Time-dependent Transmission of the 2009 H1N1 Virus in the Southern Hemisphere

Mengjiao Wang, Yiwei Gong, Zhenkun Tao

Abstract

Background

Seasonal influenza has a higher infection rate in winter.

Objective

To describe how H1N1 influenza varied in the Southern Hemisphere from April to November 2009 and examine the seasonality of the infection rate.

Methods

We used and modified Euler's methods to find a closed form of the SIR model and applied the maximum likelihood estimation (MLE) method to determine the parameters in the solution. To examine seasonality, we compared the constant infection rate case, the linear time-dependent infection rate case and the quadratic time-dependent infection rate case, using MLE and Generalised Likelihood Ratio Test.

Results

There is strong evidence (p-value= 5.0099×10^{-37}) suggesting that the infection rate α is a quadratic function of time.

Conclusions

The infection rate of H1N1 influenza is highly likely to be seasonal. This time-varying pattern of the influenza suggests seasonality need to be taken into consideration for flu preparedness planning.

1. Introduction

The H1N1 virus is the subtype of influenza A virus that was the most common cause of human influenza (flu) in 2009. It was first identified in April 2009 in Mexico and California[1][2], and became so pandemic that about 17,000 deaths were reported by the start of 2010[3]. As this virus is highly infectious, it is essential to develop strategies to mitigate and control this threat. Mathematical modelling of an epidemic has an important role in understanding the various complexities associated with an infectious disease and its control. This report is based on laboratory confirmed H1N1 cases in the Southern Hemisphere. With available data, a S-I-R (susceptible–infected–recovered) model and MLE (maximum likelihood estimation) methods have been applied to simulate the mechanisms underlying observed epidemiological patterns. A detailed analysis of the transmission among the susceptible, the infected and the recovered group—especially the first two groups, i.e. the infection rate—has been undertaken.

2. Materials and Method

2.1 Mathematical Model

The S-I-R model was developed by Kermack and McKendrick in 1927[4]. The model assumes that when an infectious disease strikes a community, the total population can be divided to three different classes: S, who are at risk of infection, I, who are infected and R, who have been removed from the first two

groups, either through recovery, death, immunity or isolation. In this paper, only the transmission between classes S and I has been analysed, thus the model is of the form

$$dS/dt = -\alpha SI \qquad S(0) = 1 \qquad (1)$$

$$dI/dt = \alpha SI - \beta I \qquad I(0) = I_0 \qquad (2)$$

Generally, by Euler's method, (1) and (2) can be derived into the form

$$S_{t+1} = S_t - \alpha S_t I_t \qquad (3)$$

$$I_{t+1} = I_t + \alpha S_t I_t - \beta I_t \qquad (4)$$

The equations can produce negative solutions, but in the current case S_t and I_t must be positive. Hence, it is better to approximate the dynamics over Δt using an approximating exponential growth or decline model. By applying this approach to (1) and (2) separately, a pair of discrete time equations has been obtained

$$S_{t+1} = S_t e^{-\alpha I_t} \qquad S(0) = 1 \qquad (7)$$

$$I_{t+1} = I_t e^{\alpha S_t} + I_t e^{-\beta} \qquad I(0) = I_0 \qquad (8)$$

When dealing with equation (2), it is of critical importance to treat αSI and $-\beta I$ separately, as the former determines the inflow of individuals from S to I, while the latter determines the outflow of individuals from I to R. Since the number of individuals leaving the susceptible class should match the number entering the infective class

$$S_t - S_{t+1} = I_t e^{\alpha S_t} \qquad (9)$$

The final discretisation has been obtained

$$S_{t+1} = S_t e^{-\alpha I_t} \qquad S(0) = 1$$

$$I_{t+1} = S_t(1 - e^{-\alpha I_t}) + I_t e^{-\beta} \qquad I(0) = I_0$$

The model has become more mature than one approximated through Euler's method, as it now can be fitted in continuous time which is in line with reality.

2.2 Distributional Assumption

The Poisson distribution is a discrete probability distribution for the counts of events that occur randomly in a given interval of time (or space). Therefore, the number of the laboratory confirmed cases in each week, y_t, can be assumed to be observations of independent $Poi(\mu_t)$ random variables. Moreover, given the information that the expected number of lab confirmed cases is proportional to the actual number of cases each week, $E(y_t) = \phi I(t)$, it is reasonable to assume that

$$y_t \sim Poisson(\phi I(t))$$

where $I(t)$ denotes the weekly predicted case and ϕ denotes some unknown parameter relating to the sizes of y_t and $I(t)$.

2.3 Maximum Likelihood Estimation

Given the model specification, the Poisson log likelihood function of parameter $\theta = \log(\phi, \alpha, \beta, I_0)^T$ can be constructed, where log parameterization ensures positive $\phi, \alpha, \beta,$ and I_0. Now the Quasi-Newton Method can be applied to minimise the negative log likelihood, which is identical to maximising the log likelihood.

From the plot of data, we can find that the number of confirmed cases first increases with time, indicating a larger rate of transmission than of recovery. Moreover, ϕ should be large enough to

compensate for the scaled number of diseased and infective individuals, $I(t)$, while I_0 should be small enough to fit y_0. Therefore, starting from an initial guess, $\theta^{[0]} = (8, 1, -1, -6)^T$, the optimized parameter values were returned as

$$\hat{\phi} = 6015.3760 \quad \widehat{\alpha_1} = 1.1715 \quad \hat{\beta} = 0.6383 \quad \widehat{I_0} = 0.0023$$

with the likelihood of $e^{-447.4638}$. **Figure a** overlays the curve of fitted numbers of cases against week over the raw data, and the model looks reasonable.

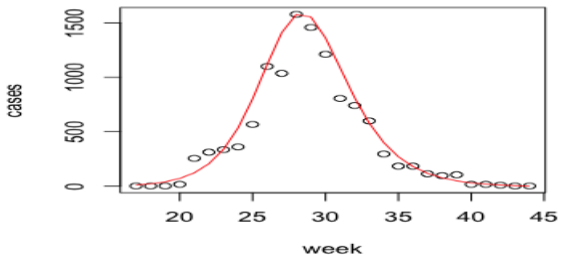

Figure a Data on number of cases against week. The symbols show raw data, while the red curve is the best fit model.

2.4 Model Checking

Before investigating the estimates further, it is important to check that the model assumptions are plausible. In the case of Poisson model, deviance residuals can be used for model checking. Deviance residuals are defined as

$$e_t^d = sign(e_t)\sqrt{d_t} \quad \text{for all t}$$

where $e_t = y_t - \widehat{\mu_t}$ are the raw residuals, and $d_t = 2\{l(y_t|y_t) - l(\widehat{\mu_t}|y_t)\}$ are the components of the deviance contributed by the t^{th} observation. For a well-fitting model, a single deviance residual should behave like a standard normal variable. The plot of deviance residuals against fitted values and the normal QQ-plot are shown in **Figure b**. The results suggest that the distributional assumption of the Poisson is rational.

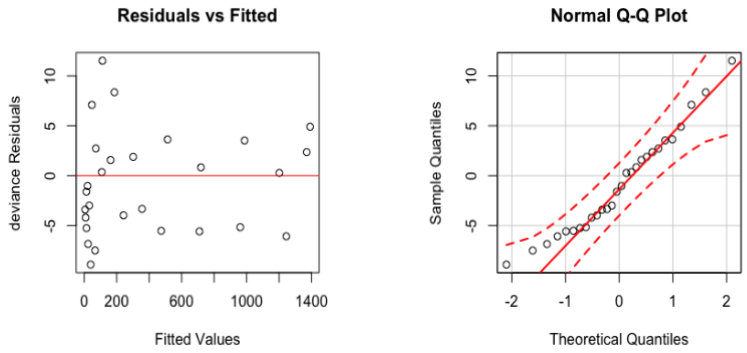

Figure b Checking plots

Left: deviance residuals against fitted values. There is no special pattern in the mean or variance. Right: normal QQ-plot of deviance residuals. It is close enough to a straight line to follow normal distribution.

3. Modified Model: Time-varying infective rate α

The previous result has been deduced based on a constant infective rate and Figure a suggests these maximum likelihood estimators regress the data well. It is possible for the infection rate to change according to time. Intuitively, people are more likely to be infected by the influenza virus in winter for many reasons, such as a weaker immune system in low temperatures. It could be examined whether this model will be improved by using a time-varying parameter α. This session will discuss two plausible forms of the infectious rate α, one from a linear function of time and the other from a quadratic function.

3.1 Modified Model I: $\alpha_2 = a_2 t + c_2$

In this case, the infective rate has been modelled as a linear function of time, denoted as α_2, where a_2 and c_2 are parameters to be estimated. The method for estimation here is again the maximum likelihood estimation and the result suggests that $\widehat{a_2}$ is approximately -0.0320 and $\widehat{c_2}$ approximately 1.3964. The associated likelihood is $e^{-441.5251}$.

3.2 Modified Model II: $\alpha_3 = a_3 t + b_3 t^2 + c_3$

In the second case, a quadratic function of time has been used to model the time-varying infection rate, which is denoted as α_3, and a_3, b_3 and c_3 are parameters to be estimated. Again, the use of MLE suggests that $\widehat{a_3} \approx -0.9713$, $\widehat{b_3} \approx 0.0838$ and $\widehat{c_3} \approx 3.5066$. The associated likelihood is $e^{-360.7154}$.

3.3 Model Comparison: constant, linear time-varying and quadratic time-varying infective rates

This section will compare the three models with different infective rates using likelihood, a graph, and generalised likelihood ratio test.

Using R, the likelihoods were calculated for all three models, which are $e^{-447.4638}$, $e^{-441.5251}$ and $e^{-360.7154}$ for the constant model, linear model and quadratic model, respectively. It is clear that the quadratic model generates the largest likelihood, followed by the linear model, and finally the constant model. This result is also illustrated in **Figure c**, where the quadratic model (the black curve) has the best fit, especially at the peak. As can be seen from the graph, Model I fitted a curve (the red line) in a similar manner to the previous result (the green line), both of which fit the data well at the tails but not so satisfactorily at the peak. A generalised likelihood ratio test has been performed comparing the two models, and it suggests there is strong evidence (p=0.00057) to believe that the infection rate is a function of time.

However, the quadratic time-dependent model clearly fits the data better than the other two options, especially at the peak. Another generalised likelihood ratio test was performed, comparing the quadratic case with the linear case. This yielded a p-value of 5.0099×10^{-37}, a relatively small value, so there is very strong evidence against the linear model. In addition, the Akaike Information Criterion has been used to examine the three models and the results (902.9277, 893.0502 and 733.4308, respectively) suggest that the quadratic infection rate, α_3, best balances the fitting result and the complexity. Therefore, the quadratic time-varying infection rate α_3 is more likely for the proposed model with better fit. The estimated infective rate $\widehat{\alpha_3}$ is shown in **Figure d**.

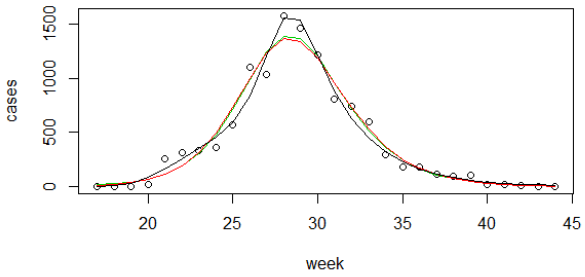

Figure c fitted curve by $\widehat{\alpha_1}$, $\widehat{\alpha_2}$, $\widehat{\alpha_3}$

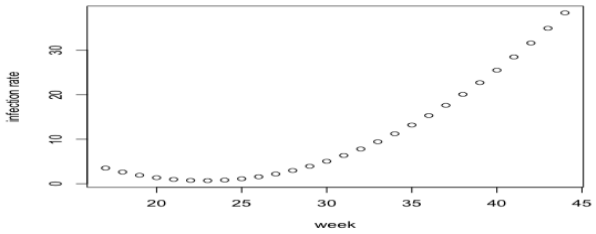

Figure d $\widehat{\alpha_3} = -0.9713(t-17) + 0.0838(t-17)^2 + 3.5066$

4. Conclusion

In this report, a SIR model has been proposed and solved. In addition, a maximum likelihood estimation has been used to find possible parameters. This was followed by model modification to confirm the existence of the time-varying infection rate. There is significant evidence showing that the infection rate is a quadratic function of time, which increases in summer and decreases in winter. This result suggests this influenza has a different seasonal pattern to common seasonal influenza, since the latter is most likely to have higher infection rate in low temperatures. This H1N1 virus had a varied influence on different age groups, attacking more children and young people than seniors, unlike seasonal influenza[5]. This might explain this unusual seasonal pattern of H1N1, or this seasonality might not be a reflection of weakened immune systems due to low temperatures. More evidence and experiments should be done to confirm this seasonality and examine the reasons for it. In addition, this unusual seasonal pattern of H1N1's infection rate should be taken into account for pandemic planning, such as vaccine and antiviral medication preparedness.

References

[1] Dawood FS, et al. (2009). Emergence of a novel swine-origin influenza A (H1N1) virus in humans. [Online]. Available: https://www.ncbi.nlm.nih.gov/pubmed/19423869. [Accessed 16-Apr-2018].

[2] Outbreak of Swine-Origin Influenza A (H1N1) Virus Infection -Mexico, March–April 2009 (2009) Morb Mortal Wkly Rep 58: 467–470.

[3] WHO. (2010). *Pandemic (H1N1) 2009 - update 100.* Available: http://www.who.int/csr/don/2010_05_14/en/. [Accessed: 19-Apr-2018].

[4] Kermack, W and McKendrick, A (1991). "Contributions to the mathematical theory of epidemics—I". *Bulletin of Mathematical Biology*. **53** (1–2): 33–55. doi:10.1007/BF02464423. PMID 2059741

[5] Roos, R (2013). "Study puts global 2009 pandemic H1N1 infection rate at 24%," Center for Infectious Disease Research and Policy, 24-Jan-2013. [Online]. Available: http://www.cidrap.umn.edu/news-perspective/2013/01/study-puts-global-2009-pandemic-h1n1-infection-rate-24. [Accessed: 19-Apr-2018].