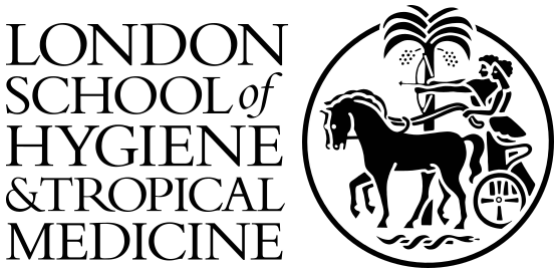

Temporal Analysis of Infectious Diseases: Influenza

Kaja Abbas, Armin R. Mikler, and Robert Gatti

Department of Computer Science and Engineering University of North Texas

Keywords: Bayesian Learning, Computational Epidemiology, Influenza

ABSTRACT

A Bayesian network is developed to embed the probabilistic reasoning dependencies of the demographics on the incidence of infectious diseases. Influenza epidemics occur every year in both hemispheres during the winter. The Bayesian learning paradigm is used to create synthetic data sets that simulate an outbreak of influenza for a geographic area. The Bayesian prior and posterior probabilities can be altered to represent an outbreak for various demographics in different geographic regions. Epidemic curves are generated, via time series analysis of the data sets, for the temporal flow of influenza on different variants of the demographics. The analysis of the demographic-based epidemic curves facilitates in the identification of the risk levels among the different demographic sections. Spread vaccination lowers the impact of the epidemic, depending on the efficacy of the vaccine. Our model is equipped to analyze the effects of spread vaccination and design vaccination strategies, that optimize the use of public health resources, by identifying high-risk demographic groups. Our results show that application of the vaccine in the order of risk levels will further lower the epidemic impact as compared to uniform spread vaccination.

1. INTRODUCTION

The incidence of infectious diseases in a population has various attributing factors, such as genetics, health history, travel history, demographics, and others, including unknown and hidden causes. It is generally assumed that an individual is protected from an infectious disease when he/she has the requisite disease immunity. Immunity can be present due to either genetic makeup, vaccination or past exposure. Vaccination is the primary policy of the public health departments in combating the spread of infectious diseases.

An epidemic of an infectious disease can be studied and analyzed by use of epidemic curves, which graph the incidence of the disease over time. In this paper, demographic based epidemic curves for a geographic region are generated. The temporal analysis of these curves shall aid in the identification of the critical groups of the population, that require primary attention in curtailing the spread of the dis- ease. Limited resources of public health departments must be optimally deployed which necessitates the identification of demographic specific groups for the surveillance, prevention, and control of infectious diseases.

The remainder of this section gives a brief sketch on influenza and the susceptibles infectives removals (SIR) model. Section 2 illustrates the utilization of the Bayesian network for influenza, to generate data sets simulating an epidemic. Section 3 discusses the temporal analysis of the demographics on the incidence of influenza, and vaccination strategies. Section 4 outlines the related work in mathematical studies of influenza and other infectious diseases, and section 5 concludes the paper.

1.1 Influenza

Influenza is an infectious disease, which is more commonly known as the flu. Symptoms include fever, respiratory symptoms, nasal discharges, cough, headache and sore throat. Incubation period is the time period between the start of infection and the onset of symptoms. Influenza has an incubation period of 1-4 days [14]. The latent period is the time between being infected and becoming infectious, that is, capable of passing on the infection to others. Influenza's latent period is a day less than the incubation period. The latent period is followed by 4-5 days of infectious period [14] . Once the infectious period ends, the recovery period starts, during which the infected individual is no more transmitting the infection to others. Fig. 1 illustrates the infection time-line for influenza.

The 1918 pandemic, caused by the H1N1 Spanish strain, is a historical event, resulting in 20 million deaths in its first year alone [14]. World Health Organization (WHO) [17] and Centers for Disease Control and Prevention (CDC) [5] are involved in influenza surveillance around the world and in the design of effective vaccination programs.

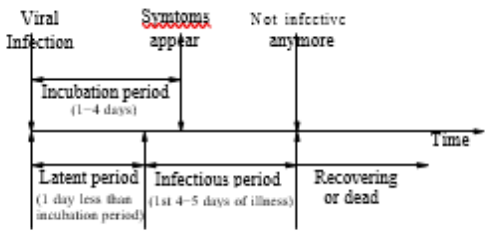

Figure 1: Infection Time-line for Influenza

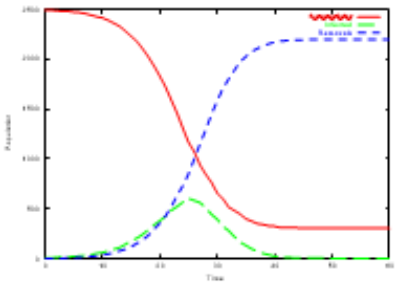

Figure 2: SIR epidemic curve

1.2 SIR Model

Mathematical modeling of infectious diseases are based on the principles of susceptibles, infectives, and removals, namely the SIR model. Susceptibles are those individuals in a population who can be infected by the disease under study. Infectives are those individuals who have been infected and are infectious. Removals include all individuals that are incapable of transmitting the infection, and are either recovering, fully recovered, or expired from the disease. In complex models, the removals who recover may revert back to susceptibles. In case of influenza, a recovered individual cannot be infected by the same influenza strain due to acquired immunity during the infection, while he/she be- comes susceptible to other influenza strains. Fig. 2 shows the transient curves for the susceptibles, infectives and removals during the course of a disease epidemic.

Kermack-McKendrick threshold theorem [3] is the basis for the SIR model. A continuous influx of susceptibles is a requisite for sustained infection in a population. The model is based on the presumption of a closed population, assuming that the epidemic spreads rapidly enough that the changes brought in by births, deaths, migration and demo- graphic changes are negligible [2].

The index case is the first infected individual and is the source of the infection. During the infectious period, the infection is passed on to some susceptibles, who interact with the index case close enough to contract the infection. This triggers the cycle of infections spreading through the population until a peak, as shown in Fig. 2. The infected individuals, on either recovery or expiry,

move over to the removals category. The rising infection on reaching the peak starts to recede due to the decrease in the number of susceptibles, and dies down eventually.

2. INFLUENZA OUTBREAK SIMULATION

The probabilistic reasoning capabilities under uncertainty are integral to Bayesian learning. The principle of Bayes' theorem is to update the beliefs of a hypothesis, given evidence. The prior probability distributions of the hypothesis transforms to posterior probability distributions upon witnessing the evidence. There are a number of good sources of study for Bayesian learning ([7],[8],[10]).

We develop a Bayesian network illustrating the probabilistic dependencies of the demographics on the incidence of influenza in a geographic region. The Bayesian network is used to generate synthetic data sets, that reverse engineer an influenza outbreak embedded in a temporal domain.

2.1 Bayesian Network

For the study, the demographics and the symptoms considered are not comprehensive to keep the Bayesian model simple. The demographic parameters are ethnicity, gender, age, income, and location; and the symptoms are cough and fever. Fig. 3 shows the Bayesian network illustrating the dependencies between demographics, symptoms, and influenza incidence. The analysis includes the absence and presence of vaccination.

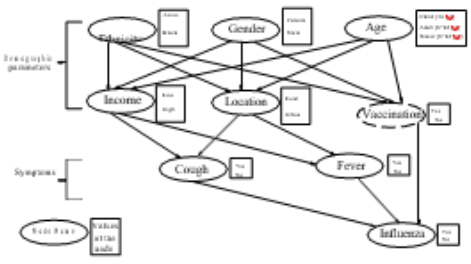

Figure 3: Bayesian Network

Table 1: Parameters' & Parameter Values' Symbols

Parameter	*Symbol*	*Parameter Value*	*Symbol*
Ethnicity	E	Asian	As
Gender	G	Black	Bl
Age	A	Hispanic	Hi
Income	I	White	Wh
Location	L	Female	Fe
Vaccination	V	Male	Ma
Cough	C	Child	Ch
Fever	F	Adult	Ad
Influenza	IN	Senior	Se
		Low	Lo
		High	Hi
		Rural	Ru
		Urban	Ur
		Yes	Ye
		No	No

2.2 Artificial Data Sets

The Bayesian network (Fig. 3), coupled with the influenza infection time-line (Fig. 1) are used to generate synthetic data sets for a population size of 100,000. The lack of avail- able real data sets for epidemic outbreaks with finer demo- graphic details necessitates the development of approaches to synthetically replicate epidemic data sets. Table 1 lists the symbols used for the random variables and their corresponding values. Table 2 defines the prior and posterior probability distributions for the random variables of the Bayesian network. For example, P(I/C,F) refers to the conditional probability of influenza, given the evidence of cough and fever. A contact rate of 10 is used, which

implies that each infected individual contacts an average of 10 other people during the infectious period of 4-5 days. A contact is considered successful if the disease is transmitted. The index case can be visualized as the root of a tree, with the infection spreading along the branches of the tree until the leaves being unable to make successful contacts with susceptibles.

Table 2: Probability Distributions

E	P(E)	G	P(G)	A	P(A)
As	0.15	Fe	0.52	Ch	0.10
Bl	0.20	Ma	0.48	Ad	0.70
Hi	0.30			Se	0.20
Wh	0.35				

E	G	A	P(I=Lo/E,G,A)	P(L=Ru/E,G,A)
As	Fe	Ch	0.95	0.20
As	Fe	Ad	0.60	0.18
As	Fe	Se	0.40	0.25
As	Ma	Ch	0.96	0.22
As	Ma	Ad	0.65	0.15
As	Ma	Se	0.50	0.27
Bl	Fe	Ch	0.96	0.30
Bl	Fe	Ad	0.55	0.35
Bl	Fe	Se	0.50	0.40
Bl	Ma	Ch	0.95	0.33
Bl	Ma	Ad	0.50	0.36
Bl	Ma	Se	0.40	0.42
Hi	Fe	Ch	0.98	0.10
Hi	Fe	Ad	0.65	0.08
Hi	Fe	Se	0.70	0.18
Hi	Ma	Ch	0.97	0.12
Hi	Ma	Ad	0.70	0.09
Hi	Ma	Se	0.75	0.20
Wh	Fe	Ch	0.97	0.60
Wh	Fe	Ad	0.40	0.55
Wh	Fe	Se	0.30	0.75
Wh	Ma	Ch	0.96	0.55
Wh	Ma	Ad	0.35	0.58
Wh	Ma	Se	0.25	0.70

I	L	P(C/I,L)	P(F/I,L)	C	F	P(IN/C,F)
Lo	Ru	0.45	0.10	Ye	Ye	0.60
Lo	Ur	0.50	0.20	Ye	No	0.25
Hi	Ru	0.20	0.05	No	Ye	0.40
Hi	Ur	0.25	0.15	No	No	0.05

3. TEMPORAL ANALYSIS

An epidemic curve visualizes the incidence (rate) that traces the number of newly affected individuals over time. Data mining is applied to synthetic data to extract pertinent information of the influenza epidemic, including demographic-based ethnic curves. The role of vaccination in curtailing the impact of the epidemic is investigated.

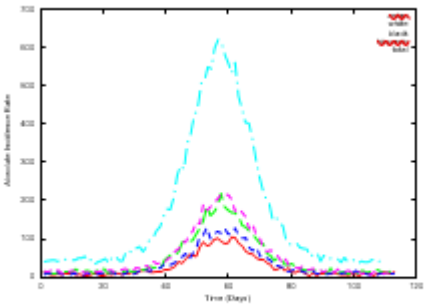

Figure 4: Ethnicity

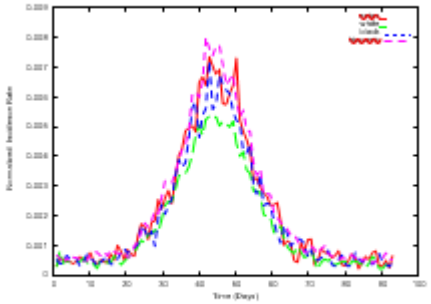

Figure 5: Ethnicity-Normalized

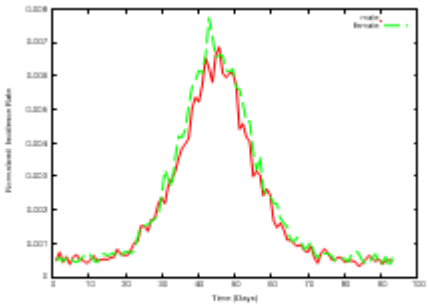

Figure 6: Gender-Normalized

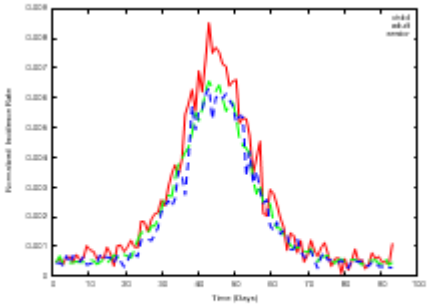

Figure 7: Age-Normalized

3.1 Epidemic Curves

Epidemic curves (Figures 4-7) are generated for the entire population, as well as for various demographic categories. Fig. 4 shows the dispersion of the epidemic among the different ethnic sub-groups. The level of impacts are ordered <hispanics, whites, blacks, asians>, thereby, placing larger number of hispanics at higher risk. Fig. 5 shows the normalized version of the epidemic spread among the ethnic sub-groups, wherein the newly affected cases are proportioned over the total population of the specific sub-group. This helps in visualizing the outbreak among the sub-groups on a balanced platform. The ordered list of epidemic im- pact is <hispanics, asians, blacks, whites>. The asians are categorized correctly at the second highest risk level, while the earlier ordering placed asians at the lowest risk. Hence, normalized curves help in perceiving the true risk groups, based on proportions, rather than epidemic curves based on absolute number of disease incidences.

Fig. 6 depicts the normalized epidemic spread among the genders, female and male. The females are observed at a higher risk level compared to males. Influenza spread does not have a gender bias. The observed result may be at- tributed to the complex social interaction patterns in the geographic region under study.

The age groups have been categorized as child (≤5 years), adult (6-64 years), and senior (≥65 years). Fig. 7 shows the normalized dispersion of the epidemic among the three age groups. Child subgroup are at the highest risk. Adults and seniors are observed at lower risk levels, closely interspersed with each other. CDC strongly recommends vaccination for seniors, since 90% of influenza related mortality occurs in the senior group. The observed lower level for the seniors in our experiments may be attributed to the specifics of the synthetic data.

3.2 Vaccination

Vaccination is the key preventive measure used by pub- lic health departments in toning down the annual influenza epidemics. A uniform random distribution of the vaccines among the population is deployed by use of spread vaccination. Limitations on public health resources prohibit the goal of herd immunity. Consequently, resources must be applied optimally in order to curtail the epidemic.

Vaccines have associated effective rates of success. An ideal vaccine shall have a 100% success in protecting a vaccinated individual from influenza. Nevertheless, influenza vaccines have an efficacy of 50%-80%. Fig. 8 compares the epidemic curves for 10% of the population being spread vaccinated with 100%, 80%, and 50% success rates respectively, with a non-vaccinated population. The epidemic impacts for vaccine scenarios are all toned down compared to no vaccination scenario, reflecting on the success of the vaccination programs. As expected, 100% vaccine efficacy yields the best result.

Fig. 9 illustrates the normalized bezier epidemic curves for ethnicity, gender and age groups. The analysis of the demographic graphs on a normalized scale aids in creation of a hierarchical list of demographic sub-groups ordered by their respective levels of risk. The ordered list in the decreasing level of risk shall be <child, hispanic, asian, female, male, black, adult, senior, white>.

In our experiments, we assume vaccines are available for 10% of the population. In the first case, these vaccines are equally deployed among the three high-risk groups <child, hispanic, asian>. In the second case, these vaccines are applied equally among the three low-risk groups <adult, senior, white>. Fig. 10 depicts the epidemic curves of both the cases, along with the curve for simple spread vaccination. We are using 80% vaccine efficacy. Vaccination, prioritized on the high-risk groups, yield the best results in curtailing the epidemic, while vaccination of low-risk groups increases the epidemic impact, as compared to naive spread vaccination. Hence, demographic based high-risk group vaccinations should be accorded higher priority to curtail the influenza epidemic. This necessitates a better understanding of disease progression in a given demographic domain.

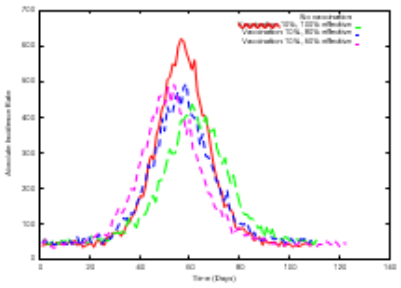

Figure 8: Varied Spread Vaccination Effective Rates

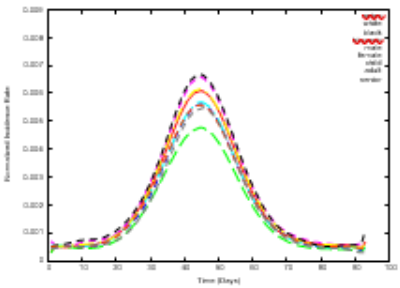

Figure 9: {Ethnicity, Gender, Age} Normalized

4. RELATED WORK

The spatial and temporal correlation of influenza epidemics in the United States, France, and Australia from 1972 to 1997 has been analyzed in [16]. The results indicate a high correlation between United States and France, but irregularity in the patterns between Australia and the other two countries. Demography is highlighted as the primary reason for the discrepancy.

Bayesian learning is used to infer the dependency of the demographics on the incidence of diseases in different geo- graphic regions [1]. A stochastic cellular automata paradigm is used to study the dynamics of infectious diseases' epidemics, and vaccination strategies in controlling them [15]. In susceptibles, infectives, and susceptibles (SIS) model, the infectives upon recovery return to the susceptibles category. The SIS model for spread of infectious diseases has been analyzed for non-uniform population densities, thereby accounting for mobility of individuals [4]. The susceptibles, exposed, infectives, and removals (SEIR) model includes the class of exposed individuals who have been exposed to the disease, but are in the latent stage, and hence are not ca- pable of infecting the susceptibles. A 4-dimensional SEIR epidemic model is considered and the criteria for stable equilibria has been established [6].

Data obtained from the WHO global influenza surveil- lance network are mined to extract qualitative and quantitative information and used in the vaccine strain selection process [12]. The effects of prior infection by influenza has been studied to determine the performance and efficacy of vaccination [13].

Hidden Markov models with an exponential-gaussian mix- ture have been used for automated detection of influenza epidemics [11]. A Monte Carlo simulation using a Markov model is implemented to study the infection models, that oc- cur naturally, such as influenza whose viral pathogen spreads through a susceptible community, or induced deliberately, as in the case of bio-terror attacks [9].

5. CONCLUSION

Bayesian probabilistic reasoning is used in developing syn- thetic data sets, portraying the outbreak of an influenza epi- demic in a geographic region. Influenza epidemic due to a single strain prevails for around eight weeks, and all the epi- demic curves generated from the synthetic data also depict similar epidemic behavior. The epidemic curve for the total population is dispersed into demographic-based epidemic curves. The temporal flow of the epidemic among the different sections of the population is analyzed. The different population sections are ordered in a hierarchical list, based on risk levels. The higher end risk spectrum of the ordered list need primary health care early, in comparison to the lower end risk spectrum, to curtail the epidemic.

Spread vaccination is applied uniformly across a proportion of the population, and is witnessed to scale down the epidemic. The higher the efficacy of the vaccine in thwarting the disease, the lower is the impact of the epidemic. Vaccination resources are alternatively applied in two different ways. In the first scenario, the high risk groups are targeted, while in the second case, the lower risk groups are targeted. As per expectations, the epidemic is reduced extensively when high-risk groups have been vaccinated. On the other hand, when applied to low-risk groups, the de- sired effects of our vaccination effort are less, as compared to naive spread vaccination. This validates the reasoning in identification of high risk demographic sections of the population, and prioritizing them in the allocation of the limited public health resources.

6. REFERENCES

[1] K. Abbas, A. Mikler, A. Ramezani, and S. Menezes. Computational epidemiology: Bayesian disease surveillance. In Proc. of the International Conference on Bioinformatics and its Applications, FL, USA, December 2004.

[2] J. Aron. Mathematical modeling: The dynamics of infection, chapter 6. Infectious Disease Epidemiology, Aspen Publishers, Gaithersburg, MD, 2000.

[3] N. Bailey. The Mathematical Theory of Epidemics. Hafner Publishing Company, NY, USA, 1957.

[4] N. Boccara and K. Cheong. Critical behavior of a probabilistic automata network SIS model for the spread of an infectious disease in a population of moving individuals. Journal of Physics A: Mathematical and General, 26(5):3707–3717, 1993.

[5] The CDC influenza web page, 2004.

[6] M. El-Sheikh and S. El-Marouf. On stability and bifurcation of solutions of an SEIR epidemic model with vertical transmission. International Journal of Mathematics and Mathematical Sciences, 2004.

[7] K. Korb and A. Nicholson. Bayesian artificial intelligence. CRC Press, 2004.

[8] R. Neapolitan. Learning Bayesian networks. Pearson Prentice Hall Series, 2004.

[9] D. O'Leary. Models of infection: Person to person. Computing in Science & Engineering, 6(1), Jan-Feb 2004.

[10] J. Pearl. Probabilistic reasoning in intelligent systems. Morgan Kaufmann, San Mateo, CA, 1988.

[11] T. Rath, M. Carrerar, and P. Sebastiani. Automated detection of influenza epidemics using hidden Markov models. In Proc. of the 5th International Symposium on Intelligent Data Analysis, Berlin, Germany, August 2003.

[12] D. Smith. Applications of bioinformatics and computational biology to influenza surveillance and vaccine strain selection. Vaccine, 21:1758–1761, 2003.

[13] D. Smith, S. Forrest, D. Ackley, and A. Perelson. Modeling the effects of prior infection on vaccine efficacy. In Proc. of the IEEE International Conference on Systems, Man, and Cybernetics, FL, USA, 1997.

[14] M. Steinhoff. Epidemiology and Prevention of Influenza, chapter 16. Infectious Disease Epidemiology, Aspen Publishers, MD, 2000.

[15] S. Venkatachalam and A. Mikler. Towards computational epidemiology: Using stochastic cellular automata in modeling spread of diseases. In Proc. of the 4th Annual International Conference on Statistics, January 2005.

[16] C. Viboud, P. Bolle, K. Pakdaman, F. Carrat, A. Valleron, and A. Flahault. Influenza epidemics in the United States, France, and Australia, 1972-1997. Emerging Infectious Diseases, 10, January 2004.

[17] WHO influenza web page, 2004.