Predicting the Peak of Influenza Cases by Geographical Zones in Alberta

by

Jeannette Amissah

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

in

BIOSTATISTICS

Department of Mathematical and Statistical Sciences
University of Alberta

Abstract

Influenza or the 'flu' can affect people from all walks of life. The burden from influenza epidemics puts tremendous pressure on health services and other resources during a flu season. To better prepare for an incoming flu season, clinicians, health services and policy makers have a great interest in developing the capacity to predict the timing of the peak of an influenza epidemic based on identified cases in the early phase of the season.

The objective of this thesis is to formulate an influenza model that can predict the week when the laboratory confirmed influenza cases peak for different geographical zones in the Province of Alberta: Edmonton, Calgary, North, Central and South zones. A Kermarck-McKendrick type compartmental model that comprises of the susceptible-infected (SI) compartments for three age groups (0-18 years, 19-64 years, and 64 years and over) is proposed. Contact mixing matrix among the age groups is computed. Estimates of model parameters are obtained by fitting the model to past influenza data (Year 2014-2015) from Alberta Health, using the nonlinear least squares method and the Mathematica software. The 95% confidence intervals for model parameters are obtained using the Markov Chain Monte Carlo method and then used for uncertainty analysis of our model predictions.

Our model predictions for the peak time have shown a good agreement with past data for all Alberta zones. The findings in this thesis provide the groundwork and insight valuable for further development of influenza forecasting models.

Acknowledgements

I would like to thank my supervisors, Dr. Brendan Pass and Dr. Michael Y. Li for their continued support and advice throughout my program. I would also like to thank them for their creative advice as well as their thoughtful suggestions on this project. Sincere thanks also go to Marie Varughese, Alberta Health Services, for her advice, help and her readiness to answer all my questions. Another sincere appreciation goes to my husband, Mr. James Eduful for his encouragements, patience, love as well as his contributions during all my time in school and also to my family and friends for their continued support.

Finally, I would also like to thank the Department of Mathematical and Statistical Sciences for their generous financial support. Thank you all for your help. I am really appreciative.

Table of Contents

1 Introduction 1

1.1 Background Information 1

1.1.1 Epidemiology of Influenza 1

1.1.2 Symptoms and Preventions 3

1.1.3 Worldwide Epidemics 4

1.1.4 Influenza Pandemics 5

1.2 Objective and Implications 5

1.3 Methodology Used 6

1.4 Limitations 7

1.5 Organization of thesis 8

2 Literature Reviews of Influenza Models 9

2.1 Compartmental Models 9

2.1.1 Henneman, Peursem and Huber (2013) 9

2.1.2 Bedada, Lemma and Koya (2015) 12

2.1.3 Varughese (2015) 14

2.2 Statistical Models 16

2.2.2 Wei (2011) 16

2.3.2 Nsoesie, Marathe, and Brownstein (2013) 17

2.3 Contact Mixing Matrix 19

3 Formulation of Influenza Model 22

3.1 Data Description 22

3.2 The Influenza Model 26

3.2.1 Introduction 26

3.2.2 Derivation of the Influenza Model 27

3.3 Derivation of the Contact Mixing Matrix 31

3.4 Statistical Analysis using Nonlinear Regression 36

3.5 Description of the Simulation code used in the analysis 39

4 Numerical Investigations 44

4.1 Results from Numerical Investigations 44

4.1.1 Results from Numerical Investigations for Central Zone 44

4.1.2 Results from Numerical Investigations for North Zone 53

4.1.3 Results from Numerical Investigations for Calgary Zone 61

4.2 Discussions from Numerical Investigation 71

5 Conclusions and Recommendations 74

5.1 Conclusions 74

5.2 Recommendations 75

Bibliography 76

List of Tables

Table 3.1.1: Parameter Descriptions and Units used in the model 31

Table 3.1.2: Daily number of contacts for the three age groups 33

Table 3.3.3: A contact mixing matrix for each age group for Calgary zone 36

Table 4.1.1: Parameter estimates for $\rho(t)$ for each age groups…….45

Table 4.1.2: 2014-2015 Population estimates for Central zone. 49

Table 4.1.3: Contact mixing matrix for Central zone. 49

Table 4.1.4: Parameter estimates for Central zone. 50

Table 4.2.1: Contact mixing matrix for North zone……...54

Table 4.2.2: 2014-2015 Population estimates for North zone. 54

Table 4.2.3: Parameter estimates for North zone. 55

Table 4.2.4: Parameter estimates for $\rho(t)$ for each age groups. 56

Table 4.3.1: Contact mixing matrix for Calgary zone……...62

Table 4.3.2: 2014-2015 Population estimates for Calgary zone. 62

Table 4.3.3: Parameter estimates for Calgary zone 63

Table 4.3.4: Parameter estimates for $\rho(t)$ for each age groups. 64

List of Figures

Figure 2.1.1: Henneman et al. Influenza Model 10

Figure 2.1.2: Bedada et. al. flow diagram of the influenza model..... 13

Figure 3.1.1: Laboratory confirmed cases for age group 0 – 18 years……..23

Figure 3.1.2: Laboratory confirmed cases for age group 19 – 64 years 24

Figure 3.1.3: Laboratory confirmed cases for age group 65 years and over..... 25

Figure 3.1.4: Laboratory confirmed cases for all age groups and all zones 26

Figure 3.1.5: Flow diagram of the proposed influenza model..... 29

Figure 4.1.1: ρt for the 0-18 years age group……45

Figure 4.1.2: ρt for the 19-64 years age group. 46

Figure 4.1.3: ρt for the 65 years and over age group. 46

Figure 4.1.4: Prediction bands for 0-18-year age group. 47

Figure 4.1.5: Prediction bands for 19-64-year age group. 48

Figure 4.1.6: Prediction bands for 65 years and over age group. 48

Figure 4.1.7: Laboratory confirmed cases for the Central zone..... 51

Figure 4.1.8: Accumulated laboratory confirmed cases for 0-18 years. 52

Figure 4.1.9: Accumulated laboratory confirmed cases for 19-64 years. 52

Figure 4.1.10: Accumulated laboratory confirmed cases for 65 years and over. . 53

Figure 4.2.1: ρt for the 0-18 years age group……56

Figure 4.2.2: ρt for the 19-64 years age group..... 56

Figure 4.2.3: ρt for the 65 years and over age group. 57

Figure 4.2.4: Prediction bands for 0-18 years' age group..... 57

Figure 4.2.5: Prediction bands for 19-64 years' age group..... 58

Figure 4.2.6: Prediction bands for 65 years and over age group. 58

Figure 4.2.7: Laboratory confirmed cases for the North zone. 59

Figure 4.2.8: Accumulated laboratory confirmed cases for 0-18 years. 60

Figure 4.2.9: Accumulated laboratory confirmed cases for 19-64 years. 60

Figure 4.2.10: Accumulated laboratory confirmed cases for 65 years and over. . 61

Figure 4.3.1: ρt for the 0-18 years age group……64

Figure 4.3.2: ρt for the 19-64 years age group. 65

Figure 4.3.3: ρt for the 65 years and over age group. 65

Figure 4.3.4: Prediction bands for 0-18 years' age group..... 66

Figure 4.3.5: Prediction bands for 19-64 years' age group..... 66

Figure 4.3.6: Prediction bands for 19-64 years' age group..... 67

Figure 4.3.7: Laboratory confirmed cases for the Calgary zone..... 68

Figure 4.3.8: Accumulated laboratory confirmed cases for 0-18 years. 69

Figure 4.3.9: Accumulated laboratory confirmed cases for 19-64 years. 69

Figure 4.3.10: Accumulated laboratory confirmed cases for 65 years and over. . 70

Chapter 1

1 Introduction

In this Chapter, a general background on the influenza virus which comprises the epidemiology of the virus, symptoms and preventions of the virus, worldwide influenza epidemics as well as some influenza pandemics that have been experienced are provided in Section 1.1. The objective of this thesis and its implication are also provided in Section 1.2. The methodology used in the thesis, the limitations encountered and the organization of the thesis are presented in Sections 1.3, 1.4 and 1.5.

1.1 Background Information

1.1.1 Epidemiology of Influenza

Influenza or the 'flu', is a respiratory disease that affects the nose, throat and occasionally the lungs of humans as well as animals like birds, dogs etc. It is caused by the influenza viruses known as orthomyxoviruses which are a family of RNA viruses made up of six genera viruses: Influenza-virus A, Influenza-virus B, Influenza-virus C, Isa-virus, Thogoto-virus and Quaranja-virus [11]. Influenza A, B, and C viruses affect humans, birds, pigs, dogs, seals as well as other mammals [11]. Influenza A virus is considered to be the most dangerous human pathogen among the three influenza types since it is the cause of most or all of the flu pandemics that occur around the world [14]. Influenza B and C are considered to be less common with Influenza C usually causing both severe illnesses and local

epidemics [16]. The inability of influenza B to mutate at a faster rate, that is, to change its antigens, ensures that pandemics of this type of influenza do not occur [25]. Types of influenza viruses can be divided into subtypes or different serotypes based on their antibody response to their viral particles [7]. For instance, Influenza A viruses are divided into subtypes based on their antibody responses to the viral surface proteins: Hemagglutin (HA or H) and neuraminidase (NA or N) [24]. H1N1, H2N2, H3N2, H5N1, H1N2, H7N9 are some of the subtypes of influenza virus A with H1N1 and H3N2 commonly found in humans. Influenza viruses' B and C are not divided into subtypes however influenza B is further divided into lineages and strains with some of the lineages being B/Yamagata and B/Victoria. Influenza viruses undergo certain antigenic changes that happen continually over time. Two processes that cause the antigens to change are the antigenic shift and antigenic drift with antigenic drift being more common than the other [4]. Antigenic drift of influenza causes the virus to change over time usually a new strain of virus evolves against the current recognised strain of virus. When this happens the antibodies that are built against the current strain of virus are not able to fight against the newer strain hence resulting in reinfection. This is why there is the need for new vaccinations every year. Due to the ability of the antigens to change over time, people usually have little or no immunity at all. The Antigenic shift on the other hand causes two different flu strains to combine and infect the same cell hence creating a new type of flu. People end up having no immunity against the new virus [4].

1.1.2 Symptoms and Preventions

Influenza is usually passed from person to person or through droplets from an infected individuals' cough or sneeze. It can also be spread by touching surfaces that have been contaminated by the virus and then touching the mouth or the eyes [13]. Influenza symptoms can be mild or severe. Some of the common symptoms include high fever, headache, aching muscles, sore throat, loss of appetite, feeling tired etc. The symptoms of influenza typically begin two days after one has been exposed to the disease [13]. According to the World Health Organization [WHO], most infected people tend to recover within a week without requiring any medical treatment; however in the young, elderly and those with serious medical conditions, influenza seems to be more severe and can results in death [10]. Influenza viruses circulate worldwide and can affect anybody in any age group: children younger than 2 years of age, adults aged 65 years or older, pregnant women and people of any age with medical conditions like liver, lung disease and kidney disease can be infected [10]. People are asymptomatic in the initial infection stage and can still infect others but as time goes on can be very infectious. Vaccination is one of the most effective way to protect people against the flu viruses. The flu vaccines are effective against the viruses when they contain either the same or related strain of the virus. For example, influenza A (H1N1) and A(H3N2) and one or two of influenza B are usually included in influenza vaccines each year [22]. However, changes in the strains of the influenza virus tend to cause any particular vaccine received to confer minimal protection.

1.1.3 Worldwide Epidemics

According to the World Health Organization [WHO], seasonal epidemics mainly occur during winter in temperate climates as compared to the tropical region where influenza occurs throughout the year. WHO also estimate that the annual influenza epidemics usually results in about 3 to 5 million cases of severe illness and about 250000 to 500000 deaths worldwide. This imposes economic burden in the form of hospital and other health cost and lost productivity on the world. In the industrialised countries, most deaths associated with influenza occur among people aged 65 years and over [10]. Flu season typically runs from August (week 35) to August (week 34). However, in Canada, active seasonal influenza occurs annually between November and March. In Canada, the number of hospitalization from flu ranges between 1000 and 8000 with death rate estimated to be between 100-800 cases per year [21]. The Community and Hospital Infection Control Association (CHICA) in Canada estimates that the numbers of influenza cases have increased from 7422 cases in 2005-2006 to 43510 cases in 2014-2015 [3]. In Alberta alone, 4,850 laboratory-confirmed cases were reported for the 2014-2015 influenza season which was 20% and 60% more than in the 2013-2014 and 2012-2013 seasons respectively [1]. The strains of the influenza virus in Alberta have been ranging between H3N2 and H1N1.

1.1.4 Influenza Pandemics

Influenza pandemic is the outbreak of the influenza virus worldwide which normally affects a large proportion of the human population. These pandemics results in high mortality among the people. It usually occurs when new strains of the flu virus are transmitted from animal species to humans [9]. Some influenza pandemics that have occurred in the world include:

- the 1889-1890 pandemic flu known as the Asiatic or Russian flu,
- the 1918-1920 flu pandemic known as the Spanish flu where H1N1 was the most predominant strain experienced,
- the Asian flu in 1957-1958 with H2N2 subtype of influenza A being the predominant strain where the elderly had the highest rates of death,
- the Hong Kong flu in 1968-1969 which was caused by the H3N2 strain of influenza A, and
- the 2009-2010 pandemic flu known as swine flu which occurred worldwide and was caused by the H1N1 influenza virus. [9].

1.2 Objective and Implications

Influenza epidemics and pandemics place a huge burden on society and individuals. Every year, influenza and its health problems put a significant burden on the health-care system as well as an economic toll on productivity through lost workforce and school absenteeism. Forecasting influenza epidemics not only helps in health resource allocation, it also helps in preparedness planning as well as educating the public on how they can adopt better personal health care around

infected individuals. In Alberta, the interest of most clinicians and policy makers is to control influenza epidemics since it put more pressure on the health systems and productivity.

The objective of this thesis is to formulate an influenza model to predict the peak influenza week. This is the week with the highest laboratory confirmed influenza cases within the geographical zones of Alberta: Edmonton, Calgary, Central, North and South zones. This is in the interest of the Alberta Health Services which is to only determine the timing of when a rise in the influenza cases should be expected based on an identified cases of the influenza season. Epidemiology has made it possible for researchers to analyse different types of diseases including influenza. Knowing the week at which each zone peaks will help in health resource allocation and will enable government, health agencies and other stakeholders to plan ahead as to what medications to introduce and how to allocate them; that is, which zones requires immediate attention. It will also help to know whether the measures that have been put in place are enough to prevent the spread of the virus from one zone to another or from person to person.

1.3 Methodology Used

The methodology used in this thesis is as follows:

- A simple deterministic compartmental model developed by Kermarck and McKendrick in 1927 [12] is used. The mathematical model is made up of three compartments namely the susceptible(S), infected but not lab confirmed(I_D) and infected but lab confirmed (I_L). The three compartments

are subdivided into three age groups 0-18 years, 19-64 years and 65 years and over.

- Numerical simulations are performed using Wolfram Mathematica Version 10.3.1 and Microsoft Excel. The simulations are conducted using data obtained from the Alberta Health services. The data comprises of all the laboratory confirmed cases obtained for each zone. Individual visits to the physicians as well as antiviral dispensing are also provided as part of the data.
- The best fit estimates are also obtained using the weighted nonlinear least squares method and the NDSOLVER in Mathematica. To ascertain that the parameters are the best estimates, the Metropolis Hasting Algorithm of the Markov Chain Monte Carlo method is used to determine the 95% confidence intervals for each of the parameters obtained. Latin Hypercube Sampling (LHS) is also used to generate 10000 samples based on the posterior distribution (the lower and upper confidence intervals and the point estimates) from the MCMC. This is to check the uncertainty in the peak week obtained.

1.4 Limitations

To obtain the best fit parameters, that is, the probability of transmission per contact for each age group, the parameters are randomly generated using the RANDOMREAL method in Mathematica where the parameters are chosen between a range of selected values from a specified sample. It is important to note

that the larger the sample size the more accurate the parameter estimates. However, as a result of computational time demand in running the program it was decided to use a sample size of 500,000 which gives a fairly accurate result as compared to using a larger sample size. Thus the parameter estimates obtained are usually local minimum instead of global minimum.

1.5 Organization of thesis

This thesis is organized into five chapters. Chapter 1 gives a general background and objective and methodology of the thesis. Chapter 2 presents literature review on some of the compartmental models as well as statistical models proposed for the influenza virus by previous researchers. A brief description of the data, statistical analysis and the formulation of the influenza model can be found in Chapter 3. Chapter 4 presents the results and discussions obtained from the simulations and analysis with Chapter 5 containing the conclusion and recommendations.

Chapter 2

2 Literature Reviews of Influenza Models

Understanding the dynamics of influenza has compelled several authors to model it using mathematical or statistical methods. In this chapter, a brief description of some of the previous mathematical and statistical models of influenza is presented. The views of some of these authors are presented below.

2.1 Compartmental Models

2.1.1 Henneman, Peursem and Huber (2013)

The authors proposed a mathematical model for influenza pandemic which takes into account that individuals are first infected with the influenza virus and then later contract a secondary bacterial infection. The authors presume that an individual with influenza become susceptible to the bacterial infection specifically bacterial pneumonia and then either contract the disease or recover. Henneman et al. proposed the influenza model based on this fact since this is the main cause of most of the deaths associated with influenza [8]. The influenza model proposed is a modification of the basic Kermarck and McKendrick SIR (Susceptible-Infected-Recovered) model with the introduction of a compartment for individuals who are susceptible to the secondary bacterial infection. The proposed model is expected to estimate the number of individuals who first become infected with the influenza virus and then become infected with the bacterial infection. The proposed model consisted of individuals who are susceptible (S) to the influenza

virus, infected with influenza virus (I_1), that is, those with symptomatic influenza, and individuals who recover from the virus and are temporarily susceptible to the bacteria infection (T). The individuals in the T compartment can either recover and move to the recovered (R) compartment or they can be infected with the secondary bacterial infection and move to the I_2 compartment. Individuals in the I_2 compartment on the other hand either die and are removed from the model or recover from the bacterial infection and move to R compartment. A schematic representation of the model proposed by Henneman et al. can be seen in Figure 2.1.1.

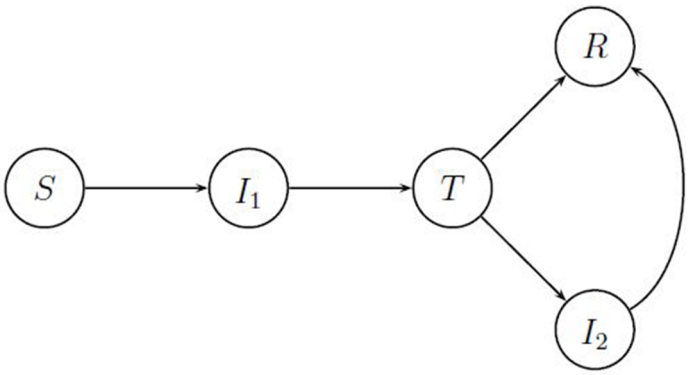

Figure 2.1.1: Henneman et al. Influenza Model

The ordinary differential equation derived by Henneman et al. [8] is as follows:

$$\frac{dS}{dt} = -\beta_1 I_1 S$$

$$\frac{dI_1}{dt} = \beta_1 I_1 S - \gamma_1 I_1$$

$$\frac{dT}{dt} = \gamma_1 I_1 - (\sigma + \beta_2 I_2) T \qquad (2.1.1a)$$

$$\frac{dI_2}{dt} = \beta_2 I_2 T - (\gamma_2 + d_2) I_2$$

$$\frac{dR}{dt} = \gamma_2 I_2 + \sigma T$$

The transmission rate of influenza is denoted by β_1, the rate at which an infected influenza patient recovers is denoted by γ_1 with σ showing the rate at which an individual looses his susceptibility from influenza to secondary infection after recovering. β_2 shows the transmission rate of the bacterial infection, with γ_2 denoting the recovery rate from the bacterial infection and d_2 representing the excess death rate due to bacterial infection. The model was constructed based on the assumption that there was no entry into or departure from the population except through an increased death rate from the secondary bacterial infection. Mathematical analysis of system (Equation 2.1.1a) shows that the maximum number of individuals infected with influenza during an epidemic is given by

$$I_{1max} = I_{10} + S_0 - \frac{\gamma_1}{\beta_1} lnS_0 - \frac{\gamma_1}{\beta_1} + \frac{\gamma_1}{\beta_1} ln\frac{\gamma_1}{\beta_1} \quad (2.1.1b)$$

The above equation predicts whether there will be an influenza epidemic or not and the severity of that epidemic at the beginning of any flu outbreak. In addition, the number of individuals that may become infected with the secondary bacterial infection can also be determined. Numerical simulations showed that the proposed model predicts the number of influenza cases obtained as well as the number of deaths.

Even though the influenza model proposed by Henneman et al. seems to predict the number of influenza cases as well as the number of deaths from the bacterial infection, the authors were mainly concerned with individuals with symptomatic influenza and did not consider individuals with asymptomatic influenza. This is because the authors presumed that those with asymptomatic influenza do not have an increased risk of getting the secondary bacterial infection hence concentrated

more on the symptomatic influenza. Another aspect that the model did not consider was the age distribution of the susceptible population and the fact that there can be cross infection among the various age groups. That is, the model proposed didn't take into considerations the various age groups specifically the at-risk populations which usually includes children, elderly individuals (65 years and older) etc. and so did not consider the contact that each of these groups make with themselves and others.

2.1.2 Bedada, Lemma and Koya (2015)

The authors proposed an SEI_SI_NR mathematical model which was an extension of the SEIR model to describe the propagation of the influenza disease among the population. According to Bedada et. al., Influenza type A, virus specifically H1N1, has a latent or exposed phase during which the individuals are said to be infected but are not showing any symptoms [2]; hence the SEIR mathematical model is modified to include these individuals. The infectious compartment of the SEIR model is segmented into the symptomatic, I_S, and the non-symptomatic I_N, compartments. Unlike the simple SEIR model, Bedada et al. included death rate in the symptomatic infectious compartment. The proposed model consists of the following compartments: susceptible (S), exposed (E), infected with symptoms (I_S), infected without symptoms (I_N), and recovered or removed (R). The model was constructed based on the assumption that individuals in the susceptible compartments are subject to infection due to contact with an infected population at a rate β. These susceptible after being infected with the disease move to the

exposed compartment where the virus multiplies for a period of time, k, and then from the exposed compartment a portion ρ of individuals enter into the infected with the symptoms compartment with a proportion (1-ρ) entering into the infected but without symptoms compartment. A pictorial view of the model proposed by Bedada et al. can be seen below:

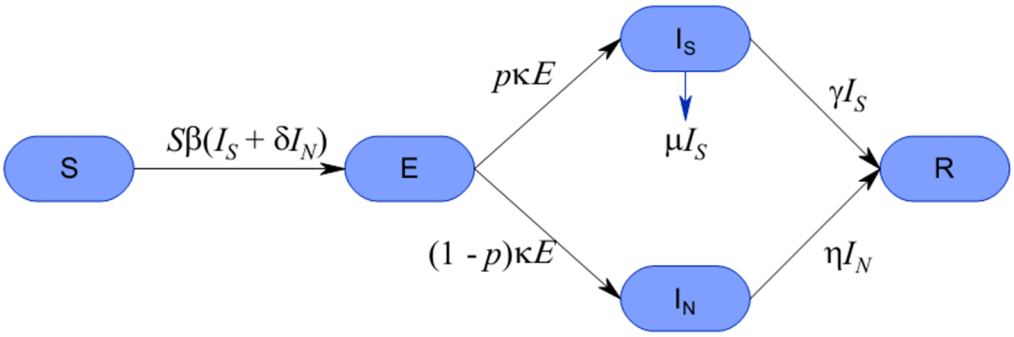

Figure 2.1.2: Bedada et. al. flow diagram of the influenza model

The mathematical formulation of the SEI_SI_NR model can be expressed as a system of differential equations as follows:

$$\frac{dS}{dt} = -\beta S(I_S + \delta I_N)$$

$$\frac{dE}{dt} = \beta S(I_S + \delta I_N) - KE$$

$$\frac{dI_S}{dt} = \rho KE - \gamma I_S - \mu I_S \qquad (2.1.2)$$

$$\frac{dI_N}{dt} = (1-\rho)KE - \eta I_N$$

$$\frac{dR}{dt} = \gamma I_S + \eta I_N$$

Here β is the transmission rate of the susceptible individual with infection, k is the latency period in the exposed class, ρ is the fraction of the exposed individuals that enter into I_S compartment, γ is the rate of recovery from the virus in the I_S

compartment, μ is the death rate with influenza in the I_S compartment, η is the rate of recovery from the disease in the I_N compartment and δ is the factor by which I_N have reduced infectivity. Simulation studies done by using different basic reproductive numbers show that the influenza disease, specifically the H1N1 influenza virus epidemic dies out within the population.

The mathematical model proposed by Bedada et al. included individuals who are infected but aren't showing symptoms (that is the asymptomatic individuals) as well as individuals that also show the symptoms; however, the proposed model did not take into consideration the age distribution of the susceptible population, that is, the various age groups specifically 0-18 years, 19-64 years and 65 years and over, where the young and elderly are usually considered to be the at risk individuals, and therefore did not consider the contact that each of these groups make with themselves and others.

2.1.3 Varughese (2015)

Varughese [23] proposed a mathematical model which was used to predict the peak of influenza in all of Alberta. The proposed mathematical model considered the three age groups 0-18 years, 19-64 years and 65 years and over. The mathematical model comprised of the susceptible compartment (S), infected but without laboratory confirmation (I_D) and infected but with laboratory confirmation (I_L). The model is constructed based on the assumption that births and deaths do not add much to the model since it just runs for a short period of time. Also, there was homogenous mixing within each compartment and by age groups (school age, workers and seniors). In the model proposed by Varughese,

there are cross infections among the various age groups implying that an individual from the younger age group (0-18years) can be infected with influenza by someone in the older group (64 years and over). The ordinary differential equations obtained from the mathematical model proposed by Varughese can be seen below

$$\frac{dS_1}{dt} = -\beta_1 S_1 I_{D1} - \beta_1 S_1 I_{D2} - \beta_1 S_1 I_{D3}$$

$$\frac{dI_{D1}}{dt} = \beta_1 S_1 I_{D1} + \beta_1 S_1 I_{D2} + \beta_1 S_1 I_{D3} - \sigma I_{D1} - f\rho_1 I_{D1} \quad (2.1.3a)$$

$$\frac{dI_{L1}}{dt} = f\rho_1 I_{D1}$$

$$\frac{dS_2}{dt} = -\beta_2 S_2 I_{D1} - \beta_2 S_2 I_{D2} - \beta_2 S_2 I_{D3}$$

$$\frac{dI_{D2}}{dt} = \beta_2 S_2 I_{D1} + \beta_2 S_2 I_{D2} + \beta_2 S_2 I_{D3} - \sigma I_{D2} - f\rho_2 I_{D2} \quad (2.1.3b)$$

$$\frac{dI_{L2}}{dt} = f\rho_2 I_{D2}$$

$$\frac{dS_3}{dt} = -\beta_3 S_3 I_{D1} - \beta_3 S_3 I_{D2} - \beta_3 S_3 I_{D3}$$

$$\frac{dI_{D3}}{dt} = \beta_3 S_3 I_{D1} + \beta_3 S_3 I_{D2} + \beta_3 S_3 I_{D3} - \sigma I_{D3} - f\rho_3 I_{D3} \quad (2.1.3c)$$

$$\frac{dI_{L3}}{dt} = f\rho_3 I_{D3}$$

Here σ is the period of infectiousness usually, 7 days or 1-week, after which the patient recovers and obtains immunity, f is a factor that is used to determine the estimate of Albertans having influenza symptoms, ρ is the weekly proportion of those individuals with symptoms who have been lab confirmed as having influenza and β is the transmission rate. Numerical simulations conducted using the above ordinary differential equations on the data obtained from the Alberta health services predicted the 52nd week with 95% confidence interval (51, 3) week

as where influenza was expected to be the highest for the next influenza season. This will help promote effective planning and allocation of resources within the province [23].

Even though Varughese's influenza model did all the predictions and considered the age groups, the transmission rate for cross infections among each of the age groups was assumed to be the same. For instance, a person from the younger age group (0-18 years) who gets infected with influenza by someone from the working class age group (19-64 years) had the same transmission rate as a person from the younger age group (0-18 years) who gets infected with influenza by someone from the old age group (65 years and over). In addition, the transmission rate measures the probability of transmitting the virus times the number of contacts made with an infected person; however, in Varughese's influenza model, the number of contacts that an individual make was not considered.

2.2 Statistical Models

There have been some methods in time series analysis like Box-Jenkins methods that apply autoregressive moving average models to find the best estimates of any time series model that can be used to forecast infectious diseases like influenza. The aim of most of these approaches is to forecast certain aspects of the influenza epidemic usually the peak time, height, magnitude and spread of the disease.

2.2.2 Wei (2011)

Authors like Qui Wei [20] used spatio-temporal modelling and cross-validated predictions in predicting clusters of influenza cases in Edmonton using real time data that has been collected over time and space from emergency department (ED)

visits by the Alberta Real Time Syndromic Surveillance Net (ARTSSN). Qui Wei uses the spatio-temporal modelling as well as pseudo likelihood estimation to estimate the parameters and also makes use of a cross validation method to validate the model. This method was used to analyze health link (HL) calls (2003-2009) and emergency departments (ED) visit data (2004-2009) as well as school absenteeism reports which were obtained by using the Alberta Real Time Syndromic Surveillance Net (ARTSSN) for the Edmonton area. Qui Wei also examined the geographic spread of influenza based on the Forward Sortation Area of residents' postal codes. Results obtained by [20] from 34,796 ED visits and 25,493 HL calls without using spatial or temporal correlations showed seasonal trend; however, incorporating spatial and temporal correlations improved the models' predictive abilities and was able to detect the peak days. For instance, using 2 weeks of data, the model used by Qui Wei was able to detect the peak days with over 30 influenza-related HL calls per day and 32 influenza-related ED visits per day. However, a problem identified with the temporal and spatial correlations was that in the case of earlier predictions, influenza peaks were not well captured. Also due to it computational complexity, the model could not be easily generalised by researchers in non-mathematical fields [20].

2.3.2 Nsoesie, Marathe, and Brownstein (2013)

Nsoesie et. al. [19], presented a framework for near real-time forecast of influenza epidemics using a simulation optimization approach. The approach consisted of the stochastic individual-based epidemiology model which was used for

simulating influenza-like disease transmission and a simple root finding optimization method for parameter estimation and forecasting that captures any ongoing disease activity. The individual-based model aims to capture the underlying process of disease transmission based on the population contact patterns which makes up the dynamics in the observed epidemic time series curve [19]. The optimization approach on the other hand is used to produce parameter values that capture the trend observed in the data and obtain new parameter values from simulated outcomes of the individual-based model. The individual-based model which consists of the dynamic social contact network and an individualized disease model makes use of the compartmental model SEIR: susceptible, exposed, infected and recovered [19]. Nsoesei et al. presumes that an infected agent moves from one compartment to another through the different transmission states based on defined incubating and infectiousness time periods which are described using discrete probability distributions [19]. The probability of transmission between the susceptible (u) and infectious (v) individuals is given by:

$$p\big(w(u,v)\big) = 1 - (1-r)^{w(u,v)} \qquad (2.2.2.a)$$

where $(w\ (u,v))$ represents the contact duration and r is the disease transmission rate which is defined as per second.

Under the optimization approach, several algorithms are usually used in the parameter search. Nsoesie et al. proposed certain assumptions under which the proposed simulation optimization technique can be used in conjunction with the individual-based model. This will be used to forecast the peak of any ongoing epidemic by minimizing the difference between the cumulative infections for the

ongoing epidemic and simulated instances for the same time period [19]. The data used was obtained from the US Outpatient Influenza-like Illness Surveillance Network (ILINet) which was provided by the Centers for Disease Control and Prevention (CDC). Results obtained for the 2007-2008 and 2012-2013 influenza season showed the true peak observed to be week 20 and week 15 respectively. The simulation approach used by Nsoesie et al. made use of the SEIR compartmental model and is able to predict the peak of influenza. However, the approach didn't consider the different age groups: the at-risk groups like the older and younger age groups.

2.3 Contact Mixing Matrix

Meltzer, Gambhir, Atkins and Swerdlow [17] proposed a method that was used in calculating the contact mixing matrix for an epidemiological model. Their aim was to model 4 epidemic curves built using a simple, deterministic model where the population was divided into 4 age groups. Thus in order to measure the risk of contact and possible onward spread between and within each age group, they developed an approach which they used in deriving the contact mixing matrix for their model. The data used was from a Polymod study that collected contact data from approximately 8000 persons living in the United Kingdom [17]. The contact data, which was separated into 5-year age groups was aggregated into 4 age groups specifically 0-10 years, 11-20 years, 21-60 years and 61 years and over. They constructed the contact mixing matrix based on the assumption that the number of contacts from one age group, for instance, age group A to another age

group, say age group B, should equal the number of contacts in the reverse direction; hence, the contact mixing matrix must be symmetric. According to Meltzer et al., it is important to note that the dimension of the Polymod matrix is, $g = 1, \dots, n$, however because fewer age groups are being considered, say age group, u, then the Polymod contact data can be made to contain narrower age groups, that is, $i = l(f)\ to\ u(f)$.

The approach consists of 4 steps that is outlined below:

Step 1: The contact rate between someone in group i and another in group g can be calculated by using

$$d_{ig} = \sum_{j=l(g)}^{u(g)} \theta_{ij}, \tag{2.3.a}$$

where θ_{ij} is the mixing matrix element of the contact matrix data from UK, $i, j = 1, \dots, m$ refer to the rows and columns respectively and m is the number of age groups in the mixing matrix.

Step 2: If the population in age group i is N_i, then calculate the population-weighted means of each of the elements d to obtain the contact rates between groups f and g. For $f = g$, the calculation is given as

$$e_{ff} = \frac{\sum_{i=l(f)}^{u(f)} N_i d_i}{\sum_{i=l(f)}^{u(f)} N_i} \tag{2.3.b}$$

Step 3: In order to have a correct number of contacts made between each age group for the off-diagonal elements, there is the need to sum them up, that is,

$$Y_{fg} = \sum_{i=l(f)}^{u(f)} N_i d_{ig} \tag{2.3.c}$$

$$Y_{gf} = \sum_{i=l(g)}^{u(g)} N_i d_{if} \tag{2.3.d}$$

Step 4: Due to the assumption of symmetry among the off diagonals, in order for those values calculated in Step 3 to be equal, there is the need to average them before the final mixing matrix elements e_{fg} and e_{gf} can be calculated. The formulae for calculating both can be seen below:

$$Z_{fg} = \frac{(Y_{fg}+Y_{gf})}{2} \tag{2.3.e}$$

$$e_{fg} = \frac{Z_{fg}}{\sum_{i=l(f)}^{u(f)} N_i} \tag{2.3.f}$$

$$e_{gf} = \frac{Z_{fg}}{\sum_{i=l(g)}^{u(g)} N_i} \tag{2.3.g}$$

Thus e_{fg} measures the rate at which an individual in age group f makes contact with an individual in age group g, per unit time and reverse order [17].

The method proposed by Meltzer et al. [17] as noted in this subsection has been adopted in calculating the contact mixing matrix for the three age groups used in this thesis.

Chapter 3

3 Formulation of Influenza Model

3.1 Data Description

The data for this project is obtained from the Provincial Laboratory for Public Health (Provlab), Alberta's Influenza Like Illness (ILI), Sentinel Physician System (TARRANT), Supplemental Enhanced Service Event (SESE), Physician claims, the Pharmacy Information Network (PIN), as well as outbreak reports and hospitalized case report forms from Alberta Health's Communicable Disease Reporting System (CDRS) from the Alberta Health Services.

In this thesis, the data considered was the laboratory confirmed influenza cases obtained for the 2014-2015 influenza season. The data comprised of the laboratory confirmed influenza cases, antiviral dispensing event as well as physician claims which were obtained for each week and age group, specifically 0-18 years, 19-64 years and 65 years and over. The confirmed cases are obtained for 52 weeks, starting from week 35 (August, 2014) and ending on week 34 (August, 2015). The data was obtained in Microsoft Excel spread sheet format. It was formatted and classified according to the various zones: Edmonton, Calgary, South, Central and North zones. A Microsoft Excel line plot of the 2014-2015 laboratory confirmed cases can be seen in Figures 3.1.1, 3.1.2, 3.1.3 for each age groups and Figure 3.1.4 for the combined data.

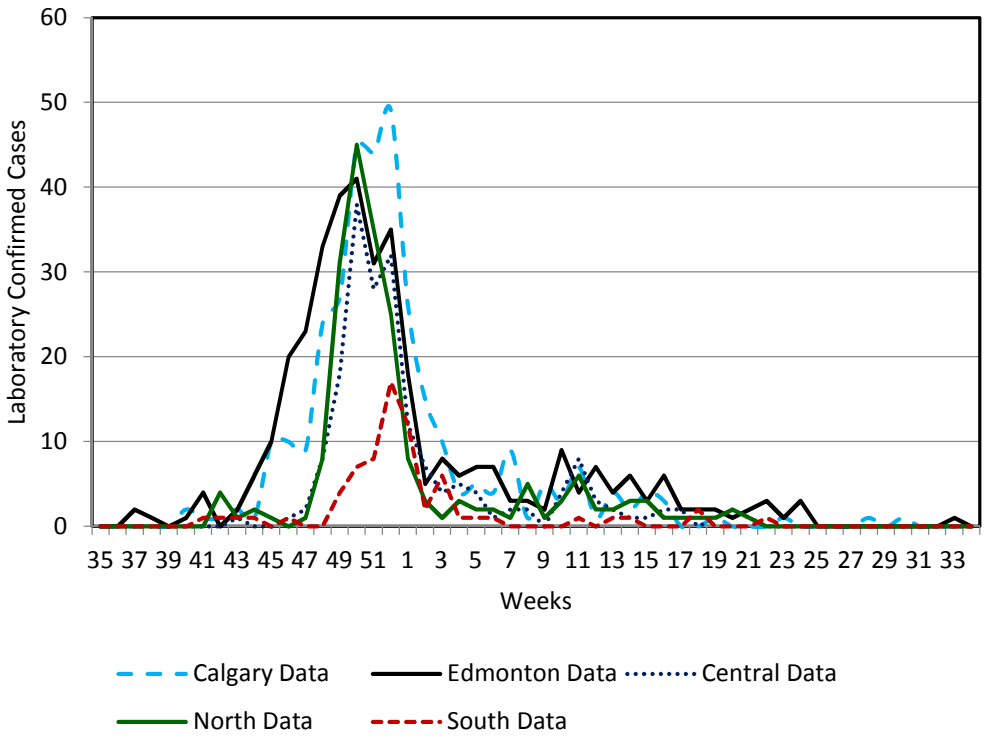

Figure 3.1.1: Laboratory confirmed cases for age group 0 – 18 years

The initial cases of influenza for the 0-18 years' age group were observed in Edmonton at the 37th week followed by Calgary in the 40th week. Edmonton, Central and North zones recorded their highest number of lab-confirmed influenza cases, 41, 38 and 45 cases respectively in the 50th week. Calgary and the South zones recorded their highest lab-confirmed influenza cases 49 and 17 respectively in the 52nd week. The peak period in influenza across the zones was observed in-between the 49th and 1st week, Figure 3.1.1. During the peak period, Calgary had the highest laboratory confirmed cases as compared to all the other zones.

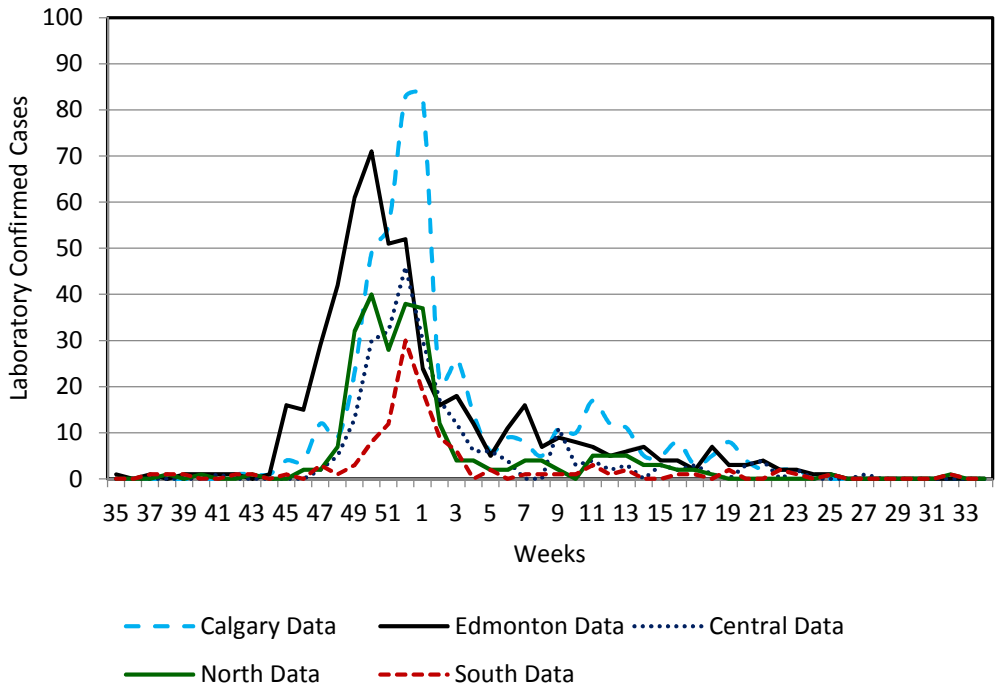

Figure 3.1.2: Laboratory confirmed cases for age group 19 – 64 years

The Edmonton and the North zones with 71 and 40 lab-confirmed influenza cases, respectively, were the first to record their maximum number of influenza cases in the 50th week for the 19-64 years' age group. In the 52nd week the Calgary, Central and South zones recorded their highest lab-confirmed influenza as follows: Calgary - 83, Central - 46, and South - 30.

Similar to the '0-18 years' age group, the peak period in influenza across the zones for the '19-64 years' age group was observed in-between the 49th and 1st week with Calgary recording the highest.

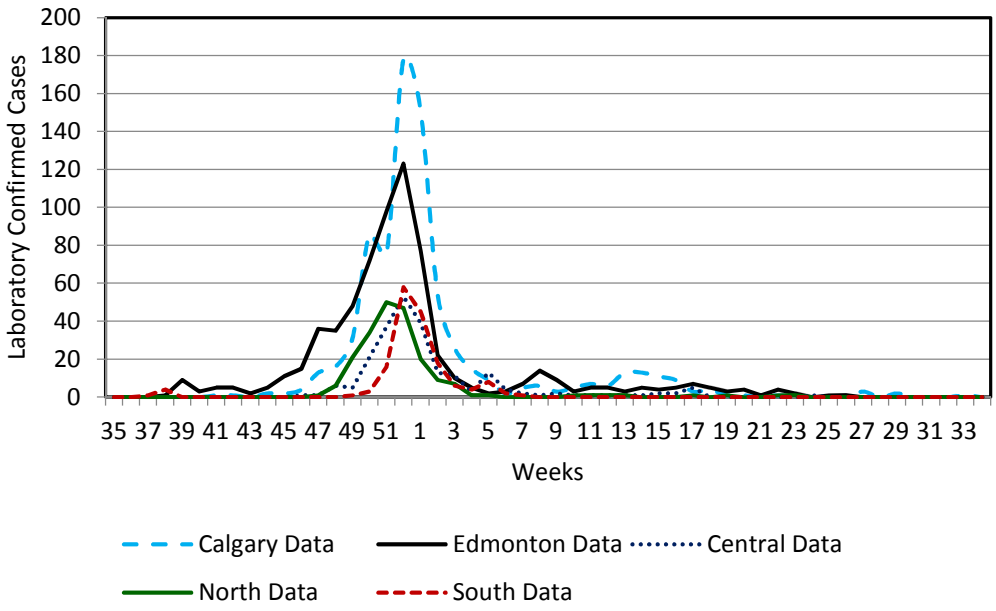

Figure 3.1.3: Laboratory confirmed cases for age group 65 years and over

Calgary, Edmonton, Central and South zones had their highest number of lab-confirmed influenza cases, that is, 178, 123, 53, 58 cases respectively in the 52nd week for the 65 years and over age group. The north zone observed it peak in the 51st week with 50 lab-confirmed influenza cases obtained. Calgary had the highest number of influenza cases followed by Edmonton zone across the zones.

It can be observed from Figures 3.1.1, 3.1.2, 3.1.3 that Calgary recorded the most lab-confirmed influenza cases among the zones. Among the three age groups, the elderly age group (65 years and over) had the highest number of cases. The elderly age groups are the at-risk individuals and they are more prone to being infected with the influenza virus, seen Figure 3.1.4.

From the surveillance report for the 2014-2015 influenza season in Alberta from the Surveillance and Assessment branch of the Alberta Health, there were high

morbidity and mortality among the seniors and the elderly in the long term care and supportive living facilities.

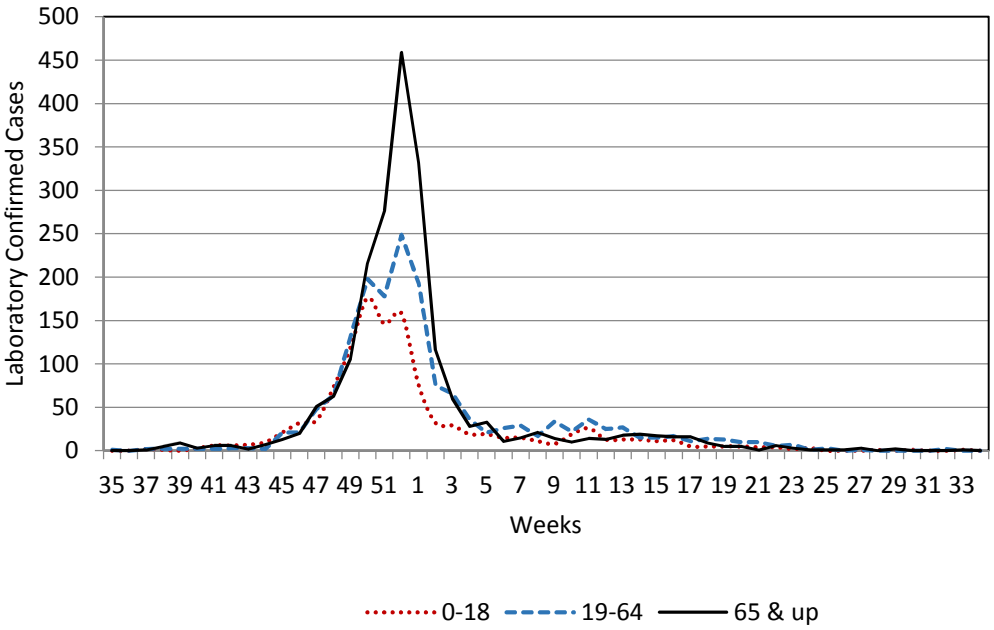

Figure 3.1.4: Laboratory confirmed cases for all age groups and all zones

3.2 The Influenza Model

3.2.1 Introduction

Influenza epidemics can persist and cause harm if left untreated and uncared for. Predicting the peak in influenza cases will not only help in resource allocation but it will help clinicians and policy makers plan ahead as to how to control the spread of the influenza. It will also help them on how to distribute medicines and educate people about the virus. In this section, the transmission mechanism of influenza is modelled using the classic epidemic theory of Kenmarck-McKendrick. The formulated influenza model is an extension of the influenza model proposed by Varughese [23] to predict the peak of influenza in Alberta.

The current model proposed in this thesis considers the asymptomatic patients and the contact mixing matrix that is obtained as a result of a susceptible individual coming into contact with an infected individual which is not considered by other models. It is usually important to consider such factors because they help in the detection of early infections and provide precise estimation.

3.2.2 Derivation of the Influenza Model

The Kenmarck-McKendrick compartmental model that was considered was the Susceptible-Infected epidemic model where the infected population is classified into two parts: those people who are infected with the influenza virus and have been lab-confirmed as having influenza, I_L and those who have been infected with the influenza virus but have not been lab-confirmed as having influenza, I_D. The individuals in the I_D compartment are classified as the asymptomatic individuals. These individuals can also recover before visiting any health departments. These individuals are difficult to account for.

The susceptible compartment comprises individuals that have not been infected with the influenza virus but are prone to be infected with the virus. The infected individuals are those individuals that have the influenza virus and can easily infect someone who does not have.

This is an age distribution model that is classified into three age groups 0-18 years, 19-64 years and 65 years and over. Simulation study will be conducted by assigning different valid values to the parameters of the model. The model is a one strain influenza SI model designed based on the assumption that the

population is a closed population. This implies that births, deaths and migrations are assumed to be negligible since the model runs for only a short period of time. A vaccination compartment is also not included in the model due to the minimal impact that it has on the overall influenza curve. The members of the population also mix homogeneously within each compartment and by age groups (school age, workers and seniors). Another assumption is that an individual is immune after he or she recovers from the virus.

We also assumed that an individual in the susceptible compartment can be infected with the flu virus after coming into contact with an infected individual at a rate of β. The susceptible population on getting the infection moves into the I_D compartment. After been lab-confirmed as having influenza, these individuals move into the I_L compartment at a rate of ρ. Within the I_D compartment, there are those infected individuals who, after a period of time, usually 1 week or 7 days, recover and move out of the compartment at a rate of σ.

f in the model acts as a scale factor to estimate the total number of Albertans showing the influenza symptoms. This number is usually under reported hence we estimate from the model. This is because, people who are sick or infected with the virus often do not go to clinics or the hospitals hence the actual total number of people showing symptoms is difficult to be accounted for. In the model, f multiplies ρ, the weekly proportion of individuals that have been lab-confirmed as having the influenza virus. Cross infections among the three age groups results in an individual from age group i infecting someone within age group i or infecting another individual in age group j. The compartmental structure and flow diagram

of the influenza model, which is an extension of the influenza model proposed by Varughese [23], can be seen in Figure 3.1.5.

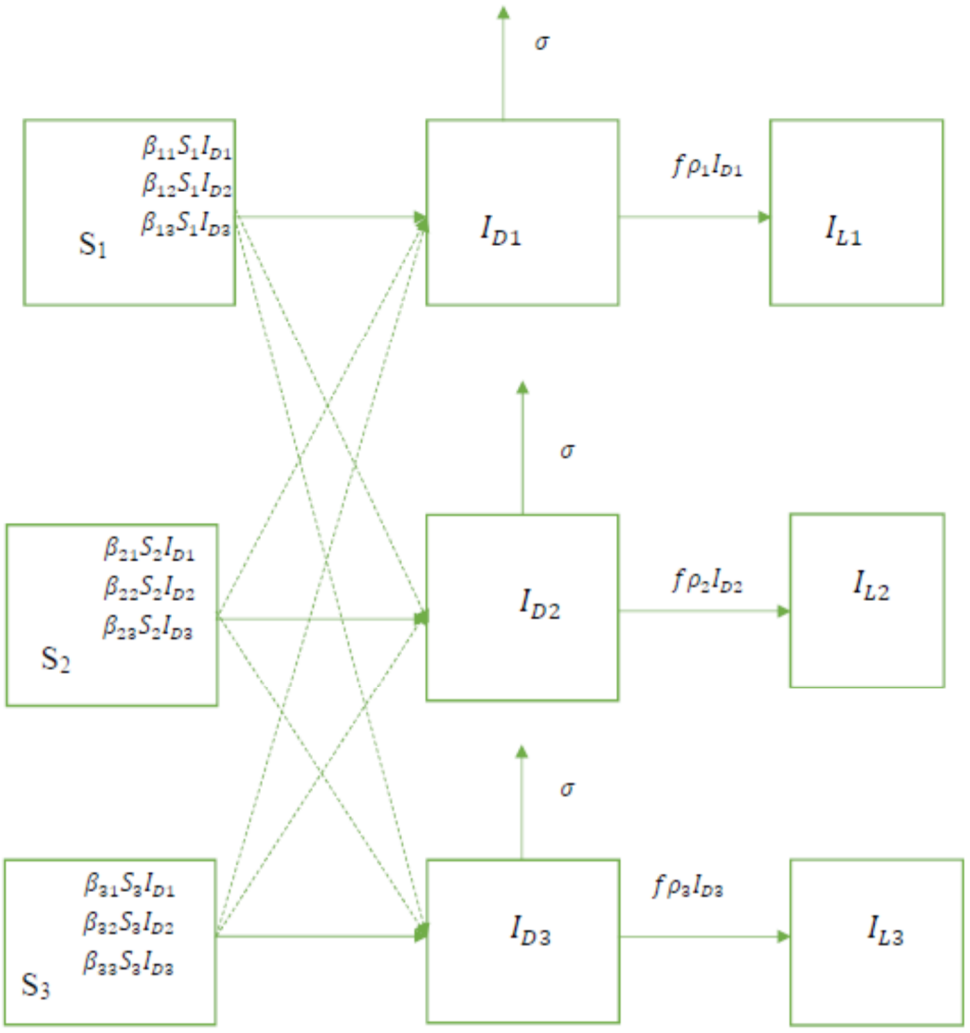

Figure 3.1.5: Flow diagram of the proposed influenza model

From the flow diagram above, it must be noted that the subscript 1, 2, and 3 represent the age groups: 0-18 years, 19-64 years and 65 years and over respectively. The time frame is in weeks. The mathematical formulation of the SI model can be expressed as a system of differential equations as follows:

$$\frac{dS_1}{dt} = -\beta_{11}S_1I_{D1} - \beta_{12}S_1I_{D2} - \beta_{13}S_1I_{D3}$$

$$\frac{dI_{D1}}{dt} = \beta_{11}S_1I_{D1} + \beta_{12}S_1I_{D2} + \beta_{13}S_1I_{D3} - \sigma I_{D1} - f\rho_1 I_{D1} \quad (3.2.2a)$$

$$\frac{dI_{L1}}{dt} = f\rho_1 I_{D1}$$

$$\frac{dS_2}{dt} = -\beta_{21}S_2I_{D1} - \beta_{22}S_2I_{D2} - \beta_{23}S_2I_{D3}$$

$$\frac{dI_{D2}}{dt} = \beta_{21}S_2I_{D1} + \beta_{22}S_2I_{D2} + \beta_{23}S_2I_{D3} - \sigma\, I_{D2} - f\rho_2 I_{D2} \quad (3.2.2b)$$

$$\frac{dI_{L2}}{dt} = f\rho_2 I_{D2}$$

$$\frac{dS_3}{dt} = -\beta_{31}S_3I_{D1} - \beta_{32}S_3I_{D2} - \beta_{33}S_3I_{D3}$$

$$\frac{dI_{D3}}{dt} = \beta_{31}S_3I_{D1} + \beta_{32}S_3I_{D2} + \beta_{33}S_3I_{D3} - \sigma\, I_{D3} - f\rho_3 I_{D3} \quad (3.2.2c)$$

$$\frac{dI_{L3}}{dt} = f\rho_3 I_{D3}$$

Then for i = 1, 2, 3, S_i represents the Susceptible population for the three age groups. I_{Di} represent the individuals or Albertans who have the influenza virus but haven't been lab-confirmed for each age group. I_{Li} represent those individuals or Albertans with influenza who have been lab confirmed as having influenza. A description of the parameters used is given in Table 3.1.1:

Table 3.1.1: Parameter Descriptions and Units used in the model

Parameters	Descriptions	Units
$\beta_{i,j}$	age specific transmission rate of influenza or the rate at which susceptible individual become infected with the influenza virus.	1/time (time in weeks)
f	Factor that estimate the individuals or the Albertans having influenza symptoms.	
σ	measures the period of infectiousness or the period when one recovers from the flu.	time (weeks)
$\rho_i(t)$	measures the weekly proportion of those individuals with symptoms who have been lab confirmed as having influenza.	1/time (time in weeks)

$\beta_{i,j}$ can be classified as the probability of transmitting the influenza virus, defined as β_i , times the average number of contacts that an individual in one age group, that is, age group i has with an individual in another age group, that is, j defined as $c_{i,j}$, $\beta_{i,j}$=$\beta_i * c_{i,j}$. In this thesis, we calculated the contacts rates, $c_{i,j}$,which are measured per unit of time (weekly). However, β_i is what we do not know and have to fit it.

3.3 Derivation of the Contact Mixing Matrix

In order to do simulation studies on the differential equations stated above, we need to derive the contacts that an individual in one age group can make with another individual in another age group. This leads to the derivation of the contact mixing matrix. The contact mixing matrix contains the elements $c_{i,j}$ that represent the average number of contacts an individual in one age group, i makes with

another individual in another age group, that is, j. The contact mixing matrix for each zone was derived using the method proposed by Meltzer et al. in Section 2.3 in Chapter 2. A sample calculation following the steps proposed using the Calgary zone is shown below.

It must be noted that the contact data used was the Polymod contact data from the United Kingdom which was used in calculating the contact mixing matrix by Meltzer et al. [17]. The contact data is shown in Table A.2 in Appendix A. However, the age specific population distribution used was obtained from the Alberta Health Services which is shown in Table A.1 in Appendix A. The steps used in calculating the contact rate is given as follows:

Step 1:

This step considers all the contact data obtained from the Polymod study [17] which is presented in the form of a matrix as seen in Table A.2 in Appendix A. The elements of this matrix show the daily number of contacts that an individual in one 5-year age group makes with another individual in another 5-year age group.

This thesis considers the three age groups 0-18 years, 19-64 years and 65 years and over, however, summing the columns of the 5-year group matrix yields the following age groups 0-19 years, 20-64 years and 65 years and over which can be seen in Table 3.1.2.

Table 3.1.2: Daily number of contacts for the three age groups

	0-19 years	20-64 years	≥65 years
0-4 years	3.2	3.9	0.4
5-9 years	9.4	5.9	0.7
10-14years	10.2	4.6	0.7
15-19years	8.3	6.6	1.1
20-24years	1.9	9.2	0.7
25-29years	2.6	8.9	1.0
30-34years	3.2	8.9	0.9
35-39years	4.4	9.7	1.2
40-44years	3.6	8.8	1.6
45-49years	2.2	8.5	1.2
50-54years	1.5	6.6	1.2
55-59years	1.1	6.1	1.2
60-64years	1.0	4.3	1.5
65-69years	0.5	2.0	1.3
70-74years	0.6	4.0	2.2

Step 2:

The 5-year age population distribution for Calgary which is obtained from the 2014-2015 population estimates from the Alberta Health Service shown in Table A.1 in Appendix A, is also used in calculating the contact matrix. This is to help in calculating the population weighted means. Then the average number of contacts that an individual in one age group can make with another individual in the same age group can be calculated by finding the total number of contacts for that age group and then dividing it by the number of individuals in that age group. for instance, the daily number of contacts between two individuals in the 0-19-year age group can be calculated as below:

$$\frac{3.2*99{,}470+9.4*94{,}689+10.2*84{,}267+8.3*88{,}696}{99{,}470+94{,}689+84{,}267+88{,}696}=7.6$$

A similar calculation done for the 20-64-year age group gave the daily number of contacts to be 8.1.

The daily number of contacts between two members of the > 65 years' age group is calculated to be

$$\frac{0.5*58,041+0.6*37,801}{58,041+37,801} = 1.7$$

Step 3:

This step involves calculating the daily number of contacts for the off diagonals; that is, the daily number of contacts that an individual in one age group, say, 0-19 years, makes with another individual in another age group, say 20-64 years. However, the assumption of symmetry ensures that the total number of contacts between those in age group 0-19 years and 20-64 years is the same as the total number of contacts between those in the 20-64 years and the 0-19 years. However, due to differences in population estimates, we don't expect them to be equal; hence we need to average these two numbers and then find the daily number of contacts made by dividing the average by the sum of the number of individuals in each age group.

For instance, calculating the average number of contacts that an individual in age group 0-19 years makes with another individual in 20-64 years can be done as follows:

- first find the total number of contacts

The total number of contacts that an individual in age group 0-19 years makes with another individual in 20-64 years is given as:

$$1.9 * 104{,}599 + 2.6 * 129{,}283 + 3.2 * 141{,}225 + 4.4 * 126{,}575 + 3.6 * 118{,}704 + 2.2 * 110{,}091 + 1.5 * 113{,}879 + 1.1 * 103{,}098 + 1.0 * 77{,}895 = 2575379.8$$

The total number of contacts that an individual in age group 20-64 years make with another individual in 0-19 years is also given as:

$$3.9 * 99{,}470 + 5.9 * 94{,}689 + 4.6 * 84{,}267 + 6.6 * 88{,}696 = 1919619.9$$

- Then averaging both total contacts above gives:

$$\frac{2575379.8 + 1919619.9}{2} = 2247499.85$$

- Now the daily number of contacts that an individual in 0-19 years make with an individual in 20-64 years is given as:

$$\frac{2247499.85}{104{,}599 + 129{,}283 + 141{,}225 + 126{,}575 + 118{,}704 + 110{,}091 + 113{,}879 + 103{,}098 + 77{,}895} = 2.2$$

- and the daily number of contacts that an individual in 20-64 years make with an individual in 0-19 years is given as:

$$\frac{2247499.85}{99{,}470 + 94{,}689 + 84{,}267 + 88{,}696} = 6.1$$

The same calculations can also be done for the other off diagonal elements.

Hence a table of the daily number of contacts that an individual in one age group makes with another individual in another age group for the Calgary zone can be seen below in Table 3.3.

Table 3.3.3: A contact mixing matrix for each age group for the Calgary zone

	Age Groups		
Age Groups	0-19 years	20-64 years	$\geq$ 65 years
0-19 years	7.6	6.1	0.4
20-64years	2.2	8.1	0.7
$\geq$ 65 years	1.6	7.6	1.7

The above steps were used to calculate the contact mixing matrices for the other zones, specifically, the Edmonton, North, South and Central zones.

It is important to note that although in this thesis the age groups we considered were 0-18 years, 19-64 years and 65 years and over, respectively, the age groups for the contact mixing data after summation is 0-19 years, 20-64 years and 65 years and over; however, we considered the differences to be negligible.

3.4 Statistical Analysis using Nonlinear Regression

Nonlinear regression is an essential form of regression analysis that is used to analyze biological data as well as many other forms of data [16]. It is used to fit data to a specified model where the interest may be in determining the best-fit parameters that define that model. The result obtained from nonlinear model fitting is used in generating a standard curve that is expected to fit our data perfectly well. Nonlinear regression is usually represented in the form

$$Y_n = f(x_n, \theta) + Z_n \tag{3.4.1a}$$

Where f is the expectation function, x_n is a vector of regressor variables or independent variables for the nth case, Y_n is a vector of the dependent variables for the nth case, Z_n is a vector of residuals or noise for the nth case [6].

Nonlinear regression has no close form solution; instead, it uses iterations to determine the solutions of interest. It normally requires a starting value for each parameter before it can iteratively run through the program to select the best estimates. These starting values are usually estimated values you can obtain by observing your data or by assuming constant values for some of the parameters. The linearization method is also another way of determining the starting values for your model. In this thesis, $\rho(t)$ which measures the weekly proportion of lab-confirmed cases for the three age groups, is fitted using a standard normal equation which is chosen based on the assumption that the data points are independent and follows the standard normal distribution. As a result, no starting values were specified; Mathematica automatically generates the starting values and then derives the best-fit parameters for the chosen model. The standard normal equation used to analyze $\rho(t)$ is given as

$$c^2 x^a e^{-bx} \tag{3.4.1b}$$

where *a, b*, and *c* are what we are interested in determining.

We also considered the 90% prediction bands which was constructed for each age group's $\rho(t)$. This was to determine the area expected to contain 90% of all data points if there were additional available data points. Prediction bands are most often wider than confidence bands because they include both the uncertainty in the true position of the curve as well as the scatter of the data around the curve. The parameters, the test statistics and the p-values can also be obtained for the estimates. The p-value shows if the parameters obtained are significant or not significant based on a specified significance level.

Obtaining the parameter estimates of interest is what is termed as nonlinear least squares fitting. The nonlinear least squares fitting involves obtaining a solution or parameters for the model that minimizes the sum of squared errors written as

$$SSE = \sum_{i=1}^{n} {e_i}^2 = \sum_{i=1}^{n} [y_i - \hat{y_i}]^2 = \sum_{i=1}^{n} [y_{curve} - y_{data}]^2 \qquad (3.4.1c)$$

Formulae 3.4.1c is used if we assume that the errors follow a normal distribution with constant or equal variance [16]. However, in the case where the variance is not constant, that is, there is variability in the data points such that the least squares assigns unequal weights, then weighted nonlinear least squares fitting is the method to use [16]. Thus minimizing the sum of squares of the relative distances of points from the curve given as

$$SSE = \sum \left(\frac{y_{curve} - y_{data}}{y_{data}} \right)^2 \qquad (3.4.1d)$$

It is also essential to note that whenever a parameter is fitted to a model, the accuracy or the precision can be expressed as a confidence interval or a prediction interval. Confidence intervals show how well the mean has been determined. The best confidence intervals are the narrower intervals since they show that uncertainty in the true value of the parameter, especially the mean, is smaller as compared to a wider confidence interval showing a larger uncertainty. The width of the confidence interval depends to a large extent on the sample size, the variation of the data, the type of the interval and the confidence level [16]. This is because larger samples tend to give more precise estimates, hence narrower confidence intervals, than smaller samples.

Hence, after determining the best fit parameters after running a sample of 500 000, these parameters are then used to determine the 95% confidence intervals and the mean. The mean is then chosen to be used for the predictions.

3.5 Description of the Simulation code used in the analysis

The simulation code used for analysing the laboratory confirmed data obtained from the Alberta Health Services was written in Mathematica. It consisted of finding $\rho(t)$, the best fit parameters, β, and also the 95% confidence intervals to ascertain if the parameters and mean obtained are the best estimates for predictions. The 95% confidence interval is obtained using the Metropolis hasting algorithm of the Markov Chain Monte Carlo Method. Latin hypercube sampling method is then used to generate 10000 sample points based on the point estimate and their 95% confidence interval. The peak is then recorded 10000 times and the minimum and maximum peak is calculated from the resulting distribution.

$\rho(t)$ measures the weekly proportion of the individuals who have been lab-confirmed as having influenza either through emergency departments visits or general practitioners' visits and antiviral dispensing. $\rho(t)$ is measured by dividing the total laboratory confirmed cases for each zone by the total physician visits from physician's claims and antiviral dispensing. In the simulation code, a nonlinear model fit in the form of $c^2 x^a e^{-bx}$ where a, b, c are constants is fitted to the data points for each age group and then plotted. This gave a perfect fit to the data points although there was variability observed in the data points.

The force of infection, $\beta_{ij}S_iI_{Dj}$, where $\beta_{ij,}i,j = 1,2,3,\ldots$ is the transmission rate which consist of the contact rate, c_{ij} and the probability of transmission, β_i is also fitted. However, in order to fit the transmission rate, the contact rates for each zone was obtained using the steps proposed by Meltzer et. al. [17]. To obtain the best-fit parameters, that is the probability of transmission for each age group, the parameters are randomly generated using RANDOMREAL in Mathematica, where the parameters are chosen between a range of selected values from a specified sample. These generated numbers are then applied to the laboratory confirmed data and the ordinary differential equations to obtain the best fit parameters for both the transmission parameters and the scale factor, f. Using nonlinear least squares fitting, the best-fit parameters that minimizes the sum of squares errors; that is, the sum of squares of the vertical distances between the curve and the data divided by the average of the data, given as equation (3.5.1a) is fitted.

$$\sum\left(\frac{Curve-data}{mean(data)}\right)^2 \tag{3.5.1a}$$

This acts as a weighting scheme to determine the best fit parameters. This is used to compensate for the variability in the data.

To ascertain if the parameters obtained are the best fit parameters, 95% confidence intervals are obtained using the Metropolis hasting algorithm of the Markov Chain Monte Carlo proposed by Gbasemi et al. [5]. The M-H algorithm is iteratively used to generate samples such that as more and more samples are being generated, the distribution of these values will generally approximate the desired distribution. According to Gbasemi et al., M-H algorithm provides a scheme that

is helpful in generating random samples from any desired posterior distribution $p(\theta|y)$. These random samples can then be used to approximate the posterior distribution from which the unknown parameters (θ) can be determined by using the minimum mean squared error which estimate the parameters by the mean or mode of the posterior distribution $p(\theta|y)$ (Gbasemi et al., 2011). The posterior distribution is given by the formula

$$p(\theta|y) = \frac{p(y|\theta)\, p(\theta)}{\int p(y|\theta)\, p(\theta) d\theta} \tag{3.5.1b}$$

where $p(y|\theta)$ estimates the likelihood function, $p(\theta)$ is the prior distribution, θ is the parameter estimate and y is the data. The M-H algorithm of the MCMC method is used to determine the posterior distribution numerically instead of equation (3.5.1b) due to the nonlinearity of the differential equations obtained, which makes determining the closed form solutions of $p(y|\theta)$ as well as the estimates analytically difficult (Gbasemi et al., 2011) .

Generally, at each iteration, the M-H algorithm selects a suitable candidate for the next sample based on the current sample and then, at a specified probability, the proposed candidate is either accepted (and used in the next iteration) or rejected (and the current sample is used instead in the next iteration).

The M-H algorithm proposed by Gbasemi et al. [5] used in this thesis includes the following steps:

Step 1:Take the parameter sample θ^i obtained in the ith iteration. This step entails starting with a random parameter value which is usually from the variable's prior distribution.

Step 2: Draw θ^* from the proposal distribution $q(\theta^*|\theta^i)$ as a proposed sample. For this step, all that needs to be done is to obtain a new sample θ^* which acts as a candidate sample from a proposal distribution which can be Gaussian, Gamma or Poisson. The gamma distribution is however chosen as the proposal distribution with the assumption that all the parameters are positive [5].

Step 3: Calculate the acceptance probability

$$\alpha = \min\left\{1, \frac{p(y|\theta^*)p(\theta^*)q(\theta^i|\theta^*)}{p(y|\theta^i)p(\theta^i)q(\theta^*|\theta^i)}\right\} \tag{3.5.1c}$$

where the likelihood function $p(y|\theta)$ is given as

$$p(y|\theta) = \frac{\left(\frac{\beta_3}{2}\right)^{\frac{\eta_3}{2}}\Gamma\left(\frac{m+\eta_3}{2}\right)}{(2\pi)^{\frac{m}{2}}\Gamma\left(\frac{\eta_3}{2}\right)}\left(\frac{M.E+\beta_3}{2}\right)^{-\frac{m+\eta_3}{2}} = \left(\frac{M.E+\beta_3}{2}\right)^{-\frac{m+\eta_3}{2}} \tag{3.5.1d}$$

and $M.E$, which is the model error, is also given as

$$M.E = \sum\left(\frac{Curve-data}{mean(data)}\right)^2 \tag{3.5.1e}$$

and $$\left(\frac{q(\theta^i|\theta^*)}{q(\theta^*|\theta^i)}\right) = \left(\frac{\theta^i}{\theta^*}\right)^{2\eta_1-1} e^{\frac{1}{\beta_1}\left(\frac{\theta^*}{\theta^i}-\frac{\theta^i}{\theta^*}\right)} \tag{3.5.1f}$$

Step 4: Then draw a random sample U (0,1) such that

$$\theta^{i+1} = \begin{pmatrix} \theta^* & \text{if} \quad U \le \alpha \\ \theta^i & otherwise \end{pmatrix} \tag{3.5.1g}$$

Steps 3 and 4 show the acceptance probability that will be used to determine whether to accept or reject the proposed (candidate) sample. The min ensures that the acceptance probability, α, is never greater than 1, if any random number drawn from U(0,1) is less than or equal to α, then the sample proposal (candidate) will be accepted, that is, $\theta^{i+1} = \theta^*$ is accepted with a probability of α. On the other hand, if the candidate is rejected, then we set $\theta^{i+1} = \theta^i$. The above step is

used to obtain the confidence intervals of the parameters by excluding the 2.5% from each end of the posterior distribution.

The peak is then obtained from the fitted parameters by determining the week with the most laboratory confirmed cases. This is done by first obtaining a list of all the weekly laboratory confirmed cases, $f\rho I_D$, for all the age groups and then determining the week where the largest number of influenza cases is obtained. Latin hypercube sampling method is then used to generate 10000 sample points based on the point estimate and their 95% confidence interval. The peak is then recorded 10000 times and the minimum and maximum peak is calculated from the resulting distribution. Latin hypercube sampling is usually used to sample random numbers that attempts to distribute samples evenly [15]. This is to check the uncertainty in the peak week obtained.

Chapter 4

4 Numerical Investigations

4.1 Results from Numerical Investigations

From the influenza model obtained in Section 3.3 in Chapter 3, first order ordinary differential equations were obtained which were used to estimate the unknown parameters of the model. These equations, alongside the laboratory confirmed data obtained from the Albertan Health Services, were employed in the simulation analysis described in section 3.5 in chapter 3. It is essential to note that the more samples you run, the better the estimates that will be obtained; hence in this thesis we chose to run the code with a sample size of about 500,000. Since the results obtained for the Edmonton and South zone are similar to the others, this thesis present results for only the Central, Calgary and North zones.

4.1.1 Results from Numerical Investigations for Central Zone

Before the parameters were fitted, $\rho(t)$, which measures the weekly proportion of individuals that have been lab-confirmed either through emergency department or general practitioner's office or antiviral dispensing departments as having influenza, is fitted to a nonlinear function. This is to determine if the nonlinear function can generate a smooth standard curve that can interpolate the 52 data points. The intent is to determine the parameters, a, b, c for each age group and then determine if the chosen nonlinear function actually produces a curve that is

smooth enough as well as comes close to our data. The parameters obtained are shown in Table 4.1.1.

Table 4.1.1: Parameter estimates for $\rho(t)$ for each age groups.

Parameter estimates	0-18 years	19-64 years	65 years and over
a	5.5445	3.89014	5.25331
b	0.291892	0.173557	0.251612
c	0.00234067	0.0051543	0.00240244

Hence Figures 4.1.1, 4.1.2, 4.1.3 show the results obtained as a result of plotting the fitted nonlinear function to $\rho(t)$ obtained for the three age groups 0-18 years, 19-64 years and 65 years and over.

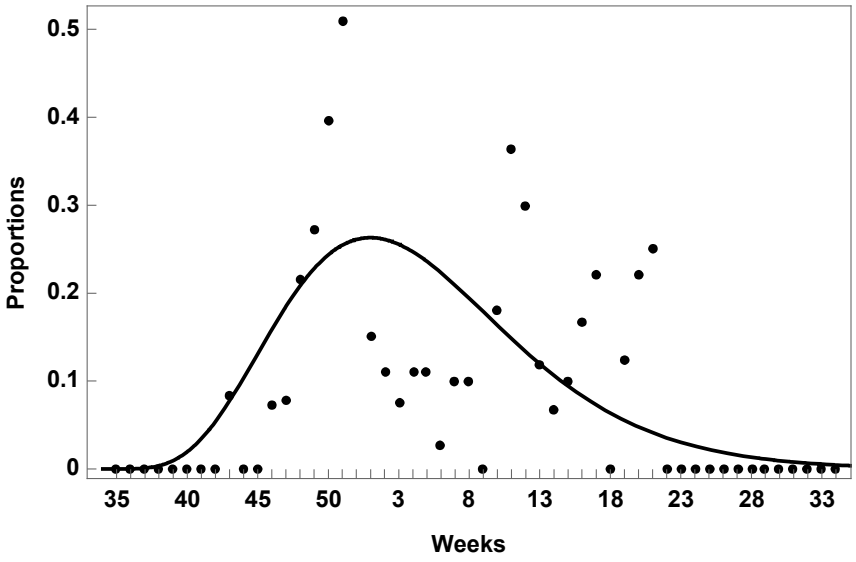

Figure 4.1.1: $\rho(t)$ for the 0-18 years age group.

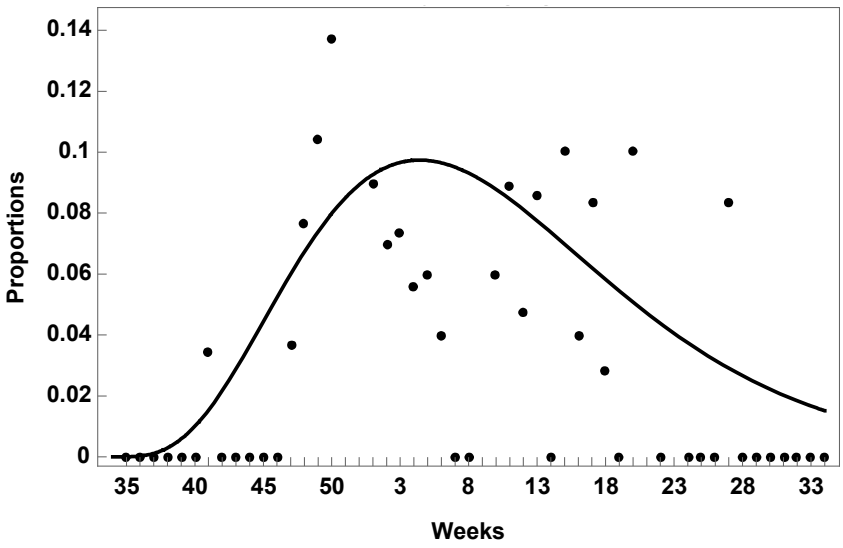

Figure 4.1.2: $\rho(t)$ for the 19-64 years age group.

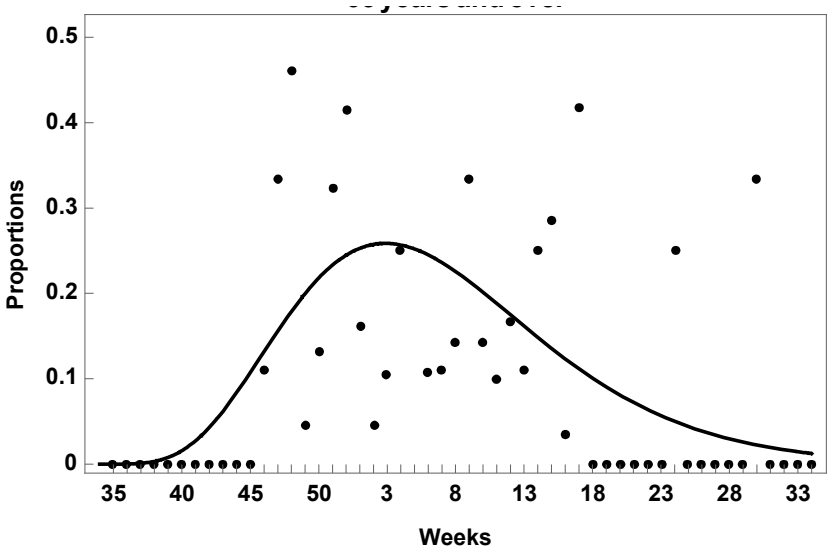

Figure 4.1.3: $\rho(t)$ for the 65 years and over age group.

Figures 4.1.1 to 4.1.3 show the $\rho(t)$ for the three age groups and it can be observed that the curves fit the data points well. The curves depict an increasing trend with one or two cases at the initial start and then a maximum is reached after which it decreases again. However, a look at the actual data shows few visits at

the start of the season but as time elapses, more visits were observed. The trend is typical of other epidemics where a limited number of cases are observed at the initial stage but as time elapses there is an sharp or steady increase in the number of cases which then declines at the end. It can also be observed that although there is variability in the dataset, that is, the average scatter among the data varies, they are more clustered in the middle. The elderly (64 years and over) and younger age group (0-18 years) seems to have had more lab-confirmed cases than the working age group as can be observed from the graph, where the highest proportions for both age group was about 50% as compared to 14% for the working age group (19-64 years).

Figures 4.1.4, 4.1.5 and 4.1.6 shows the 90% prediction bands for $\rho(t)$ obtained for the three age groups 0-18 years, 19-64 years and 65 years and over respectively.

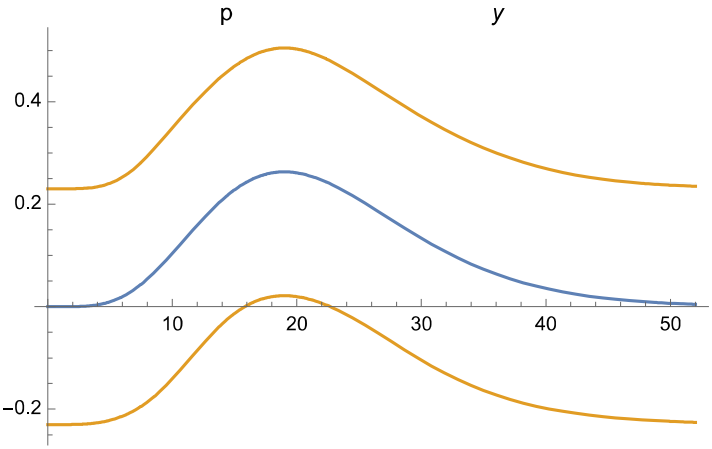

Figure 4.1.4: Prediction bands for 0-18-year age group.

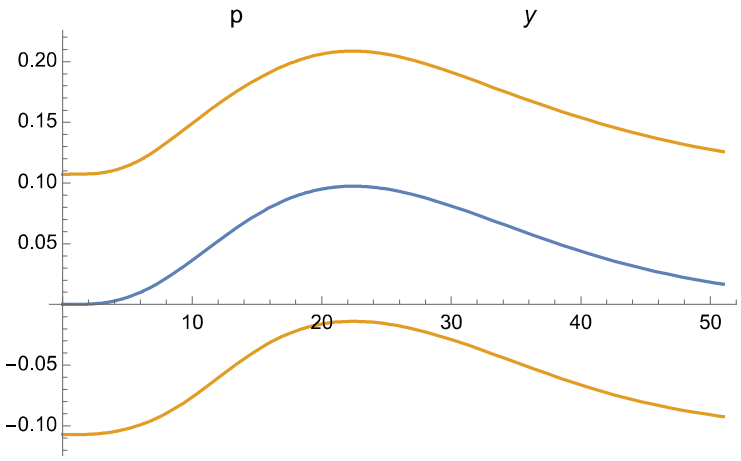

Figure 4.1.5: Prediction bands for 19-64-year age group.

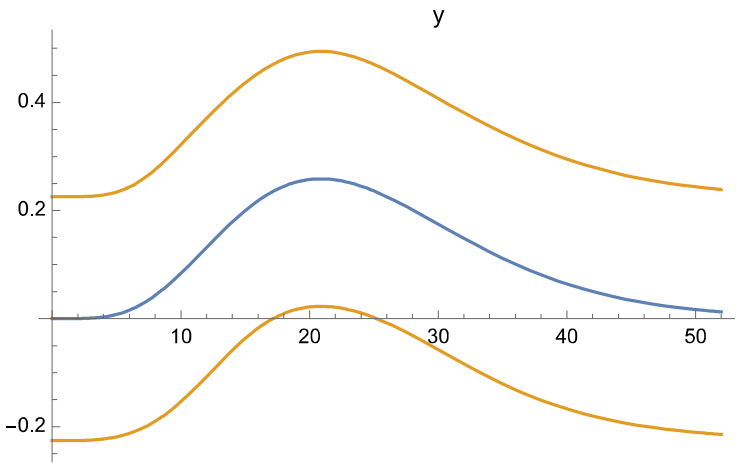

Figure 4.1.6: Prediction bands for 65 years and over age group.

The prediction bands are shown by the cream-colored curves with the actual or true best-fit curve shown by the blue curve. The prediction bands, which show the scatter of the data, mean that if more data points were to be obtained for $\rho(t)$ for each age group, then we should expect 90% of these data points to fall within these bands. It can also be observed that the prediction bands are wider because they have to account for the uncertainty of the curve as well as the scatter of the data.

The initial conditions in Table 4.1.2 shows the 2014-2015 population estimates for the Central zone which we obtained from Alberta Health Services website with $I_{Di} = 1$ and $I_{Li} = 0,\ i$=1,2,3 .

Table 4.1.2: 2014-2015 Population estimates for Central zone.

Parameter	Estimates
S_1	118444
S_2	289306
S_3	65332

The contact mixing matrix obtained for the Central zone is shown in Table 4.1.3:

Table 4.1.3: Contact mixing matrix for Central zone.

Age Groups			
Age Groups	0-19 years	20-64 years	$\geq$65 years
0-19 years	7.7	5.6	0.4
20-64years	2.3	8.0	0.8
$\geq$65 years	1.5	6.0	1.7

It can be observed that the working group (19-64-year age group) make more contacts with themselves than the elderly age group. This might be due to the fact that the elderly age group are often more confined to a fixed location for example, supportive living facilities, as compared to the working group who come into contact with people either at work, at eating places or at recreational centers etc. The younger age group also makes more contact within itself and this is usually at schools, at sleepovers etc. We also have the case where someone from the younger age group who is suffering from the influenza can infect his or her mom or grandma through visits or staying together. All these factors contribute greatly to the contacts that an individual can make.

Next, we obtained the fitted parameters with their 95% confidence intervals; this is done using the weighted nonlinear least squares method in Mathematica with the ordinary differential equations obtained from the influenza model as well as the metropolis hasting algorithm of the MCMC on the 2014-2015 laboratory confirmed cases for the Central zone. This is shown in Table 4.1.4.

Table 4.1.4: Parameter estimates for Central zone.

Parameters	Estimates	95% Confidence interval
β_1	$3.88917 * 10^{-7}$	(3.69598*10^{-7}, $4.08436 * 10^{-7}$)
β_2	6.76495* 10^{-7}	(6.64212*10^{-7}, $6.89996 * 10^{-7}$)
β_3	2.11598*10^{-6}	(1.8696*10^{-6}, $2.37174 * 10^{-6}$)
f	0.0146828	(0.000459573, 0.0153642)

The parameter estimates obtained above are as a result of running about 500 000 samples. These estimates are the mean estimates and it can be observed that they fall within the 95% confidence intervals obtained. The 95% confidence intervals are narrower, hence the uncertainty in the parameter estimates is smaller. Although the 95% confidence interval for f seems to be a little bit wider but we will use it in our predictions.

Plugging the estimates into the simulation code to determine the week where we should expect the highest or peak influenza cases for the Central zone, we obtained the results shown in Figure 4.1.7.

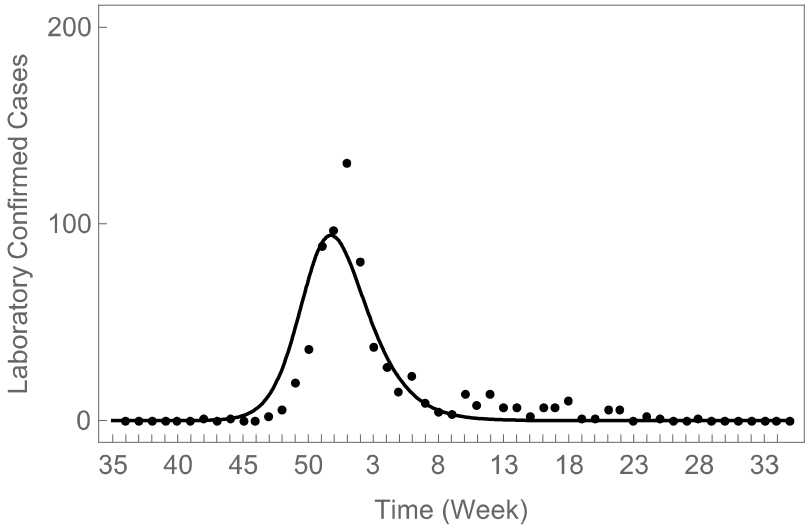

Figure 4.1.7: Laboratory confirmed cases for the Central zone.

Figure 4.1.7 shows the weekly laboratory confirmed cases for the Central zone. The dotted points show the actual laboratory confirmed data obtained from the Alberta Health Services. It can be observed from Figure 4.1.7 that even though the prediction curve predicts an earlier rise in the cases as compared to the actual data, the model predicted the 52nd week to be the peak week where the most cases of influenza were to be expected with a 95% confidence interval of (51, 1) week. The highest number of laboratory confirmed cases predicted by the model was about 100, although the actual data showed the highest number to be about 131. Hence the model does the predictions fairly well. Also, the actual data obtained shows the 1st week as where most cases of influenza was obtained, that is, the actual peak, as compared to the 52nd week predicted by the model. The best-fit curve fits the data well although a few of the data points lie outside the curve.

The accumulated laboratory confirmed cases for each of the age group was also plotted. This can be seen in Figures 4.1.8, 4.1.9 and 4.1.10, respectively:

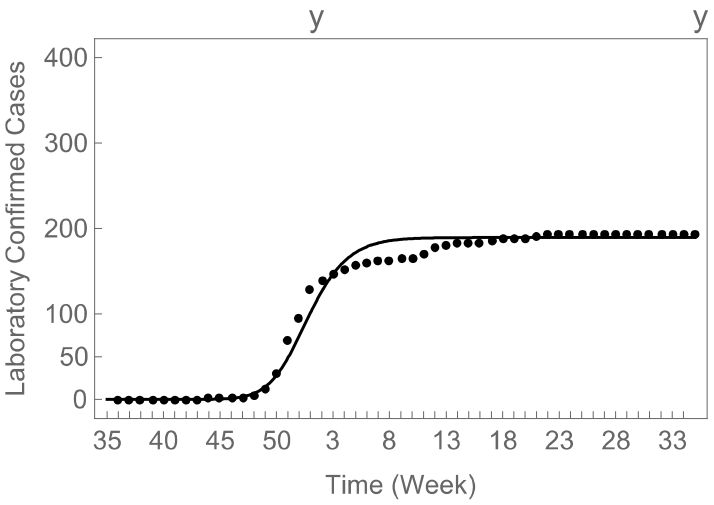

Figure 4.1.8: Accumulated laboratory confirmed cases for 0-18 years.

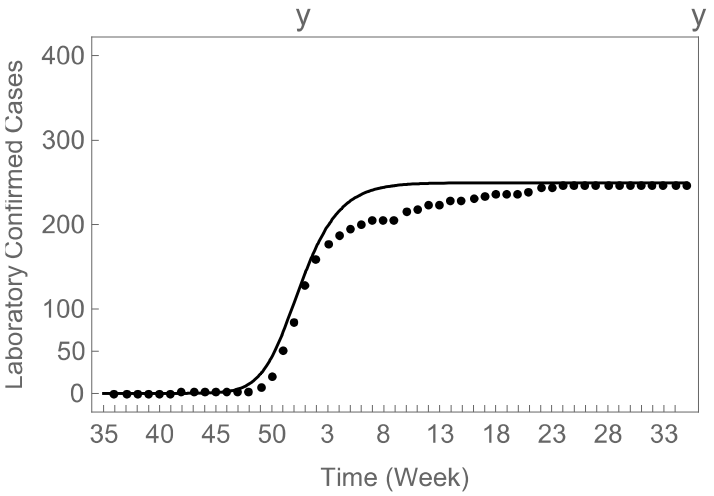

Figure 4.1.9: Accumulated laboratory confirmed cases for 19-64 years.

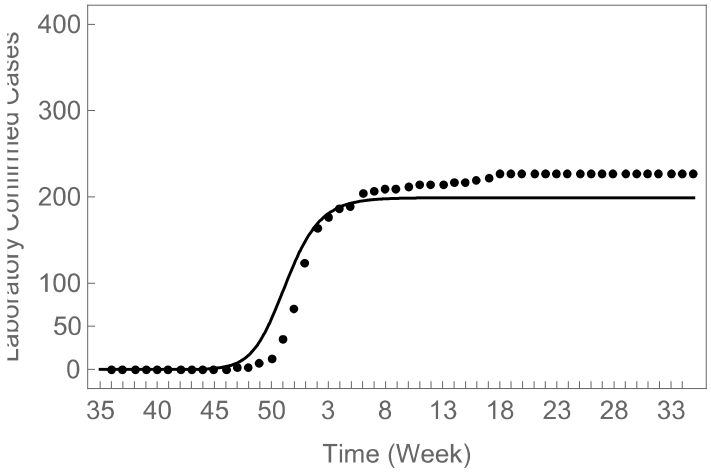

Figure 4.1.10: Accumulated laboratory confirmed cases for 65 years and over.

Figures 4.1.8, 4.1.9 and 4.1.10 show the accumulated laboratory confirmed cases for the three age groups, 0-18 years,19-64 years and 65 years and over. The dotted points are the actual accumulated cases and our model is able to fit this data well. These accumulated cases are obtained by summing up the cases week by week which results in the rise in the cases. The elderly and the working age group had more accumulated laboratory confirmed cases as compared to the younger age group.

4.1.2 Results from Numerical Investigations for North Zone

Similar to the Central zone, the contact mixing matrix for North zone was calculated using the method proposed by Meltzer et al. (2015) which was presented in Section 3.3 of Chapter 3. The contact mixing matrix obtained for the North zone is presented in Table 4.2.1.

Table 4.2.1: Contact mixing matrix for North zone.

Age Groups			
Age Groups	0-19 years	20-64 years	$\geq$65 years
0-19 years	7.6	5.5	0.4
20-64years	2.3	8.1	0.7
$\geq$65 years	2.0	8.1	1.7

It can be observed from Table 4.2.1 that the working group (19-64-year age group) makes more contacts with people within their age group as well as with the elderly age group. The elderly age group on the other hand makes less contact with people of their age as well as with the other age groups as compared to the other age groups. This might be a result of being confined, in for instance, supportive living facilities, homes etc. The younger age group also makes more contact within themselves as well as with others and this is usually at schools, at sleepovers as well as coming into contacts with people like teachers etc.

From the 2014-2015 population estimates obtained for the North zone, the initial conditions used are shown in Table 4.2.2 with $I_{Di} = 1$ and $I_{Li} = 0,\ i$=1,2,3 .

Table 4.2.2: 2014-2015 Population estimates for North zone.

Parameter	Estimates
S_1	130876
S_2	310039
S_3	43474

The fitted parameters with their 95% confidence interval are shown in Table 4.2.3. This is as a result of applying the weighted nonlinear least squares method in Mathematica with the ordinary differential equations obtained from the

influenza model as well as the metropolis hasting algorithm of the MCMC on the 2014-2015 laboratory confirmed cases for the North zone.

Table 4.2.3: Parameter estimates for North zone.

Parameters	Estimates	95% Confidence interval
β_1	$3.80171 * 10^{-7}$	$(3.20292 * 10^{-7},\ 4.43698 * 10^{-7})$
β_2	$5.9343 * 10^{-7}$	$(5.69809 * 10^{-7},\ 6.26682 * 10^{-7})$
β_3	$1.2105 * 10^{-6}$	$(8.78672 * 10^{-7},\ 1.63766 * 10^{-6})$
f	0.0218228	(0.0193009, 0.0243848)

The parameter estimates obtained above are a result of running about 500 000 samples. These estimates are the mean estimates and it can be observed that they fall within the 95% confidence intervals obtained. The 95% confidence intervals are narrower hence the uncertainty in the parameter estimates is smaller hence we will use it in our predictions. But before we get on with the predictions, we first have to also analyze $\rho(t)$ for the North zone.

The results, after running the 52 data points for $\rho(t)$, the weekly proportion of laboratory confirmed cases for the North zone, show the selected nonlinear least squares function to fit the data well as it passes through two or more points for each age group. Also the data points are widely scattered; that is, variability is observed in the data sets although it is mostly clustered in the mid-section. From Figures 4.2.1, 4.2.2 and 4.2.3, the elderly and younger age group had the highest number of visits as compared to the working age group.

The parameter estimates obtained for the nonlinear function used in analysing $\rho(t)$, are shown in Table 4.2.4.

Table 4.2.4: Parameter estimates for $\rho(t)$ for each age groups.

Parameter estimates	0-18 years	19-64 years	65 years and over
a	4.68004	5.50203	4.0924
b	0.271735	0.283306	0.196978
c	0.00567141	0.00137066	0.00648688

Then the results obtained for the nonlinear function is plotted and shown below.

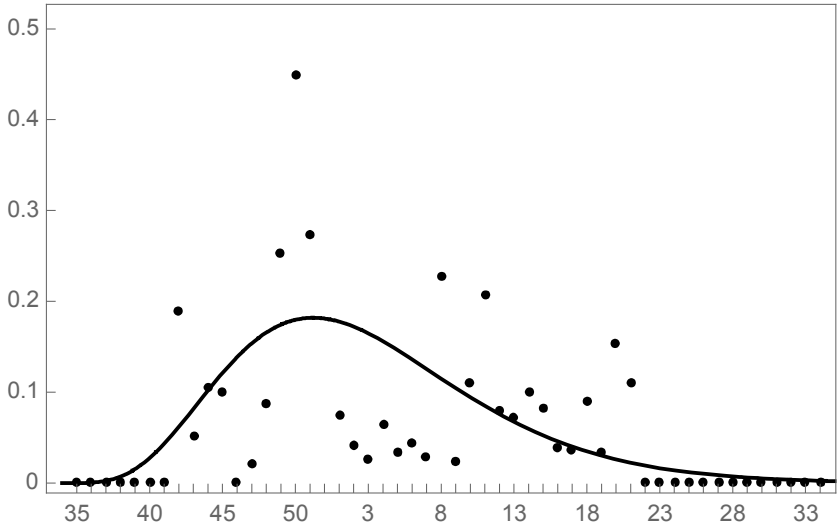

Figure 4.2.1: $\rho(t)$ for the 0-18 years age group

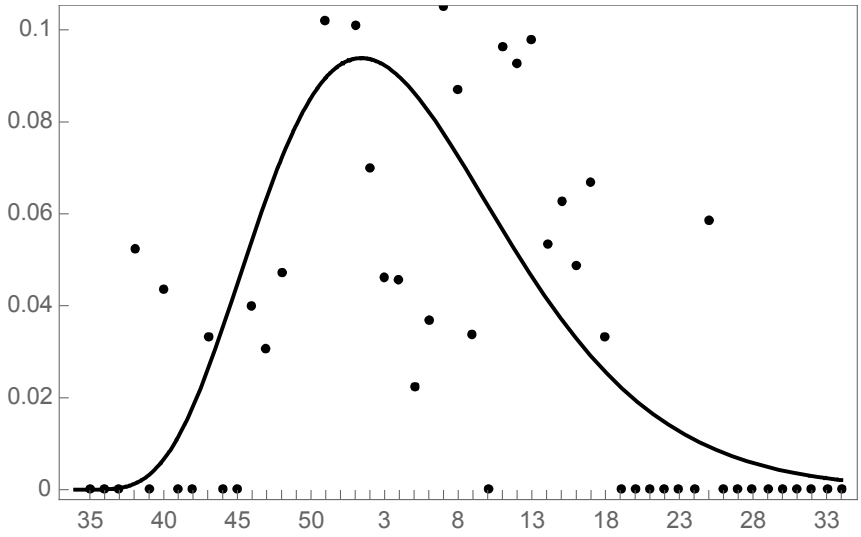

Figure 4.2.2: $\rho(t)$ for the 19-64 years age group

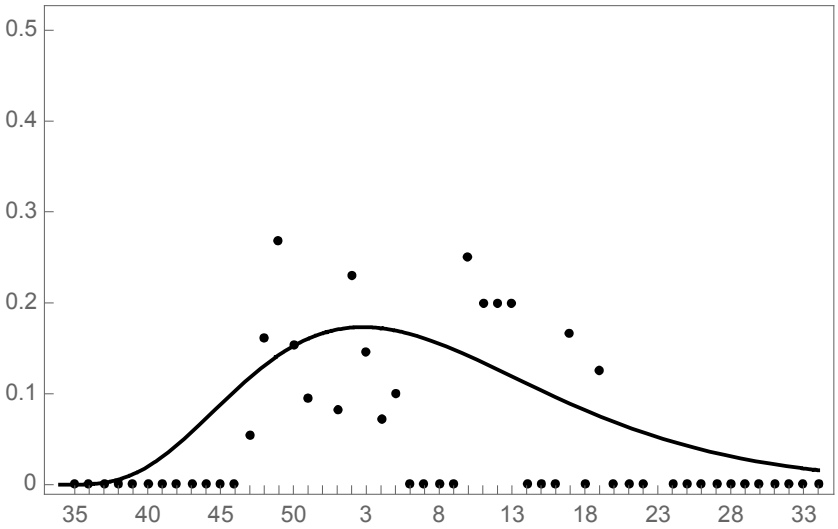

Figure 4.2.3: $\rho(t)$ for the 65 years and over age group.

Figures 4.2.4, 4.2.5 and 4.2.6 shows the 90% prediction bands for $\rho(t)$ obtained for the 0-18 years, 19-64 years and 65 years and over respectively.

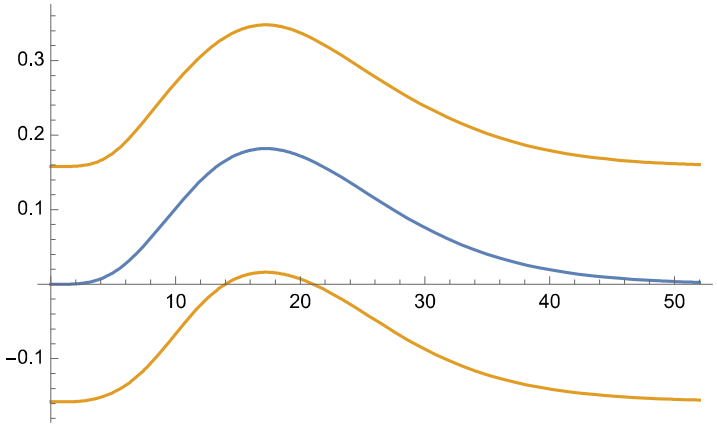

Figure 4.2.4: Prediction bands for 0-18 years' age group.

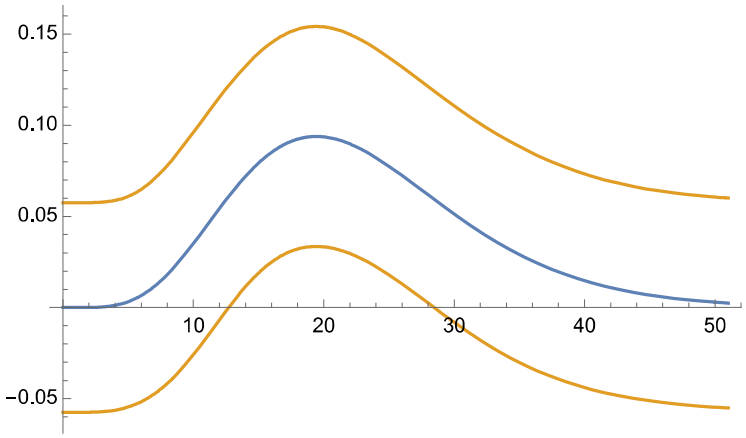

Figure 4.2.5: Prediction bands for 19-64 years' age group.

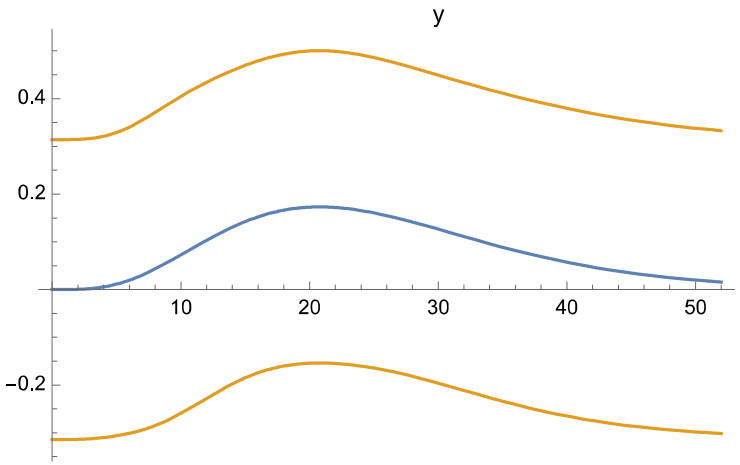

Figure 4.2.6: Prediction bands for 65 years and over age group.

The prediction bands are shown by the cream-colored curves with the actual or true best-fit curve shown by the blue curve. The prediction bands shows that if more data points were to be obtained for $\rho(t)$ for each age group, then we should expect 90% of these data points to fall within these bands. It can also be observed that the prediction bands are wider because they have to account for the uncertainty of the curve as well as the scatter of the data. The bands for the elderly age group appear flatter because the true curve also appears flatter.

From using the parameter estimates obtained in Table 4.2.3 on all the 52 weekly data points, it was established that the week where the highest number of cases of influenza was observed was the 1st week with 95% confidence interval (52, 4) week. This prediction is later than the actual week where the most lab-confirmed influenza cases was observed for the 2014-2015 influenza season which was between the 51st and 52nd week, as can be observed from the actual data. This is shown in Figure 4.2.7.

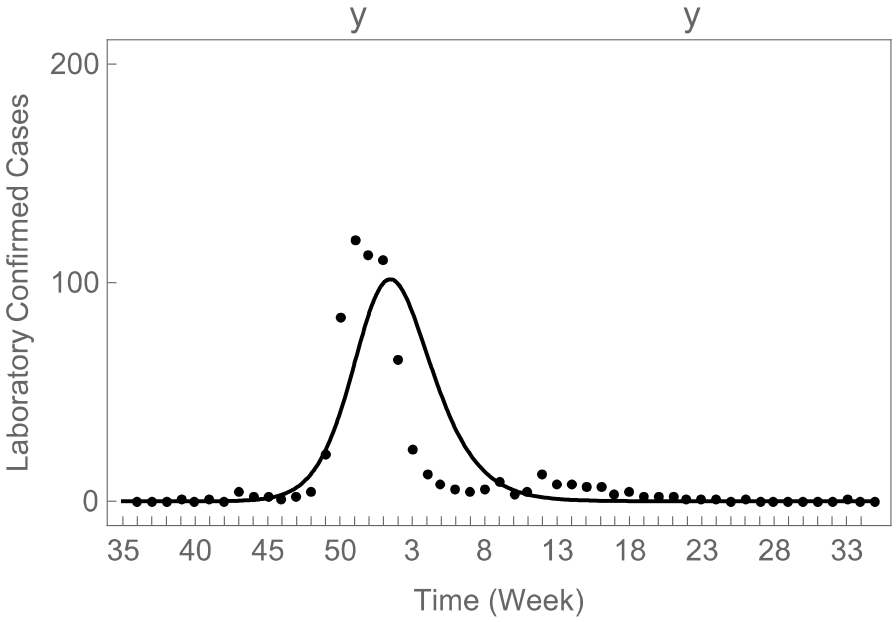

Figure 4.2.7: Laboratory confirmed cases for the North zone.

Figure 4.2.7 shows the weekly laboratory confirmed cases for the North zone where it can be observed that our influenza model is able to fit the actual data well since it was able to pass close to most of the data points. The number of laboratory confirmed cases that our model predicts is about 98 cases however the actual highest influenza case obtained was about 113. The model predicts the 1st week with 95% confidence interval (52, 4) week as the peak week. However, the

actual peak obtained from plotting the laboratory confirmed cases was between the 51st and 52nd week.

The weekly laboratory confirmed cases are then accumulated for each age group to determine the total number of laboratory confirmed cases obtained for the 2014-2015 influenza season. This is shown in Figures 4.2.8, 4.2.9, 4.2.10.

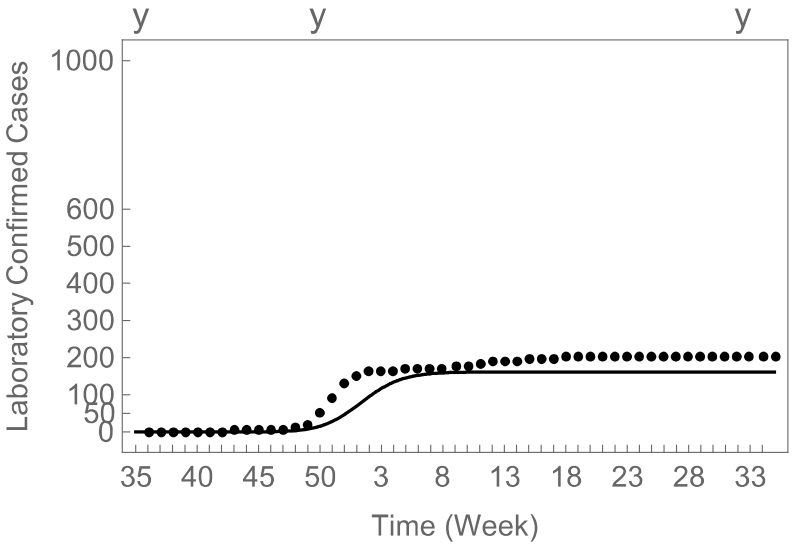

Figure 4.2.8: Accumulated laboratory confirmed cases for 0-18 years.

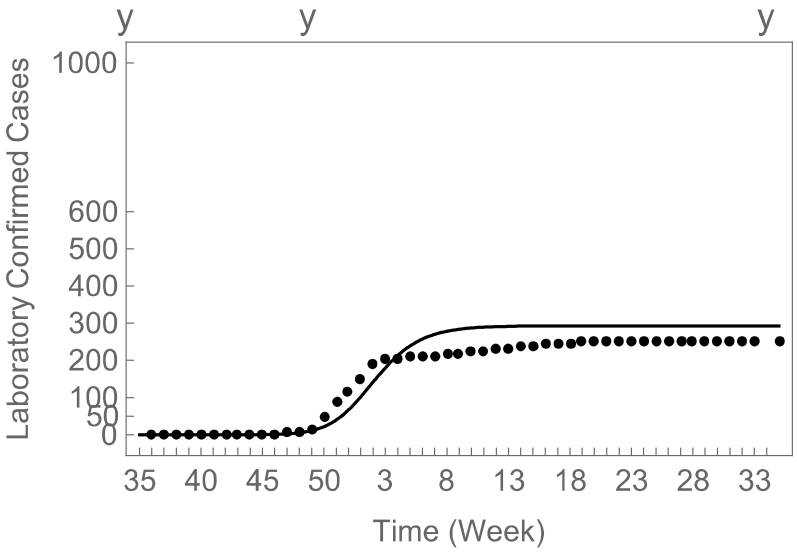

Figure 4.2.9: Accumulated laboratory confirmed cases for 19-64 years.

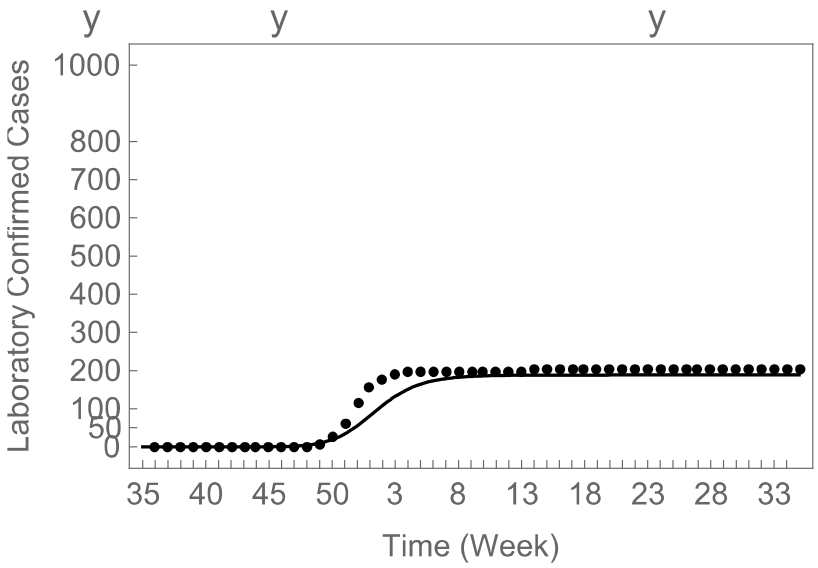

Figure 4.2.10: Accumulated laboratory confirmed cases for 65 years and over.

The above diagrams show the accumulated laboratory confirmed cases for the age groups 0-18 years, 19-64 years and 65 years and over. As can be observed from Figures 4.2.8, 4.2.9, 4.2.10., our model is able to fit the data points well; that is, the accumulated cases that is obtained as a result of using our model fits perfectly to the actual laboratory confirmed data obtained for each age group for the 2014-2015 influenza season.

4.1.3 Results from Numerical Investigations for Calgary Zone

Like the Central and North zones, the contact mixing matrix for Calgary zone was calculated using the method proposed by Meltzer et al., 2015 which was presented in section 3.3 of Chapter 3. The contact matrix obtained for the Calgary zone is presented in Table 4.3.1:

Table 4.3.1: Contact mixing matrix for Calgary zone.

Age Groups			
Age Groups	0-19 years	20-64 years	$\geq$ 65 years
0-19 years	7.6	6.1	0.4
20-64years	2.2	8.1	0.7
$\geq$ 65 years	1.6	7.6	1.7

From Table 4.3.1, it can be observed that the working class age group also makes more contact with themselves and the elderly age group. There are other interactions between the age groups including by the contacts between the working age group and the younger age group. This might be due to meeting in places like schools, playground, shops, home etc.

The initial conditions for Calgary were also obtained from the Alberta Health Services. It comprises of the 2014-2015 population estimates for Calgary with $I_{Di} = 1$ and $I_{Li} = 0$, i =1,2,3=0 . This can be seen in Table 4.3.2.

Table 4.3.2: 2014-2015 Population estimates for Calgary zone.

Parameter	Estimates
S_1	367122
S_2	1025349
S_3	164099

Like the Central and North zones respectively, the fitted parameters with their 95% confidence interval are obtained using the weighted nonlinear least squares fitting method with the ordinary differential equations in Mathematica. The 95% confidence interval is also obtained using the Metropolis hasting algorithm of the MCMC and the results obtained are shown in Table 4.3.3.

Table 4.3.3: Parameter estimates for Calgary zone

Parameters	Estimates	95% Confidence interval
β_1	$1.26558 * 10^{-7}$	(1.18168 $*10^{-7}$, $1.34082 * 10^{-7}$)
β_2	$1.82501 * 10^{-7}$	(1.79589 $*10^{-7}$, $1.85331 * 10^{-7}$)
β_3	1.01849 $*10^{-6}$	(8.51819 $*10^{-7}$, $1.19509 * 10^{-6}$)
f	0.0140083	(0.0133275, 0.0146999)

The estimates obtained in Table 4.3.3 fall within the 95% confidence intervals. It can also be observed that the confidence intervals are narrower hence the uncertainties in the estimates are lesser as compared to having a wider confidence interval. These estimates were obtained as a result of running a sample of 500000. Before we can go ahead to do any predictions, we first have to analyze $\rho(t)$.

$\rho(t)$ which measures the weekly proportion of laboratory confirmed cases including individual visits to the general practitioner's office and emergency departments, is analyzed using a specified nonlinear model. The nonlinear least squares function fits the data well as it passes through two or more points for each age group. Also the data points are widely scattere; that is, variability observed in the data sets although they are mostly clustered at the mid-section. From Figures 4.3.1, 4.3.2 and 4.3.3, the elderly and younger age group had the highest proportion of laboratory confirmed cases as compared to the working age group.

The parameter estimates obtained for the nonlinear function used in analysing $\rho(t)$, are shown in Table 4.3.4.

Table 4.3.4: Parameter estimates for $\rho(t)$ for each age groups.

Parameter estimates	0-18 years	19-64 years	65 years and over
a	13.9853	4.72771	4.3588
b	0.834321	0.223292	0.180505
c	$1.29624*10^{-6}$	0.00207674	0.00437972

From these results, we obtain the following plots below for the three age groups respectively.

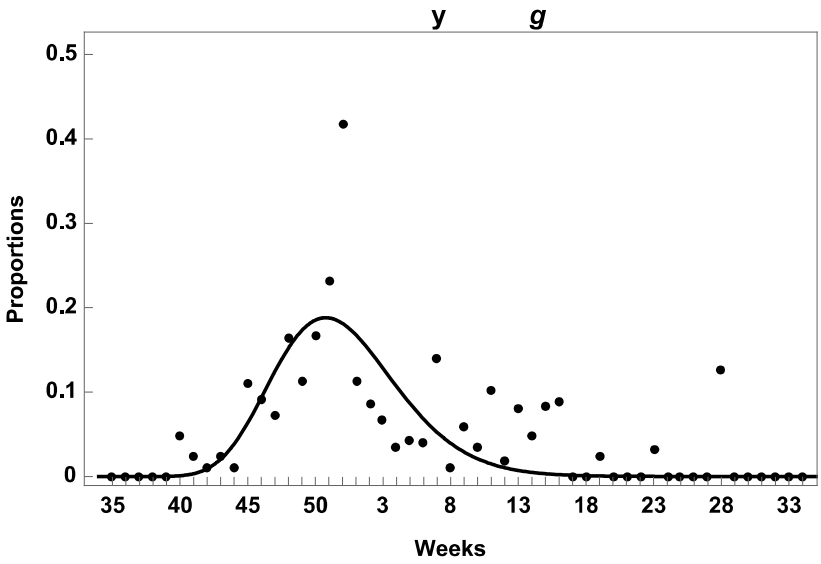

Figure 4.3.1: $\rho(t)$ for the 0-18 years age group.

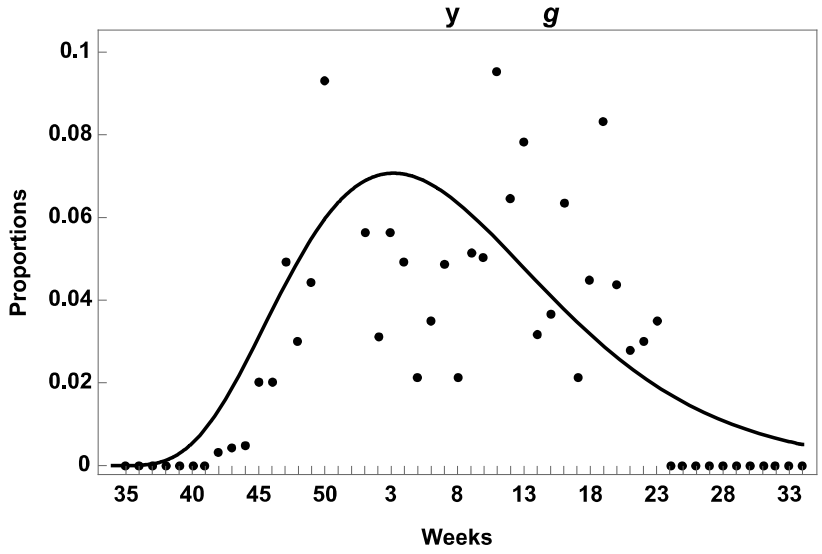

Figure 4.3.2: $\rho(t)$ for the 19-64 years age group.

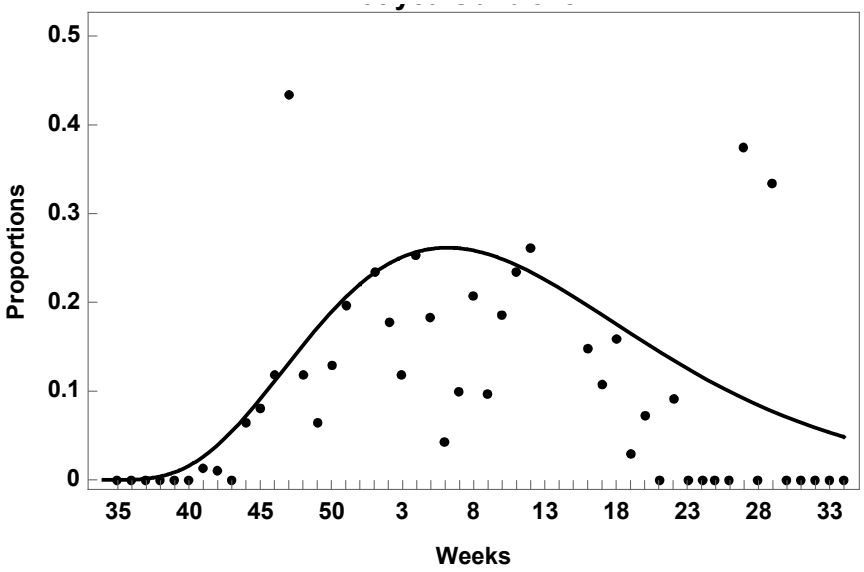

Figure 4.3.3: $\rho(t)$ for the 65 years and over age group.

Figures 4.3.4, 4.3.5 and 4.3.6 shows the 90% prediction bands for $\rho(t)$ obtained for the 0-18 years, 19-64 years and 65 years and over respectively.

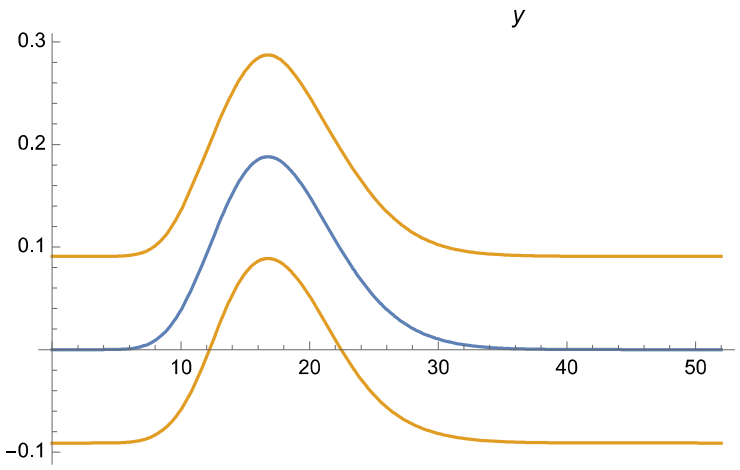

Figure 4.3.4: Prediction bands for 0-18 years' age group.

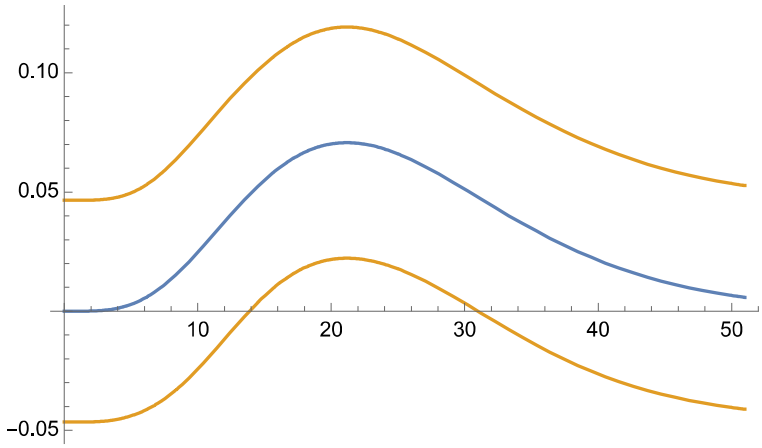

Figure 4.3.5: Prediction bands for 19-64 years' age group.

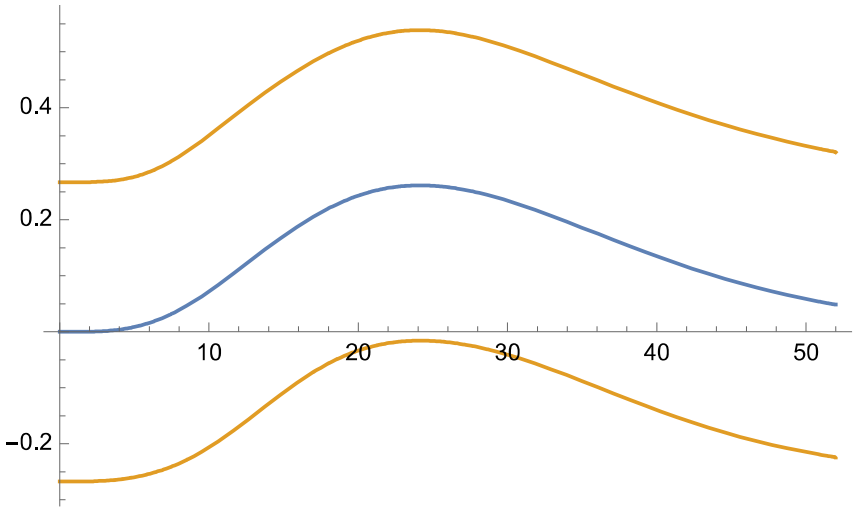

Figure 4.3.6: Prediction bands for 19-64 years' age group.

The prediction bands are shown by the cream-colored curves with the actual or true best-fit curve shown by the blue curve. The prediction bands, which show the scatter of the data, show that if more data points were obtained for $\rho(t)$ for each age group, then we should expect 90% of these data points to fall within these bands. It can also be observed that the prediction bands are wider because they have to account for the uncertainty of the curve as well as the scatter of the data

From using the parameter estimates obtained in Table 4.3.3 on all the 52 weekly data points, it was established that the week where the highest number of lab-confirmed influenza cases was observed, that is, peak in the influenza cases, was the 1st week with 95% confidence interval (52, 2) week. Our model predicted the 1st week and we can observe from the actual data that the highest case was also observed in the 52nd week. This is shown in Figure 4.3.7.

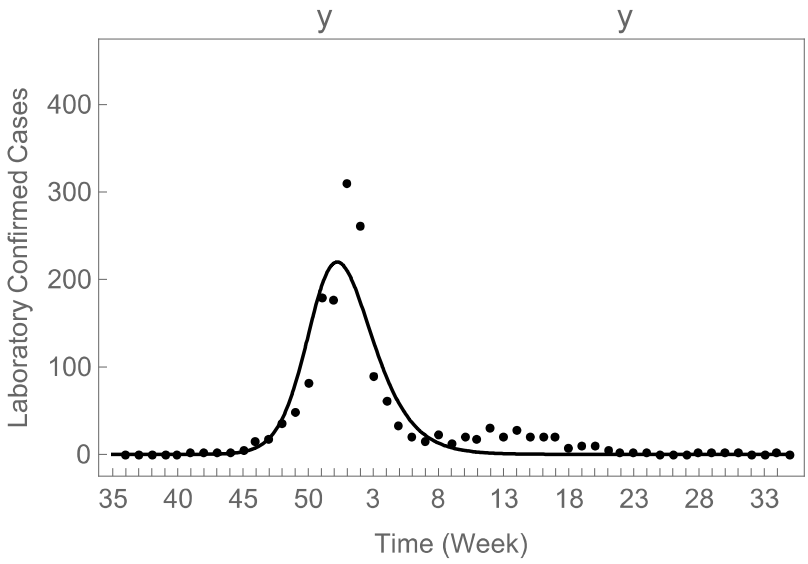

Figure 4.3.7: Laboratory confirmed cases for the Calgary zone.

Figure 4.3.7 shows the weekly laboratory confirmed cases for the Calgary zone where it can be observed that our influenza model is able to fit the actual data obtained as it comes close almost all of the data points. The number of laboratory confirmed cases that our model predicted is about 200 cases however the actual number of cases obtained was about 300. Hence our model fairly predicts the actual data well.

These weekly cases are then accumulated for each age group and then plotted to determine if our model can fit these cases as well as predict it.

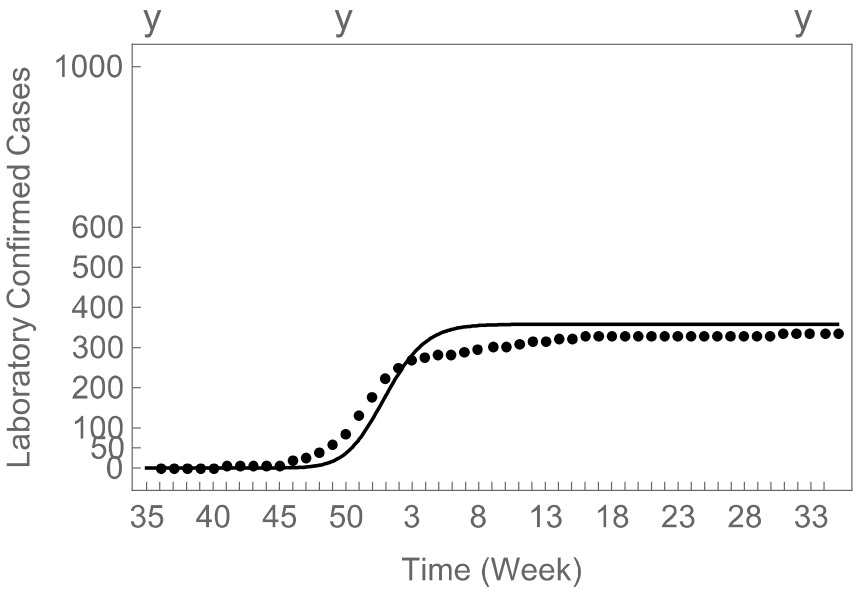

Figure 4.3.8: Accumulated laboratory confirmed cases for 0-18 years.

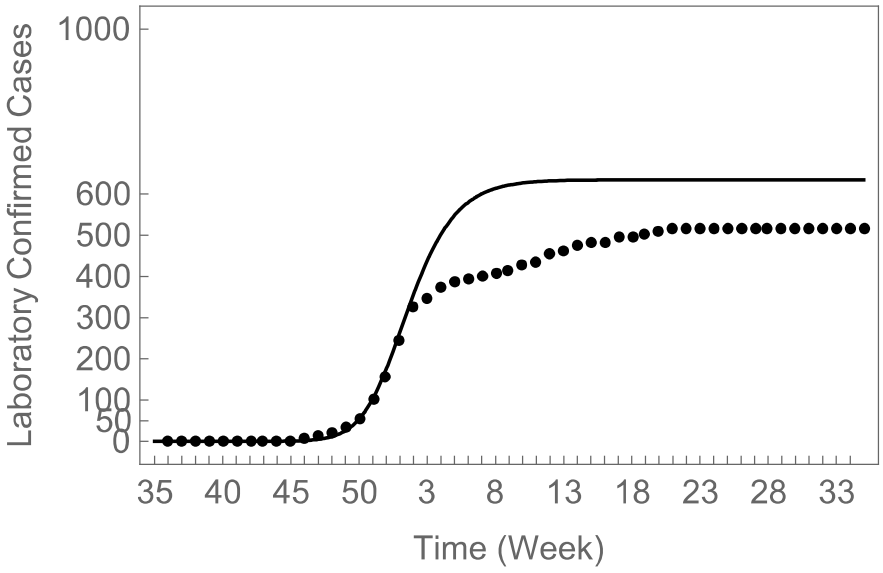

Figure 4.3.9: Accumulated laboratory confirmed cases for 19-64 years.

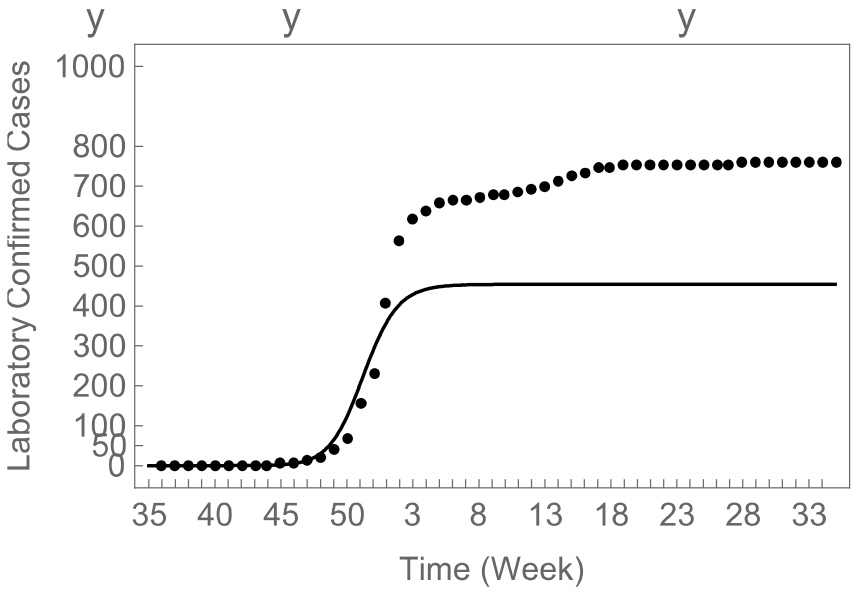

Figure 4.3.10: Accumulated laboratory confirmed cases for 65 years and over.

The accumulated laboratory confirmed cases for the age groups are shown in Figures 4.3.8, 4.3.9 and 4.3.10. It was identified that the model does fit the actual data at the initial start but not at the later part of the influenza season.

4.2 Discussions from Numerical Investigation

A question of interest that one might ask is "why the interest in predicting the peak of influenza?". Influenza epidemics, like any other viral epidemics, can persist and tend to cause more harm and havoc if more thought is not given on how the outbreak of this viral infection can be eradicated. Forecasting the peak of influenza will not only help in resource allocation but it will also help clinicians and policy makers plan ahead as to how to control the spread of the influenza, as well as how to distribute medicines and educate people on the do's and don'ts of this viral epidemic.

Our objective for this thesis was to propose a mathematical model for influenza that will be able to predict the week where there will a peak in the number of lab-confirmed influenza cases for the 2014-2015 influenza season for each zone. From the proposed influenza model, we were able to generate the ordinary differential equations which we used to analyse the laboratory confirmed cases obtained for each of the zones. These cases were analysed using weighted nonlinear least squares fitting. This was to obtain the parameters that we used to determine the week with the highest number of influenza cases, that is, the peak week for the 2014-2015 influenza season. However, before the peak week was determined, nonlinear least squares function was used to fit $\rho(t)$, the weekly proportions of laboratory confirmed cases. These are fitted for the three age groups considered in this thesis.

As was observed from the table of parameter estimates for $\rho(t)$ and the plots for $\rho(t)$ for the zones considered, Central, North and Calgary in section 4.1, the

selected nonlinear least squares function fits the data well as it passed through the points. Although the average scatter of the data was fairly wide, that is, there is variability in the data sets, it was usually clustered at the mid-section. The 90% prediction band was also plotted for $\rho(t)$ to determine if any additional points obtained for $\rho(t)$ will fall within the band. From the results it was identified that should there be any additional points obtained for $\rho(t)$, then 90% of these data points will fall within the bands.

The parameter estimates for each of the zones was obtained using the weighted nonlinear least squares fitting method in Mathematica and the 95% confidence interval was obtained using the Metropolis hasting algorithm of the MCMC. It was identified that each of the estimates for the zones fall within the 95% confidence interval. The 95% confidence interval was also identified to be narrower, hence uncertainty in the estimates was lesser as compared to a wider confidence interval.

For the Central zone, the model predicted the 52nd week with 95% confidence interval (51,1) week as the peak week where the most laboratory confirmed cases of influenza was obtained for the 2014-2015 influenza season. The actual data however showed the 1st week to be the week with the most laboratory confirmed influenza cases. The number of laboratory confirmed cases that our influenza model predicted was about 100 as compared to 131 actual cases obtained.

The influenza model also predicted the 1st week with 95% confidence intervals (52,4) week as the peak week where most of laboratory confirmed influenza cases was obtained for the 2014-2015 influenza season for the North zone. The model

predicted about 98 lab-confirmed cases whereas the actual number of cases obtained was about 113.

The Calgary zone had its peak week also being the 1st week with 95% confidence interval (52,2) week. The model predicted about 200 lab-confirmed cases whereas the actual highest number of cases obtained was about 300. Knowing this peak will help health agencies and the governments.

From the Central, North and Calgary zones results, it was identified from the actual data obtained that the actual peak week ranged between the 50th and 1st week whereas the model peak week ranged between the 52nd and 1st week. The same goes for the other zones too. The model also showed Calgary zone to have the highest number of laboratory confirmed cases. This therefore shows that our model was able to do the predictions well. Hence, given full data or real time data our model will likely be able to predict well.

Chapter 5

5 Conclusions and Recommendations

5.1 Conclusions

Forecasting the occurrence of epidemics will not only contribute to creating awareness but it will also help in resource allocation in terms of how to stock and distribute medicines as well as educate the masses on how to effectively ensure that the spread of the influenza virus is contained. In this thesis, a mathematical model for the influenza epidemic was developed and then used to predict the peak week where influenza was expected to be the highest within the geographical zones in the province of Alberta. The influenza model showed the relationship between the Susceptible (S) individuals, the infected but not lab-confirmed (I_D) individuals and the infected and lab-confirmed individuals (I_D). It was identified that within the infected but not lab confirmed compartment, there were some who were showing symptoms as well as others who were not showing any symptoms but could still infect a susceptible person. A person can also recover and leave the compartment and it was assumed that these people obtain immunity and so do not become susceptible again. From the results obtained, it was identified that using all the 52 data points given, our model was able to predict the peak week to range between the 52nd and 1st week. The model was also able to predict the number of laboratory confirmed cases which ranged between 95 and 400. Thus, the proposed influenza model is able to predict the peak week given all the data points. It is

expected that given any real time data, the proposed model should be able to predict the peak week.

5.2 Recommendations

There were many limitations that were encountered in this project. For instance, running the appropriate sample to get the appropriate parameter estimates takes about 7 days to complete and it takes 2 additional days to obtain the 95% confidence intervals. As a result of the computational demand in running the simulation code, the use of an interface like CDF player and parallel coding in Mathematica that can provide the peak prediction with their 95% confidence interval is recommended for future work on the proposed model.

The current model did not consider vaccination. However, it is recommended for future work on the proposed model to consider vaccination and how it will affect the prediction of the peak. Also the model did not consider the basic reproduction number; hence, it is recommended for future work to consider it.

Bibliography

[1]. Alberta Health Services. (2015). Seasonal Influenza in Alberta, 2014-2015 Season. *Surveillance and Assessment Brach – Alberta Health*. Retrieved July 28, 2016 from http://www.health.alberta.ca/documents/Influenza-Summary-Report-2015.pdf

[2]. Bedada, T. D., Lemma, M. N., & Koya, P. R. (2015, November). Mathematical Modeling and Simulation Study of Influenza Diseases. *Journal of Multidisciplinary Engineering Science and Technology (JMEST),* 2(11), pp. 3263 - 3269

[3]. Current Seasonal Influenza Update (n.d.). In IPAC. Retrieved July 28, 2016 from http://www.ipac-canada.org/links_flu.php

[4]. Earn, D. J., Dushoff, J., & Levin, S. A. (2002). Ecology and evolution of the flu. *Trends in ecology & evolution*, *17*(7), 334-340.

[5]. Ghasemi, O., Lindsey, M. L., Yang, T., Nguyen, N., Huang, Y., & Jin, Y. F. (2011). Bayesian parameter estimation for nonlinear modelling of biological pathways. *BMC systems biology*, *5*(Suppl. 3), S9.

[6]. Graybill, F. A., & Iyer, H. K. (1994). Regression analysis. *Duxbury Press*.

[7]. Hay, A. J., Gregory, V., Douglas, A. R., & Lin, Y. P. (2001). The evolution of human influenza viruses. *Philos Trans R Soc Lond B Biol Sci*, *356*(1416), 1861-70.

[8]. Henneman, K., Peursem, D. V., Huber, V. C. (2013, January). Mathematical Modelling of Influenza and Secondary bacterial infection. *WSEAS Transactions on Biology and Bioscience,* 1(10), pp. 1-11

[9]. Influenza Pandemic (n.d.). In *Wikipedia*. Retrieved July, 2016 from https://en.wikipedia.org/wiki/Influenza_pandemic.

[10]. Influenza - Seasonal (2014, March). In *WHO*. Retrieved June 25, 2016 from http://www.who.int/mediacentre/factsheets/fs211/en/

[11]. International Committee on Taxonomy of Viruses Index of Viruses — Orthomyxovirus (2006). In: ICTVdB—The Universal Virus Database,

version 4. Büchen-Osmond, C (Ed), Columbia University, New York, USA.

[12]. 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.

[13]. Key Facts. (n.d.). In *CDC*. Retrieved on November 26, 2014 from http://www.cdc.gov/flu/keyfacts.htm.

[14]. Klenk, H. D., Matrosovich, M., & Stech, J. (2008). Avian Influenza: Molecular Mechanisms of Pathogenesis and Host Range. *Animal Viruses: Molecular Biology*, 253.

[15]. Latin hypercube sampling. (n.d.). In *Wikipedia*. Retrieved July 10, 2016, from https://en.wikipedia.org/wiki/Latin_hypercube_sampling

[16]. Matsuzaki, Y., Sugawara, K., Mizuta, K., Tsuchiya, E., Muraki, Y., Hongo, S., & Nakamura, K. (2002). Antigenic and genetic characterization of influenza C viruses which caused two outbreaks in Yamagata City, Japan, in 1996 and 1998. *Journal of clinical microbiology*, *40*(2), 422-429.

[17]. Meltzer, M. I., Gambhir, M., Atkins, C. Y., & Swerdlow, D. L. (2015). Standardizing scenarios to assess the need to respond to an influenza pandemic. *Clinical Infectious Diseases*, *60* (suppl 1), S1-S8.

[18]. Motulsky, H., & Christopoulos, A. (2004). *Fitting models to biological data using linear and nonlinear regression: a practical guide to curve fitting*. Oxford University Press USA.

[19]. Nsoesie, E. O., Marathe, M., & Brownstein, J. S. (2013). Forecasting peaks of seasonal influenza epidemics. *PLOS Currents Outbreaks*, (1), doi: 10.1371/currents.outbreaks.bb1e879a23137022ea79a8c508b030bc.

[20]. Qiu, W. (2013). Spatio-Temporal Prediction Modeling of Clusters of Influenza Cases in Edmonton. *Master of Science in Epidemeology Thesis – University of Alberta*, 40 pp

[21]. Reported Influenza Hospitalization and Deaths in Canada: 2011-12 to 2015-16 (n.d.). In *PHAC*. Retrieved June 25, 2016 from http://www.phac-aspc.gc.ca/influenza/flu-stat-eng.php.

[22]. Types of Viruses (n.d.). In *CDC*. Retrieved July 28, 2016 from http://www.cdc.gov/flu/about/viruses/types.htm

[23]. Varugese, M. (2015). Forecasting the Peak: A model for influenza. *Presentation to Alberta Health Surveillance and Assessment Branch.*

[24]. Webster, R. G., Bean, W. J., Gorman, O. T., Chambers, T. M., & Kawaoka, Y. (1992). Evolution and ecology of influenza A virus. *Microbiological reviews*, *56*(1), 152-179.

[25]. Zambon, M. C. (1999). Epidemiology and pathogenesis of influenza. *Journal of Antimicrobial Chemotherapy*, *44*(suppl 2), 3-9.

Appendix A

Table A.1: Population Estimates for each age groups and each zones.

Age group	South zone	Calgary zone	Central zone	Edmonton zone	North zone
0-4 years	20,279	99,470	30,370	83,032	36,445
5-9 years	19,798	94,689	29,498	75,422	32,988
10-14years	18,333	84,267	28,003	68,551	29,727
15-19years	19,790	88,696	30,573	75,105	31,716
20-24years	22,174	104,599	33,741	97,128	37,306
25-29years	22,477	129,283	35,376	114,529	42,438
30-34years	22,329	141,225	34,996	115,779	41,909
35-39years	19,708	126,575	31,285	99,218	35,608
40-44years	17,800	118,704	29,428	89,888	32,755
45-49years	17,635	110,091	29,862	87,283	32,227
50-54years	20,515	113,879	35,442	94,900	35,182
55-59years	19,845	103,098	32,913	86,066	30,650
60-64years	16,495	77,895	26,263	67,723	21,964
65 - 69 years	13,381	58,041	21,125	51,171	15,595
70-74 years	9,852	37,801	15,309	34,804	10,722
75-79 years	7,431	27,585	11,510	26,147	7,383
80-84 years	5,904	20,950	8,727	20,175	5,247
85-89 years	3,811	12,908	5,366	12,653	2,928
90-94 years	2,300	6,814	3,295	7,236	1,599

Table A.2: Daily number of contacts between age groups obtained from the Polymod study conducted in United Kingdom.

	0-4 years	5-9 years	10-14 years	15-19 years	20-24 years	25-29 years
0-4 years	1.9	0.7	0.4	0.2	0.5	0.7
5-9 years	1.0	6.6	1.1	0.7	0.6	0.8
10-14years	0.5	1.3	6.9	1.5	0.3	0.3
15-19years	0.3	0.3	1.0	6.7	1.6	0.7
20-24years	0.5	0.3	0.2	0.9	2.6	1.5
25-29years	0.8	0.7	0.4	0.7	1.3	1.8
30-34years	1.0	1.1	0.6	0.5	0.9	1.2
35-39years	1.0	1.0	1.3	1.1	0.8	1.0
40-44years	0.6	1.0	1.1	0.9	0.7	0.9
45-49years	0.3	0.5	0.6	0.8	1.0	0.9
50-54years	0.3	0.4	0.4	0.4	0.4	0.9
55-59years	0.3	0.2	0.3	0.3	0.4	0.5
60-64years	0.3	0.3	0.2	0.2	0.2	0.3
65-69years	0.1	0.1	0.1	0.2	0.2	0.2
70-74years	0.1	0.2	0.2	0.1	0.2	0.2

Table A.2 Continued: Daily number of contacts between age groups obtained from the Polymod study conducted in United Kingdom.

	30-34 years	35-39 years	40-44 years	45-49 years	50-54 years	55-59 years
0-4 years	0.7	0.8	0.2	0.2	0.4	0.2
5-9 years	1.0	1.4	0.9	0.2	0.3	0.2
10-14years	0.5	0.8	1.0	0.7	0.3	0.4
15-19years	0.4	0.6	0.9	1.2	0.7	0.3
20-24years	0.8	0.6	0.8	0.9	0.9	0.6
25-29years	1.0	0.7	0.7	0.9	0.9	0.9
30-34years	1.7	0.9	1.0	0.9	0.9	0.6
35-39years	1.5	1.5	1.3	1.1	0.8	0.7
40-44years	0.7	0.9	0.8	1.2	1.4	1.3
45-49years	1.0	0.9	0.6	0.8	1.3	1.9
50-54years	0.4	0.9	0.6	0.6	0.7	1.0
55-59years	0.4	0.5	0.7	0.5	0.6	0.5
60-64years	0.2	0.3	0.4	0.4	0.5	0.6
65-69years	0.2	0.2	0.1	0.3	0.2	0.1
70-74years	0.2	0.2	0.2	0.4	0.5	0.7

Table A.2 Continued: Daily number of contacts between age groups obtained from the Polymod study conducted in United Kingdom.

	60-64 years	65 - 69 years	70-74 years
0-4 years	0.2	0.3	0.1
5-9 years	0.5	0.5	0.2
10-14years	0.3	0.4	0.3
15-19years	0.2	0.5	0.6
20-24years	0.5	0.4	0.3
25-29years	0.7	0.7	0.3
30-34years	0.8	0.6	0.3
35-39years	1.0	1.0	0.2
40-44years	0.9	0.8	0.8
45-49years	0.6	0.6	0.6
50-54years	0.6	0.6	0.6
55-59years	0.9	0.9	0.3
60-64years	0.7	0.9	0.6
65-69years	0.4	0.7	0.6
70-74years	0.5	0.7	1.5