A novel algorithm to define infection tendencies in H1N1 cases in Mainland China

Fan Ding  
*College of Veterinary Medicine, Northeast Agricultural University & Chinese Armed Police Force Classification Command College, Beijing*

Dante S. Zarlenga  
*Agricultural Research Service, U.S. Department of Agriculture, Beltsville*

Chengfeng Qin  
*State Key Laboratory of Pathogen and Biosecurity, Beijing Institute of Microbiology and Epidemiology*

Follow this and additional works at: [https://digitalcommons.unl.edu/usdaarsfacpub](https://digitalcommons.unl.edu/usdaarsfacpub)

Ding, Fan; Zarlenga, Dante S.; and Qin, Chengfeng, "A novel algorithm to define infection tendencies in H1N1 cases in Mainland China" (2010). *Publications from USDA-ARS / UNL Faculty*. 2227. [https://digitalcommons.unl.edu/usdaarsfacpub/2227](https://digitalcommons.unl.edu/usdaarsfacpub/2227)

This Article is brought to you for free and open access by the U.S. Department of Agriculture: Agricultural Research Service, Lincoln, Nebraska at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Publications from USDA-ARS / UNL Faculty by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.
A novel algorithm to define infection tendencies in H1N1 cases in Mainland China

Fan Ding a,1, Dante S. Zarlenga b, Chengfeng Qin c, Xiaofeng Ren a,∗

a Microbiology Lab, Department of Preventive Veterinary Medicine, College of Veterinary Medicine, Northeast Agricultural University, 59 Mucai Street, Xiangfang District, 150030 Harbin, China
b Animal Parasitic Diseases Laboratory and Bovine Functional Genomics Laboratory, Agricultural Research Service, U.S. Department of Agriculture, Beltsville, MD 20705, USA
c State Key Laboratory of Pathogen and Biosecurity, Beijing Institute of Microbiology and Epidemiology, Beijing 100071, China

ARTICLE INFO

Article history:
Received 28 April 2010
Received in revised form 24 September 2010
Accepted 27 September 2010
Available online 15 October 2010

Keywords:
H1N1
Mathematical model
Grey prediction theory
Infection

ABSTRACT

Incidences of H1N1 viral infections in Mainland China are collected by the Ministry of Health, the People’s Republic of China. The number of confirmed cases and the timing of these outbreaks from May 13 to July 22, 2009 were obtained and subjected to a novel mathematical model to simulate the infection profile (time vs number). The model was predicated upon the grey prediction theory which allows assignment of future trends using limited numbers of data points. During the period of our analysis, the number of confirmed H1N1 cases in Mainland China increased from 1 to 1772. The efficiency of our model to simulate these data points was evaluated using Sum of squares of error (SSE), Relative standard error (RSE), Mean absolute deviation (MAD) and Average relative error (ARE). Results from these analyses were compared to similar calculations based upon the grey prediction algorithm. Using our equation, defined herein as equation D–R, results showed that SSE = 6742.00, RSE = 10.69, MAD = 7.07, ARE = 2.47% were all consistent with the D–R algorithm performing well in the estimation of future trends of H1N1 cases in Mainland China. Calculations using the grey theory had no predictive value [ARE for GM(1,1) = −104.63%]. To validate this algorithm, we performed a second analysis using new data obtained from cases reported to the WHO and CDC in the US between April 26 and June 8, 2009. In like manner, the model was equally predictive. The success of the D–R mathematical model suggests that it may have broader application to other viral infections among the human population in China and may be modified for application to other regions of the world.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

Since the World Health Organization (WHO) first identified H1N1 influenza outbreaks in the United States and Mexico in April 2009, the worldwide incidence of this disease has risen dramatically. This infection has had substantial impact on the politics, economies and public health in now endemic regions of the world, so much so that in June 2009, the WHO raised the pandemic alert level to phase 6 (maximum) (http://www.who.int/mediacentre/news/statements/2009/h1n1_pandemic_phase6_20090611/en/index.html). This action prompted many countries to adopt counter-measures such as strengthening prevention and control, perfecting surveillance, and accelerating vaccine development. In Mainland, China, the government and related agencies took rapid and effective measures to curtail the infection. As such, the number of confirmed H1N1 cases in Mainland China increased only moderately. Nevertheless, a better model is needed to predict infection trends of H1N1 and other potential viral infections of humans in order to advance effective prevention and control practices. Precedence exists for such modeling. Several years ago, mathematical models were used to simulate the incidence of Severe Acute Respiratory Syndrome, SARS (Riley et al., 2003; Lipsitch et al., 2003; Dye and Gay, 2003).

Currently there exists established mathematical methods describing the prevalence of diseases over time such as the Time Series Analysis (TSA) which relies on ARMA (Autoregressive Moving Average) and ARIMA (Autoregressive Integrated Moving Average) models extensively (Harris and Sollis, 2003; Yu, 2005; Bell et al., 2004; Campbell et al., 2009). The disadvantage of the TSA algorithm is that the sequence data for TSA must be complete and some unrelated factors need to be excluded. This in turn can result in the lack of qualified data for testing. For this reason, i.e. incomplete data, TSA was considered not be suitable for predicting future trends in H1N1 infections. The grey prediction model overcomes this shortcoming to some extent, because it does not require complete data sequences in order to be effective (Wang et al., 2004; Campbell et al., 2009).
al., 2007; Xiong and Xu, 2005). As such, we used the grey prediction model \([GM(1,1)]\) as a basis for constructing the algorithm defined herein. In the case of H1N1, many researchers have analyzed the organism's genetics (McDonald et al., 2007), pathology (Tang and Chong, 2009), evolution and phylogeny (Dunham et al., 2009), mechanism of infection and transmission (Lange et al., 2009), and resistance to drugs (Deyde et al., 2007) to help prevent a pandemic. Here we establish the mathematic model D-R using limited data points, and fit the H1N1 epidemic in Mainland China as a test, then compare this model to the predictive capability of \([GM(1,1)]\) to show the enhanced predictive value of the new equation. As an additional proof of principle, the validity of the developed model was secondarily demonstrated using data obtained from the US. We trust that such a model will assist in designing effective surveillance strategies for this and other viral diseases in order to determine a trend line for future infections and also determine when and to what extent the infection rate deviates from the norm.

The number of confirmed H1N1 cases in Mainland China is accessible from the official website of the Ministry of Health, the People’s Republic of China (http://www.moh.gov.cn/publicfiles/business/htmlfiles/web/index.htm). It should be noted that these numbers reflect only confirmed and reported cases. The number of actual cases may be substantially higher. In addition, the number of actual cases and reported cases may differ more in the exponentially increasing period of the epidemic as compared to the start of the epidemic. However, given that these anomalies are partitioned into nearly all public databases of this type and represent an unknown, this factor was not incorporated into the model. The number of the H1N1 cases between May 13 and July 22 was collected and used in this study. During this period, the data on July 8, 10, 12, 14, 16, 18, 19 and 21 were not publicly available. Our model was substantially based upon the grey prediction model then modified using components of the temporal series’ Sliding Average Method (SAM), Weighted Average Method (WAM), ARMA and ARIMA (Harris and Sollis, 2003; Yu, 2005), and statistical regression analysis. The mathematical expression representing increasing tendency of H1N1 cases is given below:

\[
A(t) = A_0 + \alpha_1(K(A_1 - A_2) + K(1-K))(A_2 - A_3) + K(1-K)(A_3 - A_4) + \ldots + K(1-K)^{(n-1)}(A_n - A_{n+1}) + \beta \left[ \sum_{i=1}^{n} \left( A_{n-i} - A_{n-i+1} \right) \right]
\]

At \(t = 1\) or \(t = 2\) the number of cases is chosen, and \(n\) is the number of weeks. Table 1 shows the actual values of H1N1 infections at time \(t\) that were used to calculate predicted values using equation \(D-R\), the upper \((U)\) and lower \((L)\) limits, and predicted values using \([GM(1,1)]\) for the period May 17 to July 22. No data was available for the time periods defined by ‘–’. This table presents the actual values of H1N1 infections at time \(t\) (\(A_t\)) that were used to calculate predicted values using equation \(D-R\), the upper \((U)\) and lower \((L)\) limits, and predicted values using \([GM(1,1)]\) for the period May 17 to July 22. No data was available for the time periods defined by ‘–’.

| Date       | \(A_t\) | \(D-R\) U | \(L\) | \([GM(1,1)]\) |
|------------|---------|-----------|------|----------------|
| May 17     | 3.00    | 2.28      | 2.16 | 2.39           |
| May 18     | 3.00    | 3.62      | 3.44 | 3.80           |
| May 19     | 4.00    | 3.31      | 3.15 | 3.48           |
| May 20     | 4.00    | 4.62      | 4.39 | 4.85           |
| May 21     | 5.00    | 4.33      | 4.11 | 4.54           |
| May 22     | 5.00    | 5.62      | 5.34 | 5.90           |
| May 23     | 7.00    | 5.33      | 5.07 | 5.60           |
| May 24     | 7.00    | 8.00      | 7.60 | 8.40           |
| May 25     | 11.00   | 7.49      | 7.12 | 7.87           |
| May 26     | 12.00   | 12.82     | 12.18| 13.47          |
| May 27     | 12.00   | 13.27     | 12.61| 13.94          |
| May 28     | 13.00   | 12.63     | 11.99| 13.26          |
| May 29     | 21.00   | 13.76     | 13.08| 14.45          |
| May 30     | 21.00   | 24.83     | 23.59| 26.08          |
| May 31     | 26.00   | 25.28     | 21.45| 23.71          |

\(A_t\) is the actual value at time \(t\), \(D-R\) is the upper \((U)\) and lower \((L)\) limits, and \([GM(1,1)]\) is the predicted value using the grey prediction model then modified using components of the temporal series’ Sliding Average Method (SAM), Weighted Average Method (WAM), ARMA and ARIMA (Harris and Sollis, 2003; Yu, 2005), and statistical regression analysis. The mathematical expression representing increasing tendency of H1N1 cases is given below:

\[
A_{t+1} = A_t + \alpha_1(K(A_t - A_{t+2}) + K(1-K)(A_{t+2} - A_{t+3}) + K(1-K)^2(A_{t+3} - A_{t+4}) + \ldots + K(1-K)^{(n-1)}(A_n - A_{n+1}) + \beta \left[ \sum_{i=1}^{n} \left( A_{n-i} - A_{n-i+1} \right) \right]
\]

At \(t = 1\) or \(t = 2\) the number of cases is chosen, and \(n\) is the number of weeks. Table 1 shows the actual values of H1N1 infections at time \(t\) (\(A_t\)) that were used to calculate predicted values using equation \(D-R\), the upper \((U)\) and lower \((L)\) limits, and predicted values using \([GM(1,1)]\) for the period May 17 to July 22. No data was available for the time periods defined by ‘–’. This table presents the actual values of H1N1 infections at time \(t\) (\(A_t\)) that were used to calculate predicted values using equation \(D-R\), the upper \((U)\) and lower \((L)\) limits, and predicted values using \([GM(1,1)]\) for the period May 17 to July 22. No data was available for the time periods defined by ‘–’.
The percent difference between the actual and fitted curves in the mathematical model suggests that the accuracy of the fitted model is very high. The percent difference between the actual values and the fitted values. Using equation D–R, we calculated MAD to be 7.07 vs. 901.72 for GM(1,1). In general, MAD and RSE reflect the mean differences between actual and fitted curves. Using large datasets, these differences become insignificant suggesting that the predictive value of the D–R model is very high. The percent difference between the actual and fitted curves in actual value is defined by ARE. The 2.47% ARE using the D–R equation suggests that the accuracy of the fitted model is approximately 97%; the number (−104.63%) obtained for GM(1,1) is uninformative.

As shown in Table 1 and pictorially demonstrated in Fig. 1, to better display the trend lines and compared datasets, the confirmed H1N1 cases during May 17 to July 22 and those predicted by the model are very similar, with the actual curve appearing higher than the upper limit of the fitted curve at some points. After self-adapting, variation in the fitted curve realigns with the limit values. Self-adaptation is one of the key features of the D–R mathematical model where implementation of the upper and lower limits occurs when the actual values approach or exceed these limits. When the upper limit is exceeded, the infection trend is greater than the norm implying the epidemic is gaining strength. If the actual curve and fitted curve are similar, it suggests there are no significant changes in the epidemic situation. If however, the actual curve shows a gradual increase, the RSE value is only 10.69, indicating that the fitted model is very accurate and far more predictive than that calculated using GM(1,1) where the RSE = 1482.04. The average total absolute value between actual and fitted data is described by MAD. Using equation D–R, we calculated MAD to be 7.07 vs. 901.72 for GM(1,1). In general, MAD and RSE reflect the mean differences between actual and fitted curves. Using large datasets, these differences become insignificant suggesting that the predictive value of the D–R mathematical model is very high.
actual curve approaches the lower limit, it suggests the epidemic is abating. Our results show that the fitted curve derived from the model closely mirrors the actual number of H1N1 cases, indicating the utility of the model. Upon comparing the D–R and GM(1,1) algorithms the data generated using GM(1,1) are initially in line with those using D–R; however, as the dataset increases, substantial deviation occurs in the capabilities of GM(1,1) to mimic the known progression of disease. This loss of predictive value is mirrored in the calculated values for SSE, RSE, MAD and ARE.

To further validate the fitting accuracy and utility of the D–R model, the daily numbers of reported H1N1 cases in the US during the period April 26 to June 8, 2009 were collected from WHO and CDC. The fitted curve was generated using data obtained between April 30 and June 8; the period April 26–29 was used as the minimum dataset upon which to begin building the simulation model and is therefore not presented in Fig. 2. The initial data were small i.e. on April 26, there were only 20 reported cases; however, by June 8, there were 13,217 reported cases. Beginning June 8, data were provided weekly rather than daily and therefore not used in our analysis. As with the data from Mainland China, fitting was performed by means of the D–R and GM(1,1) algorithms. Calculated values for SSE, RES, MAD and ARE from both the D–R and GM(1,1) models are included in Table 2 (note: tabulated data points from the WHO and CDC and calculated values for the period April 26 to June 8 used to construct Fig. 2 are available upon request). Values calculated from the D–R model were significantly lower than those from GM(1,1) model. For example, the D–R model ARE value is 2.93% which means that the fitting accuracy reached 97.07% (high precision). In contrast, the accuracy of GM(1,1) was only 39.69%, indicating substantial deviation in the predictive value. As with the data from Mainland China, Fig. 2 demonstrates that the actual curve and D–R fitted curve exhibited similar trends and even overlapped at some time points. In contrast, GM(1,1) model was reasonably predictive during the initial phase of the reporting period similar to that observed with the data from Mainland China; however, beginning May 12, the GM(1,1) curve deviated significantly from the actual curve.

Others have modeled H1N1 virus transmission under specialized circumstances and in subpopulations of individuals. Fraser et al. (2009) were among the first to present a model of H1N1 transmission based upon data obtained from the outbreak in Mexico. They concluded that transmissibility is substantially higher than that of the seasonal flu, but comparable to previous influenza pandemics with respect to low Basic Reproduction Numbers (R0). Gojovic et al. (2009) generated a simulation model based upon combinatorial uncertainty analysis to project the effects of several strategies to mitigate transmission among Koreans and concluded that if available, massive vaccination would be optimal. Tracht et al. (2010) developed a transmission model predicated on a subpopulation (10%) of individuals that would be willing to correctly use facemasks and concluded a substantial reduction (20%) would ensue. Yet, among these and other extensive studies modeling the effects of mitigation, Coburn et al. (2009) concluded that trying to identify intervention strategies for epidemics that involve recombination of species-specific strains and cross-species transmission, i.e. H1N1 is problematic. The work herein used a subset of data to project future trends and then tested those trend lines against existing data and demonstrated good congruence. This model gives less consideration to micro-environmental factors by establishing upper and lower limits of the prediction intervals and providing self-adapting parameters to account for dominating long or short term effects. As with most models, assimilating trends to regional variation, health care resources and the public health measures to mitigate impact were not evaluated; however, our model is capable of assessing the benefits of intervention strategies.

Taken together, our novel algorithm aligns well with trends observed in the report of H1N1 cases in Mainland China and in the US. This model may not only be used to broadly predict trends in H1N1 cases, but also may be applicable for predicting other epidemics. Although the progression of epidemic diseases is often based upon several hypotheses (Gordis, 2008), for the first time we show that the infection rate of influenza H1N1 can be predicted by a mathematical model that depicts a relationship between the tendency of the data to change, and the fitted and limit values of that data.

Acknowledgements

This work was supported, in part, by the National 973 Plan of China (2010CB534002). We also acknowledge National Natural Science Foundation of China (30700590 and 30972195), Funding supported by Program for New Century Excellent Talents in Heilongjiang Provincial University (1155–NCT–005) and Heilongjiang Provincial Science and Technology Department, China (ZJN0702–01).

References

Bell, M.L., McDermott, A., Zeger, S.L., Samet, J.M., Dominici, F., 2004. Ozone and short-term mortality in 95 US Urban Communities, 1987–2000. JAMA 292, 2372–2378.
Campbell, S.M., Reeves, D., Kontopantelis, E., Sibbald, B., Roland, M., 2009. Effects of pay for performance on the quality of primary care in England. N. Engl. J. Med. 361, 368–378.
Coburn, B.J., Wagner, B.G., Blower, S., 2009. Modeling influenza epidemics and pandemics: insights into the future of swine flu (H1N1). BMC Med. 7, 30. doi:10.1186/1741–7015–3–70.
Deyde, V.M., Xu, X., Bright, R.A., Shaw, M., Smith, C.B., Zhang, Y., Shu, Y., Guabereva, LV., Cox, N.J., Klimov, A.I., 2007. Surveillance of resistance to adamantanes among influenza A(H3N2) and A(H1N1) viruses isolated worldwide. J. Infect. Dis. 196, 249–257.
Dunham, E.J., Dugan, V.G., Kaser, E.K., Sarah, E., Perkins, S.E., Brown, L.H., Holmes, E.C., Taubenberger, J.K., 2009. Different evolutionary trajectories of European Avian-Like and Classical Swine H1N1 Influenza A Viruses. J. Virol. 83, 5485–5494.
Dye, C., Gay, N., 2003. Modeling the SARS epidemic. Science 300, 1884–1885.
Fraser, C., Donnelly, C.A., Cauchemez, S., et al., 2009. Pandemic potential of a strain of influenza A (H1N1): early findings. Science 324, 1557–1561.
Gojovic, M.Z., Sander, B., Fisman, D., Krahn, M.D., Bauch, C.T., 2009. Modelling mitigation strategies for pandemic (H1N1). CMAJ 181, 673–680.
Gordis, L., 2008. Epidemiology: with STUDENT CONSULT Online Access, 4e. Saun-
ders, ISBN: 1416040021/9781416040026.
Harris, R., Solis, P., 2003. Applied Time Series Modeling and Forecasting, John Wiley & Sons Ltd.
Lange, E., Kalthoff, D., Blohm, U., Trstek, J.P., Breithaupt, A., Maresch, C., Starick, E., Feredouni, S., Hoffmann, B., Mettenleiter, T.C., Bee, M., Vahlenkamp, T.W., 2009. Pathogenesis and transmission of the novel swine-origin influenza virus A/H1N1 after experimental infection of pigs. J. Med. Virol. 81, 2119–2123 Epub2009 Jul 10.
Lipsitch, M., Cohen, T., Cooper, B., et al., 2003. Transmission dynamics and control of severe acute respiratory syndrome. Science 300, 1966–1970.
McDonald, N.J., Smith, C.B., Cox, N.J., 2007. Antigenic drift in the evolution of H1N1 influenza A viruses resulting from deletion of a single amino acid in the haemagglutinin gene. J. Gen. Virol. 88, 3209–3213.
Riley, S., Fraser, C., Donnelly, C.A., et al., 2003. Transmission dynamics of the etiological agent of SARS in Hong Kong: impact of public health interventions. Science 300, 1961–1966.

Table 2

| SSE | RSE | MAD | ARE |
|-----|-----|-----|-----|
| D–R | 7280042.99 | 484.60 | 338.51 | 2.93% |
| GM(1,1) | 92307653.18 | 5458.07 | 4171.35 | −61.31% |

Based on the H1N1 case number reported by the WHO and CDC in the US during the period April 30 to June 8, the SSE, RES, MAD and ARE values were calculated for the D–R and GM(1,1) models.
Tang, X., Chong, K.T., 2009. Histopathology and growth kinetics of influenza viruses (H1N1 and H3N2) in the upper and lower airways of guinea pigs H1N1. J. Gen. Virol. 90, 386–391.

Tracht, S.M., Del Valle, S.Y., Hyman, J.M., 2010. Mathematical modeling of the effectiveness of facemasks in reducing the spread of novel influenza A (H1N1). PLoS One 5 (February (2)), e9018.

Wang, Z.L., Xie, M.P., Wang, D.L., 2007. Application of gray dynamic forecasting model research in urban water use. Res. Soil Water Conserv. 14, 430.

World now at the start of 2009 influenza pandemic. Statement to the press by WHO Director-General Dr Margaret Chan. Geneva: World Health Organization; 11 June 2009. Available from: http://www.who.int/mediacentre/news/ statements/2009/h1n1_pandemic_phase6_20090611/en/index.html.

Xiong, H.J., Xu, H.Z., 2005. Grey control. National Defense Industry Press, ISBN: 7118041440 (in Chinese).

Pan, H., 2005. Time Series Analysis. University of International Business and Economics Press, Beijing, China, ISBN: 7-81078-567-2 (in Chinese).