Likelihood-Based Estimation of Dynamic Transmission Model Parameters for Seasonal Influenza by Fitting to Age and Season Specific ILI Data

Michael Waithaka

MedCrave
Step into the World of Research

Likelihood-Based Estimation of Dynamic Transmission Model Parameters for Seasonal Influenza by Fitting to Age and Season Specific ILI Data

Author:

Michael Waithaka

Promoters:

Dr. Nele Goeyvaerts

Prof. Dr. Niel Hens

Published By:

MedCrave Group LLC

March 24, 2017

Contents

Acknowledgement

A thesis submitted in fulfillment of the requirements for the degree of Master of Biostatistics

Abstract

Mathematical models provide important practical insights into the epidemiology of infectious diseases, and concepts derived from such models are widely used in the design of infection control programmes. This project was aimed at directly estimating the parameters of a dynamic transmission model using likelihood-based estimation methods, by fitting the model to age-specific influenza-like-illness (ILI) incidence over multiple influenza seasons. In an attempt to achieve the goal of this project, the dynamic transmission model for seasonal influenza of Vynnycky et al. [1] was adopted and the various model parameters estimated. Weighted Least Squares and Maximum Likelihood estimation methods were applied for the model parameters estimation. From the obtained estimates of these parameters, estimates for the average basic reproduction numbers, which is an important measure used in infectious disease control, immunization and eradication programmes, were also derived. This modelling approach is an improvement to the previous approaches where the parameter values of seasonal influenza models were commonly chosen ad hoc though projections based on such models heavily rely on the assumed input parameter values. Moreover, there exists considerable uncertainty over the most appropriate values for parameters for such models. The importance of parameter estimation and accounting for uncertainty when using dynamic transmission model outcomes as input for economic evaluations related to infectious diseases have already been highlighted by several previous studies [2,3].

Introduction

Influenza is a viral infection which effects human populations both through regular seasonal epidemics and occasional pandemics. Seasonal influenza is a contagious respiratory illness that strikes every year while pandemic influenza is a global outbreak. In temperate regions, seasonal influenza tends to occur as one annual epidemic that occurs in the winter months i.e. December to March in the Northern Hemisphere and June to September in the Southern Hemisphere [4]. In Europe and in Belgium, annual epidemics of influenza occur mostly during the winter months, usually between week 40 and week 20 of the following year. However the patterns of these epidemics are highly variable from year to year in terms of the beginning of the epidemic, its duration, intensity, and influenza strains that circulate. Thus, numbers of cases and deaths from influenza, as well as the most affected age groups, vary each season [5].

Seasonal influenza is a major burden on public health worldwide. In 2012, WHO estimated that annually it attacks 5-10% of adults and 20-30% of children globally and causes significant levels of illness, hospitalization and death [6]. Hanquet et al. [5] noted that the clinical and economic burden of seasonal influenza is frequently underestimated, as cases and deaths caused by influenza are rarely identified or coded as influenza outcomes, and only a minority of cases is confirmed by laboratory testing. The most common influenza-related outcomes are influenza-like illnesses (ILI), acute respiratory infections, pneumonia and all-cause deaths.

Vaccination is the most common and most effective public health responses to influenza though there exists other non-pharmaceutical interventions for the prevention and control of influenza infection such as mask use, hand hygiene, and social isolation. The effectiveness of these non-pharmaceutical interventions is uncertain and depends on behavioral responses in the general population which may vary across settings [7]. In addition, such measures are unlikely to be sufficient to prevent sustained influenza transmission both in pandemic and seasonal epidemic years [8]. Vaccination protects against influenza by stimulating an antigen-specific immune response. Two different types of influenza vaccine are available: the trivalent inactivated influenza vaccine (TIV) and the trivalent live-attenuated influenza vaccine (LAIV). TIV is administered via intramuscular or intradermal injection while LAIV is administered intranasally via a sprayer. The current Health Council recommendations for seasonal influenza vaccination in Belgium is limited to persons at higher risk of influenza complications, including, persons aged above 50 years, health care workers, pregnant women and poultry and pork farmers [9].

Various studies have been conducted to evaluate the impact of various childhood vaccination strategies. For instance, Vynnycky et al. [1] applied an age-structured model to estimate the long-term impact of vaccinating children of either pre-school or school age on the burden of seasonal influenza (A and B) in the United Kingdom and to assess the effects of different contact patterns between children and adults. Beutels et al. [10] and Goeyvaerts et al. [11] developed a dynamic transmission model for seasonal influenza to evaluate the impact of various childhood vaccination strategies by fitting to observed age and time specific Influenza-Like-Illness (ILI) incidence in Belgium. The dynamic model was implemented in MATLAB and a global search algorithm was used to estimate the model parameters by minimizing a weighted least squares criterium for the ILI data. This procedure turned out to be highly sensitive to the initial values and was only able to identify local optima. The goal of this project is to explore alternative estimation methods for the influenza model parameters in the free statistical programming language R. In this regard, we apply the Weighted Least Squares and Maximum Likelihood estimation methods.

Methodology

Data

Weekly data on Influenza-Like-Illness (ILI) are collected from a sentinel network of General Practitioners (GPs) in Belgium coordinated by the Scientific Institute of Public Health. In 2009, the network involved around 200 GPs, representing approximately 1.8% of all Belgian GPs, reporting on ILI consultations [5]. The GPs report weekly, on a standardized paper form, every patient with an influenza-like illness. For every patient, age group (<5, 5-14, 15-64, 65-84, >84), hospitalization, antiviral treatment (as of week 35), delivery of absence from work certificate, and seasonal and pandemic vaccination status (as of week 42) are also recorded [12].

In this project, the dynamic model is fitted to ILI incidence data from week 40 of year 2003 to week 35 of year 2009. It is assumed that the ILI incidence is representative of the true influenza incidence, and that there is no time or age bias. Belgian demographic data for year 2009 obtained from Eurostat [13] are used to determine the initial age-specific population distribution and to estimate an age-specific annual mortality rates.

Dynamic transmission model

Mathematical models provide important practical insights into the epidemiology of infectious diseases, and concepts derived from such models are widely used in the design of infection control programmes. The basic idea in infectious disease modeling is that the population is divided into disjoint groups, according to a few key characteristics which are relevant to the disease under consideration. Then the progress of an epidemic is modelled by keeping track of the number of individuals within each subgroup, which are called compartments. In a dynamic model there are transition processes between the compartments that specify the rate at which individuals move from one compartment to the other. The models are typically formulated as systems

of differential equations.

The dynamic model structure: In order to model the progress of an epidemic in the population, the population diversity must be classified into a few key characteristics which are relevant to the infection under consideration. In this project, the model of Vynnycky et al. [1] is adopted. The model is an age stratified SEIRS model with vaccination. The population is classified into compartments S_a (t) = number of susceptible individuals aged a years at time t, E_a (t) = number of exposed individuals aged a years at time t, I_a (t) = number of infectious individuals aged a years at time t, R_a (t) = number of recovered individuals aged a years at time t, and V_a (t) = number of vaccinated individuals aged a years at time t. In this case, the infection has a significant latent period during which the individual who has been infected is not yet infectious to others. During this period the individual is in compartment E (for exposed). Both the recovered and vaccinated individuals are assumed fully protected after infection and vaccination, respectively, until their immunity wanes [14].

We define age groups, a, of length 1 year such that we have individuals who are aged <1, 1 - <2, 2 - <3, . . . , 99 - <100. The population size and the age-specific mortality rates were assumed to be constant to avoid complications associated with modeling population growth. The model is dynamic in that the number of individuals in each compartment may fluctuate over time. Figure 1 presents the age-stratified SEIRS model with vaccination. The dynamic model states are defined in Table A.1 and the system of ordinary differential equations that characterize the model is given in (A.1) in the appendices. The model parameters are defined in Table 1.

A Realistic Age-Structured (RAS) model was assumed in which all individuals move to the next age group on August 31 of each year [14]. Individuals are born into the first stratum on August 31 of each year. Though this approach for introducing newborns into the population may be less “natural” than allowing newborns to enter the first age stratum continuously, it facilitates tracking the exact time when individuals reach the earliest age at vaccination (6 months of age) [1]. To ensure that the population size remains constant, the individuals in the last age group are removed from the population and as many births as deaths in the preceeding year are introduced into the population. Similarly, vaccination of a proportion of individuals in any age group is assumed to be completed on time t_{vacc} of each year. The influenza vaccination is given to the individuals irrespective of their vaccination or disease history. The transmission of influenza is ensured to continue each year by seeding a fraction of the susceptible in each of the age groups as newly infectious individuals at time t_{seed}. The two rates at which individuals lose their immunity, w_i and w_v, are assumed to be equal and age-independent. In the model representation, the single arrows indicate time continuous transitions while double arrows indicate instantaneous transitions on August 31 of each year. The arrows indicating influx compartments for vaccination and seeding are suppressed from the display.

Table 1: Definition of the model parameters.

Parameter	Definition
$\lambda_a(t)$	The force of infection for individuals of age a at any time t during a year.
$\overline{R_0}$	The average basic reproduction number measured at reference time t_0.
δ	Amplitude of the sinusoidal seasonality function z(t)
t_0	Reference time for the seasonality function z(t), at which the basic reproduction number equals $\overline{R_0}$
m_a	Yearly mortality rates of individuals of age a
f	The daily progression rate from exposed (latent) to infectious, (1/aver-age latent period).
r	Daily rate at which infectious individuals recover and become immune, (1/average infectious period)
t_{vacc}	Time of vaccination each year
t_{seed}	Time of the year at which newly infectious individuals are introduced as a seed into the population.
w_v	Yearly rate at which vaccinated individuals lose their immunity (1/average duration of protection)
w_i	Yearly rate at which naturally infected individuals lose their immunity, (1/average duration of immunity)

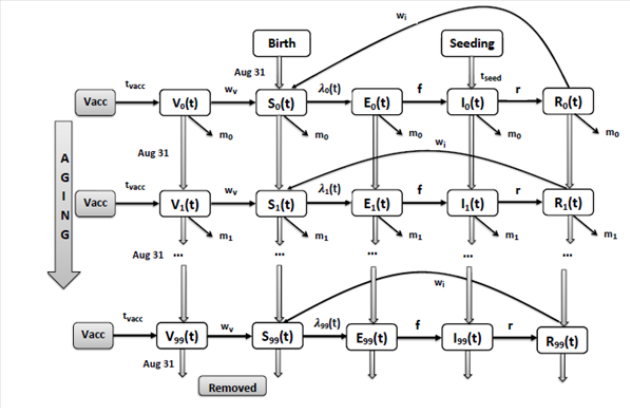


Figure 1: Age-stratified SEIRS model with vaccination.

The force of infection for individuals of age a at any time t during a year, is related to the seasonally forced age-specific transmission rates by:

$$\lambda_a(t) = Z(t)\Sigma_{a'}\beta_{a,a'}I_{a'}(t), \ldots\ldots\ldots\ldots (2.1)$$

$$Z(t) = 1 + \delta * \sin\left(\frac{2\pi(t-t_0)}{365}\right).$$

Here, $\beta_{a,a'}$ denotes the average daily per capita rate at which an individual of age a' makes effective contact with a person of age a while Z (t) reflects the relative change of the basic reproduction number at time t, $R_0(t)$, from the average basic reproduction number, R_0, measured at reference time t_0. The seasonal peak of transmission occurs three months after the reference time t0. The amplitude of the seasonality function Z(t) is bounded $0 \leq \delta \leq 1$ to ensure that $Z(t) \geq 0$, 8 $\forall$ t. It determines the peak value of the basic reproduction number. The basic reproduction number at time t, $R_0(t) = R_0 Z(t)$.

The average basic reproduction number can be defined as the average number of new infections caused by a single infected individual when introduced into a wholly susceptible population at reference time t0 over the course of the infection of this individual. In an age-stratified population the basic reproduction number depends on the duration of the infectious period, the probability that a contact between an infected and a susceptible individual leads to an infection, the contact rate, and the constant age distribution of the population. It is calculated as the dominant eigenvalue of the next generation matrix with elements $\frac{N_a\beta(a,a')}{r}$

In general, the basic reproduction number quantifies the transmission potential of the disease such that, if it falls below one ($R_0 < 1$) the infection eventually dies out. If R0 > 1 there is an epidemic in the population, whereas in the case where R0 = 1, the disease becomes endemic, meaning the disease remains in the population at a consistent rate since one infected individual transmits the disease to only one susceptible on average.

Social contacts: Modelling the spread of infectious diseases requires assumptions to be made regarding the underlying transmission process. Influenza is transmitted mainly through social interactions of which the frequency and intensity typically depend on age. To this purpose, daily rates of close contacts >15 minutes by 1-year age intervals (starting from age 0), estimated by Goeyvaerts et al. [14] from the Belgian POLYMOD contact survey conducted in 2006, were used in this project.

The age-specific transmission rates ($\beta_{a,a'}$) are assumed to be directly proportional to the age-specific rates of making social contact. This concept is known as the social contact hypothesis introduced by Wallinga et al. [15]:

$$\beta(a,a') = q.c(a,a') \ldots\ldots (2.2)$$

Where q is a constant proportionality factor which measures the disease-specific infectivity.

Vaccination: As noted earlier in the introduction, vaccination is generally considered to be an effective tool to protect against influenza disease and its complications. Although two different types of influenza vaccine are available: the trivalent inactivated influenza vaccine (TIV) and the trivalent live-attenuated influenza vaccine (LAIV), it is not until 2011 that LAIV became authorized in the European Union. Thus it was not on the Belgian market during the study period considered in this project. Therefore, all references on vaccination in this project are based only on TIV.

Our model assumed an all-or-none effect of vaccine. With an all-or-none effect of vaccine, the vaccine efficacy V E means that the vaccine is 100% efficacious in fraction V E of individuals who are vaccinated and has no effect on the remaining fraction (1-V E). The effective vaccination coverage is thus the product of the vaccine efficacy and vaccine coverage and determines the proportion of individuals that move to the vaccinated state each year. Goeyvaerts et al. [14] obtained literature-based estimates of the TIV vaccine efficacy for influenza-confirmed ILI from randomized controlled trials and observational studies as proxies for V E. The estimates classified by age and type of season are summarized in Table A.2 in the appendices. The type of season is then classified by influenza intensity and the degree of matching between the vaccine and the circulating viral strain. The following age-stratified vaccination coverage estimates obtained from the Belgian Health Interview Survey of 2008 conducted by the Scientific Institute of Public Health were used: 0:066% for 6 months - 17 years (arising from a 1% coverage in children with co-morbidities), 11% for 18-49 years, 28% for 50-64 years, 50% for 65-74 years, and 71% for $\geq$ 75 years [10,14,5].

Estimation of the dynamic model parameters

To estimate the values of the unknown parameters that are applied in the SEIR model, different approaches outlined below were used.

Weighted least squares approach: The Weighted Least Squares (WLS) technique involves least squares fitting, where the values of the model parameters which minimize the weighted squared differences between model predictions and the observed ILI incidence data are sought.

We assume that the epidemiological system is exactly described by a dynamic model together with some set of parameters, but the observed data arises from some deviation of the output of this system by observational errors. The parameter set is written as the p-element vector θ . C_a (w_i) denote the number of reported ILI cases of age a in calendar week i and Pa(w_i) denote the corresponding

denominator, i.e. the number of individuals of age a covered by the sentinel network in calendar week i. The observed age-specific ILI incidence rate in calendar week i can then expressed as: $Y_a(w_i) = C_a(w_i) = P_a(w_i)$. Thus, we can assume that the statistical model can be written as:

$$Y_i = M(t_i;\theta) + E_i$$

Where M_i is the dynamic model for the incidence evaluated at the true*parameter value, θ and the E_i depict the errors. Letting $I_a(t)$ denote the number of newly infectious individuals of age a at time t, and Na (t) denote the total number of individuals of age a at time t as predicted by the model, the model-based age-specific incidence rate in calendar week i is given as:

$$Z_a(wi) = \frac{\sum_{t\varepsilon wi} I^*_a(t)}{\frac{1}{7}\sum_{t\varepsilon wi} N_a(t)} \quad \text{..... (2.4)}$$

For known values of the weights, estimation of the model parameters proceeds by minimizing the weighted sum of squared differences between the observed ILI incidence rate and the scaled model-based incidence rate in (2.5).

$$\sum_{j=1}^{4}\sum_{i} v_{aj}(w_i)\left(Y_{aj}(w_i) - \alpha Z_{aj}(w_i)\right)^2 \quad \text{..... (2.5)}$$

The weights, V_{aj} (w_i), are taken to be proportional to the corresponding denominator P_{aj} (w_i). They account for the unequal population sizes represented by the different age groups. The scale factor α which calibrates the model-based incidence rate to the observed ILI incidence rate may absorb several effects such as the probability for an infected individual to show symptoms, the GPs consultation rate and ILI cases reporting rate. The weighted sum is taken over all weekly ILI observations, from week 40 in 2003 to week 35 in 2009, per age group aj : 0-4, 5-14, 15-64 and ≥ 65 years.

Maximum likelihood approach: In statistics, each population is identified by a corresponding probability distribution. Associated with each probability distribution is a unique set of the model's parameters. As the parameters change in value, different probability distributions are generated. Thus a model is defined as the family of probability distributions indexed by the models parameters. The likelihood is the probability of observing the data given the model and parameter values for the model. The basic idea of maximum likelihood estimation is to find the parameter set that maximizes the likelihood of observing your data.

Maximum likelihood estimation:

Let $f(y \mid \theta) = f(y_1, y_{2,\ldots} y_n \mid \theta) = \prod_{i=1}^{n} f_i(y_i \mid \theta$ denote the probability density function (pdf) that specifies the probability of observing data vector y given a vector of parameters θ. The individual observations, yi's, are assumed independent of one another. The likelihood function is defined as L(θ|y) = f (y|θ) and represents the likelihood of the parameters θ given the observed data y and as such is a function of θ. A maximum likelihood estimator (MLE) of the parameter set θ based on the observed data y is a parameter vector at which L(θ|y) attains its maximum as a function of θ, with y being held fixed. In maximum likelihood estimation we seek the value of the parameter vector that maximizes the likelihood function L(θ|y). The resulting parameter vector, which is sought by searching the multidimensional parameter space, is called the MLE, and is denoted by θˆ = θˆ1 , θˆ2 , . . . , θˆk . For computational convenience, the MLE estimate is obtained by maximizing the log likelihood function, in (L(θ|y)) ≡ l(θ|y). This is because the two functions, l(θ|y) and L(θ|y) are monotonically related to each other so the same MLE estimate is obtained by maximizing either of the two.

Application to the data: The observational model is given by

$C_a(w_i) \sim$ Binomial (n , α)……. (2.6)

Where $n = \frac{P_a(w_i)}{Na(w_i)} * I^*_a(w_i)$ and α scale factor

Since n is large, this distribution can be approximated by a Poisson distribution. Because of over dispersion of the data, this can further be generalized to an over dispersed Poisson or a Negative Binomial distribution. In this project, the Negative Binomial distribution was considered which is given as:

$$Ca(wi) \sim Negative\ Binomial\left(\mu_a(w_i),\ \mu_a(w_i) + \frac{1}{k}\mu_a(w_i)^2\right)$$

…… (2.7)

where: $\mu_a(w_i) = E(C_a(w_i)) = P_a(w_i) \times E(Y_a(w_i)) = P_a(w_i) \times \alpha Z_a(w_i)$ and 1 > 0

denotes the over dispersion parameter $\frac{1}{k} = 0$ there is no over dispersion and the Negative Binomial distribution simplifies to a Poisson distribution. In terms of the standard

parameterization of the Negative Binomial distribution using the parameters r, number of events until the experiment is stopped, and p, probability of success in each trial, where r = k and p = $1/\left(1+\frac{1}{k}\mu_a(w_i)\right)$ the distribution (2.7) can be written as:

$$C_a(w_i) \sim Negative\ Binomial\left(r=k,\ p=\frac{1}{1+\frac{1}{k}\mu_a(w_i)}\right) \quad \ldots\ldots (2.8)$$

Considering the value of the parameters r and p, and given the vector of observed data, y, the log-likelihood function is the same function as the logarithm of the probability density (2.9).

$$l(r,p \mid y) \sum_a \sum_i l(r,p \mid C_a(w_i))$$

$$= \sum_a \sum_i \ln\left(\Gamma(r+C_a(w_i))\right) + r\ln(p) + C_a(w_i)\ln(1-p) - \ln\left(\Gamma(r)\right)$$

$$-\ln\left(\Gamma(C_a(w_i)+1\right) \ldots\ldots\ldots (2.9)$$

The general mathematical technique for solving for MLEs involves differentiating the log likelihood function with respect to the parameter vector, set the resulting gradient vector to zero and then solve the system of equations. But this method only works if there is an analytical solution. Another possibility is the grid search method which involves finding the maximum of the log-likelihood function by repeated approximation and iteration. However, this method is also not practical in most cases and becomes much more difficult when the number of parameters increases beyond one or two. As a result, most statistical packages employ some kind of numerical maximization method. In this method one essentially feeds the computer with a set of starting values and let algorithms such as Newton-Raphson, Nelder-Mead, quasi-Newton and conjugate- gradient find the maximum.

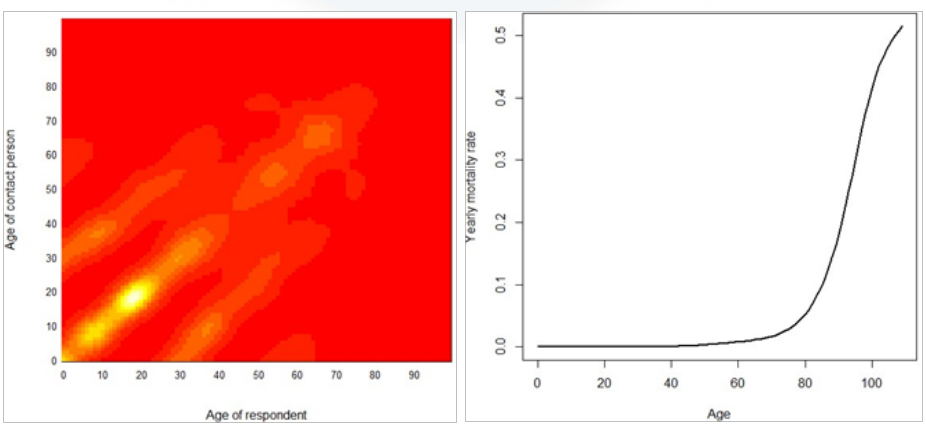
Results

Exploratory results

Figure 2a displays the contact rate matrix estimated by Goeyvaerts et al. [14] from the Belgian POLYMOD contact survey conducted in 2006 (technical details on the estimation are provided in Goeyvaerts et al. [14]. From the plot, high rates are observed on the diagonal indicating that people mostly mix with people of the same age class, particularly among the children and young adults (assortative mixing). In addition, an off-diagonal parent-child component is observed, though of weaker magnitude than the assortative structure.

The population size and mortality rates data for Belgium in 2009 obtained from Eurostat was stratified by age. Data for up to age 99 years were used for analysis in this report. This is motivated by the fact that, individuals aged 99 years and above are assumed to be removed from our population. The total Belgian population size from age 0 to 99 years was 10,751,601. The mortality rates estimated from Eurostat demographical data are shown in Figure 2b.

Figure 2 presents plots of the weekly observed ILI incidence rates stratified by age groups in Belgium. The data available was for the seasons from week 40 of year 2003 to week 35 of year 2009. Most of the data represented the weeks when an influenza epidemic can be expected (week 40 up to week 20 of the following year). The plots show similar trends in all the age groups for each influenza season but ILI incidence rates are higher for the age group 0 − 14 years than the other age groups. The incidence rates for the last age group (≥ 65) are the lowest compared to the other age groups. 2003-2004 on average had the highest observed ILI incidence rates while 2007-2008 had the lowest rates. Similar observations are made for the total population as shown in Figure A.1 in the appendices.



(a) Daily close contact rates of >15 minutes (b) Age-specific yearly mortality rates

Figure 2: Daily rates of close contacts >15 minutes as estimated by Goeyvaerts et al. (2010) and yearly mortality rates by age estimated from Eurostat data.

Dynamic model parameters estimation

The main parameters driving goodness-of-fit to the ILI incidence data as identified by Goeyvaert et al. [14] include: the average basic reproduction number (R_0), the amplitude (δ), the immunity waning rates (w_i, w_v), and the scale factor (α). Also, some of the less influential parameters include: time point of vaccination (t_{vacc}), time point at which infectious individuals are seeded into the population (t_{seed}), rate at which the exposed individuals become infectious (f), and recovery rate (r).

It was assumed that vaccination took place the same time as the spontaneous transition on August 31 each year. To ensure that transmission of influenza continued each year, 200 infectious individuals were seeded in each age class of 5-50 years at time point t_{seed}. Individuals outside this age group were not included in the seed. Vynnycky et al. [1] argued that in previous influenza pandemics, very few of the earliest cases occurred in this age range and that older individuals (arbitrarily taken to be those aged >50 years) are unlikely to be the first cases during a typical influenza season, given some immunity to circulating strains resulting from exposure to related strains.

Some parameter values such as the average duration of the latent period and the average duration of the infectious period could be assumed with reasonable confidence from empirical studies. Glasser et al. [16] suggested an average latent period of 1 day and an average infectious period of 3.8 days. Similarly, Cauchemez et al. [17] estimated the mean duration of infectiousness to be 3.8 days with a 95% credible interval of [2,8,17,18].

However, there is still considerable uncertainty over the most appropriate values for some parameters in our model: the average basic reproduction number, the amplitude of seasonal forcing, the average duration of effective immunity, and the scale factor. Therefore, solutions to the SEIRS model in Figure 1 were used to estimate values for these parameters so that the resulting dynamics exhibit the characteristics of seasonal influenza. Some of which being: the amplitude of seasonal forcing to be sufficiently strong that there is a genuine off-season with very little transmission, and that the amplitude of seasonal forcing is weak enough that the system settles down into regular annual cycles: if seasonal forcing is too strong, there are frequent years with no infections [19].

Because the timing of the epidemic peak differs substantially between seasons, the reference and seeding time points t_0 and t_{seed}, were retained as season-specific parameters. The parameters related to the characteristics of seasonal influenza, including the amplitude of seasonal forcing and the immunity waning rates were assumed constant over the seasons. The scale factor (α) was also assumed constant over the course of the sea- son (not allowed to change during an epidemic season), by age and across the various seasons.

Weighted least squares estimation: The model was pre-run over a burn-in period of five influenza seasons to generate back-ground immunity due to historical infection or vaccination. A season-based year of 364 days per year which is equivalent to 52 seven-day weeks was implemented with t = 0 being on September 1. The dynamic model was simulated using a set of starting values (initial parameter guesses) and R function optim was used in the optimization using SANN optimization method with iteration count as 10,000. SANN method performs an optimization using a stochastic optimization algorithm known as simulated annealing which is an adaptation of the Metropolis-Hastings algorithm (a Monte Carlo method) [13]. This algorithm can overcome the problem of local maxima, although the algorithm may not be a feasible option as it may take unrealistically long time to find the solution.

The Nelder-Mead algorithm was then run until convergence using the estimates from the SANN method as starting values. The Nelder-Mead method work reasonably well for non-differentiable functions but though it is relatively slow, convergence is attained relatively faster as compared to the SANN method. The weighted least squares function used enabled estimation by finding parameter values minimizing the weighted sum of squared differences between the observed ILI incidence rate and the scaled model-based incidence rate given in (2.5). Table 2 shows the parameter estimates obtained. The dynamic model was then re-simulated using the parameter estimates and Figure 3 shows the fit of the dynamic model to the age-stratified ILI incidence data.

Maximum likelihood estimation: As was the case in the weighted least squares approach, a function for simulating the dynamic model was used. In addition, a function to return the log-likelihood of the data, (2.9), given some combination of parameters was also used. The optimization procedure used in the weighted least squares estimation was followed. In this scenario, maximizing the likelihood function determines the parameters that are most likely to produce the observed data. The parameter estimates obtained are given in Table 2, Figure 4 shows the fit of the dynamic model to the age-stratified ILI incidence data after re-simulating the dynamic model using the parameter estimates obtained.

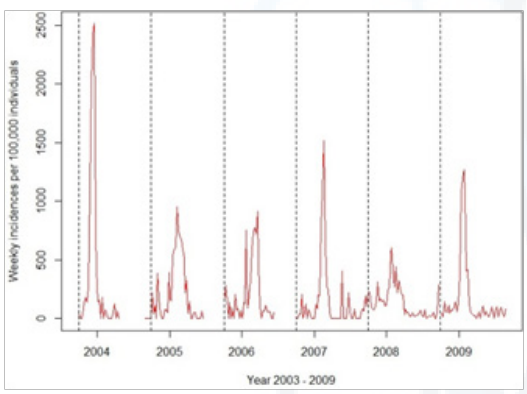
Based on the model under consideration, and conditionally on the values of the other parameters obtained, it takes approximately 2.27 years for naturally infected and vaccinated individuals to lose immunity and become susceptible again. As mentioned earlier in this report, the scaling factor may absorb several effects such as ILI cases reporting rate, the probability for an infected individual to show symptoms, the GP consultation rate and also it might also absorb incorrect modelling assumptions. The $\overline{R_0}$ is estimated to be highest in 2003-2004 and lowest in 2005-2006, which very well corresponds with the seasonal classification of influenza intensity in Table A.2. The peak

value of the basic reproduction number, determined by δ is estimated to be approximately 1.21 or 1.20 times the average basic reproduction number using the maximum likelihood and weighted least squares approaches respectively. The reference time is estimated to be in September - October for the different seasons, which means that the seasonal peak of transmission would occur in December-January since the seasonal peak of transmission occurs three months after the reference time. The estimated seeding time is mainly September-October for the different seasons.

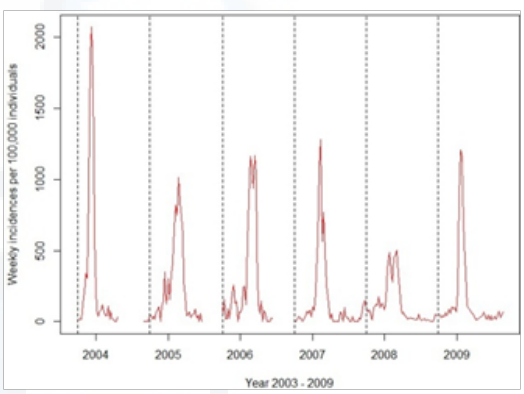
From Figure 3 and Figure 4 it is clear that the dynamic model tend to very well approximate the incidence rates for the total population. Similarly, the incidence rates for the age group of 15-64 years are also very well approximated by the model. On the other hand, the model tend to underestimate the total incidence for individuals aged 0-14 years.

Table 2: Weighted Least Squares estimates for the dynamic transmission model parameters.

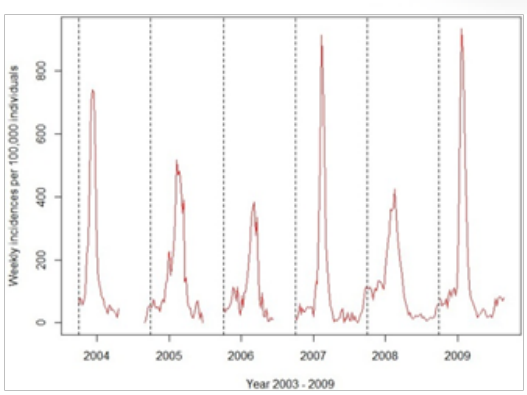
Season	δ	$w_v = w_i$	α	t_0	tseed	q	$\overline{R_0}$
2003-2004	0.201	0.440	0.212	Oct 05	Sept 21	0.170	5.002
2004-2005	''	''	''	Sept 17	Sept 14	0.120	3.530
2005-2006	''	''	''	Sept 01	Oct 11	0.101	2.942
2006-2007	''	''	''	Sept 30	Sept 02	0.140	4.119
2007-2008	''	''	''	Sept 07	Nov 15	0.110	3.236
2008 -2009	''	''	''	Oct 26	Sept 04	0.160	4.707



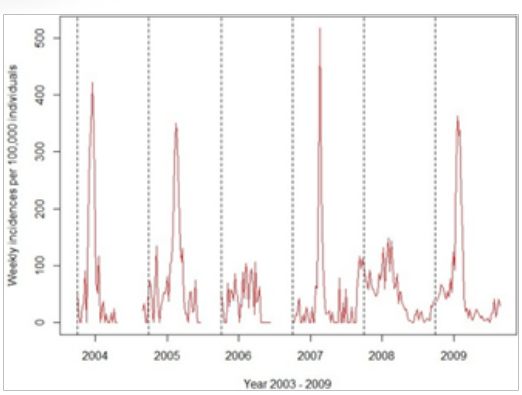
(a) Incidence rates for age group 0- 4 years



(b) Incidence rates for age group 5-14 years

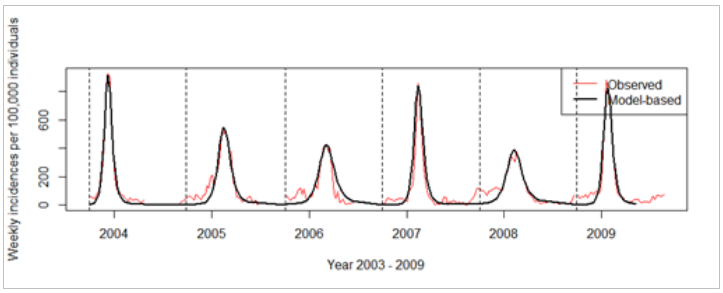


(c) Incidence rates for age group 15- 64 years

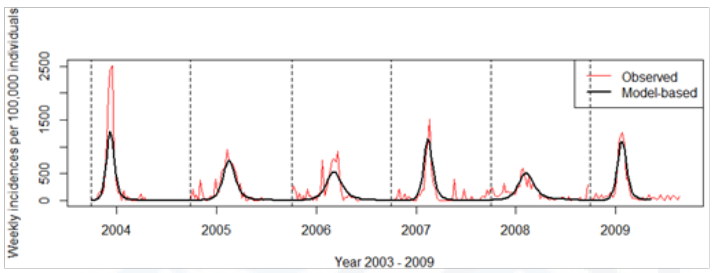


(d) Incidence rates for age group ≥65 years

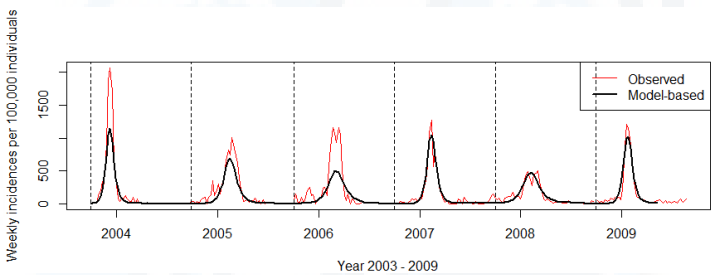
Figure 3: Observed ILI incidence rates stratified by age groups in Belgium 2003-2009.



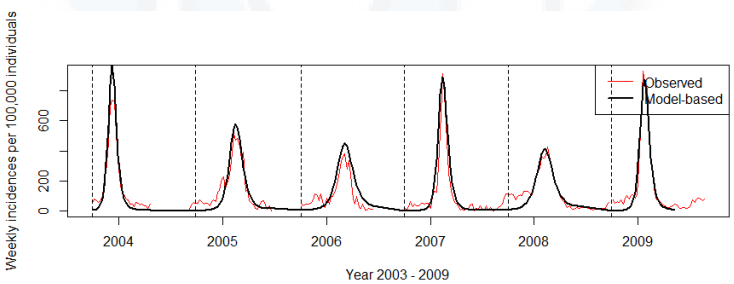
(a) Incidence rates for the total population.



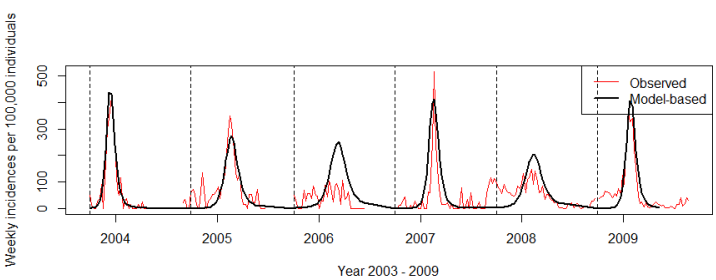
(b) Incidence rates for age group 0-4 years



(c) Incidence rates for age group 5-14 years



(d) Incidence rates for age group 15-64 years



(e) Incidence rates for age group _65 years

Figure 4: Observed ILI incidence rates and the corresponding model-based estimates (using WLS parameter estimates) in Belgium 2003-2009.

Table 3: Maximum likelihood estimates for the dynamic transmission model parameters.

Season	δ	$w_v = w_i$	α	t_0	tseed	q	$\overline{R_0}$
2003-2004	0.201	0.439	0.230	Oct 05	Sept 24	0.168	4.943
2004-2005	‘’	‘’	‘’	Sept 13	Sept 27	0.116	3.413
2005-2006	‘’	‘’	‘’	Sept 02	Oct 29	0.100	2.942
2006-2007	‘’	‘’	‘’	Oct 05	Sept 03	0.140	4.119
2007-2008	‘’	‘’	‘’	Sept 11	Dec 05	0.110	3.236
2008 -2009	‘’	‘’	‘’	Oct 26	Sept 05	0.161	4.737

Discussion and Conclusion

Since influenza is transmitted mainly through social interactions of which the frequency and intensity typically depend on age, the daily rates of close contacts >15 minutes by 1-year age intervals estimated by Goeyvaerts et al. (2010) from the Belgian POLYMOD contact survey were used. These estimates depicted an assortative mixing especially among the children and young adults. In addition, an off-diagonal parent-child component was observed, though of weaker magnitude than the assortative structure. Belgian demographic data on population size and mortality rates for year 2009 obtained from Eurostat were also used to determine the initial age-specific population distribution and to estimate age-specific annual mortality rates. The total Belgian population size from age 0 to 99 years was 10,751,601. Data on the weekly observed ILI incidence rates stratified by age groups showed that ILI incidence rates are highest for the age group 0-14 years and lowest for age group ≥ 65 years. In addition, 2003 - 2004 on average had the

Highest observed ILI incidence rates while 2007 - 2008 had the lowest rates.

To estimate the values of the unknown parameters of the dynamic model, Weighted Least Squares and Maximum Likelihood estimation methods were applied. In the Weighted Least Squares method, we assumed that the epidemiological system is exactly described by a dynamic model together with some set of parameters but the observed data arises from some deviation of the output of this system by observational errors. Then the values of the model parameters which minimizes the weighted squared errors (differences between the model predictions and the observed ILI incidence data) were sought, with the weights taken to be proportional to the corresponding denominator in each age group to account for the unequal population sizes represented by the different age groups. In the Maximum Likelihood estimation methods, it was assumed that the observations are Negative binomial distributed. This seemed appropriate since though it is natural to approximate a Binomial distribution by a Poisson distribution, the Poisson distribution can further be generalized to a Negative Binomial distribution for over dispersed data. Maximizing the likelihood function determines the parameters that are most likely to produce the observed data.

Depending upon the choice of the initial parameter values, the Nelder-Mead algorithm could prematurely stop and return a sub-optimal set of parameter values. Thus the dynamic model was first simulated using a set of starting values and SANN optimization method. This method is advantageous since it can overcome the problem of local maxima. In contrast, the algorithm may not be a feasible option as it may take unrealistically long time to find the solution. Thus the iteration count for the algorithm was set to 10,000. The Nelder-Mead algorithm was then run until convergence using estimates obtained using the SANN method as starting values. This optimization procedure was employed in both estimation approaches.

The parameter estimates obtained using the two different approaches did not differ much. Indeed, the values obtained for the amplitude, immunity waning rates and the proportionality factor were approximately the same. Though some differences were observed in the parameter estimates for the scaling factor, reference time and the seeding time, the differences were not so much pronounced. Thus the choice between the two methods of estimation can have non-trivial consequences.

References

1. E Vynnycky, R Pitman, R Siddiqui, N Gay, WJ Edmunds (2008) Estimating the impact of childhood influenza vaccination programmes in England and Wales. Vaccine 26(41): 5321-5330.
2. J Bilcke, P Beutels, M Brisson, M Jit (2011) Accounting for methodological, structural, and parameter uncertainty in decision-analytic models: a practical guide. Medical Decision Making 31(4): 675-692.
3. M Jit, M Brisson (2011) Modelling the epidemiology of infectious diseases for decision analysis: a primer. PharmacoEconomics 29(5): 371-386.
4. JD Tamerius, J Shaman, WJ Alonso, WJ Alonso, K Bloom Feshbach, et al. (2013) Environmental predictors of seasonal influenza epidemics across temperate and tropical climates. PLoS Pathog 9(3): e1003194.

5. G Hanquet, P Jonckheer, JV Yen, F Vrijens, NY Thir, et al. (2011) Vaccinatie tegen seizoensinfluenza: prioritaire doelgroepen. Deel I KCE reports 162A, Belgian Health Care Knowledge Centre, Belgium.

6. WHO (2012) Vaccines against influenza. World Health Organization position paper 87(47): 461-476.

7. T Suess, C Remschmidt, SB Schink, B Schweiger, A Nitsche, et al. (2012) The role of facemasks and hand hygiene in the prevention of influenza transmission in households: results from a cluster randomised trial; Berlin, Germany, 2009-2011. BMC Infect Dis 12: 26.

8. JE Aledort, N Lurie, J Wasserman, SA Bozzette (2007) Non-pharmaceutical public health interventions for pandemic influenza: an evaluation of the evidence base. BMC Public Health 7: 208.

9. Belgian Health Council (2010) Vaccination contre la grippe saisonniere saison hivernale2010 - 2011. Belgian Health Council, Belgium.

10. P Beutels, Y Vandendijck, L Willem, N Goeyvaerts, A Blommaert, et al. (2013) Seasonal influenza vaccination: prioritizing children or other target groups? part II: cost-effectiveness analysis. Belgian Health Care Knowledge Centre (KCE).

11. N Goeyvaerts, L Willem, KV Kerckhove, Y Vandendijck, G Hanquet, et al. (2014) Estimating dynamic transmission model parameters for seasonal influenza by fitting to age and season specific Influenza-Like-Illness. Epidemics 13: 1-9.

12. V Van Casteren, K Mertens, J Antoine, S Wanyama, I Thomas, et al. (2010) Clinical surveillance of the influenza A (H1N1) 2009 pandemic through the network of Sentinel General Practitioners. Archives of Public Health 68(2): 62-67.

13. Eurostat (2010) Population table for Belgium, Luxembourg, Belgium.

14. N Goeyvaerts, N Hens, B Ogunjimi, M Aerts, Z Shkedy, et al. (2010) Estimating infectious disease parameters from data on social contacts and serological status. Journal of the Royal Statistical Society: series C: applied statistics 59(2): 255-277.

15. J Wallinga, P Teunis, M Kretzschmar (2006) Using data on social contacts to estimate age-specific transmission parameters for respiratory-spread infectious agents. Am J Epidemiol 164(10): 936-944.

16. J Glasser, D Taneri, Z Feng, JH Chuang, P Tull, et al. (2010) Evaluation of targeted influenza vaccination strategies via population modeling. PLoS ONE 5(9).

17. S Cauchemez, F Carrat, C Viboud, AJ Valleron, PY Boelle (2004) A Bayesian MCMC approach to study transmission of influenza: application to household lon- gitudinal data. Statistics in Medicine 23: 3469-3487.

18. O Diekmann, JAP Heesterbeek, JAJ Metz (1990) On the definition and the computation of the basic reproduction ratio R0 in models for infectious diseases in heterogeneous populations. J Math Biol 28: 365-382.

19. KO Kwok, GM Leung, S Riley (2011) Modelling the proportion of influenza infections within households during pandemic and non-pandemic years. PLoS ONE 6(7): e22089.

20. S Kirkpatrick, CD Gelatt, MP Vecchi (1983) Optimization by simulated annealing. Science 220(4598): 671-680.