



Article

Predicting Seasonal Influenza Based on SARIMA Model, in Mainland China from 2005 to 2018

Jing Cong ¹, Mengmeng Ren ¹, Shuyang Xie ² and Pingyu Wang ^{1,2,*}

¹ Department of Epidemiology, Binzhou Medical University, YanTai 264003, China; cjboy123@163.com (J.C.); 17853589692@163.com (M.R.)

² Department of Biochemistry and Molecular Biology, Binzhou Medical University, YanTai 264003, China; shuyangxie@aliyun.com

* Correspondence: wpingyugirl@163.com; Tel.: +86-535-691-3217

Received: 12 October 2019; Accepted: 25 November 2019; Published: 27 November 2019



Abstract: Seasonal influenza is one of the mandatorily monitored infectious diseases, in China. Making full use of the influenza surveillance data helps to predict seasonal influenza. In this study, a seasonal autoregressive integrated moving average (SARIMA) model was used to predict the influenza changes by analyzing monthly data of influenza incidence from January 2005 to December 2018, in China. The inter-annual incidence rate fluctuated from 2.76 to 55.07 per 100,000 individuals. The SARIMA (1, 0, 0) × (0, 1, 1) 12 model predicted that the influenza incidence in 2018 was similar to that of previous years, and it fitted the seasonal fluctuation. The relative errors between actual values and predicted values fluctuated from 0.0010 to 0.0137, which indicated that the predicted values matched the actual values well. This study demonstrated that the SARIMA model could effectively make short-term predictions of seasonal influenza.

Keywords: SARIMA model; influenza; prediction

1. Introduction

Influenza remains an increasing public health problem worldwide, especially since the 1918 influenza pandemic. Although many efforts have focused on measures and strategies to prevent and control influenza [1–4], it still results in significant mortality, health care capacity, and economic costs to society annually. A modeling study estimated that 291,243 to 645,832 seasonal influenza-associated respiratory deaths (4.0 to 8.8 per 100,000 individuals) occurred annually worldwide from 1999 to 2015 [5]. It estimated that 11.5% of lower respiratory tract infection episodes were attributable to influenza [6]. The expected annual losses from pandemic risk are about 0.6% of the global income [7]. The burden caused by influenza is higher in seasons dominated by A (H3N2) or A (H1N1) pdm 2009 influenza viruses but is lower in seasons where pre-pandemic A (H1N1) or influenza B accounts for the majority of cases [8]. In 2019, the WHO and partners launched the Global Influenza Strategy for 2019 to 2030 to strengthen seasonal prevention and control of future pandemics [9]. This strategy includes improving the influenza model and forecast. Therefore, it is important to use influenza surveillance data to model and forecast the influenza pandemics.

Time series analysis of infection data from analysis of specific models can improve the prevention system and forecast future values based on the previously observed values [10,11]. These models include seasonal autoregressive integrated moving average (SARIMA), neural network model, exponential smoothing, grey swing model, etc. [12–15]. The SARIMA model takes both overall trends and seasonal changes into account, which is widely used in modeling time series [13–15].

Our previous study demonstrated that there was a trend of increasing influenza incidence from 2005 to 2015, in mainland China [16]. China has a population of about 1.3 billion people, and the

disease burden caused by influenza is still heavy. The mortality caused by influenza is around 8% of all respiratory deaths in China [17]. Considering that seasonal influenza is one of the notifiable infectious diseases in China, it is helpful to model and forecast influenza by analyzing the influenza surveillance data. Furthermore, the country-specific estimate should be updated periodically, and a certain model should be developed. Therefore, the SARIMA model was performed to analyze the changes of influenza from recent surveillance data of influenza from January 2005 to December 2018, in China.

2. Materials

2.1. Data Collection

The total number of influenza cases and the monthly data of influenza incidence from January 2005 to December 2018 was provided by the website of the National Health Commission of the People's Republic of China [18].

The Chinese government has established a web-based national notifiable infectious disease surveillance system for 39 infectious diseases, including influenza since 2003. Clinicians complete a standard case report card when they identify any probable, clinical, or laboratory-confirmed case of seasonal influenza-A and influenza-B. Then, the local epidemiologists do a field investigation when they receive the disease card using a standardized form, which improves the data accuracy [16,19].

2.2. SARIMA Model

The data of influenza incidence from January 2005 to June 2018 was used as a training dataset, and that data from July 2018 to December 2018 was used as the forecasting dataset. We established and selected the best SARIMA model $(p, d, q) \times (P, D, Q)$ according to the steps introduced by Box and Jenkins [20] (Figure 1). Autoregressive lags, moving average lags, seasonal autoregressive lags, and seasonal moving average lags are indicated by $p, q, P,$ and $Q,$ respectively.

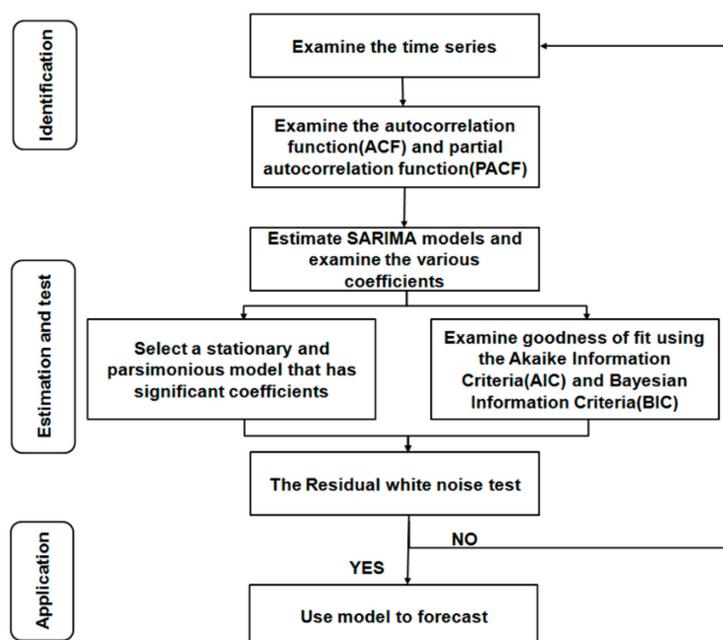


Figure 1. The process and method of seasonal autoregressive integrated moving average (SARIMA) model.

2.3. Statistical Analysis

STATA 15.0 (Stata Corp., College Station, TX, USA) and SPSS 22.0 (SPSS Inc., Chicago, IL, USA) were performed to create the SARIMA model.

3. Results

3.1. General Trend of Influenza Incidence

A total of 2,686,180 influenza cases were reported in mainland China from January 2005 to December 2018. The annual incidence rate fluctuated from 2.76 to 55.07 per 100,000 individuals. Influenza occurred throughout the year, with two peaks in winter and spring (Figure 2).

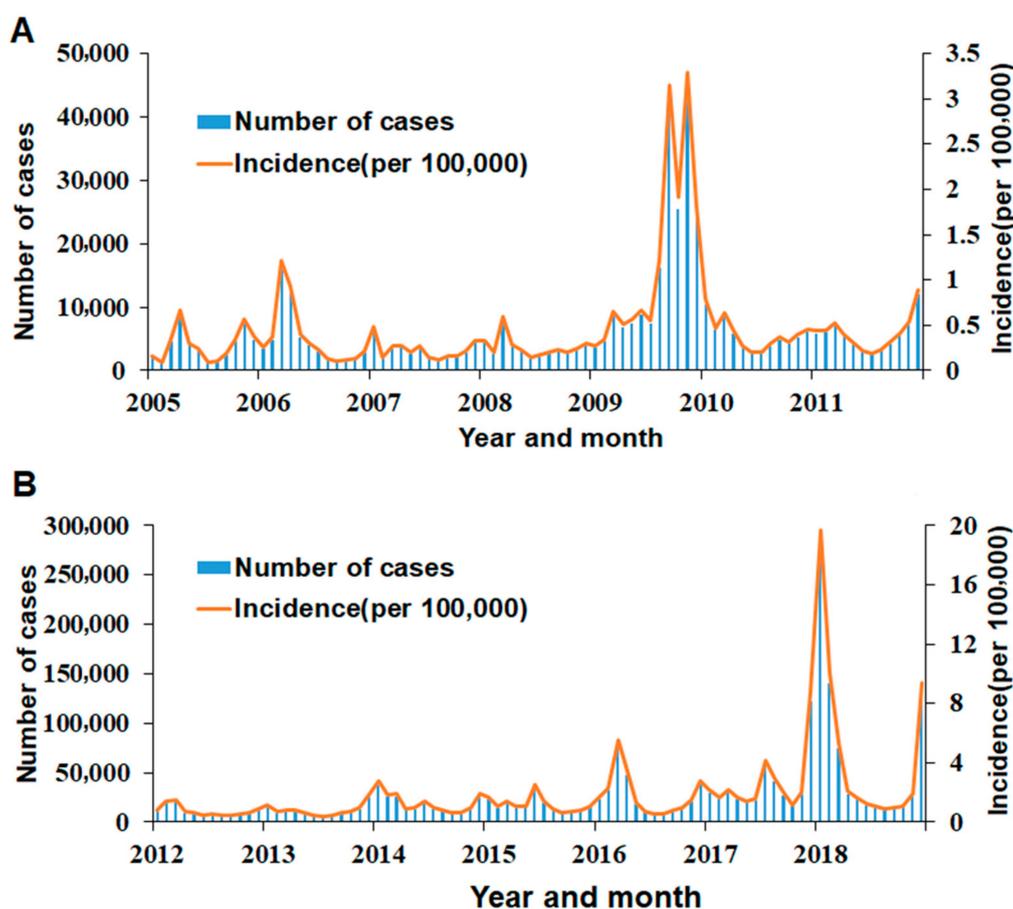


Figure 2. The influenza incidence in mainland China from 2005 to 2018: (A) Influenza cases and incidence from January 2005 to December 2011 and (B) influenza cases and incidence from January 2012 to December 2018.

3.2. SARIMA Model

Using the raw training data from January 2005 to June 2018, trend difference ($d = 0$) and seasonal difference ($D = 1$) were completed. The augmented Dickey–Fuller method was used to determine whether the sequence was stationary, and the result supported that the data was a stationary time series ($t = -9.247, p < 0.001$).

The auto-correlation Function (ACF) and partial correlation function (PACF) graphs were used to estimate the parameter ranges of p , P and q , Q (Figure 3). The ACF graph of one-order seasonal difference data (Figure 3C) and the PACF graph of one-order seasonal difference data (Figure 3D) showed better than the others. Then, some candidate SARIMA models were assessed to forecast future values based on the previously observed values (Table 1). On the basis of the results of the

goodness-of-fit test statistics, SARIMA (1, 0, 0) × (0, 1, 1) 12 model was found to be the optimal model, which had the lowest Akaike information criterion (AIC = 535.2955) and Bayesian information criterion (BIC = 544.3274). This model also passed the Ljung–Box Q Test ($z = 25.607, p = 0.060$). All the parameter estimates were significant (Table 1).

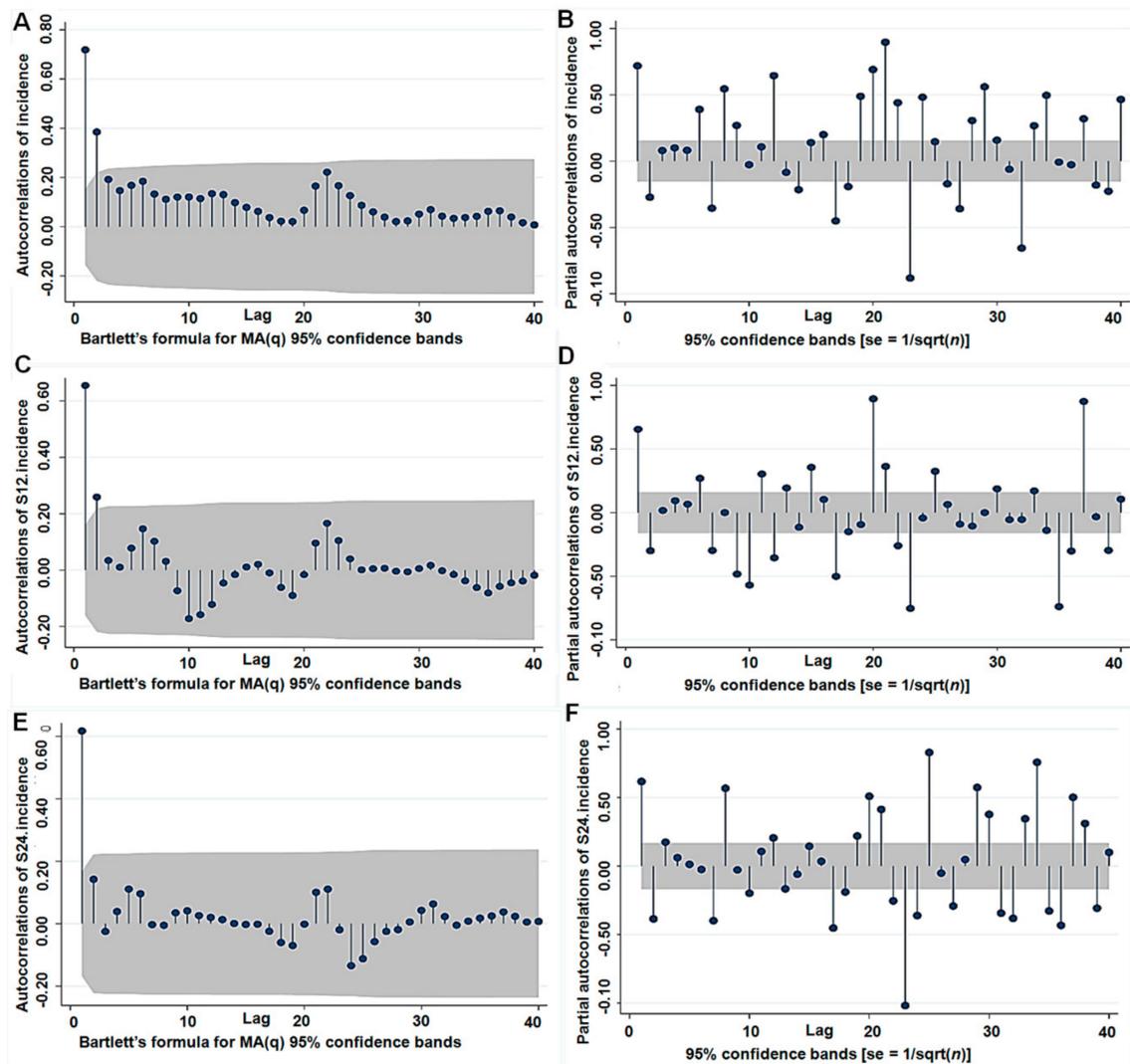


Figure 3. The ACF and PACF graphs for estimating the parameter: (A) The ACF graph of the raw data ($d = 0$ and $D = 0$), (B) the PACF graph of the raw data ($d = 0$ and $D = 0$), (C) the ACF graph of one-order seasonal difference data ($d = 0$ and $D = 1$), (D) the PACF graph of one-order seasonal difference data ($d = 0$ and $D = 1$), (E) the ACF graph of two-order seasonal difference data ($d = 0$ and $D = 2$), and (F) the PACF graph of two-order seasonal difference data ($d = 0$ and $D = 2$).

The model forecasting effect was tested by comparing predicted values with the actual values. The results showed SARIMA (1, 0, 0) × (0, 1, 1) 12 model fitted the seasonal fluctuation well (Figure 4). Then, the model was used to forecast the influenza incidence from July to December 2018. The relative errors between actual values and predicted values fluctuated from 0.0010 to 0.0137. All actual values are among 95% CI of predicted values (Table 2).

Table 1. Comparison of candidate SARIMA models.

Model	Estimate	Z	p-Value	Ljung-Box Q Test			AIC	BIC	RMSE	MAPE
				Statistics	DF	p-Value				
SARIMA (0,0,1) (0,1,1) ₁₂	-	-	-	22.753	16	0.121	541.661	550.692	1.439	48.744
q	0.654	11.00	0.000	-	-	-	-	-	-	-
Q	-0.415	-2.17	0.030	-	-	-	-	-	-	-
SARIMA (1,0,0) (0,1,1) ₁₂	-	-	-	25.607	16	0.060	535.296	544.327	1.407	44.280
p	0.668	25.68	0.000	-	-	-	-	-	-	-
Q	-0.445	-2.24	0.025	-	-	-	-	-	-	-
SARIMA (1,0,1) (0,1,1) ₁₂	-	-	-	8.157	15	0.917	523.172	535.214	1.345	44.137
p	0.481	3.15	0.002	-	-	-	-	-	-	-
q	-0.393	-3.772	0.074	-	-	-	-	-	-	-
Q	0.473	-2.53	0.012	-	-	-	-	-	-	-
SARIMA (1,0,1) (1,1,1) ₁₂	-	-	-	7.916	14	0.894	525.083	540.136	1.348	44.021
p	0.476	2.93	0.003	-	-	-	-	-	-	-
q	0.399	1.76	0.078	-	-	-	-	-	-	-
P	-0.080	-0.10	0.923	-	-	-	-	-	-	-
Q	-0.425	-0.50	0.615	-	-	-	-	-	-	-

AIC: Akaike information criterion; BIC: Bayesian information criterion; RMSE: root mean squared error; MAPE: mean absolute percent error; DF: degree of freedom.

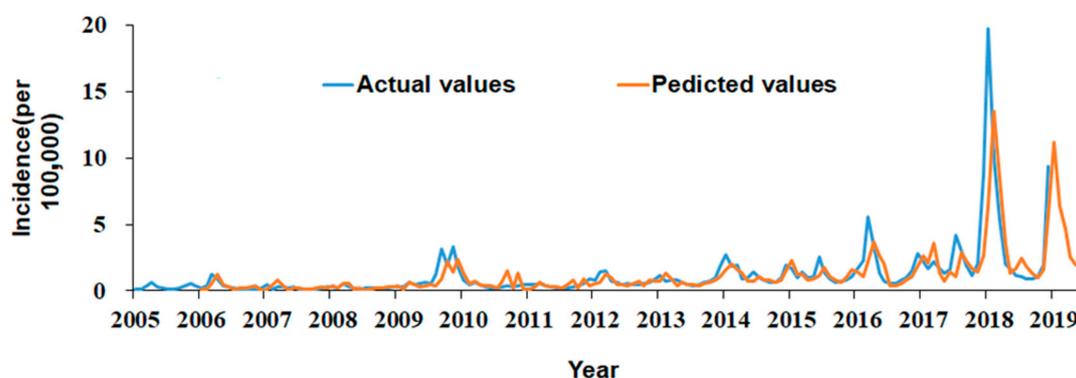


Figure 4. Comparison of actual and predicted incidence of influenza in mainland China.

Table 2. Comparison of predicted values and actual values form July to December 2018 (per 100,000 population).

Month	Actual Value	Predicted Value	Relative Error	95%CI	
				LCL	UCL
July	1.04	2.47	0.0137	-0.49	4.82
August	0.88	1.82	0.0106	-1.83	5.22
September	0.95	1.34	0.0042	-2.4	4.99
October	1.06	0.96	0.0010	-2.79	4.69
November	1.93	1.61	0.0017	-2.15	5.34
December	9.35	5.72	0.0039	1.86	9.36

LCL: lower confidence limit; UCL: upper confidence limit.

4. Discussion

Seasonal influenza is an acute respiratory infection caused by influenza viruses. Influenza surveillance to improve the influenza is the basis of influenza prevention and control. Global influenza surveillance has been conducted through the WHO’s Global Influenza Surveillance and Response System (GISRS) since 1952. The GISRS Network remains alert to timely recognize potential threats and minimize the impact of influenza epidemics and pandemics [21].

In many countries, more attention has been paid to influenza surveillance. The surveillance data types are usually used to establish a variety of influenza surveillance systems, including influenza-likely illness (ILI), acute respiratory infection (ARI), influenza cases, laboratory-confirmed influenza, Google

flu trends, protein sequences, etc. [5,10,22–25]. In this study, the current influenza surveillance system covers the following contents: (i) ILI, (ii) ARI, (iii) outbreak surveillance, and (iv) notifiable infectious disease surveillance. We used reportable contagious disease surveillance data to analyze the trend of influenza, and our results demonstrated that seasonal influenza is one of the web-based national notifiable infectious diseases in China, since 2003. The surveillance data of this study are of high accuracy by quality control, which ensures the authenticity of these results. Since 2003, the Chinese government has improved the surveillance system. When clinicians identify any case of seasonal influenza-A and influenza-B, since 2003, they report it through the web-based national notifiable infectious disease surveillance system within 24 hours. Epidemiologists evaluate the report rigorously, which is helpful to reduce the surveillance bias and enhance the data accuracy.

Influenza is affected by many biological, behavioral, and environmental factors, which lead to a seasonal variation. Previous studies have assumed that influenza is an annual spring or winter epidemic in some cities of China [26,27]. The data from the 14-year surveillance in this study are in agreement that influenza has seasonal variation in winter or spring. Seasonality exists with two peaks in winter and spring, consistent with well-documented peaks for influenza-A and influenza-B [25,26]. The reasons for seasonal epidemics may be related to factors such as a vast population, high residential density, and crowded living conditions, the variability of influenza viruses, diversity of geography, cold winter weather, low vaccination rate, etc. [27–29].

The multiple factors cause difficulties in modeling the influenza pandemic. Several approaches are applied to make these models. These approaches can be categorized as follows: time series models, compartmental modes, agent-based models, met population models, and approaches in meteorology. Time series analysis has the advantage of forecasting the incidence without focusing on specific risk factors. It uses the number of patients in the past as features to forecast the number of patients in the future as the response. The SARIMA model is performed over a time series in an automated fashion to maximize prediction accuracy. In addition, it takes both overall trends and seasonal changes into account and has been widely used for time series analysis. The SARIMA model typically assumes that future values in about three to six months can be predicted based on previously observed values [29]. Accordingly, we constructed the SARIMA $(1, 0, 0) \times (0, 1, 1)_{12}$ model to forecast influenza incidence. This model forecasted that the influenza incidence from July 2018 to December 2018 was similar to that of previous years, and there was also a seasonal variation during winter. Our results demonstrated that the predicted values matched the actual values well, supporting that the SARIMA model is effective in the prevention and control of influenza. It can capture trends and periodic changes.

5. Limitations

Several limitations should be noted in this study. First, only the SARIMA model was used and we assumed that there was a linear relationship between influenza incidence and its factors, such as exposure, susceptibility, access to care, etc. Many environmental and natural factors are dynamic, so the parameters of the SARIMA model should be periodically reassessed according to continuously updated data. Second, the surveillance data of this study cannot exclude surveillance bias in spite of quality control, which may affect our results to some extent. Third, we only used the data of all mainland China for prediction, analysis of subgroups (the South of China and the North of China) could be more reasonable. Fourth, we collected only monthly data, and weekly reporting could have better accuracy.

6. Conclusions

This work demonstrates that influenza occurred throughout the year with two peaks in winter and spring, in mainland China, which reminds us that influenza never goes away. Additional practical efforts should focus on reducing the burden of seasonal influenza. Our results also indicate that the SARIMA model can make short-term predictions of seasonal influenza effectively, and it is helpful to decision makers to allocate public health resources.

Author Contributions: P.W. conceived and designed the study; J.C. and M.R. performed the experiments and analyzed the data; S.X. reviewed and edited the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China (no. 81772281), Shandong province Taishan Scholar Program (ts201712067), and the Yantai Science and Technology Committee (2018XSCC051).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Feng, S.; Chiu, S.S.; Chan, E.L.Y.; Kwan, M.Y.W.; Wong, J.S.C.; Leung, C.; Chung Lau, Y.; Sullivan, S.G.; Malik Peiris, J.S.; Cowling, B.J. Effectiveness of influenza vaccination on influenza-associated hospitalisations over time among children in Hong Kong: A test-negative case-control study. *Lancet Respir. Med.* **2018**, *6*, 925–934. [[CrossRef](#)]
2. Huang, Q.S.; Bandaranayake, D.; Wood, T.; Newbern, E.C.; Seeds, R.; Ralston, J.; Waite, B.; Bissielo, A.; Prasad, N.; Todd, A.; et al. Risk Factors and Attack Rates of Seasonal Influenza Infection: Results of the Southern Hemisphere Influenza and Vaccine Effectiveness Research and Surveillance (SHIVERS) Seroepidemiologic Cohort Study. *J. Infect. Dis.* **2019**, *219*, 347–357. [[CrossRef](#)] [[PubMed](#)]
3. Moser, J.S.; Galindo-Fraga, A.; Ortiz-Hernández, A.A.; Gu, W.; Hunsberger, S.; Galán-Herrera, J.; Guerrero, M.L.; Ruiz-Palacios, G.M.; Beigel, J.H. Underweight, overweight, and obesity as independent risk factors for hospitalization in adults and children from influenza and other respiratory viruses. *Influenza Other Respir. Viruses* **2019**, *13*, 3–9. [[CrossRef](#)] [[PubMed](#)]
4. Mulpuru, S.; Li, L.; Ye, L.; Hatchette, T.; Andrew, M.K.; Ambrose, A.; Boivin, G.; Bowie, W.; Chit, A.; Dos Santos, G.; et al. Effectiveness of Influenza Vaccination on Hospitalizations and Risk Factors for Severe Outcomes in Hospitalized Patients With COPD. *Chest* **2019**, *155*, 69–78. [[CrossRef](#)] [[PubMed](#)]
5. Iuliano, A.D.; Roguski, K.M.; Chang, H.H.; Muscatello, D.J.; Palekar, R.; Tempia, S.; Cohen, C.; Gran, J.M.; Schanzer, D.; Cowling, B.J.; et al. Estimates of global seasonal influenza-associated respiratory mortality: A modelling study. *Lancet* **2018**, *391*, 1285–1300. [[CrossRef](#)]
6. Troeger, C.E.; Blacker, B.F.; Khalil, I.A.; Zimsen, S.R.M.; Albertson, S.B.; Abate, D.; Abdela, J.; Adhikari, T.B.; Aghayan, S.A.; Agrawal, S.; et al. Mortality, morbidity, and hospitalisations due to influenza lower respiratory tract infections, 2017: An analysis for the Global Burden of Disease Study 2017. *Lancet Respir. Med.* **2019**, *7*, 69–89. [[CrossRef](#)]
7. Fan, V.Y.; Jamison, D.T.; Summers, L.H. Pandemic risk: How large are the expected losses? *Bull. World Health Organ.* **2018**, *96*, 129–134. [[CrossRef](#)]
8. Chaves, S.S.; Aragon, D.; Bennett, N.; Cooper, T.; D’Mello, T.; Farley, M.; Fowler, B.; Hancock, E.; Kirley, P.D.; Lynfield, R.; et al. Patients hospitalized with laboratory-confirmed influenza during the 2010–2011 influenza season: Exploring disease severity by virus type and subtype. *J. Infect. Dis.* **2013**, *208*, 1305–1314. [[CrossRef](#)]
9. Organization, G.W.H. *Global Influenza Strategy 2019–2030*; World Health Organization: Geneva, Switzerland, 2019.
10. Song, X.; Xiao, J.; Deng, J.; Kang, Q.; Zhang, Y.; Xu, J. Time series analysis of influenza incidence in Chinese provinces from 2004 to 2011. *Medicine* **2016**, *95*, e3929. [[CrossRef](#)]
11. Zeng, Q.; Li, D.; Huang, G.; Xia, J.; Wang, X.; Zhang, Y.; Tang, W.; Zhou, H. Time series analysis of temporal trends in the pertussis incidence in Mainland China from 2005 to 2016. *Sci. Rep.* **2016**, *6*, 32367. [[CrossRef](#)]
12. Kane, M.J.; Price, N.; Scotch, M.; Rabinowitz, P. Comparison of ARIMA and Random Forest time series models for prediction of avian influenza H5N1 outbreaks. *BMC Bioinform.* **2014**, *15*, 276. [[CrossRef](#)] [[PubMed](#)]
13. Xu, Q.; Gel, Y.R.; Ramirez Ramirez, L.L.; Nezafati, K.; Zhang, Q.; Tsui, K. Forecasting influenza in Hong Kong with Google search queries and statistical model fusion. *PLoS ONE* **2017**, *12*, e176690. [[CrossRef](#)] [[PubMed](#)]
14. Wang, K.W.; Deng, C.; Li, J.P.; Zhang, Y.Y.; Li, X.Y.; Wu, M.C. Hybrid methodology for tuberculosis incidence time-series forecasting based on ARIMA and a NAR neural network. *Epidemiol. Infect.* **2017**, *145*, 1118–1129. [[CrossRef](#)] [[PubMed](#)]
15. Wang, H.; Tian, C.W.; Wang, W.M.; Luo, X.M. Time-series analysis of tuberculosis from 2005 to 2017 in China. *Epidemiol. Infect.* **2018**, *146*, 935–939. [[CrossRef](#)] [[PubMed](#)]

16. Sun, S.; Fu, C.; Cong, J.; Li, Y.; Xie, S.; Wang, P. Epidemiological features and trends of influenza incidence in mainland China: A population-based surveillance study from 2005 to 2015. *Int. J. Infect. Dis.* **2019**, *89*, 12–20. [[CrossRef](#)]
17. Li, L.; Liu, Y.; Wu, P.; Peng, Z.; Wang, X.; Chen, T.; Wong, J.; Yang, J.; Bond, H.S.; Wang, L.; et al. Influenza-associated excess respiratory mortality in China, 2010–2015: A population-based study. *Lancet Public Health* **2019**, *4*, e473–e481. [[CrossRef](#)]
18. The National Health Commission of the People’s Republic of China. Available online: <http://www.nhc.gov.cn/jkj/s3578/201904/050427ff32704a5db64f4ae1f6d57c6c.shtml> (accessed on 2 May 2019).
19. Makridakis, S.M.H. ARMA Models and the Box–Jenkins Methodology. *J. Forecast.* **1997**, *16*, 147–163. [[CrossRef](#)]
20. Ziegler, T.; Mamahit, A.; Cox, N.J. 65 years of influenza surveillance by a World Health Organization-coordinated global network. *Influenza Other Resp. Viruses* **2018**, *12*, 558–565. [[CrossRef](#)]
21. Tizzoni, M.; Bajardi, P.; Poletto, C.; Ramasco, J.J.; Balcan, D.; Goncalves, B.; Perra, N.; Colizza, V.; Vespignani, A. Real-time numerical forecast of global epidemic spreading: Case study of 2009 A/H1N1pdm. *BMC Med.* **2012**, *10*, 165. [[CrossRef](#)]
22. Nishiura, H. Real-time forecasting of an epidemic using a discrete time stochastic model: A case study of pandemic influenza (H1N1-2009). *Biomed. Eng. Online* **2011**, *10*, 15. [[CrossRef](#)]
23. He, Z.; Tao, H. Epidemiology and ARIMA model of positive-rate of influenza viruses among children in Wuhan, China: A nine-year retrospective study. *Int. J. Infect. Dis.* **2018**, *74*, 61–70. [[CrossRef](#)] [[PubMed](#)]
24. Chrysostomou, C.; Partaourides, H.; Seker, H. Prediction of Influenza A virus infections in humans using an Artificial Neural Network learning approach. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* **2017**, *2017*, 1186–1189. [[PubMed](#)]
25. Ye, C.; Zhu, W.; Yu, J.; Li, Z.; Zhang, Y.; Wang, Y.; Gu, H.; Zou, W.; Hao, L.; Hu, W. Understanding the complex seasonality of seasonal influenza A and B virus transmission: Evidence from six years of surveillance data in Shanghai, China. *Int. J. Infect. Dis.* **2019**, *81*, 57–65. [[CrossRef](#)] [[PubMed](#)]
26. Cheng, X.; Tan, Y.; He, M.; Lam, T.T.; Lu, X.; Viboud, C.; He, J.; Zhang, S.; Lu, J.; Wu, C.; et al. Epidemiological Dynamics and Phylogeography of Influenza Virus in Southern China. *J. Infect. Dis.* **2013**, *207*, 106–114. [[CrossRef](#)]
27. Ye, C.; Zhu, W.; Yu, J.; Li, Z.; Hu, W.; Hao, L.; Wang, Y.; Xu, H.; Sun, Q.; Zhao, G. Low coverage rate and awareness of influenza vaccine among older people in Shanghai, China: A cross-sectional study. *Hum. Vaccin. Immunother.* **2018**, *14*, 1–7. [[CrossRef](#)]
28. McLeod, C.; Adunuri, N.; Booth, R. Risk factors and mitigation of influenza among Indigenous children in Australia, Canada, United States, and New Zealand: A scoping review. *Perspect. Public Health* **2019**, *139*, 228–235. [[CrossRef](#)]
29. Nsoesie, E.O.; Brownstein, J.S.; Ramakrishnan, N.; Marathe, M.V. A systematic review of studies on forecasting the dynamics of influenza outbreaks. *Influenza Other Resp.* **2014**, *8*, 309–316. [[CrossRef](#)]

