

Estimating Transmissibility of Seasonal Influenza Virus by Surveillance Data

Shenghai Zhang
Public Health Agency of Canada

Abstract: It is important to estimate transmissibility of influenza virus during its growing phase for understanding the propagation of the virus. The estimation procedures of the transmissibility are usually based on the data generated in flu seasons. The data-generating process of the outbreak of influenza has many features. The data is generated by not only a biological process but also control measures such as flu vaccination. The estimation is discussed by considering the aspects of the data-generating process and using the model to capture the essential characteristics of flu transmission during the growing phase of a flu season.

Key words: Reproduction number, seasonal influenza, surveillance data, transmissibility of flu virus.

1. Introduction

Influenza, or the flu, leads to an average of 20,000 hospitalization yearly in Canada. Influenza and its complications results in 2,000 to 8,000 deaths in Canada each year (PHAC (2009b)). A novel influenza A(H1N1) virus variant has spread globally since its first appearance in April 2009 (Fraser (2009) and Team (2009)) and after nearly 30,000 confirmed cases had been reported in 74 countries as of 11 June 2009, World Health Organization (WHO) declared “the start of the 2009 influenza pandemic (Cohen and Enserink (2009)). However, the new strain of influenza A (H1N1) is causing mild to moderate disease so far (WHO (2009)). People are very concerned that if the novel influenza A (H1N1) virus will replace the human H1 virus as the seasonal influenza virus and evolve antigenic variants every year (Belshe (2009)). Seasonal influenza has already struck the elderly and infant population hard (Dolin, 2005). Therefore, seasonal influenza surveillance will play an very important role in monitoring spread of flu virus.

The aim of this article is to study the transmissibility of the seasonal influenza by modeling influenza transmission by using surveillance data. A measure of

transmissibility and the stringency of control measures required to stop the epidemic could be determined by the magnitude of the basic reproduction number (R_0) (Anderson and May (1991)), which is the average number of secondary cases generated by a primary case. It plays a central role in tracking infectious disease outbreak (Anderson and May (1991)). This quantity has been used for discussing infectious disease dynamics as well as the control strategy. For example, a control programme could include the vaccination to produce immunity to the influenza virus and the education at the same time to reduce the contact rate. Based on the value of R_0 , it is possible to know the minimum effort of the programme to eliminate or eradicate the disease, which is $1 - R_0^{-1}$ and the minimum proportional reduction of the susceptible fraction by the public health policy. Therefore, a control method for reducing the parameters by about 50% may be efficient for eliminating one disease when the values of R_0 between 1 and 2, however, it may be unsuccessful in another situation where the $R_0 > 3$ (Diekmann, Heesterbeek and Metz (1998) and Diekmann and Heesterbeek (2000)).

There are many ways to estimate the R_0 . If the contacts are clearly defined such that they can be counted, then the estimation of R_0 can be based on the individual parameters: the number of persons contacted per unit of time by one infectious individual, denoted by κ ; given a contact, the proportion of contacted persons who are infected, λ ; the duration of the infectious period, $1/\gamma$ (γ could be explained as the rate of transfer from a infectious state to a non-infectious state). Then, the reproduction number R_0 can be expressed by $R_0 = \kappa\lambda/\gamma$. The actual average number of secondary cases per primary case at time t after the disease outbreak is called effective reproduction number R_t . The purpose of this article is to estimate the effective reproduction number for influenza during its growing phase, based on the surveillance data. Usually, the surveillance data is not sufficient to validate a detailed model for studying the transmission of influenza, because the individual parameters may not be available. Therefore, the estimation and analysis procedures have to rely on methods with broader model assumptions which can be applied to the data from surveillance.

2. Data and Models

Surveillance of seasonal influenza-like illness is conducted in many countries (Ginsberg *et al.* (2009)). The surveillance of influenza infection is conducted in Canada for many years. FluWatch is a national surveillance system of Canada which collects data and information from various sources in order to provide a national picture of influenza activity. The FluWatch program monitors the spread of flu and flu-like illnesses on an on-going basis (PHAC (2009a)). It is important to analyze routinely collected surveillance data to fulfill the purpose of such surveillance for detection of epidemic, especially detection in the early

stage of the epidemic. Fig. 1 and Fig. 2 show number of weekly reported flu tests and number of weekly reported positive influenza specimens from 2001 to 2006, respectively.

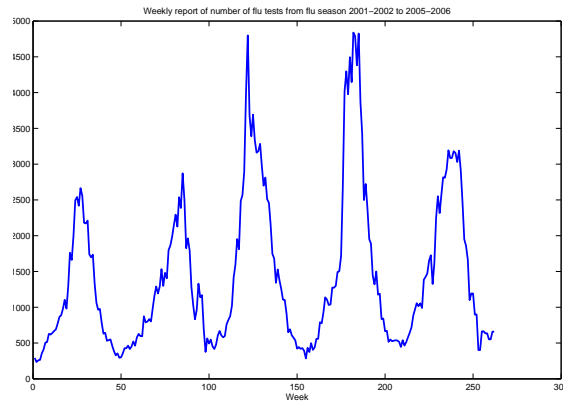


Figure 1: The number of weekly reported influenza tests from 2001-2006

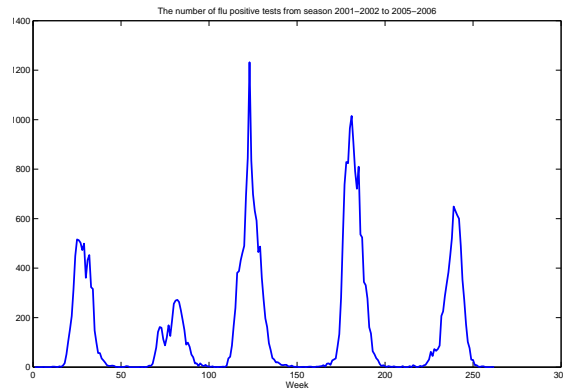


Figure 2: The number of weekly reported influenza positive tests from 2001-2006

Many studies show that the time series of weekly incident rates have a clear mixture of two dynamics: a low-level dynamic with incidence rates that vary according to a seasonal pattern (the non-epidemic dynamic) and a high-level dynamic, in which the incidence rate increases sharply at irregular intervals (the epidemic dynamic). Based on the surveillance data, the reproduction number can be estimated to detect which state the influenza activity are currently going through. The transmission dynamics of the disease will be described to obtain the estimator of the transmissibility of seasonal flu.

If a susceptible individual is infected, she/he would be in the latent period and then could infect other people. If an infected person could be recovered or dead, then individuals can be classified as susceptible (S), exposed (E), infectious (I), recovered (R) and dead (D). The contact rate at which infective individuals encounter and infect susceptible individuals is denoted by ρ . The infectious individuals either to recover or to die from the disease; let γ and δ denote rates for the discovery and the death, respectively. The rate for latent individuals to progress to the infectious class is denoted by ζ . Let $S(t)$, $E(t)$, $I(t)$, $R(t)$ and $D(t)$ denote number of susceptible individuals, number of exposed individuals, number of infectious individuals, number of recovered individuals, number of death from the disease in the time t , respectively. An infection at time t is an event which leads to $S(t) \rightarrow S(t) - 1$ & $E(t) \rightarrow E(t) + 1$; similarly, the event $E(t) \rightarrow E(t) - 1$ & $I(t) \rightarrow I(t) + 1$ corresponding to an end of a latent stage; etc. It is clear that all these events are random. Therefore, $S(t)$, $E(t)$, $I(t)$, $R(t)$ and $D(t)$ are stochastic processes. However, it is difficult to know distributions of these processes. It is assumed that the first moments of the random processes are continuous differential and they follow the well-known compartmental model of SEIR-type (Stone, Olink and Huppert (2007), Deguen, Thoms and Chau (2000) and Li and Muldowney (1995)).

$$\begin{cases} \frac{d\mathbb{E}(S(t))}{dt} = -\rho\mathbb{E}(S(t))\mathbb{E}(I(t)), \\ \frac{d\mathbb{E}(E(t))}{dt} = \rho\mathbb{E}(S(t))\mathbb{E}(I(t)) - \zeta\mathbb{E}(E(t)), \\ \frac{d\mathbb{E}(I(t))}{dt} = \zeta\mathbb{E}(E(t)) - (\gamma + \delta)\mathbb{E}(I(t)), \\ \frac{d\mathbb{E}(R(t))}{dt} = \gamma\mathbb{E}(I(t)), \\ \frac{d\mathbb{E}(D(t))}{dt} = \delta\mathbb{E}(I(t)). \end{cases} \quad (2.1)$$

Let $C(t)$ denote the number of cumulative cases at time t , that is, $\mathbb{E}(C(t)) = N - \mathbb{E}(S(t))$, where \mathbb{E} denotes the expectation of a random variable and N is a constant for the number of total population. The growing phase is defined as time period $T = (t_0, t_1)$, where $\log(\mathbb{E}(C(t+1) - C(t))/\mathbb{E}(C(t) - C(t-1))) > 0$ for all $t \in T$. For seasonal influenza, the phase of the number of cases increasing in an exponential way is the growing phase. A growing phase of seasonal influenza is far more rapid than the vital dynamics of a population, thus, the birth-death processes is neglected in the model described by (2.1). In practice, the individuals who are in the latent period are not observed from the surveillance data. Therefore, the concerned population are divided into three parts: susceptible ones, infected ones and recovered or dead. The model reduces to *SIR* model (Anderson and May (1979) and Laura and Guy (2006)) and $\mathbb{E}(S(t) + I(t) + R(t) + D(t)) = N$. Then,

$$\frac{d\mathbb{E}(I(t))}{dt} = \left[\rho - \gamma - \delta - \rho\mathbb{E}(I(t)) - \rho\gamma \int_0^t \mathbb{E}(I(w))dw \right] \mathbb{E}(I(t)).$$

In the growing phase of the disease, for example, in the first week of outbreak of the disease, it can be assumed that the infected individuals are always in their infectious state. Therefore, $\frac{d\mathbb{E}(I(t))}{dt} = \rho(1 - \mathbb{E}(I(t)))\mathbb{E}(I(t))$, when epidemiological reports are given by number of cases rather than in terms of infectious individuals, which at the time of reporting may have progressed. In this case, it is possible to use the following approximation (Vynnycky, Trindall and Mangtani (2007) and Chowell, Nishiura and Bettencourt (2007)). First, it is noticed that $\frac{d\mathbb{E}(C(t))}{dt} = \rho\mathbb{E}(S(t))\mathbb{E}(I(t))$. Considering $\mathbb{E}(S(t + \tau)) \approx \mathbb{E}(S(t))$ and $\frac{d\mathbb{E}(C(t))}{dt} \approx \frac{\mathbb{E}(C(t) - C(t - \tau))}{\tau}$, when τ is small, it can be obtained that $\mathbb{E}(C(t + \tau) - C(t)) \approx \tau\rho\mathbb{E}(S(t))\mathbb{E}(I(t + \tau))$, and $\mathbb{E}(C(t) - C(t - \tau)) \approx \tau\rho\mathbb{E}(S(t))\mathbb{E}(I(t))$. Therefore, it follows from $R_t = \frac{\rho}{\gamma} \frac{\mathbb{E}(S(t))}{N}$ that

$$\mathbb{E}(C(t + \tau) - C(t)) \approx \mathbb{E}(C(t) - C(t - \tau))e^{\tau\gamma(R_t - 1)}.$$

Considering weekly reported data, the effective reproduction number R_t in the growing phase can be approximated by the following formula:

$$1 + \frac{1}{\gamma} \log \frac{\mathbb{E}(W(t + 1))}{\mathbb{E}(W(t))},$$

where $W(t)$ is the number of reported cases in week t .

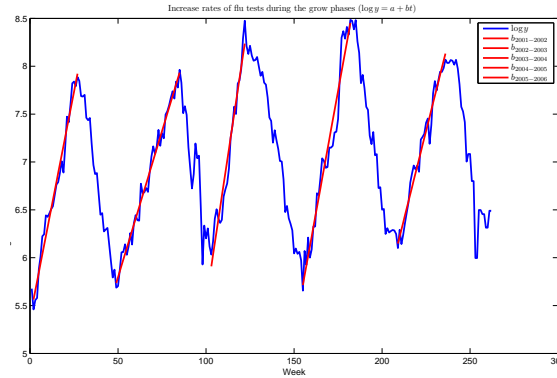


Figure 3: The increase rates of weekly reported influenza tests from 2001-2006

3. Application to Flu Surveillance Data

FluWatch program in Canada disseminates information through weekly reports during the active influenza season and biweekly reports during the low season (mid-May to September). Reports are available to health professionals

and the public. It is noticed that during the growing phase of a flu season the log transformation of the number of flu tests can be approximated by a linear regression line (see Fig. 3). Let $z(t)$ be the numbers of flu positive tests in the week t during the growing phases. Then, $\log z(t)$ can be modeled as $\log z(t) = \beta_0 + \beta t + \epsilon$, where $\epsilon \in N(0, \sigma^2)$ and β can be estimated for each flu season and is shown in the Table 1. The Fig. 4 shows the residual case order plot: the case order error-bar plot of confidence intervals on residuals from estimating β .

Table 1: Estimates of growth rates of flu **positive** tests during growing phases

Parameter	Estimate	CI
$\hat{\beta}_{2001-2002}$	0.6148	(0.4581 0.7715)
$\hat{\beta}_{2002-2003}$	0.2325	(0.1577 0.3073)
$\hat{\beta}_{2003-2004}$	0.3282	(0.2407 0.4157)
$\hat{\beta}_{2004-2005}$	0.3509	(0.3129 0.3888)
$\hat{\beta}_{2006-2007}$	0.2698	(0.2371 0.3025)

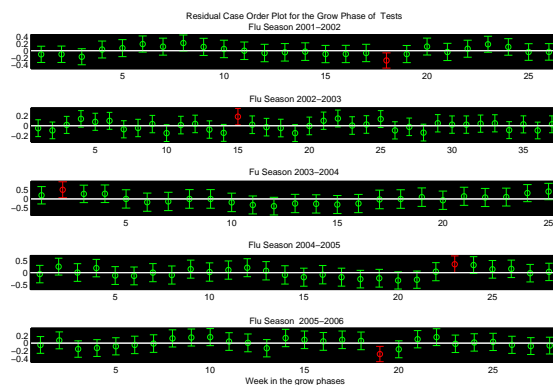


Figure 4: Errorbar plot of confidence intervals on residuals from regression for growth rates of flu tests

Similarly, number of flu tests in the growing phase of flu season can be approximated by $\log y(t) = b_0 + bt + \epsilon$, where $y(t)$ is the number of reported flu tests in the growing phase and $\epsilon \in N(0, \sigma^2)$. The estimated b for each flu season is shown in the Table 2. The Fig. 5 shows the residual case order error-bar plot of confidence intervals on residuals from each regression.

Table 2: Estimates of growth rates of flu tests during growing phase of flu season

Parameter	Estimate	CI
$b_{2001-2002}$	0.0943	(0.0878 0.1009)
$b_{2002-2003}$	0.0611	(0.0583 0.0638)
$b_{2003-2004}$	0.0973	(0.0828 0.1118)
$b_{2004-2005}$	0.1014	(0.0924 0.1104)
$b_{2005-2006}$	0.0735	(0.0682 0.0789)

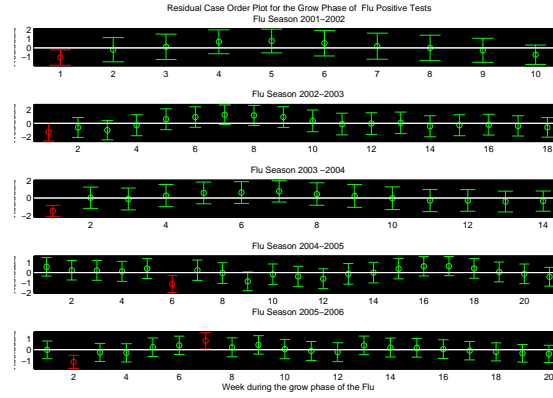


Figure 5: Errorbar plot of confidence intervals on residuals from regression for growth rates of flu positive tests

Because $R_t = 1 + \frac{1}{\hat{\gamma}}(\log z(t+1) - \log z(t))$, where $z(t)$ is the weekly number of reported flu positive tests, the effective reproduction numbers in the growing phases can be estimated by $\hat{R}_t = 1 + \frac{1}{\hat{\gamma}}\hat{\beta}$. The $\hat{\gamma}$ is determined by infectious period of influenza. The period is thought to be an average of four and a half days (Longini *et al.* (2004)), that is, 0.6 week, we have $1/\hat{\gamma} = 0.6$. It is seen that the R_t is free of time variable t in the growing phase of each flu season. The estimated reproduction number in the growing phase of each flu season from 2001 to 2006 are shown in the Table 3.

Table 3: Estimates of effective reproduction numbers in the growing phases of flu season from 2001 to 2006

Flu season	\hat{R}	CI
2001-2002	1.4304	(1.3207 1.5401)
2002-2003	1.1627	(1.1104 1.2151)
2003-2004	1.2297	(1.1685 1.2910)
2004-2005	1.2456	(1.2191 1.2722)
2005-2006	1.1889	(1.1660 1.2117)

4. Discussion

In summary, a method is given to estimate the effective reproduction number of influenza in the growing phase of a flu season. The effective reproduction number is the average number of new infectious cases infected by an index case in the susceptible population during the growing phase of the flu season. However, it is known that not everyone can be susceptible, say, because they can be treated prophylactically with antiviral drug or (if available) a vaccine. Thus the effective reproductive number will be less than basic reproduction number R_0 . Also, it is clear that not everyone is equally likely to be in contact with the index case, not everyone has the same duration of infectiousness. Therefore, the models discussed in this article belongs to the “simple but tractable” variety which average over demographic structure and social mixing patterns. The aim of this study is to find methods, which can be rapidly applied (say on a daily basis) in an emerging outbreak situation, to inform policy.

Acknowledgements

The author thanks Dr. P. Yan and B. Winchester from Public Health Agency of Canada for their helps and the author also acknowledges reviewers for suggestions on the manuscript. This article only represents the opinion of the author and it does not present any official views of an organization.

References

- Anderson, R. M. and May, R. M. (1979). Population biology of infectious diseases: Part I. *Nature* **280**, 361-367.
- Anderson, R. M. and May, R. M. (1991). *Infectious Diseases of Humans: Dynamics and Control*. Oxford University Press.
- Belshe, R. B. (2009). Implications of the Emergence of a Novel H1 Influenza Virus. *N. Engl. J. Med.* **360**, 2667-2668.
- Chowell, G., Nishiura, G. H. and Bettencourt, L. M. A. (2007). Comparative estimation of the reproduction number for pandemic influenza from daily case notification data. *J. R. Soc. Interface* **4**, 155-166.
- Cohen, J. and Enserink, M. (2009). After Delays, WHO Agrees: The 2009 Pandemic Has Begun. *Science* **324**, 1496-1497.
- Deguen, S., Thoms, G. and Chau, N. P. (2000). Estimating of the contact rate in a seasonal SEIR model: application to chickenpox incidence in France. *Statistics in Medicine* **19** 1207-1216.

- Diekmann, O., Heesterbeek, J. and Metz, J. (1998). On the definition and computation of the basic reproduction ratio R_0 in models for infectious diseases in heterogeneous populations. *J. Math. Biol.* **28**, 365-382.
- Diekmann, O. and Heesterbeek, J. (2000). *Mathematical Epidemiology of Infectious Disease; Model building, Analysis and Interpretation*. Wiley.
- Dolin, R. (2005). Influenza inter-pandemic as well as pandemic disease. *N. Engl. J. Med.* **353**, 2535-2537.
- Fraser, C. E. A. (2009). Pandemic Potential of a Strain of Influenza A (H1N1): Early Findings. *Nature* **324**, 1557-1561.
- Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S. and Brilliant, L. (2009). Detecting influenza epidemics using search engine query data. *Nature* **457**, 1012-1014.
- Laura, T. and Guy, T. (2006). Estimation of balanced simultaneous confidence sets for SIR models. *Communications in Statistics: Simulation and Computation* **35**, 803-812.
- Li, M. Y. and Muldowney, J. (1995). Global stability for the SEIR model in epidemiology. *Math. Biosci.* **125**, 155-164.
- Longini, I. J., Halloran, M. E., Nizam, A. and Yang, Y. (2004). Containing pandemic influenza with antiviral agents. *Am. J. Epidemiol.* **159**, 623-633.
- PHAC (2009a). Public Health Agency of Canada. <http://www.phac-aspc.gc.ca/fluwatch/index-eng.php> accessed 24 June 2009.
- PHAC (2009b). Public Health Agency of Canada: Influenza Immunization. <http://www.phac-aspc.gc.ca/im/iif-vcg/index-eng.php> accessed 18 September 2009.
- Stone, L., Olink, R. and Huppert, A. (2007). Seasonal dynamics of recurrent epidemics. *Nature* **446**, 533-536.
- Team, (2009). Novel Swine-Origin Influenza A (H1N1) Virus Investigation Team: Emergence of a Novel Swine-Origin Influenza A (H1N1) Virus in Humans. *N. Engl. J. Med.* **360**, 2605-2615.
- Vynnycky, E., Trindall, A. and Mangtani, P. (2007). Estimates of the reproduction numbers of Spanish influenza using morbidity data. *International Journal of Epidemiology* **36**, 881-889.
- WHO, (2009). World Health Organization: Influenza A(H1N1) press briefings. <http://www.who.int/mediacentre/influenzaAH1N1-presstranscript-20090611> accessed 19 June 2009.

S. Zhang
Centre for Communicable Diseases and Infection Control
Public Health Agency of Canada
100 Eglantine Driveway, Ottawa, Canada K1A 0K9
shzhang@hc-sc.gc.ca