Research Article

A Bayesian Inferential Approach to Quantify the Transmission Intensity of Disease Outbreak

Adiveppa S. Kadi and Shivakumari R. Avaradi

Department of Studies in Statistics, Karnatak University, Dharwad 580003, India

Correspondence should be addressed to Shivakumari R. Avaradi; shiv.avaradi@gmail.com

Received 30 October 2014; Revised 16 January 2015; Accepted 20 January 2015

Academic Editor: Xiaojun Yao

Copyright © 2015 A. S. Kadi and S. R. Avaradi. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Background. Emergence of infectious diseases like influenza pandemic (H1N1) 2009 has become great concern, which posed new challenges to the health authorities worldwide. To control these diseases various studies have been developed in the field of mathematical modelling, which is useful tool for understanding the epidemiological dynamics and their dependence on social mixing patterns. Method. We have used Bayesian approach to quantify the disease outbreak through key epidemiological parameter basic reproduction number ($R_0$), using effective contacts, defined as sum of the product of incidence cases and probability of generation time distribution. We have estimated $R_0$ from daily case incidence data for pandemic influenza A/H1N1 2009 in India, for the initial phase. Result. The estimated $R_0$ with 95% credible interval is consistent with several other studies on the same strain. Through sensitivity analysis our study indicates that infectiousness affects the estimate of $R_0$. Conclusion. Basic reproduction number $R_0$ provides the useful information to the public health system to do some effort in controlling the disease by using mitigation strategies like vaccination, quarantine, and so forth.

1. Introduction

Influenza is an emerging infectious disease and influenza-like illness (ILI) is a clinical illness caused by the influenza virus, which gave rise to human pandemics such as 1918 Spanish flu (H1N1), 1957 Asian flu (H3N2), 1968 Hong Kong flu (H2N2), and most recently H1N1 pandemic 2009. Influenza A (H1N1) was originally referred to as “swine flu” because laboratory testing showed that many of the genes in this new virus were very similar to those found in pigs in North America. Further on, it has been found that this new virus has gene segments from the swine, avian, and human flu virus genes. The scientists call this a “quadruple reassortant” virus and hence this new (novel) virus is christened “influenza-A (H1NI) virus” [1].

A reassorted influenza was first detected in Mexico on March 18, 2009, and rapidly spread to the United States, Canada, and subsequently all regions worldwide including India. The first case of H1NI was reported in India on May 17, 2009, at Hyderabad Airport in a young boy who travelled from the USA and later it spread throughout the country [2]. According to World Health Organization (WHO) update of November 20, 2009, virus spread across more than 206 countries resulting in 6770 deaths [3]. In June 2009 World Health Organization (WHO) raised the level of pandemic alert phase 5 to phase 6 [4]. According to the Directorate General of Health Services, Government of India, New Delhi, update on November 19, 2009, there have been 16044 laboratories confirming cases recorded in India resulting in 537 deaths [5].

To formulate the valid and reliable estimate of transmissibility and spread of an outbreak we have utilised statistical modelling, which facilitates our understanding of mechanism of disease spread. To access the intensity of an outbreak, transmission potential can be quantified by reproduction number $R_0$, that is, average number of secondary cases generated by a single primary case in a completely susceptible population [6, 7]. The importance of basic reproduction number $R_0$ becomes more apparent when an emerging infectious disease strikes a population which is a key concept in the epidemic theory. If $R_0$ is less than or equal to one, then transmission in the population goes stochastically extinct with probability of one after a small number of infections. If $R_0$
is greater than 1, then there is a positive probability of a large epidemic. Statistical estimation of $R_0$ is used to understand the transmission dynamics and evolution of the infectious disease which facilitate designing the effective public health intervention strategies and mitigation policies [8].

The aim of present study is to quantify the intensity of pandemic influenza A/H1N1 2009 in India. To achieve this goal we have calculated basic reproduction number from time series dataset of H1N1 through Bayesian approach to contact patterns. We have estimated basic reproduction number $R_0$ with 95% credible interval.

2. Material and Methods

2.1. Data Sources. Time series data for the 2009 influenza A/H1N1 pandemic in India was obtained from the Ministry of Health and Family Welfare (MoHFW), Government of India [9]. 31,924 infected cases with 1525 deaths were reported during May 17, 2009, to May 17, 2010, with two complete waves of epidemic. We have represented the data of the first wave from May to October 2009 (see Figure 1). For the analysis we have used daily reported cases (incidence) for initial phase of an epidemic from June 11 to August 13, 2009 [10].

2.2. Statistical Method. Statistical inference of $R_0$ is still in progress, and it is recognized that the estimate is very sensitive to dispersal of the disease progression [11]. Estimation of $R_0$ can be illustrated by employing time-since-infection model and suggests origin of transmission of infectious diseases which is the counterpart of compartmental models like SI, SIS, and so forth. Both models are originated in the basic paper of Kermack and McKendrick [12], and both the SIR model and the simplest time-since-infection model are known as "the Kermack-McKendrick age structure model." It is used to identify key epidemiological parameter by using a simple renewal process which adheres to the basic reproduction number $R_0$.

Let $j(t)$ represent the number of new infections or incidence at calendar time $t$; that is, each infected individual on an average generates secondary cases at a rate $A(\tau)$ at time-since-infection $\tau$ which is a specific case of renewal equation of birth process [13–15]. Consider

$$j(t) = \int_0^\infty A(\tau) j(t-\tau) \, d\tau. \quad (1)$$

Since $R_0$ represents average number of secondary cases that a primary case generates during entire his/her infectious period, the estimate is given by [16, 17]

$$R_0 = \int_0^\infty A(\tau) \, d\tau. \quad (2)$$

Under Kermack and McKendrick assumption "single infection causes an independent process of infection with host" which allows an age representation for the state of infection, that is, infectivity of an individual. The time elapsed since infection is called the infection age or time since infection, whereas $A(\tau)$ is expected infectivity of an individual with time-since-infection $\tau$ whereas $A(\tau)$ becomes the rate of secondary transmission per single primary case at time-since-infection $\tau$ [18]:

$$A(\tau) = \beta(\tau) \Gamma(\tau), \quad (3)$$

where $\beta(\tau)$ is the transmission rate which depends on frequency of contact and infectiousness at infection age $\tau$ and $\Gamma(\tau)$ is the probability of being infectious at infection age $\tau$. Substituting $A(\tau)$ into (1) we get

$$j(t) = \int_0^\infty \beta(\tau) \Gamma(\tau) j(t-\tau) \, d\tau. \quad (4)$$

Further, we consider a probability density of the generation time where generation time is defined as time from infection of an individual to the infection of a secondary case by that individual, denoted by $g(\tau)$ through normalized density of secondary transmission [16, 17, 19]. We have

$$g(\tau) = \frac{\beta(\tau) \Gamma(\tau)}{\int_0^\infty \beta(\tau) \Gamma(\tau) \, d\tau} = \frac{\beta(\tau) \Gamma(\tau)}{R_0}. \quad (5)$$

Using (5) in (4) we get

$$j(t) = \int_0^\infty R_0 g(\tau) j(t-\tau) \, d\tau. \quad (6)$$

Then, the basic reproduction number $R_0$ is as in

$$R_0 = \frac{\int_0^\infty j(t) \, g(\tau) j(t-\tau) \, d\tau}{\int_0^\infty g(\tau) j(t-\tau) \, d\tau}. \quad (7)$$

In reality, the case incidences are rather in discrete form as daily/weekly reports. The discretized analogy of (7) can be derived as

$$R_0 = \frac{\sum_{s=0}^{\infty} w_s j_{t-s}}{\sum_{s=0}^{\infty} w_s j_{t-s}}. \quad (8)$$
By taking the inverse of both sides of (8) we get

\[
\frac{1}{R_0} = \sum_{s=0}^{\infty} w_s j_{i-s} / j_i,
\]

where \(s\) is the discretized form of time since infection (usually in days), \(i\) is discretized calendar time (usually in days/weeks), and \(w_s\) is the generation time distribution for an infectious disease which is the probability distribution function for the time from infection of an individual to the infection of a secondary case by that individual [19, 20].

From Figure 2 the concept of \(R_0\) through contact patterns has been illustrated here before taking up the Bayesian method of estimation. From Figure 2(a), the transmission tree with \(R_0 = 2\) represents who infected whom, where each primary case on an average generates “2” secondary cases. Secondary transmissions from primary to secondary cases are given by the basic reproduction number \(R_0 = 2\). From Figure 2(b) restructuring the transmission tree, given that all the potential contacts made by primary cases with the probability of each possible contact resulted in a secondary transmission, is \(1/R_0\). This type of transmission tree represents who infected whom which is unobservable in nature unless rigorous contact tracing is implemented [21, 22].

The numerator of the right hand side of (9) represents the total number of effective contacts made by a possible primary case in day/weeks which have an equal probability of resulting in the secondary transmission; that is, the probability that a secondary case is linked to an effective contact made by a single primary case at time \(i\) is given by \(1/R_0\). Effective contacts lead to potential secondary cases with equal chance of getting the infection from the primary cases at time \(i\). It has been seen that every contact does not lead to successful transmission of infection; that is, the effective contacts are uncertain which is defined as the contact that is sufficient to lead to the transmission of infection between infectious and susceptible population. The total number of effective contacts made by a potential primary case at time \(i\) is the sum of the product of incidence cases and generation time distribution during the generation interval of length \(m\). This indicates a simple binomial law of uncertainty in effective contacts [22].

The likelihood function of \(1/R_0\) is as follows:

\[
L \left( \frac{1}{R_0} \right) = \prod_{i=1}^{n} \left( \sum_{j=0}^{m} w_j j_{i-j} \right) \cdot \left( \frac{1}{R_0} \right)^{j_{i-C_i}} \left( 1 - \frac{1}{R_0} \right)^{C_i - j_{i-C_i}}
\]

Here, Bayesian inferential approach is used to estimate parameter \(R_0\) which provides us with different but related estimate by combining prior belief and the evidence observed. As more evidence is gathered the prior distribution is modified into the posterior distribution that represents the uncertainty over the parameter values. Posterior distribution is derived from the Bayes formula [23]

\[
\pi(\theta/x) = \frac{f(x/\theta) \pi(\theta)}{\int_{\Theta} f(x/\theta) \pi(\theta) d\theta},
\]

where \(\Theta\) is the parameter space.
Table 1: Sensitivity analysis of basic reproduction number $R_0$ is depending on generation time distribution as Weibull distribution for time since infection $s$ for 7 days as well as 10 days.

| Prior distribution for $s = 7$ days | Basic reproduction number $R_0$ (with 95% CrI) | Prior distribution for $s = 10$ days | Basic reproduction number $R_0$ (with 95% CrI) |
|-------------------------------------|-----------------------------------------------|-------------------------------------|-----------------------------------------------|
| Beta(1, 1)                          | 1.2548 (1.2223, 1.2923)                        | Beta(2, 1)                          | 1.3392 (1.3128, 1.3938)                        |
| Beta(4, 2)                          | 1.2543 (1.2250, 1.2830)                        | Beta(3, 4.6, 5.2)                   | 1.3323 (1.2962, 1.3762)                        |
| Beta(3.46, 5.2)                     | 1.2578 (1.2309, 1.2909)                        | Beta(4, 4.2)                        | 1.3296 (1.2895, 1.3695)                        |
| Beta(1.75, 3.5)                     | 1.2569 (1.2279, 1.2879)                        | Beta(7, 3.5)                        | 1.3303 (1.2969, 1.3669)                        |

where $\int_{\theta} f(x/\theta) \pi(\theta) d\theta$ is a normalization constant, $x$ indicated data, $\theta$ is the unknown quantity, $\pi(\theta)$ is the prior distribution, $f(x/\theta)$ is the likelihood function, and posterior distribution $\pi(\theta/x)$ completely describes the uncertainty. There are two key advantages of Bayesian theory: (i) once the uncertainty in the posterior distribution is expressed via probability distribution then the statistical inference can be automated and (ii) available prior information is reasonably incorporated into the statistical model. Now, the posterior estimate of parameter $R_0$ is derived using its prior information, where likelihood function follows binomial distribution with conjugate prior as beta distribution of first kind [24]. That is, $1/R_0 \sim \text{Beta}(a, b)$, with realistic choice of parameters ($a, b$):

$$
\pi \left( \frac{1}{R_0} \right) = \frac{1}{\beta(a, b)} \left( \frac{1}{R_0} \right)^{a-1} \left( 1 - \frac{1}{R_0} \right)^{b-1}
$$

(13)

Now posterior distribution of $1/R_0$ is proportional to the likelihood times prior. Therefore,

$$
\text{Posterior} = \frac{\text{likelihood} \times \text{prior}}{\text{normalization constant}}
$$

$$
= \left[ \prod_{i=1}^{n} \left( \frac{j_i}{\theta} \right) \right] \left( \frac{1}{R_0} \right)^{\sum_{i=1}^{n} j_i} \left( 1 - \frac{1}{R_0} \right)^{\sum_{i=1}^{n} \sum_{j=1}^{i} j_i} \left( \frac{1}{\beta(a, b)} \right) \left( \frac{1}{R_0} \right)^{a-1} \left( 1 - \frac{1}{R_0} \right)^{b-1} \int_{1/R_0} [\prod_{i=1}^{n} \left( \frac{j_i}{\theta} \right) \right] \left( \frac{1}{R_0} \right)^{\sum_{i=1}^{n} j_i} \left( 1 - \frac{1}{R_0} \right)^{\sum_{i=1}^{n} \sum_{j=1}^{i} j_i} \left( \frac{1}{\beta(a, b)} \right) \left( 1/R_0 \right)^{a-1} \left( 1 - 1/R_0 \right)^{b-1} d(1/R_0)
$$

(14)

We are interested in estimating $R_0$ with its 95% credible interval (CrI) which has been derived by considering the sampling transformation of $1/R_0$ through simulation from the above posterior distribution with different choices of priors for beta distribution. Theoretically, it is hard to find the posterior distribution of $R_0$ where $1/R_0$ is a beta variable. Through simulation we have generated 10000 samples from beta posterior distribution and also estimated 95% CrI. We have considered beta distribution with several combinations of mean (ranges from 0.4 to 0.8) and accordingly we chose different values of $(a, b)$. The second and fourth column of Table 1 represent our estimates of $R_0$ which are posterior means along with 95% CrI. Figure 3 displays the posterior distribution of $R_0$ with different prior choices.

Generation time is another most important characteristic in infectious disease epidemiology, since $R_0$ indicates only the average number of secondary infections one primary infection produces in one disease generation. When we consider disease transmission in real time scale such as days or weeks, it matters a lot how long one disease generation lasts. Generation time is the average time taken for secondary infections produced by a primary infection [25]. Generation time interval or generation time distribution is assumed to be known as Weibull distribution which is a biologically plausible choice [20, 26–28] with a mean of 1.78 and 2.48 days and a standard deviation (SD) of 0.66 and 1.06 days for $s = 7$ and 10 days [29, 30] (Tables S1 and S2 in Supplementary Material available online at http://dx.doi.org/10.1155/2015/256319).
Figure 3: Histogram of posterior distribution of $R_0$ by using different values of prior choices for beta distribution.
All the sensitivity analysis was done by using MATLAB (Supplementary Material, Algorithm).

2.3. Sensitivity Analysis. Sensitivity analysis is a statistical technique which provides insight into how uncertainty in input variables affects the model outputs and which input variable tends to derive variation in the outputs [31]. We performed sensitivity analysis to quantify the effect of changes on \( R_0 \). It has been used to determine how sensitive an estimate of the parameter is. It is usually performed as series of tests in which one can use different set of hyperparameter values to see the change in the estimate. Our analysis is based on the pandemic influenza A/H1N1 in India 2009 through the Bayesian estimates of basic reproduction number; we used the daily reported cases to calculate effective contacts. We have calculated posterior distribution of \( R_0 \) using prior as beta distribution with different values of parameter choices. From Figure 1 we have seen that as prior choice changes the shape of the posterior distribution also changes.

3. Results and Discussion
The estimates of \( R_0 \) for the 2009 H1N1 influenza pandemic were mainly reported based on the data obtained in the first few months of pandemic or based on whole first wave data. Most of these \( R_0 \) estimates ranges from 1.1 to 2 [32–37]. Our estimated value of the basic reproduction number indicates the milder intensity of disease transmission in India. Interestingly, this estimated \( R_0 \) with 95% credible interval is consistent with several other studies on the same strain [35], along with many European countries [28]. Notably, it has a smaller credible length which is more reliable estimate; see Table 1. Statistical inference of \( R_0 \) is based on incidence (reported cases) and known generation time distribution. Some differences among these estimates are due to the choice of generation time distribution because \( R_0 \) estimation relies much on the assumptions of the generation time distribution [38]. In general, shorter mean generation time may lead to smaller \( R_0 \) estimates. Since, the estimate of \( R_0 \) crucially depends on generation time distribution. From Table 1, we conclude that generation time or infectiousness of an individual affects the basic reproduction number. This method does not require exponential growth assumption. Still our estimate is greater than one so one has to make effort in controlling the disease through control strategies, which are typically targeted to bring this number below one and maintain it, as this will lead to eventual extinction of the epidemic.

3.1. Limitations. This method is applied only for initial stage of the epidemic (exponential phase) when there is no intervention like quarantine, isolation vaccination, and so forth. If basic reproduction number is \( R_0 < 1 \), then the probability \( 1/R_0 \) terminates because it exceeds the law of probability.

Conflict of Interests
The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments
The authors thank the University Grants Commission (UGC) through Research Fellowship in Science for Meritorious Students (RFMS) and DST (Science & Engineering Research Board) Project (no. SR/S4/MS: 396/10) New Delhi, India, for research funding support. They are thankful to Sheikh Taslim Ali for his motivation and suggestions.

References
[1] Factsheet influenza, http://www.pib.nic.in/h1n1/factsheet.pdf.
[2] Pandemic Influenza: A H1N1 Clinical Management Protocol and Infection Control, Guidelines, http://www.mohfw-h1n1.nic.in/guidelines.html.
[3] WHO, “pandemic H1N1 2009-update,” 2009, http://who.int/csr/don/2009/enindex.html.
[4] World Health Organization, Pandemic (H1N1) 2009—Update 99, 2010, http://www.who.int/csr/don/2010_05_07/en/index.html.
[5] Ministry of Health and Family Welfare and Government of India, Pandemic Influenza A/H1N1, 2009, http://mohfw-h1n1.nic.in.
[6] O. Diekmann, J. A. P. Heesterbeek, and J. A. J. Metz, “On the definition and the computation of the basic reproduction ratio \( R_0 \) in models for infectious diseases in heterogeneous populations,” Journal of Mathematical Biology, vol. 28, no. 4, pp. 365–382, 1990.
[7] R. M. Anderson and R. M. May, Infectious Diseases of Humans: Dynamics and Control, Oxford University Press, Oxford, UK, 1991.
[8] K. Dietz, “The estimation of the basic reproduction number for infectious diseases,” Statistical Methods in Medical Research, vol. 2, no. 1, pp. 23–41, 1993.
[9] Ministry of Health and Family Welfare, Government of India, May 2010, http://www.mohfw-h1n1.nic.in/.
[10] S. R. Gani, S. T. Ali, and A. S. Kadi, “The transmission dynamics of pandemic influenza A/H1N1 2009-2010 in India,” Current Science, vol. 101, no. 8, 2011.
[11] J. M. Heffernan and L. M. Wahl, “Improving estimates of the basic reproductive ratio: using both the mean and the dispersal of transition times,” Theoretical Population Biology, vol. 70, no. 2, pp. 135–145, 2006.
[12] W. O. Kermack and A. G. McKendrick, “A contribution to the mathematical theory of epidemics,” Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, vol. 115, no. 772, pp. 700–721, 1927.
[13] W. Feller, “On the integral equation of renewal theory,” Annals of Mathematical Statistics, vol. 12, pp. 243–267, 1941.
[14] B. R. Levin, J. J. Bull, and E. M. Stewart, “The intrinsic rate of increase of HIV/AIDS: epidemiological and evolutionary implications,” Mathematical Biosciences, vol. 132, no. 1, pp. 69–96, 1996.
[15] J. A. P. Heesterbeek and K. Dietz, “The concept of \( R_0 \) in the epidemic theory,” Statistica Neerlandica, vol. 50, no. 1, pp. 89–110, 1996.
[16] J. A. J. Metz and O. Diekmann, The Dynamics of Physiologically Structured Populations, Springer, Berlin, Germany, 1986.
[17] J. Wallinga and M. Lipsitch, “How generation intervals shape the relationship between growth rates and reproductive numbers,” Proceedings of the Royal Society B, vol. 274, no. 1609, pp. 599–604, 2007.
[18] G. Chowell and H. Nishiura, “Quantifying the transmission potential of pandemic influenza,” Physics of Life Reviews, vol. 5, no. 1, pp. 50–77, 2008.

[19] A. Svensson, “A note on generation times in epidemic models,” Mathematical Biosciences, vol. 208, no. 1, pp. 300–311, 2007.

[20] J. Wallinga and P. Teunis, “Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures,” The American Journal of Epidemiology, vol. 160, no. 6, pp. 509–516, 2004.

[21] D. T. Haydon, M. Chase-Topping, D. J. Shaw et al., “The construction and analysis of epidemic trees with reference to the 2001 UK foot-and-mouth outbreak,” Proceedings of the Royal Society of London B: Biological Sciences, vol. 270, no. 1511, pp. 121–127, 2003.

[22] H. Nishiura, “Correcting the actual reproduction number: a simple method to estimate R0 from early epidemic growth data,” International Journal of Environmental Research and Public Health, vol. 7, no. 1, pp. 291–302, 2010.

[23] C. P. Robert and G. Casella, Monte Carlo Statistical Methods, Springer Texts in Statistics, Springer, New York, NY, USA, 2004.

[24] F. J. Anscombe and T. Bayes, Bayesian Statistics, Academic Class Notes, 1992.

[25] R. M. Anderson, C. Fraser, A. C. Ghani et al., “Epidemiology, transmission dynamics and control of SARS: the 2002-2003 epidemic,” Philosophical Transactions of the Royal Society B: Biological Sciences, vol. 359, no. 1447, pp. 1091–1105, 2004.

[26] L. E. White and M. Pagano, “A likelihood based method for real time estimation of the serial interval and reproductive number of an epidemic,” Harvard University Biostatistics Working Paper Series 50, 2006.

[27] T. Garske, P. Clarke, and A. C. Ghani, “The transmissibility of highly pathogenic avian influenza in commercial poultry in industrialised countries,” PLoS ONE, vol. 2, no. 4, article e349, 2007.

[28] B. J. Cowling, V. J. Fang, S. Riley, I. S. M. Peiris, and G. M. Leung, “Estimation of the serial interval of influenza,” Epidemiology, vol. 20, no. 3, pp. 344–347, 2009.

[29] S. Cauchemez, C. A. Donnelly, C. Reed et al., “Household transmission of 2009 pandemic influenza A (H1N1) virus in the United States,” The New England Journal of Medicine, vol. 361, no. 27, pp. 2619–2627, 2009.

[30] N. M. Ferguson, D. A. T. Cummings, S. Cauchemez et al., “St rates for containing an emerging influenza pandemic in Southeast Asia,” Nature, vol. 437, no. 7056, pp. 209–214, 2005.

[31] M. D. Samsuzzoha, M. Singh, and D. Lucy, “Uncertainty and sensitivity analysis of the basic reproduction number of a vaccinated epidemic model of influenza,” Applied Mathematical Modelling, vol. 37, no. 3, pp. 903–915, 2013.

[32] Y. Yang, J. D. Sugimoto, M. Elizabeth Halloran et al., “The transmissibility and control of pandemic influenza a (H1N1) virus,” Science, vol. 326, no. 5953, pp. 729–733, 2009.

[33] A. Ghani, M. Baguelin, J. Griffin et al., “The early transmission dynamics of H1N1pdm influenza in the United Kingdom,” PLoS Currents, Article ID RRN130, 2010.

[34] B. Pourbohloul, A. Ahued, B. Davoudi et al., “Initial human transmission dynamics of the pandemic (H1N1) 2009 virus in North America,” Influenza and Other Respiratory Viruses, vol. 3, no. 5, pp. 215–222, 2009.

[35] M. Baguelin, A. J. V. Hoek, M. Jit, S. Flasche, P. J. White, and W. J. Edmunds, “Vaccination against pandemic influenza A/H1N1v in England: a real-time economic evaluation,” Vaccine, vol. 28, no. 12, pp. 2370–2384, 2010.