical population biology*, **57** (3), 235-247.

[77] Jones, J. H. (2007). *Notes on R0*. Califonia: Department of Anthropological Sciences.

[78] Brauer, F. (2012). *Deterministic compartmental disease transmission models*.

[79] Al-Sheikh, S. A., (2012) Modeling and analysis of an seir epidemic model with a limited resource for treatment. *Global Journal of Science Frontier Research Mathematics and Decision Sciences* **12** (14).

[80] Zhang, J., Li, J., Ma, Z. (2006). Global dynamics of an seir epidemic model with immigration of different compartments. *Acta Mathematica Scientia* **26** 551-567

[81] Li, M. Y., Graef, J. R., Wang, L., Karsai, J. (2010). Global dynamics of a seir model with varying total population size. *Mathematical Biosciences, elsevier* **160** 191-213.

[82] Yi, N., Zhang, Q., Mao, K., Yang, D., , Li, Q. (2009). Analysis and control of an seir epidemic system with non-linear transmission rate. *Mathematical and computer modelling* **50** 1498-1513.

[83] Shu, H., Fan, D., Global, J. W. (2012). Stability of multi-group seir epidemic models with distributed delays and nonlinear transmission. *Nonlinear Analysis: Real World Applications* **13** (4), 1581- 1592.

[84] Maliki, S. O. (2011). Analysis of Numerical and Exact solutions of certain SIR and SIS Epidemic models. *Journal of Mathematical Modelling and Application*, **1** (4), 51-56.

[85] Chasnov, J. R. (2009). *Introduction to Differential Equations*. The Hong Kong University of Science and Technology, Department of mathematics, 2012.

[86] Van den Driessche P., Watmough J. (2002) Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. *Math. Biosci.* **180**, 29–48.

[87] Cantwell MF, Snider DE Jr, Cauthen GM, Onorato IM. (1994). Epidemiology of tuber-culosis in the United States, 1985 through 1992.. *JAMA*. **272** (7), p535-539.

[88] Jia Z, Cheng S, Jia X (2011) A Mathematical model for evaluating tuberculosis screening strategies. *J Evid Based Med*. doi: 10.1111/j.1756–5391.2011.01116.x.

[89] Liu L, Zhao X, Zhou Y (2010) A tuberculosis model with seasonality. *Bull Math Biol* **72** 931–952.

[90] Blower, S.M. (1995). The intrinsic transmission dynamics of tuberculosis epidemics. *Nat. Med* **1**, 815–821.

[91] Blower, S.M., Small, P.M., Hopewell, P.C. (1996). Control strategies for tuberculosis epidemics: new models for old problems. *Science* **273**, 497–500.

[92] Ziv, E., Daley, C.L., Blower, S.M. (2001). Early therapy for latent tuberculosis infection. *Am. J. Epidemiol.* **153**, 381–385.

[93] Jia, Z. W., Tang, G. Y., Jin, Z., Dye, C., Vlas, S. J., Li, X. W., ... & Cao, W. C. (2008). Modeling the impact of immigration on the epidemiology of tuberculosis. *Theoretical population biology* **73** (3), 437-448.

[94] Sandgren, A., Schepisi, M. S., Sotgiu, G., Huitric, E., Migliori, G. B., Manissero, D., ... & Girardi, E. (2014). Tuberculosis transmission between foreign-and native-born populations in the EU/EEA: a systematic review. *European Respiratory Journal* **43** (4), 1159-1171.

[95] Falzon, D., & Aït-Belghiti, F. (2007). What is tuberculosis surveillance in the European Union telling us?. *Clinical infectious diseases*, **44** (10), 1261-1267.

[96] Central Statistics Office. Census of Ireland 2011.

[97] Health Service Executive. *Reports on the Epidemiology of TB in Ireland 1991-2002*. Retrieved August 24, 2016, from `http://www.hpsc.ie/A-Z/VaccinePreventable/TuberculosisTB/Epidemiology/AnnualReports/`.

[98] World Health Organization. (2014). *Framework for tuberculosis elimination in low-incidence countries*. Geneva: World Health Organization.

[99] Cleveland, W. S., & Loader, C. (1996). Smoothing by local regression: Principles and methods. *Statistical theory and computational aspects of smoothing* (pp. 10-49). Physica-Verlag HD.

[100] Neyrolles, O., & Quintana-Murci, L. (2009). Sexual inequality in tuberculosis. *PLoS Med*, **6** (12), e1000199.

Health Service Executive. *Reports on the Epidemiology of TB in Ireland 2007*. Retrieved August 24, 2016, from `http://www.hpsc.ie/A-Z/VaccinePreventable/TuberculosisTB/Epidemiology/AnnualReports/`.

[101] Health Service Executive. *Reports on the Epidemiology of TB in Ireland 2007*. Retrieved August 24, 2016, from `http://www.hpsc.ie/A-Z/VaccinePreventable/TuberculosisTB/Epidemiology/AnnualReports/`.

[102] Health Service Executive. *Reports on the Epidemiology of TB in Ireland 2012*. Retrieved August 24, 2016, from http://www.hpsc.ie/A-Z/VaccinePreventable/TuberculosisTB/Epidemiology/AnnualReports/.

[103] Lalor, M. K., Pedrazzoli, D., Davidson, J. A., Anderson, L. F., Shaji, K., Mohiyuddin, T., ... & Thomas, H. L. (2014). Tuberculosis in the UK, 2014 report. *Tuberculosis in the UK, 2014 report.*

[104] European Centre for Disease Prevention and Control/WHO Regional Office for Europe (2014). *Tuberculosis Surveillance and Monitoring in Europe 2014*. Stockholm, European Centre for Disease Prevention and Control, 2014.

[105] Alami, N. N., Yuen, C. M., Miramontes, R., Pratt, R., Price, S. F., Navin, T. R., & Centers for Disease Control and Prevention. (2014). Trends in tuberculosis—United States, 2013. *MMWR Morb Mortal Wkly Rep*, **63** (11), 229-233.

[106] The United Nations Refugee Agency (2002), *The 2002 Global Report*.

[107] The United Nations Refugee Agency (2014). Syrian Refugee Response - Regional Overview. Available at: http://data.unhcr.org/syrianrefugees/regional.php. [Accessed 22 September 2016].

[108] Fisman, D. N. (2007). Seasonality of infectious diseases. *Annu. Rev. Public Health* **28**, 127-143.

[109] Geyer, C. J. (2010). *Introduction to Markov chain Monte Carlo*. In Handbook of Markov Chain Monte Carlo. CRC, London.

[110] Box, G. E. P., and Jenkins, G. (1976), *Time Series Analysis: Forecasting and Control*, Holden-Day.

[111] Cleveland, R. B., Cleveland, W. S., McRae, J. E., & Terpenning, I. (1990). STL: A seasonal-trend decomposition procedure based on loess. *Journal of Official Statistics*, **6**(1), 3-73.

[112] Health Service Executive. *Reports on the Epidemiology of TB in Ireland 2011*. Retrieved August 24, 2016, from `http://www.hpsc.ie/A-Z/VaccinePreventable/TuberculosisTB/Epidemiology/AnnualReports/`.

[113] McKenna, M. T., McCray, E., & Onorato, I. (1995). The epidemiology of tuberculosis among foreign-born persons in the United States, 1986 to 1993. *New England Journal of Medicine*, **332** (16), 1071-1076.

[114] Bureau of Labour Statistics (2015) *Foreign-Born Workers: Labor Force Characteristics - 2015*, US Department of Labour.

[115] World Health Organisation (2010) *Tuberculosis country work summary Lithuania*, World Health Organisation.

[116] Thorpe, L. E., Frieden, T. R., Laserson, K. F., Wells, C., & Khatri, G. R. (2004). Seasonality of tuberculosis in India: is it real and what does it tell us?. *The Lancet*, **364** (9445), 1613-1614.

[117] Nagayama, N., & Ohmori, M. (2006). Seasonality in various forms of tuberculosis. The International Journal of Tuberculosis and Lung Disease, 10(10), 1117-1122.

[118] Schaaf, H. S., Nel, E. D., Beyers, N., Gie, R. P., Scott, F., & Donald, P. R. (1996). A decade of experience with Mycobacterium tuberculosis culture from children: a seasonal influence on incidence of childhood tuberculosis. *Tubercle and Lung Disease*, **77** (1), 43-46.

[119] Douglas, A. S., Strachan, D. P., & Maxwell, J. D. (1996). Seasonality of tuberculosis: the reverse of other respiratory diseases in the UK. *Thorax*, **51** (9), 944-946.

[120] Leung, C. C., Yew, W. W., Chan, T. Y. K., Tam, C. M., Chan, C. Y., Chan, C. K., ... & Law, W. S. (2005). Seasonal pattern of tuberculosis in Hong Kong. textbfInternational journal of epidemiology **34** (4), 924-930.

[121] Willis, M. D., Winston, C. A., Heilig, C. M., Cain, K. P., Walter, N. D., & Mac Kenzie, W. R. (2012). Seasonality of tuberculosis in the United States, 1993–2008. *Clinical infectious diseases* **54** (11), 1553-1560.

[122] Davies, P. D. O. (1985). A possible link between vitamin D deficiency and impaired host defence to Mycobacterium tuberculosis. *Tubercle*, **66** (4), 301-306.

[123] Rios, M., Garcia, J. M., Sanchez, J. A., & Perez, D. (2000). A statistical analysis of the seasonality in pulmonary tuberculosis. *European journal of epidemiology*, **16** (5), 483-488.

[124] Health and Safety Executive. 2015. Tuberculosis Vaccination. [ONLINE] Available at: `https://www.hse.ie/eng/health/az/B/-tuberculosis-vaccination/`. [Accessed 22 September 2016].

[125] Li, X. X., Wang, L. X., Zhang, H., Du, X., Jiang, S. W., Shen, T., ... & Zeng, G. (2013). Seasonal variations in notification of active tuberculosis cases in China, 2005–2012. *PLoS One*, **8** (7), e68102.

[126] Hethcote, H. W., Stech, H. W., & van den Driessche, P. (1981). Periodicity and stability in epidemic models: a survey. Differential Equations and Applications in Ecology, *Epidemics and Population Problems*, Academic Press, New York, 65-82.

[127] Health Service Executive. *Reports on the Epidemiology of TB in Ireland 2014*. Retrieved August 24, 2016, from `http://www.hpsc.ie/A-Z/VaccinePreventable/TuberculosisTB/Epidemiology/AnnualReports/`.ce

[128] Yeh, Y. P., Luh, D. L., Chang, S. H., Suo, J., Chang, H. J., & Chen, T. H. H. (2005). Tuberculin reactivity in adults after 50 years of universal bacille Calmette—Guérin vaccination in Taiwan. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **99** (7), 509-516.

[129] Tanaka MM, Francis AR, Luciani F, Sisson SA. (2006) Using approximate Bayesian computation to estimate tuberculosis transmission parameters from genotype data. *Genetics* **173** 1511-20.

[130] Streftaris, G. and Gibson, G. J. (2001). Bayesian inference for stochastic epidemics in closed populations. *Statistical Modelling* **4**, 63-75.

[131] Smith, R. C. (2013). *Uncertainty quantification: theory, implementation, and applications* (Vol. **12**). SIAM.

[132] Parrinello, C. M., Crossa, A., & Harris, T. G. (2012). Seasonality of tuberculosis in New York City, 1990–2007. *The International Journal of Tuberculosis and Lung Disease*, **16** (1), 32-37.

[133] Diekmann O, Heesterbeek JA, Roberts MG. (2010) The construction of next-generation matrices for compartmental epidemic models. *J. R. Soc. Interface* **7** 873–85.

[134] Anh, D. T., Bonnet, M. P., Vachaud, G., Van Minh, C., Prieur, N., & Duc, L. V. (2006). Biochemical modeling of the Nhue River (Hanoi, Vietnam): Practical identifiability analysis and parameters estimation. *Ecological modelling* **193** (3), 182-204.

[135] Marino, S., Hogue, I. B., Ray, C. J., & Kirschner, D. E. (2008). A methodology for performing global uncertainty and sensitivity analysis in systems biology. *Journal of theoretical biology*, **254** (1), 178-196.

[136] Mukaka, M. M. (2012). A guide to appropriate use of Correlation coefficient in medical research. *Malawi Medical Journal* **24** (3), 69-71.

[137] Aldridge, R. W., Yates, T. A., Zenner, D., White, P. J., Abubakar, I., & Hayward, A. C. (2014). Pre-entry screening programmes for tuberculosis in migrants to low-incidence countries: a systematic review and meta-analysis. *The Lancet Infectious Diseases*, **14** (12), 1240-1249.

[138] Weisstein, E. W. (2002). Heaviside step function.

[139] Dorgan, S. (2006). How Ireland became the Celtic tiger. *The Heritage Foundation Backgrounder* 1945.

[140] Chin, D. P., DeRIEMER, K. A. T. H. R. Y. N., Small, P. M., De Leon, A. P., Steinhart, R., Schecter, G. F., ... & Agasino, C. B. (1998). Differences in contributing

factors to tuberculosis incidence in US-born and foreign-born persons. *American journal of respiratory and critical care medicine* **158** (6), 1797-1803.

[141] Salinas, J. L. (2016). *Leveling of tuberculosis incidence—United States, 2013–2015*. MMWR. Morbidity and mortality weekly report, 65.

[142] Varughese, M. B., Langlois-Klassen, D., Long, R., & Li, M. (2014). Preventing tuberculosis in the foreign-born population of Canada: a mathematical modelling study. *The International Journal of Tuberculosis and Lung Disease* **18** (4), 405-412.

[143] Varughese, M. B., Langlois-Klassen, D., Long, R., & Li, M. (2014). Preventing tuberculosis in the foreign-born population of Canada: a mathematical modelling study. *The International Journal of Tuberculosis and Lung Disease* **18** (4), 405-412.

[144] Nnadi, C. D., Anderson, L. F., Armstrong, L. R., Stagg, H. R., Pedrazzoli, D., Pratt, R., ... & Moonan, P. K. (2016). Mind the gap: TB trends in the USA and the UK, 2000–2011. *Thorax*, thoraxjnl-2015.

[145] Odone, A., Tillmann, T., Sandgren, A., Williams, G., Rechel, B., Ingleby, D., ... & McKee, M. (2015). Tuberculosis among migrant populations in the European Union and the European Economic Area. *European journal of public health*, **25** (3), 506-512.

[146] Health Information and Quality Authority (2015) *Health technology assessment of a selective BCG vaccination programme*, Health Information and Quality Authority.

[147] Health Protection Surveillance Centre - `http://www.hpsc.ie/A-Z/VaccinePreventable/TuberculosisTB/`

[148] Klotz, A., Harouna, A., & Smith, A. F. (2013). Forecast analysis of the incidence of tuberculosis in the province of Quebec. *BMC public health*, **13**(1), 1.

[149] Okuonghae, D., & Aihie, V. U. (2010). Optimal control measures for tuberculosis mathematical models including immigration and isolation of infective. *Journal of Biological Systems*, **18** (01), 17-54.

[150] Varughese, M. B., Langlois-Klassen, D., Long, R., & Li, M. (2014). Preventing tuberculosis in the foreign-born population of Canada: a mathematical modelling study. *The International Journal of Tuberculosis and Lung Disease*, **18**(4), 405-412.

[151] Zhou, Y., Khan, K., Feng, Z., & Wu, J. (2008). Projection of tuberculosis incidence with increasing immigration trends. *Journal of theoretical biology*, **254**(2), 215-228.

[152] Koppaka, V. R., Harvey, E., Mertz, B., & Johnson, B. A. (2003). Risk factors associated with tuberculin skin test positivity among university students and the use of such factors in the development of a targeted screening program. *Clinical infectious diseases*, **36**(5), 599-607. Chicago.

[153] Sobol, I. M. (1998). On quasi-monte carlo integrations. *Mathematics and Computers in Simulation*, **47**(2), 103-112.

[154] Dockery, D., & Pope, A. (1996). Epidemiology of acute health effects: summary of time-series studies. *Particles in our air: Concentrations and health effects*, 123-147.

[155] Boutayeb, A., & Chetouani, A. (2006). A critical review of mathematical models and data used in diabetology. *Biomedical engineering online*, **5** (1), 1. Chicago

[156] Romanus, V. (2005). Selective BCG vaccination in a country with low incidence of tuberculosis. *Euro surveillance: European communicable disease bulletin*, **11** (3), 14-17.

[157] F. J. Bentley, S. Grzybowski, and B. Benjamin, Tuberculosis in Childhood and Adolescence, 1954, National Association for the Prevention of Tuberculosis, London, UK, 1954.

[158] Rebelo, C., Margheri, A., and Bacaër, N. (2012). Persistence in seasonally forced epidemiological models. Journal of Mathematical Biology, 64(6), 933-949.

[159] B. J. Marais and P. R. Donald, "The natural history of tuberculosis infection and disease in children," in Tuberculosis: A Comprehensive Clinical Reference, H. S. Schaaf and A. Zumla, Eds., Elsevier Health Sciences, 2009.

[160] Ortblad, K. F., Salomon, J. A., Bärnighausen, T., & Atun, R. (2015). Stopping tuberculosis: a biosocial model for sustainable development. The Lancet, 386(10010), 2354-2362.

[161] World Health Organisation - STOP TB Campaign (2002), *What is the relationship between TB and poverty?*.

[162] Jensen, P. A., Lambert, L. A., Iademarco, M. F., Ridzon, R., & Centers for Disease Control and Prevention. (2005). Guidelines for preventing the transmission of Mycobacterium tuberculosis in health-care settings, 2005.

[163] Software, P. (n.d.). Tuberculosis. Retrieved September 27, 2017, from `http://www.hse.ie/eng/health/Immunisation/pubinfo /babychildimm/vaccprevdisease/tb/#Who should get BCG vaccine?`

[164] Briggs, A. D., Wolstenholme, J., Blakely, T., & Scarborough, P. (2016). Choosing an epidemiological model structure for the economic evaluation of non-communicable disease public health interventions. Population health metrics, 14(1), 17.

[165] Box, G. E. (1976). Science and statistics. Journal of the American Statistical Association, 71(356), 791-799.

[166] Bartlett, M. S. (1956). Deterministic and stochastic models for recurrent epidemics. In Proceedings of the third Berkeley symposium on mathematical statistics and probability (Vol. 4, No. 81, p. 109).

[167] Silverman, B. G., Hanrahan, N., Bharathy, G., Gordon, K., & Johnson, D. (2015). A systems approach to healthcare: agent-based modeling, community mental health, and population well-being. Artificial intelligence in medicine, 63(2), 61-71.

[168] Roberts, M., Andreasen, V., Lloyd, A., & Pellis, L. (2015). Nine challenges for deterministic epidemic models. Epidemics, 10, 49-53.

[169] Gumel, A. B., Ruan, S., Day, T., Watmough, J., Brauer, F., Van den Driessche, P., ... & Wu, J. (2004). Modelling strategies for controlling SARS outbreaks. Proceed-

ings of the Royal Society of London B: Biological Sciences, 271(1554), 2223-2232. Chicago

[170] Hawke, I., Husa, S., & Szilágyi, B. (2004). Numerical methods for ODEs. Chicago

[171] Bowong, S., & Kurths, J. (2011). Modeling and Parameter Estimation of Tuberculosis with Application to Cameroon. International Journal of Bifurcation and Chaos, 21(07), 1999-2015.

[172] Hu, X. (2012). Threshold dynamics for a tuberculosis model with seasonality. Mathematical biosciences and engineering: MBE, 9(1), 111-122.

[173] Leung, C. C., Yew, W. W., Chan, T. Y. K., Tam, C. M., Chan, C. Y., Chan, C. K., ... & Law, W. S. (2005). Seasonal pattern of tuberculosis in Hong Kong. International journal of epidemiology, 34(4), 924-930. Chicago

[174] Parrinello, C. M., Crossa, A., & Harris, T. G. (2012). Seasonality of tuberculosis in New York City, 1990–2007. The International Journal of Tuberculosis and Lung Disease, 16(1), 32-37.

[175] Soetens, L. C., Boshuizen, H. C., & Korthals Altes, H. (2013). Contribution of seasonality in transmission of Mycobacterium tuberculosis to seasonality in tuberculosis disease: a simulation study. American journal of epidemiology, 178(8), 1281-1288.

[176] Sun, G. Q., Bai, Z., Zhang, Z. K., Zhou, T., & Jin, Z. (2013). Positive periodic solutions of an epidemic model with seasonality. The Scientific World Journal, 2013.

[177] Bowong, S., & Tewa, J. J. (2010). Global analysis of a dynamical model for transmission of tuberculosis with a general contact rate. Communications in Nonlinear Science and Numerical Simulation, 15(11), 3621-3631.

[178] Cohen, T., & Murray, M. (2005). Incident tuberculosis among recent US immigrants and exogenous reinfection. Emerging infectious diseases, 11(5), 725.

[179] Colijn, C., Cohen, T. E. D., & MURRAY, M. (2007). Mathematical models of tuberculosis: accomplishments and future challenges. In BIOMAT 2006 (pp. 123-148).

[180] Denholm, J. T., & McBryde, E. S. (2014). Can Australia eliminate TB. Modelling immigration strategies for reaching MDG targets in a low transmission setting. Australian and New Zealand journal of public health, 38(1), 78-82.

[181] Guo, H. O. N. G. B. I. N., & Li, M. Y. (2011). Global stability of the endemic equilibrium of a tuberculosis model with immigration and treatment. Canad. Appl. Math. Quart, 19, 1-18.

[182] Guo, H., & Wu, J. (2011). Persistent high incidence of tuberculosis among immigrants in a low-incidence country: impact of immigrants with early or late latency. Math Biosci Eng, 8, 695-709.

[183] Hill, A. N., Becerra, J. E., & Castro, K. G. (2012). Modelling tuberculosis trends in the USA. Epidemiology & Infection, 140(10), 1862-1872.

[184] Klotz, A., Harouna, A., & Smith, A. F. (2013). Forecast analysis of the incidence of tuberculosis in the province of Quebec. BMC public health, 13(1), 400.

[185] Li, L., Sun, G. Q., & Jin, Z. (2009). Traveling pattern induced by migration in an epidemic model. Journal of Biological Systems, 17(02), 319-328.

[186] Ma, Z. (2009). Dynamical modeling and analysis of epidemics. World Scientific.

[187] Okuonghae, D., & Aihie, V. U. (2010). Optimal control measures for tuberculosis mathematical models including immigration and isolation of infective. Journal of Biological Systems, 18(01), 17-54.

[188] Varughese, M. B., Langlois-Klassen, D., Long, R., & Li, M. (2014). Preventing tuberculosis in the foreign-born population of Canada: a mathematical modelling study. The International Journal of Tuberculosis and Lung Disease, 18(4), 405-412.

[189] Wolleswinkel-van den Bosch J, H., Nagelkerke, N. J., Broekmans, J., & Borgdorff, M. (2002). The impact of immigration on the elimination of tuberculosis in The Netherlands: a model based approach. The International Journal of Tuberculosis and Lung Disease, 6(2), 130-136.

[190] Yang, Y., Wu, J., Li, J., & Ma, Z. (2010). Global dynamics–convergence to equilibria–of epidemic patch models with immigration. Mathematical and Computer Modelling, 51(5), 329-337.

[191] Zhou, Y., Khan, K., Feng, Z., & Wu, J. (2008). Projection of tuberculosis incidence with increasing immigration trends. Journal of theoretical biology, 254(2), 215-228.

[192] World Health Organization, 2015. Assessing and Improving the Accuracy of Target Population Estimates for Immunization Coverage. World Health Organisation. Available at:
http://www.who.int/immunization/monitoring_surveillance/data/Denominator_guide.pdf?ua=1 [Accessed: 10 Aug 2017]

[193] World Health Organisation - Case Notification Data. (2013). Retrieved from http://www.who.int/tb/country/data/download/en/

[194] Glaziou, P., Sismanidis, C., Zignol, M., & Floyd, K. Methods used by WHO to estimate the global burden of TB disease.

[195] Diekmann, O., & Heesterbeek, J. A. P. (2000). Mathematical epidemiology of infectious diseases: model building, analysis and interpretation (Vol. 5). John Wiley & Sons.

[196] The World Bank, World Development Indicators (2012). [Data file]. Retrieved from http://data.worldbank.org/indicator/

[197] Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. Statistical science, 457-472.

[198] Soetaert, K. E. R., Petzoldt, T., & Setzer, R. W. (2010). Solving differential equations in R: package deSolve. Journal of Statistical Software, 33.

[199] Dormand, J. R., & Prince, P. J. (1980). A family of embedded Runge-Kutta formulae. Journal of computational and applied mathematics, 6(1), 19-26.

[200] National Disease Surveillance Centre. (2003). Case definitions for notifiable diseases. Dublin: National Disease Surveillance Centre.

[201] Central Statistics Office. Census of Ireland 2011.

[202] Public Health England. Tuberculosis in the UK: 2014 report. London: Public Health England; 2014