HELLENIC REPUBLIC
National and Kapodistrian
University of Athens

EPIDEMIC MODELS AND THEIR APPLICATION IN THE ANALYSIS OF INFLUENZA OUTBREAKS

Student

KOYTSOSPYROY CHRISTINA

ATHENS, 2020

The present dissertation was prepared in the context of the studies for the acquisition of the Postgraduate Specialization Diploma in

BIOSTATISTICS

awarded by the Medical School and the Department of Mathematics of the National & Kapodistrian University of Athens.

Approved on *29/03/2021* by the selection board:

NAME	DEGREE	SIGNATURE
N.DEMIRIS(Supervisor)	**AS.PROFESSOR**	
G.TOULOUMI	**PROFESSOR**	
V.SYPSA	**AS.PROFESSOR**	

Table of Contents

1. Introduction

Epidemiology as a branch of the Mathematical science was developed in the 20th century, with the parallel development of technology and mathematical concepts in general. The hypothetical modeling in a strict mathematical framework led to a plethora of substantial results in the epidemiology science.

The mathematical tools that are used are many, some of them include the study of dynamic systems, the resolving of differential equations, the study of continuous functions, the theoretical computation e.t.c (Kermak et al., 1927).

In many different parts of biology, mathematical models have been developed. These models are tools that are applied to many scientific field studies to calculate a variety of circumstances. Modeling of diseases and epidemics were present many years ago, but nowadays they are applied to internet viruses, marketing and data mining. This is due to the nature of the continuous development of Internet and because the mechanics of the virus spreading and the nature of the media are different. Therefore, the modeling of epidemics in computer networks can affect the modeling of epidemics in human biology (Dushoff et al., 2004).

Epidemic models have two main targets. The first is concerned with the representation of the spreading mechanism and estimation of the relevant parameters. The second target is to use such estimates for prediction spreading and control. (Stone et al., 2007).

The mechanics of spreading define how exactly a virus spreads in the general population. For example, in some diseases the infections can be spread through the air while in other it can be spread through saliva or blood. In this basis, the most studied models of epidemics are SI (Sensitive – Infected), SIS (Sensitive – Infected – Sensitive) and SIR (Sensitive – Infected – Recovered). In the SI model, the part of the population that is sensitive in infections, when infected they remain in this situation for ever. In the SIS model, the persons that fall in this category come through three situations, form sensitive they become infected and then again sensitive. The spreading of the common influenza virus can be modeled with the SIS model. In the case of the SIR model, a person spends some time in the infected state and then the

person either dies or recovers his immune system in such a degree that he cannot spread the disease further. In other words three situations exist for a person in these models: Sensitive – Infected – Recovered with immunity (Chowell et al., 2006).

When a disease is modeled, first we must have a strict description of the sum of the characteristics of it spreading in the population and secondly, we must choose a certain mathematical approach to describe the spreading. Even though the spreading of a disease is a statistical process by nature, there are deterministic models that are often used to predict its outcome.

We want to estimate the number of cases of seasonal influenza in Greece based on the package fluevidenceSynthesis. It is important to have an idea of how many people get sick each year in our country. With the advent of the corona virus we need at all times free beds of intensive care units, so it would be very useful to reduce the inflows to them from other diseases. Therefore, knowing the estimated number of patients, we can adjust the vaccination scenario accordingly, in order to reduce the inflows from influenza to the intensive care units and to the hospitals in general. By using this package we also try to predict the efficacy of the different vaccination scenarios and to infer epidemiological parameters.

1.1 Main Concepts

Epidemic modeling of diseases in a population is based on the mass action law. This law declares that in a mixed population, the number of contacts is proportional to the product of the infected and the sensitive. In other words, the rate of the contacts that affects the infection is directly related to the sum of sensitive persons and the sum of the infected persons. In our analysis the contacts we used are from the United Kingdom. The mass action law is a superposition of all possible situation and the partial characteristics to calculate the contacts (infected with healthy persons). In case there are more than one process that are linked to the calculation of the contacts, then these processes are added to the variables of the system. It is therefore important to understand some basic epidemiological concepts (Mills et al., 2004).

Initially, a model can be deterministic or stochastic. Stochastic is the model that calculates all the possible outcomes based on the possibilities. This kind of model

is applied to small populations where there are fewer factors that affect the environment and there is a smaller possibility for changes from external factors. In a deterministic model, advanced mathematics used to split the population in smaller parts for more specific results and calculates different stages of a disease. The model we use is deterministic since we divide the population into different age groups. In table 1 we describe the most important symbolisms that are described in the epidemic models and their meaning.

Table 1: Description of symbolisms used in an epidemic model

R_0	Basic reproduction number
S	Sensitive
E	Exposed in the incubation period
I	Infected
R	Recovered and with immunity
β	Possibility of contact
μ	Average of mortality
$1/\varepsilon$	Average period of incubation
$1/\gamma$	Average period with risk of contagion
N	Total sum of the studied population
f	Mean of immunity elimination for persons that were infected

In terms of demographic data, the Wi number of people in the population aged i is obtained from the Hellenic Statistical Authority (https://www.statistics.gr/). 5 age groups (1-5, 5-15, 15-45, 45-65, 65+) were examined divided into high and low risk. People at risk of chronic respiratory, heart or kidney disease, diabetes or immunosuppression due to illness or treatment are considered at high risk.

We used English contact data collected from a survey conducted in 2006 and participants recorded their contacts in one day. The age of the contact, the nature of the contact (conversation or physical contact) and the nature of the day (daily, weekend or holidays) were also recorded.

Through the system of observational diseases in primary health care, diseases are observed for diseases that are similar to the flu but are not (ILI). Volunteer private physicians specializing in pathology, pediatrics or general practitioners from all over the country participate in the system, where they report supervised diseases on a weekly basis. Here we have the results from the 1st to the 52nd week of the year 2013.

1.2 Basic rate of reproduction

The basic rate of reproduction is used to describe if a disease is able to cause an outbreak. It is a mathematical symbol that helps us understand how contagious a disease is. There are three possible values for the basic rate of reproduction.

1. $R_0 < 1$. In this case the virus is not a danger for epidemic
2. $R_0 = 1$. Theoretically, in this case only 1 person is infected by an infected person.
3. $R_0 > 1$. There is a danger for an epidemic because each infected person can infect one or more healthy persons.

In the last case the danger is obvious, because the possibilities of the disease spreading are increased exponentially.

Generally, the most important factors that we must take into account when we calculate this number is:

The infection period. This factor is defined by the dynamics of the virus spreading. Some diseases are more contagious than others and for different time frames. Furthermore, an important factor is the population that the virus is incubated. For example, the influenza virus can be maintained for 8 days in adults and for 2 weeks in children. For this reason, when a virus infects a child, it is more possible to infect more persons (Vynnycky et al., 2008).

Possibility of contact. This is basically the possibility of an infected person to contact a healthy person. For this reason, in order to avoid pandemics the patients must be confined in closed spaces.

Transmission media. The media is the means that the virus uses to be transferred from one person to another. The diseases that are transmitted faster are the ones that are transmitted through the air, as the influenza virus. Physical contact is not necessary and for this reason the possibilities of transmission are increased exponentially. The disease that require physical contact for their transmission, for example the HIV virus, are mostly unlikely to lead to a pandemic (Ferguson et al., 2006).

2. Greek Influenza data

Influenza is one of the biggest problems because it is one of the main causes of death in developed countries, with more than 1,000 deaths in 1,000,000 of the population. The population percentage that is 65 years old or older has greater chance to die from influenza infection. Furthermore, an influential pandemic increases the rate of hospitalization and the days that people are absent from their jobs (EODY, 2019).

Influenza is an acute disease of the respiratory system and is usually consists of the types A, B and C. It is a contagious disease in birds, mammals and humans. The most common symptoms of the disease include fever, sore throat, muscle pains, severe headache, cough, weakness and general unwellness. It cannot be differentially diagnosed with other acute infections of the respiratory system according to clinical symptoms. For this reason, laboratory examinations are important for the disease's confirmation (Urban et al., 2009).

Every winter in Greece, like every other country, has an increase in the activity of influenza virus. The increase in morbidity and mortality rates varies each year, depending on the characteristics of the virus and the degree of immunity that certain groups have towards this specific virus type or types.

The influenza virus is constantly mutating. If the changes in the gene profile of the virus are substantial, there is no immunity in the population, and this particular stem of the virus can create an outbreak. A pandemic is a situation where a large number of people are infected at the same time, and it creates problems in the National Health System while it also obstructs the economic and social activity. In order to have a

picture about the activity of the influenza in Greece, EODY has systems of epidemic observation of influenza. These systems are supplementary with each other because influenza affects the population in various degrees. The levels of these systems are depicted in Table 2.

Table 2: Levels of the supervisory systems of the influenza virus in Greece

Level	Description
Asymptomatic patients	Supervision of vaccination coverage
Symptomatic patients that did not ask for medical attention	Laboratory supervision
Patients that asked for medical attention	Laboratory supervision
Patients inside a hospital	Sentinels of morbidity system, laboratory supervision
Patients in Emergency Treatment	Laboratory supervision, supervision of serious cases
Deaths	Supervision of general morbidity rates

Some of the above systems are in function the whole year (like the sentinel system) but the majority of them function, as in most European countries from the 40th week of the current year until the 20th week of the next year. This time period is also known as influenza supervision period. In Greece the influenza virus supervision has shown that the time of year with increased activity is during the months February – March. We must, however note that influenza is unpredictable and the start and the duration of the seasonal outbreak differs from time to time. Furthermore, we must mention that influenza activity is never zero, because there are cases of the disease throughout the year (Pogka et al., 2011).

In Figure 1 we can see the number of influenza cases per 1.000 hospital visitations every week in the years 2017-2018. The influenza activity was increased in the time period from week 52/2017 until week 8/2018. In comparison to the previous year the epidemical raise was milder and it peaked 7 weeks later (EODY, 2019).

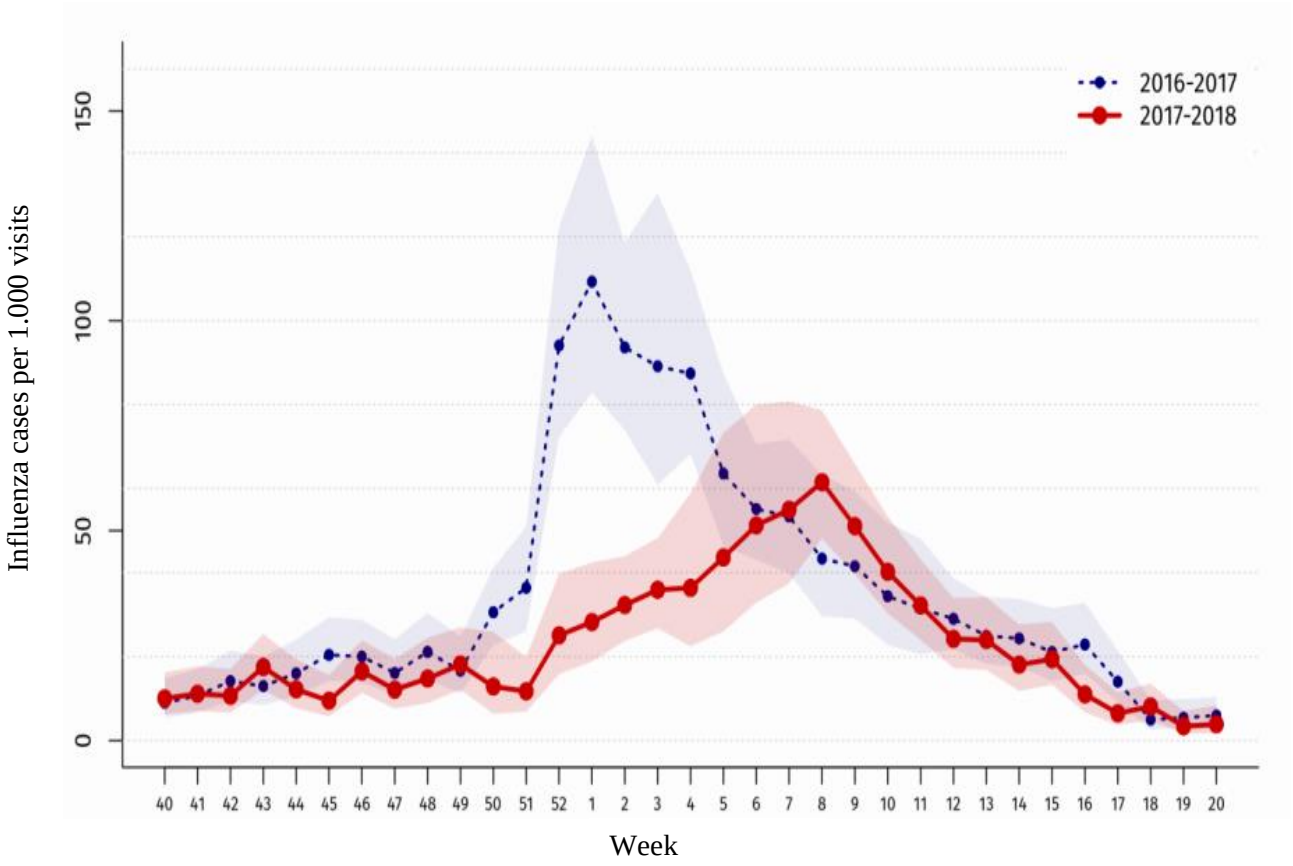

Figure 1: Estimation of influenza cases per 1.000 hospital visits per week. In the vertical axis the number of cases are depicted and in the horizontal axis the weeks.

In Figure 2 there is the same estimation as in Figure 1 but according to the different age groups. From the figure we can see that the activity of the influenza virus was similar in almost all cases (EODY, 2019).

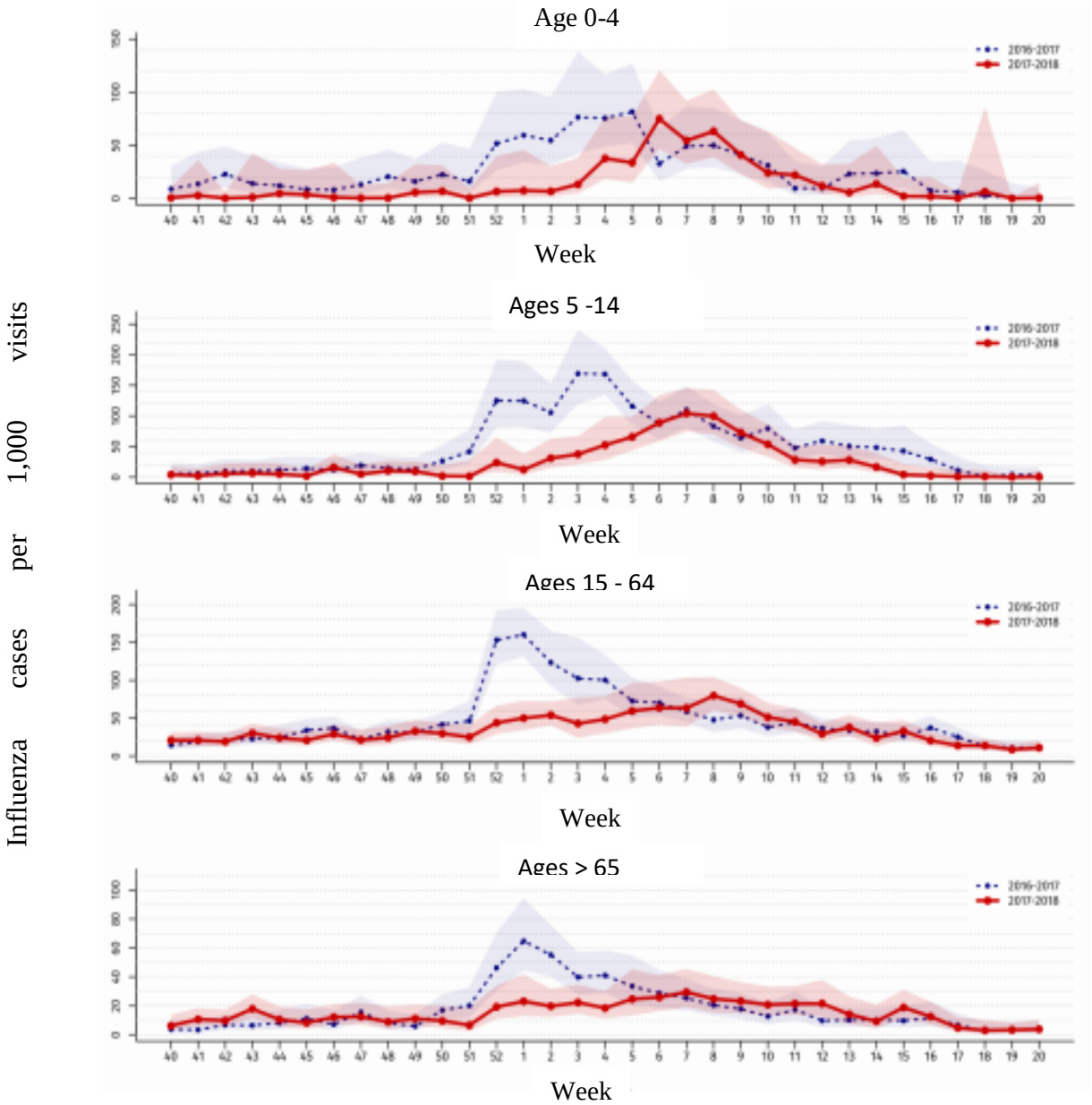

Figure 2: Estimation of influenza cases per 1.000 hospital visits per week in different age groups. In the vertical axis the number of cases are depicted and in the horizontal axis the weeks

During this period 2956 clinical samples were checked in the laboratories and more specifically 2694 from public hospital laboratories and 262 form the Sentinel network. 504 (17,1 %) from them were positive in influenza viruses and 145 (28,8%) were positive in the type A of the virus while 359 (71,2%) were positive in

the type B of the virus. The 125 stems of the A virus were further identified and 17 (13,6%) were of the A (H3N2) subtype and 108 (86,4%) of the A (H1N1) subtype.

In Figure 3 we can see that until the 10th week of 2018, the type of the virus that was the most prevalent was B. In the following weeks, the influenza virus activity was decreased, but there were also cases of type B virus, subtype A (H1N1) and subtype A(H3N2) (EODY, 2019).

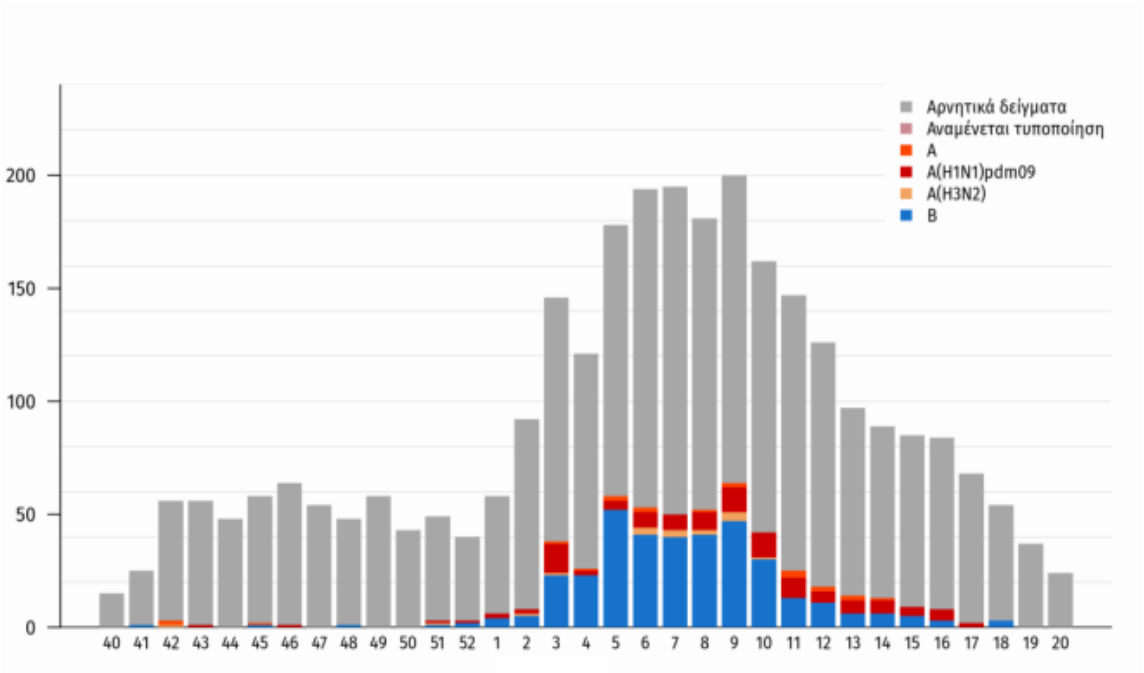

Figure 3: Total number of pharyngeal samples in the laboratories used to monitor the influenza virus activity. The negative samples are depicted in grey color, the A type with red, the A (H1N1) with the pink color and the B type with the blue color.

For the Influenza like Illness (ILI), which is a flu that looks like seasonal influenza but it isn't, we used data from 2013 that were available in EODY. The total number of visits to doctors in the whole country that year was 145.432 and the ILI was 3.120. In Figure 4 we have the percentage of the ILI cases.

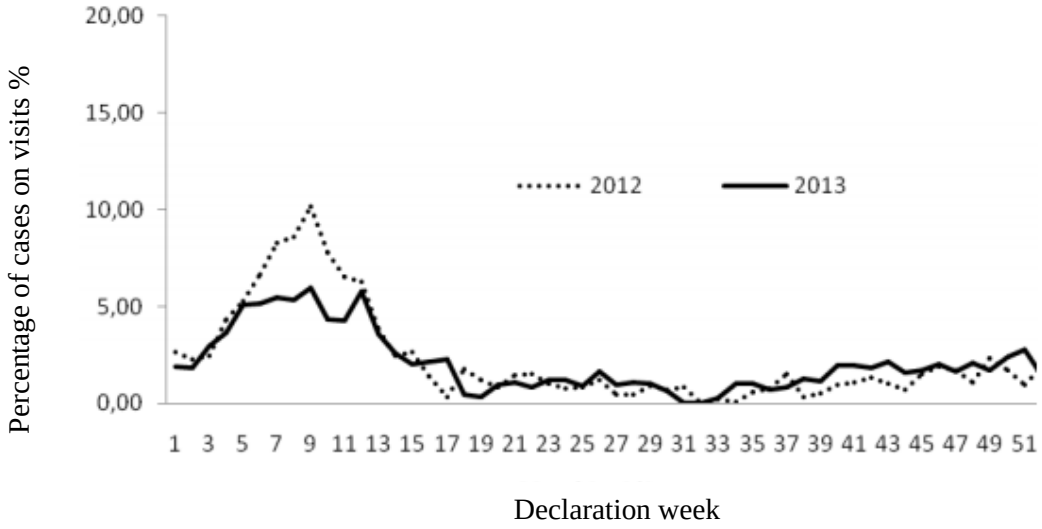

Figure 4: Percentage of ILI cases, in the total number of visits per week of declaration in Greece in 2012-2013 (EODY, 2019)

In Figure 5 we see the number of the ILI cases for the different age groups in 2013.

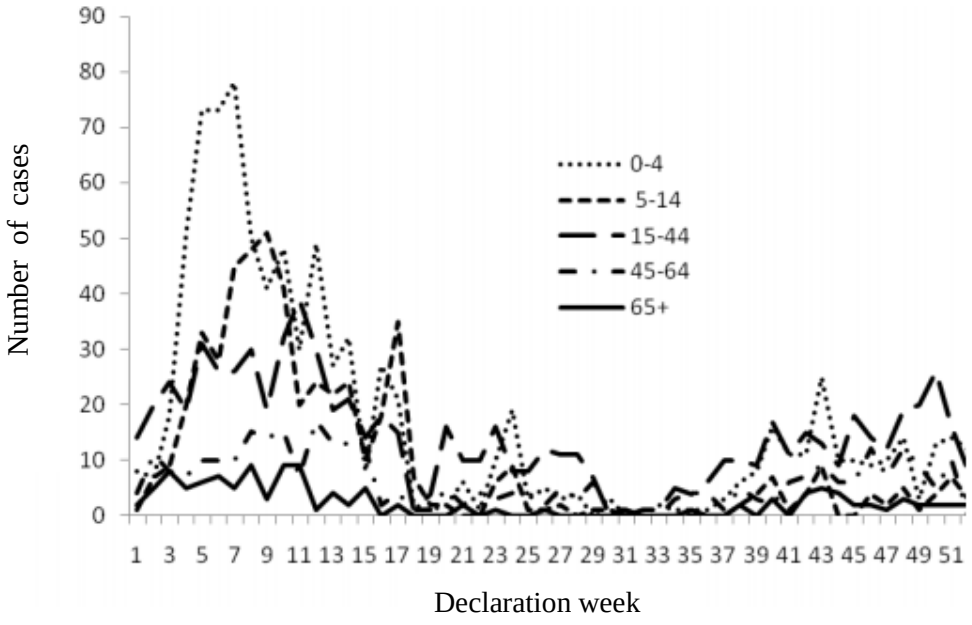

Figure 5: Number of ILI cases per age group and week of declaration in Greece (EODY, 2019).

3. Modeling framework

The overall epidemiological model is seen in Figure 6 and consists of the transmission matrix, the rate of the immunization, the profile of those susceptible to the infection, and the initial number of infections. A Bayesian approach to statistical conclusions was adopted. We still used adaptive techniques of the Markov Chain Monte Carlo (MCMC). The combination of the Bayesian approach with the MCMC methodology is suitable for the natural propagation of uncertainty.

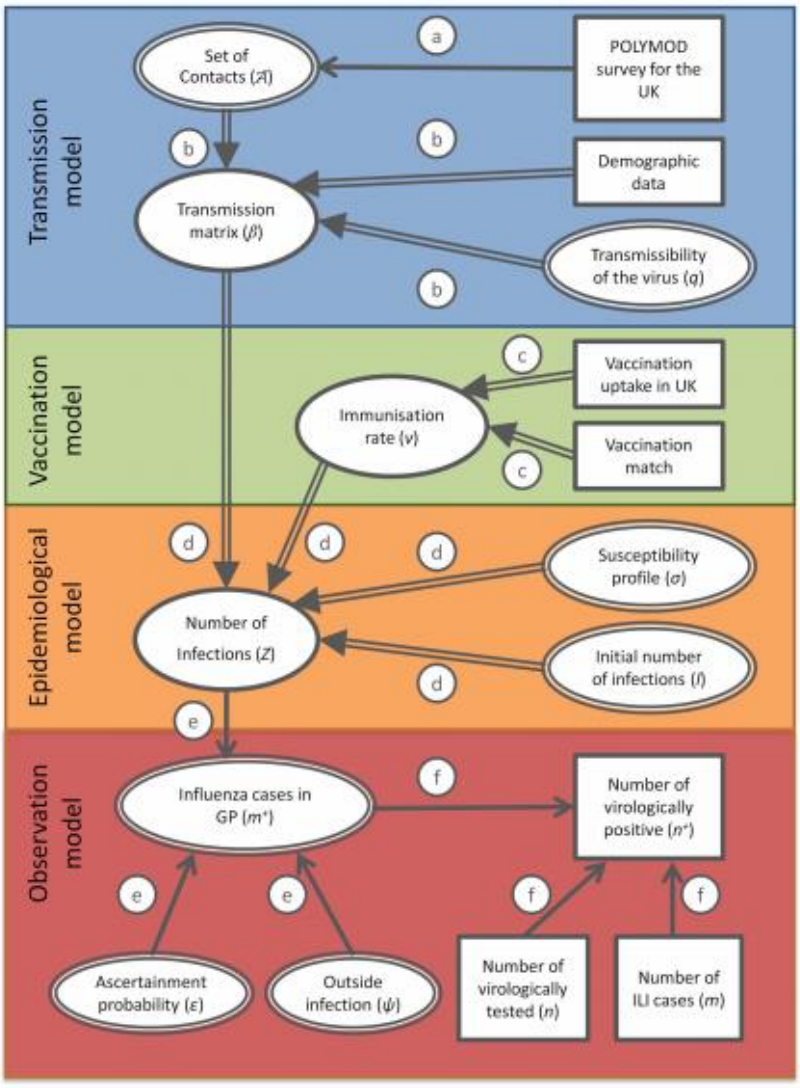

Figure 6: Flowchart that shows the connection between the different modeling components.

3.1 Transmission Model

The transmission of infectious diseases is linked to the form of the contacts within the community. The mathematical modeling assumes that when a healthy individual comes in contact with an infected one, there is a probability (q) of being infected. This probability is affected by a number of factors such as the type of contact (physical contact is more contagious than a conversational contact in the case of a virus that spreads through the respiratory system) and the type of the pathogen.

The equation that describes the re-normalized average number of contacts per day from a certain group is standardized for age and weekdays (1).

$$d_{ij} = \frac{\sum_{k:A_k \in j} N_k^i w_k}{\sum_{k:A_k \in j} w_k} \quad (1)$$

The contact is a process defined by symmetrical characterization and therefore, the number of contacts, of group *i* coming in contact with group *j* is the same as the number of contacts of group *j* with group *i*. Furthermore, if we define the possibility that two random persons from the two groups come into contact as c_{ij} then we must also take note that $c_{ij} = c_{ji}$. To further correct the probability of contact between an infectious person and a healthy person (the symmetry is usually not achieved due to participation biases) we provide the equation (2) where T_i and T_j are the number of participants into two groups and d_{ij} and d_{ji} are the number of contacts per day of in the corresponding participant group (Wallinga et al., 2006).

$$c_{ij} = \frac{1}{2}\left(\frac{d_{ij}}{T_i} + \frac{d_{ji}}{T_j}\right) \quad (2)$$

3.2 Vaccination Model

The model of the influenza epidemiology is further divided into two groups based on the vaccination (we define as N the non – vaccinated group and V the vaccinated group). However, since the vaccine is never 100% successful there is a factor a_i in the vaccinated group that is not immune. The rest of the population in the vaccinated group is protected ($1-a_i$). The effectiveness of the vaccination process is affected by a number of factors such as the mean age in the group, the virus strain and the match between the vaccine and the virus strain (Baguelin et al., 2010).

There is a period of two weeks between someone being infected and developing antibodies for this particular virus strain. Develop of the antibodies are not included in the mathematical modeling because it mainly focus on the vaccination and not the infected individuals (Miller et al., 2009).

The resulting rate is the immunization rate v_{ik} that is produced by combination of the vaccination intake and the vaccination success in the groups i and k (risk group). This rate is over a monthly period.

3.3 Epidemiological Model

The epidemic model that is used is characterized as SEIR (susceptible – exposed-infected-recovered) model. It is also assumed that the inherent population immunity does not fully protect the virus spread but rather reduces the possibilities of infection. The model assumes that during an influenza outbreak that is characterized as epidemic a portion of each of the studied groups is affected and the remaining population is subject to an infection. The original portion of the group that is infected is obtained by modifying the initial population by a factor I. A profile due to age susceptibility is also assumed $\{\sigma_i\}$. The profile characteristics are taken from the influenza outbreaks that took place in the previous years. The groups are split according to age categories of the sample population. More specifically to avoid overflowing the model with many groups, only 3 age groups are defined: children (0-14), young adults (15-64) and elderly population (over 65) (Johnson et al., 2009).

The equations of the epidemiological model are given below:

$$\lambda_i = q\sigma_i \sum_{j=1}^{7} \sum_{k=1}^{2} \sum_{X=\{N,V\}} c_{ij}\left(I_{jk}^{1X} + I_{jk}^{2X}\right) \quad (3)$$

Where q is the parameter that describes transmission, c_{ij} is the rate of the contacts between those in age group i and age group j and σ_i is how susceptible is the i group.

The new infection relative incidence in age group i and risk group k in week n is given by the equation:

$$Z_{ik}(n) = \int_{7(n-1)}^{7n} \gamma_1\left(E_{ik}^{2V} + E_{ik}^{2N}\right) dt \quad (4)$$

There is a version of the SEIR model, the SEEIIR resulting in a more realistic gamma distributed waiting time between exposed and infected, and between infected and recovered. Its form is:

$$\frac{dS_{ik}}{dt} = -\lambda_i S_{ik}$$

$$\frac{dE_{ik}^1}{dt} = \lambda_i S_{ik} - \gamma_1 E_{ik}^1$$

$$\frac{dE_{ik}^2}{dt} = \gamma_1\left(E_{ik}^1 - E_{ik}^2\right)$$

$$\frac{dI_{ik}^1}{dt} = \gamma_1 E_{ik}^2 - \gamma_2 I_{ik}^1$$

$$\frac{dI_{ik}^2}{dt} = \gamma_2\left(I_{ik}^1 - I_{ik}^2\right)$$

$$\frac{dR_{ik}}{dt} = \gamma_2 I_{ik}^2 \quad (5)$$

Where S_{ik} is the number of sensitive in the age groups i and risk group k, E_{ik}^1 and E_{ik}^2 are two groups that correspond to two groups of exposed but not infected individuals. Furthermore, I_{ik}^1 and I_{ik}^2 are infected individuals and R_{ik} are the immune individuals. The force of infection is λ_i is depended of age and is given by the following equation:

$$\lambda_i = \sigma_i \sum_{j=1}^{x} \sum_{k=1}^{y} \beta_{ij} \left(I_{jk}^1 + I_{jk}^2 \right) \qquad (6)$$

Where β_{ij} is the effective rate between individuals in age group i and age group j and σ_i is the susceptibility of age group i.

Finally the interference is given by a likelihood function that is incorporated into the package:

$$L(n_i^+, n_i, m_i | \varepsilon_i, \psi, \theta_i) = \sum_{m_i^+} L(n_i^+, n_i, m_i | m_i^+, \theta_i) L(m_i^+, m_i | \varepsilon_i, \psi, \theta_i) \qquad (7)$$

3.4 Observation Model

The final model had to connect the investigation data with the number of infections that occur due to the spreading of the influenza in the studied groups. We must also note that although data obtained for General Practitioners are used to monitor an epidemic of influenza, these data are not always correct. There are many cases that patients are recorded as influenza patients because of their symptoms but they are not actually affected by influenza. If all of the patients recorded by the doctors as having influenza are examined by PCR for actual virus DNA a certain number will not be positive. For this reason, statistical model is necessary to connect the data between surveillance and infections in an influenza epidemic (Carrat et al., 2008).

The sets of the surveillance system, are depicted in Figure 7.

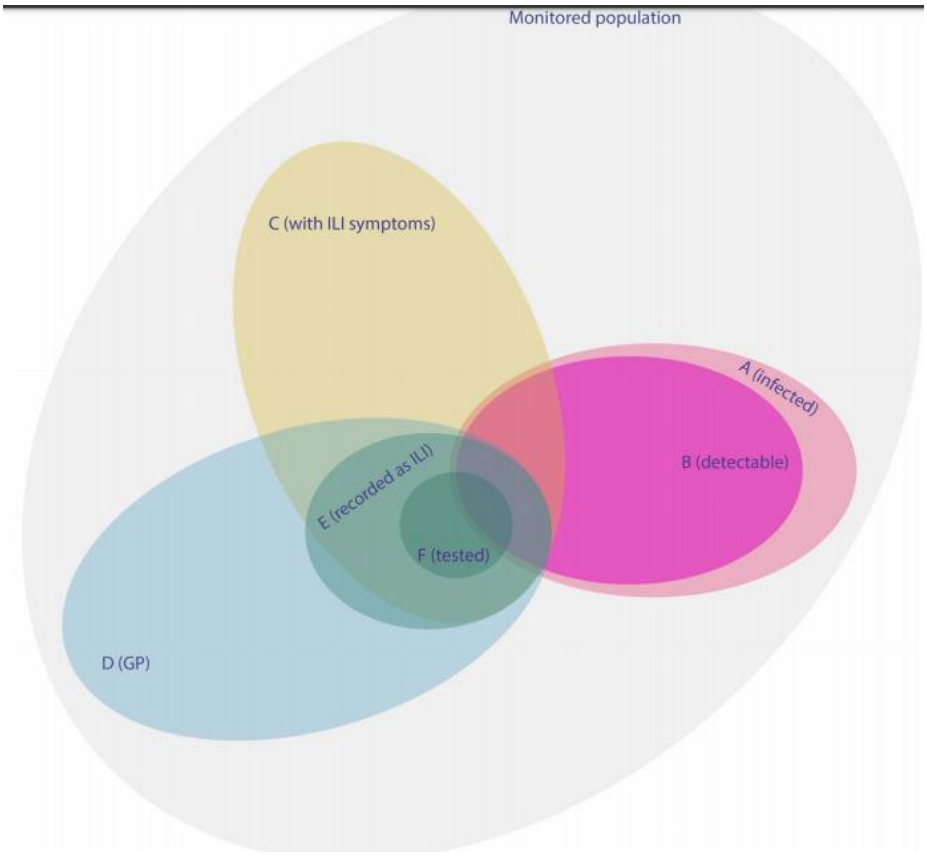

Figure 7: A schematic illustration of the various parts of the surveillance system.

In the countries that have temperate climate, the reoccurrence of influenza every year is determined by the traveling population. This fraction of the population is infected with new variants of virus strains that are globally transmitted. The global influenza outbreaks seem to be consistent with the reoccurrence of outbreaks in certain regions. It is proven that the epidemics of the influenza virus in each country are directly correlated with the re-infection of the population by people that travel in other countries (Flasche et al., 2011).

For this reason the probability of each group being infected is directly associated with a fraction of the population that travels in other countries or with an outbreak of the virus inside the country that is not linked to an international outbreak. This risk is defined as ψ and is not age or time dependent.

$$z_{ij}^{\theta} = \left\| \frac{N_{ij}^{mon}}{N_{ij}^{tot}} \sum_{k=1}^{2} Z_{ik}(j) \right\| \quad (8)$$

Where θ are the parameters of the model and the incidence of the population of group i at week j of new cases, is defined by the equation (8). The parameters that describe the infection are the q, σ_i, I and A, and N_{ij}^{mon} and N_{ij}^{tot} are the monitored and total population sizes in age group i and the week j.

There is also the probability of infection from the traveling population that is given by the equation (9).

$$z_{ij}^{outside} \sim Binomial\left(N_{ij}^{mon}, \psi\right) \quad (9)$$

In order to certify that a patient is infected by the influenza virus two steps must be considered. The first step is to be characterized as an influenza case by the GP and the second step is to be confirmed by laboratory examinations. We can see that these steps have as a result that the actual cases of influenza are a much smaller number that the cases recorded. These number of cases, that are infected outside of the influenza are described by the equation (10).

$$m_{ij}^{outside} \sim Binomial\left(N_{ij}^{mon}, \psi_{\varepsilon_i}\right) \quad (10)$$

Where, ψ_{ε_i} is a small number and N_{ij}^{mon} is big. The probability can be defined by a Poisson distribution with rate the product of $N_{ij}^{mon}\psi_{\varepsilon_i}$

$$m_{ij}^{outside} \sim Poisson\left(\psi_{\varepsilon_i} N_{ij}^{mon}\right) \quad (11)$$

In a certain group we are interested in the m_{ij}^{+} persons that are categorized as influenza cases in week j and group i. These persons n_{ij}^{+} are detected and found positive for an influenza strain. The positive number of cases that are expected are described by a hypergeometric equation and the model is described by (12) system of equations (Baguelin et al., 2013).

$$\begin{cases} m_{ij}^{+} \sim Binomial\left(z_{ij}^{\theta}, \varepsilon_i\right) + Poisson\left(\psi_{\varepsilon_i} N_{ij}^{mon}\right) \\ n_{ij}^{+} \sim Hypergeometric\left(n_{ij}, m_{ij}^{+}, m_{ij}\right) \end{cases} \quad (12)$$

Finally, the ε can be described as a product of epidemiological quantities.

$$\varepsilon \approx P(E|A \cap C \cap D)P(D|A \cap C)P(C|A)P(B|A \cap C) \quad (13)$$

We can see an example of the results of the computational model when applied to the influenza outbreak in Figure 14 in the Appendix.

4. Computation

4.1 MCMC

The Markov Chain Monte Carlo algorithm (MCMC) methods were invented in the 1990s when computers increased the computing power. There are numeral ways to implement an MCMC algorithm but the most important algorithms estimate posterior distributions of Parameters in a Bayesian model and are the following:

1. Metropolis
2. Gibbs
3. Hamiltonian.

The algorithm is used as a way to correct the shortcomings of the grid approximation error techniques. More specifically, grid approximation does not scale well with the number of parameters in a vector of parameters to compute the posterior distributions. Moreover, quadratic approximation even though it can be scaled better than the grid approximation it also has shortcomings with complex, hierarchical models.

The main concept behind the MCMC algorithm is that if we design and implement a strategy of careful planning, we can be sure that the sample distribution is representative of the target posterior distribution. This strategy is similar to the sampling methods that are applied to the design of a survey or a poll in political and social sciences (Fork et al., 2018).

In other words, the MCMC methods are used to determine the posterior distribution of a parameter that we are interested in by sampling with a certain strategy in a probabilistic scale. This is done by number of simulations that repeatedly generate random numbers. The simulations approximate a parameter in cases where the calculation of this parameter is very difficult. The second parameter that is important to MCMC algorithms are Markov chains. Markov Chains are sequences of events that are related to each other through probabilities. Each event results from a collection of effects and each effect defines the next result, in regards to a certain set of probabilities.

The MCMC method can also be used to estimate the posterior distribution of more than one parameters. For a set of k parameters, in a space with k – dimensions there are areas with high probabilities. These sets of parameter values explain a number of observed data. The MCMC methods are a way collect samples from a probabilistic space to approximate the posterior distribution (Figure 8) (Hill et al., 2019, Cowles et al., 1996).

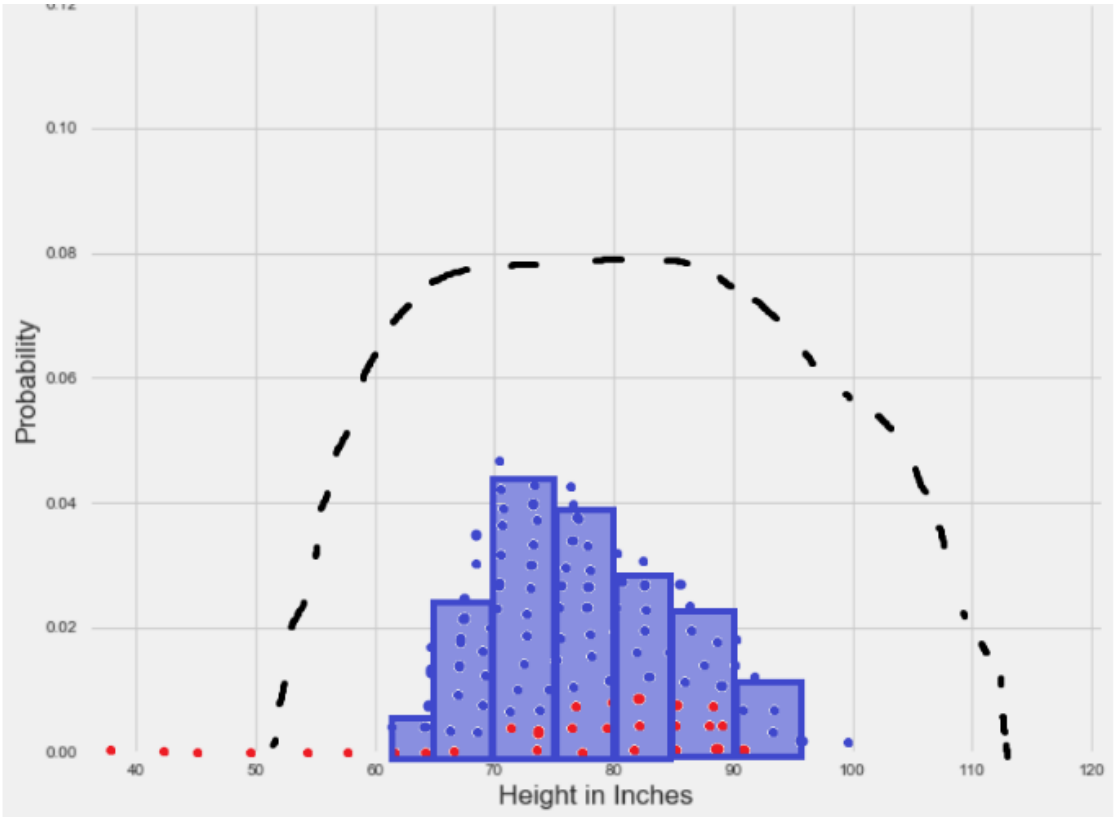

Figure 8: MCMC example where the parameters values (x-axis) exhibit areas of low and high probability. After convergence, MCMC gives a set of values (points) that represent samples from a posterior distribution. In this example the distribution depicts the average human height in inches.

4.2 R package

The package that we will be using is an R based package. The fluEvidenceSynthesis package is a method to analyze epidemiological outbreaks. In Figure 9 there is a flowchart that depicts the general data.

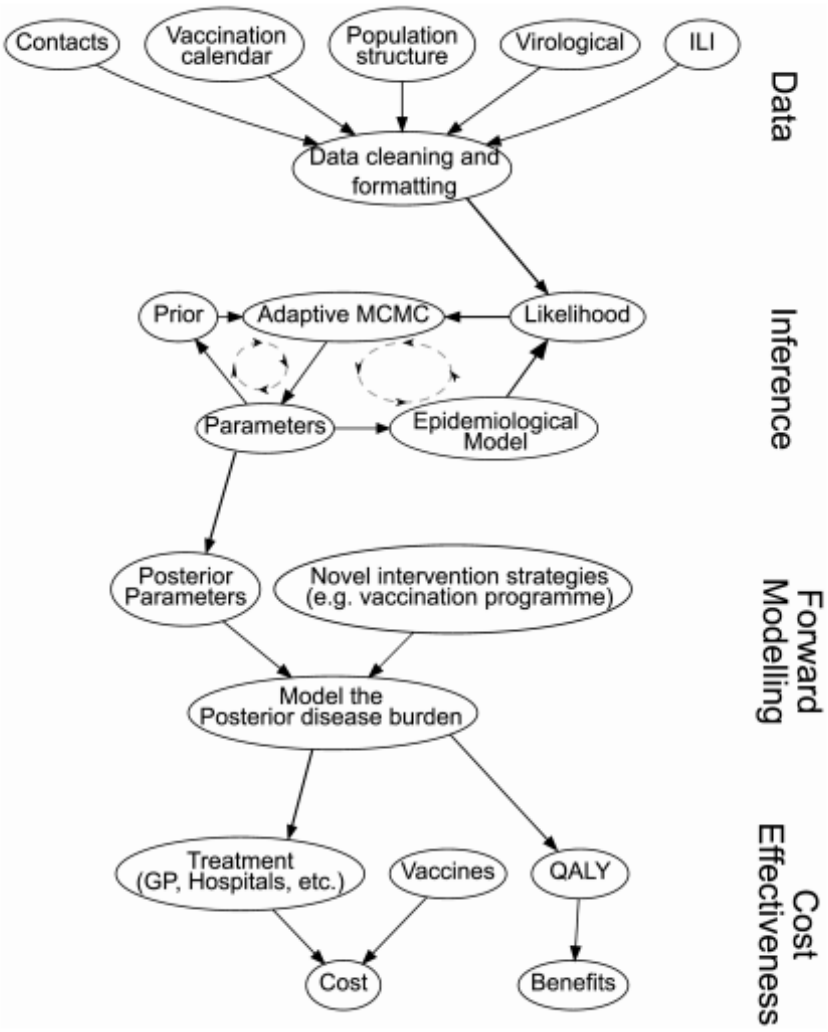

Figure 9: The package initially takes the influenza data that relate to the outbreak as an algorithm input. These data are contacts, the vaccination calendar, the population structure (demographic data), the virological data (the type of viruses) and the ILI cases (Influenza Like Illness). The next step is the calculation of likelihood to observe these data. Then, when a certain set of data is finally selected the algorithm uses an MCMC model to find a set of parameters that correspond to the epidemiological model. These parameters are the output of the package and can be used to plan treatment strategies. These treatment strategies are selected as a cost effective interventions that provide the highest benefits (Leewuen et al., 2017).

The input data of the algorithm are the demography, which is a vector with the number of individuals of the United Kingdom, the coverage, that is data about the vaccination cover in the period 2007-2008, the polymod_UK, that are contact data for the United Kingdom, the ILI, which is the number of ILI cases per week, the confirmed samples, which is the number of confirmed positive samples per week and

finally the vaccine calendar, which is the vaccination rate in the United Kingdom in 1999. The input data are organized in a weekly basis. The data are represented as a table where each row corresponds to a week and each column separated by age group or by risk factors, if there are any to take into account. A typical set of data that can be used as an input can be

- Weekly ILI counts organized by age groups
- Virological data organized by age groups
- Vaccination data organized by risk group
- Population size organized by age
- Contact data.

The first step of the algorithm is to organize the data in different groups. This separation is made in regards to age and risk groups. Then, the algorithm defines vaccination methods by taking into account the effectiveness of the vaccine against the most prominent virus strain. The effectiveness also depends on the age (less effective in older age groups). The coverage of the population is different in every country because each country has a different vaccination program and policy. There is an appropriate function in the algorithm that calculates the effectiveness of the vaccination (Balguerin et al., 2015).

The package also provides a number of functions that estimates the cost effectiveness of the therapy. These functions are based on the existed data of mortality rates and cases that have to be hospitalized. The number of vaccines that need to be administered is also calculated. However, many of these costs differ according to the country and the vaccination program of each healthcare system (Sherlock et al., 2010).

5. Application to the Greek influenza data

In the current thesis we apply a dataset of cases in Greece in the period of the first 32 weeks of 2017 – 2018 to the fluEvidenceSynthesis package, to estimate the number of patients with influenza. We have used the Greek demographic data, the confirmed positive cases and the ILI from the year 2013. For the rest of the input data we used what the package provided. Our data is organized in 5 age groups (0-5, 5-15,

15-45, 45-65 and 65+). We also provide statistical evidence for the number of deaths of the patients that were vaccinated. Also there is demographic evidence about the sex, the type of virus that the patients were infected with (A, A-H1N1, A-H3N2, B) and how much of the patients needed to be hospitalize in emergency care. In Table 4 in the Appendix we provide a sample of our data.

To predict the effectiveness of different vaccination scenarios the SEEIIR model is used, with exposed and infected divided into two different groups. We also look at how many have been vaccinated or not and the effectiveness of the vaccine. We sort the population by age and risk group, and depending on the group they belong they are being vaccinated in different rates. So we apply the epidemiological model with parameters: population, initial infected, vaccine calendar, contacts, susceptibility, transmissibility, infection delays and an interval of days.

- The population, is the Greek demographic data stratified by age (<65, 65+) and by risk (where in the high risk we have 1% for the under 65, and 40% for the 65+)

- The initial infected, is the number of individuals being infected at the beginning of the season stratified by age (<65, 65+) and by risk (again for the high risk we have 1% for the non-elderly and 40% for the elderly). We assumed that the infected at the beginning of the season are 1.000 for each of the two age groups.

- We create the vaccine calendar, which is a list with the calendar and the efficacy of the vaccine for that year (0.7 for the non-elderly and 0.3 for the elderly in each risk group), by assuming that a constant percentage is vaccinated for four months, where the elderly and the high risk groups are being vaccinated at the highest rate and the low risk of non-elderly are not being vaccinated.

- The contacts, is a matrix of the rates of contacts of the different age groups, created by the Greek demographic data with the contacts of England.

- The epidemiological parameters are the susceptibility, where we assume that it is different for the different age groups (0.7 and 0.3 for the non-elderly and the elderly respectively). The transmissibility, where we assume that is 0.17 for both age groups. Finally, the infection delays is the average time from exposed to infected (0.8 days) and from infected to recovered (1.8 days).

- The interval is the time we want to integrate so we put it 7 days because we have a data point each week and we want to model all infections during that week.

For the analysis and the comparison of the efficacy of the different vaccination programs about influenza we have two steps. First, is the inferences of the parameters by using the existing model and second, the simulation of the different possible vaccination strategies using these inferred parameters.

For the first step, we use combined data to calculate the likelihood of the predicted number of cases from influenza in a given week. We use MCMC to take the posterior distribution of the parameters of the epidemiological model. Two functions are defined from the package to execute inference parameters, one that returns the log likelihood for given parameter values depend on the data and one that returns the log prior probability of the parameters. After that, these functions are transported to the adaptive MCMC, which returns a posterior sample for the parameter values. We run the model for 7 age groups and 3 risk groups, given the parameters. Then, we convert the age groups from 7 to 5 so that the structure matches with the ILI and the confirmed samples data. Last, we calculate the probability of the results given the ILI and the confirmed data.

For the second step, we create a vaccination scenario for the first 4 months which determines the percentage of the vaccine per day at each age and risk group, and the efficacy of the vaccine during a specific period. This vaccination calendar is made for 7 age groups and 2 risk groups by assuming that high risk young children and over 65 are being vaccinated. So we set the efficacy at 0.7 for the under 65 and 0.3 for the 65+ in all risk groups. We also define the percentage of the coverage of the vaccine for the high risk young children (0-5, 5-15) which are 0.62 for the first 3 months, and the elderly (65+) which is 0.62 the first month, 0.77 the second and 0.925 the third month. For this vaccination scenario we get the full posterior of the cases and then we set new vaccination coverage, up to 80% for all age and risk groups.

5.1 Data sources

Our analysis is based on the fluEvidenceSynthesis package, so we tried to replace as much input data it was possible, from those used in that package. In the table below you can see what replacements have been made and where we found our Greek data. Some of our data made approximately from graphs.

Table 3: Used data for our analysis

Input data	English	Greek
Demography		✓ [1]
Polymod_UK	✓	
ILI		✓ [2]
Confirmed Samples		✓ [3]
Coverage	✓	
Vaccine Calendar	✓	

[1] https://www.statistics.gr/el/statistics/-/publication/SAM03/-

[2] https://eody.gov.gr/wp-content/uploads/2019/01/ekthesi_SENTINEL_2013.pdf

[3] https://eody.gov.gr/epidimiologika-statistika-dedomena/evdomadiaies-ektheseis/evdomadiaies-ektheseis-epochiki-gripi/

5.2 Results

By implementing the SEEIIR model for the given parameters we mentioned earlier, we take the number of the new cases after each period during the year. We create the fraction of the infected for each age group (<65, 65+) and risk group (low and high risk) and the plot is shown in Figure 10 below.

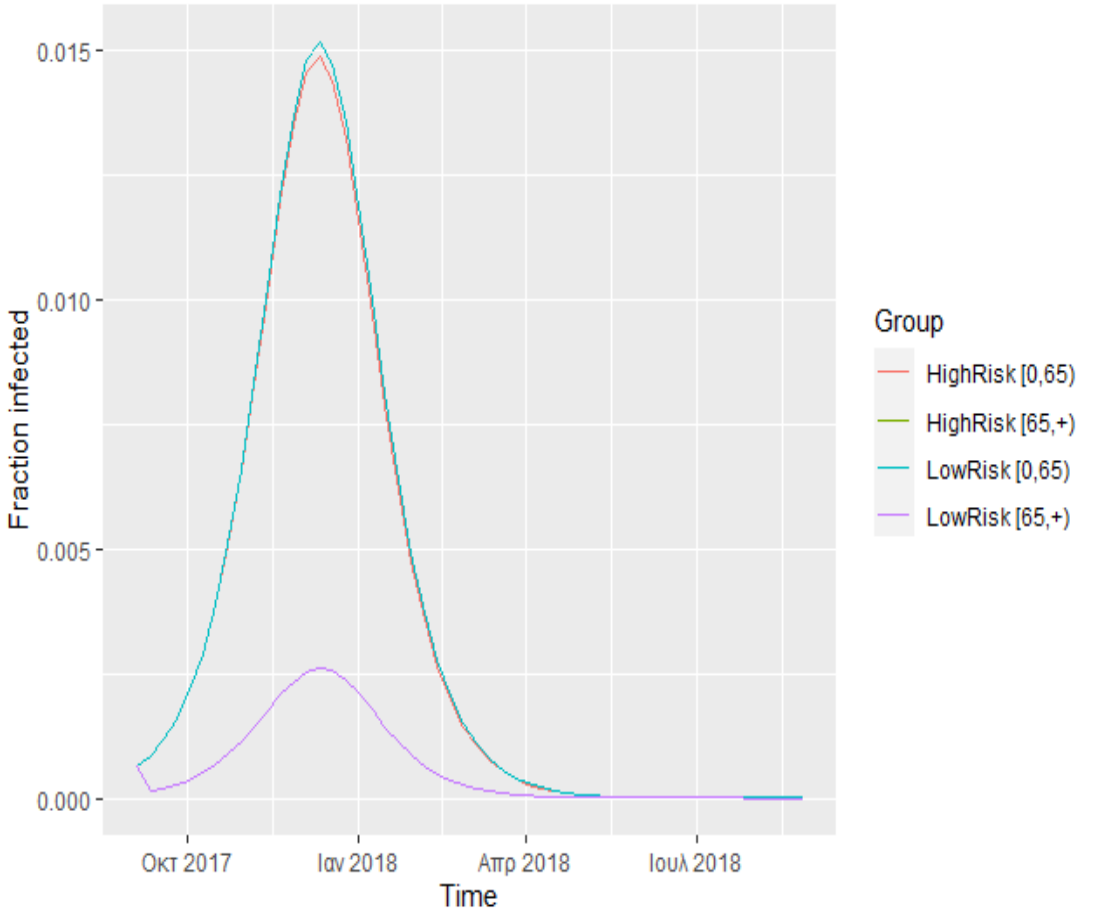

Figure 10: Fraction infected for each group with the ages 65+ covering each other because they have the same results.

Note the differences in the y axis scale. The low risk age group below 65 (Age1Risk1) has the largest population and also the largest incidence level.

The functions needed to perform the inference of the parameters have been defined. Using MCMC we get a posterior sample for the parameter values and we see the results in figure 11, which are not realistic because there are for a short MCMC run. There are five main parameters in our model:

- Ascertainment probability for 5 age groups (0-5, 5-15, 15-45, 45-65 and 65+) (ε_i)
- Outside infection (ψ)
- Transmissibility (q)
- Susceptibility for 5 age groups (σ_i)
- Initial number of infections (log transformed) (I)

The ascertainment probability and susceptibility are age group specific, so we need two new parameters to define them. The exact values are not that relevant, but the closer they are to the correct values the faster the interference will converge. We also have two risk groups (low and high). This means that our epidemiological data is mapped to 5 different groups (the age groups). We also reduce the complexity of the model by assuming that the first two age groups have the same ascertainment rate and susceptibility.

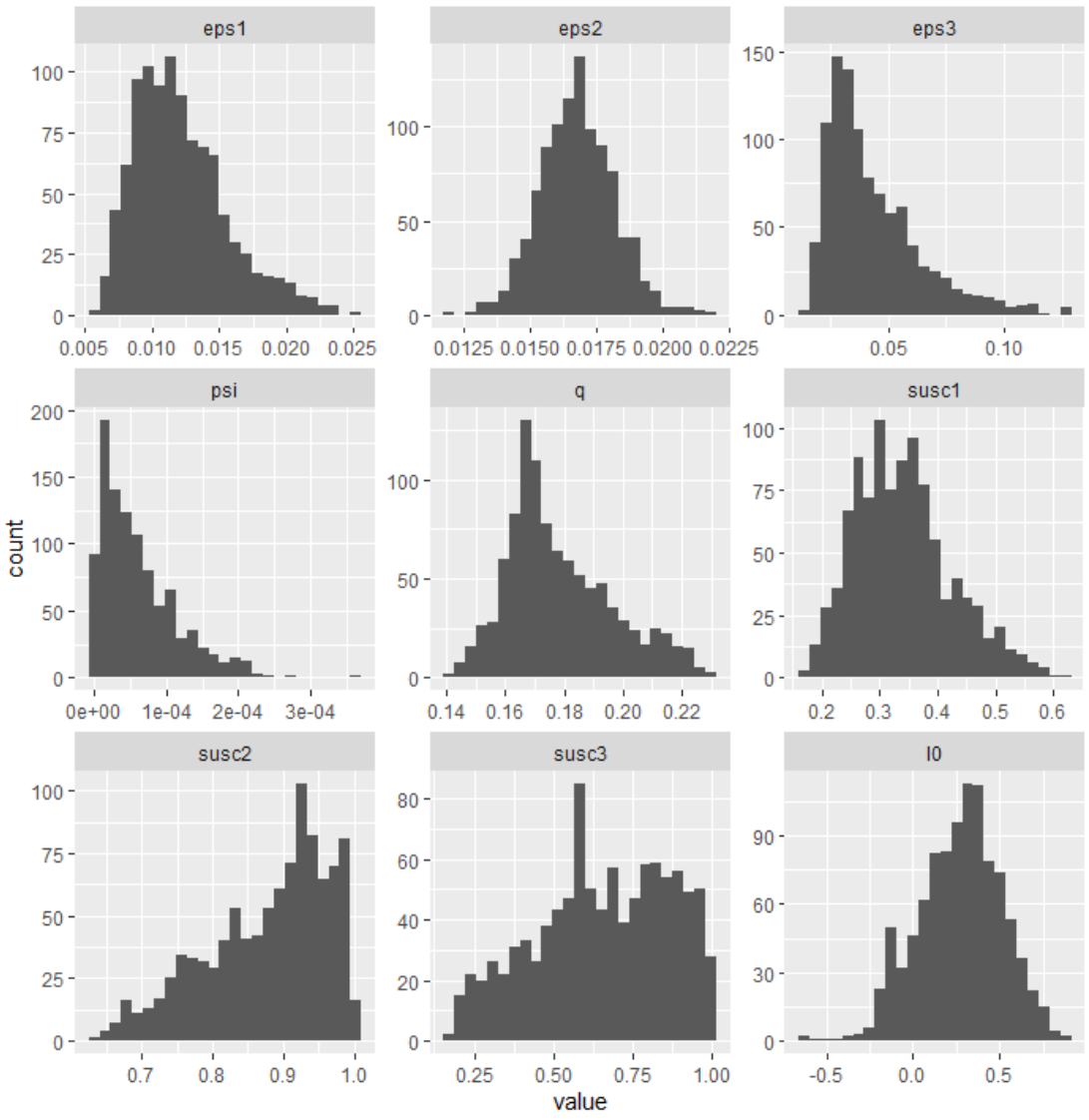

Figure 11: Plotting the resulting posterior parameter values

After that we create the credibility intervals of our models and we plot it, given a set of parameters (ε_i, ψ, q, σ_i, I), for each time point. Each plot is the result for

one of the 7 age groups and shows the estimated number of cases in each week (Figure 12). We sum the low and high risk group for simplicity and so each age group classified as high risk (RG1).

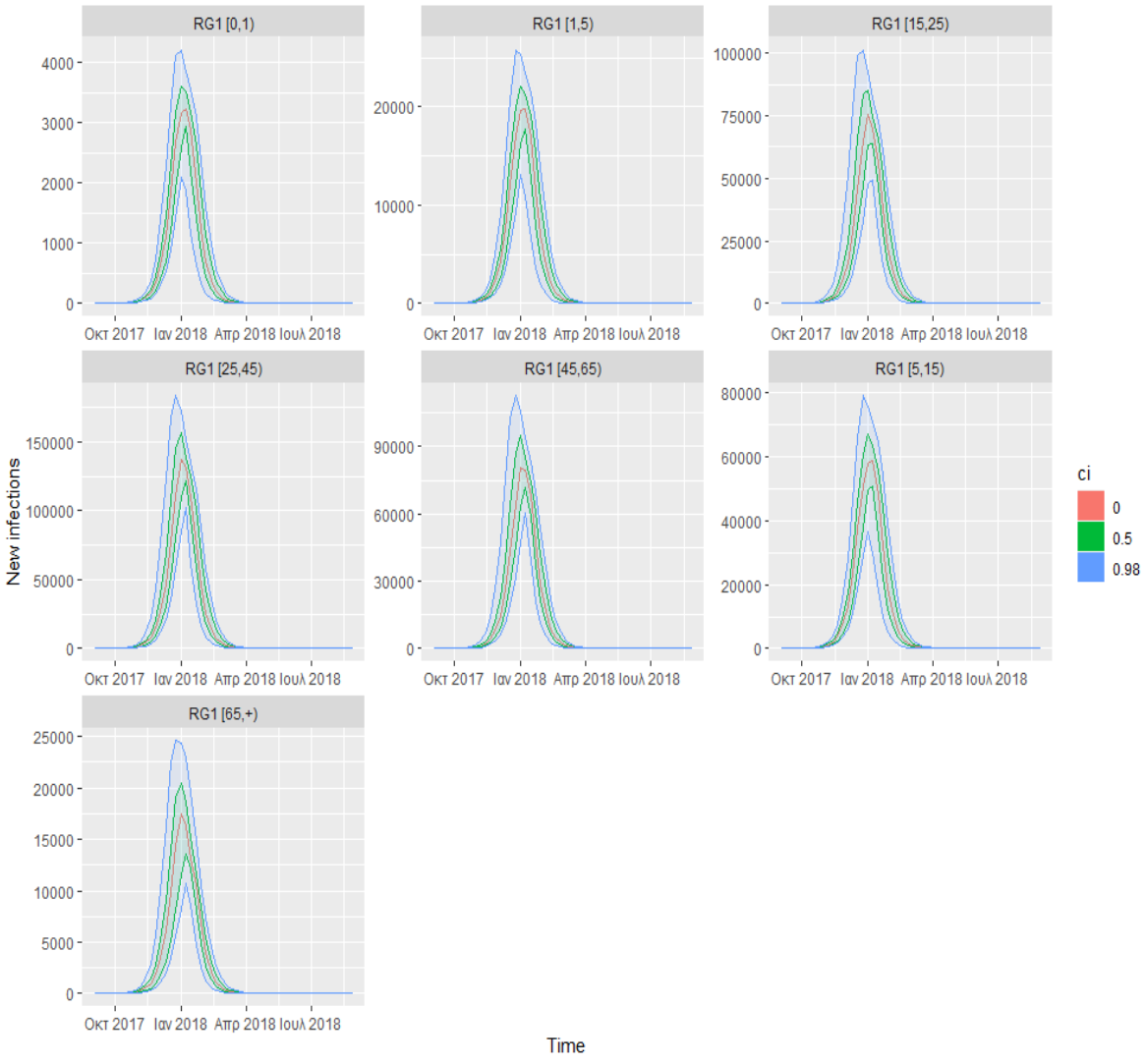

Figure 12: Credibility intervals for each one of the seven age groups with the estimated number of cases in each week.

We can see that in every age group of the high risk population, the peak of the estimated individuals is around February and March. The most cases estimated to be in age group (25-45) and the less in infants (0-1).

Finally we quote an example (figure 13) of how the number of cases with influenza changes depending on the different vaccination scenarios.

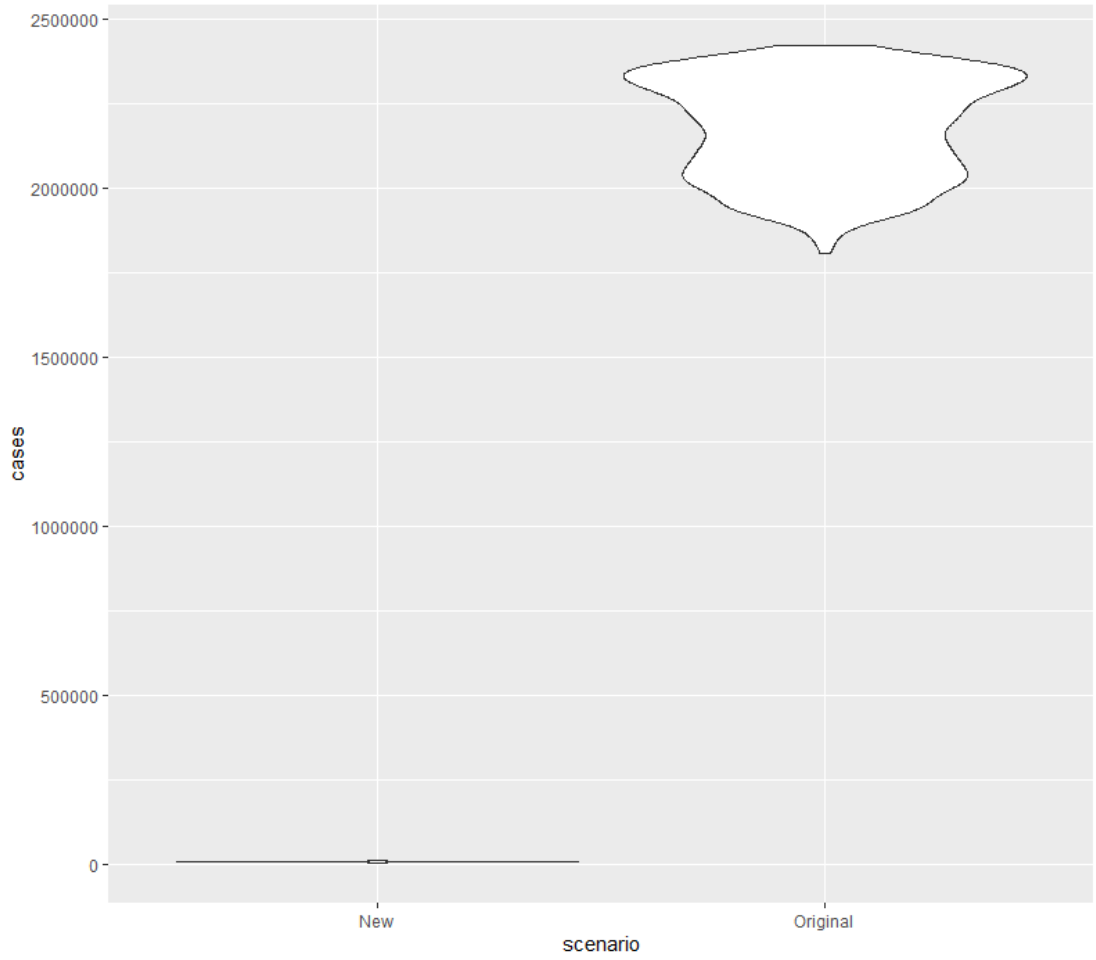

Figure 11: Different vaccination scenarios. Original is the current scenario and new is the change to the uptake rate to 80% in all age and risk groups.

For the original scenario we assume that the high risk young children and all 65 years and older get a vaccine. The rate of the vaccine uptake is different for each month in the first 3 months for the elderly, and same rate for the first 3 months for the high risk young children. The new scenario is with new vaccination coverage up to 80% for all age groups. There is a huge different between the two scenarios which might have been different if we could replace more input data from the beginning.

6. Discussion

In the current thesis we used a SEEIIR epidemic model to analyze the spread of the influenza virus and to estimate the number of cases through the year in Greece. The data we used is from the period of the first 51 weeks of 2017-2018. By taking the resulting values of the posterior of the parameters we calculate the incredibility intervals for each time point (0, 50%, 98%) to see the new infections over time for the high risk of all of the 7 age groups. We see a big reduction in the number of flu cases for different vaccination scenarios. First we set a coverage rate, different for non-elderly and elderly and in the second phase these rates increase to 80% for all age and risk groups. It is estimated that we have more cases in ages between 25 and 45 and less in infants.

The interaction between individuals is not following a random pattern but depend on the physical presence of the individual and his contacts that vary according to his/her location. We can compare out data to the data of another European country that is modeled by similar calculation by Kiesha et al. To the study by Kiesha et al, they try to predict the contact matrices between 152 countries. We observe that percentage of reduction of infection is similar to that of Germany. This is probably due to school closure and the resulting social distances of the individuals in the corresponding age groups. ([5-15],[15-45]). The contact patterns are varied according to age (different age groups) and locations (schools, working places etc). The contacts are associated with age and these variations led to the differences in the contacts of the different age groups in our model. These contacts play a significant role in the modeling of the virus transmissions because they use contact rates to predict the spread of the contact transmissible diseases. In our estimation the contact data we used is from the United Kingdom so our results may not be very representatively.

ΠΕΡΙΛΗΨΗ

Το εμβόλιο γρίπης μπορεί να χρησιμοποιηθεί για να μεγιστοποιήσει τα οφέλη για την υγεία μέσω στοχευμένων πολιτικών εμβολιασμού. Γενικά, οι εμβολιασμοί στοχεύουν συγκεκριμένες ομάδες που είναι πιο ευαίσθητες στον ιό της γρίπης. Αυτή η διαδικασία συνδυάζεται με την ανάγκη για οικονομική αποδοτικότητα και αυτό μπορεί να επιτευχθεί με την προσομοίωση επιδημίας. Τα επιδημικά μοντέλα έχουν δύο βασικούς σκοπούς. Το πρώτο αφορά την αναπαράσταση του μηχανισμού διάδοσης και την αξιολόγηση των σχετικών παραμέτρων. Ο δεύτερος στόχος είναι η χρήση τέτοιων εκτιμήσεων για την πρόβλεψη της εξάπλωσης και του ελέγχου των ασθενειών. Σε αυτό το έργο προσπαθούμε να ανοικοδομήσουμε την ελληνική επιδημία γρίπης μιας συγκεκριμένης περιόδου. Χρησιμοποιώντας μια σύγχρονη προσέγγιση τεκμηρίωσης-σύνθεσης, χρησιμοποιούμε βιολογικά, κλινικά, επιδημιολογικά και συμπεριφορικά δεδομένα για να αναπτύξουμε ένα στρωματοποιημένο μοντέλο μετάδοσης ηλικίας και κινδύνου που αναπαράγει τη συμπεριφορά της γρίπης σε συγκεκριμένα στελέχη για την περίοδο 2017-2018 στην Ελλάδα, έχοντας υπόψη τον εμβολιασμό κατά τη διάρκεια αυτής της περιόδου . Εκτιμούμε τον αριθμό των μολύνσεων όπως προέκυψαν από το πρόγραμμα ελέγχου ιστορικού, συγκριτικά με τον μη-εμβολιασμό και τη μείωση, εφαρμόζοντας διαφορετικές πολιτικές κατά την διάρκεια της περιόδου. Για το σκοπό αυτό, χρησιμοποιούμε το πακέτο fluEvidenseSynthesis και αντικαθιστούμε αρκετές πηγές δεδομένων με δεδομένα της ελληνικής γρίπης από την περίοδο 2017-2018 για να υπολογίσουμε τον αριθμό των ατόμων που έχουν μολυνθεί με τη γρίπη εκείνο το έτος. Τέτοιες εκτιμήσεις μπορούν να χρησιμοποιηθούν για να καθοδηγήσουν τις προσπάθειες ελέγχου για τη μείωση του βάρους της γρίπης ενόψει της επόμενης σεζόν.

Abstract

Influenza vaccine can be used to maximize health benefits through targeted vaccination policies. In general, vaccinations target specific target groups that are more susceptible to influenza virus. This process is combined with the need for financial efficiency and this can be achieved by epidemic simulation forward. Epidemic models have two main purposes. The first concerns the representation of the propagation mechanism and the assessment of the relevant parameters. The second goal is to use such estimates to predict the spread and control of diseases. In this work we try to reconstruct the Greek flu epidemic of a specific period. Using a modern documentation-synthesis approach, we use biological, clinical, epidemiological and behavioral data to develop a stratified age and risk transmission model that reproduces influenza behavior in specific strains for the 2017-2018 season in Greece, having consider getting vaccinated during this period. We estimate the number of infections as resulted from the historical check program compared to non-vaccination and the reduction, implementing different policies during the period. For this purpose, we use the fluEvidenseSynthesis package and replace several data sources with Greek flu data from the period 2017-2018 to calculate the number of people infected with the flu in that year. Such estimates can be used to guide control efforts to reduce the weight of influenza in view of the coming season.

References.

1. Baguelin M, Hoek AJV, Jit M, Flasche S, White PJ, et al. (2010) Vaccination against pandemic influenza A/H1N1v in England: a real-time economic evaluation. Vaccine 28: 2370–2384.

2. Baguelin M, Flasche S, Camacho A, Demiris N, Miller E, Edmunds WJ (2013) Assessing Optimal Target Populations for Influenza Vaccination Programmes: An Evidence Synthesis and Modelling Study. PLoS Med 10(10): e1001527.

3. Carrat F, Vergu E, Ferguson NM, Lemaitre M, Cauchemez S, et al. (2008) Time lines of infection and disease in human influenza: a review of volunteer challenge studies. Am J Epidemiol 167: 775–785.

4. Chowell G, Miller MA, Viboud C: Seasonal influenza in the United States, France and Australia: transmission and prospects for control. Epidemiol Infect 2007, 136:852-864

5. Cowles, M.K.; Carlin, B.P. (1996). "Markov chain Monte Carlo convergence diagnostics: a comparative review". Journal of the American Statistical Association. 91 (434): 883–904

6. Dushoff J, Plotkin JB, Levin SA, Earn DJD: Dynamical resonance can account for seasonal influenza epidemics. Proc Natl Acad Sci USA 2004, 101:16915-16916

7. Ferguson NM, Cummings DA, Fraser C, Cajka JC, Cooley PC, Burke DS: Strategies for mitigating an influenza pandemic. Nature 2006, 442:448-452.

8. Flasche S, Hens N, Boëlle P-Y, Mossong J, van Ballegooijen WM, et al. (2011) Different transmission patterns in the early stages of the influenza A(H1N1)v pandemic: a comparative analysis of 12 European countries. Epidemics 3: 125–133.

9. Fork P. MCMC Algorithms . CRC Press, 2016

10. Johnson BF, Wilson LE, Ellis J, Elliot AJ, Barclay WS, et al. (2009) Fatal cases of influenza a in childhood. PLoS ONE 4: e7671

11. Hill, S. D.; Spall, J. C. (2019). "Stationarity and Convergence of the Metropolis-Hastings Algorithm: Insights into Theoretical Aspects". *IEEE Control Systems Magazine*. **39** (1): 56–67.

12. Kermack WO, McKendrick AG: A contribution to the mathematical theory of epidemics. Proc Roy Soc Lond 1927, 115:700-721

13. Kiesha P., Cook A., Jit M., Projecting social contact matrices in 152 countries using contact surveys and demographic data. 2017. Comput. Biol. 13(9): e1005697.

14. Leeuwen E., Klepac P., Thorrington D., Pobody R., Baguelin M. fluEvidenceSynthesis: an R package for evidence synthesis based analysisi of epidemiological outbreaks. 2017. PLOS Computational Biology 13(11): 1-12.

15. Miller E, Hoschler K, Hardelid P, Stanford E, Andrews N, et al. (2010) Incidence of 2009 pandemic influenza A H1N1 infection in England: a cross-sectional serological study. Lancet 375: 1100–1108.

16. Mills CE, Robins JM, Lipsitch M: Transmissibility of 1918 pandemic influenza. Nature 2004, 432:904-906.

17. Pogka, V., Kossivakis, A., Kalliaropoulos, A., Moutousi, A., Sgouras, D., Panagiotopoulos, T., Chrousos, G.P., Theodoridou, M., Syriopoulou, V.P. and Mentis, A.F. (2011), Respiratory viruses involved in influenza-like illness in a Greek pediatric population during the winter period of the years 2005–2008. J. Med. Virol., 83: 1841-1848

18. Sherlock C, Fearnhead P, Roberts GO. The Random Walk Metropolis: Linking Theory and Practice. Through a Case Study. Statistical Science. 2010; 25: 172–190

19. Stone L, Olinky R, Huppert A: Seasonal dynamics of recurrent epidemics. Nature 2007, 446:533-536.

20. Urban Marguerite. Influenza, Viral infections: Merck Manual Home edition. 2009. Available at

21. **https://doi.org/10.1002/jmv.22173** (accessed 15/10/2020)

22. Vynnycky E, Edmunds WJ: Analyses of the 1957 (Asian) influenza pandemic in the United Kingdom and the impact of school closures. Epidemiol Infect 2008, 136:166-179.

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

Appendix

Figure 14:

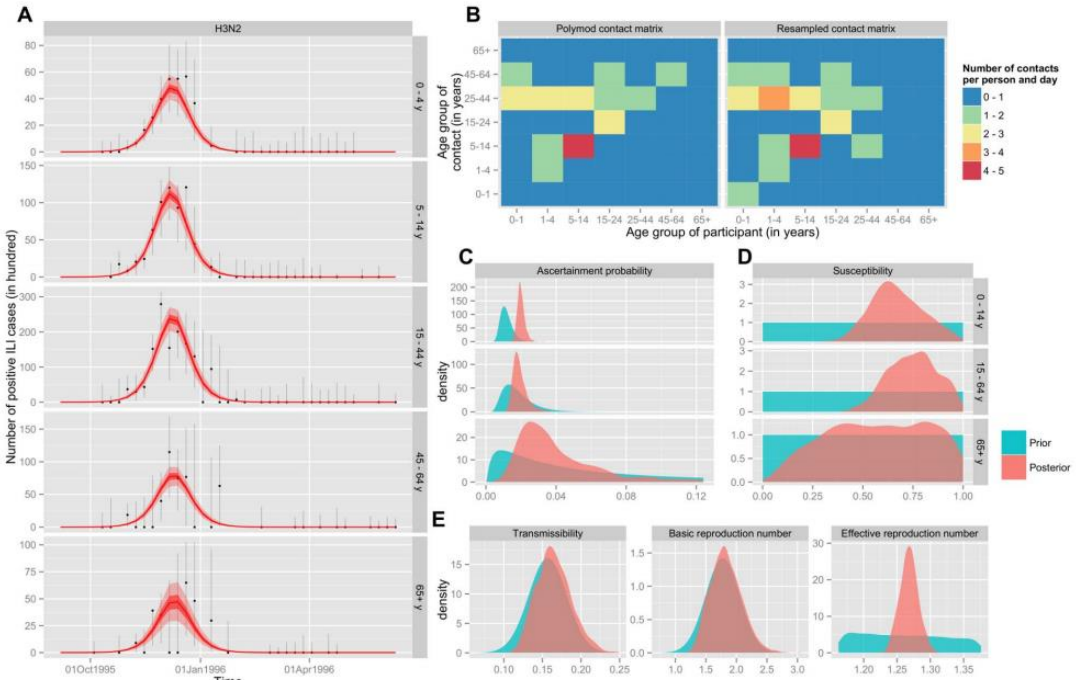

Figure 14: The results of the computational model when applied to the influenza outbreak (H3N2 stem) during the 1995/1996 season. A: We can see the five age groups and the number of positive Influenza – Like Illness cases in each group. The red line represents the mean. B: On the left we can see the contact matrix of the POLYMOD and the resembled likelihood matrix. C: The probability of someone being recorded as ILI case in all age groups. D: The sensitivity of each group at the beginning of the influenza season. E: Transmission coefficient (q - left), basic (R_0 - middle) and effective (Re(t=0) - right) reproduction numbers.

Table 4: Sample data table

Week	0-5	5-15	15-45	45-65	65+	Positive for influenza	Hospital samples	A positive	A-H1N1 positive	A-H3N3 positive	B positive	Deaths	Deaths of vaccinated people	Female ICU	Male ICU	High risk ICU
42-17	1	0	1	0	0	1	32	0	0	0	1	0	0	0	0	0
43-17	0	0	1	4	3	1	39	1	0	0	0	0	0	0	0	0
6-18	0	1	1	2	4	8	57	2	1	5	0	0	0	0	0	0
7-18	1	2	5	0	6	4	80	2	2	0	0	0	0	0	0	0
8-18	0	0	1	2	3	48	174	10	7	3	38	4	0	2	2	4

Π.Μ.Σ. ΒΙΟΣΤΑΤΙΣΤΙΚΗΣ

Συμμετέχοντα Τμήματα:
Ιατρική Σχολή Πανεπιστημίου Αθηνών
Μαθηματικών Πανεπιστημίου Αθηνών

ΒΕΒΑΙΩΣΗ

Η μεταπτυχιακή φοιτήτρια *ΚΟΥΤΣΟΣΠΥΡΟΥ ΧΡΙΣΤΙΝΑ* ολοκλήρωσε τη διπλωματική εργασία του με τίτλο *EPIDEMIC MODELS AND THEIR APPLICATION IN THE ANALYSIS OF INFLUENZA OUTBREAKS* στα πλαίσια των σπουδών του για το Διατμηματικό Μεταπτυχιακό Δίπλωμα Ειδίκευσης στη **"Βιοστατιστική"** των Τμημάτων Μαθηματικών και Ιατρικής Σχολής του Πανεπιστημίου Αθηνών.

Την εργασία αυτή παρουσίασε σε δημόσια διάλεξη στις 25/05/2020 *στο Τμήμα* ΜΑΘΗΜΑΤΙΚΩΝ *του Πανεπιστημίου Αθηνών.*

Μετά από προφορική εξέταση που ακολούθησε τη διάλεξη, η τριμελής εξεταστική επιτροπή, ενέκρινε ομόφωνα τη διπλωματική αυτή εργασία.

Η Εξεταστική Επιτροπή

Ονοματεπώνυμο	**Βαθμίδα**	**Υπογραφή**
1. Νικόλαος… …Δεμίρης…	…Επ. Καθηγητής…	
2. Γιώτα Τουλούμη	**Καθηγήτρια**	
3. Βασιλική-Αναστασία Σύψα	**Αναπλ. Καθηγήτρια**	