Estimation of the Basic Reproduction Number and Vaccination Coverage of Influenza in the United States (2017-18)

Roya Nikbakht (MSc)1, Mohammad Reza Baneshi (PhD)1, and Abbas Bahrampour (PhD)1*

1 Department of Biostatistics and Epidemiology, Modeling in Health Research Center, Faculty of Health, Institute for Futures Studies in Health, Kerman University of Medical Sciences, Kerman, Iran

ABSTRACT

Background: Determining the epidemic threshold parameter helps health providers calculate the coverage while guiding them in planning the process of vaccination strategy. Since the trend and mechanism of influenza is very similar in different countries, we planned a study with the mentioned goal by using data of US from 2017 to 2018.

Study design: A secondary study.

Methods: R0 and corresponding vaccination coverage are estimated using the national and state-level data of the US from the 40th in 2017 to the 5th week in 2018. Four methods maximum likelihood (ML), exponential growth (EG), time-dependent reproduction numbers (TD), and sequential Bayesian (SB) are used to calculate minimum vaccination coverage. The gamma distribution is considered as the distribution and the generation of time.

Results: The peak of epidemic in most states has occurred in the 15th week after the beginning of the epidemics. The generation time obey the Gamma distribution with mean and standard deviation of 3.6 and 1.6, respectively, was utilized for the generation time. The R0 (vaccination coverage) equaled 1.94 (48.4%), 1.80 (44.4%), 3.06 (67.3%), and 2.11 (52.6%) for EG, ML, SB, and TD methods at the national level, respectively.

Conclusion: The R0 estimations were in the range of 1.8-3.06, indicating that an epidemic has occurred in the US (R0>1). Thus, it is required to vaccinate at least 44.4% to 67.3% to prevent the next epidemics of influenza. The findings of this study assist futures studies to apply disease control by vaccination strategies in order to prevent a national disaster.

Introduction

Clinical research and studies of communicable (infectious) diseases follow different purposes such as determining the trend, epidemic threshold parameters, and vaccination coverage. The epidemic threshold parameter, R0, plays a key role in the diagnosis of suggested control strategies in order to apply interventions or vaccination preventative strategies. In biostatistics and epidemiology, epidemic threshold parameter (whose special form is known as the basic reproduction number or reproductive ratio) defined as the mean number of secondary cases infected by initial cases in a fully susceptible community1,2. The basic reproduction number is generally compared with unity to assess the spread of infectious diseases to the population. An R0 greater than unity (R0>1) means an epidemic has occurred and each infected individual generates more than one new case3. In addition, the epidemic likely fades out when the basic reproduction number becomes less than unity (R0<1) and the R0 equals1, leading to an epidemic1,2,3.

Several approaches have been suggested to estimate R0. These include maximum likelihood (ML), exponential growth rate (EG), estimation of time-dependent reproduction numbers (TD), attack rate, gamma-distributed generation time, the final size of epidemic, and Richard model4-14. The type of approach depends on the type of data which is being studied (the type of household or daily incidence data). In each method, the basic reproduction numbers are reviewed to assess the intensity of interventions and vaccination strategies that can estimate the vaccination coverage of an infectious disease. Therefore, vaccination strategies are introduced in order to reduce and prevent the risk of transmission of infectious diseases in the target community. In addition, R0 and vaccination coverage have a direct impact on each other, which means that any variation in R0 leads to a corresponding variation in vaccination coverage and vice versa. Therefore, a larger number of people should be vaccinated in a susceptible population where the estimation of basic reproduction number is a larger number.

The estimation of the basic reproduction number (R0) has been addressed in various infectious diseases, including influenza, HIV, SARS, smallpox, malaria, yellow fever, measles, and Ebola15-21. In particular, influenza is a leading cause of mortality, with a considerable number of annual deaths in the world22. Several influenza epidemics have occurred worldwide from 2009 to 2017 during which a substantial number of people died annually22. For example, the number of deaths caused by “Asian flu” and “Hong Kong flu” is estimated at 1 to 4 million23. On the other hand, the annual
deaths attributed to influenza are estimated at nearly 250,000 to 500,000 globally. Moreover, United States flu (pH1N1) killed about 12,469 persons in 2009.

An epidemic has recently occurred in the US. Of course, the epidemic of influenza has been recorded in the US every year over the past 20 years, showing a yearly seasonal threat of the influenza epidemic (Table 1). A few of these epidemics are listed below:

In Philadelphia and New York (Sep 14 to Oct 17, 1918), the estimated R0 (95% CI) and generation time were 2.14 (1.88, 2.39) and 2.5 d, respectively. The formula used for estimating R0 was $R_t = \sum_{i>0} IR_{t+i}W_i$. In the USA (1972-2002) reported the R0 (95% CI) of 1.30 (1.20, 1.40) with generation time of 5.5 d and formula $R_0 = \beta/(\gamma + \delta)$. Yang et al. estimated R0 of influenza using epidemiological surveys and mathematical modeling approaches in two periods of 2003-2004 and 2012-2013 are taken into account generation time of 4.2 and 4.8 respectively. The R0 (95% CI) in this study were estimated 2.04 (1.84, 2.21) and 1.97 (1.84, 2.21)\(^{35}\). The estimated R0 in the year 2009 (Mar 28- Mar 04) using likelihood-based method had a range 2.50 to 3.48 with corresponding 95% CI (1.80, 2.16) and (1.84, 2.13)\(^{36}\).

The R0s reported above have been obtained using different methods and generation times and are, therefore, impossible to be compared. Therefore, another aim of the present study was to calculate the R0 of influenza for a given set of data (USA data) considering the same distribution for generation time by different methods to be able to compare various approaches. The number of cases and generation time distribution is needed which is gamma distribution with mean and SD 3.6 and 1.6 respectively based on similar study.

In addition, determine the epidemic threshold, helps health providers calculate the coverage while guiding them in planning the process of vaccination strategy. Vaccination coverage is directly computed by R0, called indirect effect/herd protection\(^{31}\). Indirect vaccination coverage is not only economic but also prevent epidemic which its effect exceed the direct effect\(^{34}\).

Since the trend of influenza is very similar in different years, we conducted this study to determine the epidemic threshold parameters and, consequently, the vaccination coverage in the US from 2017 to 2018.

**Methods**

**Statistical Analysis**

In the first step, the classic SIR (susceptible, infected/infectious, and removal) compartmental model used to describe the process of influenza was implemented to determine the protocol of the person transmission indifferent states.

Next, four methods were developed to estimate the R0 based on cumulative case count data in R statistical software (version 3.4.2) with R0 packages, including ML, EG, sequential Bayesian method (SB), and TD. All the methods were used in different papers for influenza data, so for this type of data all the methods could be used to estimate the R0.

It is necessary to have a distribution for the generation time in each method defined below. Moreover, a brief overview of each method is presented.

### Generation time

The length of time between infection in a primary infection and a secondary infection is defined as generation time or serial interval\(^{37}\).

**EG:** The R0 formula in the exponential growth rate method is $\mu_t = R(\sum_{i=1}^t N_{t-i}w_i)$ where $R = \frac{1}{M(r)}$. Here, "r", "M", and $N_t$ demonstrate the growth rate of the infection population, the moment generating function of the generation time distribution, and cases over a consecutive time unit, respectively, and parameter "w" represents the generation time.

In order to estimate the growth rate parameter, Poisson regression method is applied\(^{1}\).

**ML:** In this method, the distribution of secondary cases infected by primary cases assume Poisson with mean R. Suppose $N_{t\rightarrow t+1}$ represent cases over time and parameter w shows generation time. Then, the log-likelihood function based on Poisson distribution is as follows:

$$LL(R) = \sum_{t=1}^{T} \exp(-\mu_t)R_t^{N_t}/N_t!$$

$$\mu_t = R \sum_{i=1}^{t} N_{t-i}w_i$$

The maximum of log-likelihood function gives the reproduction number (R0)\(^{37}\).

**SB:** Suppose $N(t+1)$ denotes incidence in time $(t+1)$ for the SIR model where we have an approximate Poisson distribution with mean $N(t)e^{\gamma(R-1)}$ (γ shows the average generation time). A non-informative prior for R is used in the Bayesian framework. The posterior distribution of R in the previous day is applied as the prior distribution for R in the new day. The posterior distribution for R is as follows:

$$P(R|N_{t-1}, ..., N_t) = \frac{P(N_{t+1}|R, N_0, ..., N_t)P(N_0, ..., N_t)}{P(N_0, ..., N_t)}$$

The exponential distribution applies for generation time in this method\(^{11}\).
**TD:** $R_0$ can be estimated by TD with formula $R_t = \frac{1}{N_t} \sum_{t_j=t} R_j$ where $R_j = \sum_i P_{ij}$ and $P_{ij} = \frac{N_i w(t_i-t_j)}{\sum_i N_i w(t_i-t_k)}$. In this formula, $P_{ij}$ demonstrates the probability of infection transmission from case $i$ (in time $t_i$) to case $j$ (in time $t_j$). $R_t$ is the mean of all $R_j$ computed by all networks of observed cases.

**Vaccination coverage**

To compute the percent of vaccination coverage, we need to estimate $R_0$. Therefore, the critical vaccination coverage, i.e. the proportion of people who receive vaccines, is obtained by the reproduction number using the following formula:

$$v = 1 - \frac{1}{R_0}$$

Vaccination coverage is also defined as the reduction in the probability of infection risk which is a value between 0 and 1.

Vaccine efficacy for reducing transmission can be achieved by $\vartheta = \frac{R_0-1}{R_0-1} = \frac{1-1/R_0}{1-ab}$ that $a$ and $b$ represents susceptibility effect and infectivity effect respectively.

**Data**

The four above-mentioned models were fitted to the US 2017-18 pH1N1 data, applying FluView weekly report achieved from the Centers for Disease Control and Prevention (CDC) website. The data of Surveillance Network (ILINet) were implemented, reporting influenza cases in all 47 states, the District of Columbia, New Dakota, New York City, Puerto Rico, and the U.S. Virgin Islands for each age group in the 40th week in 2017 to 5th week in 2018.

For all states, we estimated the $R_0$ based on four methods. Afterward, we calculated the variances of $R_0$s for each method to check for variability among different states, and then hierarchical cluster analysis was applied as an explorative technique to specify the number of clusters in the K-means clustering method to cluster the states. Cluster analysis was performed in Minitab 17 statistical software (Minitab Inc., State College, PA) and IBM SPSS Statistics 22 (Chicago, IL, USA).

**Results**

The incidence data are presented on a weekly basis and all dates of USA data are based on week/year from the 40th week in 2017 to the 5th week in 2018. The peak of epidemy in most states has occurred in the 15th week after start of the epidemics. The number of infected cases at the national level is provided in Figure 1.

The number of infected cases was plotted at the national level in five age groups. Maximum numbers of cases were in the age group of 5 to 24 yr and the peak incidence of influenza in this category occurred in the 19th week (Figure 2).

The gamma distribution with the mean of 3.6 d and standard deviation of 1.6 d has been used as the distribution of the generation time. The results will be presented in three parts: national level, state level, and comparison of methods.

**National level**

The national $R_0$ and vaccination coverage are summarized in Table 1 based on four methods (ML, EG, SB, and TD) while assuming a wholly susceptible population before the start of the epidemic.

At the national level, the highest value of $R_0$ was attributed to SB. Indeed, the estimated $R_0$ at the national level by SB was quite different compared to other methods (3.057 95% CI: 3.037, 3.08). Moreover, the estimated $R_0$s for EG, ML, and TD were 1.939 (1.937, 1.940), 1.80 (1.789, 1.802), and 2.111 (2.102, 2.120), respectively.
In addition, the estimates of vaccination coverage varied for the four methods, from 44.4% to 67.3%. The lowest and highest vaccination coverage values in this setting were associated with and SB methods, respectively (Table 1).

**States Level**

The computed \( R_0 \) (95% CI) and vaccination coverage for all states by EG, ML, SB, and TD are summarized in Table 1 in the appendix. In general, the estimates of \( R_0 \) at state level ranged from 1.55 to 2.79 for EG, 1.48 to 2.65 for ML, 1.62 to 2.46 for SB, and 1.67 to 2.73 for TD.

Examples of states with \( R_0 \) equal to 2.74 (95% CI: 2.14, 3.48) were estimated by ML, whereas the corresponding estimates for EG, SB, and TD were 2.73 (95% CI: 2.12, 3.35), 2.71 (95% CI: 2.37, 3.13), and 2.72 (95% CI: 2.39, 3.15), respectively. Thus, the estimates of \( R_0 \) among these four approaches were 0.1, 0.2, and 0.3, indicating variability. Cluster analysis based on the mentioned methods resulted in two clusters. The first cluster included EG and ML, while the second cluster comprised SB and TD. This analysis confirmed the above findings.

**Discussion**

The simple SIR compartmental model is used as the transition model, indicating that the estimation of \( R_0 \) using ML, EG, SB, and TD varied in different states due to the difference in the number of infectious cases during the outbreak. The variability of \( R_0 \) depends on many factors, including location, estimation method, generation time, and pandemic wave. The virus and network size are also influential factors in influenza transmission. The peak value of outbreak was the same in most states. Moreover, a sharp peak was observed in the incidence of H1N1 at the national level (Figure 1).

We have found variation in the estimation of \( R_0 \) using ML, EG, SB, and TD implemented by the "\( R_0 \) package". A quantitative comparison of findings revealed that the estimations of \( R_0 \) in SB are approximately close to those of TD. Besides, EG and ML yielded almost identical results, and cluster analysis based on the four methods confirmed this hypothesis.

The estimated epidemic threshold values based on three methods (EG, ML, and TD) in the first cluster were higher than those of the second cluster. Consequently, states in the first cluster have a higher risk of epidemic and require more vaccination coverage.

Generally, the \( R_0 \) associated with ILINet using four methods was greater than the one at the national level (winter of 2017) as well as state level (in both clusters), representing the epidemic of influenza. Therefore, it seems necessary to consider appropriate solutions to control, decrease, and prevent the epidemic or pandemic of influenza. An effective way to protect people from the attack rate of influenza is vaccination. Annual vaccination against seasonal influenza...
provides protection in high-risk groups (elderly people, ill persons, pregnant woman, and children) and can also reduce mortality rate, the incidence of disease, exacerbations, hospitalizations, and costs.

In determining vaccination coverage, \( R_0 \) plays a key role because the estimation of vaccination coverage is affected by \( R_0 (v=1-1/R_0) \). In other words, the percentage of a community vaccinated against influenza can be represented in terms of \( R_0 \). Vaccination coverage and directly impact each other, which means that, with an increase in \( R_0 \), vaccination coverage increases, and vice versa. The present study also provides an estimate of vaccination coverage for both national and state levels, which is one of the strengths of this study.

The \( R_0 \) of influenza for USA ranged between 1.3 and 3.1 from 1918 to 2013 using various methods. In our study, similar values were obtained for \( R_0 \). For example, at the national level, the \( R_0 \) was estimated using four methods (ML, EG, SB, and TD) and their values were in the range of 1.8 to 3.06, indicating that an epidemic occurred in USA (\( R_0 >1 \)).

Various studies have employed different methods and generation times to estimate the threshold of epidemics. Thus, it would be illogical to make such comparisons. In our study, the \( R_0 \) of influenza for USA data considering the same distribution for the generation time in different methods so that various approaches can be compared, which is another strength of this study.

A weakness of this study was that, although cluster analysis determined similar methods, there was no exact criterion for determining the best method.

**Conclusion**

The findings of our study can be used to improve policy-making, health care, and public health not only in the USA but also in other parts of the world. These results can be extended to other countries with similar epidemics. As the transmission mechanism is the same, the influenced parameters of the disease should be the same. In other words, the epidemic of influenza is similar in all countries so, in our country by at least 44.4% of vaccination can prevent the flu outbreak. Hence, awareness of the \( R_0 \) of influenza as a highly infectious disease is helpful for future studies to apply disease control through vaccination strategies in order to prevent a national disaster. Indirect vaccination coverage is not only economic but also prevent epidemic which its effect exceed the direct effect.

Influenza would become re-epidemic. Therefore, a more comprehensive study is needed to deal with this dangerous virus.

**Acknowledgments**

This research is part of Roya Nikbakht’s PhD dissertation.

**Conflict of interest statement**

The authors declare no conflict of interest.

**Funding**

There was no funding resource for this study.

---

**Highlights**

- The \( R_0 \) (vaccination coverage) equaled 1.94 (48.4%), 1.80 (44.4%), 3.06 (67.3%), and 2.11 (52.6%) for EG, ML, SB, and TD methods at the national level, respectively.
- The \( R_0 \) estimations indicating that an epidemic has occurred in the US (\( R_0 >1 \)).
- It is required to vaccinate at least 44.4% to prevent the next epidemics of influenza in the US.
- The estimated \( R_0 \) using SB were consistent with those calculated using TD.

**References**

1. Bahrampour A. On Multi-type Branching Process in determining reproduction number under a markovian model movement. J Stat Theory Appl. 2005; 4(1): 57-66.
2. Becker NG, Bahrampour A. Preventing epidemics with age-specific vaccination schedules. Math Biosci. 1997; 142(2): 63-77.
3. Farrington C, Kanaan M, Gay N. Estimation of the basic reproduction number for infectious diseases from age-stratified serological survey data. Appl Stat. 2001; 50(3): 251-92.
4. Khan A, Hassan M, Imran M. Estimating the basic reproduction number for single-strain dengue fever epidemics. Infect Dis Poverty. 2014; 3(1): 12-29.
5. Safi MA, Garba SM. Global stability analysis of SEIR model with holling type II incidence function. Comput Math Methods Med. 2012; 2012(2012): 1-8.
6. Arruda AG, Alkhamis MA, VanderWaal K, Morrison RB, Perez AM. Estimation of time-dependent reproduction numbers for porcine reproductive and respiratory syndrome across different regions and production systems of the US. Front Vet Sci. 2017; 4(46), 1-9.
7. Blumberg S, Lloyd-Smith J. Comparing methods for estimating \( R_0 \) from the size distribution of subcritical transmission chains. Epidemics. 2013; 5(3): 131-45.
8. Chowell G, Nishiura H. Quantifying the transmission potential of pandemic influenza. Phys Life Rev. 2008; 5(1): 50-77.
9. Griffin JT, Garske T, Ghani AC, Clarke PS. Joint estimation of the basic reproduction number and generation time parameters for infectious disease outbreaks. Biostatistics. 2010; 12(2): 303-12.
10. Haghdoost A, Baneshi MR, Zolala F, Farvahari S, Safizadeh H. Estimation of basic reproductive number of Flu-like syndrome in a primary school in Iran. Int J Prev Med. 2012; 3(6): 408-13.
11. Obadia T, Haneef R, Boëlle P-Y. The \( R_0 \) package: a toolbox to estimate reproduction numbers for epidemic outbreaks. BMC Med Inform Decis Mak. 2012; 12(1): 147-56.
12. Stadler T, Koyou R, von Wyl V, Yerly S, Böni J, Bürgisser P, et al. Estimating the basic reproductive number from viral sequence data. Mol Biol Evol. 2011; 29(1): 347-57.
13. Wallinga J, Lipsitch M. How generation intervals shape the relationship between growth rates and reproductive numbers. Proc R Soc Lond B Biol Sci. 2007; 274(1609): 599-604.
14. Wang X-S, Wu J, Yang Y. Richards model revisited: Validation by and application to infection dynamics. J Theor Biol. 2012; 313 (21): 12-19.
15. Althaus CL. Estimating the reproduction number of Ebola virus (EBOV) during the 2014 outbreak in West Africa. PLoS Curr. 2014; 6: 1-9.

16. Chowell G, Castillo-Chavez C, Fenimore PW, Kribs-Zaleta CM, Arriola L, Hyman JM. Model parameters and outbreak control for SARS. Emerg Infect Dis. 2004; 10(7): 1258-63.

17. Eichner M, Dietz K. Transmission potential of smallpox: estimates based on detailed data from an outbreak. Am J Epidemiol. 2003; 158(2): 110-7.

18. Furushima D, Kawano S, Ohno Y, Kakehashi M. Estimation of the Basic Reproduction Number of Novel Influenza A (H1N1) pdm09 in Elementary Schools Using the SIR Model. Open Nurs J. 2011; 11: 64-72.

19. Johansson MA, Arana-Vizcarrondo N, Biggerstaff BJ, Gallagher N, Marano N, Staples JE. Assessing the risk of international spread of yellow fever virus: a mathematical analysis of an urban outbreak in Asuncion, 2008. Am J Trop Med Hyg. 2012; 86(2): 349-58.

20. Nsubuga RN, White RG, Mayanja BN, Shafer LA. Estimation of the HIV basic reproduction number in rural South West Uganda: 1991–2008. PloS One. 2014; 9(1): 1-9.

21. Smith DL, McKenzie FE, Snow RW, Hay SI. Revisiting the basic reproductive number for malaria and its implications for malaria control. PLoS Biol. 2007;5(3):e42.

22. Karami M, Zahraei SM, Saboury A, Soltanshahi R, Biderafsh A, Piri N, et al. Documentation of Measles Elimination in Iran: Evidences from 2012 to 2014. J Res Health Sci. 2017; 17(3): e00387.

23. Cheng PY, Palekar R, Azziz-Baumgartner E, Iuliano AD, Alencar AP, Bresee J, et al. Burden of influenza-associated deaths in the Americas, 2002–2008. Influenza Other Respir Viruses. 2015; 9(1): 13-21.

24. Committee on Global Health and the Future of the United States. Global health and the future role of the United States. Washington DC: National Academies Press; 2017.

25. Yamamoto T. Pandemic Control Measures. Asian Med J. 2013; 56(1): 51-4.

26. Clem A, Galwankar S. Seasonal influenza: waiting for the next pandemic. J Glob Infect Dis. 2009; 1(1): 51-6.

27. Shrestha SS, Swerdlow DL, Borse RH, Prabhu VS, Finelli L, Atkins CY, et al. Estimating the burden of 2009 pandemic influenza A (H1N1) in the United States (April 2009–April 2010). Clin Infect Dis. 2011; 52(suppl1): S75-82.

28. Doshi P. Trends in recorded influenza mortality: United States, 1900–2004. Am J Public Health. 2008; 98(5): 939-45.

29. Goldstein E, Dushoff J, Ma J, Plotkin JB, Earn DJ, Lipsitch M. Reconstructing influenza incidence by deconvolution of daily mortality time series. Proc Natl Acad Sci USA. 2009; 106(51): 21825-9.

30. Chowell G, Miller M, Viboud C. Seasonal influenza in the United States, France, and Australia: transmission and prospects for control. Epidemiol Infect. 2008; 136(6): 852-64.

31. Yang W, Lipsitch M, Shaman J. Inference of seasonal and pandemic influenza transmission dynamics. Proc Natl Acad Sci U S A. 2015; 112(9): 2723-8.

32. White LF, Wallinga J, Finelli L, Reed C, Riley S, Lipsitch M, et al. Estimation of the reproductive number and the serial interval in early phase of the 2009 influenza A/H1N1 pandemic in the USA. Influenza Other Respir Viruses. 2009; 3(6): 267-76.

33. Scherer A, McLean A. Mathematical models of vaccination. Br Med Bull. 2002; 62(1): 187-99.

34. Eichner M, Schwemm M, Eichner L, Gerlier L. Direct and indirect effects of influenza vaccination. BMC Infect Dis. 2017; 17(1): 308.

35. Vink MA, Bootsma MCJ, Wallinga J. Serial intervals of respiratory infectious diseases: a systematic review and analysis. Am J Epidemiol. 2014; 180(9): 865-75.

36. Fine P, Eames K, Heymann DL. “Herd immunity”: a rough guide. Clin Infect Dis. 2011;52(7): 911-16.

37. Forsberg White L, Pagano M. A likelihood-based method for real-time estimation of the serial interval and reproductive number of an epidemic. Stat Med. 2008; 27(16): 2999-3016.

38. CDC FluView Weekly Report. National Regional and State level outpatient illness and Viral Surveillance; 2018; Available from: https://gis.cdc.gov/grasp/fluview/fluportal dashboard.html.

39. Gumel AB, Nuño M, Chowell G. Mathematical assessment of Canada’s pandemic influenza preparedness plan. Can J Infect Dis Med Microbiol. 2008; 19(2): 185-92.